A Distributional Framework for Matched Employer Employee Data*

Stéphane Bonhomme, University of Chicago
Thibaut Lamadon, University of Chicago
Elena Manresa, New York University

Revised Draft: December 2018

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 and complementarities in earnings. 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 show that this framework nests a number of structural models of wages and worker mobility. We establish identification in short panels, and develop tractable two-step estimators where firms are classified in a first step. Applying our method to Swedish administrative data, we find that log-earnings are approximately additive in worker and firm heterogeneity. Our estimates imply the presence of strong sorting patterns between workers and firms, and a small contribution of firms – net of worker composition – 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 previous employer.

JEL codes: J31, J62, C23.

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

*We thank the co-editor and four anonymous referees, our discussants Pat Kline, Chris Hansen and Mikkel Sølvsten, and Joe Altonji, Bryan Graham, Jeremy Lise, Rafael Lopes de Melo, Costas Meghir, Magne Mogstad, Derek Neal, Luigi Pistaferri, Jean-Marc Robin, Adam Rosen, Raffaele Saggio, Uta Schoenberg, Robert Shimer, Chris Taber, and audiences at various places for useful 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. 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 one to quantify their respective contributions to earnings dispersion, as well as correlations between worker and firm effects. The AKM method is widely used in labor and other fields in economics.[1] A second literature tackles similar issues from a structural perspective, by developing and estimating fully specified theoretical models of sorting in 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. This is at odds with the theoretical literature which, since Becker (1973), has 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. Such a static model is not able to account for a number of mechanisms that have been emphasized in the dynamic structural literature.

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 worker 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? And how much of the results are driven by functional form assumptions?

In this paper, we introduce an empirical framework with two-sided 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 (2018), among others. The AKM estimator has been used in a variety of other fields, for example to link banks to firms, teachers to schools or students, and 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 Lopes de Melo (2018), Lise et al. (2016), Bagger and Lentz (2014), Hagedorn et al. (2017), and Lamadon et al. (2013).

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 by specifying dynamic persistence as first-order Markov.

We provide conditions for identification in short panels 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 even when matched employer employee data sets with a long panel dimension are available, assuming time-invariant heterogeneity of either workers or firms over long periods may be unattractive. Our analysis shows that mobility and heterogeneity patterns play a key role in identifying complementarities.

We define the relevant level of firm unobserved heterogeneity as the *class* of a firm. We model worker types as draws from a discrete distribution, and allow for unrestricted interactions between worker types and firm classes. In principle, a class could be a firm itself. 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 in estimation.[3] 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 consistency of the classification under discrete firm heterogeneity.[4]

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 maximum likelihood, conditional on the estimated firm classes. Estimating firm classes in a first step is helpful for tractability. We verify in simulations that our estimator performs well in data sets

[3]See Abowd et al. (2004), Andrews et al. (2008, 2012), and recently Kline et al. (2018) for methods to address incidental parameter bias in fixed-effects regressions.

[4]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, known number of classes. In Bonhomme et al. (2017) we consider a 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.

similar to the one of our application. 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.

We take our approach to Swedish matched employer employee panel data, focusing on males for the 2002-2004 period. The estimates of our static model imply that an additive specification provides a good first-order approximation to log-earnings, although our results also highlight the presence of some complementarities between firms and lower-type workers. Between-firm differences explain 38% of the overall log-earnings variance. However our estimates imply that, net of the effect of worker composition, firm heterogeneity accounts for less than 5% of the overall variance (that is, less than 13% of the between-firm variance). The largest share of the variance is explained by worker heterogeneity. In addition, we find a strong association between worker and firm heterogeneity, with a correlation ranging between 30% and 50% depending on the specification.

These results suggest that similar workers are not paid very differently across employers, although different workers tend to work in very different firms. The presence of strong sorting, together with the absence of strong complementarities in wages, are difficult to reconcile with models where sorting is driven by complementaries in production, as in Becker (1973). Alternative explanations for sorting have been proposed, such as the presence of amenities, peer effects or more complex heterogeneity, although our findings might also be partly driven by specificities of the Swedish labor market.

The estimates of our dynamic model in the 2001-2005 period, besides being in line with the cross-sectional variance decomposition implied by our static model, 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 make workers more likely to move. This violation of exogenous mobility may indicate the presence of match heterogeneity. We also find evidence of an effect of the previous employer on current earnings, conditional on the current firm's class. This state dependence effect could be rationalized by existing theories, such as the offer and counteroffer mechanism of Postel-Vinay and Robin (2002).

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 use such an approach in models with two-sided heterogeneity. Our approach to identification and estimation of mixture models has a number of precedents in the literature, such as Hall and Zhou (2003), Hu (2008), Henry et al. (2014), Levine et al. (2011), Bonhomme et al. (2014), and Hu and Schennach (2008) and Hu and Shum (2012) for continuous mixtures. Our conditional mixture approach is also related to mixed membership models (Blei et al., 2003, Airoldi et al., 2008).

Compared to this previous work, we rely on a hybrid "one-sided correlated random-effects" approach, where we model the firm classes as discrete fixed-effects, and the worker types as (discrete or continuous) random-effects correlated with the firm classes. 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. Bonhomme (2017) reviews existing econometric methods for bipartite network data.

Also related, Abowd et al. (2018) 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. Hagedorn et al. (2017) 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. In contrast, while our framework nests a number of theoretical models of wages and mobility, it is not tied to a specific structural model.

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. The Supplemental Material contains details on computation and several extensions, and an exercise on simulated data generated using a theoretical sorting model.

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 $k_{it}=k(j_{it})$ in $\{1,...,K\}$ the class of firm j_{it}. The K classes form a partition of the set of firms. 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 firm heterogeneity has a finite, known number of points of support in the population.

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 specification. In addition to their unobserved types, workers may also differ in terms of their observable characteristics X_{it}.

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

In this framework we will be interested in recovering the distributions of log-earnings for workers of type α in firms of class k, and the proportions of type-α workers in class-k firms. Earnings distributions will be informative about complementarities, while type proportions will be informative about sorting patterns. In addition, within our framework we will be able to document transition probabilities and other 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 a 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}.

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 AKM model in the absence of interaction effects, i.e. when $b_t(k) = 1$, and firms j_{it} and classes k_{it} coincide.[5]

2.2 Dynamic model

There are two main differences between the dynamic model and the static model. 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

[5]While both parts in Assumption 1 are needed to identify the full model, restrictions on dependence are not needed to identify parameters such as $a_t(k)$, $b_t(k)$ and c_t in (1).

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}.

(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.[6] 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 $k_{i,t-1}$. 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,

[6]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 can be extended to this case, and estimation could allow for sequentially exogenous individual characteristics, such as job tenure.

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 Assumption 1 of our static model.

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 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. (2017). 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. Alternatively, ε_t may be thought of as a scalar human capital process. 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 wage Y_t and firm class k_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. Related examples are contract posting models (as in Burdett and Coles, 2003, Shi, 2008), where the optimal contract is a tenure contract.

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. However, recovering underlying primitives such as distributions of wage offers Y^*_{t+1} would require making additional 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, 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 our framework, 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 periods 1 and 2. Here we study identification in a population where there is a continuum of workers moving from k to k'.[7] 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.[8]

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 condition 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).

[7] 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.

[8] 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).

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. This analysis clarifies what can be learned from graphical illustrations of mean log-earnings before and after a job move event, which are often used to support additive specifications (e.g., Card et al., 2013). Strictly speaking, documenting symmetric wage gains and losses is not 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 (see Figure S2 in the Supplemental Material).

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 a rank condition akin to (5). Such a condition will be satisfied quite generally. For example, it is satisfied in an extension of the model of Shimer and Smith (2000) with on-the-job search. Below we use this theoretical model as a laboratory to evaluate the performance of our estimator on simulated data.

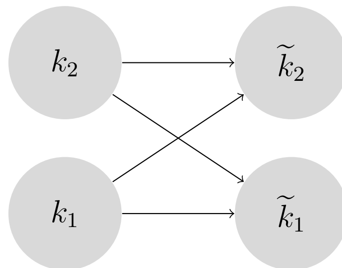
3.2 Identification with discrete worker types

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. While assuming discrete worker types is not crucial for identification, discreteness is helpful for tractability, and we will use a finite mixture specification in our empirical implementation.

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.

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. In the application we will check sensitivity by varying L. The static model imposes the following restrictions on the bivariate log-earnings

Figure 1: A connecting cycle of length $R=2$



distribution for job movers:

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

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. 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:

$$\Pr\left[Y_{i1}\leq y_1\,|\,k_{i1}=k\right] \;=\; \sum_{\alpha=1}^{L} F_{k\alpha}(y_1)q_k(\alpha). \tag{8}$$

We now provide conditions under which all parameters appearing in (7) and (8) are identified. We start with a definition.

Definition 1. *A* connecting cycle *of length R is a pair of sequences of firm classes $(k_1,...,k_R)$ in period 1, and $(\widetilde{k}_1,...,\widetilde{k}_R)$ in period 2, 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 a connecting 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 a connecting 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 $A(k_{r+1}, \widetilde{k}_r)$ have rank L, where $A(k, k')$ has (y_1, y_2) element:

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

Assumption 3 requires that any two firm classes k and k' belong to a connecting cycle. An example is given in Figure 1, where the presence of a connecting cycle requires that there is a positive proportion of every worker type among job 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 implies graph connectedness, in the sense of AKM (Abowd et al., 2002). However, connectedness here is at the firm class level. A specific feature of our nonlinear setting is the need for every firm class to contain job movers of all types of workers. This may be demanding empirically, and the condition may fail in some models of sorting. The requirement on cycles can be relaxed, at the cost of loosing point-identification of some of the quantities of interest. Alternatively, one may impose more structure, as in the interactive regression model (1), for example.

Assumption 3 (i) requires some asymmetry in worker type composition between different firm classes. This condition requires non-random mobility, since 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, for example. In the mixture model we focus on here, the presence of asymmetric job movements between firm classes is crucial for identification. This is similar to the case of the interactive regression model we analyzed above, see equation (6).

Finally, 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 can be uniquely recovered. The intuition for the result is similar to that in the interactive regression model. Due to the discrete heterogeneity setting, identification is up to an arbitrary choice of labeling of the latent worker types. Proofs are given in Appendix A.

Theorem 1. *Let $T = 2$, and 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. Lastly, the type proportions $q_k(\alpha)$ in the first period are all identified, again for the same labeling.*

We now turn to the dynamic model, where we focus 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. We again abstract from covariates.

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

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

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 $(\alpha_i, Y_{i2}, k_{i1}, k_{i2}, m_{i1})$.

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

$$\Pr\left[Y_{i1} \leq y_1, Y_{i2} \leq y_2 \,|\, k_{i1} = k_{i2} = k, m_{i1} = 0\right] = \sum_{\alpha=1}^{L} G^f_{y_2,k\alpha}(y_1) F_{k\alpha}(y_2) q_k(\alpha). \tag{10}$$

The mathematical structure of (9)-(10) is analogous to that of (7)-(8). 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}, even though in this model earnings have a direct effect on job mobility and respond dynamically to lagged earnings and previous firm classes.

We start by listing the assumptions, which are stronger than in the static case.

Definition 2. *An* augmented connecting 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 4. *(mixture model, dynamic)*

(i) For any two firm classes k, k' *in* $\{1, ..., K\}$ *and any two log-earnings values* y, y', *with* $(k, y) \neq (k', y')$, *there exists an augmented connecting 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 connecting 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* $A(y_{r+1}, \widetilde{y}_r, k_{r+1}, \widetilde{k}_r)$ *have rank* L, *where* $A(y, y', k, k')$ *has* (y_1, y_4) *element:*

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

We then have the following nonparametric identification result for the dynamic model under discrete worker heterogeneity.

Theorem 2. *Let* $T = 4$, *and let Assumptions 2 and 4 hold. 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), $p_{y_2 y_3, kk'}(\alpha)$ *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.*

4 Two-step grouped fixed-effects 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 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 ill-behaved 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 computationally tractable two-step grouped fixed-effects estimator, 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

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 $k = k(j)$:

$$\Pr\left[Y_{i1} \leq y_1 \,|\, j_{i1} = j\right] = \sum_{\alpha=1}^{L} F_{k\alpha}(y_1) q_k(\alpha), \tag{11}$$

where the left-hand side thus only depends on $k = k(j)$. This observation motivates classifying firms into classes using 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 (μ is supported on a finite grid in our application), $k(1), ..., k(J)$ denotes a partition of firms into K classes, and $H_1, ..., H_K$ are cdfs. Here we take the number K of firm classes as known. In the empirical analysis we will check robustness around the baseline value $K = 10$. 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.

To provide a formal justification for the classification, we consider a setting where the model (either static or dynamic) is well-specified, and there exists a partition of 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 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 is not affected by the estimation of firm classes. In Appendix B we provide details on this analysis.

Two remarks are in order. First, the labeling of the classes is arbitrary. This labeling does not affect variance decomposition exercises. However, a structural interpretation of the firm classes in terms of productivity would require additional assumptions. Eeckhout and Kircher (2011) show that in some cases it can be impossible to recover such a productivity ordering from wage data only.

Second, the classification fails to be identified when two firm classes have identical earnings distributions in the cross-section. This can happen if one firm offers a higher earnings schedule but has lower-type workers, and another one offers a lower earnings schedule but has higher-type workers, in such as way that the two firms have exactly the same earnings distributions. In some environments without firm capacity constraints, such as Postel-Vinay and Robin (2002), the upper bound of earnings in the firm is increasing in firm productivity, so firm-specific distributions are all different and firms may be consistently classified based on their earnings distributions. It is difficult to obtain similar guarantees in models with capacity constraints. Nevertheless, note that identification can survive even when some firms have identical mean earnings, provided the *distributions* of earnings differ.

In the empirical analysis we attempt to bring additional information beyond cross-sectional earnings to learn about firm heterogeneity, in several ways. Going beyond wage information has been advocated in the structural literature (e.g., Eeckhout and Kircher, 2011, Bagger and Lentz, 2014, Hagedorn et al., 2017, Bartolucci et al., 2015). As robustness checks, we use re-classification and random-effects methods, which rely on longitudinal information on both earnings and mobility. We also experiment, albeit without a structural model, with bringing other information on the firm in order to inform the classification, such as firm value added and measures of worker flows.

4.2 Recovering the model's parameters in a second step

From the first step, we obtain estimates of firm classes $\widehat{k}(j)$ for all firms j in the sample. Given those, we then impute a class $\widehat{k}_{it} = \widehat{k}(j_{it})$ to each worker-period observation, and we estimate the model conditional on the $\widehat{k}_{it}$'s. To describe this second 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 estimate on Swedish data.

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 log-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.

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.[9] 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 standard 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 for computation (Dempster et al., 1977).

Dynamic model. We use a similar approach for the dynamic finite mixture model using 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 be relaxed with a larger sample.

[9]Proceeding in this way has the advantage of recovering earnings parameters from job movements directly, albeit with some efficiency loss. 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 $q_k(\alpha)$.

In addition, we specify the mean of (Y_{i2}, Y_{i3}) for job movers between classes k and k' as $(\mu_{2k\alpha} + \xi_2(k'), \mu_{3k'\alpha} + \xi_3(k))$. The term $\xi_2(k')$ 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. We specify the mean of (Y_{i2}, Y_{i3}) for job stayers in a firm of class k as $(\mu^s_{2k\alpha}, \mu^s_{3k\alpha})$. Lastly, for both stayers and movers we let the covariance matrices vary with firm classes.[10]

Given estimates $\widehat{\rho}_{4|3}$ and $\widehat{\rho}_{1|2}$ of the persistence parameters, the other parameters can be estimated using a 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 the Supplemental Material.

The estimation approach outlined in this section can be modified in several ways that we implement empirically. A first extension is a model-based re-classification. Given estimates

[10] We impose that the best linear predictors in the regressions of Y_{i3} on Y_{i2}, for both stayers (denoted as $\rho^s_{3|2}$) and movers ($\rho^m_{3|2}$), 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.

of the θ parameters one may re-classify every firm j to the class $k = \widetilde{k}(j)$ that corresponds to the maximal value of the k-specific likelihoods of firm j's observations. This approach can be iterated further. In addition, while we have described estimation in the context of finite mixture models, the two-step approach can also be used in regression models, such as the AKM model and its interactive counterparts (1) and (2) that allow for complementarities or dynamics. In such models, the two-step grouped fixed-effects method deliver computationally convenient estimation algorithms based on mean and covariance restrictions. We will use estimates of interactive regression models to show the robustness of our results.

Experiments on simulated data. To evaluate the ability of our estimator to recover the contributions of worker and firm heterogeneity to earnings dispersion, we consider two simulation experiments. We first analyze an extension of the model of Shimer and Smith (2000) with on-the-job search, using parameter values that imply either positive or negative assortative matching. We find that our finite mixture estimator of the static model recovers wage functions and variance contributions well (see Figure S3 and Table S1 in the Supplemental Material). We also set up a second simulation experiment, where we use our empirical estimates on Swedish data as "true" parameter values. We find that the Monte Carlo distributions are approximately centered around true values and have low dispersion, for both the static model and the dynamic model (see Tables S5 and S6). In the simulation based on the static model, we also compute means of AKM fixed-effects estimates across Monte Carlo distributions (fifth row in Table S5). We find that AKM estimates of variance components are severely biased. Although such biases may not be surprising, given the short length of the panel and the low mobility rate, note that our estimator does recover the magnitudes of variance components in this setting.

5 Empirical results I: Static model

We now present results for the static model in the Swedish 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 working in the private sector 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. In Appendix C we provide details on the Swedish labor market, and on sample construction. We define job movements in a conservative way, which we describe in detail in Appendix C. This construction

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	32.30	60.08	51.24	54.16	50.86	37.59
number of firms $\geq$ 10 (actual size)	160	1,034	1,519	1,357	1,192	930	999	855	632	415	9,093
number of firms $\geq$ 50 (actual size)	7	87	260	225	270	162	245	183	147	52	1,638
% 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%	59.0%	71.2%	22.7%
% workers younger than 30	24.3%	19.5%	19.8%	17.5%	18.6%	15.4%	13.8%	14.3%	15.0%	14.3%	16.8%
% workers between 31 and 50	54.1%	54.6%	55.0%	56.2%	56.0%	57.6%	58.5%	58.9%	60.0%	64.2%	57.2%
% workers older than 51	21.7%	25.9%	25.1%	26.3%	25.5%	27.0%	27.6%	26.8%	25.0%	21.5%	26.0%
% workers in manufacturing	24.3%	39.3%	46.8%	53.0%	51.5%	52.0%	53.0%	40.3%	31.5%	7.6%	45.4%
% workers in services	39.3%	32.1%	23.3%	19.7%	14.4%	15.0%	16.0%	29.7%	52.1%	72.6%	25.3%
% workers in retail and trade	26.4%	19.0%	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	10.50	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
skewness of log-earnings	-1.392	-0.709	0.345	0.019	0.576	0.433	0.474	0.703	0.385	1.001	0.582
kurtosis of log-earnings	7.780	14.093	9.017	15.565	7.788	14.763	10.033	8.141	6.651	6.984	7.400
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	12.40	12.58	12.69	12.69	12.84	12.75	12.87	12.94	13.03	13.18	12.74

Notes: The table corresponds to males fully employed in the same firm in 2002 and 2004, for firms that are continuously present in the sample. The "actual size" is the number of workers per firm in our sample. All numbers in the table correspond to 2002.

results in low mobility rates, with a ratio 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 with $10,000$ randomly generated starting values. 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.

In Table 1 we provide summary statistics about the estimated firm classes for our baseline choice of number of classes $K = 10$. We order firm classes 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 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. Classes differ also in terms of second, third, and fourth-order log-earnings moments. In addition, there are 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.

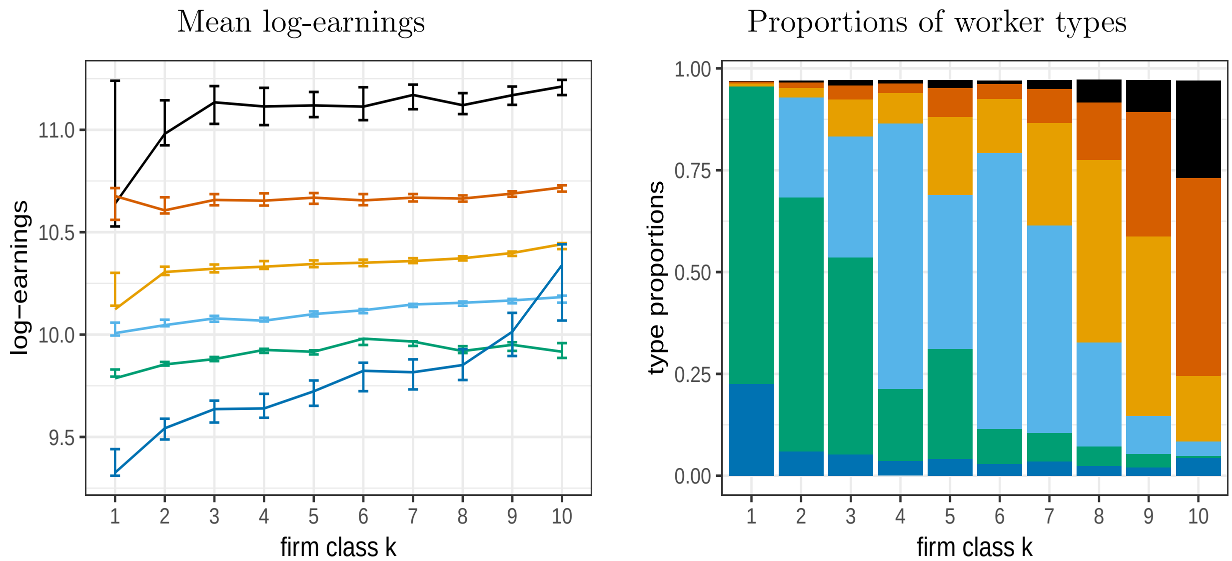
Firm size reported by the firm tends to increase with firm class, although the relationship is not monotonic. Classes 1 and 2 contain smaller firms than the other classes. There is also evidence of both between- and within-sector variation between classes, which is not monotonic in mean earnings. For example, the proportion of workers in services is U-shaped in firm class, while it is inverse U-shaped in manufacturing. Lastly, log value added per worker tends to increase with firm class, although classes explain only 18% of the between-firm variance in log value added per worker.

There is substantial worker mobility between firm classes in our sample, especially between adjacent classes (see Table S4 in the Supplemental Material). This is important since our identification strategy is based on exploiting mobility. In addition, means of log-earnings for workers moving between different firm classes – e.g., moving "up" or "down" – tend to be associated with different earnings levels (see Figure S6). We emphasized the importance of such asymmetric patterns 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. We check below how sensitive our main results are to varying these two numbers. As we explained in Section 4, we estimate earnings distributions and type proportions of job movers using the sample of job movers between 2002-2004. We then estimate type proportions 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, some local maxima have poor connectedness, in the sense that some worker types only move within a subset of firm classes, which results in unstable parameter estimates. We use the EM algorithm to explore the likelihood function, and use a measure of network connectedness recently studied in Jochmans and Weidner (2017) to select our preferred estimates; see the Supplemental Material for details.

In the left panel of Figure 2 we plot estimates of the means of log-earnings, for each firm

Figure 2: Main parameter estimates of the static model



Notes: Estimates of the static model, on 2002-2004. In the left graph we plot estimates of means of log-earnings, by worker type and firm class. We order the $K = 10$ firm classes (on the x-axis) by mean log-earnings. On the y-axis we report estimates of mean log-earnings for the $L = 6$ worker types. In the right graph we show estimates of the proportions of worker types in each firm class. In the left graph, the brackets indicate pointwise parametric bootstrap 2.5%–97.5% quantile bands (computed using 200 replications).

class and each worker type. On the x-axis, firm classes are ordered by mean log-earnings. The brackets show 95% confidence intervals based on the parametric bootstrap.[11] The results show clear evidence of worker heterogeneity. They also show some variation in log-earnings between firm classes, although to a lesser extent. Moreover, lower-type workers (where "lower" types refer to low 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.

In 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

[11]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.

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

Variance decomposition (×100)				
$\frac{Var(\alpha)}{Var(y)}$	$\frac{Var(\psi)}{Var(y)}$	$\frac{2Cov(\alpha,\psi)}{Var(y)}$	$\frac{Var(\varepsilon)}{Var(y)}$	$Corr(\alpha,\psi)$
60.03	2.56	12.17	25.24	49.13
(0.85)	(0.16)	(0.39)	(0.59)	(0.86)
Reallocation exercise (×100)				
Mean	Median	10%-quantile	90%-quantile	Variance
0.50	0.58	2.60	-1.24	-1.12
(0.10)	(0.11)	(0.19)	(0.31)	(0.11)

Notes: Estimates of the static model, on 2002-2004. In the top panel, α denotes the worker effect, and ψ denotes the firm effect, in the linear regression $Y = \alpha + \psi + \varepsilon$. In the bottom panel we report differences in means, quantiles, and variances of log-earnings between two samples: a counterfactual sample where workers are randomly reallocated to firms, and the original sample. The results are obtained using $1,000,000$ simulations, and we report parametric bootstrap standard errors in parentheses (computed using 200 replications).

classes is mainly due to firms employing different workers, rather than differences in earnings for a given worker type.

Variance decomposition and reallocation. We next report the results of several exercises that illustrate how earnings and heterogeneity relate to each other. We start with a decomposition of the variance of 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 α (that is, coefficients of worker type indicators), the variance of firm effects ψ (i.e, coefficients of firm class indicators), twice the covariance between the two, and the variance of residuals ε. In our nonlinear model we perform a similar decomposition by working with a linear projection of log-earnings on worker type indicators and firm class indicators, in a regression without interactions. The results of the decomposition reported in the top panel of Table 2 show two main features. First, worker heterogeneity explains substantially more variation in earnings than firm heterogeneity. Differences in firm classes only account for 2.6% of the variance, compared to 60% 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%, which suggests the presence of strong sorting between workers and firms. This is in line with the evidence documented in 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 those enter the regression interactively as opposed to additively. 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%. Hence, while the left panel of Figure 2 suggests the presence of some complementarity between firms and lower-type workers, those complementarities explain only a small part of the variance of log-earnings.

We next consider the impact on log-earnings of a reallocation exercise where workers are randomly allocated to firms. With this exercise, our aim is to assess the contribution of sorting to the distribution of earnings. We assume that earnings functions, for all worker types and firm classes, are not affected by the reallocation, thus abstracting from equilibrium effects. We show the results of the reallocation in the bottom panel of Table 2. In the first column we report the estimate of the difference in mean log-earnings between two samples: a counterfactual sample where workers are randomly allocated among firms, and our original sample. 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, using our estimates that account for the presence of complementarities we find a positive mean impact (.5%), which suggests that the effect of complementarities on average log-earnings is statistically significant but quantitatively small. The positive sign is in line with the fact that the mean worker type tends to increase with firm class k, while complementarities are somewhat stronger in low-wage firm classes.[12]

Moreover, in our distributional framework we are able to estimate the entire counterfactual earnings 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 counterfactual random allocation and our sample. 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, whereas 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, as well as the presence of complementarities at the bottom of the distribution.

Interpretation. The coexistence of strong sorting and weak complementarities may be surprising in the perspective of the search matching models inspired by Becker (1973). In such

[12] To provide an intuition, consider the regression model (1). In this specification the difference between mean outcomes in a population where workers are randomly allocated to firms and in our data is, abstracting from time indices for clarity: $\mathbb{E}^{random}(Y_i) - \mathbb{E}(Y_i) = -\operatorname{Cov}(b(k_i), \mathbb{E}(\alpha_i \mid k_i))$.

models, assortative matching patterns arise from complementarities in production, which are often reflected in non-monotonic wages. For example, this happens in the model of Shimer and Smith (2000) with on-the-job search that we have simulated. More generally, the results in this section are difficult to reconcile with models based on revealed preferences for wages only. This may indicate that workers or firms care about other job attributes (Hwang et al., 1998), or that workers of a similar type enjoy working together as in the peer effects literature. Workers and firms might also be more complex than assumed in most models of sorting (Lindenlaub, 2017, Lise and Postel-Vinay, 2015). More specifically to the Swedish context, as we discuss in Appendix C, institutions such as unions may contribute to explain why different firms do not pay similar workers very differently.

Robustness analysis. In order to explore the robustness of our results, we perform several exercises that we summarize in Table 3. In panels B and C of the table we show the results of the variance decomposition for different values of the numbers of firm classes and worker types, respectively. The results are quite stable across K values. In addition, while taking $L = 3$ seems to understate the contribution of worker heterogeneity and overstate that of firm heterogeneity, the results are stable between $L = 5$ and $L = 9$.

In panel D we report results corresponding to several additional specifications. In the first row 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. The results are very similar to the baseline. In the second row we report the results of a variance decomposition applied to residuals of log-earnings on a third-order polynomial in age, three education categories, three sectors (manufacturing, retail and services), and time indicators, with a set of interactions. The R^2 in the regression is 21%. We estimate the parameters on the regression residuals using our two-step approach, including the k-means step. We find that the contribution of worker unobserved heterogeneity is 50% of the variance, compared to 60% in the baseline. Moreover, the relative contribution of firm heterogeneity is somewhat larger than in the baseline. In rows three and four we report decompositions estimated on smaller firms (less than 50 workers per firm in 2002) and large firms (more than 50). In smaller firms the contribution of worker heterogeneity is reduced compared to that of firm heterogeneity and the covariance term, while in larger firms worker heterogeneity accounts for a greater share of the log-earnings variance. In the fifth row we show that the variance decomposition implied by a fully nonstationary model gives very close results to our baseline estimates.

Table 3: Variance decomposition ($\times 100$), static model, robustness checks

	$\frac{Var(\alpha)}{Var(y)}$	$\frac{Var(\psi)}{Var(y)}$	$\frac{2Cov(\alpha,\psi)}{Var(y)}$	$\frac{Var(\varepsilon)}{Var(y)}$	$Corr(\alpha,\psi)$
	A. Baseline model				
	60.0	2.6	12.2	25.2	49.1
	B. Varying the number of firm classes				
$K = 3$	65.9	1.4	8.7	24.1	45.3
$K = 20$	57.6	2.9	12.2	27.3	47.4
	C. Varying the number of worker types				
$L = 3$	13.5	13.4	12.2	60.8	45.4
$L = 5$	55.5	3.2	12.8	28.5	47.9
$L = 9$	53.3	3.5	13.1	30.1	48.2
	D. Other mixture specifications				
mixture-of-mixtures	62.6	2.4	11.3	23.8	46.4
log-earnings residuals	49.6	3.5	10.6	36.3	40.0
firms with ≤ 50 workers	52.6	4.0	17.0	26.4	59.1
firms with > 50 workers	60.1	2.6	7.9	29.4	31.6
fully nonstationary	60.5	2.8	12.5	24.3	48.3
	E. Regression models				
interactive	56.5	2.1	10.8	30.6	50.2
linear	60.6	1.7	10.0	27.6	48.5
	F. Other classifications				
classify using means	57.1	4.2	13.7	25.0	43.8
split by percent of movers	67.2	1.5	8.5	22.9	42.5
split by value added	55.4	3.4	12.5	28.7	46.0
	G. Re-classification (starting at baseline)				
1 iteration	55.9	4.1	13.3	26.7	43.9
10 iterations	56.3	4.1	12.5	27.2	41.2

Notes: Estimates of the static model, on 2002-2004. α denotes the worker effect, and ψ denotes the firm effect, in the linear regression $Y = \alpha + \psi + \varepsilon$. The results are obtained using $1,000,000$ simulations. The various specifications are described in the text.

In panel E of Table 3 we show results based on the interactive regression model (1).[13] This model is a valuable complement to our baseline model for several reasons. First, in the regression

[13]We do not account for covariates (i.e., $c_t = 0$), and we keep $b_t(k)$ constant over time.

model we do *not* assume that worker types are discrete. Second, estimation only relies on mean and covariance restrictions, not on other features of the 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. In the second row we report the results for a regression model where we additionally impose that $b(k) = 1$ for all k. The results in panel E show similar variance decompositions as in our baseline model.

In panel F we consider several specifications in which we vary the way we classify firms. In the first row we use deciles of mean log-earnings for the classification. In the next two rows, we split each class in two equally-sized subclasses according to the percentage of job movers in the firm (second row), and value added (third row). These specifications capture other information, beyond differences in log-earnings distributions, contained in mobility patterns and value added. The results show some differences with our baseline estimates. For example, when splitting the classes according to value added, the firm effects variance becomes 3.4%, compared to 2.6% in the baseline. The three specifications in panel F also give slightly smaller correlations between worker and firm effects. However, the differences compared to our baseline estimates are small.

Lastly, in panel G of Table 3 we show the results of a model-based re-classification. Unlike our baseline classification of firms, this method incorporates information from both periods, including earnings associated with job mobility. We report the result of one and ten iterations, starting from the baseline parameter values. The iterated estimates tend to imply a somewhat larger firm effect and smaller correlation coefficient than our baseline two-step results, although the differences are quantitatively small. We have also performed a similar re-classification starting from different initial classifications: deciles of the firm effects estimates of Abowd et al. (1999), deciles of log value added per worker, deciles of the poaching rank measure of Bagger and Lentz (2014), and deciles of the firm-specific shares of job movers. We have found that, while the initial values of the variance components differ substantially, the values after ten iterations are quite close to each other (see Table S7 in the Supplemental Material).

The magnitudes we find differ from the fixed-effects estimates of Abowd et al. (1999). In particular, the AKM estimate of the variance of firm effects is 32% on our sample, and the estimated correlation coefficient is negative. In short panels with low mobility rates, it is well-documented that the AKM estimator may be biased (Andrews et al., 2008). Our experiments on simulated data, which we described in Section 4, suggest that biases may be substantial in our case. In additional robustness checks, we have found that the large AKM estimate of the firm effects variance is mostly driven by firms with very small number of job movers, and that the variance drops substantially when using existing bias-correction methods (see Table S2 in the Supplemental Material). These various exercises support our finding that, once netted out of incidental parameter bias, the variance of firm effects is small in the Swedish data.

We have performed two additional robustness checks. To further address the robustness of our classification of firms, we have estimated a two-sided random-effects specification where both worker types and firm classes are treated as stochastic. Compared to our two-step approach, this specification is more challenging to estimate. To check sensitivity to departures from discrete firm heterogeneity, we have computed estimates allowing for within-class firm heterogeneity. In both cases we have found magnitudes in line with our baseline estimates. We provide details on these robustness checks in the Supplemental Material.

In all our specifications, we consistently find magnitudes that are in line with our baseline estimates. In particular, firm effects net of worker composition explain less than 5% of the log-earnings variance, and the correlation between worker and firm effects ranges between 30% and 50%. Compared to estimates based on AKM – not corrected for bias – in other countries, our findings on Swedish data suggest a smaller contribution of firms to wage dispersion and a stronger association between worker and firm heterogeneity. We are aware of only a handful of structural estimates in the literature. Notably, Bagger and Lentz (2014) find a variance of firm effects of 11% and a correlation coefficient of 32% on Danish data, and Hagedorn et al. (2017) estimate a correlation between worker and firm types of 75% on a German sample.

6 Empirical results II: Dynamic model

In this section we present empirical results for our dynamic model. We start by describing parameter estimates, and we then assess their implications in terms of dynamic patterns.

6.1 Parameter estimates

We estimate the dynamic model on 2001-2005, focusing on males fully-employed in the same firm in 2001-2002 and 2004-2005. In order to estimate firm classes, we use the same weighted k-means algorithm as in the static model. 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 those of job stayers, in turn.

In Table 4 we show estimates of several parameters of the dynamic model. The parameter $\xi_2(k')$ is the effect of firm class k' on the mean log-earnings in period 2 of a worker moving from

Table 4: Parameter estimates of the dynamic model

Earnings effects $\xi_2(k')$ of future firm classes								
$k'=2$	$k'=3$	$k'=4$	$k'=5$	$k'=6$	$k'=7$	$k'=8$	$k'=9$	$k'=10$
-0.005	0.004	0.005	0.022	0.002	0.015	0.009	0.016	0.023
(0.008)	(0.009)	(0.011)	(0.012)	(0.011)	(0.010)	(0.011)	(0.012)	(0.012)
Earnings effects $\xi_3(k)$ of past firm classes								
$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
0.051	0.038	0.045	0.061	0.040	0.072	0.058	0.087	0.090
(0.016)	(0.015)	(0.015)	(0.016)	(0.017)	(0.015)	(0.018)	(0.016)	(0.017)
Persistence parameters ρ								
	$\rho_{1\|2}$		$\rho^m_{3\|2}$		$\rho^s_{3\|2}$		$\rho_{4\|3}$	
	0.227		0.246		0.681		0.651	
	(0.009)		(0.044)		(0.022)		(0.004)	

Notes: Estimates of the dynamic model, on 2001-2005. $\rho^m_{3|2}$ is the autoregressive coefficient of log-earnings for job movers between 2002 and 2004, and $\rho^s_{3|2}$ is the coefficient for job stayers. $\xi_2(k')$ and $\xi_3(k)$ denote the mean effects on log-earnings before and after a job move between firm classes k and k', respectively. $k'=1$ and $k=1$ are the omitted categories. We report parametric bootstrap standard errors in parentheses (computed using 200 replications).

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. This effect is quantitatively small, with at most a 2% effect relative to the omitted class $k'=1$. Note that strict exogeneity of mobility also imposes that earnings realizations are independent of the subsequent decision to move. We find stronger support against this implication of exogenous mobility, see Table 7 below.

The parameter $\xi_3(k)$, in turn, captures the effect of firm class k on the mean log-earnings in period 3 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 previous employer 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% wage premium in the highest classes. This suggests that past firms have an impact on future earnings.

In the bottom panel of Table 4 we report the estimates of earnings persistence parameters. Persistence estimates are higher for job stayers than for job movers. Note that the autoregressive

Table 5: Variance decomposition and reallocation exercise in the dynamic model

Variance decomposition $(\times 100)$				
$\frac{Var(\alpha)}{Var(y)}$	$\frac{Var(\psi)}{Var(y)}$	$\frac{2Cov(\alpha,\psi)}{Var(y)}$	$\frac{Var(\varepsilon)}{Var(y)}$	$Corr(\alpha,\psi)$
60.27	4.24	13.40	22.09	41.90
(1.30)	(0.65)	(0.37)	(0.69)	(2.35)
Reallocation exercise $(\times 100)$				
Mean	Median	10%-quantile	90%-quantile	Variance
0.26	0.80	2.57	-3.24	-1.05
(0.28)	(0.19)	(0.70)	(0.57)	(1.03)

Notes: Estimates of the dynamic model, on 2001-2005. See the notes to Table 2.

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 an important robustness check, we have estimated the persistence parameters based on covariance restrictions in first differences, in addition to the specification in levels shown in Table 4. The literature has documented differences between level estimates and first difference estimates of the dynamics of earnings in several data sets (Daly et al., 2016). We find $\rho_{1|2} = .506$, $\rho_{4|3} = .451$, $\rho^m_{3|2} = .194$, and $\rho^s_{3|2} = .605$. Despite the differences in persistence estimates, variance decomposition results are similar in this case.

The estimates of the dynamic model deliver similar cross-sectional patterns for log-earnings and sorting as in the static case. In particular, the model implies 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 (see Figure S7 in the Supplemental Material). In Table 5 we report the cross-sectional variance decomposition and reallocation exercise based on the dynamic model. The variance decomposition is quite similar to the one we obtained using the static model, with some differences: the contribution of firm effects increases from 2.6% to 4.2%, and the correlation between worker and firm effects decreases from 49% to 42%. As in the static model, adding interactions between worker types and firm classes has small effects on the R^2 of the regression (i.e., 78.5% versus 77.9%). In the bottom panel of Table 5, we show 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.

In Table 6 we show that our estimates are stable across a range of specifications, when

Table 6: Variance decomposition (×100), dynamic model, robustness checks

	$\frac{Var(\alpha)}{Var(y)}$	$\frac{Var(\psi)}{Var(y)}$	$\frac{2Cov(\alpha,\psi)}{Var(y)}$	$\frac{Var(\varepsilon)}{Var(y)}$	$Corr(\alpha,\psi)$
	A. Baseline model				
	60.3	4.2	13.4	22.1	41.9
	B. Persistence parameters estimated in first differences				
	58.7	4.5	13.0	23.9	40.1
	C. Varying the number of firm classes				
$K = 3$	63.5	3.2	10.9	22.4	38.4
$K = 15$	58.7	4.9	13.8	22.6	40.8
	D. Varying the number of worker types				
$L = 3$	18.7	16.6	10.7	54.0	30.2
$L = 5$	57.8	5.5	13.6	23.2	38.0
$L = 9$	60.0	4.8	13.4	21.9	39.5
	E. Regression models				
interactive	60.8	4.9	14.9	19.4	43.1
linear	72.9	2.9	11.2	13.0	38.9
	F. Re-classification (starting at baseline)				
1 iteration	58.0	5.8	12.9	23.3	35.0
10 iterations	57.9	7.1	11.5	23.3	28.4

Notes: Estimates of the dynamic model, on 2001-2005. See the notes to Table 3.

estimating the persistence parameters in differences instead of levels (panel B), varying the numbers of firm classes and worker types (panels C and D), and estimating regression models (panel E). Using the dynamic model to re-classify firms, we obtain a larger variance contribution of firm effects and a lower correlation coefficient (panel F). This suggests that re-classifying firms does make a difference in the dynamic case, although the overall magnitudes are still quite comparable to the baseline two-step estimates of the dynamic model. Lastly, our framework delivers estimates of workers' mobility patterns among firms, which we report in the Supplemental Material (see Figure S8).

6.2 Dynamic effects

While the dynamic and static models have similar cross-sectional implications, the richer setting we consider in this section allows us to study dynamic aspects of worker mobility and earnings,

Table 7: Transition probabilities (×100), by conditional decile of previous earnings

	All			
	All movers	k'=1−3	k'=4−7	k'=8−10
k=1−3	2.20	0.84	1.06	0.30
	(0.03)	(0.02)	(0.03)	(0.02)
k=4−7	1.86	0.39	1.03	0.44
	(0.03)	(0.02)	(0.02)	(0.01)
k=8−10	2.90	0.47	1.00	1.43
	(0.06)	(0.06)	(0.03)	(0.02)
	First conditional decile of earnings			
	All movers	k'=1−3	k'=4−7	k'=8−10
k=1−3	3.41	1.53	1.52	0.37
	(0.21)	(0.09)	(0.12)	(0.05)
k=4−7	3.20	0.77	1.74	0.69
	(0.17)	(0.09)	(0.09)	(0.05)
k=8−10	4.92	0.93	1.74	2.25
	(0.31)	(0.15)	(0.14)	(0.13)
	Tenth conditional decile of earnings			
	All movers	k'=1−3	k'=4−7	k'=8−10
k=1−3	2.76	0.95	1.42	0.40
	(0.20)	(0.08)	(0.13)	(0.06)
k=4−7	1.82	0.35	1.05	0.42
	(0.10)	(0.04)	(0.06)	(0.03)
k=8−10	2.03	0.29	0.71	1.03
	(0.14)	(0.06)	(0.06)	(0.08)

Notes: We report the probability of changing job, overall and by destination firm class k', for each origin firm class k. In the top panel we show results for all workers. In the middle and bottom panels we show results for workers in the first and tenth deciles of log-earnings Y_{i2}, respectively, conditional on worker type α_i and current firm class $k_{i2} = k$. The estimates are obtained using the dynamic model, with $10,000,000$ simulations. We report parametric bootstrap standard errors in parentheses (computed using 200 replications).

which are interesting from both empirical and theoretical perspectives. We discuss endogenous mobility and dynamic dependence in turn.

Endogenous mobility. We first study how the current wage affects a worker's decision to move, and which firm she moves to. In models where there is no match heterogeneity and mobility is efficient, the wage before the move should not have an effect on the propensity to move, since a worker would always move towards higher-surplus firms irrespective of her current wage. In contrast, in wage posting models with match-specific heterogeneity, a worker is more (respectively, less) likely to accept job offers when her current match is of low (resp. high) quality. In addition, endogenous mobility is ruled out in AKM fixed-effects regressions.

In Table 7 we report the estimated probability of a job move conditional on the firm class at origin, overall and by firm class at destination. In the top panel we show those numbers in the full sample, while in the middle and bottom panels we select workers for whom the earnings rank before the move, given firm class and worker type, is below 10 or above 90 percents, respectively. In the first column we see that while the overall mobility rate lies between 2% and 3%, it is substantially higher for workers who had a low earnings realization in 2002. This suggests that workers are more likely to move when paid less. Such evidence of endogenous mobility is in line with estimates in Abowd et al. (2018), and it is consistent with the predictions of wage posting models with match-specific heterogeneity. In contrast, our estimates in the bottom panel suggest that high earnings realizations do not 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 $\xi_2(k')$ that we found.

State dependence and network effects. We next use our dynamic estimates to study how a worker's wage is affected by the firms she works for over time. Specifically, we focus on how much of the wage dispersion among similar workers is explained by the workers' previous employers. There are two different reasons why the previous employer may matter. First, working in a high-wage firm, say, may make a worker more likely to move to another high-wage firm. We refer to this effect, which is driven by mobility patterns across the network of firms, as the *network effect*. Second, the past firm may have a direct effect on the worker's wage after a job move. We call this the *state dependence effect*.

Network and state dependence effects are present in many models of earnings and mobility. Consider as an example sequential bargaining models where firms are characterized by their productivities (e.g., Postel-Vinay and Robin, 2002, Lise et al., 2016, Bagger and Lentz, 2014). Due to the on-the-job search process, the distribution of productivities of future employers depends on the productivity of the current employer (network effect). Moreover, due to the offer and counteroffer mechanism, 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). In AKM and our static model, network effects

Table 8: Decomposition of the share of variance of log-earnings of job movers explained by the firm class of their previous employer ($\times$100)

	total	**network effect**	**state dependence**
year after the move (2004)	2.57 (0.75)	0.84 (0.14)	1.74 (0.67)
two years after the move (2005)	2.04 (0.49)	0.98 (0.16)	1.05 (0.39)

Notes: Estimates of the dynamic model, on 2001-2005. In the first column we show the share of the within-worker-type variance of log-earnings of job movers explained by the firm class of their previous employer. In the second and third columns we decompose this share of variance into two components, "network effect" and "state dependence", which we define in the text. The estimates are based on $1,000,000$ *simulations. We report parametric bootstrap standard errors in parentheses (computed using* 200 *replications).*

are allowed for, but there is no state dependence.

Formally, we decompose the log-earnings variance explained by the previous firm class as follows, holding worker types constant and omitting them from the notation for conciseness:

$$\begin{aligned} \underbrace{\text{Var}\left(\mathbb{E}\left(Y_{i3}\,|\,k_{i2}\right)\right)}_{\text{total}} &= \text{Var}\Big(\,\mathbb{E}\Big[\mathbb{E}(Y_{i3}\,|\,k_{i3},k_{i2})\,\Big|\,k_{i2}\Big]\Big) \\ &= \underbrace{\text{Var}\Big(\,\mathbb{E}\Big[\mathbb{E}(Y_{i3}\,|\,k_{i3})\,\Big|\,k_{i2}\Big]\Big)}_{\text{network effect}} \\ &\qquad + \underbrace{\text{Var}\left(\mathbb{E}\left(Y_{i3}\,|\,k_{i2}\right)\right) - \text{Var}\Big(\,\mathbb{E}\Big[\mathbb{E}(Y_{i3}\,|\,k_{i3})\,\Big|\,k_{i2}\Big]\Big)}_{\text{state dependence effect}}. \end{aligned}$$

Note that the first term in the decomposition (i.e., the network effect) ignores the direct effect that the previous firm class k_{i2} could have on the wage Y_{i3} after the move, given the current class k_{i3}. Hence, this network effect measures the dependence of wages on the past firm class k_{i2} that is solely due to the dependence of k_{i3} on k_{i2}.

In Table 8 we report the results of this decomposition. We focus on the part of the variance of log-earnings following a job move that is net of worker heterogeneity. This within-type variance amounts to 40% of the overall variance after the move. In the first column we show that 2.6% of the log-earnings variance immediately after the move is explained by the previous employer.[14] Two years after the move (that is, in 2005), previous employers still explain 2% of

[14]Interestingly, variation in previous employers also explains .94% of the variance within the same *current*

the log-earnings variance.

In the second and third columns of Table 8 we show how much of the effect of the previous employer is due to network effect and state dependence, respectively. We find that, immediately after the move, one third of the total effect is due to the network effect (0.8%), while two thirds reflect state dependence (1.7%). Two years after the move (that is, in 2005), the total effect of the previous employer is split approximately equally between network effect and state dependence. Hence, according to our results, state dependence is of a similar order of magnitude (in fact, larger in the short run) compared to the network effect.

These findings suggest that, while it is important to study how workers move between firms, it is equally important to understand how wages are set dynamically around such moves. These issues have been studied theoretically in structural settings through the use of specific contracting environments. The results in this section should be of interest for the empirical modeling of mobility and earnings since, in addition to relying on exogenous mobility, standard static regression models rule out state dependence while leaving network effects fully unrestricted.

7 Conclusion

In this paper we propose a framework that allows for two-sided unobserved heterogeneity in matched employer employee data sets. We introduce empirical models that 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 several estimators.

Our application to Swedish administrative data shows that an additive model provides a good first-order approximation to 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. These findings, which are robust to a wide variety of specification checks, 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.

Using our dynamic model we find that endogenous mobility, by which earnings shocks affect mobility decisions, and state dependence and network effects, by which past firms have an impact on earnings after a job move, are features of our data. These findings support mechanisms

employer. This represents approximately 10% of the contribution of the firm class of the current employer.

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 models of workers and firms. In particular, it would be interesting to study how, under certain structural assumptions, our estimates could be used to reveal sorting patterns in terms of firm and worker productivity. Another interesting extension of our results would be to allow for time-varying processes of worker types that could respond to firm-level shocks.

Lastly, in this paper we have proposed 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 hope that our methods will 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., K. L. McKinney, and I. M. Schmutte (2018): "Modeling endogenous mobility in earnings determination," *Journal of Business & Economic Statistics*, 1–14.

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. 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.

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.*

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.

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 (2017): "Identifying equilibrium models of labor market sorting," *Econometrica*, 85, 29–65.

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.

Hwang, H.-s., D. T. Mortensen, and W. R. Reed (1998): "Hedonic wages and labor market search," *Journal of Labor Economics*, 16, 815–847.

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*.

Kline, P., R. Saggio, and M. Sølvsten (2018): "Leave-out estimation of variance components," *ArXiv e-prints*.

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.

Lindenlaub, I. (2017): "Sorting multidimensional types: Theory and application," *The Review of Economic Studies*, 84, 718–789.

Lise, J., C. Meghir, and J.-M. Robin (2016): "Matching, sorting and wages," *Review of Economic Dynamics*, 19, 63–87.

Lise, J. and F. Postel-Vinay (2015): "Multidimensional skills, sorting, and human capital accumulation," *Manuscript, University of Minnesota.*

Lise, J. and J.-M. Robin (2013): "The Macro-Dynamics of Sorting Between Workers and Firms," *to appear in American Economic Review.*

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

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.

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.

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. (2018): "Ranking firms using revealed preference," *The quarterly journal of economics*, 133, 1331–1393.

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

A.1 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:

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

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

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

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 a connecting 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 $F_{k_2,\sigma(\alpha)}$ 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 $F_{k',\sigma(\alpha)}$ is identified. Moreover, for each k' (possibly equal to k), using a connecting cycle as in the second part of Assumption 3 (i) we obtain by a similar argument that $F^m_{k',\sigma(\alpha)}$ is identified.

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

To show the last part of Theorem 1, note that by the first part of the proof 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 $\Pr[Y_{i1} \leq y_1 \,|\, k_{i1} = k]$, the $L \times 1$ vector $P(k)$ has generic element $q_k(\sigma(\alpha))$, and the columns of $F(k)$ have been ordered with respect to σ. By Assumption 3 (ii), $F(k)$ has rank 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.

A.2 Proof of Theorem 2

Part (i) is a direct application of Theorem 1, under Assumption 4.

For part (ii) we have, from (10):

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

where:

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

are the posterior probabilities of worker types given $Y_{i2} = y_2$, $k_{i2} = k$, and $m_{i1} = 0$, with $f_{k\alpha}$ denoting the 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:

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

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

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

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:

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

B Asymptotic properties

We consider a setting where the model is well-specified and there exists a partition of 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_k^0 denote the population class-specific cdfs, and $\|H\|^2 = \sum_{d=1}^{D} H(y_d)^2$.

Assumption B1. *(clustering)*

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

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

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

(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 B1 (i) could be relaxed to allow for some form of weak dependence 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 B1 (ii) and (iii) require that the clusters be large and well-separated in the population. Assumption B1 (iv) allows for asymptotic sequences where the number of workers per firm grows polynomially more slowly than the number of firms.

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).

Proposition B1. *Let Assumption B1 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 since $\mathbf{1}\{Y_{i1} \leq y_1\}$ is bounded. Assumptions 1d, 1e, and 1f hold because of Assumption B1 (i). Assumptions 2a and 2b hold by Assumptions B1 (ii) and (iii). Finally, Assumptions 2c and 2d are also satisfied by Assumption B1 (i) and boundedness of $\mathbf{1}\{Y_{i1} \leq y_1\}$. Theorems 1 and 2 in Bonhomme and Manresa (2015) and Assumption B1 (iv) then imply the result.

∎

We next turn to 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 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 B2. *(infeasible estimator)*

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

$$\sqrt{N}\left(\widetilde{\theta} - \theta^0\right) \xrightarrow{d} \mathcal{N}(0, \Omega).$$

Since $\widetilde{\theta}$ is a standard finite-dimensional maximum likelihood estimator, and observations are independent across individuals, Assumption B2 is not restrictive. Under correct specification, Ω 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 B1 and B2 hold. Then, as N tends to infinity:*

$$\sqrt{N}\left(\widehat{\theta} - \theta^0\right) \xrightarrow{d} \mathcal{N}(0, \Omega).$$

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 firm 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 Ω is different.

C 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 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, since 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, we discard all workers from the public sector or self-employed. 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 that 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. If a worker returns in 2004 to the firm she worked for in 2002, we do not consider this worker to be a mover (4.3% of the sample).

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 have conducted robustness checks, in our preferred specification we do not include these job moves since 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 in the 2002-2004 and 2001-2005 samples, as well as in the subsamples of job movers. The numbers can be found in Table C1. The 2002-2004 sample contains about 600,000 workers and 44,000 firms. Hence, the average number of workers per firm is 13.7. The mean firm size as reported by the firm is higher, 37.6, due to our sample selection that focuses on fully employed male workers. In the 2001-2005 sample, the mean number of workers and mean reported size are 12.3 and 37.1, respectively. The distribution of firm size is skewed, and medians are smaller. At the same time, reported firm sizes in the subsamples of movers are substantially higher.

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.[15] To understand

[15] See their Figure 7.14. Skans et al. (2009) report the fraction of workers employed in plants with more

how our sample selection influences the mobility rate, we have computed similar descriptive statistics in the entire 2002-2004 sample, without imposing that workers are fully employed in 2002 and 2004 or that firms exist in the two periods. 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. Although we prefer to focus on a more restrictive definition for estimation, as robustness checks 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-firm 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 and 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 C1, we see that job movers are on average younger and more educated than workers who remain in the same firm. They also tend to work more in service sectors as opposed to manufacturing. At the same time, characteristics of job movers and stayers show substantial overlap.

than 25 employees in years $t-1$ and t who changed plant between $t-1$ and t. As a comparison, in Germany Fitzenberger and Garloff (2007) report yearly between-employers transition rates of 7.5% in the period 1976 to 1996 for male workers.

Table C1: 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
% 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

Notes: Descriptive statistics for a sample from the Swedish registry data that includes 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), for firms that are continuously present in the sample. The numbers in the table correspond to 2002.