IFAU
Institute for Evaluation of Labour Market and Education Policy

A Distributional Framework for Matched Employer Employee Data

Stéphane Bonhomme
Thibaut Lamadon
Elena Manresa

WORKING PAPER 2017:24

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

A Distributional Framework for Matched Employer Employee Data*

Stéphane Bonhomme
University of Chicago

Thibaut Lamadon
University of Chicago

Elena Manresa
New York University

December 10, 2017

Abstract

We propose a framework to identify and estimate earnings distributions and worker composition on matched panel data, allowing for two-sided worker-firm unobserved heterogeneity. We introduce two models: a static model that allows for nonlinear interactions between workers and firms, and a dynamic model that allows in addition for Markovian earnings dynamics and endogenous mobility. We establish identification in short panels, and develop tractable two-step estimators where firms are classified into heterogeneous classes in a first step. Applying our method to Swedish administrative data, we find that log-earnings are approximately additive in worker and firm heterogeneity, with a strong association between workers and firms, and a small relative contribution of firm heterogeneity to earnings dispersion. In addition, we document that wages have a direct effect on mobility, and that, beyond their dependence on the current firm, earnings after a job move also depend on the past firm.

JEL codes: J31, J62, C23.

Keywords: two-sided heterogeneity, bipartite networks, matched employer employee data, sorting, job mobility.

---

*First draft: November 2014. We thank our discussants Pat Kline and Chris Hansen. We also thank Joe Altonji, Bryan Graham, Rafael Lopes de Melo, Costas Meghir, Luigi Pistaferri, Jean-Marc Robin, Adam Rosen, Raffaele Saggio, Uta Schoenberg, Robert Shimer, and Chris Taber for useful comments. We thank seminar audiences at the 2015 NBER Summer conferences, the 2015 SED conference, the 2015 Interactions conference in Chicago, and audiences at various places for comments. We thank the IFAU for access to, and Lisa Laun for help with, the Swedish administrative data. The authors acknowledge support from the NSF grant number SES-1658920. The usual disclaimer applies.

Contact: sbonhomme@uchicago.edu; lamadon@uchicago.edu; em1849@nyu.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 and correlations between worker and firm heterogeneity. The AKM method is widely used in labor economics and outside.[1] A second literature tackles similar issues from a structural perspective, by developing and estimating fully specified theoretical models of sorting on the labor market.[2]

Reconciling these reduced-form and structural literatures has proven difficult. 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 interactions between worker and firm attributes restricts complementarity patterns in earnings. However, since Gary Becker's work, numerous theories have emphasized the link between complementarity and sorting (Shimer and Smith, 2000, Eeckhout and Kircher, 2011). In addition, the AKM model is static, in the sense that worker mobility does not depend on earnings realizations conditional on worker and firm heterogeneity, and that earnings after a job move do not depend on the previous firm's attributes. These assumptions 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 dozens of thousands of firms in the presence of both firm and worker unobserved heterogeneity?

[1] Applications of the method to earnings data include Gruetter and Lalive (2009), Mendes et al. (2010), Woodcock (2008), Card et al. (2013), Goldschmidt and Schmieder (2015), Song et al. (2015), and Sorkin (2016), among others. The AKM estimator 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 (e.g., Kramarz et al., 2015, Jackson, 2013, Finkelstein et al., 2016).

[2] Many structural models proposed in the literature build on Becker (1973). Examples are De Melo (2009), Lise et al. (2008), Bagger et al. (2014), Hagedorn et al. (2014), Lamadon et al. (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.

And how informative are functional form assumptions in these often tightly parameterized models?

In this paper we introduce an empirical framework with two-sided unobserved heterogeneity that nests a range of theoretical mechanisms emphasized in the literature. While allowing for rich patterns of complementarities, sorting, and dynamics, the framework preserves parsimony using a dimension reduction technique to model firm heterogeneity. We propose two models, static and dynamic, which allow for interaction effects between worker and firm heterogeneity. In the dynamic model we let job mobility depend on earnings realizations in addition to worker and firm attributes, and we allow earnings after a job move to depend on attributes of the previous firm beyond those of the current one. Dynamic persistence is specified as first-order Markov.

We provide conditions for identification under discrete worker heterogeneity. The primary source of identification is given by job movers. For the static model we rely on two periods, while we use four periods to identify the dynamic model. The ability of our method to deal with short panels is important, since assuming time-invariant heterogeneity over long periods may be unattractive. 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 complementarities, there must be variation in the latent types of job movers between different firms. In particular, worker-firm interaction effects would not be identified if workers' allocation to firms was fully random.

We define the relevant level of firm unobserved heterogeneity as the *class* of a firm. In principle, these classes could be the firms themselves. However, in typical matched employer employee data sets the number of job movers per firm tends to be small, which creates an incidental parameter bias of a similar nature as in fixed-length panel data.[4] In such environments, reducing the number of classes can alleviate small-sample biases. We use a k-means clustering estimator to classify firms based on how similar their earnings distributions are. The classification may also be based on mobility patterns or longitudinal earnings information, and it can be modified to incorporate firm characteristics such as value added. We establish the asymptotic consistency of the classification under discrete firm heterogeneity by providing conditions under which the main theorems in Bonhomme and Manresa (2015) hold. Under these conditions, estimation error in classification does not affect inference on parameters estimated

[4]See Abowd et al. (2004) and Andrews et al. (2008, 2012) for illustrations of incidental parameter bias in fixed-effects regressions.

in a second step. This provides a formal justification for clustering in our models.[5]

We use a two-step approach for estimation. In the *classification* step we group firms into classes using k-means clustering, and in the *estimation* step we estimate the model by allowing for firm class heterogeneity. We model worker types as discrete and allow for unrestricted interactions between worker and firm heterogeneity. We use maximum likelihood for estimation. We verify in simulations that our estimator performs well in data sets similar to the one of our application. In addition, we also confirm the ability of our estimator to recover wage functions in data sets generated according to the theoretical model of Shimer and Smith (2000), extended to allow for on-the-job search, under both positive and negative assortative matching. Finally, we show that, when the model is specified as a static or dynamic extension of the AKM regression model that allows for interaction effects between firms and workers, parameters can be estimated using simple linear instrumental variables and covariance-based estimators conditional on the firm classes. We analyze such regression-based estimators, and use them to show the robustness of our main results.

We take our approach to Swedish matched employer employee panel data on the 2002-2004 period. The estimates of our static model imply that an additive model of log-earnings provides a good first-order approximation to the variance structure of log-earnings. However, in both the static and dynamic models we find the presence of stronger complementarities between firms and lower-type workers than with workers of higher types. We show that those are quantitatively relevant in a reallocation exercise where we shut down the association between firm and worker heterogeneity and assess the distributional impacts of the reallocation.

We find that firm heterogeneity (net of the effect of worker composition) accounts for less than 5% of the variance of log-earnings, most of the variance being explained by worker heterogeneity. In addition, we find a substantial remaining contribution due to the strong association between worker and firm heterogeneity, with a correlation ranging between 40% and 50%. Our estimates on Swedish data suggest a stronger association between worker and firm heterogeneity, and a smaller role for firm heterogeneity, compared to many empirical estimates in the literature. We estimate a battery of additional specifications that show the robustness of our estimates and highlight the ability of our method to deal with the low mobility rates in the

[5]Similarly as in most of the literature on discrete estimation, this result is derived under the assumption that the population of firms consists of a finite number of classes. In Bonhomme et al. (2017) we consider a more general setting where the discrete modeling is viewed as an approximation to an underlying, possibly continuous, distribution of firm unobserved heterogeneity, and we provide consistency results and rates of convergence. In this alternative asymptotic framework, estimation error in the classification generally affects post-classification inference. These results provide further justification for the use of clustering methods.

data and the resulting biases.

While they are in line with the cross-sectional variance decomposition implied by our static model, the estimates of our dynamic model on the 2001-2005 period shed light on several mechanisms that have been emphasized in the structural literature. In particular, we find that low earnings realizations, conditional on worker and firm heterogeneity, tend to increase the propensity to change job, hence challenging the strict exogeneity assumption often made in the literature. We also find evidence of an effect of the past firm's class on current earnings, conditional on the current firm's class.

Literature and outline. The methods we propose contribute to a large literature on the identification and estimation of models with latent heterogeneity. 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). The k-means clustering algorithm we use to classify firms is widely used in a number of fields, and efficient computational routines are available (Steinley, 2006). Here we apply such an approach to models with two-sided heterogeneity.

Nonparametric identification and estimation of finite mixtures, which we use to specify worker heterogeneity, have been extensively studied, see for example Hall and Zhou (2003), Hu (2008), Henry et al. (2014), Levine et al. (2011), or Bonhomme et al. (2014). Our framework can also accommodate continuous worker heterogeneity. 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 et al., 2003, Airoldi et al., 2008).

Compared to this previous work, we rely on a hybrid "one-sided random-effects" approach that models 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 class membership will be accurately estimated. In contrast, the number of observations for a given worker is typically small. This approach can alleviate the incidental parameter bias of fixed-effects estimators, particularly in short panels. It also offers a tractable way of allowing for complementarities and dynamics.

Lastly, two recent innovative contributions rely on estimation methods for models with latent heterogeneity to study questions related to worker-firm sorting on the labor market. Abowd et al. (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 a two-sided random-effects approach for estimation. Hagedorn et al. (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 exploits the implications of a structural model to identify firm heterogeneity.[6] In contrast, while our framework nests a number of theoretical models of wages and mobility it is not tied to a specific structural model. In addition, we provide conditions for identification and consistency in short panels. Bonhomme (2017) reviews existing econometric methods for bipartite network data.

The outline of the paper is as follows. In Section 2 we present the framework. In Sections 3 and 4 we study identification and estimation. In Sections 5 and 6 we show empirical results based on the static and dynamic models. Lastly, we conclude in Section 7. A supplementary appendix with additional results can be found on the authors' web pages.

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 $m_{it} = 1$.

Heterogeneity across firms is characterized by their *class*. We denote as k_{it} in $\{1, ..., K\}$ the class of firm j_{it}. Classes form a partition of the set of firms into K classes, and k_{it} is a shorthand for $k(j_{it})$.[7] There may 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 describe a method to consistently estimate the latent classes k_{it} from the data, under the assumption that population heterogeneity has a finite number of points of support.

Workers are also heterogeneous, and we denote the *type* of worker i as α_i. These types can 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}.[8]

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 $(j_{i1}, m_{i1}, ..., j_{i,T-1}, m_{i,T-1}, j_{iT})$,

[6] Also related is Bartolucci et al. (2015), who rank firms according to their profits, and find evidence of positive sorting using Italian data.

[7] In other words, $k : \{1, ..., J\} \mapsto \{1, ..., K\}$ maps firm j to firm class $k(j)$.

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

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.[9]

In this framework we will be interested in recovering the conditional distribution of log-earnings for a worker of type α in a firm of class k, and the proportion of type-α workers in a class-k firm. Conditional earnings distributions will be informative about the form of the earnings function, in particular complementarities, while type proportions will be informative about sorting patterns. In addition, within our framework we will be able to document dynamic aspects.

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} up to period t.

2.1 Static model

There are 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 are also strictly exogenous. Second, log-earnings after a job move are not allowed to depend on previous firm classes or previous earnings, conditional on the worker type and the new firm's class.

Before stating the assumptions formally let us describe the model's timing. 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 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 (while also varying with k'). 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}.

[9]Incorporating an extensive employment margin within our framework could be done by adding a "non-employment" firm class $k_{it} = 0$.

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}.

Formally the two main assumptions are thus as follows.

Assumption 1. *(static model)*

(i) (mobility determinants) m_{it}, $k_{i,t+1}$ *and* $X_{i,t+1}$ *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 $\mathbb{E}\left(\varepsilon_{it} \,|\, \alpha_i, k_i^T, m_i^T, X_i^T\right) = 0$. This model simplifies to the one of Abowd et al. (1999) in the absence of interaction effects, i.e. when $b_t(k) = 1$, and firms j_{it} and classes k_{it} coincide.[10]

2.2 Dynamic model

There are two main differences between the dynamic and static models. 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}. Second, 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. As we discuss in the next subsection, allowing for these features is important in order to nest a number of structural models of wage and employment dynamics that have been proposed in the literature. Formally we make the following assumptions.

Assumption 2. *(dynamic model)*

(i) (mobility determinants) m_{it}, $k_{i,t+1}$ *and* $X_{i,t+1}$ *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}.

[10] While both parts in Assumption 1 are needed to identify the full model, 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.

(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.[11] 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 not on 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 (i.e., $m_{i,t-1}=1$):

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

where $\mathbb{E}\left(v_{it}\,|\,\alpha_i, k_i^t, m_i^{t-1}, Y_i^{t-1}, X_i^t\right) = 0$. Here log-earnings after a job move may depend on earnings and firm class in the previous job.

2.3 Links with theoretical models

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

Models where the 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, say ε_t, where ε_t does not affect mobility decisions. This structure is compatible for instance with wage posting models (as in Burdett and Mortensen, 1998, Delacroix and Shi, 2006, or Shimer, 2005), where the wage paid to a worker does not have any history dependence and ε_t is classical measurement error or an i.i.d. match effect realized after mobility. This means that, while allowing for rich mobility and earnings patterns, such 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 under certain conditions on the worker's

[11] Assumption 2 (i) allows $X_{i,t+1}$ to 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 could allow for predetermined individual characteristics, such as job tenure.

outside option. For example, this is the case in Shimer and Smith (2000), where the outside option is unemployment since workers always go through unemployment before finding a new job; see also Hagedorn et al. (2014). In such sorting models, specifying the wage function in a way that allows for interactions between worker types and firm classes is key, since earnings may be non-monotonic in firm productivity and different workers rank identical firms differently. Our static model can accommodate both features.

Models with Markovian match effects and state dependence. In dynamic models workers often move based on the realization of the match effect ε_t, which is allowed to be serially correlated. This is compatible with the assumptions of our dynamic model provided ε_t is first-order Markov, see Assumption 2. For example, in a wage posting model with match-specific heterogeneity workers may observe potential wages before deciding whether or not to move. While incompatible with Assumption 1, this is perfectly consistent with the dynamic model's assumptions provided mobility, the new firm's class, and the new wage are *jointly* first-order Markov.

To see this formally, consider an agent in period t with firm class k_t and wage Y_t. She draws an offer, (Y_{t+1}^*, k_{t+1}^*), jointly with a potential wage $\widetilde{Y}_{t+1}$ she would get should she decide not to move, all of which may depend on the current wage Y_t, firm class k_t, and type α. The decision to move is based on all this information. The realized firm class is then either $k_{t+1} = k_t$ with associated wage $\widetilde{Y}_{t+1}$, or $k_{t+1} = k_{t+1}^*$ with wage Y_{t+1}^*, depending on the outcome of the mobility decision. Assumption 2 is satisfied in this model, since the effective conditioning set is (α, Y_t, k_t).

Our dynamic model encompasses other mechanisms, such as endogenous search intensity along the lines of Bagger and Lentz (2014), where the previous wage may affect offers through an endogenous search decision. It also encompasses sequential contracting as in Postel-Vinay and Robin (2002), where the Bertrand competition is captured by the fact that the outside offer Y_{t+1}^* and the firm's wage counteroffer $\widetilde{Y}_{t+1}$ may depend on each other, and (α, Y_t, k_t) are sufficient statistics for the history.[12]

In the setting of Assumption 2 the wage conditional on moving depends on the past wage and the past firm. Our dynamic model allows for these selection effects.[13] However, recovering underlying primitives such as distributions of wage offers Y_{t+1}^* would require making additional

[12]Related examples are contract posting models (as in Burdett and Coles, 2003, or Shi, 2008), where the optimal contract is a tenure contract.

[13]Alternatively, ε_t may be thought of as a first-order Markov, one-dimensional human capital accumulation process.

assumptions. In the absence of those, our framework allows one to identify the distributions of *realized* wages for job movers and stayers, as a function of worker and firm heterogeneity.

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.

Outside our framework. However, non-Markovian earnings structures will violate the assumptions of our dynamic model. This will happen if the structural model allows for permanent-transitory earnings dynamics conditional on worker types, as in Hall and Mishkin (1982) for example. This will also happen in models that combine a sequential contracting mechanism (à la Postel-Vinay and Robin, 2002) with a match-specific effect. In this case agents need to keep track of both the match effect and the bargaining position, so the one-to-one mapping between earnings and the value to the worker no longer holds, making mobility decisions potentially dependent on the whole history of wages. Such environments are not nested in a framework such as ours, which only allows for uni-dimensional time-varying effects ε_t.

3 Identification

In this section we provide conditions for identification of earnings distributions for all worker types and firm classes, and worker type distributions for all firm classes, given two periods in the static model and four periods in the dynamic model. The analysis is conditional on a partition of firms into classes. In the next section we will show how to consistently estimate class membership $k(j)$, for each firm j.

3.1 Intuition in an interactive regression model

We first provide an intuition for identification of complementarities in a stationary specification of the interactive regression model of equation (1) with $T = 2$ periods, where we abstract from covariates. 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 from k to k'.[14] Log-earnings in each period are given by:

$$Y_{i1} = a(k) + b(k)\alpha_i + \varepsilon_{i1}, \quad Y_{i2} = a(k') + b(k')\alpha_i + \varepsilon_{i2}. \tag{3}$$

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.[15]

Consider now job movers from a firm in class k' to a firm in class k. Their log-earnings are given by:

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

By comparing differences in log-earnings in each class between these two subpopulations of job movers, we obtain:

$$\frac{b(k')}{b(k)} = \frac{\mathbb{E}_{kk'}(Y_{i2}) - \mathbb{E}_{k'k}(Y_{i1})}{\mathbb{E}_{kk'}(Y_{i1}) - \mathbb{E}_{k'k}(Y_{i2})}, \tag{4}$$

provided that the following holds:

$$\mathbb{E}_{kk'}(\alpha_i) \neq \mathbb{E}_{k'k}(\alpha_i), \tag{5}$$

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 (5) holds, then $b(k')/b(k)$ is identified from mean restrictions on job movers between k and k'. Conversely, if (5) does not hold then $b(k')/b(k)$ is not identified based on those restrictions. Note that (5) requires the types of workers moving from k to k' and from k' to k to differ. If $b(k') + b(k) \neq 0$, (5) is equivalent to:

$$\mathbb{E}_{kk'}\left(Y_{i1} + Y_{i2}\right) \neq \mathbb{E}_{k'k}\left(Y_{i1} + Y_{i2}\right), \tag{6}$$

so it can be empirically tested (under the maintained hypothesis of exogenous mobility).

An implication is that, when (5) does not hold, additivity of log-earnings in worker and firm attributes (that is, the $b(k)$'s being equal in all firms) is not testable based on mean restrictions. Graphical illustrations of mean log-earnings before and after a job move event, as introduced in Card et al. (2013), are a popular way of providing suggestive evidence for additivity. Our analysis implies that documenting symmetric wage gains and losses is not

[14] This intuitively means that this analysis will be relevant for data sets with a sufficient number of workers moving between firm classes. Our grouping of firms into classes is motivated by the incidental parameter bias due to low mobility. We will return to this issue in the estimation section.

[15] Model (3) 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).

sufficient to demonstrate that wage functions are additive. As an example, in the theoretical model of Shimer and Smith (2000) wage gains and losses are symmetric around a job move, yet the wage function can feature *any* degree of complementarity between worker types and firm classes. In Figure E3 in Appendix E we illustrate this point by simulating the Shimer Smith model under positive assortative matching, and showing the corresponding event study graph around job mobility.

A main goal of this paper is to establish that, by fully exploiting earnings information before and after a job move, complementarities can be identified and consistently estimated under suitable rank conditions such as (5). In fact, such conditions will be satisfied in an extension of the model of Shimer and Smith (2000) with on-the-job search. In Appendix C we evaluate the performance of our estimator to recover the contributions of worker and firm heterogeneity to earnings dispersion, when the data generating process follows this theoretical model. We show that our estimator recovers the wage functions and contributions of firms and workers to earnings dispersion, both under positive and negative assortative matching.

3.2 Identification in finite mixture models

In this subsection we consider the 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 outline an extension to continuously distributed worker types. We start by presenting the main identifying equations.

3.2.1 Identifying equations

Static model on two periods. We first consider the static model on $T = 2$ periods, which 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 $F^m_{k'\alpha}(y_2)$ 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 exogenous covariates X_{i1} and X_{i2}, although we omit the conditioning for conciseness.[16]

[16] 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 nonparametrically link

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

$$\Pr\left[Y_{i1}\leq y_1, Y_{i2}\leq y_2 \,|\, k_{i1}=k, k_{i2}=k', m_{i1}=1\right] \;=\; \int F_{k\alpha}(y_1)F^m_{k'\alpha}(y_2)p_{kk'}(\alpha)d\alpha. \tag{7}$$

To see why (7) holds, note that Y_{i1} is independent of (k_{i2}, m_{i1}) conditional on (α_i, 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}, k_{i1}) conditional on $(\alpha_i, k_{i2}, m_{i1} = 1)$. 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:

$$\Pr\left[Y_{i1} \leq y_1 \,|\, k_{i1} = k\right] \;=\; \int F_{k\alpha}(y_1)q_k(\alpha)d\alpha. \tag{8}$$

The parameters in (7) and (8) 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 cross-sectional sorting. Moreover, from the $q_k(\alpha)$, $p_{kk'}(\alpha)$, and data on transitions between classes, one can recover estimates of type-specific transition probabilities between classes, which are informative about dynamic sorting patterns. We will report estimates of all these quantities in the empirical analysis.

Dynamic model on four periods. In 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 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. For conciseness we again omit the conditioning on covariates.

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

type probabilities across covariates values, due to a labeling indeterminacy. This issue echoes the impossibility to identify the coefficients of time-invariant regressors in fixed-effects panel data regressions.

$$\Pr\left[Y_{i1}\leq y_1, Y_{i4}\leq y_4 \,|\, Y_{i2}{=}y_2, Y_{i3}{=}y_3, k_{i1}{=}k_{i2}{=}k, k_{i3}{=}k_{i4}{=}k', m_{i1}{=}0, m_{i2}{=}1, m_{i3}{=}0\right]$$
$$= \int G^f_{y_2,k\alpha}(y_1) G^b_{y_3,k'\alpha}(y_4) p_{y_2y_3,kk'}(\alpha) d\alpha. \qquad (9)$$

Equation (9) is a consequence of Assumption 2, which is a first-order Markov assumption on the process $(Y_{it}, k_{it}, m_{i,t-1})$, where in addition m_{it} can only depend on Y_{it} and k_{it} but not on $m_{i,t-1}$. In particular, by Assumption 2 (ii), Y_{i4} is independent of past mobility, firm classes, and earnings, conditional on $(\alpha_i, Y_{i3}, k_{i4}, k_{i3}, m_{i3})$. Similarly, Y_{i1} can be shown to be independent of future classes, earnings and mobility conditional on $(\alpha_i, Y_{i2}, k_{i1}, k_{i2}, m_{i1})$.[17]

In addition, here $F_{k\alpha}$ denotes 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, $m_{i1} = 0$), while $q_k(\alpha)$ denotes the distribution of α_i for these workers. The joint cdf of log-earnings in periods 1 and 2 is:

$$\Pr\left[Y_{i1}\leq y_1, Y_{i2}\leq y_2 \,|\, k_{i1}=k_{i2}=k, m_{i1}=0\right] = \int G^f_{y_2,k\alpha}(y_1) F_{k\alpha}(y_2) q_k(\alpha) d\alpha. \qquad (10)$$

The mathematical structure of (9) is analogous to that of (7). This is useful to analyze the static and dynamic models using similar methods. 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.

3.2.2 Results under discrete worker types

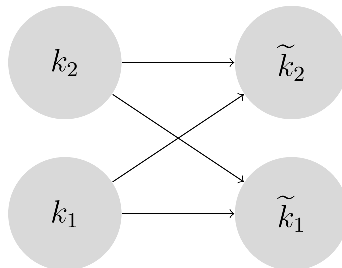
We start by considering the static model on two periods. Since the dynamic model has a similar mathematical structure the identification arguments will be closely related. Let L be the number of points of support of worker types, and let us denote the types as $\alpha_i \in \{1, ..., L\}$. We assume that L is known.[18] All distributions below may be conditional on (X_{i1}, X_{i2}), although we omit the conditioning for conciseness.

In this finite mixture model, (7) and (8) 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 provide conditions under which these quantities are identified. We start with a definition.

[17] To see this, note that, by Assumption 2 (i), Y_{i1} is independent of (m_{i2}, k_{i3}) conditional on $(\alpha_i, Y_{i2}, k_{i1}, k_{i2}, m_{i1})$; by Assumption 2 (ii), Y_{i1} is independent of Y_{i3} conditional on $(\alpha_i, Y_{i2}, k_{i1}, k_{i2}, k_{i3}, m_{i1}, m_{i2})$; and, by Assumption 2 (i), Y_{i1} is independent of (m_{i3}, k_{i4}) conditional on $(\alpha_i, Y_{i2}, Y_{i3}, k_{i1}, k_{i2}, k_{i3}, m_{i1}, m_{i2})$.

[18] 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 on the number of types. In the application we will check sensitivity by varying L (and also the number K of firm classes).

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



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, static)*
(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 $a(1), ..., a(L)$ 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) There exist finite sets of M values for y_1 and y_2 such that, for all r in $\{1, ..., R\}$, the matrices $A(k_r, \widetilde{k}_r)$ and $A(k_r, \widetilde{k}_{r+1})$ have rank L, where:

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

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 graph connectedness in AKM (Abowd et al., 2002). As in AKM, in our setup identification will fail in the presence of completely segmented labor markets where firms are not connected between markets via job moves. One difference with AKM is that, in our nonlinear setting, 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 non-random mobility, as it fails when $p_{kk'}(\alpha)$ does not depend on (k,k'). Also, part (i) fails when $p_{kk'}(\alpha)$ is symmetric in (k,k'). This situation arises in the model of Shimer and Smith (2000) in the absence of on-the-job search, as we discuss in Appendix C. 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 regression model studied above, see (6). In the empirical analysis we will provide evidence of such asymmetry.[19]

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 $F^m_{\widetilde{k}_r,1}$, ..., $F^m_{\widetilde{k}_r,L}$.

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 an arbitrary and irrelevant choice of labeling 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 are observed. Then, up to labeling 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, for the same labeling.*

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 are observed. Then the type proportions $q_k(\alpha)$ are identified for the same labeling 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 (7) and (9) on the one hand, and (8) and (10) on the other hand, we obtain the following

[19] The requirements on cycles can be relaxed, at the cost of loosing point-identification of some of the quantities of interest. In Supplementary Appendix S1 we illustrate this 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.

corollary to Theorem 1 and Corollary 1. The required assumptions, particularly on the existence of cycles, are more stringent than in the static case.

Corollary 2. *Let $T = 4$. Consider the joint distribution of log-earnings of job movers between periods 2 and 3. 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 complete formulation. Suppose that firm classes are observed. Then, up to labeling 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, type-specific transition probabilities between firm classes are identified.

3.3 Continuous worker types

While our empirical results will be based on models with discrete worker heterogeneity, our framework can accommodate the presence of continuous worker types. An example is the regression model (1), which is of interest in its own right. In Supplementary Appendix S3 we describe this model in detail, as well as its dynamic extension (2), and we provide conditions for their identification. In particular, we exploit the fact that mean restrictions are linear in parameters to derive simple mean and covariance restrictions. In turn, those restrictions lead to convenient estimators, which we also describe in detail in the supplement. Hence, in such regression models that feature complementarities and possibly dynamics, the simplicity of the AKM regression approach is preserved. We will estimate such regression models as robustness checks for our main empirical results.[20]

In more general models with continuous worker types, identification can be established under conditions that have been used in the literature on nonlinear measurement error models, notably Hu and Schennach (2008) and Hu and Shum (2012). In Supplementary Appendix S1 we outline an identification argument in the static model when α is continuously distributed. Although we do not pursue this route in this paper, it would be interesting to adapt our estimation strategy to estimate mixture models with continuously distributed α's.

[20] In addition, note that (4) holds irrespective of the serial dependence properties of ε_{it}. In regression models it is thus not necessary to assume that the log-earnings of job movers are serially independent conditional on worker type and firm classes, in order to identify the $a(k)$'s, $b(k)$'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 of job movers, such as independence between ε_{i1} and ε_{i2} when $T = 2$.

4 Estimation

In the previous section we have provided conditions under which earnings distributions are 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 complementarities, dynamics, and two-sided heterogeneity may be impractical due to the incidental parameter biases caused by the large number of firm-specific parameters that are solely identified from job movements. For this reason, we use a dimension reduction method to partition firms into classes. We now describe a two-step grouped fixed-effects approach, where we classify firms in a first step and estimate earnings and mobility parameters in a second step.

4.1 Recovering firm classes using k-means 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. For example, in (8) the first period's distribution of log-earnings in firm j does not depend on j beyond its dependence on firm class $k = k(j)$:

$$\Pr\left[Y_{i1} \leq y_1 \,|\, j_{i1} = j\right] = \int F_{k\alpha}(y_1) q_k(\alpha) d\alpha, \tag{11}$$

where the left-hand side thus only depends on $k = k(j)$. This observation motivates classifying firms into classes in terms of their earnings distributions, as we now explain.

We propose partitioning the J firms in the sample into classes by solving the following weighted k-means problem:

$$\min_{k(1),...,k(J),H_1,...,H_K} \sum_{j=1}^{J} n_j \int \left(\widehat{F}_j(y) - H_{k(j)}\left(y\right)\right)^2 d\mu(y), \tag{12}$$

where $\widehat{F}_j$ denotes the empirical cdf of log-earnings in firm j, n_j is the number of workers in firm j, μ is a discrete or continuous measure, $k(1), ..., k(J)$ denotes a partition of firms into K classes, and $H_1, ..., H_K$ are cdfs. We minimize (12) with respect to all possible partitions and to class-specific cdfs. While global minima in k-means may be challenging to compute, k-means algorithms are widely used in many fields and efficient heuristic computational methods have been developed (e.g., Steinley, 2006).

Through the classification in (12) we estimate 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 we will correlate the estimated classes *ex-post* to firm observables.[21]

To provide a formal justification for the classification, in Appendix B 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 both the number of firms and the number of workers per firm tend to infinity. Using a result from Bonhomme and Manresa (2015) we show that estimated firm classes, $\widehat{k}(j)$, converge uniformly to the population ones up to an arbitrary labeling as the sample size grows. As a result, the asymptotic distribution of parameter estimates in the second step (which are the main quantities of interest in this paper, see the next subsection) is not affected by the estimation of firm classes.

Using other moments for classification. Consistency of the classification still holds, under analogous assumptions, if instead of empirical cdfs other moments of the log-earnings distribution are used in k-means, provided these moments are informative about firm heterogeneity. For example one could use firm-specific means and variances. A potential issue with identifying firm classes from cross-sectional observations only is that two cross-sectional earnings distributions might be identical between two firms of different classes, if for example one offers a higher earnings schedule but has low-type workers, and the other one offers a lower earnings schedule but has high-type workers. This possibility has been emphasized in the theoretical sorting literature (Eeckhout and Kircher, 2011). It is reflected in the violation of one of the assumptions for classification consistency (see Assumption B2 (*iii*) in Appendix B). 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. For this reason, adding other moments to classify firms, beyond cross-sectional earnings moments, can be beneficial. In Supplementary Appendix S1 we outline a bi-clustering method which fully exploits longitudinal information on earnings and mobility to classify firms. A different approach is to use other information on the firm. In the empirical analysis we will refine our earnings-based classification using firm value added

[21]In longer panels, the clustering method could be generalized to account for time-varying classes, allowing one to document how the evolution of the classes relates to time variation in observables such as firm size. We outline this extension in Supplementary Appendix S1.

and mobility measures.

Properties under continuous firm heterogeneity. The asymptotic justification for k-means-based classification does not necessarily hinge on discreteness. Indeed, second-step parameters estimated from the classification are still consistent even when firm heterogeneity is continuous in the population, although asymptotically exact classification no longer holds. In Bonhomme et al. (2017) we study asymptotic properties of k-means clustering and two-step methods in settings where the K classes are viewed as approximating a possibly continuous heterogeneity structure. This analysis provides a justification for such discrete methods in continuous settings too.

4.2 Two-step grouped fixed-effects estimation

Our two-step grouped fixed-effects estimation strategy is as follows. In the first step (classification) we estimate the firm classes $\widehat{k}(j)$ for all firms j in the sample, by solving a classification problem such as (12). In the second step (estimation) 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.

Static model. To describe the estimation step we consider a specification where workers belong to L latent types, and the model is parametric given worker and firm heterogeneity. We focus on a two-period version of the static model and a four-period version of the dynamic model, both of which we will estimate on Swedish data.[22] In the static case we let $f_{k\alpha}(y;\theta_f)$ (first-period earnings), $f^m_{k\alpha}(y;\theta_{f^m})$ (second-period earnings for job movers), $q_k(\alpha;\theta_q)$ (worker-type proportions), and $p_{kk'}(\alpha;\theta_p)$ (worker-type proportions for job movers) be indexed by parameter vectors $\theta_f, \theta_{f^m}, \theta_q, \theta_p$. In our baseline specification we will let both earnings densities be log-normal with (k,α)-specific means and variances. That is, means and variances of earnings are allowed to differ between all combinations of worker types and firm classes. In addition, in the time dimension we will allow for full interactions between firm classes and time indicators, as well as unrestricted non-stationary variances. Lastly, we will treat all $q_k(\alpha)$ and $p_{kk'}(\alpha)$ as unrestricted parameters.[23]

[22] In Supplementary Appendix S1 we describe estimation on T periods.

[23] Identification in Theorem 1 and Corollaries 1 and 2 does not rely on functional form assumptions. Several methods have recently been proposed to estimate finite mixture models while treating the type-conditional distributions nonparametrically (e.g., Levine et al., 2011, Bonhomme et al., 2014). It would be interesting to use such methods to estimate our models. In the empirical analysis we will report results based on mixture of normals, in addition to those based on normal distributions.

Following the spirit of the identification strategy, we first estimate log-earnings densities using job movers only, and we then estimate worker type proportions in the first period using both job movers and job stayers.[24] Under the assumption that worker types and earnings realizations are independent across workers *conditional* on mobility indicators and firm classes, the log-likelihood of job movers conditional on mobility patterns and estimated firm classes takes the following form (N_m denoting the number of job movers):

$$\sum_{i=1}^{N_m}\sum_{k=1}^{K}\sum_{k'=1}^{K}\mathbf{1}\{\widehat{k}_{i1}=k\}\mathbf{1}\{\widehat{k}_{i2}=k'\}\ln\left(\sum_{\alpha=1}^{L}p_{kk'}(\alpha;\theta_p)f_{k\alpha}(Y_{i1};\theta_f)f^m_{k'\alpha}(Y_{i2};\theta_{f^m})\right). \tag{13}$$

In turn, the log-likelihood of all workers in period 1 is:

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{1}\{\widehat{k}_{i1}=k\}\ln\left(\sum_{\alpha=1}^{L}q_k(\alpha;\theta_q)f_{k\alpha}(Y_{i1};\widehat{\theta}_f)\right). \tag{14}$$

Hence, conditional on the estimated firm classes, (13) and (14) are conventional, single-agent correlated random-effects log-likelihood functions. We estimate $\widehat{\theta}_f, \widehat{\theta}_{f^m}, \widehat{\theta}_p$ by maximizing (13), and then $\widehat{\theta}_q$ by maximizing (14). We use the EM algorithm (Dempster et al., 1977) for computation.

Exogenous worker covariates can be readily incorporated in estimation, by modifying the form of the likelihood. In the empirical analysis, given the short length of the panel we will use a nonstationary specification, and relate the latent worker types to time-invariant covariates *a posteriori*. This will allow us to account for sorting on observables such as education or cohort. In Supplementary Appendix S1 we explain how we modify (14) for this purpose.

Dynamic model. We use a similar approach for the dynamic finite mixture model on four periods, see equations (9) and (10). In this case we specify the conditional mean of Y_{i4} given Y_{i3} and worker and firm heterogeneity as $\mu_{4k'\alpha}+\rho_{4|3}Y_{i3}$, where $\mu_{4k'\alpha}$ is a (k',α)-specific intercept. Likewise, the conditional mean of Y_{i1} given Y_{i2} and worker and firm heterogeneity is $\mu_{1k\alpha}+\rho_{1|2}Y_{i2}$. The parameters $\rho_{4|3}$ and $\rho_{1|2}$ capture the persistence of log-earnings within job. For parsimony we have imposed that those parameters are homogeneous across worker types and firm classes, although this could easily be relaxed with a larger sample.

In addition we specify the mean of (Y_{i2}, Y_{i3}) for job movers between classes k and k' as $(\mu_{2k\alpha}+\xi_2(k'), \mu_{3k'\alpha}+\xi_3(k))$, and we let its covariance matrix depend on (k,k'). The term $\xi_2(k')$

[24] Proceeding in this way has the advantage of recovering earnings parameters from job movements directly, albeit at some efficiency cost. In practice we estimate the type proportions of job stayers in the last step, and combine them with the $p_{kk'}(\alpha)$ to recover the unconditional proportions $q_k(\alpha)$.

reflects that, conditional on moving between k and k', mean log-earnings before the move can differ with the firm of destination, due to the presence of *endogenous mobility*. The term $\xi_3(k)$ reflects that the previous firm is allowed to have a direct effect on log-earnings after a move, through the presence of *state dependence*. Neither of those effects is allowed for in the static version of the model. Lastly, we specify the mean of (Y_{i2}, Y_{i3}) for job stayers in a firm of class k as $(\mu^s_{2k\alpha}, \mu^s_{3k\alpha})$, and we let its covariance matrix depend on k.[25]

Given estimates $\widehat{\rho}_{4|3}$ and $\widehat{\rho}_{1|2}$ of the persistence parameters, the other parameters can be estimated using a very similar approach as in the static case, based on the following log-likelihood functions:

$$\sum_{i=1}^{N_m}\sum_{k=1}^{K}\sum_{k'=1}^{K}\mathbf{1}\{\widehat{k}_{i2}=k\}\mathbf{1}\{\widehat{k}_{i3}=k'\}\times ... \\ \ln\left(\sum_{\alpha=1}^{L}p_{kk'}(\alpha;\theta_p)f^f_{Y_{i2},k\alpha}(Y_{i1};\widehat{\rho}_{1|2},\theta_{f^f})f^m_{kk'\alpha}(Y_{i2},Y_{i3};\theta_{f^m})f^b_{Y_{i3},k'\alpha}(Y_{i4};\widehat{\rho}_{4|3},\theta_{f^b})\right), \quad (15)$$

and:

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{1}\{\widehat{k}_{i2}=k\}\ln\left(\sum_{\alpha=1}^{L}q_k(\alpha;\theta_q)f^f_{Y_{i2}k\alpha}(Y_{i1};\widehat{\rho}_{1|2},\widehat{\theta}_{f^f})f^s_{k\alpha}(Y_{i2},Y_{i3};\theta_{f^s})f^b_{Y_{i3},k'\alpha}(Y_{i4};\widehat{\rho}_{4|3},\widehat{\theta}_{f^b})\right). \quad (16)$$

We estimate $\widehat{\theta}_p, \widehat{\theta}_{f^f}, \widehat{\theta}_{f^m}, \widehat{\theta}_{f^b}$ based on (15), and then $\widehat{\theta}_q, \widehat{\theta}_{f^s}$ based on (16), using the EM algorithm in both cases.

While it is in principle possible to estimate $\rho_{4|3}$ and $\rho_{1|2}$ by maximizing a joint likelihood function across movers and stayers with respect to all parameters, doing so would be computationally cumbersome. A convenient alternative, which we adopt in the empirical analysis, is to estimate these parameters in an initial step based on covariance restrictions. Under the assumption that the effect of worker types on mean log-earnings is constant over time within firm, simple restrictions on the ρ's can be obtained by exploiting the particular form of the conditional means of Y_{i4} given Y_{i3} and Y_{i1} given Y_{i2}, respectively. We provide details on the covariance-based estimation of $\rho_{4|3}$ and $\rho_{1|2}$ in Supplementary Appendix S1.

Related estimators. The estimation approach outlined in this section can be modified in several ways that we will implement empirically. A first extension is a model-based re-classification. Given estimates of the θ parameters one may re-classify every firm j to the class $k = \widetilde{k}(j)$ which

[25] We impose that the best linear predictors in the regressions of Y_{i3} on Y_{i2}, for both stayers and movers, do not depend on worker types or firm classes, and that the residual variances in the case of movers only depend on k'. This could be relaxed with a large enough sample.

corresponds to the maximal value of the k-specific likelihoods of firm j's observations. This approach can be iterated further.

A second related estimation strategy is the use of half-sample estimates and bias reduction. Fixed-effects estimators and discrete grouped fixed-effects estimators can be subject to incidental parameter biases. In the present case biases can arise if the number of job movers per firm, or the number of workers per firm, is relatively small so the corresponding parameters are poorly estimated. We use the following simple check which closely mirrors the method of Dhaene and Jochmans (2015). Let $\widehat{\theta}$ be an estimator of θ (or a function of θ such as a variance component) computed on the full data. We will also estimate θ on two half-samples where workers are selected with probability 1/2 within every firm and mobility status (i.e., mover and stayer). Letting $\widehat{\theta}_1$ and $\widehat{\theta}_2$ denote the half-sample estimates we will then compare $\widehat{\theta}$ to $\widetilde{\theta} = 2\widehat{\theta} - (\widehat{\theta}_1 + \widehat{\theta}_2)/2$. Under the assumption that observations are independent across workers within firms, we will interpret a discrepancy between $\widehat{\theta}$ and $\widetilde{\theta}$ as indicative of bias.

Lastly, while we have described estimation in the context of finite mixture models the two-step approach can be used in other settings. Relevant examples are regression models such as the AKM model and its interactive counterparts (1) and (2) that allow for complementarities or dynamics. In such models two-step grouped fixed-effects methods deliver computationally convenient estimation algorithms based on mean and covariance restrictions, as we show in detail in Supplementary Appendix S3.

5 Empirical results I: Static model

We now present results for the static model on the Swedish data. We first report estimates of firm classes, worker types and earnings based on our preferred specification, and then study the robustness of our findings.

5.1 Results

Data. We use administrative data covering the entire working age population in Sweden between 1997 and 2008. We follow Friedrich et al. (2014) for sample selection and construction of monthly log-earnings. We estimate the static model on males in 2002 and 2004. We keep workers who are both fully employed in the same firm in 2002 and fully employed in the same firm in 2004, and firms with at least one fully-employed worker during the period. Descriptive statistics for our sample are shown in Table E2 in Appendix E. In Appendix D we provide details on the Swedish context and sample construction. We define job movements in a conservative way, which we describe in detail in the appendix. This results in low mobility rates, with a proportion of job movers to job stayers of 3.3% in the sample.

Firm classes. As described in Section 4, we estimate firm classes using a weighted k-means algorithm. We use firms' cdfs of 2002 log-earnings on a grid of 20 percentiles of the overall log-earnings distribution. We weight measurements by firm size, and only include job stayers in the classification.[26]

Table 1: Data description, by estimated firm classes

class:	1	2	3	4	5	6	7	8	9	10	all
number of workers	16,868	50,906	74,073	76,616	80,562	66,120	105,485	61,272	47,164	20,709	599,775
number of firms	5,808	6,832	4,983	5,835	3,507	4,149	3,672	3,467	2,886	2,687	43,826
mean firm reported size	12.43	20.92	42.68	28.47	65.06	32.3	60.08	51.24	54.16	50.86	37.59
number of firms ≥ 10 (actual size)	160	1,034	1,519	1,357	1,192	930	999	855	632	415	9,093
number of firms ≥ 50 (actual size)	7	87	260	225	270	162	245	183	147	52	1,638
firm actual size for median worker	4	13	39	47	121	100	429	112	134	22	72
% high school drop out	28.5%	27.8%	25.9%	26.8%	22.2%	23.8%	18.9%	12.9%	6.1%	3.2%	20.6%
% high school graduates	61.3%	63.4%	62.3%	63.3%	59.1%	62.7%	58.4%	49.3%	34.9%	25.6%	56.7%
% some college	10.2%	8.8%	11.8%	9.9%	18.7%	13.5%	22.8%	37.8%	59%	71.2%	22.7%
% workers younger than 30	24.3%	19.5%	19.8%	17.5%	18.6%	15.4%	13.8%	14.3%	15%	14.3%	16.8%
% workers between 31 and 50	54.1%	54.6%	55%	56.2%	56%	57.6%	58.5%	58.9%	60%	64.2%	57.2%
% workers older than 51	21.7%	25.9%	25.1%	26.3%	25.5%	27%	27.6%	26.8%	25%	21.5%	26%
% workers in manufacturing	24.3%	39.3%	46.8%	53%	51.5%	52%	53%	40.3%	31.5%	7.6%	45.4%
% workers in services	39.3%	32.1%	23.3%	19.7%	14.4%	15%	16%	29.7%	52.1%	72.6%	25.3%
% workers in retail and trade	26.4%	19%	24.9%	10.6%	29.3%	7.9%	8.4%	17.7%	14.8%	18.7%	16.7%
% workers in construction	9.9%	9.6%	5.1%	16.8%	4.9%	25.1%	22.5%	12.3%	1.5%	1.1%	12.6%
mean log-earnings	9.69	9.92	10.01	10.06	10.15	10.16	10.24	10.36	10.5	10.77	10.18
variance of log-earnings	0.101	0.054	0.085	0.051	0.102	0.051	0.077	0.096	0.109	0.173	0.124
between-firm variance of log-earnings	0.0462	0.0044	0.0036	0.0018	0.0032	0.0016	0.0016	0.0045	0.0057	0.0435	0.0475
mean log-value-added per worker	14.48	14.97	15.54	15.21	15.82	15.26	15.61	15.69	15.76	15.78	15.3

Notes: Males, fully employed in the same firm 2002 and 2004, continuously existing firms. Actual size is the number of workers per firm in our sample. Figures for 2002.

Table 1 provides summary statistics on the estimated firm classes for our baseline choice of number of classes $K = 10$. We order firm classes from 1 to 10 according to mean log-earnings in each class. Classes capture substantial heterogeneity between firms. The between-firm-class log-earnings variance is 0.0421, that is, 89% of the overall between-firm variance. This

[26]We use the Hartigan-Wong algorithm in the code "kmeansW" in the *R* package "FactoClass", with $10,000$ randomly generated starting values.

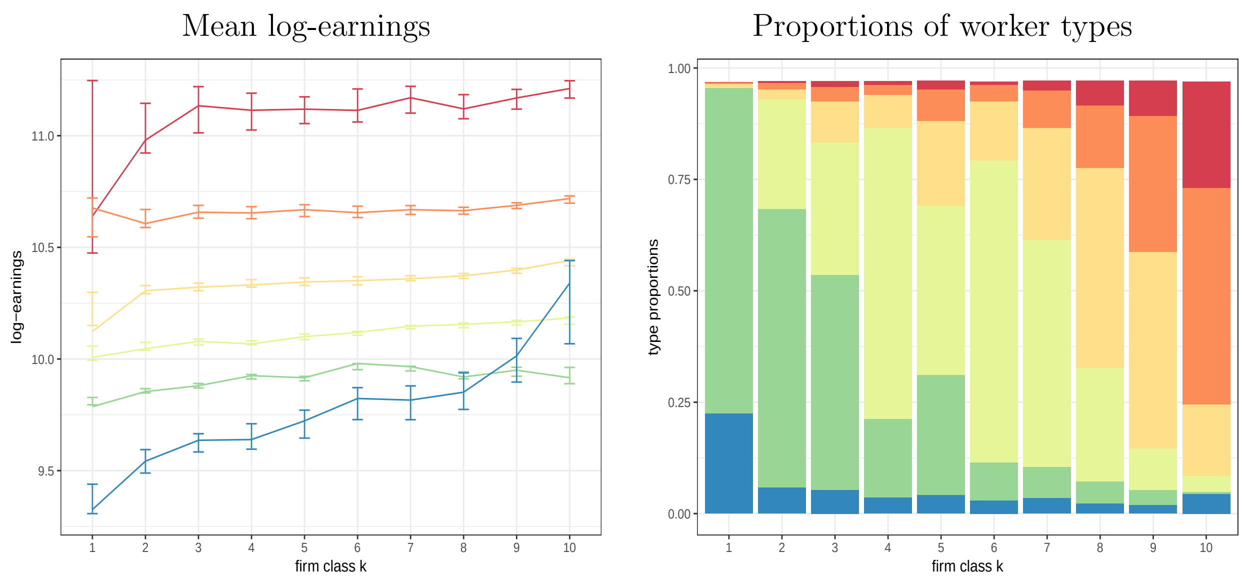
is important since it suggests that assuming homogeneity within each of the 10 classes might not result in major losses of information, at least in terms of variance of log-earnings. There are also substantial differences between classes in terms of worker characteristics. While lower classes (according to their mean log-earnings) show high percentages of high school dropouts and low percentages of workers with some college, higher classes show the opposite pattern. Lower classes also 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.

Firm size reported by the firm tends to increase with firm class, although the relationship is not monotonic. Classes 1 and 2 contain smaller firms than the other classes. There is also evidence of both between- and within-sector variation between classes, which is not monotonic in mean earnings. For example, the proportion of workers in services is U-shaped in firm class, and that in manufacturing is inverse U-shaped. Lastly, log value added per worker tends to increase with firm class, although again there is not a monotonic relationship. Moreover, classes explain only 13.2% of the between-firm variance in log value added per worker.

We next describe some patterns of mobility and earnings 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, especially between adjacent classes. This is important since our identification strategy is based on exploiting mobility. Figure E4 in Appendix E shows means of log-earnings for workers moving from class k to k' (x-axis) and for those moving from k' to k (y-axis), for each pair of firm classes (k, k') with $k < k'$. The graph shows that on average workers "moving up" (i.e., from k to k') tend to have lower earnings than workers "moving down" (from k' to k). We emphasized the importance of such empirical differences in our identification analysis, see equation (6).

Wages, worker heterogeneity and firm heterogeneity. Our baseline estimates are based on a Gaussian finite mixture model with $L = 6$ types of workers and $K = 10$ firm classes. As explained in Section 4 we estimate earnings distributions on the sample of job movers between 2002-2004, as well as the type proportions of movers. We then estimate proportions of worker types of job stayers in 2002. Maximum likelihood estimation of finite mixture models is often subject to local maxima, and our setting is no exception. In addition, in some of the locally optimal solutions some worker types only move within a subset of firm classes, resulting in unstable parameter estimates. In Supplementary Appendix S2 we describe how we use the EM algorithm to explore the likelihood function. We also explain how we use a measure of network connectedness recently studied in Jochmans and Weidner (2017) to select our preferred estimates.[27]

Figure 2: Main parameter estimates of the static model



Notes: Static model, 2002-2004. The left graph plots estimates of the means of log-earnings distributions, by worker type and firm class. The $K = 10$ firm classes (on the x-axis) are ordered by mean log-earnings. On the y-axis we report estimates of mean log-earnings for $L = 6$ worker types. The right graph shows estimates of the proportions of worker types in each firm class. Left: brackets indicate parametric bootstrap 2.5% and 97.5% quantiles (100 replications).

On the left panel of Figure 2 we plot estimates of the means of log-earnings for each firm class and each worker type.[28] On the x-axis, firm classes are ordered by mean log-earnings. The brackets show 95% confidence intervals based on the parametric bootstrap.[29] The results show clear evidence of worker heterogeneity. They also show some variation in log-earnings

[27]Among the 10 best local maximum likelihood values out of 50 we select the solution where the minimum connectedness measure across worker types is the highest. This strategy mainly improves stability across bootstrap repetitions and has little impact on the main estimates. We also computed our main results at the best likelihood value and found very similar results, in both the static and dynamic models.

[28]In Figures S3, S4 and S5 we report measures of fit of the static model on earnings distributions and covariances.

[29]The bootstrap draws are conditional on worker and firm links in the data, and firm classes are re-estimated in each replication. This bootstrap procedure provides a measure of parameter uncertainty that accounts for uncertainty in firm classes. In Appendix B we derive the asymptotic distribution of the estimators. See Supplementary Appendix S2 for implementation details.

between firm classes, although to a lesser extent. Moreover, lower-type workers (where "lower" and "higher" types refer to low and high mean log-earnings) appear to gain the most from working in a higher-wage firm. This suggests the presence of some complementarity between firms and lower-type workers, which we will further explore below.

On the right panel of Figure 2 we report the estimated proportions of worker types in each firm class. The results show how the composition of worker types differs markedly across firm classes. For example, the lowest-class firms (in terms of mean log-earnings) employ mostly the bottom two worker types, while the highest-class firms employ mostly the top three worker types. Overall, the two graphs in Figure 2 suggest that variation in log-earnings between firm classes is mainly due to firms employing different workers, rather than differences in earnings for a given worker type.

Table 2: Variance decomposition and reallocation exercise in the static model

Variance decomposition ($\times 100$)				
$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
80.3 (.8)	3.4 (.2)	16.3 (.6)	49.1 (.9)	74.8 (.6)
Reallocation exercise ($\times 100$)				
Mean	Median	10%-quantile	90%-quantile	Variance
.5 (.09)	.6 (.10)	2.7 (.20)	-1.2 (.30)	-1.1 (.11)

Notes: Static model, 2002-2004. Top panel: α is the worker effect, ψ is the firm effect, variance decomposition based on a linear regression of simulated 2002 log-earnings on α and ψ. Bottom panel: differences in means, quantiles and variances of log-earnings between a sample where workers are randomly reallocated to firms and the original sample. 1,000,000 *simulations. Parametric bootstrap standard errors in parentheses (*100 *replications).*

Variance decomposition and reallocation. We next report the results of two exercises that illustrate how earnings and heterogeneity relate to each other. We start with a decomposition of the variance of 2002 log-earnings. In the literature since Abowd et al. (1999) it is common to decompose the variance of log-earnings, net of observed covariates, into four components: the variance of worker effects (that is, coefficients of worker type indicators), the variance of firm effects (i.e, coefficients of firm class indicators), twice the covariance between the two, and the variance of residuals. In our nonlinear model a similar decomposition can be performed by working with a linear projection of log-earnings on worker type indicators and firm class indicators, in a regression without interactions.[30] In the top panel of Table 2 we show percentages of explained variance due to worker and firm heterogeneity and due to the covariance between the two. In addition we report the correlation between worker and firm effects and the R^2 in the linear regression. The results show two main features. First, worker heterogeneity explains substantially more variation in earnings than firm heterogeneity. Differences in firm classes only account for 3.4% of the explained variance, compared to 80.3% for the part due to differences in worker types. The second main finding is that the part explained by the covariance is substantial. The correlation between worker and firm effects is 49.1%, which suggests the presence of strong sorting between workers and firms. This is in line with the evidence documented on the right panel of Figure 2.

As a first way to quantify the economic magnitude of complementarities, we next assess the explanatory power of worker types and firm classes when entered interactively as opposed to additively in the regression. The R^2 coefficient in the linear regression is 74.8%, while in the regression that includes all interactions between worker type indicators and firm class indicators the R^2 is 75.8%. This suggests that, while the left panel of Figure 2 shows the presence of some complementarity between firms and low-type workers, those complementarities explain only a small part of the overall variance of log-earnings.

We next consider the impact on log-earnings of a reallocation exercise where workers are allocated randomly to firms. Such an exercise aims at assessing the contribution of sorting to the distribution of earnings.[31] We show the results of the reallocation on the bottom panel of Table 2. In the first column we report the estimate of the difference in mean log-earnings between a counterfactual sample where workers are randomly allocated among firms and our sample, using our estimates that account for the presence of complementarities. In an additively separable economy between workers and firms, such as under the AKM model, there should be no effect of the reallocation on mean log-earnings (e.g., Graham et al., 2014). However we find a positive mean impact (.5%), which, albeit moderate, suggests that the effect of complementarities on average log-earnings is not insignificant.

To provide an intuition on the mean effect of the reallocation, consider the regression model (1). In this specification the difference between mean outcomes in a population where workers

[30] We compute the variance decomposition by simulation. See Supplementary Appendix S2 for details.

[31] In the exercise we assume that earnings functions, for all worker types and firm classes, are not affected by the reallocation. Hence this exercise abstracts from equilibrium effects through which changes in composition could affect the form of the conditional earnings distributions.

are randomly allocated to firms and in our data is, abstracting from time indices for clarity:

$$\mathbb{E}^{random}(Y_i) - \mathbb{E}(Y_i) = -\operatorname{Cov}\left(b(k_i), \mathbb{E}(\alpha_i \mid k_i)\right). \tag{17}$$

In particular the reallocation has no effect on the mean if $b(k)$ does not depend on k, or more generally if complementarities $b(k)$ are uncorrelated with the mean worker type in the firm class $\mathbb{E}(\alpha_i \mid k)$. Our results on Swedish data show that, while the mean worker type tends to increase with firm class k, complementarities are stronger in low firm classes, yielding a negative covariance. This explains why the effect of the reallocation is positive.

Moreover, in our distributional framework we are able to estimate the entire distribution corresponding to a given reallocation of workers to firms. In columns 2 to 4 of the bottom panel of Table 2, we show the differences in medians and 10% and 90% percentiles of log-earnings between the random allocation and our sample. Those are "quantile treatment effects" corresponding to the change. We also report differences in variances in the last column. We see that, while the median effect is in line with the mean effect, the bottom of the distribution would tend to benefit in the random allocation, while the top would be hurt. Those differences reflect both the fact that log-earnings are less dispersed in the random allocation, as shown by the reduction in variance, and the presence of complementarities at the bottom of the distribution.

In addition, our framework allows us to analyze dynamic aspects of worker-firm interactions. We will present such analyses in the next section, using our dynamic model.

5.2 Robustness

In this subsection we present several exercises to show the robustness of our main results. For brevity, most results can be found in Supplementary Appendix S4.

Firm classes and worker types. The number K of firm classes is an important input for the two-step grouped fixed effects method. In Table S2 we show the results of the variance decomposition when varying K between 3 and 20 classes. The results are quite stable across K values.[32] In Table S4 we show the results of varying the number of worker types between $L = 3$ and $L = 9$. The results show that taking $L = 3$ or $L = 4$ seems to understate the contribution

[32]We also computed the BIC criterion of Schwarz (1978) based on the overall likelihood function. BIC points to a relatively small number ($K = 7$) of classes. In Bonhomme et al. (2017) we used a different rule to select K motivated by an asymptotic theory where firm heterogeneity may not be discrete in the population, and taking too small a K may lead to a poor approximation. That rule also gave $K = 7$.

of worker heterogeneity and overstate that of firm heterogeneity. The results are very stable between $L = 5$ and $L = 9$.[33]

We next consider a series of specifications in which firm classification uses additional information beyond earnings distributions. In the top panel of Table S3 we report variance decomposition results for three different splits within the 10 classes we estimated using k-means. We split each class in two subclasses according to a firm-specific "mobility rank" (first row), the percentage of job movers in the firm (second row), and value added per worker (third row). We construct the mobility rank of a firm as the number of workers who move to that firm from a firm whose mean log-earnings belongs to the lowest tercile, divided by the total number of movers to the firm.[34] We split the class in two depending on the firm ranking above or below the median value of the measure in the class. These specifications capture other information, beyond differences in log-earnings distributions, which may be contained in mobility patterns and value added.[35] The results show some differences with our baseline estimates. For example, when splitting the classes according to value added the firm effects variance becomes 4.7%, compared to 3.4% in the baseline. The three specifications also give slightly smaller correlations between worker and firm effects. However the differences are small, suggesting that the initial classification captures most of the firm heterogeneity which is relevant to earnings variation.

In the second panel of Table S3 we re-classify firms into classes based on the mixture model given our estimates. Unlike our baseline classification, this method incorporates information from both periods, including earnings associated with job mobility. To alleviate biases associated with low mobility rates (see below) we only use half of the job movers within each firm for classification in addition to job stayers, and the other half for estimation. We report results based on one and five iterations. Compared to the baseline results based on k-means clustering of cross-sectional log-earnings cdfs, the results using the re-classification are similar except for a slightly smaller correlation parameter. In our setting, where mobility is unrestricted as a function of worker types and firm classes, the model-based re-classification has little impact on the results compared to two-step grouped fixed-effects. In fact the correlation between firm classes in the two cases is 94% after five iterations.

[33] BIC points to $L = 6$ in this case.

[34] This measure is closely related to the poaching rank introduced by Bagger and Lentz (2014), except that it combines mobility and earnings information. Bagger and Lentz' poaching rank uses workers coming from unemployment.

[35] However, using any of the measures to rank firms directly would not be well-suited for the exercise we conduct. For example, the mobility rank explains less than 20% of the variance of log-earnings.

Other specifications. In Table S5 we report variance decomposition results corresponding to several additional specifications. In the second row of the table we show the results for a model where log-earnings are distributed as three-component mixtures of Gaussians conditional on worker types and firm classes. This provides additional flexibility in earnings distributions, at the cost of adding parameters to an already richly parameterized model. The results are very similar to the baseline. In the third row we show the results of the decomposition obtained from our baseline mixture model, when controlling for the effect of two worker covariates: age and education. We find that, once worker covariates are controlled for, the contribution of worker unobserved heterogeneity is 67% of the variance explained by worker and firm unobservables, as opposed to 80% in the baseline. The relative contribution of firm heterogeneity is larger than in the baseline.[36] In rows four and five we report decompositions estimated on smaller firms (less than 50 workers per firm in 2002) and large firms (more than 50). While still qualitatively similar, the results show interesting heterogeneity: in smaller firms the contribution of worker heterogeneity is reduced compared to that of firm heterogeneity and the covariance term. In comparison, in larger firms worker heterogeneity accounts for a greater share (i.e., 85% versus 72%) of the log-earnings variance. Lastly, in the sixth row we show that the variance decomposition implied by a fully nonstationary model gives very close results to our baseline estimates.

In panel B of Table S5 we show the results for the interactive regression model (1).[37] This model is a valuable complement to our baseline model for several reasons. First, worker types are *not* discretely distributed in the regression model. Second, estimation only relies on mean and covariance restrictions, not on other features of the multivariate distribution of the data. For example, the parameters $a_t(k)$ and $b(k)$ are estimated from mean restrictions alone, which do not rely on assumptions on serial dependence. In addition, the regression estimator is straightforward to compute. On the other hand, unlike our baseline specification the model's functional forms restrict the shape of interaction effects between worker and firm heterogeneity. The results show a very similar variance decomposition as in our baseline model, the only notable difference being that the R^2 is lower. Finally, in the last row of the table we report the results for a regression model where we additionally impose that $b(k) = 1$ for all k. This corresponds to a one-sided random-effects version of the additive model of Abowd et al. (1999).

[36] In Figure S6 in Supplementary Appendix S4 we show how the proportions of worker types by firm classes within age×education cells (9 categories). In Supplementary Appendix S1 we explain how we estimate those parameters.

[37] We abstract away from covariates (i.e., $c_t = 0$) and we assume that the worker-firm interaction coefficients $b_t(k)$ are constant over time. We explain in detail how we estimate the model in Supplementary Appendix S3.

However, our estimator based on classification and mean and covariance restrictions differs from the AKM estimator, and is arguably more robust to incidental parameter bias in our data (see the next paragraph). Overall, these estimates confirm the stability of the variance decomposition results.

Incidental parameter bias. We next investigate the bias of parameter estimates and variance decomposition. We first document the bias of the estimator of Abowd et al. (1999) of the variance of firm effects on samples with different selections on the number of movers per firm.[38] The solid line with circles in Figure E5 in Appendix E shows that the firm effects variance decreases substantially and monotonically, from .011 to 0.006 when the number of movers per firm increases from ≥ 4 to ≥ 40. On the same figure the solid line with triangles shows the firm effects variance estimated using a bias-corrected estimator.[39] The dashed line varies much less with the number of movers, and gives a firm effects variance of .004 in the subsample with more than 40 movers per firm. This suggests the presence of substantial low mobility bias in this data. Interestingly, .004 is close to the number implied by our baseline estimates (in Table 2), which we report on the dotted horizontal line. Hence, with respect to the variance contribution of firm effects, there is little difference between our estimate of the variance contribution of firms and an estimate based on a bias-corrected AKM estimator, although there are more substantial differences with the results using the original AKM estimator.[40]

We next move on to study the finite sample bias of our estimator. In Figure S7 we first report results based on two half-subsamples of the data, where each firm is randomly split in two.[41] We implement this procedure in an attempt to assess the incidental parameter bias due to the fact that some firms are small or have few job movers, resulting in noisy estimates of some of the parameters. We see that the results are similar to the full-sample results of Figure

[38] In each subsample we focus on the largest connected set of firms. To increase the stability of the results, for this exercise we work with a larger sample constructed by pooling the 2002-2004 sample together with similar samples from 1997-1999 to 2006-2008 (while not keeping track of workers' identities across subpanels). Our main results are very similar in this larger sample, and they are available upon request.

[39] The estimator is based on half-sample estimation (as in Dhaene and Jochmans, 2015) where, within each firm, job stayers and job movers out of the firm are split into two random subsamples of equal sizes. The bias-corrected estimate is equal to twice the full-sample estimate minus the mean of the half-sample estimates.

[40] In a previous version of the paper we also estimated fixed-effects AKM regressions on data simulated from our mixture model. We found large incidental parameter biases in the variance decomposition. For example, we found a negative correlation between worker and firm effects (around -25%), compared to a population parameter of more than 40%. At the same time, the biases were substantially reduced when we let the number of job movers per firm increase tenfold and we increased the lengths of job spells.

[41] We use the same procedure to generate the two random subsamples as described in footnote 39.

2. This suggests that biases due to low firm sizes and low mobility rates have little impact on our results. This is also confirmed by the results of the variance decompositions in the two half-samples, which are very similar to the decomposition on the full sample, see the last two rows of panel A in Table S5. The results point to a slight underestimation of the firm contribution and the covariance component, and a slight overestimation of the worker contribution, but the biases appear very small.[42]

Lastly, we show in Table S6 the means and 2.5%-97.5% percentiles of the 100 replications of the parametric bootstrap, for each of the main results of Table 2. The bootstrap means do not exactly coincide with our estimates, and standard errors are small given the large sample size. The presence of bias is not surprising in a nonlinear model with rich two-sided heterogeneity and a large number of parameters, including not only earnings parameters but also type proportions by mobility and firm classes. Nevertheless, the biases of the variance decomposition and reallocation effects are economically small.

6 Empirical results II: Dynamic model

In this section we present empirical results for our dynamic model. We first report parameter estimates, and then analyze cross-sectional and dynamic features of the model in turn.

6.1 Parameter estimates, cross-sectional decomposition and reallocation

We estimate the dynamic model on 2001-2005, focusing on males both fully-employed in the same firm in 2001-2002, and fully-employed in the same firm in 2004-2005. In order to estimate firm classes we use the same weighted k-means algorithm as in the static model.[43] We then estimate the model in three steps, as explained in Section 4: we estimate the earnings persistence parameters $\rho_{4|3}$ and $\rho_{1|2}$, the wage functions and type probabilities of job movers, and finally those of job stayers.

Table 3 shows estimates of several parameters of the model.[44] The parameter $\xi_2(k')$ is the

[42]A bias-corrected estimate of the firm component can be constructed as $2 \times 3.4 - 1/2 \times (3.0 + 3.0) = 3.8\%$. Likewise, bias-corrected worker component and covariance component are 78.7% and 17.6%, respectively. Those are close to the un-corrected estimates.

[43]We show descriptive statistics in Table E2 in Appendix E. In Appendix D we provide details on sample selection. Summary statistics on the partition of firms are shown in Table S7 in Supplementary Appendix S4.

[44]In Figures S8, S9 and S10 in Supplementary Appendix S4 we report measures of fit of the dynamic model.

Table 3: Parameter estimates of the dynamic model

	Earnings effects $\xi_2(k')$ of future firm classes								
$k' =$	2	3	4	5	6	7	8	9	10
estimate	$-.005$	.004	.005	.022	.003	.015	.009	.016	.023
	(.008)	(.009)	(.011)	(.012)	(.011)	(.010)	(.011)	(.011)	(.011)
	Earnings effects $\xi_3(k)$ of past firm classes								
$k =$	2	3	4	5	6	7	8	9	10
estimate	.051	.038	.045	.061	.040	.072	.058	.087	.090
	(.015)	(.015)	(.014)	(.015)	(.018)	(.015)	(.016)	(.015)	(.016)
	Persistence parameters ρ								
	$\rho_{1\|2}$	$\rho^m_{3\|2}$	$\rho^s_{3\|2}$	$\rho_{4\|3}$					
estimate	.227	.246	.681	.651					
	(.009)	(.040)	(.023)	(.004)					

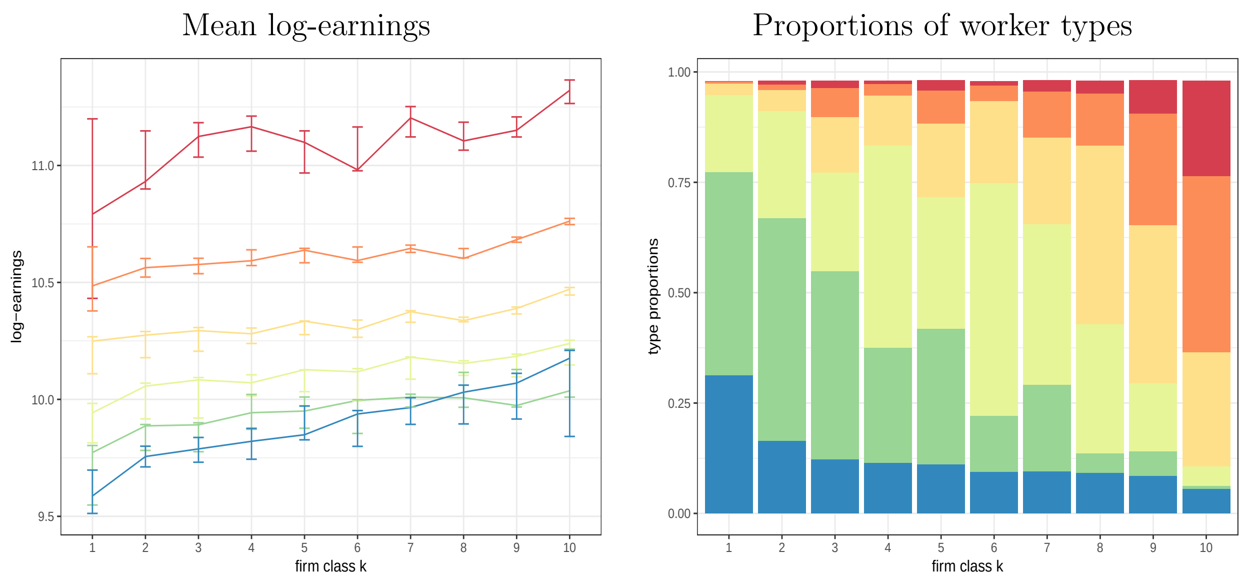
Notes: Dynamic model, 2001-2005. $\rho^m_{3|2}$ is the autoregressive coefficient of log-earnings for job movers between 2002 and 2004; $\rho^s_{3|2}$ is the coefficient for job stayers. $\xi_2(k')$ and $\xi_3(k)$ are the mean effects on log-earnings before and after a job move between firm classes k and k', respectively. $k' = 1$ (resp., $k = 1$) is the omitted category. Parametric bootstrap standard errors in parentheses (100 replications).

effect of firm class k' on the mean log-earnings in period 2 of a worker moving from k in period 2 to k' in period 3. It would be zero for all k' under the strict exogeneity assumption on mobility, which is imposed in our static model and many models in the literature.[45] We see that the effects are quantitatively small, with at most a 2% effect relative to the omitted class $k' = 1$. The parameter $\xi_3(k)$, in turn, is the *state dependence* effect of firm class k on the mean log-earnings of a worker moving between k and k'. Our static model, and many models in the literature, would also rule out the presence of a direct effect of the past firm's class on current earnings. This effect appears empirically quite large in the dynamic model. It is approximately monotonic in firm class, and amounts to a 9% effect in the highest classes. This suggests that past firms have an impact on future earnings.

In the bottom panel of Table 3 we report the estimates of earnings persistence parameters. Persistence estimates are higher for job stayers than for job movers. Notice however that

[45] Strict exogeneity of mobility also imposes that earnings realizations are independent of the subsequent decision to move, conditional on firm and worker heterogeneity, irrespective of the class of the firm the worker wishes to move to. This assumption cannot be tested from Table 3, but we will check its empirical plausibility in Table 5 below.

Figure 3: Parameter estimates of the dynamic model (continued)



Notes: Dynamic model, 2001-2005. See notes to Figure 2.

the autoregressive coefficient of .246 upon job move is significantly different from zero, which suggests that the conditional independence assumption of the static model does not hold in our data. As a robustness check we estimated the persistence parameters based on covariance restrictions in first differences, as opposed to levels. The literature has documented differences between level estimates and first difference estimates of the dynamics of earnings in several data sets, see for example Daly et al. (2016). We found $\rho_{1|2} = .506$, $\rho_{4|3} = .451$, $\rho^m_{3|2} = .194$, and $\rho^s_{3|2} = .605$. Despite the differences in persistence estimates, variance decomposition results are similar in this case.[46]

Next, in Figure 3 we show estimates of mean log-earnings for each worker type and firm class (left panel) and type proportions in each firm class (right panel). The results correspond to 2002. We see similar cross-sectional patterns as for the static model, with approximate additivity of log-earnings in worker types and firm classes, relatively small differences across firms for all worker types except the lowest one, and strong evidence of association between worker types and firm classes. This suggests that the dynamic and static models have similar

[46] The results of the variance decomposition for $\rho_{4|3}$ and $\rho_{1|2}$ estimated in differences are shown in the fourth row of Table S11 in the supplement. The dynamic implications of the model (i.e., endogenous mobility and state dependence) are also similar. In Figures S11 and S12 we show the fit to covariances in levels and covariances in differences of the models estimated in levels and first differences, respectively.

implications for cross-sectional earnings distributions.

Table 4: Variance decomposition and reallocation exercise in the dynamic model

Variance decomposition ($\times 100$)				
$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
77.4 (1.2)	5.5 (.9)	17.2 (.5)	41.9 (2.4)	77.9 (.7)
Reallocation exercise ($\times 100$)				
Mean	Median	10%-quantile	90%-quantile	Variance
.3 (.35)	.8 (.18)	2.5 (.73)	-3.0 (.47)	-1.0 (.88)

Notes: Dynamic model, 2001-2005. α is the worker effect, ψ is the firm effect. In the reallocation exercise we randomly assign workers to firms. See the notes to Table 2.

We next report the results of a cross-sectional variance decomposition exercise in the top panel of Table 4. As in the static case we perform the decomposition on 2002 earnings. The variance decomposition is quite similar to the one we obtained from the static model, with some differences: the contribution of firm effects increases from 3.4% to 5.5%, and the correlation between worker and firm effects decreases from 49.1% to 41.9%. As in the static model, adding interactions between worker types and firm effects has relatively small effects on the R^2 of the linear regression (i.e., 78.5% versus 77.9%).

The bottom panel of Table 4 shows that the distributional effects of randomly reallocating workers across firms are also in line with the static results. The reallocation has a positive effect on mean (now insignificant) and median (significant) log-earnings, with asymmetric effects on the two tails of the distribution.[47]

Our estimates are stable across a range of specifications. We report the results of the following robustness checks in Supplementary Appendix S4: varying the number of firm classes and worker types (see Tables S9 and S10), using an interactive regression estimator estimated from mean and covariance restrictions, and estimating the model on random half-samples (see Table S11 and Figure S13). We also report means and percentiles of the parametric bootstrap in Table S12.

[47]Note that the variance effect takes into account both between type-and-class and within type-and-class dispersion, since our earnings model allows for heteroskedasticity. The between component, which can directly be compared to the covariance term in the first row of Table 4, decreases by 1.7% (with a standard error of .45%), while the effect of the within component is insignificantly positive at .7% (with a standard error of .7%).

6.2 Dynamic effects: endogenous mobility and state dependence

While the cross-sectional variance decomposition of log-earnings and the reallocation exercise give comparable results as in the static case, the estimates of our dynamic model challenge some assumptions commonly made in empirical work, such as exogenous mobility and the absence of state dependence in firm effects. Moreover, these findings based on an empirical model with worker and firm heterogeneity are suggestive of theoretical mechanisms that have been emphasized in the structural literature.

Table 5: Transition probabilities (×100) by conditional decile of previous earnings

	All movers	$k' = 1-3$	$k' = 4-7$	$k' = 8-10$
All				
$k = 1-3$	2.2 (.03)	.8 (.02)	1.1 (.03)	.3 (.02)
$k = 4-7$	1.9 (.03)	.4 (.02)	1.0 (.02)	.5 (.01)
$k = 8-10$	2.8 (.05)	.5 (.06)	1.0 (.03)	1.4 (.02)
First conditional decile of earnings				
	All movers	$k' = 1-3$	$k' = 4-7$	$k' = 8-10$
$k = 1-3$	3.3 (.22)	1.4 (.09)	1.5 (.13)	.4 (.05)
$k = 4-7$	3.3 (.17)	.9 (.08)	1.7 (.09)	.7 (.04)
$k = 8-10$	5.0 (.29)	.9 (.16)	1.8 (.13)	2.2 (.12)
Tenth conditional decile of earnings				
	All movers	$k' = 1-3$	$k' = 4-7$	$k' = 8-10$
$k = 1-3$	3.1 (.20)	1.0 (.08)	1.6 (.13)	.5 (.06)
$k = 4-7$	1.9 (.10)	.4 (.04)	1.0 (.06)	.5 (.03)
$k = 8-10$	2.0 (.13)	.3 (.06)	.7 (.06)	1.0 (.07)

Notes: Probability of moving, overall and by destination firm class k', for each origin firm class. Top panel: all workers. Middle and bottom panels: first and tenth decile of log-earnings Y_{i2} conditional on worker type α_i and current firm class $k_{i2} = k$. Dynamic model, $10,000,000$ simulations. Standard errors based on the parametric bootstrap (100 replications) in parentheses.

Endogenous mobility. We start by computing the overall probability of a job move between 2002 and 2004 conditional on the firm class at origin, and the job move probability by firm class at destination.[48] In the top panel of Table 5 we show those numbers in the full sample, while in the middle and bottom panels we select workers for whom the conditional earnings rank in 2002, given firm class and worker type, is below .10 and above .90, respectively.[49] In the first column we see that while the overall 2002-2004 mobility rate lies between 2% and 3% it is substantially higher for workers who had a low earnings realization in 2002. This suggests the presence of endogenous mobility, in line with estimates in Abowd et al. (2015) and with the predictions of wage posting models with match-specific heterogeneity (for example). In contrast, high earnings realizations do not seem to strongly affect mobility. Lastly, results by firm class at destination do not seem to vary much with earnings realizations, which is in line with the small estimates of $\xi_2(k')$ in Table 3.

State dependence. We next document the dynamics of earnings after a job move. For this, we decompose the within-worker-type variances of log-earnings of job movers in 2004 and 2005[50] and compare, within type, the earnings effect of the current firm class to the effect of the past firm class. Unlike most models (such as AKM and our static model), our dynamic model allows past firm classes to affect earnings.[51]

In the first two rows of Table 6 we decompose the within-worker type variance of 2004 and 2005 log-earnings into three terms: a between-current firm class component (first column), a within-current and between-past firm class component (second column), and a residual component within both classes (third column). We find that, in the year after the move, the past firm class contributes to .86% of the within-type variance. This is 10% of the contribution of the current firm class. One year after (that is, in 2005) the contribution of the past firm class is twice as small (.37%). This suggests that the contribution of state dependence is not negligible, but tends to decrease over time.

[48] The joint distribution of firm classes upon job mobility is shown in Figure E6 in Appendix E.

[49] We compute probabilities of job mobility by simulation, drawing earnings for movers and stayers according to the model and weighting movers and stayers according to their empirical frequencies, see Supplementary Appendix S2. Note that the bootstrap procedure that we use to compute standard errors holds the links between firms and workers fixed across replications. To account for variability in the links we also computed bootstrap standard errors where we resampled individual sequences of firm indicators independently. We found qualitatively similar, albeit slightly larger standard errors compared to Table 5.

[50] Within-type variances amount to about 40% of the overall variances.

[51] In fact, our model also allows the *future* firm class to affect earnings *before* the move (in 2001 and 2002). However, those effects are empirically very small, in line with our estimates of the parameters $\xi_2(k')$ in Table 3.

Table 6: Decomposition of the within-worker-type variance of log-earnings (×100) implied by the dynamic model

	Within type		
year		within current firm	
	between current firm	between past firm	residual
2004	8.6 (.87)	.86 (.46)	90.6 (1.2)
2005	10.0 (1.0)	.37 (.21)	89.6 (1.2)

	Within type and past firm		
	total	network effect	state dependence
2004	2.6 (.76)	.87 (.13)	1.7 (.68)
2005	2.0 (.51)	1.0 (.15)	1.0 (.40)

Notes: All numbers are in percentage of the within-type variance of log-earnings. Job movers between 2002 and 2004 only. Bottom panel: "network effect" and "state dependence" effects are defined in the text. Dynamic model, $1,000,000$ *simulations. Standard errors based on the parametric bootstrap (100 replications) in parentheses.*

To better understand the relevance of state dependence effects we perform a last decomposition. In our dynamic model there are two reasons why the earnings of a worker of a given type may depend on the firm where she previously worked. First, working in a particular firm may make her more likely to move to particular firms. This *network effect* is also present in AKM and our static model. In fact, in those models mobility can depend in an unrestricted way on all firms where the worker ever worked or will work. In addition, in our dynamic model the past firm can have a direct effect on the worker's earnings after a job move. This *state dependence effect*, which is represented by the coefficients $\xi_3(k')$, is absent from AKM and our static model. Both effects are theoretically well-grounded. Consider as an example the sequential bargaining model of Postel-Vinay and Robin (2002), where firms are characterized by their productivities. Due to the process of offers and counter-offers, workers tend to work in similarly productive firms over time (network effect). Moreover, a worker coming from a more productive firm is able to extract a higher share of the surplus from the poaching firm, compared to a worker coming from a less productive firm (state dependence).

To study those effects, we focus on 2004 and 2005 log-earnings of job movers; that is, after the job move. We compute the following counterfactual variance where we shut down the effect of state dependence (omitting the conditioning on worker types for simplicity):

$$\text{Var}^{netw}\left(Y_{i3}\right) = \text{Var}\Big(\,\mathbb{E}\Big[\mathbb{E}(Y_{i3}\,|\,k_{i3})\,\Big|\,k_{i2}\Big]\,\Big).$$

In $\text{Var}^{netw}\left(Y_{i3}\right)$, the dependence on the past firm class k_{i2} only reflects the dependence of k_{i3} on k_{i2}; that is, the network effect associated to mobility across classes being non-random. Hence, in a static model $\text{Var}^{netw}\left(Y_{i3}\right)$ would be identical to the between-k_{i2} variance of Y_{i3}. In our dynamic model, the difference between the two reflects the degree of state dependence by which the earnings of a worker of a certain type in a current firm class are affected by the past firm class. The results in the last two rows of Table 6 show that, in the year 2004 after a job move, the total effect of the past firm amounts 2.6% of the within-type variance. One third of it is due to the network effect, while two thirds reflect state dependence. In 2005, the relative contribution of state dependence is smaller. Those findings should be of interest for the modeling of mobility and earnings since they show that, at least in the short run, state dependence is of a similar order of magnitude (in fact, larger) compared to the network effect, although standard static empirical models rule out the former and leave the latter fully unrestricted.

7 Conclusion

In this paper we propose a framework to allow for two-sided unobserved heterogeneity in matched employer employee data sets. We introduce empirical models which allow for worker-firm interactions and dynamics, hence for mechanisms that have been emphasized in theoretical work. We provide conditions for identification in short panels, and develop estimators for finite mixtures and regression specifications.

Our application to Swedish administrative data shows that an additive model provides a good first-order approximation to the variance structure of log-earnings, while at the same time showing a strong association between worker and firm heterogeneity and a small relative contribution of firms to earnings dispersion. The magnitudes we find differ from many estimates of variance components in the literature. A recent paper by Borovickova and Shimer (2017) proposes a different measure of sorting and also finds a strong worker-firm association on Austrian data. Another recent paper by Lentz et al. (2017) uses an estimator related to ours to study wages and mobility using Danish administrative data while accounting for unemployment.

Our results show complementarities between lower-type workers and firm heterogeneity. Although they do not have large effects on the variance structure of log-earnings, these nonlinearities matter to quantify the impact of worker reallocations on the mean and other features of the distribution. Using our dynamic model we also find that endogenous mobility, by which earnings shocks affect mobility decisions, and state dependence, by which past firms have a direct impact on earnings after a job move, are features of our data. These findings support mechanisms that have been emphasized in the structural literature. At the same time, our estimates call for theoretical models that, unlike standard sorting models where complementaries between agents drive the nature of the allocation, can rationalize the presence of a relatively small firm effect and a strong association between worker and firm heterogeneity.

Our two-step estimation approach preserves parsimony by reducing the dimension of firm heterogeneity to a smaller number of classes, and modeling the conditional distributions of worker types. We show this strategy is helpful in alleviating small-sample biases arising from low mobility rates. In companion work (Bonhomme et al., 2017) we further study the theoretical properties of approaches based on an initial clustering step, viewing discrete estimation as an approximation to individual or firm heterogeneity.

Two-step estimation could be useful in structural settings too, where joint estimation of the distribution of two-sided heterogeneity and the structural parameters may be computationally prohibitive. 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. Our identification results could also prove useful for structural work on workers and firms. In this light, an interesting extension of our results would be to allow for time-varying processes of worker types that could vary in response to firm-level shocks.

Lastly, this paper proposes a portable methodology for empirical work. Our methods may reveal interesting patterns of sorting and complementarities in other studies of workers and firms, including in relatively small samples such as a particular occupation or a short period of time (e.g., around a recession), where dimension reduction is likely to be particularly helpful. More generally, we expect our methods to be useful in other settings involving matched panel data, for example in economics of education, urban economics, or finance.

References

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

Abowd, J. M., F. Kramarz, P. Lengermann, and S. Pérez-Duarte (2004): "Are Good Workers Employed by Good Firms? A Test of a Simple Assortative Matching Model for France and the United States," *Unpublished Manuscript*.

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

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

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, 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, 673–697.

——— (2012): "High Wage Workers Match with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias," *Econ. Lett.*, 117, 824–827.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J. M. Robin (2014): "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104, 1551–1596.

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." *Unpublished Manuscript*.

Bartolucci, C., F. Devicienti, and I. Monzon (2015): "Identifying Sorting in Practice," *Unpublished Manuscript.*

Baumgarten, Daniel, G. F. and S. Lehwald (2014): "Drivers of Wage Inequality in Germany: Trade, Technology, or Institutions?" *Unpublished Manuscript.*

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–46.

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

Bonhomme, S. (2017): "Econometric Analysis of Bipartite Networks," *to appear in "The Econometric Analysis of Network data," edited by B. Graham and A. De Paula.*

Bonhomme, S., K. Jochmans, and J. M. Robin (2014): "Nonparametric Estimation of Finite Mixtures from Repeated Measurements," *Journal of the Royal Statistical Society, Series B, forthcoming.*

Bonhomme, S., T. Lamadon, and E. Manresa (2017): "Discretizing Unobserved Heterogeneity," *Unpublished Manuscript.*

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

Borovickova, K. and R. Shimer (2017): "High Wage Workers Work for High Wage Firms," *Unpublished Manuscript.*

Burdett, K. and M. G. Coles (2003): "Wage/Tenure Contracts and Equilibrium Search," *Econometrica*, 71, 1377–1404.

Burdett, K. and D. T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *Int. Econ. Rev.*, 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, 967–1015.

Daly, M., D. Hryshko, and I. Manovskii (2016): "Improving the Measurement of Earnings Dynamics," *Unpublished Manuscript.*

De Melo, R. L. (2009): "Sorting in the Labor Market: Theory and Measurement," *Unpublished Manuscript.*

Delacroix, A. and S. Shi (2006): "Directed Search on the Job and the Wage Ladder," *Int. Econ. Rev.*, 47, 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, 1–38.

Dhaene, G. and K. Jochmans (2015): "Split Panel Jackknife Estimation," *Review of Economic Studies*, 82, 991–1030.

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

Finkelstein, A., M. Gentzkow, and H. Williams (2016): "Sources of Geographic Variation in Health Care: Evidence from Patient Migration," *Quarterly Journal of Economics*, 131, 1681–1726.

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

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

Goldschmidt, D. and J. Schmieder (2015): "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure," *to appear in Quarterly Journal of Economics.*

Graham, B. S., G. W. Imbens, and G. Ridder (2014): "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis," *Quantitative Economics*, 5, 29–66.

Gruetter, M. and R. Lalive (2009): "The Importance of Firms in Wage Determination," *Labour Econ.*, 16, 149–160.

Hagedorn, M., T. H. Law, and I. Manovskii (2014): "Identifying Equilibrium Models of Labor Market Sorting," *to appear in Econometrica.*

Hahn, J. and H. R. Moon (2010): "Panel Data Models with Finite Number of Multiple Equilibria," *Econometric Theory*, 26, 863–881.

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

Hall, R. E. and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50, 461–481.

Henry, M., Y. Kitamura, and B. Salanié (2014): "Partial Identification of Finite Mixtures in Econometric Models," *Quant. Econom.*, 5, 123–144.

Hu, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27–61.

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

Hu, Y. and M. Shum (2012): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *J. Econom.*, 171, 32–44.

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

Jochmans, K. and M. Weidner (2017): "Fixed-Effect Regressions on Network data," *Unpublished Manuscript*.

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 (2015): "Using Compulsory Mobility to Identify School Quality and Peer Effects," *Oxford Bulletin of Economics and Statistics*, 77, 566–587.

Lamadon, T., J. Lise, C. Meghir, and J. M. Robin (2013): "Matching, Sorting, Firm Productivity and Wages," *Unpublished Manuscript*.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2017): "On Worker and Firm Heterogeneity in Wages and Employment Mobility: Evidence from Danish Register Data," *Unpublished Manuscript*.

Levine, M., D. R. Hunter, and D. Chauveau (2011): "Maximum Smoothed Likelihood for Multivariate Mixtures," *Biometrika*, 98, 403–416.

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.

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," *to appear in American Economic Review.*

Mendes, R., G. J. van den Berg, and M. Lindeboom (2010): "An Empirical Assessment of Assortative Matching in the Labor Market," *Labour Econ.*, 17, 919–929.

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

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

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

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

Shimer, R. (2005): "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," *Journal of Political Economy*, 113, 996–1025.

——— (2006): "On-the-Job Search and Strategic Bargaining," *Eur. Econ. Rev.*, 50, 811–830.

Shimer, R. and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68, 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, University of Chicago Press, 217–260.

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

Sorkin, I. (2016): "Ranking Firms Using Revealed Preference," *Unpublished Manuscript.*

Steinley, D. (2006): "K-means Clustering: a Half-Century Synthesis," *Br. J. Math. Stat. Psychol.*, 59, 1–34.

Woodcock, S. D. (2008): "Wage Differentials in the Presence of Unobserved Worker, Firm, and Match Heterogeneity," *Labour Econ.*, 15, 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 (7) we have, considering workers who move from k_r to $\widetilde{k}_{r'}$ 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 sets of M values for y_1 and y_2 that satisfy Assumption 3 *ii*). Note that one can augment those sets with a finite number of other values, including $+\infty$, while preserving the rank condition in 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 $M \times L$ with element $F_{k_r,\alpha}(y_1)$, $F^m(\widetilde{k}_{r'})$ is $M \times L$ with element $F^m_{\widetilde{k}_{r'},\alpha}(y_2)$, $D(k_r, \widetilde{k}_{r'})$ is $L \times L$ diagonal with element $p_{k_r, \widetilde{k}_{r'}}(\alpha)$, and $A^\intercal$ denotes the transpose of matrix A.

Note that $A(k_r, \widetilde{k}_{r'})$ has rank L 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 $L \times L$ 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 right-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 right-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 labeling, where $\lambda_\alpha \neq 0$ is a scale factor. As pointed out above, without loss of generality we can assume that the set of y_1 values contains $y_1 = +\infty$. This implies that λ_α is identified, so $F_{k\alpha}(y_1)$ is identified up to labeling. 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 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 L:

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

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 $F_{k',\sigma(\alpha)}$ are identified. Moreover, for each k' (possibly equal to k), using an alternating cycle as in the second part of Assumption 3 i) we obtain by a similar argument 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 L 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 (8) for the L worker types and 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 $\Pr\left[Y_{i1}\leq y_1\,|\,k_{i1}=k\right]$, 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 L, from which it follows that:

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

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) There exist finite sets of M values for y_1 and y_4 such that, 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 L, 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 (10):

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

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 density 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 $M \times L$ matrix with generic element $G^f_{y_2,k\alpha}(y_1)$, which is identified up to labeling of α, $\pi_{y_2,k}(\alpha)$ are thus identified up to the same labeling. 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 labeling. 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 labeling. Similarly, the log-earnings cdfs in all other periods can be uniquely recovered up to labeling, 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:

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

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 also recover unconditional probabilities by also using job movers in the first periods ($m_{i1} = 1$), although we do not provide details here.

B Consistency of firm classification and asymptotic distribution of parameter estimates

Firm classification. We consider a setting where the model is well-specified and there exists a partition of the J firms into K classes in the population. We focus on 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 μ is a discrete measure on $\{y_1, ..., y_D\}$, $k^0(j)$ denote firm classes in the population, H^0_k denote the population class-specific cdfs, and $\|H\|^2 = \sum_{d=1}^{D} H(y_d)^2$.

Assumption B2. *(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* $\{1, ..., K\}$, $\left\|H^0_k - H^0_{k'}\right\| > 0$.

(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 B2 (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 analyzed panel data on individuals over time as opposed to workers within firms. Parts B2 (ii) and (iii) require that the clusters be large and well-separated in the population. Assumption B2 (iv) allows for asymptotic sequences where the number of workers per firm grows polynomially more slowly than the number of firms.[52]

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 classes up to an arbitrary labeling. 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 log-earnings in the population class k (that is, the true one).[53]

Proposition B1. *Let Assumption B2 hold. Then, up to labeling of the classes k:*

(i) $\Pr\left(\widehat{k}(j) \neq k^0(j) \text{ for some } j \leq J\right) = o(1)$.

(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: $N_k = \sum_{i=1}^{N} \mathbf{1}\left\{k^0(j_{i1}) = k\right\}$.*

Proof. Note that (12) is equivalent to the following weighted 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: $n = \min_j n_j$. Their 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 B2 (i). Assumptions 2a and 2b hold by Assumptions B2 (ii) and (iii). Finally, Assumptions 2c and 2d are also satisfied by Assumption B2 (i) and boundedness of $\mathbf{1}\{Y_{i1} \leq y_1\}$. Theorems 1 and 2 in Bonhomme and Manresa (2015) and Assumption B2 (iv) then imply the result.

∎

Parameter estimates. We next consider second-step estimation of parameters. In the static model the likelihood function of log-earnings Y_i conditional on mobility m_i, firm indicators j_{i1}, j_{i2}, and population firm classes $k^0(j)$, takes the form:

$$f\left(Y_1, ..., Y_N \,|\, m_1, ..., m_N, j_{11}, j_{12}..., j_{N1}, j_{N2}, k^0(1), ..., k^0(J); \theta\right) = \prod_{i=1}^{N} f\left(Y_i \,|\, m_i, k^0(j_{i1}), k^0(j_{i2}); \theta\right),$$

where θ is a finite-dimensional vector of parameters with population value θ^0. Conditional independence follows from the assumption that worker types and idiosyncratic shocks to log-earnings are

[52] 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.

[53] While here we prove a pointwise result for $\widehat{H}_k$ the equivalence also holds uniformly in y.

independent across workers, conditionally on firm classes and mobility indicators. The likelihood function takes a similar form in the dynamic model.

Let us define the following infeasible parameter estimate:

$$\widetilde{\theta} = \underset{\theta}{\text{argmax}} \sum_{i=1}^{N} \ln f\left(Y_i \,|\, m_i, k^0(j_{i1}), k^0(j_{i2}); \theta\right).$$

Assumption B3. *(infeasible estimator)*

There is a positive-definite matrix V such that, as N tends to infinity:

$$\sqrt{N}\left(\widetilde{\theta} - \theta^0\right) \overset{d}{\to} \mathcal{N}(0, V).$$

Since $\widetilde{\theta}$ is a standard finite-dimensional maximum likelihood estimator, and observations are independent across individuals, Assumption B3 is not restrictive. Under correct specification V is the inverse of the Hessian matrix.

Let now:

$$\widehat{\theta} = \underset{\theta}{\text{argmax}} \sum_{i=1}^{N} \ln f\left(Y_i \,|\, m_i, \widehat{k}(j_{i1}), \widehat{k}(j_{i2}); \theta\right)$$

denote the second-step parameter estimate given the estimated firm classes. The following result shows that $\widehat{\theta}$ and $\widetilde{\theta}$ have the same asymptotic distribution. In practice this means that, under those assumptions, one can treat the estimated firm classes as known when computing standard errors of estimators based on them.

Proposition B2. *Let Assumptions B2 and B3 hold. Then, as N tends to infinity:*

$$\sqrt{N}\left(\widehat{\theta} - \theta^0\right) \overset{d}{\to} \mathcal{N}(0, V).$$

Proof. This is immediate since:

$$\Pr\left(\sqrt{N}\left(\widehat{\theta} - \theta^0\right) \neq \sqrt{N}\left(\widetilde{\theta} - \theta^0\right)\right) \leq \Pr\left(\widehat{k}(j) \neq k^0(j) \text{ for some } j \leq J\right),$$

which is $o(1)$ by Proposition B1. See Hahn and Moon (2010) for a similar argument.

∎

Under Proposition B2, asymptotically valid confidence intervals for θ^0 (or smooth functions of θ^0 such as variance components) can be obtained using analytical methods or the parametric bootstrap, without the need to account for the uncertainty arising from the classification. However, in our experience estimating classes tends to add finite sample noise to the parameter estimates. As an attempt to account for this finite sample variability we re-classify firms into classes in each bootstrap replication.

Lastly, here we have provided a result for a maximum likelihood estimator. Our estimator is slightly different since it is based on a sequential approach: estimating first some parameters using job movers only, and then estimating other parameters using job stayers. Asymptotic equivalence still goes through in this case, although the analytical form of the matrix V is different.

C Estimation on data from a theoretical model

In this section 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-period 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:

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

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

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

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

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

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

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

where:

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

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

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

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

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 our framework. From there we can recover our static 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:

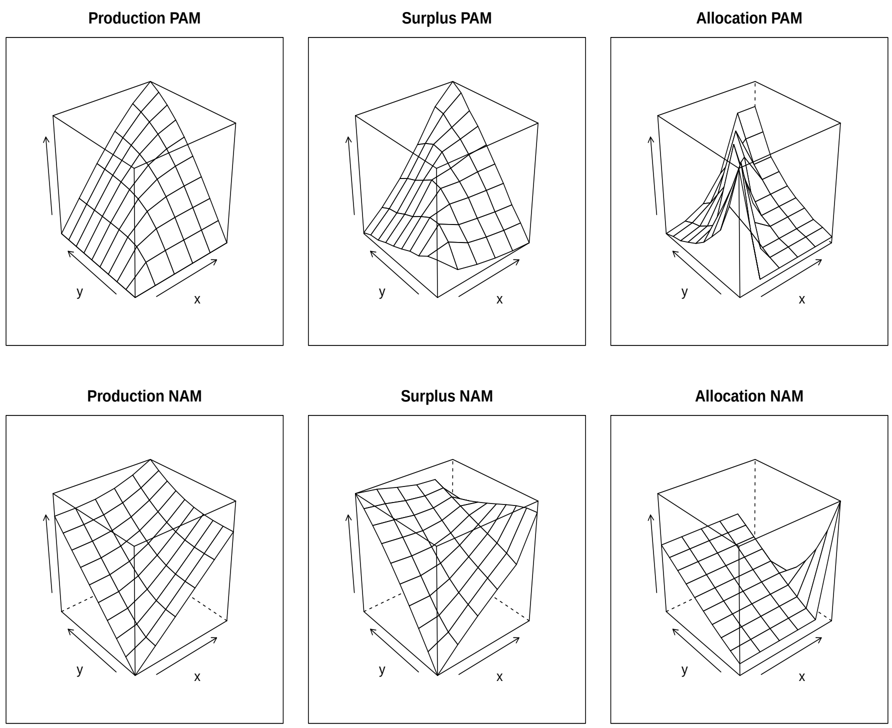
$$p_{yy'}(x) = \frac{M(x,y)M(x,y')u_{1/2}(x)}{\int M(\tilde{x},y)M(\tilde{x},y')u_{1/2}(\tilde{x})\mathrm{d}\tilde{x}}. \qquad \text{(PX-YY')}$$

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). Symmetry occurs because, in that case, 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 E4 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:

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

Figure C1: Model solutions: production, surplus and allocation



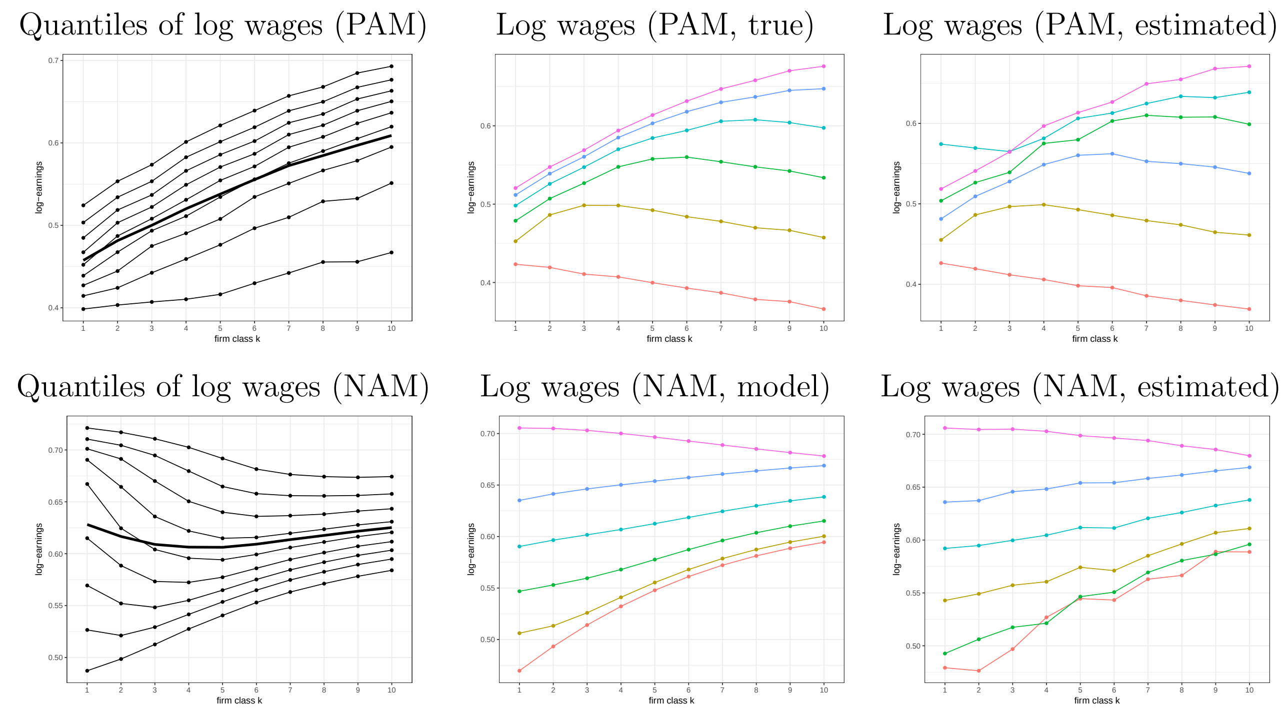
Notes: The graphs show the model solution in terms of production $f(x, y)$, surplus $S(x, y)$, and allocation $h(x, y)$. Positive assortative matching (top panel), and negative assortative matching (bottom panel).

where we set $\nu = 0.5$ and $a = 0.7$. The relative variance of measurement error is set to 10%. Finally we consider $\rho = -3$ (PAM) and $\rho = 3$ (NAM).

We simulate a sample of 500,000 individuals working in 5,000 firms, in an economy with $K = 10$ firm classes and $L = 6$ worker types. In Figure C1 we plot the model solutions, in terms of production, surplus, and allocation. On the left graph of Figure C2 we show means and quantiles of log wages in the simulated samples. We see that, while mean log wages are monotonic in firm productivity under PAM they are non-monotonic under NAM. However, as we will see, there is sufficient variation in wage distributions to separate firm classes. On the middle graph of Figure C2 we report the wage functions for the different worker types. We see clear differences between PAM and NAM. Lastly, on the right graph we show the wage functions as estimated by our static model. In estimation we use the same procedure as on the Swedish data, with $K = 10$ and $L = 6$. In particular, firm classes are estimated using k-means clustering on empirical cdfs of log wages evaluated at 20 grid points. The estimates seem to capture nonlinearities in log wages remarkably well.[54]

[54] Note that the ordering of firm classes on the x-axis is arbitrary, since the ranking of firms in terms of productivity (that is, y) is not identified using wage information only. However the variance decompositions

Figure C2: True and estimated wage functions, and quantiles of log wages in a sample simulated according to the theoretical model



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

On the first four rows of Table C1 we next report the results of variance decompositions on the samples generated according to the theoretical model, and based on estimates from our static model. We see that the decomposition is very well reproduced under both PAM and NAM. On the last four rows we report the results of an exercise where we randomly reallocate workers to firms. We see that, under PAM, the reallocation has a negative effect of mean log-earnings. This differs from the sign of the reallocation we estimated on the Swedish data. Under NAM the mean effect is also negative, though smaller. The results are again well reproduced by our static model.

Overall this set of results shows that our method is able to accurately recover the link between wages and worker-firm heterogeneity in simulated economies that feature positive or negative assortative matching.

---

and reallocation results we report below are not affected by this labeling indeterminacy.

Table C1: Variance decomposition and reallocation exercise on a sample simulated according to the theoretical model

	Variance decomposition (×100)				
	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
	PAM				
True	78.8	5.5	15.6	37.5	81.9
Estimated	79.4	5.1	15.5	38.4	81.7
	NAM				
True	107.6	12.4	-20.1	-27.4	83.9
Estimated	106.3	10.9	-17.2	-25.3	84.2

	Reallocation exercise (×100)				
	Mean	Median	10%-quantile	90%-quantile	Variance
	PAM				
True	-1.1	-.7	-1.5	-1.2	-.1
Estimated	-1.2	-.8	-1.6	-2.1	-.1
	NAM				
True	-.5	-.5	-.3	-.7	.0
Estimated	-.4	-.7	.3	-.8	.0

Notes: Variance decomposition and reallocation effects based on data generated from the theoretical sorting model with positive (PAM) or negative (NAM) assortative matching. See notes to Table 2.

D Data

We use a match of four different databases from Friedrich et al. (2014) covering the entire working age population in Sweden between 1997 and 2008. 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/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. 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 et al. (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.[55]

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. Although we conduct robustness checks, 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. Figures can be found in Table E2. 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

[55]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.

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.

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 et al. (2009), who document between-plant mobility rates ranging between 4% and 6% between 1986 and 2000.[56] 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 S1 in Supplementary Appendix S4. 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.[57] Although we prefer to focus on a more restrictive definition for estimation, as a robustness check we have also estimated the models on this larger sample, finding comparable results.

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 et al. (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 and Lehwald (2014) or Akerman et al. (2013), for example. In the US, Barth et al. (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 et al., 2009, and Akerman et al., 2013). It is important to keep these features of the Swedish labor market in mind when interpreting our results.

Finally, comparing the first two columns (or the last two columns) of Table E2 shows that job movers are on average younger and more educated than workers who remain in the same firm. They

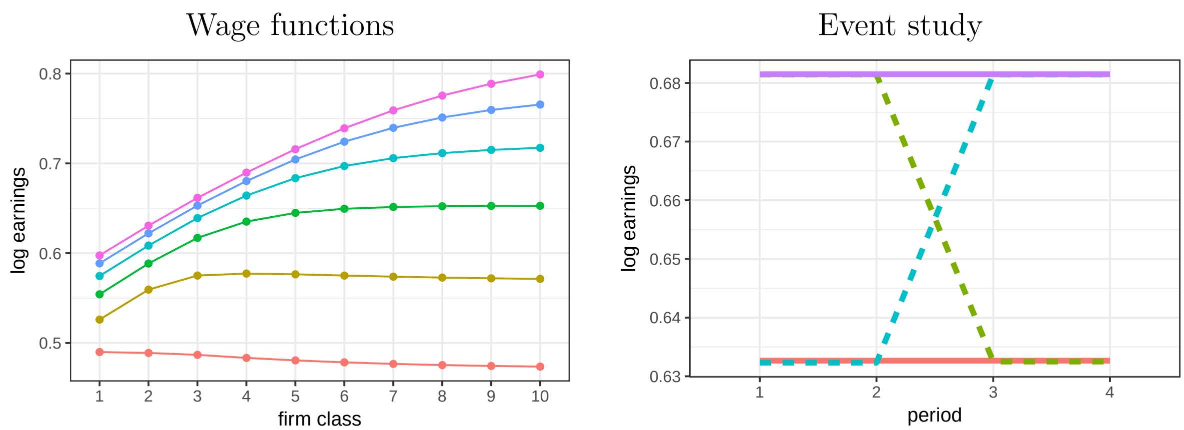
[56] See their Figure 7.14. Skans et al. (2009) report the fraction of workers employed in plants with more than 25 employees in years $t-1$ and t who changed plant between $t-1$ and t.

[57] 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.

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. At the same time, characteristics of job movers and stayers show substantial overlap.

E Additional tables and figures

Figure E3: Event study in the Shimer-Smith model in the presence of complementarities



Notes: Sample generated according to the model of Shimer and Smith (2000), without on-the-job search. Parameter values imply positive assortative matching, see Appendix C for details on the model and simulation. Left: log-wage functions for each worker type (y-axis), by firm class (x-axis). Right: log-wages of workers moving between firms within classes 4 and 10 (solid), and moving between firms between classes 4 and 10 (dashed), between periods 2 and 3.

Table E2: 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

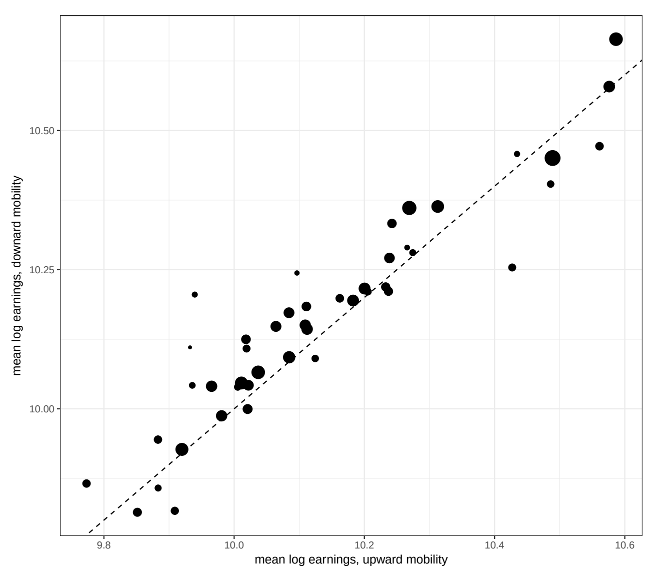
Notes: Swedish registry data. Males, fully employed in the same firm in 2002 and 2004 (columns 1 and 2), and fully employed in the same firm in 2001-2002 and 2004-2005 (columns 3 and 4), continuously existing firms. Figures for 2002. Mean log value added per worker reported for firms with positive value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

Table E3: Number of job movers between firm classes

firm class in period 1	firm class in period 2: 1	2	3	4	5	6	7	8	9	10
1	76	120	95	87	94	50	63	42	29	10
2	129	348	352	237	241	164	147	76	62	18
3	128	292	417	349	399	203	217	146	126	49
4	59	318	304	356	249	303	210	102	68	24
5	60	190	502	294	424	235	271	198	139	64
6	39	115	154	267	172	230	275	128	79	32
7	48	158	204	253	355	355	363	331	457	100
8	14	315	145	110	243	157	348	461	609	258
9	11	77	114	187	217	195	323	384	368	402
10	12	21	83	39	114	27	161	229	313	369

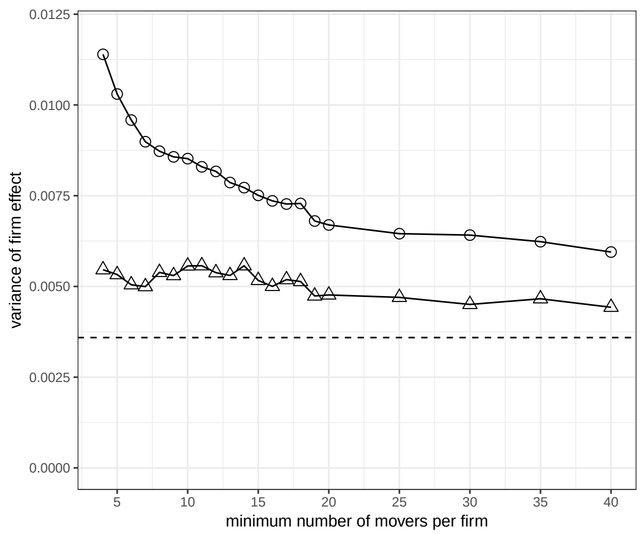
Notes: Males, fully employed in the same firm in 2002 and in 2004, continuously existing firms. Movers from firm class in 2002 (vertical axis) to firm class in period 2.

Figure E4: Earnings of job movers



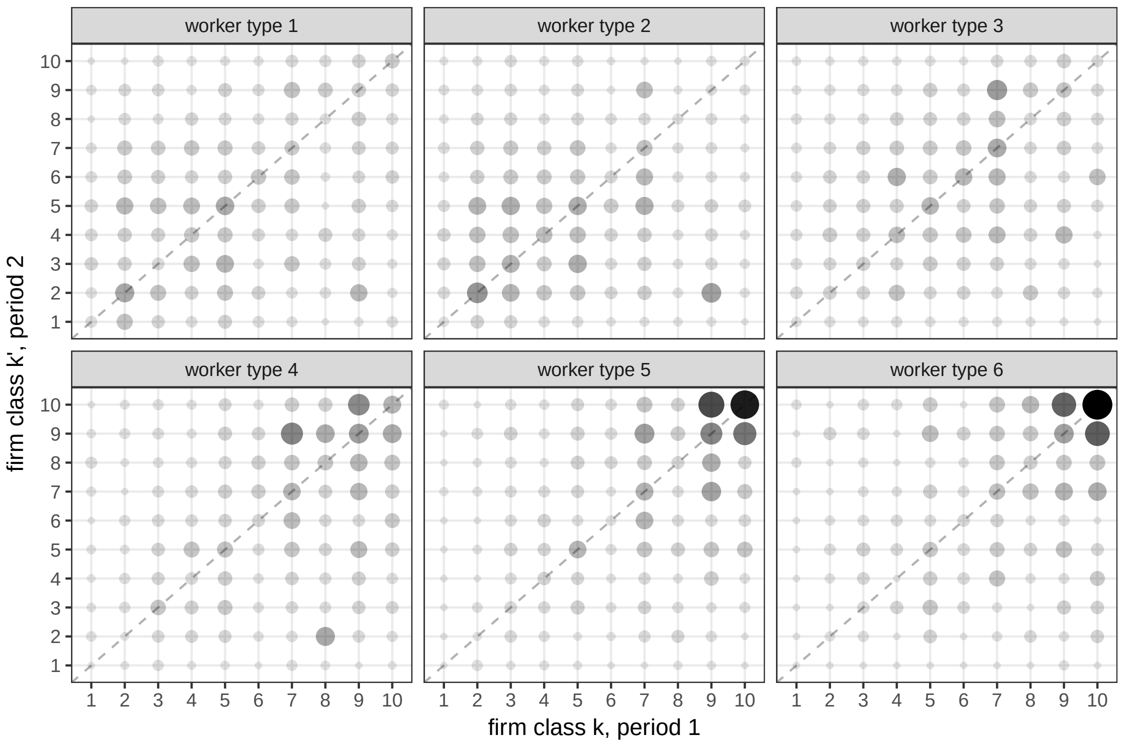
Notes: Mean log-earnings over 2002 and 2004 of movers from firm class k to firm class k' (x-axis), and of movers from k' to k (y-axis), where $k < k'$. The size of the dots is proportional to the number of job movers in the cells.

Figure E5: Fixed-effects and bias-corrected fixed-effects estimates of the variance of firm effects when varying the number of job movers per firm



Notes: The solid line with circles shows the variance of the firm fixed-effects estimates of Abowd et al. (1999), in subsamples from the 2002-2004 Swedish sample. Each point on the x-axis corresponds to selecting out firms which have less than 4 *job movers,* 5 *movers, and so on. The solid line with triangles shows half-sample bias-corrected estimates, see the text for a description. The dotted line shows our baseline estimate of the variance of firm effects, scaled by the variance of the sample with at least* 4 *movers per firm. The number of movers per firm in the whole sample is* 0.45%, *which would be outside the figure.*

Figure E6: Mobility across firm classes, dynamic model



Notes: Joint probability of firm classes in 2002 (x-axis) and 2004 (y-axis) for job movers. The size of the dots is proportional to the number of job movers in the cells. Dynamic model, 2001-2005.

Supplementary Appendix to

“A Distributional Framework for Matched Employer Employee Data”

Stéphane Bonhomme, Thibaut Lamadon and Elena Manresa

S1 Complements to the main analysis

S1.1 Accounting for worker covariates

In order to account for time-invariant worker covariates X_i we modify (14) and maximize, in the last estimation step:

$$\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbf{1}\{\widehat{k}_{i1}=k\}\ln\left(\sum_{\alpha=1}^{L}q_{k,X_i}(\alpha;\theta_q)f_{k\alpha}(Y_{i1};\widehat{\theta}_f)\right), \tag{S1}$$

with respect to θ_q, where $q_{kx}(\alpha;\theta_q)$ denotes the proportion of type α workers in class k with covariate x. In practice we use the EM algorithm to maximize (S1).

We estimate the proportions $q_{kx}(\alpha)$ in (S1) by allowing them to depend on time-invariant worker covariates X_i: education and age in the first period (9 categories). Given those probabilities we simulate $1,000,000$ observations from the model and run a linear regression of log-earnings on type indicators, class indicators, and covariates indicators. The results of the variance decomposition are shown on the third row of Table S5.

Note that this specification also allows us to distinguish sorting in terms of x from sorting in terms of unobservables. For example, for all (k,α) we can write:

$$q_k(\alpha) \quad = \quad \sum_x p(x)q_{kx}(\alpha) + \sum_x \left(p_k(x)-p(x)\right)q_{kx}(\alpha), \tag{S2}$$

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 (S2) 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 (S2) reflect the contribution of unobservables and observables, respectively, to differences in worker type composition across firm classes.

Lastly, note that it would be straightforward to also introduce observable characteristics in $f_{k\alpha}$, $f^m_{k\alpha}$, and $p_{kk'}(\alpha)$ in (13), although we do not estimate such a specification in our application on short panel data.

S1.2 Estimation of $\rho_{4|3}$ and $\rho_{1|2}$ in the dynamic model

Consider the dynamic model under the specification described in Subsection 4.2. Note that the unconditional means of log-earnings of job stayers of type α in class k are: $\mu_{1k\alpha}+\rho_{1|2}\mu^s_{2k\alpha}$, $\mu^s_{2k\alpha}$, $\mu^s_{3k\alpha}$, and $\mu_{4k\alpha}+\rho_{4|3}\mu^s_{3k\alpha}$, respectively. We make the following assumption, for all worker types α, α' and all firm classes k:

$$\begin{aligned}\mu^s_{2k\alpha'}-\mu^s_{2k\alpha} &= \mu_{1k\alpha'}+\rho_{1|2}\mu^s_{2k\alpha'}-(\mu_{1k\alpha}+\rho_{1|2}\mu^s_{2k\alpha})\\ &= \mu^s_{3k\alpha'}-\mu^s_{3k\alpha}\\ &= \mu_{4k\alpha'}+\rho_{4|3}\mu^s_{3k\alpha'}-(\mu_{4k\alpha}+\rho_{4|3}\mu^s_{3k\alpha}).\end{aligned} \tag{S3}$$

(S3) imposes that the effect of worker heterogeneity on mean log-earnings is constant over time *within firm*. On the other hand it allows for unrestricted interactions between firm classes and time.

When (S3) holds, the persistence parameters $\rho_{4|3}$ and $\rho_{1|2}$ can be estimated using simple covariance restrictions, as we now explain. The four periods' log-earnings of a job stayer of type α in class k can be written as:

$$\begin{aligned}Y_{i1} &= c_{1k}+(1-\rho_{1|2})\mu^s_{2k\alpha}+\rho_{1|2}Y_{i2}+\nu_{i1},\\ Y_{i2} &= c_{2k}+\mu^s_{2k\alpha}+\nu_{i2},\\ Y_{i3} &= c_{3k}+\mu^s_{2k\alpha}+\nu_{i3},\\ Y_{i4} &= c_{4k}+(1-\rho_{4|3})\mu^s_{2k\alpha}+\rho_{4|3}Y_{i3}+\nu_{i4},\end{aligned}$$

where ν_{i1} is independent of $(\nu_{i2},\nu_{i3},\nu_{i4})$, ν_{i4} is independent of $(\nu_{i3},\nu_{i2},\nu_{i1})$, and (taking as reference type $\alpha'=1$): $c_{1k}=\mu_{1k1}+\rho_{1|2}\mu^s_{2k1}-\mu^s_{2k1}$, $c_{2k}=0$, $c_{3k}=\mu^s_{3k1}-\mu^s_{2k1}$, and $c_{4k}=\mu_{4k1}+\rho_{4|3}\mu^s_{3k1}-\mu^s_{2k1}$.

The within-firm covariances between Y_{i1} and $Y_{i2}-Y_{i3}$ and between Y_{i4} and $Y_{i3}-Y_{i2}$ then deliver consistent estimators under standard rank conditions. As an example the model implies the panel-IV restriction $\mathrm{Cov}(Y_{i4},Y_{i3}-Y_{i2}\,|\,k)=\rho_{4|3}\,\mathrm{Cov}(Y_{i3},Y_{i3}-Y_{i2}\,|\,k)$. Notice that here $\mu^s_{2k\alpha}$ plays the role of a "fixed effect" within firm class k. In practice we combine those restrictions with all other covariance restrictions, hence also estimating the within-firm variances of the ν's and covariances of (ν_{i2},ν_{i3}) (in particular the parameter $\rho^s_{3|2}$). We estimate the parameters by minimum-distance with equal weights within firm classes, weighting each firm class according to the number of firms in the class.

S1.3 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 there 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, we assume that all matrices $F(k)$ and $F^m(k')$ have full-column rank L, for all k, k'. We also assume that the required conditions on alternating cycles are satisfied, in particular regarding the $a(\alpha)$ coefficients.

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

Likewise, $\operatorname{rank} A(2,2) = \operatorname{rank} A(2,3) = \operatorname{rank} A(3,2) = \operatorname{rank} A(3,3) = 3$. It follows that, for some $(\alpha_1, \alpha_2, \alpha_3)$, $(F_{2\alpha_1}, F_{2\alpha_2}, F_{2\alpha_3})$ and $(F_{3\alpha_1}, F_{3\alpha_2}, F_{3\alpha_3})$ 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^m_{11}, F^m_{12}), (F_{12}, F_{22}, F_{32}), $(F^m_{12}, F^m_{22}, F^m_{32})$, and so on, until $(F_{K-1,K-2}, F_{K-1,K-1}, F_{K-1,K})$, $(F^m_{K-1,K-2}, F^m_{K-1,K-1}, F^m_{K-1,K})$, $(F_{K,K-1}, F_{KK})$, and finally $(F^m_{K,K-1}, F^m_{KK})$.

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.

S1.4 Nonparametric identification for continuous worker types

Here we outline an extension where worker types α_i are continuously distributed. 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':

$$f_{kk'}(y_1, y_2) = \int f_{k\alpha}(y_1) f^m_{k'\alpha}(y_2) p_{kk'}(\alpha) d\alpha, \tag{S4}$$

where the f's are densities corresponding to the cdfs in (7). The structure of (S4) is closely related to the one analyzed in HS. Indeed, Assumption 1 implies that Y_{i1} and Y_{i2} are independent conditional on $(\alpha_i, k_{i1}, k_{i2}, m_{i1} = 1)$. 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 structure of workers' movements between firms it is possible to establish nonparametric identification using similar arguments as in the proof of Theorem 1.

To proceed with the identification argument let us define the following operators, in analogy with HS:

$$\begin{aligned}
\mathcal{L}_{kk'} g(y_1) &= \int f_{kk'}(y_1, y_2) g(y_2) dy_2, \\
\mathcal{A}_k h(y_1) &= \int f_{k\alpha}(y_1) h(\alpha) d\alpha, \\
\mathcal{B}^m_{k'} g(\alpha) &= \int f^m_{k'\alpha}(y_2) g(y_2) dy_2, \\
\mathcal{D}_{kk'} h(\alpha) &= p_{kk'}(\alpha) h(\alpha).
\end{aligned}$$

In operator form, (S4) becomes:

$$\mathcal{L}_{kk'} = \mathcal{A}_k \mathcal{D}_{kk'} \mathcal{B}^m_{k'}. \tag{S5}$$

Consider an alternating cycle of length $R = 2$. Suppose that $\mathcal{A}_k$ are $\mathcal{B}^m_{k'}$ 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 that, for all $\alpha \neq \alpha'$:

$$\frac{p_{k_1\widetilde{k}_1}(\alpha) p_{k_2\widetilde{k}_2}(\alpha)}{p_{k_1\widetilde{k}_2}(\alpha) p_{k_2\widetilde{k}_1}(\alpha)} \neq \frac{p_{k_1\widetilde{k}_1}(\alpha') p_{k_2\widetilde{k}_2}(\alpha')}{p_{k_1\widetilde{k}_2}(\alpha') p_{k_2\widetilde{k}_1}(\alpha')}. \tag{S6}$$

A condition similar to (S6) 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 analog of a rank condition. However, injectivity or completeness are high-level conditions that may be difficult to test formally (Canay et al., 2013). With $T = 2$, injectivity requires α_i to be one-dimensional.

Under these assumptions the operators $\mathcal{L}_{kk'}$, $\mathcal{A}_k$, $\mathcal{B}^m_{k'}$, and $\mathcal{D}_{kk'}$ are invertible. Moreover, analogously to HS one can show that the following spectral decomposition is unique:

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

This implies that the density $f_{k\alpha}(y_1)$ is identified up to a one-to-one transformation of α. Identification holds if there exists a known functional $\mathcal{F}$ such that $\mathcal{F}f_{k\alpha}$ is monotonic in α. As an example, identification is achieved if $\mathbb{E}[Y_{i1}\,|\,\alpha_i=\alpha,k_{i1}=k]$ is monotonic in α. In that case one may normalize worker types as $\alpha_i=\mathbb{E}[Y_{i1}\,|\,\alpha_i,k_{i1}=k]$. 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.[1]

This shows that earnings and worker type distributions can be identified based on an alternating cycle of length $R=2$. In the empirical analysis we do not attempt to estimate a nonparametric mixture model with continuous worker types, although this would be an interesting question for future work. On the other hand, note that worker types are continuous in regression models such as (1), which we estimate empirically.

S1.5 A bi-clustering classification method

Here we describe a classification approach which consists in clustering firms based on longitudinal information. Focusing on the static model on two periods we have, by (7) and due to the class-specific nature of firm heterogeneity:

$$\Pr\left[Y_{i1}\leq y_1,Y_{i2}\leq y_2\,|\,j_{i1}=j,j_{i2}=j',m_{i1}=1\right] \;=\; \int F_{k\alpha}(y_1)F^m_{k'\alpha}(y_2)p_{kk'}(\alpha)d\alpha, \tag{S7}$$

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:

$$\min_{k(1),\ldots,k(J),H_{11},\ldots,H_{KK}} \sum_{i=1}^{N_m}\int\int \left(\mathbf{1}\{Y_{i1}\leq y_1\}\mathbf{1}\{Y_{i2}\leq y_2\}-H_{k(j_{i1}),k(j_{i2})}(y_1,y_2)\right)^2 d\mu(y_1,y_2), \tag{S8}$$

for a bivariate measure μ, where the first N_m individuals in the sample are the job movers between periods 1 and 2.[2] Algorithms to solve (S8) have been comparatively less studied than k-means classification problems such as (12); see Bonhomme (2017) for references.

Under discrete worker heterogeneity, the separation condition for consistency of classification in (S8) is weaker than in the cross-sectional case of Assumption B2 (*iii*). To see this, let $G_{kk'}$ denote

[1]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 labeling of the latent types.

[2]Job stayers can be added to (S8), under suitable assumptions on within-job earnings dynamics. Also, the two objective functions in (12) and (S8) can be combined, thus incorporating both cross-sectional and longitudinal information into the classification.

the bivariate cdf on the left-hand side of (S7). In (S7) 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 B2 (*iii*).

Corollary S1. *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 L 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 L. 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 L we obtain that $F^m_{k\alpha} = F^m_{k'\alpha}$. ∎

Corollary S1 implies that, if type-specific earnings distributions are identified, then some of the G's must differ. In that case the bi-clustering method will reveal firm classes asymptotically. Hence, information from the earnings sequences of job movers can allow one to identify firm classes even when the cross-sectional earnings information is insufficient.

S1.6 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 (12). In the second period, one can estimate the period-specific classification by solving the following k-means problem:

$$\min_{k_2(1),\dots,k_2(J),H_{11},\dots,H_{KK}} \sum_{j=1}^{J} n_j \int \left(\widehat{F}_{2j}(y) - H_{\widehat{k}_1(j),k_2(j)}(y) \right)^2 d\mu(y), \tag{S9}$$

where $\widehat{F}_{2j}$ denotes the log-earnings cdf in period 2, and $\widehat{k}_1(j)$ are estimates from (12). This may be iterated until the last period of the panel.[3]

[3]Alternatively, a multi-clustering approach may be used, as in (S8).

S1.7 Estimation on T periods

Here we outline estimation in models with T periods. We focus on the dynamic model, given that the static model is a particular case. Consider the dynamic model with a finite mixture specification for worker types. The classification step is as in (12). 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 different components 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: $\Pr\left[\alpha_i = \alpha \,|\, k_{i1}, X_{i1}; \theta_1\right]$.
- Initial condition, log-earnings: $\Pr\left[Y_{i1} \leq y_1 \,|\, \alpha_i, k_{i1}, X_{i1}; \theta_2\right]$.
- Transitions, mobility: $\Pr\left[m_{it} = m \,|\, Y_{it}, \alpha_i, k_{it}, X_{it}, t; \theta_3\right]$.
- Transitions, classes: $\Pr\left[k_{i,t+1} = k' \,|\, Y_{it}, \alpha_i, k_{it}, X_{it}, m_{it} = 1, t; \theta_4\right]$.
- Transitions, log-earnings: $\Pr\left[Y_{i,t+1} \leq y_{t+1} \,|\, Y_{it}, \alpha_i, k_{i,t+1}, k_{it}, X_{i,t+1}, m_{it} = m, t; \theta_5\right]$.

S2 Computation

In this section we provide details on the computation procedure we use to estimate the model and compute bootstrap replications and model simulations.

S2.1 Estimation of parameters

Given the presence of local optima in our finite mixture model, the choice of initial conditions and exploration strategy is important.[4] We next describe how we explore the likelihood function to obtain our baseline estimates of log-earnings in the static model, based on job movers. The final step, in which we estimate worker type proportions in the cross-section, is numerically well-behaved as it is based on a concave objective. We use similar exploration strategies in the dynamic case.

[4]The computational challenges in the estimation of our mixture models motivated us to also develop interactive regression models, which do not suffer from local optima issues and are straightforward to compute. In Section S3 we provide details on the identification and estimation of interactive regressions. We use these estimators as robustness checks for our main results.

Exploration of the likelihood function Although we have experimented with different choices of starting values, we here describe the strategy that, in our experience, has consistently yielded higher likelihood values. Our estimates (and estimates within the bootstrap) are based on 50 starting values.

To obtain a starting value we first draw L wages from a Gaussian distribution with mean equal the mean of log-earnings in period 1 and standard deviation equal twice the standard deviation of log-earnings in the same period. Using the EM algorithm, holding mean log-earnings fixed across firms, we then compute estimates of proportions of worker types and type-and-class-specific log-earnings variances. We use these estimates as starting values for another preliminary estimation where mean log-earnings are held constant across firms, and are estimated jointly with log-earnings variances and type proportions. The estimates obtained with this second estimation are then used as initial conditions for our estimation based on job movers. The resulting parameter estimates are then used in the final estimation step based on job stayers.

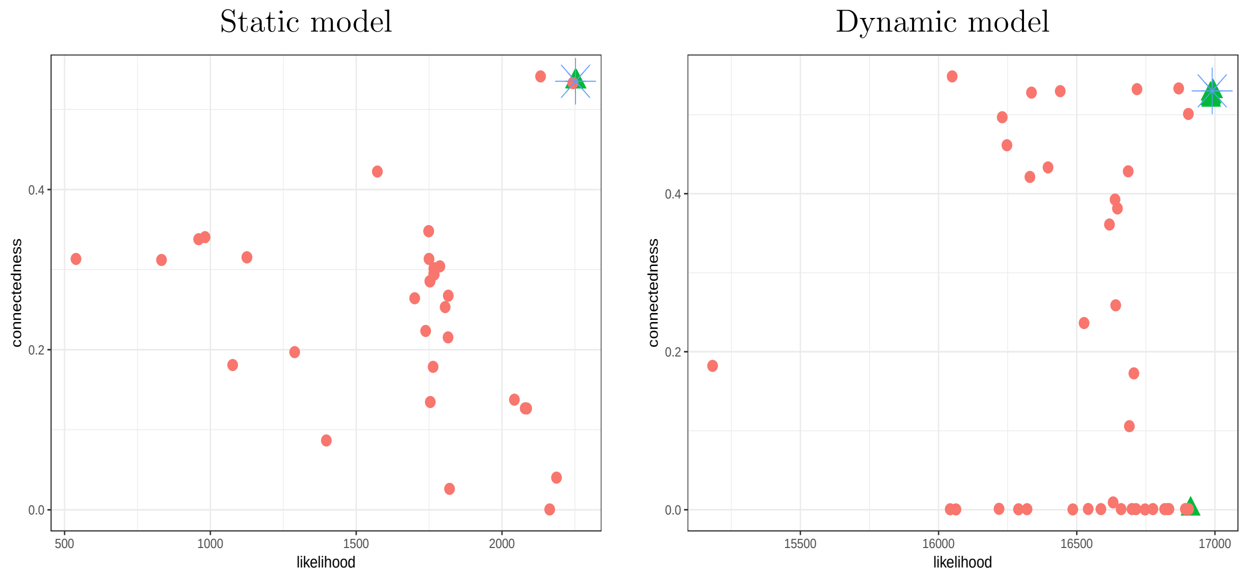
Graph connectedness. Our identification results emphasize the importance of connectedness, through the presence of alternating cycles for each worker type. The mobility patterns of workers define a graph across firm classes. A measure of connectedness of a graph is the smallest non-zero eigenvalue of its normalized Laplacian, as recently studied in Jochmans and Weidner (2017).[5] We observed that the local optima of the likelihood function tended to vary substantially in terms of their connectedness, some of the solutions having types with very low connectedness. To discriminate between estimates that have similar likelihood values we favor estimates with higher connectedness. Our main estimates are based on the best connected solution out of the ones yielding the 10 highest likelihood values.

On the left panel of Figures S1 we plot the likelihood value against the connectedness measure for the static model. In this case the solutions yielding the highest likelihood values (depicted as triangles in the figure) coincide with the one showing highest connectedness (the star). On the right panel we show the same relationship for the dynamic model. In this case there is more uncertainty about the exact location of the highest likelihood value. We see that our solution (the star) not only has high likelihood but also high connectedness.[6]

[5]Empirically, we measure connectedness as the minimum, across all worker types, of the smallest non-zero eigenvalues of the normalized Laplacians of the type-specific graphs (weighted by number of movers), where the graphs are at the firm-class level.

[6]In both cases the solutions using the maximum likelihood estimates are very similar to the ones we report.

Figure S1: Likelihood and connectedness of locally optimal solutions



Notes: The dots show the likelihood values (x-axis) and connectedness measures (y-axis) corresponding to all local optima of the job mover part of the likelihood function starting at each of the 50 starting values. The triangles show the 10 best likelihood values. The stars show our selected values.

Estimation of persistence parameters in the dynamic model. The covariance structure estimation of $\rho_{4|3}$ and $\rho_{1|2}$, which we describe in Section S1.2, is numerically well-behaved. In Figure S1.2 we plot the shape of the objective function in our baseline specification of the dynamic model, marginalized with respect to both parameters.

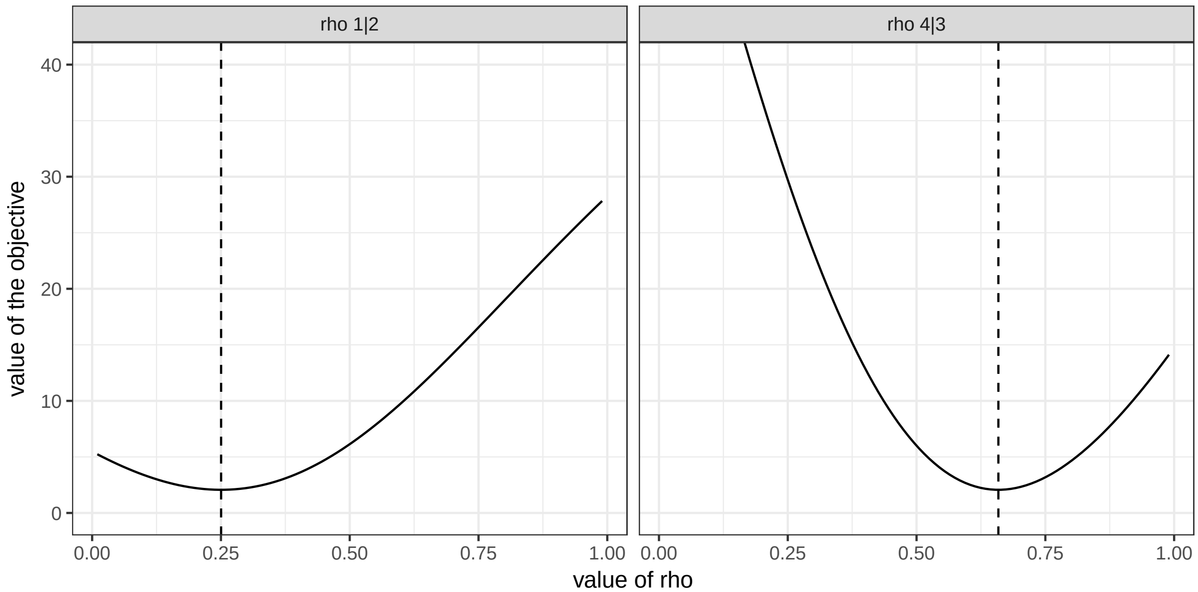
S2.2 Model simulation

We use simulations for two different purposes: to compute variance decompositions, reallocation exercises and dynamic effects, and to draw simulated samples in the parametric bootstrap. Here we explain the simulation strategy for the bootstrap. The simulation algorithm for the decompositions and reallocations is related but simpler since the identity of the firm is irrelevant in that case.[7]

The simulation for the bootstrap is conditional on firm classes and the mobility links between firms and workers, including the size of firms. We describe the simulation algorithm for the static model,

[7] In particular, to compute the frequencies in Table 5 we draw mobility status and firm classes in both periods by sampling with replacement from their empirical distribution, and then draw worker types and earnings realizations conditional on those values.

Figure S2: Estimation of ρ parameters



Notes: Using the full covariance structure for job stayers described in Section S1.2, we estimate $\rho_{4|3}$ and $\rho_{1|2}$. The figure plots the value of the objective function as each ρ deviates from the optimal value. Left is $\rho_{1|2}$; right is $\rho_{4|3}$.

the case of the dynamic model being analogous.

1. (Job stayers) For each firm in a given class we first draw, independently, the latent types of job stayers in the firm according to the distribution of types in the class.

   Given a worker of a given type and a firm of a given class, log-earnings are then independently drawn across workers from the corresponding conditional distribution. In the baseline specification this distribution is Gaussian with class-and-type-specific mean and variance.

2. (Job movers) For each pair of firms in periods 1 and 2 in given classes, the latent types of job movers between those firms are drawn according to the their distribution in the pair of classes.

   Given a worker of a given type and a pair of firms of given classes, log-earnings in periods 1 and 2 are drawn according to their conditional distribution. In the static model log-earnings are drawn independently across periods.

Given a simulated sample, we then estimate the parameters as described above. In particular, firm classes are re-estimated in each bootstrap replication using k-means, with 500 starting values.

S3 Interactive regression models

In this section we study identification and estimation in regression models, (1) and (2), which feature worker-firm interactions in log-earnings.

S3.1 Models and identification

S3.1.1 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:

$$\mathbb{E}\left[Z_i\left(\tau_2(k_{i2})Y_{i2} - \tau_1(k_{i1})Y_{i1} - \widetilde{a}_2(k_{i2}) + \widetilde{a}_1(k_{i1}) - X'_{i2}\widetilde{c}_2(k_{i2}) + X'_{i1}\widetilde{c}_1(k_{i1})\right) \,|\, m_{i1} = 1\right] = 0, \tag{S10}$$

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.[8]

Note that (S10) is linear in parameters. Linearity is important 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 (S10), with d_x denoting the dimension of X_{it}. The order condition for identification in (S10) requires $K \geq 4$. We have the following result.

Theorem S1. *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.

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

Proof. Part (i). If A has maximal rank then (S10) 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.[9] Identification of the means of α_i conditional on $m_{i1}=0$ or $m_{i1}=1$ then follows directly. For example, we have:

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

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

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

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

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

is thus identified.

∎

S3.1.2 Dynamic model

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

$$\begin{aligned} Y_{it} &= a_t^s(k)+b_t(k)\alpha_i+\varepsilon_{it}, \quad t=1,...,4, \\ & \text{if} \quad m_{i1}=0,\, m_{i2}=0,\, m_{i3}=0, \end{aligned} \tag{S11}$$

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

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

a firm of class k' in periods 3 and 4. We specify their log-earnings as follows:

$$\begin{aligned}
Y_{i1} &= a_1^s(k) + \rho_{1|2}(a_2^m(k) - a_2^s(k)) + \rho_{1|2}\xi_2(k') + b_1(k)\alpha_i + \varepsilon_{i1}, \\
Y_{i2} &= a_2^m(k) + \xi_2(k') + b_2(k)\alpha_i + \varepsilon_{i2}, \\
Y_{i3} &= a_3^m(k') + \xi_3(k) + b_3(k')\alpha_i + \varepsilon_{i3}, \\
Y_{i4} &= a_4^s(k') + \rho_{4|3}(a_3^m(k') - a_3^s(k')) + \rho_{4|3}\xi_3(k) + b_4(k')\alpha_i + \varepsilon_{i4}, \\
&\quad \text{if} \quad m_{i1} = 0,\, m_{i2} = 1,\, m_{i3} = 0,
\end{aligned} \tag{S12}$$

where "m" stands for "movers". In (S11) and (S12) 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. Likewise, one can let the b_t's differ between stayers in (S11) and movers in (S12), 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 (S11)-(S12), 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 (S12), 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, the 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 (S12) we have, for job movers between periods 2 and 3:

$$\begin{aligned} Y_{i1} - \rho_{1|2}Y_{i2} &= a_1^s(k) - \rho_{1|2}a_2^s(k) + \left[b_1(k) - \rho_{1|2}b_2(k)\right]\alpha_i + \varepsilon_{i1} - \rho_{1|2}\varepsilon_{i2}, \\ Y_{i4} - \rho_{4|3}Y_{i3} &= a_4^s(k') - \rho_{4|3}a_3^s(k') + \left[b_4(k') - \rho_{4|3}b_3(k')\right]\alpha_i + \varepsilon_{i4} - \rho_{4|3}\varepsilon_{i3}. \end{aligned} \tag{S13}$$

Equation (S13) has a similar structure as the static model on two periods. As a result, one can derive moment restrictions analogous to (S10). 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 (S13), 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 S2. *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 S1 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 $\mathbb{E}(\alpha_i \mid k_{i2}, k_{i3}, m_{i2})$, $\mathrm{Var}(\alpha_i \mid k_{i2}, k_{i3}, m_{i2})$, $\mathrm{Var}\left(\varepsilon_{i1} - \rho_{1|2}\varepsilon_{i2} \mid k_{i2} = k\right)$, and $\mathrm{Var}\left(\varepsilon_{i4} - \rho_{4|3}\varepsilon_{i3} \mid k_{i3} = k\right)$, are all identified.

(ii) If, in addition to (i), the indicators $\mathbf{1}\{k_{i2} = k\}$, $\mathbf{1}\{k_{i3} = k'\}$, and the products $\mathbf{1}\{k_{i2} = k\} \times \mathbb{E}(\alpha_i \mid k_{i2}, k_{i3}, m_{i2} = 1)$ 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 S1.

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:

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

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

$V_{3kk'} = \text{Var}\left(\varepsilon_{i3}|k_{i2}=k, k_{i3}=k', m_{i2}=1\right)$. Hence:

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

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

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

from which we recover $V_{2kk'}$, $C_{23kk'}$ and $V_{2kk'}$. The variances $\text{Var}\left(\varepsilon_{i1}|k_{i2}=k, k_{i3}=k', m_{i2}=1\right)$ and $\text{Var}\left(\varepsilon_{i4}|k_{i2}=k, k_{i3}=k', m_{i2}=1\right)$ 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, as in Section S1.2. 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 $\rho_{4|3}$:

$$\text{Cov}\left(Y_{i1}-\rho_{1|2}Y_{i2}, Y_{i2}-Y_{i3} \,|\, k_{i1}=k_{i2}=k_{i3}=k_{i4}=k, m_{i1}=m_{i2}=m_{i3}=0\right)=0,$$
$$\text{Cov}\left(Y_{i4}-\rho_{4|3}Y_{i3}, Y_{i3}-Y_{i2} \,|\, k_{i1}=k_{i2}=k_{i3}=k_{i4}=k, m_{i1}=m_{i2}=m_{i3}=0\right)=0. \qquad \text{(S14)}$$

These are familiar covariance restrictions in autoregressive models with fixed-effects. For example, the second equation in (S14) 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 $\text{Cov}\left(Y_{i3}, Y_{i3}-Y_{i2}\,|\,k_{i1}=k_{i2}=k_{i3}=k_{i4}=k, m_{i1}=m_{i2}=m_{i3}=0\right)\neq 0$. 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 (S14) 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 S2, $b_1^m(k)-\rho_{1|2}b_2^m(k)$ and $b_4^m(k')-\rho_{4|3}b_3^m(k')$ are identified, together with $\mathbb{E}\left(\alpha_i\,|\,k_{i2},k_{i3},m_{i2}\right)$. 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 S2 *(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 $\mathbb{E}\left(\alpha_i \,|\, k_{i2}=k, m_{i2}=0\right)$ are identified from Theorem S2. 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:

$$\begin{aligned}
&\mathrm{Var}\left(\left(\begin{array}{c} Y_{i1}-\rho_{1|2}Y_{i2} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4}-\rho_{4|3}Y_{i3} \end{array}\right) \middle| k_{i2}=k_{i3}=k, m_{i2}=0\right) \\
&=\left(\begin{array}{c} b_1^s(k)-\rho_{1|2}b_2^s(k) \\ b_2^s(k) \\ b_3^s(k) \\ b_4^s(k)-\rho_{4|3}b_3^s(k) \end{array}\right) \times \mathrm{Var}\left(\alpha_i \,|\, k_{i2}=k_{i3}=k, m_{i2}=0\right) \times \left(\begin{array}{c} b_1^s(k)-\rho_{1|2}b_2^s(k) \\ b_2^s(k) \\ b_3^s(k) \\ b_4^s(k)-\rho_{4|3}b_3^s(k) \end{array}\right)' \\
&\quad + \left(\begin{array}{cccc} \mathrm{Var}\left(\varepsilon_{i1}-\rho_{1|2}\varepsilon_{i2}|k_{i2}=k\right) & 0 & 0 & 0 \\ 0 & V_{2k}^s & C_{23k}^s & 0 \\ 0 & C_{23k}^s & V_{3k}^s & 0 \\ 0 & 0 & 0 & \mathrm{Var}\left(\varepsilon_{i4}-\rho_{4|3}\varepsilon_{i3}|k_{i3}=k\right) \end{array}\right), \qquad \text{(S15)}
\end{aligned}$$

where: $V_{2k}^s = \mathrm{Var}\left(\varepsilon_{i2}|k_{i2}=k_{i3}=k, m_{i2}=0\right)$, $C_{23k}^s = \mathrm{Cov}\left(\varepsilon_{i2}, \varepsilon_{i3}|k_{i2}=k_{i3}=k, m_{i2}=0\right)$, and $V_{3k}^s = \mathrm{Var}\left(\varepsilon_{i3}|k_{i2}=k_{i3}=k, m_{i2}=0\right)$. 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 (S15) 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 (S15), and for movers in the following system:

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

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

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

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

S3.2 Estimation algorithms in interactive regression models

Consider the static regression model (1) on two periods. The mean restrictions in (S10) 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.[10]

The identification results in Theorem S1 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 (S10), and estimate intercepts, slopes, and coefficients associated with observables using LIML, see part (i) in Theorem S1. 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

[10] Specifically, let us write the moment restrictions in (S10) as $\mathbb{E}(Z_i' W_i \beta) = 0$, and stack the Z_i and W_i into matrices Z and W, respectively. Then the LIML estimator is given, up to scale, by:

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

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

(*ii*) and (*iii*) in Theorem S1. The latter restrictions are also linear in parameters, so computation is straightforward.[11]

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

S3.2.1 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 (S10) with scale and location normalizations.[12] 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:
$$\frac{1}{2}\sum_{t=1}^{2} \widehat{\tau}_t(\widehat{k}_{it})Y_{it} - \widehat{\widetilde{a}}_t(\widehat{k}_{it}) - X'_{it}\widehat{\widetilde{c}}_t(\widehat{k}_{it}) \tag{S16}$$
given $\widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1$. Estimate $\widehat{\mu}_k$ as the mean of (S16) 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 $\mathrm{Var}\left(\alpha_i \,|\, k_{i1} = k, k_{i2} = k', m_{i1} = 1\right) = v^m_{kk'}$, $\mathrm{Var}\left(\varepsilon_{i1} \,|\, k_{i1} = k\right) = V_{1k}$, and $\mathrm{Var}\left(\varepsilon_{i2} \,|\, k_{i2} = k\right) = V_{2k}$ by minimizing:
$$\sum_{k=1}^{K}\sum_{k'=1}^{K} N^m_{kk'} \left\| \begin{pmatrix} \widehat{\mathrm{Var}}\left(\widetilde{Y}_{i1} \,|\, \widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1\right) \\ \widehat{\mathrm{Cov}}\left(\widetilde{Y}_{i1}, \widetilde{Y}_{i2} \,|\, \widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1\right) \\ \widehat{\mathrm{Var}}\left(\widetilde{Y}_{i2} \,|\, \widehat{k}_{i1} = k, \widehat{k}_{i2} = k', m_{i1} = 1\right) \end{pmatrix} - \begin{pmatrix} \widehat{b}_1(k)^2 v^m_{kk'} + V_{1k} \\ \widehat{b}_1(k)\widehat{b}_2(k')v^m_{kk'} \\ \widehat{b}_2(k')^2 v^m_{kk'} + V_{2k'} \end{pmatrix} \right\|^2,$$

[11] 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.

[12] 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.

subject to all $v^m_{kk'} \geq 0$, $V_{1k} \geq 0$, $V_{2k'} \geq 0$, where $\widehat{\text{Var}}$ and $\widehat{\text{Cov}}$ denote empirical variances and covariances, respectively, and $N^m_{kk'}$ denotes the number of job movers from $\widehat{k}_{i1} = k$ to $\widehat{k}_{i2} = k'$. Lastly, estimate $\text{Var}\left(\alpha_i \,|\, k_{i1} = k\right) = v_k$ by minimizing:

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

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.

S3.2.2 Dynamic case

We next consider estimation in the dynamic interactive model on four periods (S11)-(S12). 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:

$$Q(\rho_1, \rho_4) = \min \sum_{k=1}^{K} N^s_k \left|\left| \begin{pmatrix} \widehat{\text{Var}}\left(Y_{i1} - \rho_1 Y_{i2} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\text{Cov}}\left(Y_{i1} - \rho_1 Y_{i2}, Y_{i2} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\text{Cov}}\left(Y_{i1} - \rho_1 Y_{i2}, Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\text{Cov}}\left(Y_{i1} - \rho_1 Y_{i2}, Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\text{Cov}}\left(Y_{i2}, Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\text{Cov}}\left(Y_{i3}, Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \\ \widehat{\text{Var}}\left(Y_{i4} - \rho_4 Y_{i3} \,|\, \widehat{k}_{i2} = \widehat{k}_{i3} = k, m_{i2} = 0\right) \end{pmatrix} - \begin{pmatrix} (1-\rho_1)^2 \widetilde{v}^s_k + V_{1k} \\ (1-\rho_1)\widetilde{v}^s_k \\ (1-\rho_1)\widetilde{v}^s_k \\ (1-\rho_1)(1-\rho_4)\widetilde{v}^s_k \\ (1-\rho_4)\widetilde{v}^s_k \\ (1-\rho_4)\widetilde{v}^s_k \\ (1-\rho_4)^2 \widetilde{v}^s_k + V_{4k} \end{pmatrix} \right|\right|^2, \qquad \text{(S17)}$$

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:

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

- 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 $\widetilde{\tau}_1(k) = 1/\widehat{d}_1(k)$, $\widetilde{\tau}_4(k) = 1/\widehat{d}_4(k)$, $\widehat{\widetilde{c}}_1(k) = \widehat{c}_1(k)/\widehat{d}_1(k)$, and $\widehat{\widetilde{c}}_4(k) = \widehat{c}_4(k)/\widehat{d}_4(k)$ using LIML based on:

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

imposing scale and location normalizations and replacing $\rho_{1|2}$ and $\rho_{4|3}$ by $\widehat{\rho}_{1|2}$ and $\widehat{\rho}_{4|3}$, and Z_i by $\widehat{Z}_i$. 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 $\mu_{kk'}^m = \mathbb{E}(\alpha_i \,|\, k_{i2} = k, k_{i3} = k', m_{i2} = 1)$. Estimate $\widehat{\mu}_{kk'}^m$ as the mean of:

$$\frac{1}{2}\left(\widehat{\tau}_1(\widehat{k}_{i2})\left(Y_{i1} - \widehat{\rho}_{1|2} Y_{i2}\right) - \widehat{\widetilde{a}}_1(\widehat{k}_{i2}) + \widehat{\tau}_4(\widehat{k}_{i3})\left(Y_{i4} - \widehat{\rho}_{4|3} Y_{i3}\right) - \widehat{\widetilde{a}}_4(\widehat{k}_{i3})\right) \tag{S18}$$

given $\widehat{k}_{i2} = k, \widehat{k}_{i3} = k', m_{i2} = 1$. Let $\mu_k^s = \mathbb{E}(\alpha_i \,|\, k_{i2} = k_{i3} = k, m_{i2} = 0)$. Estimate $\widehat{\mu}_k^s$ as the mean of (S18) 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:

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

which is a quadratic objective function.

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

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

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

- Let $V^{\alpha s}_k = \text{Var}\left(\alpha_i \,|\, k_{i2}=k_{i3}=k, m_{i2}=0\right)$, $V^s_{2k} = \text{Var}\left(\varepsilon_{i2}|k_{i2}=k_{i3}=k, m_{i2}=0\right)$, $C^s_{23k} = \text{Cov}\left(\varepsilon_{i2}, \varepsilon_{i3}|k_{i2}=k_{i3}=k, m_{i2}=0\right)$, and $V^s_{3k} = \text{Var}\left(\varepsilon_{i3}|k_{i2}=k_{i3}=k, m_{i2}=0\right)$. Estimate

$\widehat{V}_k^{\alpha s}$, $\widehat{V}_{2k}^s$, $\widehat{C}_{23k}^s$, and $\widehat{V}_{3k}^s$ by minimizing:

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

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 $\text{Cov}\,(\varepsilon_{i3}, \varepsilon_{i4} \,|\, k_{i2}{=}k_{i3}{=}k, m_{i2}{=}0)$.

S3.2.3 Stationary interactions and profiling

Estimating complementarities based on mean restrictions alone may lead to unstable parameters when the conditions for identification are (nearly) violated in the data. Consider as an example the static model (1). Identification relies on log-earnings in period 1 varying with future firm class conditional on the current class, and analogously for log-earnings in period 2. As shown by Figure S14, there appears to be more underlying firm variation in period 2 log-earnings than in period 1 log-earnings. In our experience on the Swedish data this causes some instability in the estimation of interaction parameters $b_1(k)$ corresponding to the first period. To overcome this issue we impose that interaction parameters $b_t(k)$ do not vary with time t, while allowing for class-and-time-specific intercepts. In our experience, orthogonally projecting log-earnings in period 1 and exploiting the variation from period 2 log-earnings led to reliable results while preserving computational feasibility. An interesting question for future work will be to study the performance of an estimator of the $b_t(k)$ parameters that combines mean and covariance restrictions.

S3.3 Interactive models on T periods

In this last part we consider the dynamic regression 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 focus on 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 $s > 0$:

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

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:

$$\begin{aligned} W_{it}'\gamma_t &= a_{tt}(k_{it}, m_{i,t-1}, m_{it}) + X_{it}'c_t \\ &\quad + \sum_{s=1}^{T-t} \Big(\rho_{t|t+s} a_{t+s,t+s}(k_{i,t+s}, m_{i,t+s-1}, m_{i,t+s}) + \rho_{t|t+s-1} \xi^f_{t+s}(k_{i,t+s}, m_{i,t+s-1}) \Big) \\ &\quad + \sum_{s=1}^{t-1} \Big(\rho_{t|t-s} a_{t-s,t-s}(k_{i,t-s}, m_{i,t-s-1}, m_{i,t-s}) + \rho_{t|t-s-1} \xi^b_{t-s}(k_{i,t-s}, m_{i,t-s}) \Big). \end{aligned}$$

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.

S4 Additional empirical results

Table S1: 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. Number of firms $\geq 10, 50$ *is according to reported firm size. Figures for 2002.*

Table S2: Variance decomposition (×100), static model, varying the number of firm classes

	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
$K=3$	86.8	1.8	11.4	45.3	75.9
$K=4$	84.9	2.3	12.7	45.3	76.0
$K=5$	83.4	2.7	13.9	46.5	75.9
$K=6$	82.7	2.7	14.5	48.2	75.4
$K=7$	81.9	3.1	15.0	47.0	76.9
$K=8$	82.4	3.0	14.6	46.3	76.6
$K=9$	82.2	2.9	14.7	46.0	76.4
$K=10$	80.3	3.4	16.3	49.1	74.8
$K=12$	82.3	3.0	14.7	46.6	76.5
$K=15$	80.1	3.6	16.2	47.4	75.1
$K=20$	79.2	4.0	16.8	47.4	72.7

Notes: Static model, 2002-2004. α is the worker effect, ψ is the firm effect. Variance decomposition based on a linear regression of simulated 2002 log-earnings on α and ψ. $1,000,000$ simulations. The number of worker types is fixed to $L=6$.

Table S3: Variance decomposition (×100), static model, refined firm classification

	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
	Within-cluster splits				
by mobility rank	82.7	3.8	13.5	38.1	70.7
by percent of movers	87.1	1.9	11.0	42.5	77.1
by value added	77.7	4.7	17.6	46.0	71.3
	Likelihood-based re-classification				
iterated once	80.9	3.4	15.7	47.1	75.9
iterated 5 times	81.6	3.7	14.7	42.4	77.3

Notes: Static model, 2002-2004. α is the worker effect, ψ is the firm effect. Variance decomposition based on a linear regression of simulated 2002 log-earnings on α and ψ. $1,000,000$ simulations.

Table S4: Variance decomposition (×100), static model, varying the number of worker types

	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
$L=3$	34.5	34.3	31.2	45.4	39.2
$L=4$	47.6	20.9	31.6	50.1	46.3
$L=5$	77.6	4.5	17.9	47.9	71.5
$L=6$	80.3	3.4	16.3	49.1	74.8
$L=7$	83.6	2.6	13.8	46.7	77.7
$L=8$	80.0	3.7	16.3	47.6	76.0
$L=9$	76.2	5.0	18.8	48.2	69.9

Notes: See the notes to Table S2. The number of firm classes is fixed to $K=10$.

Table S5: Variance decomposition (×100), static model, other specifications

	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
	A. Mixture models				
baseline	80.3	3.4	16.3	49.1	74.8
mixture-of-mixtures	82.1	3.1	14.8	46.4	76.2
net of age and education	67.4	9.0	23.6	47.8	61.1
firms with ≤ 50 workers	71.5	5.4	23.1	59.1	73.6
firms with > 50 workers	85.1	3.7	11.2	31.6	70.6
fully nonstationary	79.9	3.7	16.5	48.3	75.7
random splits: first split	81.9	3.0	15.1	47.9	76.3
random splits: second split	82.0	3.0	15.0	48.0	76.9
	B. Regression models				
interactive regression model	81.4	3.0	15.6	50.2	69.4
linear regression model	83.7	2.4	13.8	48.5	72.4

Notes: Static model, 2002-2004. α is the worker effect, ψ is the firm effect. Variance decomposition based on a linear regression of 2002 log-earnings on α and ψ. $1,000,000$ simulations. "Mixture-of-mixtures" refers to a nonstationary model where, conditional on firm classes and worker types, log-earnings are distributed as three-component mixtures of Gaussians. The interactive regression model is given by (1), with $c_t = 0$ and $b_t(k)$ constant over time. The linear regression model has in addition $b_t(k) = 1$ for all k. Identification and estimation of the regression models is detailed in Section S3.

Table S6: Variance decomposition and reallocation exercise, static model, parametric bootstrap

	Variance decomposition (×100)				
	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
estimate	80.3	3.4	16.3	49.1	74.8
bootstrap mean	81.7	2.7	15.6	52.1	75.0
bootstrap 2.5%-quantile	80.1	2.3	14.4	50.4	73.9
bootstrap 97.5%-quantile	83.2	3.2	16.7	53.6	76.0
	Reallocation exercise (×100)				
	Mean	Median	10%-quantile	90%-quantile	Variance
estimate	.5	.6	2.7	-1.2	-1.1
bootstrap mean	.6	.6	2.8	-1.0	-1.1
bootstrap 2.5%-quantile	.5	.4	2.5	-1.5	-1.3
bootstrap 97.5%-quantile	.8	.8	3.2	-.4	-.8

Notes: See notes to Table 2. Mean and percentiles of the parametric bootstrap distribution, 100 replications.

Table S7: Data description, by estimated firm classes, in the sample used to estimate the dynamic model

class:	1	2	3	4	5	6	7	8	9	10	all
number of workers	13,535	39,048	45,293	50,435	63,433	45,175	81,878	34,098	49,187	20,675	442,757
number of firms	5,017	5,565	3,390	4,695	2,836	3,763	2,316	3,300	3,160	2,886	36,928
mean firm reported size	13.37	21.24	50	28.05	70.57	30.1	89.69	30.36	63.28	54.43	39.67
number of firms $\geq$ 10 (actual size)	125	814	938	871	948	693	773	470	632	369	6,633
number of firms $\geq$ 50 (actual size)	2	78	155	137	214	110	190	81	134	58	1,159
firm actual size for median worker	4	13	36	37	109	59	399	79	130	25	64
% high school drop out	30.4%	28.9%	26.8%	27.6%	23.8%	25.4%	19.1%	18.5%	9.2%	3.9%	21.5%
% high school graduates	60.6%	62.9%	61%	63.4%	59.9%	63.4%	58.6%	57.4%	40.6%	29%	57%
% some college	9%	8.2%	12.2%	9%	16.3%	11.3%	22.2%	24.1%	50.2%	67.1%	21.4%
% workers younger than 30	18.9%	15.8%	16.8%	15.4%	15.5%	13.2%	12.4%	10.5%	11.6%	11%	13.9%
% workers between 31 and 50	57.6%	56.6%	57.8%	58.3%	58.1%	59.2%	60.6%	59.8%	62%	64.5%	59.4%
% workers older than 51	23.5%	27.6%	25.4%	26.3%	26.4%	27.6%	27%	29.8%	26.4%	24.5%	26.7%
% workers in manufacturing	29.7%	44.5%	51.1%	53.4%	54.9%	50.7%	58.1%	47%	40.1%	11.4%	48.5%
% workers in services	32.1%	29%	17.7%	19.9%	15.9%	15.7%	9.8%	23.9%	37.7%	65.1%	22.4%
% workers in retail and trade	27.1%	17.2%	27.8%	9.4%	25.1%	8.8%	9.7%	10.9%	16.8%	22.2%	16.3%
% workers in construction	11.1%	9.4%	3.3%	17.2%	4.2%	24.7%	22.4%	18.1%	5.5%	1.2%	12.8%
mean log-earnings	9.75	9.94	10.04	10.06	10.15	10.15	10.24	10.3	10.45	10.73	10.19
variance of log-earnings	0.073	0.044	0.085	0.043	0.086	0.04	0.08	0.056	0.105	0.152	0.113
between-firm variance of log-earnings	0.0337	0.0033	0.0033	0.0016	0.0022	0.0013	0.0019	0.0023	0.0063	0.0436	0.0441
mean log-value-added per worker	14.57	15.05	15.65	15.22	15.95	15.26	16.06	15.33	15.85	15.89	15.37

Notes: Males, fully employed in the same firm in 2001-2002 and 2004-2005, continuously existing firms. Figures for 2002.

Table S8: Number of job movers between firm classes, in the sample used to estimate the dynamic model

firm class in period 1	firm class in period 2: 1	2	3	4	5	6	7	8	9	10
1	20	46	49	49	40	23	29	24	17	8
2	64	193	101	114	136	83	69	39	44	11
3	51	141	158	124	170	98	103	48	64	38
4	31	129	116	156	157	157	109	55	30	19
5	35	107	199	155	262	118	140	83	108	50
6	15	63	59	91	97	130	94	74	56	14
7	21	68	104	138	198	226	247	140	350	78
8	12	131	40	61	75	46	66	80	141	67
9	10	168	67	134	141	56	158	152	219	274
10	6	25	30	28	89	108	76	80	203	269

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).

Table S9: Variance decomposition (×100), dynamic model, varying the number of firm classes

	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
$K=3$	81.8	4.1	14.1	38.4	77.6
$K=4$	80.0	5.0	15.1	37.8	77.7
$K=5$	78.7	5.2	16.2	40.0	77.7
$K=6$	77.0	5.8	17.2	40.8	78.0
$K=7$	77.2	5.5	17.2	41.7	77.5
$K=8$	78.0	5.6	16.4	39.2	78.5
$K=9$	76.1	6.7	17.2	38.2	76.9
$K=10$	77.4	5.5	17.2	41.9	77.9
$K=12$	75.9	6.1	18.0	41.7	76.7
$K=15$	75.9	6.3	17.8	40.8	77.4

Notes: Dynamic model, 2001-2005. α is the worker effect, ψ is the firm effect. Variance decomposition based on a linear regression of 2002 log-earnings on α and ψ. $1,000,000$ simulations. The number of worker types is fixed to $L=6$.

Table S10: Variance decomposition (×100), dynamic model, varying the number of worker types

	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
$L=3$	40.7	36.1	23.2	30.2	46.0
$L=4$	52.5	21.9	25.6	37.7	54.8
$L=5$	75.2	7.2	17.7	38.0	76.8
$L=6$	77.4	5.5	17.2	41.9	77.9
$L=7$	75.0	6.4	18.6	42.5	77.4
$L=8$	77.0	5.7	17.4	41.6	77.4
$L=9$	76.8	6.1	17.1	39.5	78.1

Notes: See the notes to Table S9. The number of firm classes is fixed to $K=10$.

Table S11: Variance decomposition (×100), dynamic model, other specifications

$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
A. Mixture models				
Baseline model				
77.4	5.5	17.2	41.9	77.9
Random splits: first split				
76.7	5.7	17.5	41.9	77.1
Random splits: second split				
78.4	5.8	16.1	38.4	77.2
Persistence parameters estimated in first differences				
77.1	5.9	17.1	40.1	76.1
B. Regression models				
Interactive regression model				
75.4	6.1	18.5	43.1	80.6
Linear regression model				
83.8	3.3	12.9	38.9	87.0

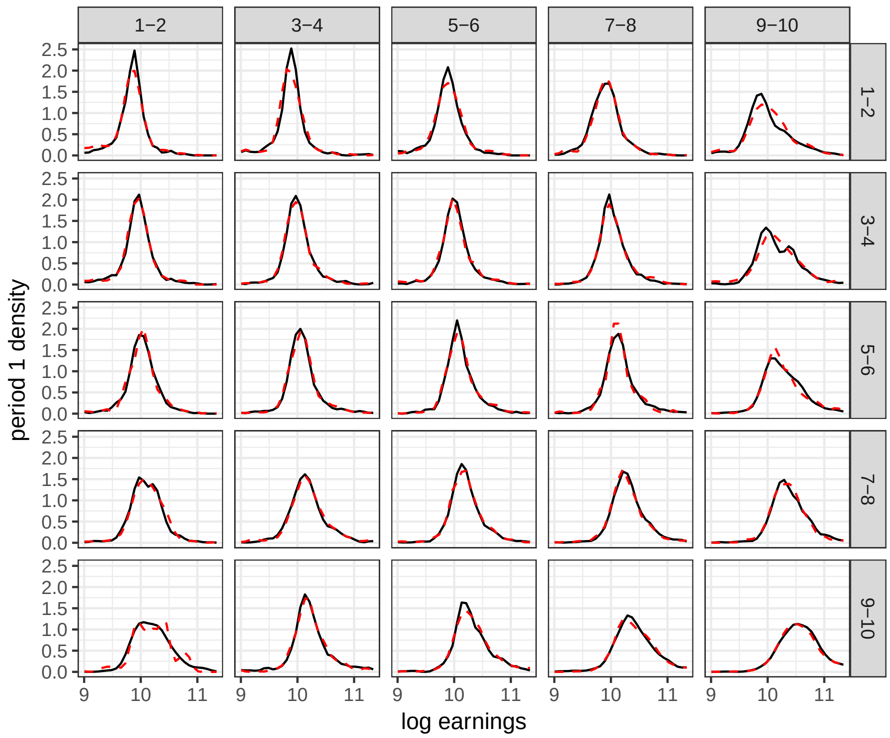
Notes: Dynamic model, 2001-2005. α is the worker effect, ψ is the firm effect. Variance decomposition based on a linear regression of 2002 log-earnings on α and ψ. $1,000,000$ simulations. The interactive regression model is given by (2), with $c_t = 0$ and $b_t(k)$ constant over time. The linear regression model has in addition $b_t(k) = 1$ for all k. Identification and estimation of the regression models is detailed in Section S3.

Table S12: Variance decomposition and reallocation exercise, dynamic model, parametric bootstrap

	Variance decomposition (×100)				
	$\frac{Var(\alpha)}{Var(\alpha+\psi)}$	$\frac{Var(\psi)}{Var(\alpha+\psi)}$	$\frac{2Cov(\alpha,\psi)}{Var(\alpha+\psi)}$	$Corr(\alpha,\psi)$	R^2
estimate	77.4	5.5	17.2	41.9	77.9
bootstrap mean	77.1	5.3	17.7	44.2	78.3
bootstrap 2.5%-quantile	73.3	4.5	16.5	37.0	76.1
bootstrap 97.5%-quantile	78.8	8.5	18.6	46.2	79.1
	Reallocation exercise (×100)				
	Mean	Median	10%-quantile	90%-quantile	Variance
estimate	.3	.8	2.5	-3.0	-1.0
bootstrap mean	.2	.8	2.5	-3.1	-.8
bootstrap 2.5%-quantile	-.1	.3	.8	-4.0	-1.6
bootstrap 97.5%-quantile	.4	1.2	3.1	-1.8	2.0

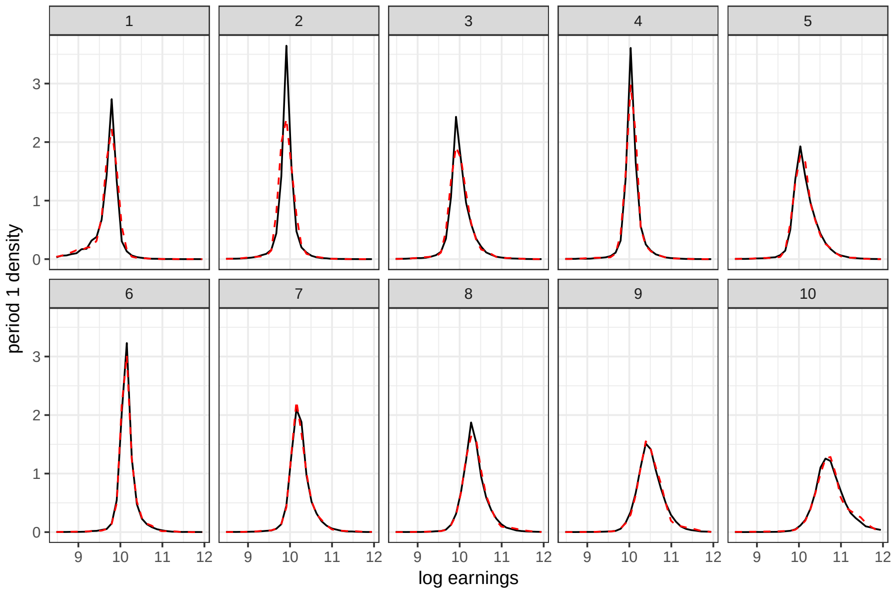
Notes: See notes to Table S6. Dynamic model, 2001-2005. Mean and percentiles of the parametric bootstrap distribution, 100 replications.

Figure S3: Fit, static model, distributions of log-earnings of job movers



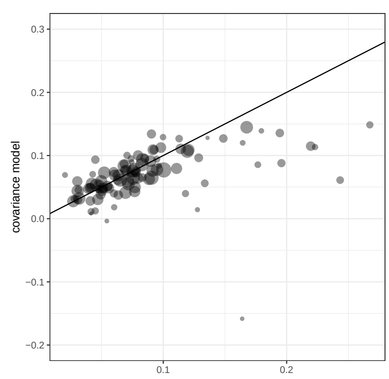
Notes: Distributions of log-earnings for workers moving from k to k' between 2002 and 2004, for all k, k'. We merge two adjacent firm classes for readability. Solid is data, dashed is model. Kernel density estimates on data simulated from the static model. $1,000,000$ *simulations.*

Figure S4: Fit, static model, distributions of log-earnings of job stayers



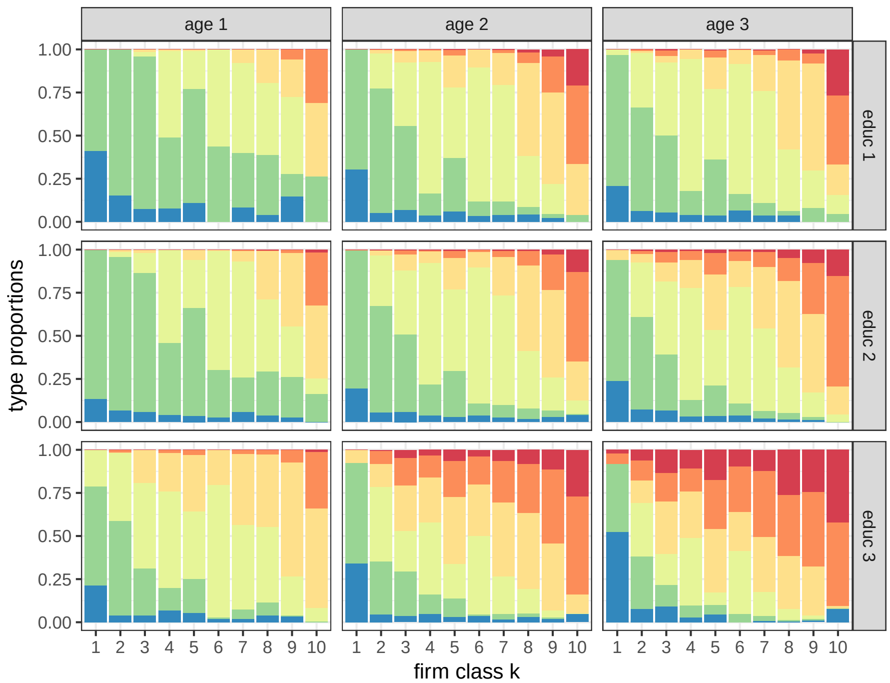
Notes: Distributions of log-earnings for workers staying in a firm in k between 2002 and 2004, for all k. Solid is data, dashed is model.

Figure S5: Fit, static model, covariance of log-earnings of job movers within each pair of firm classes



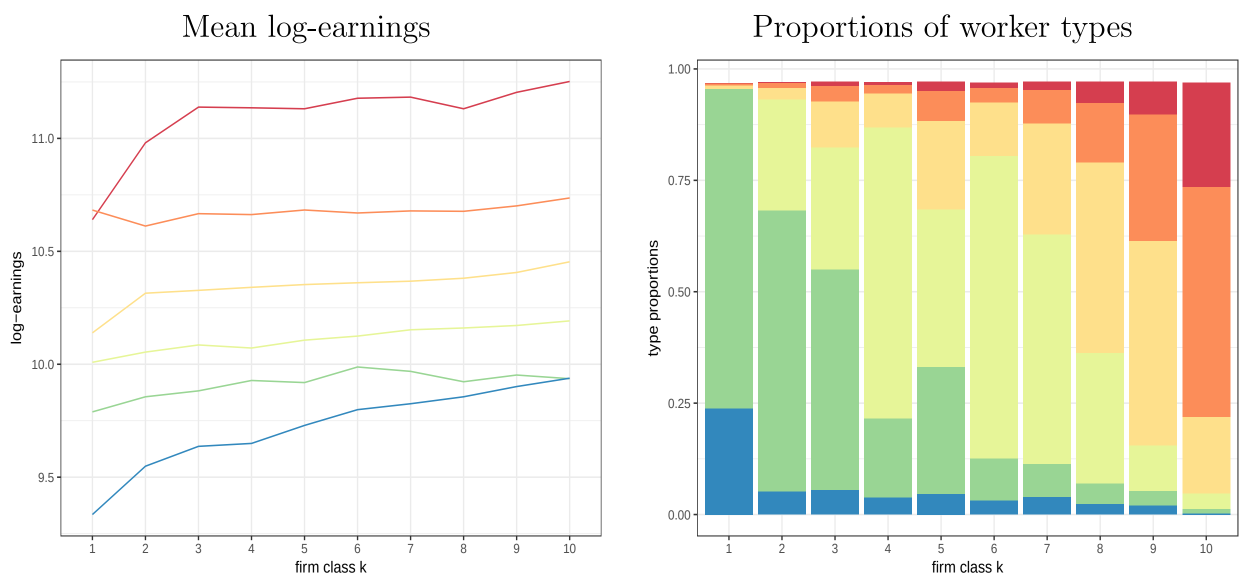
Notes: Covariance of log-earnings of job movers between k and k' of log-earnings for all k, k'. Each dot corresponds to a k, k' pair. x-axis is data, y-axis is model. The size of the dots is proportional to the number of job movers in the cells.

Figure S6: Proportions of worker types by firm classes within age×education cells



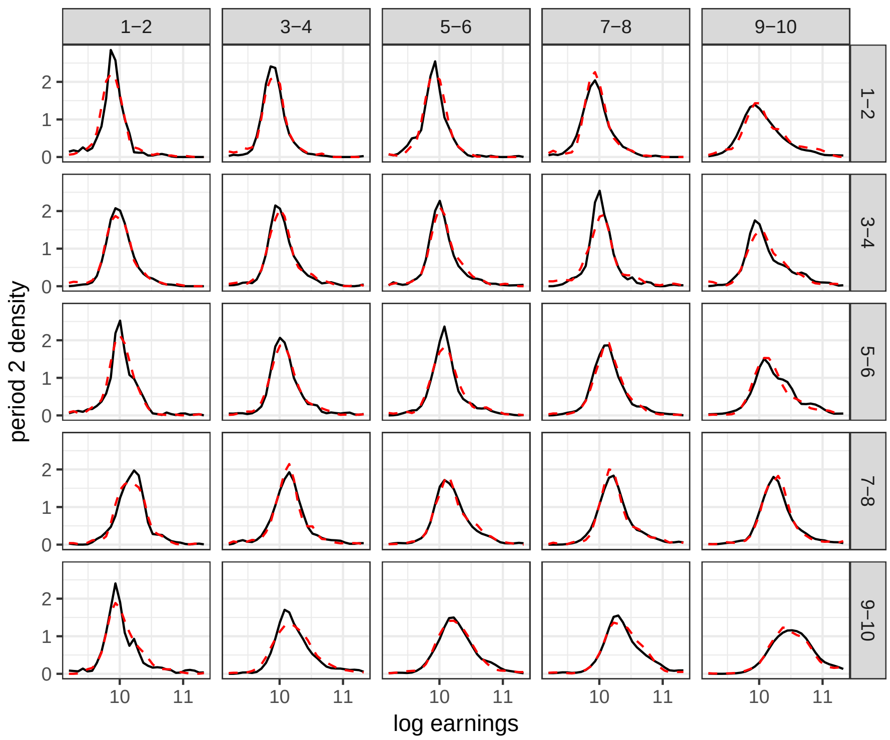
Notes: Static model, 2002-2004. Type proportions (y-axis) and firm classes (x-axis). The three age categories are: less than 30, between 30 and 50, and more than 50. The education categories are: high-school dropouts, high-school graduates, and some college.

Figure S7: Static model estimates, half-sample estimation within firm



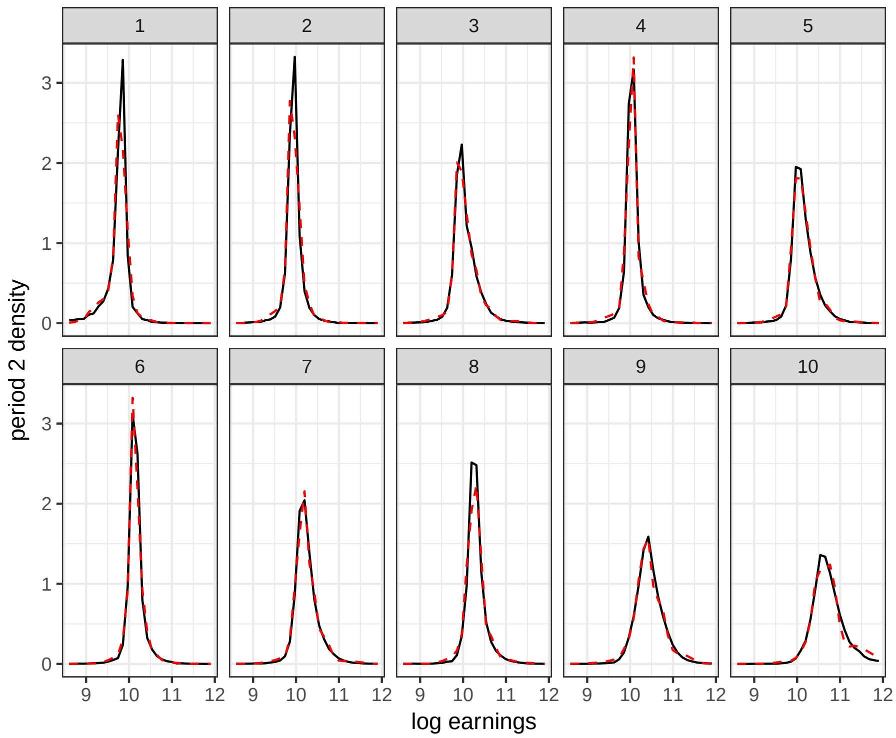
Notes: See the notes to Figure 2. The model is estimated on two random halves of the data, split within firm and mobility status. The figure shows parameter estimates in the first split.

Figure S8: Fit, dynamic model, distributions of log-earnings of job movers



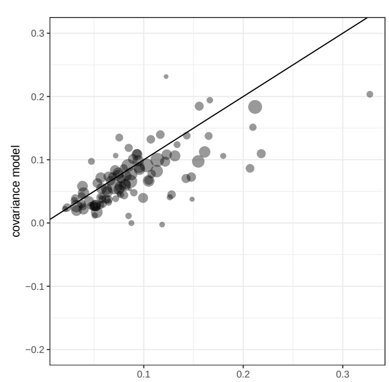
Notes: Distributions of log-earnings for workers making a move from k to k' between 2002 and 2004, for all k, k'. We merge two adjacent firm classes for readability. Solid is data, dashed is model. Kernel density estimates on data simulated from the dynamic model. $1,000,000$ *simulations.*

Figure S9: Fit, dynamic model, distributions of log-earnings of job stayers



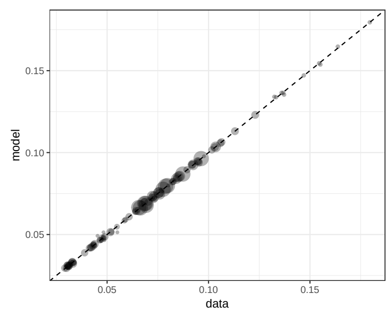
Notes: Dynamic model, 2001-2005. Distributions of log-earnings (2002) for workers staying in a firm in k between 2001 and 2005, for all k. Solid is data, dashed is model.

Figure S10: Fit, dynamic model, covariance of log-earnings of job movers within each pair of firm classes



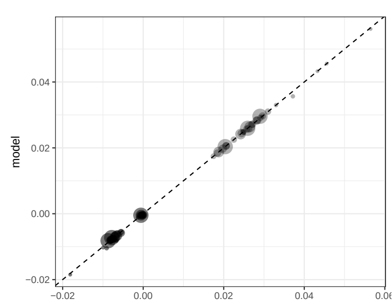
Notes: Covariance of log-earnings of job movers between k and k' of log-earnings for all k, k'. Each dot corresponds to a k, k' pair. x-axis is data, y-axis is model. The size of the dots is proportional to the number of job movers in the cells.

Figure S11: Fit of covariances, level estimation of $\rho_{4|3}$ and $\rho_{1|2}$ in the dynamic model



Notes: Covariance of log-earnings of job stayers in class k in levels, data (x-axis) and model (y-axis). Estimation in levels. The size of the dots is proportional to the number of job stayers in the cells.

Figure S12: Fit of covariances, first-differenced estimation of $\rho_{4|3}$ and $\rho_{1|2}$ in the dynamic model

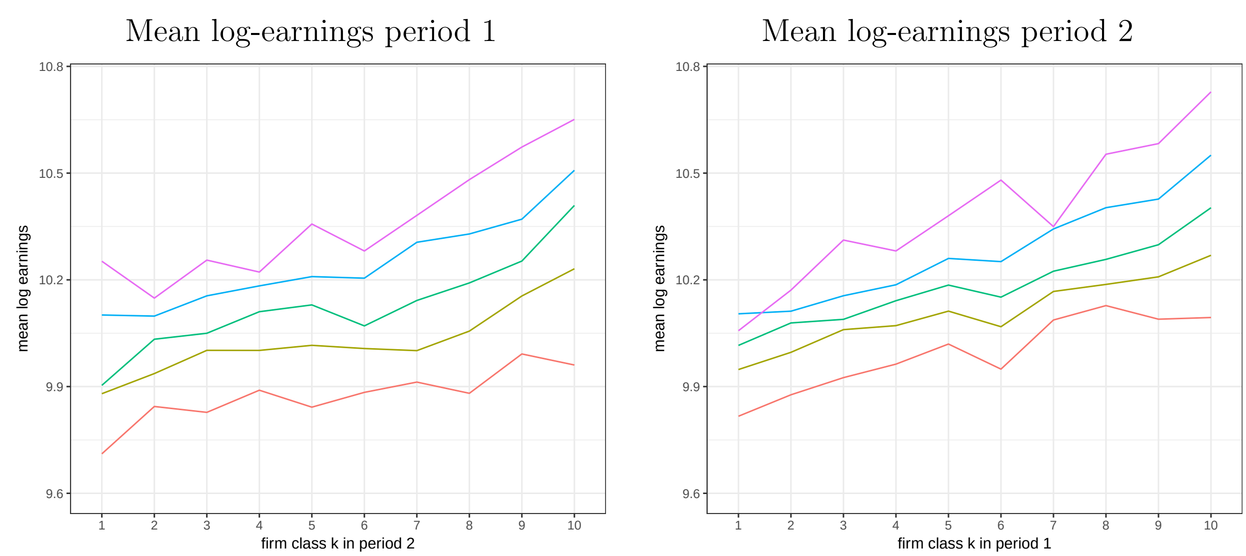


Notes: Covariance of log-earnings of job stayers in class k in first differences, data (x-axis) and model (y-axis). Estimation in first differences. The size of the dots is proportional to the number of job stayers in the cells.

Figure S13: Dynamic model estimates, half-sample estimation within firm

Notes: See the notes to Figure 3. Dynamic model, 2001-2005. The model is estimated on two random halves of the data, split within firm and mobility status. The figure shows parameter estimates in the first split.

Figure S14: Mean log-earnings of job movers by firm classes



Notes: Swedish sample, 2002-2004. Firm classes are aggregated into 5 groups on the y-axis.