# How Much Should we Trust Estimates of Firm Effects and Worker Sorting?

Stéphane Bonhomme Kerstin Holzheu Thibaut Lamadon
Elena Manresa Magne Mogstad Bradley Setzler

June 7, 2022

**Abstract**

Many studies use matched employer-employee data to estimate a statistical model of earnings determination with worker and firm fixed effects. Estimates based on this model have produced influential yet controversial conclusions. The objective of this paper is to assess the sensitivity of these conclusions to the biases that arise because of limited mobility of workers across firms. We use employer-employee data from the US and several European countries while taking advantage of both fixed-effects and random-effects methods for bias-correction. We find that limited mobility bias is severe and that bias-correction is important.

---

Bonhomme: University of Chicago (sbonhomme@uchicago.edu); Holzheu: Sciences Po (kerstin.holzheu@gmail.com); Lamadon: University of Chicago, NBER, IFAU and IFS (lamadon@uchicago.edu); Manresa: New York University (em1849@nyu.edu); Mogstad: University of Chicago, Statistics Norway, NBER, and IFS (magne.mogstad@gmail.com); Setzler: Pennsylvania State University (bradley.setzler@gmail.com). We thank David Wiczer, István Boza, Raffaele Saggio, and participants at various conferences and seminars for useful comments. The opinions expressed in this paper are those of the authors alone and do not reflect the views of the Internal Revenue Service or the US Treasury Department. We thank SOI, Statistics Norway, the IFAU and Statistics Sweden, the Fondazione Rodolfo De Benedetti, as well as the Austrian Labor Market Service and the Austrian Federal Ministry for Social Affairs, Health, Care and Consumer Protection for granting access to all data sources. This work is a component of a larger project on income risk in the United States, conducted through the SOI Joint Statistical Research Program. Mogstad and Setzler acknowledge funding from NSF Grant SES-1851808. Bonhomme, Lamadon and Manresa acknowledge funding from NSF Grant SES-1658920.

# 1 Introduction

Abowd, Kramarz, and Margolis (1999) (AKM hereafter) proposed a statistical model that uses employer-employee data to quantify the contributions of workers and firms to earnings inequality. In the AKM model, log-earnings are expressed as a sum of worker effects, firm effects, covariates, and idiosyncratic error terms. AKM showed how to estimate worker and firm effects using a fixed-effects (FE) estimator. The resulting estimates can then be used to decompose the variance of log-earnings into the contributions of worker heterogeneity, firm heterogeneity, and sorting of high-wage workers to high-paying firms.

Over the past two decades, the AKM model and FE estimator have been frequently used to analyze earnings inequality in many developed countries.[1] This work has produced several influential yet controversial conclusions, summarized in the review article by Card et al. (2018). One key conclusion is that firm-specific wage settings are important for earnings inequality. Card et al. (2018) concludes, "This literature also finds that firms play an important role in wage determination, with a typical finding that about 20% of the variance of wages is attributable to stable firm wage effects." Another key conclusion is that the correlation between firm and worker effects is often small and sometimes negative, indicating little if any sorting of high-wage workers to high-paying firm. At the same time, evidence from Germany (Card et al., 2013) and the US (Song et al., 2019) indicate that worker sorting has been increasing over time, driving much of the rise in earnings inequality in these countries.

These empirical findings have been important, not only for quantifying the sources of earnings inequality, but also for how economists model the labor market. For example, if firm effects are a key source of inequality, then it is natural to ask why similar workers are paid differently. Indeed, the evidence of significant firm effects was instrumental in the development of labor market models with frictions (Mortensen, 2003). Furthermore, if better workers do not sort to more productive firms, then one might question the empirical importance of production complementarities for

[1] See, among many others, Gruetter and Lalive (2009), Mendes et al. (2010), Card et al. (2013), Goldschmidt and Schmieder (2017), Card et al. (2016), Sorkin (2018), and Song et al. (2019). The AKM approach has also been widely used in contexts other than firms and workers, including teachers and students (e.g., Rockoff, 2004), hospitals and patients (e.g., Finkelstein et al., 2016), and banks and firms (e.g., Amiti and Weinstein, 2018).

the matching of workers and firms (Shimer and Smith, 2000, Eeckhout and Kircher, 2011).

Motivated by the importance of the findings from AKM, we ask the question: How much should we trust the FE estimates of firm effects and worker sorting? We focus on the problem of estimation, taking as given the AKM model. In particular, we assume that mobility is conditionally exogenous given worker and firm effects, and we rule out the presence of dynamics and worker-firm complementarities. Other work has examined and relaxed these assumptions (e.g., Abowd et al., 2018, Bonhomme et al., 2019).[2] Our goal is to assess the sensitivity of the FE estimator to the incidental parameter problem that arises in the AKM model, often referred to as "limited mobility bias".

Limited mobility bias is due to the large number of firm-specific parameters that are solely identified from workers who move across firms. Abowd et al. (2004) and Andrews et al. (2008, 2012) highlighted this problem, and the simulations reported in Andrews et al. (2008) suggest the bias can be substantial. If firms are weakly connected to one another because of limited mobility of workers across firms, FE estimates of the contribution of firm effects to wage inequality are biased upwards while FE estimates of the contribution of the sorting of workers to firms are biased downwards.

Even though researchers have been aware of limited mobility bias for nearly two decades, very few papers have used methods for bias correction. None of the empirical papers in the survey by Card et al. (2018) correct for limited mobility bias.[3] There could be a variety of reasons for this. As Card et al. (2018) point out, bias correction necessarily involves making potentially restrictive assumptions about the model. In addition, exact computation of fixed-effects corrections could be costly, and possibly prohibitive in large data sets. As a result, there is yet no consensus about the mag-

[2]In addition, recent work has studied worker sorting with two-sided heterogeneity using different approaches (e.g., Bagger and Lentz, 2019, Hagedorn et al., 2017, Lentz et al., 2017, and Borovickova and Shimer, 2017).

[3]Several papers published since Card et al. (2018) did not use bias correction (see e.g. Song et al. 2019, Sorkin 2018, and Gerard et al. 2021). Notable exceptions include Kline et al. (2020), and Lachowska et al. (Forthcoming), who apply fixed-effects methods for bias correction to data from two regions of Italy and one US state (Washington), as well as Bonhomme et al. (2019) and Lamadon et al. (2022), who have estimated linear and nonlinear models with discrete firm heterogeneity using data from Sweden and the US. Here we develop a correlated random-effects estimator and apply both fixed-effects and random-effects methods to data from a wide range of countries.

nitude of the biases, and how they might alter conclusions about labor markets and inequality.

To investigate the importance of limited mobility bias, we use a variety of data sets and methods. Empirically, we study matched employer-employee data from Austria, Italy, Norway, Sweden, and the US. These countries have different wage structures and labor market institutions. By comparing the results across countries, we shed light on whether our findings are specific to the US or common across several Western economies that could potentially differ in the importance of firm-specific wage-setting and the patterns of worker mobility across firms.

Methodologically, we take advantage of the availability of econometric techniques for bias-correction. We implement fixed-effects methods for bias-correction, originally proposed by Andrews et al. (2008) and developed further by Kline et al. (2020). In addition, we propose a correlated random-effects method that builds on Woodcock (2008) and Bonhomme et al. (2019). There are advantages and disadvantages to both the random- and fixed-effects methods for bias-correction. By comparing the results across the methods, we learn whether the conclusions about limited mobility bias are sensitive or robust to the alternative approaches to bias-correction. To improve researcher accessibility to bias-correction methods, we have released a comprehensive, user-friendly software package for implementing all of the bias-correction methods shown in the paper at https://github.com/tlamadon/pytwoway. It is written in Python, but includes a Stata extension so that Stata users can perform the bias corrections as well.

Our analyses deliver several important conclusions for empirical work using the AKM model. First, we show in simulations based on real data that limited mobility bias is empirically important and existing methods for bias correction perform well even as mobility becomes very limited. Second, in all the countries we consider, we find that limited mobility bias is a major empirical issue for studies using FE to document firm effects and worker sorting. Once bias is accounted for, firm effects dispersion matters less for earnings inequality and worker sorting becomes always positive and typically strong. Third, alternative methods for bias correction based on different assumptions tend to produce broadly similar results. This is reassuring, as bias correction necessarily involves making restrictive assumptions about the model and/or limiting the set of firms under consideration.

To preview our estimates and put our results into perspective, Figure 1 compares our bias-corrected estimates to existing FE estimates of the contribution from firm effects and worker sorting to wage or earnings inequality. We report FE estimates from previous studies in white bars. Then, for each of the five countries of study, we report estimates based on our correlated random-effects (CRE) method using the firm grouping of Bonhomme et al. (2019) in black, and estimates based on the heteroskedastic fixed-effects method (FE-HE) of Kline et al. (2020) in grey.[4] In Subfigure 1a, we focus on the contribution of firm effects. The interquartile range of estimates of the variance of firm effects in previous studies is from 16% to 25%, while the range of our bias-corrected estimates is from 5% to 13% using CRE and 6% to 16% using FE-HE. In Subfigure 1b, we shift attention to the contribution of sorting. The interquartile range of estimates of the contribution of sorting in previous studies is from -2% to 18%, while the range of our bias-corrected estimates is from 10% to 20% using CRE, and 5% to 13% using FE-HE.

# 2 Data

For each country, we now discuss the data sources that we use, before reporting specific sample selection rules due to data structure and variable availability. Next, we describe the procedure we use to harmonize the sample and variable definitions across countries.

## 2.1 Data Sources

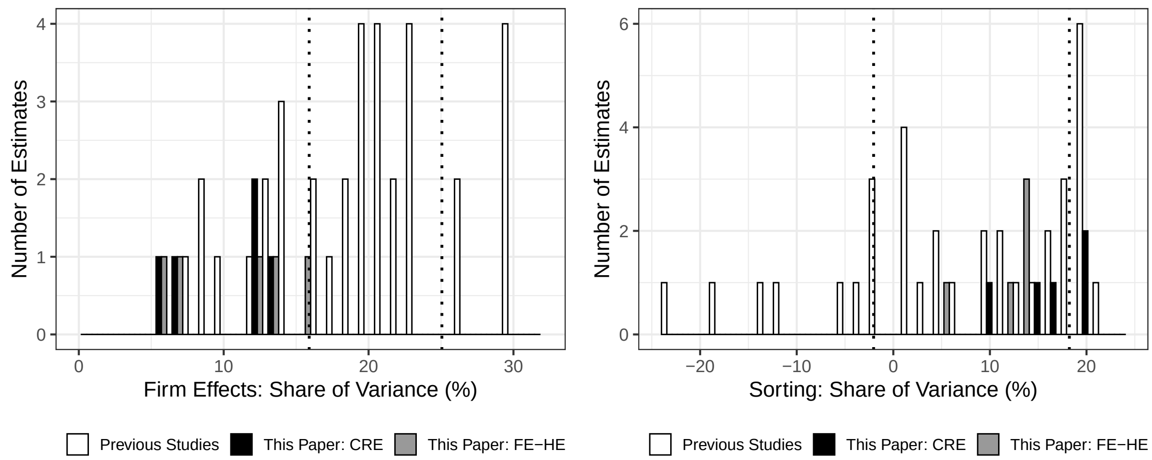
### 2.1.1 United States

The US data are constructed by linking Treasury business tax filings with worker-level filings. Our sample spans 2001-2015 and our main results focus on 2010-2015. We express all monetary variables in 2015 dollars, adjusting for inflation using the

[4] See Appendix Table F1 for a list of the 18 studies and 37 FE estimates used in this comparison. These estimates are directly comparable to those based on CRE, as both FE and CRE use the largest set of firms that are connected through at least one mover. We also report the results using FE-HE, which restricts attention to a subset of firms that remain connected after any given mover is removed from the sample (a leave-one-out connected component).

Figure 1: Comparison to Existing Studies

(a) Firm effects (b) Sorting




*Notes:* FE estimates from previous studies in white bars. Correlated random-effects (CRE) bias-corrected estimates from this paper based on the grouping of Bonhomme et al. (2019) in black. Heteroskedastic fixed-effects (FE-HE) bias-corrected estimates from this paper using the method of Kline et al. (2020) in greu. The vertical dotted lines indicate the interquartile range of estimates in previous studies.

CPI. Earnings data are based on taxable remuneration for labor services reported on form W-2 for direct employees. Earnings include wages and salaries, bonuses, tips, exercised stock options, and other sources of income deemed taxable by the IRS. These forms are filed by the firm on behalf of the worker and provide the firm-worker link. We exclude workers who are employed in the public or non-profit sector by requiring that their employers file tax form 1120 (C-corporations), 1120S (S-corporations), or 1065 (partnerships). In the US data, we do not observe any information about the duration of the spell within the year.[5] To construct a comparable sample to previous studies in other countries and in the US, we apply a full-time equivalence earnings threshold, as we describe in detail in the next subsection.

[5] For this reason, our main analysis allows for workers to change employers during the year. However, we observe start and end dates of employment spells for the European countries and can use these dates to restrict the sample to full-year employees. Appendix Figure F16 presents results for the sample of workers that are employed by a single firm for the entire calendar year in each European country. The bias-corrected estimates are similar when using this alternate sample definition.

### 2.1.2 European countries

Each European country allows for the construction of a matched employer-employee data set with information on total annual earnings paid to each worker by each employer. This measure of earnings includes both direct wage payments and other sources of labor income. All data sources include information on the worker's age and gender. Countries differ in the level of detail regarding the duration of the employment spells as well as the calendar years over which data are available. In each country, we focus on 6-year panels in the main analysis and provide results from 3-year panels for comparison, and we adjust all monetary variables for inflation.

**Austria.** The Austrian data, called the Arbeitsmaktdatenbank (AMDB), are co-constructed by the Austrian Labor Market Service and the Federal Ministry for Social Affairs, Health, Care and Consumer Protection using worker-level social security records. Our sample spans the years 2010-2015. A similar vintage of these data have been used by Borovickova and Shimer (2017). For each job, the data include information on start and end dates as well as total annual earnings. Given this information, we construct the daily average wage as our main outcome of interest.

**Italy.** The Italian data, known as the Veneto Worker Histories, were developed by the Economics Department at Università Ca' Foscari Venezia under the supervision of Giuseppe Tattara. These data are constructed by tracking all workers in the provinces Treviso and Vicenza even if they move to other provinces in Italy. Our sample spans 1996-2001. These data have been used for instance by Kline et al. (2020). For each job, it includes information on number of days worked in the year and annual earnings. Given this information, we construct the daily average wage as our main outcome of interest.

**Norway.** The Norwegian data come from the State Register of Employers and Employees, which covers the universe of workers and firms. Our sample spans 2009-2014. For each job, the data include information on start and end dates, annual earnings, and contracted hours. We construct the daily average wage as our main outcome of interest. Because the Norwegian data also provide hours worked per day, we construct the average hourly wage as a secondary outcome.

**Sweden.** The Swedish data we use build on the sample from Friedrich et al. (2019), and we focus on 2000-2005. The employee-employer link is built from the Register-Based Labor Market Statistics (RAMS), with access provided by the Institute for Evaluation of Labour Market and Education Policy (IFAU). The data cover the universe of workers and firms, but the sample available to us is limited to employment spells of at least two months. The sample contains information about yearly earnings, employer identifiers and month of start and end of each spell. Given this information, we construct average monthly earnings as our main outcome of interest.

## 2.2 Sample Harmonization and Construction

To harmonize the data across countries, we apply five steps. First, as is common in the literature, whenever a worker is employed by multiple employers in the same year, we focus on the employer associated with the greatest annual earnings. Second, we restrict attention to workers employed in the private sector. Third, we restrict attention to workers who are between 25 and 60 years of age. Fourth, we adjust for differences in age and time by regressing the outcome measure on calendar year indicators and an age profile. We follow Card et al. (2018) in specifying the age profile as a third-order polynomial which is flat at age 40.

Lastly, we restrict attention to full-time equivalent (FTE) workers. Since we do not observe hours worked in US data, or a formal measure of full-time employment, we follow Lamadon et al. (2022) in defining a worker as FTE if annual earnings exceed \$15,000, which is approximately the annualized minimum wage and corresponds to 32.5% of the national average. To harmonize the sample selection across countries, we similarly restrict the European samples to workers with annual earnings above 32.5% of the national average.[6] In Subsection 6.2, we assess the sensitivity of the Norwegian estimates to using annual earnings (as in the US), daily wages (as in Italy, Sweden, Austria) and hourly wages as the outcome variable.

Given these harmonized samples, we prepare them for estimation by collapsing the annual observations over each 6-year panel into employment spells. Since we do not want to make assumptions about serial correlation within employment spells, we only

[6] In Appendix Figure F2, we consider a range of FTE thresholds from \$3,750 (about 25% of the annualized minimum wage) to \$15,000 (about 100% of the annualized minimum wage). As shown in this figure, our findings about limited mobility bias are robust to the choice of FTE threshold.

use the mean log earnings within a spell, which is sufficient to construct our estimators of interest. This approach allows for partial-year employment when constructing spells. In Appendix Figures F3 (for the US) and F16 (for the European countries), we apply sample restrictions meant to capture only full-year employment in these spells, finding that the conclusions are unchanged. For workers that move across employers, we further reshape the spell data into an event study format that compares the spell-level log earnings or wage measures before and after a job change. A worker that does not move across employers has only one observation. This structure effectively reduces the data to the information needed for the identification of firm effects and sorting. See Appendix A for additional details.

## 2.3 Descriptive Statistics

We next present descriptive information about sample sizes, distributions of moves, and earnings or wage inequality. Table 1 provides descriptive statistics for the five countries we study. It characterizes the full population (first column, under each country), the connected set (second column), and the leave-one-out set (third column). The bias-correction methods recover variance components on these two sets, as we explain in the next two sections. These sets are constructed by computing the largest set of firms that are connected by at least one mover (connected set), and the largest set of firms that remain connected after any given mover is removed from the sample (leave-one-out set). The rows report information on the number of firms and workers, the distribution of the number of moves per firm, and certain moments of the distribution of log earnings or wages.

In Table 1 and in our main analysis, a mover is defined as a worker that is employed by at least two different firms during the sample period. In Appendix Figure F3, we consider a stricter mover definition in which a worker must be employed for at least 3 consecutive years at the first firm and at least 3 consecutive years at the second firm, only measuring earnings during intermediate years within these 3-year spells. This does not materially alter our conclusions. In Subsection 6.2, we further discuss the impact of the definition of job movers on the results.

Table 1 highlights several key features of the data. First, we see that at least 93% of workers belong to the connected set in each country and at least 87% belong to

Table 1: Sample Characteristics

| | Austria | | | Italy | | | Norway | | | Sweden | | | US | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Set:* | | | | | | | | | | | | | | | |
| Baseline Years | | 2010-2015 | | | 1996-2001 | | | 2009-2014 | | | 2000-2005 | | | 2010-2015 | |
| Full Set | ✓ | × | × | ✓ | × | × | ✓ | × | × | ✓ | × | × | ✓ | × | × |
| Connected Set | × | ✓ | × | × | ✓ | × | × | ✓ | × | × | ✓ | × | × | ✓ | × |
| Leave-one-out Set | × | × | ✓ | × | × | ✓ | × | × | ✓ | × | × | ✓ | × | × | ✓ |
| *Sample Counts (1,000):* | | | | | | | | | | | | | | | |
| Unique Firms | 446 | 206 | 140 | 198 | 92 | 61 | 233 | 114 | 78 | 136 | 63 | 52 | 7,565 | 2,568 | 1,689 |
| (Share of Full Set) | (100%) | (46%) | (31%) | (100%) | (47%) | (31%) | (100%) | (49%) | (34%) | (100%) | (46%) | (38%) | (100%) | (34%) | (22%) |
| Unique Workers | 3,582 | 3,396 | 3,240 | 1,188 | 1,111 | 1,034 | 1,379 | 1,286 | 1,199 | 1,979 | 1,921 | 1,850 | 59,621 | 55,464 | 52,484 |
| (Share of Full Set) | (100%) | (95%) | (90%) | (100%) | (94%) | (87%) | (100%) | (93%) | (87%) | (100%) | (97%) | (93%) | (100%) | (93%) | (88%) |
| *Distribution of Moves:* | | | | | | | | | | | | | | | |
| Moves per Firm | 2 | 5 | 8 | 2 | 4 | 6 | 2 | 5 | 7 | 4 | 10 | 11 | 2 | 6 | 8 |
| Worker-weighted quantiles: | | | | | | | | | | | | | | | |
| 10th Quantile | 4 | 4 | 5 | 3 | 3 | 4 | 3 | 3 | 4 | 4 | 5 | 6 | 3 | 4 | 5 |
| 50th Quantile | 52 | 51 | 56 | 22 | 22 | 25 | 26 | 26 | 29 | 77 | 77 | 82 | 56 | 58 | 67 |
| 90th Quantile | 605 | 605 | 629 | 313 | 311 | 326 | 397 | 399 | 420 | 2,354 | 2,352 | 2,484 | 4,214 | 4,304 | 4,676 |
| *Log Earnings Distrib.:* | | | | | | | | | | | | | | | |
| Variance | 0.195 | 0.187 | 0.182 | 0.169 | 0.167 | 0.168 | 0.241 | 0.239 | 0.236 | 0.164 | 0.164 | 0.164 | 0.413 | 0.414 | 0.416 |
| Between-firm Share | 43% | 46% | 44% | 46% | 46% | 45% | 47% | 47% | 46% | 31% | 32% | 31% | 40% | 40% | 39% |

*Notes:* This table displays descriptive statistics on the baseline panel data from the US and four European countries. For each country, it provides information on the full set, connected set, and leave-one-out set.

the leave-one-out set. By contrast, less than half of all firms belong to the connected set, and far fewer belong to the leave-one-out set. This indicates that, within each country, a large share of firms are very small, account for little of overall employment, and are not connected to other firms by movers. In Section 5, we further discuss the differences between the connected and leave-one-out sets.

Second, while each country has a large number of moves for the median firm, a substantial share of firms have a small number of moves. For example, in the US, the majority of firms have at least 58 moves in the connected set and 67 moves in the leave-one-out set. However, ten percent of firms have only 4 moves in the connected set and 5 moves in the leave-one-out set.

Third, while earnings or wage inequality varies substantially across countries, the between-firm share of variance tends to be more similar, ranging from 30% in Sweden to 45% in Austria and Italy. The between-firm component captures differences across firms in mean log earnings or wages. Thus, it may reflect firm effects or systematic heterogeneity in the workers that firms hire. To disentangle these two components, the AKM model takes advantage of workers moving across firms, as formalized in Section 3.

Before describing the AKM model and estimator, in Appendix Figure F4 we present an event study of the earnings changes experienced by workers moving between different types of firms, in the US sample. Following Card et al. (2013) and Card et al. (2018) we define firm groups based on the average pay of coworkers. As in previous studies, we find that workers who move to firms with more highly-paid coworkers experience earnings raises while those who move in the opposite direction experience earnings decreases of similar magnitude, and that the gains and losses for movers in opposite directions between any two groups of firms seem fairly symmetric. By comparison, earnings do not change materially when workers move between firms with similarly paid coworkers. In addition, the earnings profiles of the various groups are all relatively stable in the years before and after a job move. This lends some support to the mobility assumption in the AKM model that workers do not select their firms based on idiosyncratic earnings growth.

# 3 AKM Estimator and Limited Mobility Bias

In this section, we first describe the AKM estimator of Abowd et al. (1999), and we then provide initial evidence on the presence of bias in the US and Sweden.

## 3.1 Model, Estimator and Biases

The AKM model is

$$Y_{it} = X_{it}'\beta + \alpha_i + \psi_{j(i,t)} + \varepsilon_{it}, \tag{1}$$

where $Y_{it}$ are the log-earnings of worker $i$ in period $t$, $X_{it}$ are exogenous covariates such as age or calendar time, $\alpha_i$ is the unobserved worker effect, $j(i,t)$ is the firm where $i$ works at $t$, $\psi_{j(i,t)}$ is the unobserved firm effect, and $\varepsilon_{it}$ is an idiosyncratic error term. We denote as $N$ the number of workers, $J$ the number of firms, and $T$ the number of time periods. Following AKM, we assume that the following mean independence condition holds:

$$\mathbb{E}(\varepsilon_{it} \,|\, X_{11}, ..., X_{NT}, j(1,1), ..., j(N,T), \alpha_1, ..., \alpha_N, \psi_1, ..., \psi_J) = 0. \tag{2}$$

Throughout this paper, we assume that (2) holds in model (1), so the AKM model is correctly specified. This assumption allows for unrestricted dependence of job mobility on firm and worker effects. For instance, high-wage workers may be more likely to move to higher-paying firms than low-wage workers. However, assuming that shocks $\varepsilon_{it}$ are mean independent of past and future firm indicators rules out endogenous mobility with respect to shocks and state dependence, which are important in dynamic models with wage posting or sequential bargaining. In addition, (1) and (2) imply that the conditional mean of log-earnings is additive in worker and firm effects. Additivity rules out interactions between worker effects $\alpha_i$ and firm effects $\psi_{j(i,t)}$ that may be economically relevant (e.g., Abowd et al., 2018, Bonhomme et al., 2019).

In this model, we focus on the contributions of firm effects and sorting in the following variance decomposition

$$\mathrm{Var}(Y_{it} - X_{it}'\beta) = \underbrace{\mathrm{Var}(\alpha_i)}_{\text{Worker effects}} + \underbrace{\mathrm{Var}(\psi_{j(i,t)})}_{\text{Firm effects}} + \underbrace{2\mathrm{Cov}(\alpha_i, \psi_{j(i,t)})}_{\text{Sorting}} + \underbrace{\mathrm{Var}(\varepsilon_{i,t})}_{\text{Residual}}. \quad (3)$$

We now describe the AKM estimator of the "Firm effects" and "Sorting" components in this decomposition.

The AKM or "fixed-effects" (FE) estimator treats $\alpha = (\alpha_1, ..., \alpha_N)'$ and $\psi = (\psi_1, ..., \psi_J)'$ as parameter vectors. It is convenient to write (1) and (2) in vector form, as

$$Y = X\beta + A\gamma + \varepsilon, \quad \mathbb{E}(\varepsilon \,|\, X, A, \gamma) = 0, \quad (4)$$

where $Y$ and $\varepsilon$ are $NT \times 1$, $X$ is a matrix with $NT$ rows, and $A$ is a matrix with $NT$ rows and $N + J$ columns.[7] The vector $\gamma = (\alpha', \psi')'$ includes worker and firm effects, and the matrix $A = [A_W \; A_F]$ depends on worker and firm indicators.

The slope parameter $\beta$ can be estimated using OLS after partialling out worker and firm indicators.[8] For simplicity, in the presentation we treat $\beta$ as known, and redefine $Y_{it} - X_{it}'\beta$ as the outcome variable. That is, we work with the model

$$Y = A\gamma + \varepsilon, \quad \mathbb{E}(\varepsilon \,|\, A, \gamma) = 0. \quad (5)$$

[7]Note the conditioning on $\alpha$ and $\psi$ in (4) is not necessary here, since we are treating them as deterministic parameters. In random-effects methods below we treat $\alpha$ and $\psi$ as random.

[8]Formally, denote as $A^\dagger$ the Moore-Penrose inverse of $A$, and as $M_A = I - AA^\dagger$ the residual "hat" projection matrix. The FE estimator of $\beta$ is $\widehat{\beta} = (X'M_AX)^{-1}(X'M_AY)$. When $A'A$ is non-singular, $M_A = I - A(A'A)^{-1}A'$, however $M_A$ remains well-defined under singularity.

We start by assuming that $A'A$ is non-singular. This requires working within a connected component of the firm-worker graph (Abowd et al., 2002), and imposing one normalization on $\gamma$, e.g., one of the firm effects being equal to zero. With some abuse of notation we still denote as $A$ the resulting selection of rows and columns of the $A$ matrix, and we redefine $N, J, T$ appropriately. Then, the FE estimator of worker and firm effects is the least-squares estimator

$$\widehat{\gamma} = (A'A)^{-1}A'Y.$$

As in other studies using the AKM model, we are interested in the variance components in (3), such as the variance of firm effects $\text{Var}(\psi_{j(i,t)})$ and the covariance between worker and firm effects $\text{Cov}(\alpha_i, \psi_{j(i,t)})$. Variance components can be written as quadratic forms in $\gamma$; that is, $V_Q = \gamma' Q \gamma$ for some matrix $Q$. Note that $Q$ typically depends on $A$, although we leave the dependence implicit in the notation. The FE estimator of $V_Q$ is then

$$\widehat{V}_Q^{\text{FE}} = \widehat{\gamma}' Q \widehat{\gamma}.$$

To see that $\widehat{V}_Q^{\text{FE}}$ is biased, note that

$$\mathbb{E}[\widehat{V}_Q^{\text{FE}}] = V_Q + \underbrace{\mathbb{E}[\varepsilon' A(A'A)^{-1}Q(A'A)^{-1}A'\varepsilon]}_{=\text{Bias}_Q}, \tag{6}$$

where the expectations are conditional on $A$ and $\gamma$. The expected FE estimator $\mathbb{E}[\widehat{V}_Q^{\text{FE}}]$ differs from the true variance component $V_Q$ in general, due to the presence of the bias term $\text{Bias}_Q$. Note that the bias is due to $V_Q$ being quadratic in $\gamma$. In contrast, the FE estimates $\widehat{\gamma}$ of the level of worker and firm effects are unbiased under (4).

As explained by Andrews et al. (2008), the bias intuitively arises from an insufficient number of job movers in the firm. As a result of "limited mobility bias", the FE variance of firm effects tends to be overstated. In turn, the covariance between worker and firm effects tends to be negatively biased, since worker effects and firm effects enter (1) additively. Jochmans and Weidner (2019) show that the magnitude of the bias is inversely related to the degree of connectivity of the firm-worker graph. A limiting case is when the graph is disconnected, i.e., when $A'A$ is singular. Within a connected component, the bias can still be large when connectivity is weak. An implication of their analysis is that the structure of the bias is complex, since it depends on the (large) matrix $A$ of worker and firm indicators. Hence, the magnitude of the bias is ultimately an empirical question.

## 3.2 Empirical Illustration of Limited Mobility Bias

To get a sense of the scope for limited mobility bias, an informal approach is to apply the estimator of Abowd et al. (1999) (FE) to alternative samples of workers and firms that are comparable except for the number of movers per firm. Figure 2 present the results from such an analysis for Sweden using a subsampling strategy inspired by Andrews et al. (2008, 2012).[9]

In Figure 2, we randomly remove movers from firms while keeping the connected set of firms fixed in order to understand how the FE estimator responds to reduced worker mobility. To do so, we begin by considering the set of firms in Sweden with a relatively large number of movers; that is, at least 15 movers per firm over a six year period. Next, we remove movers randomly within each firm based on a pre-specified sampling probability, resulting in various subsamples in which mobility is more limited. Then, we restrict each of these subsamples to the set of firms that belongs to the connected set when imposing the smallest sampling probability (which is 10% in practice). This ensures that the set of firms is kept fixed as we compare across samples.[10] Lastly, we apply the FE estimator to each simulated subsample. For completeness, we repeat this exercise for the leave-one-out connected set of firms.[11] For each simulated subsample, we also report bias-corrected estimates both for the correlated random-effects and the fixed-effects methods. We discuss these bias-corrected estimates in Section 4.

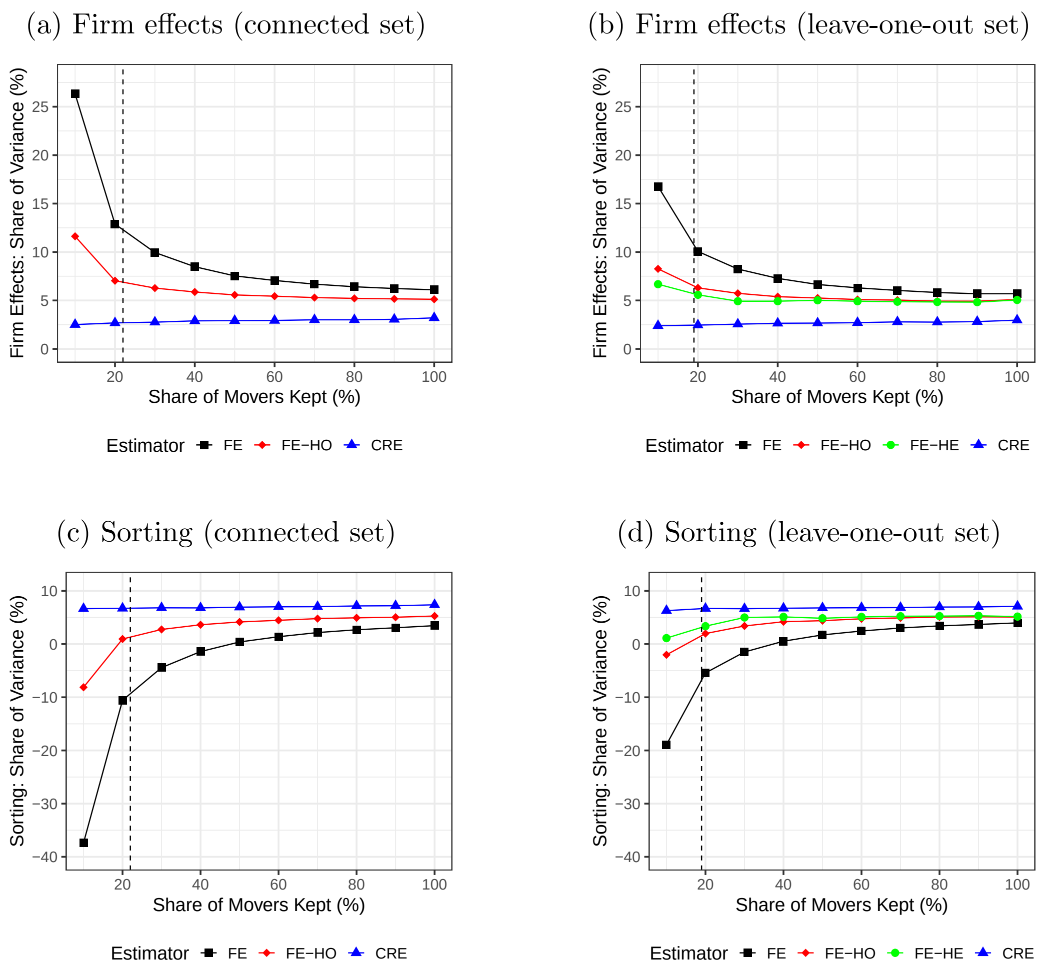
Subfigure 2a provides estimates of the contribution of firm effects to earnings inequality, i.e., $\text{Var}(\psi_{j(i,t)})/\,\text{Var}(Y_{it})$, for the connected set. Focusing on the FE estimates in the black line, we find that the variance of firm effects declines monotonically as the number of movers per firm increases. Consistent with limited mobility bias,

[9]See Appendix Figure F5 for a similar analysis for the US.

[10]In our working paper (Bonhomme et al., 2020), we do not impose this restriction, instead allowing the set of firms inncluded in the connected set to become smaller as movers are removed, finding similar results.

[11]The heteroskedastic fixed-effects method for bias-correction of Kline et al. (2020) recovers estimates of the variance components on the leave-one-out connected set.

Figure 2: Evidence on Limited Mobility Bias in Sweden




*Notes:* In this figure, we consider the subset of firms in Sweden with at least 15 movers. We randomly remove movers within each firm and re-estimate the variance of firm effects and covariance between firm and worker effects using the various estimators. For each estimator, we repeat this procedure twenty times then average the estimates across repetitions. The procedure allows us to keep the connected or leave-one-out set of firms the same and examine the bias that results from having fewer movers available in estimation. The vertical dashed line approximates the point at which movers per firm in this sample matches movers per firm in the full sample.

the fewer the number of movers per firm, the larger the variance of firm effects. For the same set of firms, the estimated variance of firm effects is about twice as large (13%) if we only keep 20% of the movers within each firm (on average, 4 movers per firm) as compared to the estimate of 6% we obtain if we keep all the movers per firm (at a minimum 15 and, on average, 45 movers per firm). By way of comparison, there are around 10 movers per firm in the full estimation sample, which roughly corresponds to the number of movers per firm when randomly keeping 22% of movers in the sample with originally 15 or more movers per firm, as indicated by a dashed vertical line. Subfigure 2b repeats this analysis for the leave-one-out set. The results are similar for the leave-one-out set, though FE is subject to less limited mobility bias, reflecting that the leave-one-out set has more movers per firm.

Subfigure 2c provides estimates of the contribution of worker sorting to earnings inequality, i.e., $2\,\mathrm{Cov}(\alpha_i, \psi_{j(i,t)})/\,\mathrm{Var}(Y_{it})$, for the connected set. Focusing again on the FE estimates in the black line, we find that the covariance between worker and firm effects increases monotonically as the number of movers per firm increases. For the same set of firms, the FE estimate of the contribution of worker sorting to earnings inequality is about 3% when we keep all movers per firm. However, if we only keep 20% of the movers within each firm, the FE estimates turn negative and large in magnitude. Subfigure 2d repeats this analysis for the leave-one-out set. The results are again broadly similar for the leave-one-out set.

# 4 Bias-correction: methods and illustration

In this section, we describe the fixed-effects and random-effects methods we use for bias-correction and illustrate the methods empirically.

## 4.1 Methods

**Fixed-effects.** Andrews et al. (2008) note that the bias in (6) can be written as

$$\mathrm{Bias}_Q = \mathrm{Trace}\left(A(A'A)^{-1}Q(A'A)^{-1}A'\Omega(A)\right),$$

where $\Omega(A) = \mathrm{Var}(\varepsilon \,|\, A)$ is the covariance matrix of errors. Andrews et al. (2008) propose an estimator of the bias in the homoskedastic case, under the assumption that $\Omega(A) = \sigma^2 I$, for $I$ the identity matrix. Specifically, they construct

$$\widehat{\text{Bias}}_Q^{\text{FE-HO}} = \widehat{\sigma}^2 \, \text{Trace}\left((A'A)^{-1}Q\right),$$

using an unbiased estimator of the variance.[12] Under homoskedastic, independent observations, $\widehat{\text{Bias}}_Q^{\text{FE-HO}}$ is unbiased for $\text{Bias}_Q$, so a bias-corrected estimator of $V_Q$ is

$$\widehat{V}_Q^{\text{FE-HO}} = \widehat{V}_Q^{\text{FE}} - \widehat{\sigma}^2 \, \text{Trace}\left((A'A)^{-1}Q\right).$$

In a recent contribution, Kline et al. (2020) propose a heteroskedastic generalization. Under the assumption that $\Omega(A)$ is diagonal, they estimate its diagonal elements using the jackknife, as

$$\widehat{\sigma}_{it}^2 = Y_{it}(Y_{it} - \widehat{\alpha}_i^{-(i,t)} - \widehat{\psi}_{j(i,t)}^{-(i,t)}),$$

where $\widehat{\alpha}_i^{-(i,t)}$ and $\widehat{\psi}_{j(i,t)}^{-(i,t)}$ are FE estimates on a subsample where observation $(i,t)$ has been taken out. In particular, computing the estimator requires focusing on a "leave-one-out" set that remains connected when any $(i,t)$ observation has been taken out. Hence, for this method the estimand changes relative to FE. Letting $\widehat{\Omega}(A)$ be the diagonal matrix with diagonal elements $\widehat{\sigma}_{it}^2$, the following estimator is unbiased under heteroskedastic, independent observations:

$$\widehat{V}_Q^{\text{FE-HE}} = \widehat{V}_Q^{\text{FE}} - \text{Trace}\left(A(A'A)^{-1}Q(A'A)^{-1}A'\widehat{\Omega}(A)\right).$$

Kline et al. (2020) provide conditions under which $\widehat{V}_Q^{\text{FE-HE}}$ is consistent, and they derive its limiting distribution.

When implementing these methods, we collapse observations at the spell level. This ensures the above estimators are unbiased in the presence of serial correlation within spell, under homoskedasticity and heteroskedasticity respectively.[13] However,

[12]The estimator $\widehat{\sigma}^2 = (NT - N - J)^{-1}Y'(I - A(A'A)^{-1}A')Y$ is unbiased for $\sigma^2$ when observations are independent and homoskedastic.

[13]As pointed out by Kline et al. (2020), when $T$=2, FE-HE estimators of firm effects and sorting components are also robust to the presence of serial correlation between spells. In the empirical analyses, we focus on 6 year panels and collapse earnings observations at the spell level (e.g., a stayer spell is collapsed into a single observation). See Appendix B for details. Without this collapsing approach, when using more than two periods of data, the FE-HE method is generally not robust to the presence of serial correlation within spells.

in practice, exact computation of $\widehat{V}_Q^{\text{FE-HO}}$ and $\widehat{V}_Q^{\text{FE-HE}}$ requires computing the trace of a large matrix inverse, which is prohibitive in most samples we use. For this reason, our empirical implementations rely on approximation methods (Gaure, 2014, Kline et al., 2020); see Section 6 and Appendix B.1.

**Random-effects.** Random-effects methods are popular in many panel data applications, yet they are rarely used in matched employer-employee settings. Here we introduce a correlated random-effects (CRE) estimator for variance components. Compared to fixed-effects estimators, the CRE estimator requires modeling the means and covariances of worker and firm effects. However, CRE depends on a smaller number of parameters. This parsimony is helpful for computational tractability, and to obtain more precise estimates.

Our starting point is the random-effects specification in Woodcock (2008). Woodcock postulates that the conditional distribution of worker and firm effects $\gamma = (\alpha', \psi')'$ given worker and firm indicators $A$ has mean $\mu(A)$ and variance $\Sigma(A)$.[14] In his specification, neither $\mu$ nor $\Sigma$ depend on $A$, and $\Sigma$ is diagonal. Woodcock uses this model as a prior for the worker and firm effects, and computes posterior estimates. We relax this specification in two ways. First, we allow $\Sigma(A)$ to be non-diagonal. Second, we allow $\mu(A)$ and $\Sigma(A)$ to depend on $A$. It is important to observe that assuming that $\alpha$ and $\psi$ are independent of $A$ would be restrictive. For example, this would require mobility across firms not to depend on worker or firm effects.

To build a flexible specification, we allow $\mu(A)$ and $\Sigma(A)$ to depend on $A$ by using the grouping strategy of Bonhomme et al. (2019). Specifically, we cluster firms into $K$ groups on the basis of their empirical earnings distributions. We use the k-means clustering algorithm for the grouping, and use $K = 10$ in our baseline specification. Given this grouping, we allow the means and variances of worker and firm effects to depend on the groups, but not on the worker and firm identities within these groups. Similarly, we allow the covariances in $\Sigma(A)$ to depend on the groups (or pairs of groups), while imposing some homogeneity assumptions so as to limit the number of parameters; see Appendix B.2 for a detailed description. The CRE model still has many fewer parameters than the AKM fixed-effects model.

[14]The model in Woodcock (2008) also accounts for covariates, which we abstract from in the presentation.

We estimate the CRE parameters by minimum distance based on mean restrictions and cross-worker covariance restrictions that are linear in parameters, so implementation is straightforward. We describe the moment restrictions and provide details on the estimation strategy in Appendix B.2. We report CRE estimates of the variance components,

$$\widehat{V}_Q^{\text{CRE}} = \widehat{\mu}(A)'Q\widehat{\mu}(A) + \text{Trace}(\widehat{\Sigma}(A)Q). \tag{7}$$

When the firm groups are defined in terms of observable categories such as industry or commuting zone, consistency of CRE follows from standard conditions for minimum distance. In addition, efficiency could be achieved using optimal weights. In our implementation, we tailor the groups to the data and construct them based on earnings using the k-means algorithm. In single-agent panel data models, Bonhomme et al. (Forthcoming) provide conditions for consistency of k-means clustering and estimators based on the estimated clusters under continuous heterogeneity. Consistency requires $K$ to tend to infinity with the sample size. We provide an analogous consistency argument for the AKM model in Appendix C. In the main analysis, we report results based on $K = 10$ groups. We document robustness with respect to this choice for a range of $K$.[15]

## 4.2 Empirical Illustration of Bias-correction

In Figure 2, we illustrate empirically the homoskedastic fixed-effects bias-correction method of Andrews et al. (2008) (FE-HO), and our correlated random-effects method based on the firm grouping of Bonhomme et al. (2019) (CRE), and compare them to FE in the Swedish data.[16] As described in Section 3, this figure considers the subsample of firms with at least 15 movers. Next, we remove movers randomly within firms before applying the FE, FE-HO, and CRE estimators to each random subsample, keeping the connected set of firms the same. We repeat this exercise

[15]In addition, in some specifications, we report posterior estimates in the spirit of empirical Bayes shrinkage. Interpreting our CRE model as a prior on the worker and firm effects, and under additional Gaussianity assumptions, we compute posterior estimates of the variance of firm effects. This provides a useful check, since under correct specification CRE and posterior estimates should be similar; see Appendix B.2 for the formula of the posterior estimator of $V_Q$.

[16]See Appendix Figure F5 for a similar analysis for the US.

for the leave-one-out set of firms, which allows us to also compare results to the heteroskedastic fixed-effects bias-correction method of Kline et al. (2020) (FE-HE).

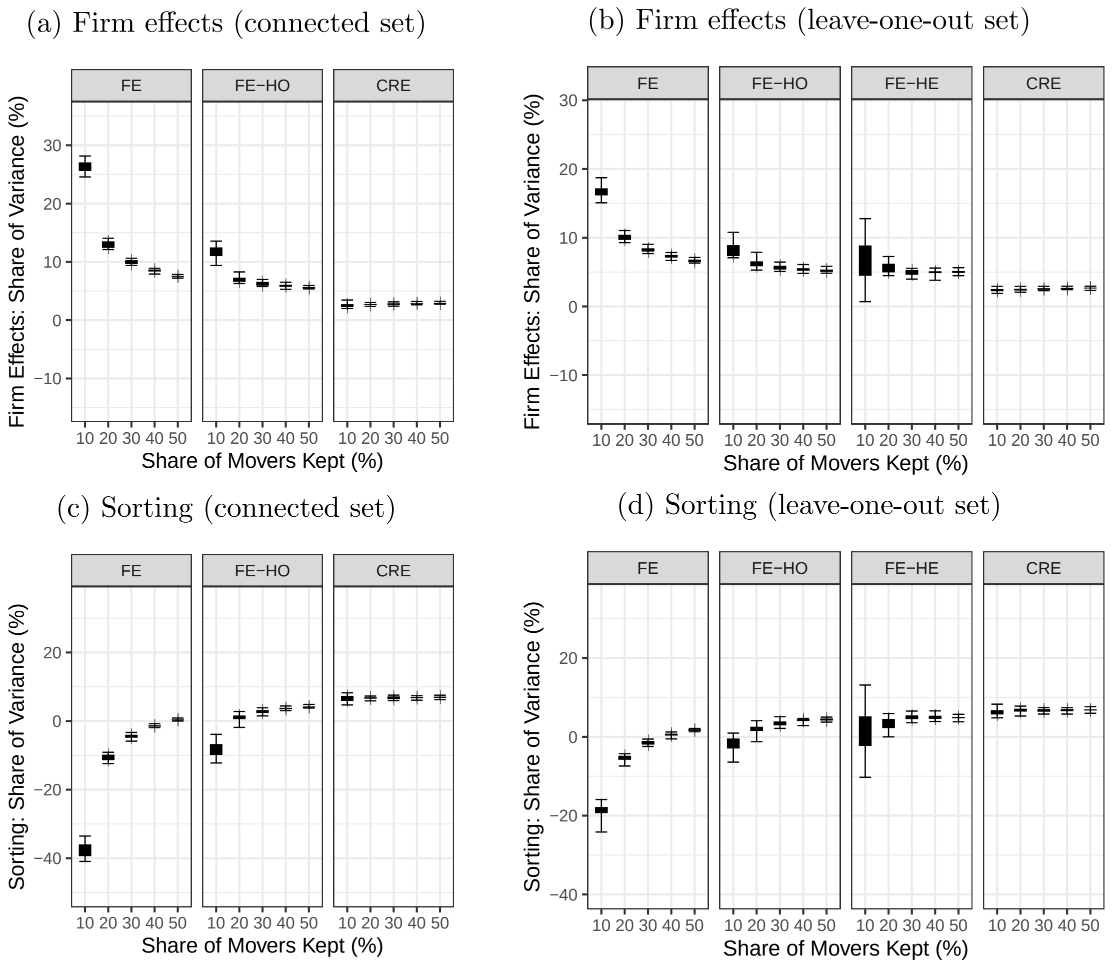
Consider again Subfigure 2a, but now focusing on the blue (CRE) and red (FE-HO) lines. The FE estimator and the bias-corrected estimators are similar when including all movers per firm, but become more dissimilar when there are fewer movers per firm. In contrast with FE, the two bias-corrected estimates remain nearly identical as the number of movers per firms declines, suggesting that the bias-corrected estimators are robust to the number of movers per firm. At the same time, the CRE estimates tend to be smaller than the FE-HO estimates. In Subfigure 2b we repeat this analysis for the leave-one-out set, which allows us to include the FE-HE bias-correction (in green). In this case also, all three bias-correction methods behave similarly, and in sharp contrast with FE, these estimators seem approximately insensitive to limited mobility bias.

Turning to Subfigure 2c, focusing on the blue and red lines, we see that CRE and FE-HO bias-corrected estimates of the contribution of worker sorting to earnings inequality are also quite similar to each other, and the estimates do not vary much with the sample. In particular, bias-corrected estimates are always positive while FE estimates in samples with few movers are negative. In Subfigure 2d we repeat this analysis for the leave-one-out set, finding that the three bias-corrected estimators, now including FE-HE, behave quite similarly, albeit with some quantitative differences.

In the next sections, we report results based on both the fixed- and random-effects methods for bias-correction. The rationale for using a variety of methods is that they rely on different modeling strategies. While FE-HO and FE-HE involve a very large number of worker and firm fixed-effects, CRE depends on a smaller number of parameters and therefore can be more precise. To illustrate this, Figure 3 presents the range (whiskers) and the interquantile range (solid bar) of the estimates from the random draws of Swedish data. Whereas Figure 2 presents the mean across random draws of the data, Figure 3 presents the variability across these random draws.

The findings from Figure 3 suggest that the CRE estimates of firm effects and worker sorting are less variable than those produced by the FE-HO and FE-HE estimates across the random draws of the data. For example, Subfigure 3a considers the variability in the estimates of the contribution of firm effects to wage inequality in

Figure 3: Evidence on Variability of the Estimators in Sweden




(a) Firm effects (connected set)

(b) Firm effects (leave-one-out set)

(c) Sorting (connected set)

(d) Sorting (leave-one-out set)

*Notes:* In this figure, we consider the subset of firms in Sweden with at least 15 movers. We randomly remove movers within each firm and re-estimate the variance of firm effects and covariance between firm and worker effects using the various estimators. For each estimator, we repeat this procedure twenty times, and report the overall range (whiskers) and interquartile range (solid bar) of estimates across these repetitions. The procedure allows us to keep the connected or leave-one-out set of firms the same and examine the variability in the estimators when there are fewer movers available in estimation.

the connected set, finding much lower variation for the CRE than the FE and FE-HO estimators when a small share of movers is kept.

# 5 Empirical Findings

We now present results on firm effects and sorting for Austria, Italy, Norway, Sweden, and the US. As described in Section 2, the sample selection and variable definitions are harmonized, to the extent possible, across countries. We compare firm effect and sorting estimates across bias-correction methods and samples. Given that some studies have used relatively short panels with no more than 3-years (e.g. Kline et al. 2020), while others have used longer panels with at least 6-years (e.g. Song et al. 2019), we present results for both 3-year and 6-year panels.

**Limited mobility bias leads to large upward bias in firm effects**

Figure 4 presents the main results for the connected set in the various countries. Subfigure 4a focuses on estimates of the share of earnings inequality due to firm effects for the 6-year panel.

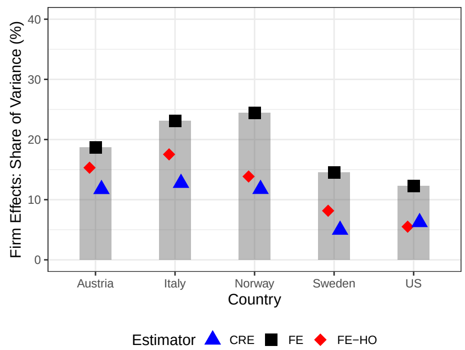
In the US, the fixed-effects (FE) estimator suggests that 12% of all earnings variation is due to firm effects.[17] When applying a bias-corrected estimator, this number falls to 5% when using the homoskedastic fixed-effects (FE-HO) approach, and 6% when using our correlated random-effects (CRE) approach. These estimates suggest that limited mobility bias accounts for at least half of the FE estimates of the contribution of firm effects to wage inequality.

For the European countries in Subfigure 4a, the FE estimator suggests 23-24% of earnings variance is due to firm effects in Italy and Norway whereas 15-18% is due to firm effects in Austria and Sweden. When using the FE-HO bias-correction, we find a range of reductions in the estimates from about one-fifth (Austria and Italy) to about one-half (Norway and Sweden) relative to FE. The bias-correction becomes stronger when using CRE, with estimates across countries in the 5-13% range, implying reductions in the estimates ranging from about one-half to about two-thirds relative to FE.

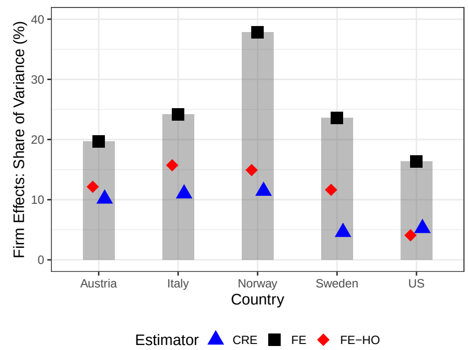
[17] Song et al. (2019) and Sorkin (2018) use the FE estimator to estimate firm effects in the US. Our finding that 12% of all earnings variation is due to firm effects falls between the estimates of Song et al. (2019) and Sorkin (2018) of 14% and 9%, respectively. In Appendix E, we explore the sources of these discrepancies, showing that they differ in part due to the choice of minimum earnings threshold used to define full-time equivalence and in part due to differences in minimum firm size thresholds.

Figure 4: Firm Effects and Sorting across Countries

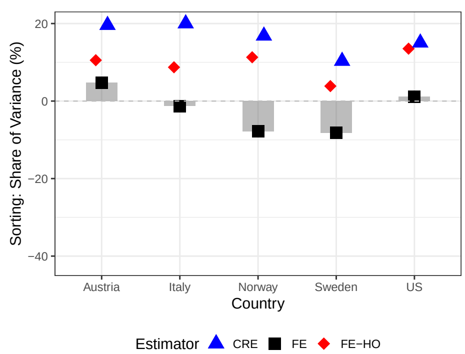
(a) Firm Effects (6-year panel)



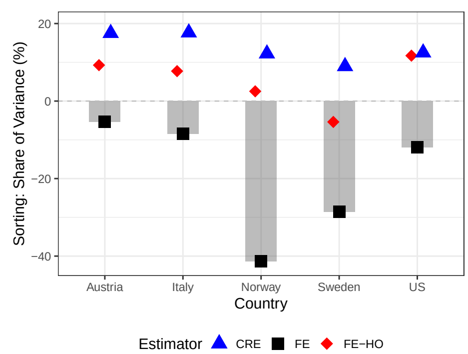

(b) Firm Effects (3-year panel)




(c) Sorting (6-year panel)




(d) Sorting (3-year panel)




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings or wage inequality of firm effects (Subfigures a and b) and the sorting of workers to firms (Subfigure c and d) in Austria, Italy, Norway, Sweden, and the US. We consider the connected set of firms within each country for 6-year panels (Subfigures a and c) and 3-year panels (Subfigures b and d).

Subfigure 4b repeats these estimates for the shorter 3-year panel. The FE estimator suggests an even greater role for firm effects when the panel is shorter. In the US, the FE estimator suggests that 16% of all earnings variation is due to firm effects, compared to 12% for the 6-year panel. The FE estimates are also larger in each of the European countries when considering a shorter panel, with firm effects explaining at least 20% of variance in each country with an upper estimate of about 38% (Norway). However, the CRE estimates remain in the 5-13% range for the US and each European country, suggesting the FE estimator is much more biased in shorter panels with fewer movers per firm.

In sum, we conclude there is substantial upward-bias in the FE estimator of firm effects in each country, FE is more biased in shorter panels, and the share of earnings variance due to firm effects is substantially smaller compared to what the FE estimator suggests. These conclusions hold true for both bias-correction methods.

**Limited mobility bias leads to large downward bias in sorting**

Subfigure 4c provides the main results on the contribution to inequality of the sorting of workers to firms in the various countries. When using FE, we find a negative estimate of the share of earnings variation due to sorting in all but the US and one European country. FE estimates range from -8% (Norway and Sweden) to 5% (Austria). However, when using either the FE-HO or CRE bias-correction, all of the sorting estimates become positive. In the US, the FE-HO estimator finds a sorting contribution of 13%, while the CRE estimate is 15%. In the European countries, FE-HO finds estimates of the sorting contribution ranging from 4% (Sweden) to 11% (Austria and Norway), while CRE finds estimates ranging from 10% (Sweden) to 20% (Austria and Italy).

Subfigure 4d repeats this analysis for the shorter 3-year panels. FE suggests a negative contribution of sorting in each country, while CRE finds nearly the same estimates as in the longer 6-year panel, reflecting that limited mobility bias is more severe in shorter panels. Comparing the contribution to inequality of firm effects (Subfigure 4a) to that of sorting (Subfigure 4c), the FE estimates suggest that firm effects explain a larger share of inequality than sorting. However, once one corrects for bias using the CRE estimator, it becomes evident that sorting is more important than firm effects.

When translating the estimates of sorting into correlations, it is important to observe that estimating the correlation between worker and firm effects requires estimating the variance of worker effects, and stronger assumptions would be needed to recover the variance of worker effects (for example, one could assume a particular dependence structure within and between job spells). However, as long as the covariance is positive, it is easy to compute the following lower bound on the correlation,

$$\text{Corr}(\alpha_i, \psi_{j(i,t)}) \geq \frac{\text{Cov}(\alpha_i, \psi_{j(i,t)})}{\sqrt{\text{Var}(\psi_{j(i,t)})}\sqrt{\text{Var}(Y_{it}) - \text{Var}(\psi_{j(i,t)}) - 2\,\text{Cov}(\alpha_i, \psi_{j(i,t)})}}. \tag{8}$$

Using this lower bound, the above results for the US translate into correlations between worker and firm effects of 0.32 when using FE-HO and 0.34 when using CRE. By contrast, FE suggests only a correlation of 0.02. In the European countries, the CRE estimator of sorting translates into a lower bound on the correlation between worker and firm effects (given by equation 8) ranging from 0.24 (Sweden) to 0.34 (Austria and Italy).

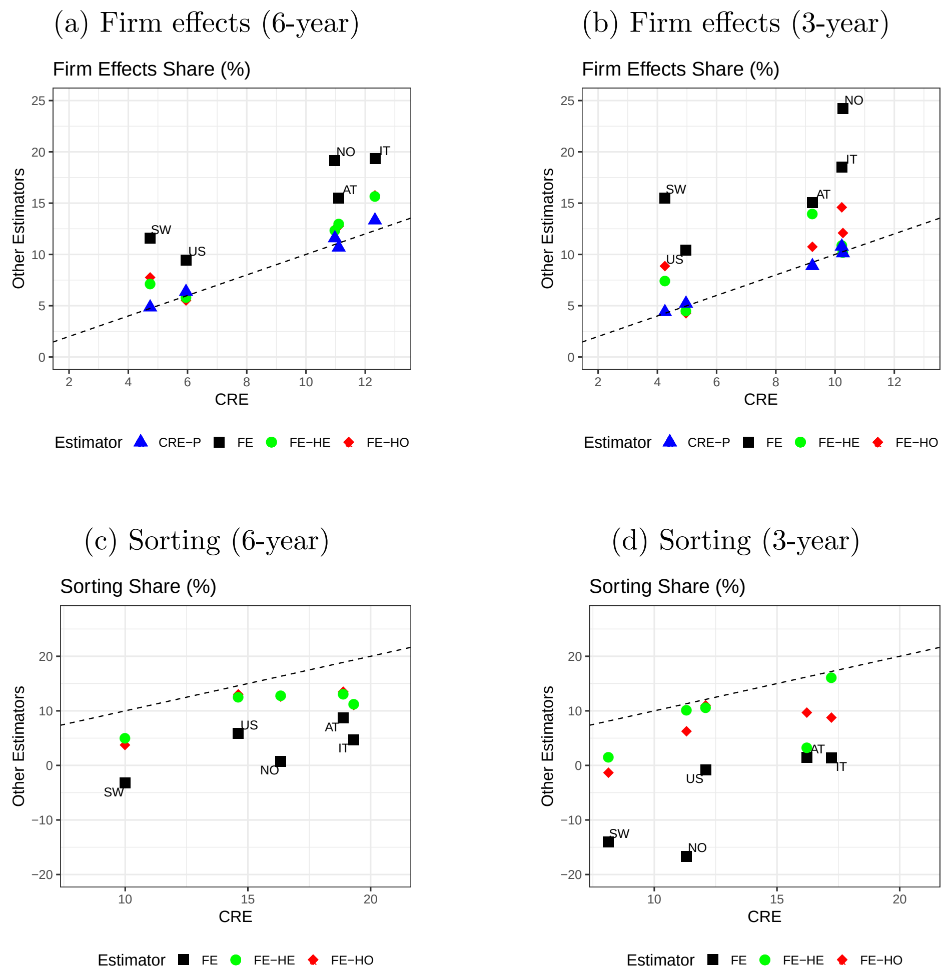
Overall, we conclude the FE estimator for sorting is downward-biased and typically of the wrong sign, the biases are more severe in shorter panels, and the bias-corrected share of earnings variance due to sorting tends to be substantial and of similar or larger magnitude compared to the share due to firm effects. These conclusions hold true for both fixed-effects and random-effects bias-correction methods.

**Comparison between connected set and leave-one-out set**

To apply the heteroskedastic fixed-effects (FE-HE) bias-correction method (and compare it to the other bias-correction methods), it is necessary to focus on the leave-one-out connected set of firms. In Table 1, we saw that most workers from the connected set are also included in the leave-one-out set. However, around half of all firms in the connected set are excluded from the leave-one-out set. A natural concern is that the leave-one-out set differs from the connected set in the composition of workers, moves, and firms. As shown in Table 1, larger firms are over-represented in the leave-one-out connected set.

In Figure 5, we consider the leave-one-out set. We plot the CRE estimator on the x-axis and various alternate estimators on the y-axis, so that the 45-degree line represents equality between CRE and the alternate estimators. The FE estimator is denoted by grey squares, FE-HE by green circles, and FE-HO by red diamonds. The blue triangles denote posterior CRE estimates that we discuss in Section 6. In Subfigure 5a, we provide estimates of the share of earnings variance due to firm effects for the longer 6-year panel. We see that the FE estimator is much higher than CRE

Figure 5: Leave-one-out Set: Various Countries




*Notes:* In this figure, we provide FE, FE-HO, FE-HE, and CRE estimates of the contribution to earnings or wage inequality of firm effects (Subfigures a and b) and the sorting of workers to firms (Subfigure c and d) in Austria, Italy, Norway, Sweden, and the US. We consider the leave-one-out set of firms within each country for 6-year panels (Subfigures a and c) and 3-year panels (Subfigures b and d). CRE estimates are displayed on the x-axis, and the dashed 45-degree line represents equality between CRE and the alternate estimators. The posterior CRE estimator (CRE-P) for firm effects is also displayed (Subfigures a and b).

in each country. By way of comparison, FE-HO and FE-HE line up well along the 45-degree line for some countries, while the estimators are somewhat larger than CRE in other countries. We repeat this analysis for the 3-year panel in Subfigure 5b, finding a similar pattern but the FE estimates are even further from the 45-degree line.

In Subfigures 5c and 5d, we provide estimates of the share of earnings variance due to the sorting of workers to firms using the 6-year and 3-year panels, respectively. We see that the FE estimates are far below the CRE estimates, while FE-HO and FE-HE produce estimates that are close to the CRE estimates across the various countries, albeit somewhat lower.[18]

# 6 Practical Considerations for Bias Correction

In our last set of results, we turn to some practical considerations for bias correction, including an assessment of situations in which limited mobility bias is likely to be a problem and an examination of some potential implementation issues. We briefly summarize the insights here, and refer to our working paper (Bonhomme et al., 2020) for further details.

## 6.1 When is limited mobility bias (un)likely to be an important problem?

It is difficult to know ex ante whether or not limited mobility bias is likely to be empirically important, as it depends on a large matrix of worker and firm indicators. Nevertheless, there are reasons to suspect that limited mobility bias is a more important problem in some settings than others.

**2-year panels** One might conjecture that limited mobility bias is more likely to be severe if one uses very short panels, since a longer time period helps to observe more workers moving across firms. However, a shorter time period may also have

[18] In Appendix E, we relate our sample and findings to those in Kline et al. (2020) on Italian data. In addition, in Appendix Figure F14 we show FE and bias-corrected estimates for the 20 smallest US states. The results show that bias-corrected estimates are very similar across methods, and that the FE estimates of firm effects (worker sorting) are severely upward (downward) biased as compared to their bias-corrected counterparts.

advantages, such as making the assumption of time-invariant worker effects more plausible. Furthermore, very short panels can be particularly useful to study the evolution of firm effects and sorting over time as in the study of Washington state by Lachowska et al. (Forthcoming).

In Figure 4, we compared 6-year panels to 3-year panels for each country, finding that the FE estimate was more biased in the 3-year panels while the CRE results were nearly identical. In Appendix Figure F6, we investigate further the performance of the estimators in short panels by splitting our baseline sample from the US during 2010-2015 into each two-year time interval and applying our estimators to these 5 short panels. We find that the FE estimator becomes much more biased, with the share of variance due to firm effects rising from 12% in the 6-year panel to more than 20% in the 2-year panels, and the share of variance due to sorting falling from 1% in the 6-year panel to below -20% in the 2-year panels. Reassuringly, the bias-corrected estimates do not materially change when shortening the panel.

**Small firms** One possible strategy to reduce limited mobility bias is to restrict firm size. Large firms tend to have more movers and, therefore, are better connected. For example, Song et al. (2019) and Bassier et al. (2021) restrict to firms with at least 20 workers, and Sorkin (2018) restricts to firms with at least 15 workers. In Appendix Figure F7 we explore this possibility in our US sample by restricting the sample to firms with at least 10, 20, 30, 40, or 50 workers. This corresponds to an increase in the number of movers per firm from about 5 (baseline) to about 45 (minimum 50 workers per firm).

As expected, we find that the bias in the FE estimates diminishes as the minimum firm size rises.[19] However, it is necessary to exclude a large share of workers and firms to reduce limited mobility bias. For the share of variance due to firm effects, there is little remaining bias when minimum firm size is 30. For the share of variance due to sorting, there is non-neglible bias even when minimum firm size is 50.

When interpreting results, it is important to observe that such restrictions change the population of study. Indeed, only 2 in 3 workers, 1 in 3 moves, and 1 in 20

[19]Note that, while biases tend to be smaller for larger firms in our US sample, there is no theoretical guarantee this will happen in other samples, since the structure of the bias depends on the network of workers and firms in complex ways (Jochmans and Weidner, 2019).

firms remain in the sample when the minimum firm size is 50 workers. Economic theory suggests that the distribution of firm effects in larger firms is likely to differ systemically from the distribution in smaller firms.[20] Thus, changing the population of study to minimize limited mobility bias introduces another form of bias, namely, sample selection bias. In Appendix D, we characterize analytically and numerically the bias introduced by approximating the variance of firm effects in the population using estimates for a selected subsample. We find that modest sample restrictions based on firm size can lead to substantial bias in the estimated variance of firm effects, even if there is no limited mobility bias.

**Changes over time** One situation in which one may be worried about limited mobility bias is when studying changes over time in the wage distribution. In Germany, Card et al. (2013) find that a rise in the variance of firm effects as well as increased sorting over time have contributed substantially to recent increases in wage inequality. In the US, Song et al. (2019) find that the contribution of the variance of firm effects to earnings inequality has declined over time while increased sorting over time has contributed substantially to earnings inequality. However, these studies rely on FE estimation and do not perform formal bias correction. One may be concerned that *changes* in the bias of the FE estimator over time explain the findings on changes in the role of sorting over time.

We now investigate changes over time in the contribution of firm effects and sorting to earnings inequality in the US. We compare our baseline estimates from the final years in our sample window, 2010-2015, to the estimates we obtain for 2001-2006. The results are presented in Appendix Figure F8. The main insight from this figure is that bias-correction is important for obtaining reliable estimates of the contribution of firm effects and sorting to earnings inequality in a given time period but not for capturing how their contribution to inequality changes across time periods. The reason is that limited mobility bias, while sizable, does not change materially over time in our US sample. This conclusion is consistent with the conjecture by Card et al. (2013) that limited mobility bias may be less important for studying inequality over time due to

[20]For example, with imperfect competition in the labor market, larger firms need to bid up wages to hire the additional workers, and, as a result, these firms may have larger firm effects on average (see for example Lamadon et al. 2022 and Kroft et al. 2021).

limited mobility bias being similar in different time periods.

## 6.2 Possible implementation issues for bias correction

**Mover definition** In the European countries, our data includes start and end dates of employment spells, so we know the year in which a move occurs. However, we do not observe start and end dates in the US. To harmonize the mover definition across countries, in the analysis above, we defined a change in primary employer across years as a move and measured earnings across all years during which the firm was the primary employer. As a check on the importance of this mover definition, we consider a stricter mover definition for the US in which a worker must be employed for at least 3 consecutive years at the first firm and at least 3 consecutive years at the second firm, only measuring earnings during intermediate years in these multi-year spells. Appendix Figure F9 provides a diagram to help visualize the difference in these mover definitions and the timing of earnings measurement.

Imposing the strict mover definition in the US sample substantially decreases the number of movers during our sample period. Only 1 in 60 moves satisfies this particular "3-year/3-year" structure of FTE employment spells during 2010-2015. Appendix Figure F3 compares the estimates obtained under the baseline and strict definitions of movers. The FE estimate of the contribution of firm effects to earnings variation rises from 12% to 17% (the bias-corrected estimates are both around 5%), and the FE estimate of the contribution of sorting to earnings variation decreases from about 1% to about -17% (the bias-corrected estimates are both around 14%). Yet, the CRE estimates are nearly identical under the two definitions, despite the substantial change in sample composition.

**Annual earnings, daily wages, and hourly wages** In many employer-employee data sets, one does not observe hourly wages but instead observes annual earnings or average earnings over an employment spell. When applying the FE estimation, one must then take a stand on the proper measure of wages or earnings. The data from Norway is an exception, as we have accurate measures of days and hours worked in this data set.

In Appendix Figure F10, we compare results on annual earnings, daily wages, and hourly wages for the same set of workers in the Norwegian data. We provide the comparison for the 6-year and 3-year panels. The FE estimate of the contribution of firm effects rises substantially when using a higher-frequency measure. In the 6-year (3-year) panel, it rises from about 19% (30%) for annual earnings to about 31% (48%) for hourly wages. The three bias-correction methods yield similar results across outcome measures. In the 6-year (3-year) panel, the CRE estimate of the contribution of firm effects rises from about 9% (8%) for annual earnings to about 13% (12%) for hourly wages. These estimates imply that FE is more biased when using higher-frequency outcome measures, and the bias-corrected estimate of the contribution of firm effects to inequality remains economically modest and somewhat greater for higher-frequency measures. A similar pattern is observed for the estimates of sorting, where FE suggests much stronger negative sorting when using hourly wages, but CRE finds substantial positive sorting with similar point estimates across outcome measures.

**FE-HO and FE-HE exact vs approximate estimators** Due to the large sample size in the US, we cannot compute the FE-HO and FE-HE estimators exactly, and the estimates are computed using an approximate method following Gaure (2014) and Kline et al. (2020). A natural worry is that the approximation may perform poorly. In order to investigate this possibility, we apply the estimators to 20 small US states where we can feasibly compute the exact and approximate solutions. In Appendix Figure F11 we plot exact versus approximate FE-HO estimators in the connected set in panel (a), and exact versus approximate FE-HE estimators in the leave-one-out set in panel (b). The results show that exact and approximate solutions are close to each other, suggesting that at least in these samples the numerical approximation works well.

**CRE number of clusters and posterior estimators** In our baseline CRE estimation, we cluster firms into 10 groups. One may worry that 10 groups is too restrictive. Appendix Figure F12 compares CRE estimates by number of groups in our US sample. We find that, as we increase the number of groups from 10 to 50, the estimates remain nearly identical for the earnings variation due to firm effects and sorting.

Above, we reported CRE estimates of variance components based on (7). We can also compute posterior estimates using the CRE specification as a Bayesian prior. Such estimates enjoy robustness properties when the CRE model is misspecified (Bonhomme and Weidner, Forthcoming). In Figures 5(a-b), we compare the posterior CRE (CRE-P) estimator to our other estimators for the variance of firm effects. We find that CRE-P is almost identical to CRE for both the 6-year and 3-year panels. As shown in Appendix Figure F14, the same holds true when we compare CRE and CRE-P separately for the 20 small US states. This is to be expected if the CRE model is correctly specified. Lastly, in Appendix Figure F13, we report posterior estimates for a random-effects specification that does not condition on firm groups. We estimate the firm effects variance to be less than half the CRE estimate. This suggests that accounting for the firm groups in the random-effects specification is important.

# 7 Broad Lessons for Empirical Work using AKM

Over the past two decades, a large body of work has used the AKM model and FE estimator to analyze earnings inequality in many developed countries. The results from these studies have been important, not only for quantifying the sources of earnings inequality, but also for how economists model the labor market. In this paper, we assessed the sensitivity of FE estimates to the incidental parameter problem that arises in the AKM model, often referred to as "limited mobility bias". Researchers have long been aware of the problem of limited mobility bias. Despite this awareness and the availability of bias-correction methods, relatively few studies correct for bias.

In our analyses, we use employer-employee data from the US and several European countries while taking advantage of both fixed-effects and random-effects methods for bias-correction. Our analyses deliver several important conclusions for empirical work using the AKM model. First, we show in simulations based on real data that limited mobility bias can be empirically important and existing methods for bias correction perform well even as mobility becomes very limited. Given their good performance, there is no need to resort to informal strategies based on sample restrictions (see for example Song et al. 2019, Sorkin 2018, and Bassier et al. 2021), which may introduce sample selection bias. One should instead implement theoretically justified bias-correction methods in empirical studies based on the AKM model.

Second, we find in Austria, Italy, Norway, Sweden, and the US that limited mobility bias is a major empirical issue for studies using FE to document firm effects and worker sorting. Once bias is accounted for, firm effects dispersion matters much less for earnings inequality and worker sorting becomes always positive and typically strong. Thus, we argue that it is important for empirical work using FE to perform bias correction of the estimates, especially when working with short panels.

Third, alternative methods for bias correction based on different assumptions and different cuts of the data (e.g. varying the number of periods or imposing a minimum firm size) tend to produce broadly similar results to one another. This is reassuring, as bias correction necessarily involves making restrictive assumptions about the model or limiting the set of firms under consideration. Furthermore, we find that the bias-correction methods are fairly robust to several possible specification and computational issues related to numerical approximation or discretization.

It is important to observe, however, that these conclusions rely on correctly specifying the model of earnings and the processes of worker and firm heterogeneity. There are several reasons why the AKM model may be misspecified, for example, both the assumptions that earnings are log-additive and that worker and firm heterogeneity are constant over time may be violated. To address these concerns, one possibility is to develop methods for bias-correction that are robust to misspecification. Another possibility is to enrich the model by, for example, incorporating worker-firm interactions and dynamic processes of worker and firm productivity. In this spirit, Bonhomme et al. (2019) estimate worker-firm interactions while allowing for state dependence and endogenous mobility in Sweden, while Lamadon et al. (2022) allow for worker-firm interactions and dynamic productivity processes of workers and firms in their study of the US labor market.

# References

**Abowd, John M., Robert H. Creecy, and Francis Kramarz.** 2002. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data.

Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

**Abowd, John M., Francis Kramarz, Paul Lengermann, and Sébastien 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, John M., Francis Kramarz, and David N. Margolis.** 1999. High wage workers and high wage firms. *Econometrica* 67 (2): 251–333.

**Abowd, John M., Kevin L. McKinney, and Ian M. Schmutte.** 2018. Modeling endogenous mobility in earnings determination. *Journal of Business & Economic Statistics* 1–14.

**Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser.** 2018. Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics* 10 (1): 149–89.

**Amiti, Mary, and David E. Weinstein.** 2018. How much do idiosyncratic bank shocks affect investment? evidence from matched bank-firm loan data. *Journal of Political Economy* 126 (2): 525–587.

**Andrews, Martyn J., Len Gill, Thorsten Schank, and Richard Upward.** 2008. High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (3): 673–697.

**Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward.** 2012. High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters* 117 (3): 824–827.

**Bagger, Jesper, and Rasmus Lentz.** 2019. An empirical model of wage dispersion with sorting. *The Review of Economic Studies* 86 (1): 153–190.

**Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu.** 2021. Monopsony in movers: The elasticity of labor supply to firm wage policies. *Journal of Human Resources* 0319–10111R1.

**Bonhomme, Stephane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler.** 2020. How much should we trust estimates of firm effects and worker sorting? Working Paper w27368, National Bureau of Economic Research.

**Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa.** 2019. A distributional framework for matched employer employee data. *Econometrica* 87 (3): 699–739.

**Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa.** Forthcoming. Discretizing unobserved heterogeneity. *Econometrica*.

**Bonhomme, Stéphane, and Martin Weidner.** Forthcoming. Posterior average effects. *Journal of Business & Economic Statistics*.

**Borovickova, Katarina, and Robert Shimer.** 2017. High wage workers work for high wage firms. Working Paper w24074, National Bureau of Economic Research.

**Card, David, Ana R. Cardoso, Joerg Heining, and Patrick Kline.** 2018. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics* 36 (S1): S13–S70.

**Card, David, Ana R. Cardoso, and Patrick Kline.** 2016. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics* 131 (2): 633–686.

**Card, David, Jörg Heining, and Patrick Kline.** 2013. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics* 128 (3): 967–1015.

**Drineas, Petros, Malik Magdon-Ismail, Michael W. Mahoney, and David P. Woodruff.** 2012. Fast approximation of matrix coherence and statistical leverage. *Journal of Machine Learning Research* 13 (Dec): 3475–3506.

**Eeckhout, Jan, and Philipp Kircher.** 2011. Identifying sorting—in theory. *The Review of Economic Studies* 78 (3): 872–906.

**Engbom, Niklas, and Christian Moser.** 2021. Earnings inequality and the minimum wage: Evidence from brazil. Working Paper w28831, National Bureau of Economic Research.

**Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams.** 2016. Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics* 131 (4): 1681–1726.

**Friedrich, Benjamin, Lisa Laun, Costas Meghir, and Luigi Pistaferri.** 2019. Earnings dynamics and firm-level shocks. Working Paper w25786, National Bureau of Economic Research.

**Gaure, Simen.** 2014. Correlation bias correction in two-way fixed-effects linear regression. *Stat* 3 (1): 379–390.

**Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card.** 2021. Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in brazil. *American Economic Review* 111 (10): 3418–57.

**Goldschmidt, Deborah, and Johannes F. Schmieder.** 2017. The rise of domestic outsourcing and the evolution of the german wage structure. *The Quarterly Journal of Economics* 132 (3): 1165–1217.

**Goux, Dominique, and Eric Maurin.** 1999. Persistence of interindustry wage differentials: A reexamination using matched worker-firm panel data. *Journal of Labor Economics* 17 (3): 492–533.

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

**Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii.** 2017. Identifying equilibrium models of labor market sorting. *Econometrica* 85 (1): 29–65.

**Hutchinson, Michael F.** 1990. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation* 19 (2): 433–450.

**Iranzo, Susana, Fabiano Schivardi, and Elisa Tosetti.** 2008. Skill dispersion and firm productivity: An analysis with employer-employee matched data. *Journal of Labor Economics* 26 (2): 247–285.

**Jochmans, Koen, and Martin Weidner.** 2019. Fixed-effect regressions on network data. *Econometrica* 87 (5): 1543–1560.

**Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten.** 2020. Leave-out estimation of variance components. *Econometrica* 88 (5): 1859–1898.

**Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler.** 2021. Imperfect competition and rents in labor and product markets: The case of the construction industry. Working Paper w27325, National Bureau of Economic Research.

**Lachowska, Marta, Alexandre Mas, Raffaele D. Saggio, and Stephen A. Woodbury.** Forthcoming. Do firm effects drift? evidence from washington administrative data. *Journal of Econometrics*.

**Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler.** 2022. Imperfect competition, compensating differentials, and rent sharing in the US labor market. *American Economic Review* 112 (1): 169–212.

**Lentz, Rasmus, Suphanit Piyapromdee, and Jean-Marc Robin.** 2017. On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data. *Unpublished Manuscript*.

**Lopes de Melo, Rafael.** 2018. Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy* 126 (1): 313–346.

**Mendes, Rute, Gerard J. van den Berg, and Maarten Lindeboom.** 2010. An empirical assessment of assortative matching in the labor market. *Labour Economics* 17 (6): 919–929.

**Mortensen, Dale.** 2003. *Wage Dispersion: Why are Similar Workers Paid Differently?*. MIT Press.

**Rockoff, Jonah E.** 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94 (2): 247–252.

**Shimer, Robert, and Lones Smith.** 2000. Assortative matching and search. *Econometrica* 68 (2): 343–369.

**Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter.** 2019. Firming up inequality. *The Quarterly Journal of Economics* 134 (1): 1–50.

**Sorkin, Isaac.** 2018. Ranking firms using revealed preference. *The Quarterly Journal of Economics* 133 (3): 1331–1393.

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

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

# ONLINE APPENDIX

## A Construction of event study data

In this section we describe the procedure we employ to go from an unbalanced panel of data over $T$ years to an event study format at the spell level, with earnings before and after a move for movers, and one earning per spell for stayers.

1. **Original data:** The raw data across countries contains the variables (worker ID, firm ID, year, log earnings, spell length information). A unique row of data is defined by a (worker ID, employer ID, year) triplet. The spell length information has a different level of precision in different countries; for example, in Sweden the data has monthly spell information, the US has no spell information, and Italy has the number of days worked.

2. **Select largest earning employer:** As is common in the literature, in the event that a worker receives earnings from multiple firms within a given year, we start by selecting the (employer ID) within each (worker ID, year) associated with the highest annual earnings.

3. **Construct log-earnings measures:** We construct an earnings measure as the reported yearly earnings divided by the reported spell length. In the US, this does not change the measure in any way since the reported spell length is the same for all spells. In other countries we get a measure of monthly-earnings or daily-earnings respectively.

4. **Residualize log-earnings measures:** We residualize log earnings using OLS regression on calendar year indicators and a third-order polynomial in age. Following Card et al. (2018), the age profile is restricted to be flat at age 40.

5. **Collapse years into spells:** We assign a unique (spell ID) to each time-consecutive sequence of (worker ID, employer ID) pairs. We collapse the data by taking the mean of the residualized log-earnings within each spell ID. The resulting data has variables (worker ID, employer ID, spell ID, begin year of spell, end year of spell, log-earnings). A unique row of data is defined by a

(worker ID, spell ID) pair, or alternatively, a unique (worker ID, begin year of spell) pair.

6. **Extract stayer spells and mover spell pairs:** We collect all workers with only one spell in a dataset of stayers with (worker ID, employer ID, log-earnings, begin year of spell, end year of spell). Next, we collect all pairs of consecutive spells into a movers event-study dataset where the variables are (worker ID, employer ID 1, employer ID 2, log-earnings 1, log-earnings 2). Employer ID 1 and employer ID 2 are the employer identifiers at two consecutive spells for a given worker. These employers ID's are different by construction. Log-earnings 1 is the mean log-earnings at employer ID 1, before the job change, and log-earnings 2 is the mean log-earnings at the second employer. Employer ID 1 and employer ID 2 are defined in chronological order based on spell begin year.

7. **Weighting used in variance decompositions:** We compute the variance decompositions weighted by person-event as constructed in the previous step. This means that each move is counted once and each stayer is counted once. Given that in most of our samples individuals rarely have more than one move, this is almost identical to weighting by individuals.

# B Estimation and computation

In what follows we describe the approach when working with an event-study data format. This means that each worker $i$ is either a stayer with one log-earnings (at the only employer), or he is a mover with at most two log-earnings (one at the employer before the move, and one at a different employer after the move). An advantage of this data structure, relative to other panel data formats, is that it does not require the researcher to make assumptions about serial correlation within job spells. Given this data structure, we next describe fixed-effects and random-effects methods in turn.

## B.1 Fixed-effects methods

**Estimation of FE-HO.** We follow Andrews et al. (2008). The first step in the estimation procedure is to extract the variance $\sigma^2$ of the residual. As noted in the text we use the following expression which provides an unbiased estimator under homoskedasticity:

$$\widehat{\sigma}^2 = (NT - N - J)^{-1} Y'(I - A(A'A)^{-1}A')Y.$$

Importantly, job stayers do not contribute to the estimation of this variance since they only have a single spell observation per individual. This is because the data are in event-study form. If this were not the case, one might worry about the fact that the formula assumes away serial correlation within job spells.

The next step is to compute the trace formula. When the design matrix $A$ is not too large, we directly invert the matrix and compute:

$$\widehat{\text{Bias}}_Q^{\text{FE-HO}} = \widehat{\sigma}^2 \,\text{Trace}\left((A'A)^{-1}Q\right).$$

**Estimation of FE-HO: Approximation.** When the design matrix is too large to be fully inverted we rely on trace approximation methods. To be precise, we use the Hutchinson stochastic trace estimator introduced in Hutchinson (1990), and proposed in the present context in Gaure (2014) and Kline et al. (2020), whereby the trace is approximated by

$$T_p = \frac{1}{p}\sum_{i=1}^{p} r_i'(A'A)^{-1}Qr_i,$$

where the $r_i$ are i.i.d. Rademacher random vectors. This procedure only requires solving $p$ linear systems, instead of trying to invert the matrix. It can be easily parallelized and in practice only a few draws seem to be sufficient to approximate the trace well.

**Estimation of FE-HE.** We refer to Kline et al. (2020) for a full description of their approach. Here we first outline the method while abstracting from computational feasibility concerns. The first step requires computing the leverage coefficients for each spell observation $(i,t)$. This is done by computing:

$$\widehat{\sigma}_{it}^2 = \frac{Y_{it}\left(Y_{it} - \widehat{\alpha}_i - \widehat{\psi}_{j(i,t)}\right)}{1 - P_{it,it}},$$

where

$$P_{it,it} = A_{it}\left(A'A\right)^{-1}A'_{it}.$$

This expression however does not recover the $\widehat{\sigma}_{it}^2$ for the stayers since they only have one spell-observation. In order to be able to compute the trace correction for the covariance in a sample that includes both stayers and movers, we then make an homogeneity assumption that $\sigma_{it}^2$ for stayers is equal to the average among movers at the same firm $j(i,t)$; that is,[21]

$$[\widehat{\sigma}_{it}^2]^{stayer} = \widehat{\mathbb{E}}_{i'}\widehat{\sigma}_{i't}^2 \text{ for movers } i' \text{ in } j(i,t)\ .$$

Next, we construct the trace correction expression

$$\text{Trace}\left[A\left(A'A\right)^{-1}Q\left(A'A\right)^{-1}A'\widehat{\Omega}(A)\right],$$

where $\widehat{\Omega}(A) = \text{diag}[\widehat{\sigma}_{it}^2]$. We compute this formula directly whenever inverting the matrix $A'A$ is computationally feasible.

**Estimation of FE-HE: Approximation.** There are two computational bottlenecks when computing the FE-HE estimator. One is the computation of the trace expression, for which we rely on the same Hutchinson trace estimator described above. This approximation performs very well in our experience.

The second computational bottleneck is the computation of $P_{it,it}$, which requires effectively inverting the $A'A$ matrix. This expression does not benefit from the same aggregation property that computing the trace does. Indeed, the $P_{it,it}$ enter the expression of $\widehat{\sigma}_{it}^2$ as inverses. This is a difficult computational problem that is actively

[21] As an alternative one could consider the following. First, compute the variance of firm effects in differences using movers and re-weight. Second, compute the covariance among movers using the leave-one-out procedure. Finally, compute the covariance for the stayers by using the covariance of their log-earnings with the estimated firm effects.

researched (Drineas et al., 2012). We decided to apply the procedure described in the computational appendix of Kline et al. (2020). Since we have $P_{it,it} = A_{it}\left(A'A\right)^{-1}A'_{it}$, if we could solve for $Z$ in

$$(A'A)Z = A',$$

we would simply get $P_{it,it} = A'_{it}Z_i$. We draw a set of $p$ random vectors $r_i$ as in the Hutchinson approach, and to combine them into a matrix $R_p$ with $p$ columns, and solve instead

$$(A'A)\tilde{Z} = (R_p A)',$$

and use $\tilde{P}_{it,it} = A'_{it}\tilde{Z}_i$. We thus use the following approximation:

$$\tilde{P}_{it,it} = A'_{it}(A'A)^{-1}A'R'_p,$$

which requires solving only $p$ linear system instead of inverting $A'A$ fully.

In practice, using a small $p$ tends to give some estimates $\tilde{P}_{it,it}$ that are not strictly less than 1. Since $(1-P_{it,it})$ enters in the denominator of $\widehat{\sigma}^2_{it}$, this can cause unbounded $\widehat{\sigma}^2_{it}$'s. We choose to increase $p$ until all $\tilde{P}_{it,it}$'s are $< 1$. This requires $p$ to be in the order of thousands.

## B.2 Correlated random-effects

**Overview.** The correlated random-effects (CRE) method consists of two steps. In the first step, group firms using a k-means clustering approach. In the second step, estimate the parameters of the grouped random-effects model by computing simple means, variances and covariances of log-earnings within and between groups. The first step relies on a standard Lloyd's algorithm for k-means. The second step involves mean and covariance restrictions that are linear in parameters. With a moderate number of parameters, estimation in the second step is thus straightforward. A fast implementation of the CRE estimator is provided at https://github.com/tlamadon/pytwoway.

**Estimating firm groups.** Let us first describe how we estimate the firm groups that we use to build the CRE specification. Accounting for the groups allows one to correlate worker and firm effects to mobility patterns, as we explain in the next paragraph. To estimate the firm grouping $\{k_j, j = 1, ..., J\}$, we follow Bonhomme et al. (2019) and cluster firms together based on earnings information. For example, using mean log-earnings one can estimate the partition by minimizing

$$\sum_{j=1}^{J} n_j(\overline{Y}_j - \mu(k_j))^2,$$

with respect to $\mu(1), ..., \mu(K)$ and $k_1, ..., k_J$, where $n_j$ is firm size, and $\overline{Y}_j$ is the mean log-earnings in firm $j$. In practice we add information beyond means by including the full earnings distribution function, evaluated at a grid of 20 points (20 percentiles of the overall earnings distribution). For computation we use Lloyds' algorithm for k-means, with 30 starting values. Consistency of k-means is not straightforward to establish in this context, due to the presence of within-$k$ firm heterogeneity. In single-agent panel data, Bonhomme et al. (Forthcoming) provide conditions for consistency and asymptotic normality of functions of the heterogeneity such as variance components as $K$ tends to infinity together with the sample size. In Appendix C, we provide a consistency argument in the present matched employer-employee setting.

**Overview of the model.** In CRE, we impose three orthogonality conditions on $\Sigma(A)$ and the covariance matrix $\Omega(A)$ of $\varepsilon_{it}$:

$$\text{Cov}(\alpha_i, \psi_j) = 0 \text{ for } (i, j) \in \mathcal{S}_1, \tag{B1}$$

$$\text{Cov}(\psi_j, \psi_{j'}) = 0 \text{ for } (j, j') \in \mathcal{S}_2, \tag{B2}$$

$$\text{Cov}(\varepsilon_{it}, \varepsilon_{i't'}) = 0 \text{ for } t\,,\, t',\, i \neq i', \tag{B3}$$

where all covariances are conditional on $A$ but we omit the dependence in the notation. Here $\mathcal{S}_1$ contains worker-firm pairs $(i, j)$ such that $i$ never works in $j$ at any point in the sample, and $\mathcal{S}_2$ contains firm pairs $(j, j')$ where $j \neq j'$.

Equations (B1) and (B2) are conditions about the covariance structure of worker and firm effects. Such conditions are not needed in fixed-effects approaches. Allowing the mean vector $\mu(A)$ and the variance matrix $\Sigma(A)$ to depend on worker and firm indicators $A$ will be helpful to relax these conditions by restricting the sets $\mathcal{S}_1$ and $\mathcal{S}_2$. Indeed, assuming that (B2) holds for all firm pairs may be empirically strong, if for example firms $j$ and $j'$ that are close to each other in economic distance have correlated effects $\psi_j$ and $\psi_{j'}$ because they share the same suppliers. In our implementation, we group firms and we only assume that $\psi_j$ and $\psi_{j'}$ are uncorrelated conditional on $j$ and $j'$ belonging to different firm groups.[22] Likewise, we only assume that $\alpha_i$ and $\psi_j$ are uncorrelated in (B1) when $i$ never visits the group of firm $j$. In turn, (B3) is an assumption on the covariance structure of $\varepsilon_{it}$. Note that this condition does not restrict the covariance matrix $\Omega(A)$ beyond cross-worker covariances.

Based on (B1)-(B2)-(B3), if one is willing to assume in addition that $\alpha_i$, $\psi_j$, and $\varepsilon_{it}$ are independent of $A$, one can build a simple CRE specification that depends on only three parameters: the variance of firm effects and the covariance between worker and firm effects, which are our parameters of interest, and the covariance between the worker effects of two workers who are employed in the same firm at some point in time. Hence this model is very parsimonious. Moreover, the parameters can be recovered from cross-worker covariance restrictions.

As an example, consider two workers $i$ and $i'$ who work in the same firm in period $t$. Both $i$ and $i'$ move between $t$ and $t'$, and $i'$ (respectively, $i$) moves to a firm where $i$ (resp., $i'$) never works. In this case the variance of firm effects can be recovered from

$$\begin{aligned} \text{Cov}(Y_{it'} - Y_{it}, Y_{i't'} - Y_{i't}) &= \text{Cov}(\psi_{j(i,t')} - \psi_{j(i,t)} + \varepsilon_{it'} - \varepsilon_{it}, \\ &\qquad \psi_{j(i',t')} - \psi_{j(i',t)} + \varepsilon_{i't'} - \varepsilon_{i't}) \\ &= \text{Cov}(\psi_{j(i,t')} - \psi_{j(i,t)}, \psi_{j(i',t')} - \psi_{j(i',t)}) \\ &= \text{Cov}(\psi_{j(i,t)}, \psi_{j(i',t)}) \\ &= \text{Var}(\psi_{j(i,t)}), \end{aligned} \tag{B4}$$

and the covariance between worker and firm effects can be recovered from

$$\begin{aligned} \text{Cov}(Y_{it'} - Y_{it}, Y_{i't'}) &= \text{Cov}(\psi_{j(i,t')} - \psi_{j(i,t)} + \varepsilon_{it'} - \varepsilon_{it}, \alpha_{i'} + \psi_{j(i',t')} + \varepsilon_{i't'}) \\ &= \text{Cov}(\psi_{j(i,t')} - \psi_{j(i,t)}, \alpha_{i'} + \psi_{j(i',t')}) \\ &= \text{Cov}(\psi_{j(i,t')} - \psi_{j(i,t)}, \alpha_{i'}) \\ &= -\text{Cov}(\psi_{j(i',t)}, \alpha_{i'}). \end{aligned} \tag{B5}$$

[22] A related approach would be to only consider firms $j$ and $j'$ in $\mathcal{S}_2$ that do not directly share a worker (i.e., a mover), although they might share workers indirectly through other firms $j''$.

To derive both (B4) and (B5) we have used the model in the first line, (2) and (B3) in the second line, and (B2) in the third line. In the last line, we have used that $j(i,t) = j(i',t)$ to derive (B4), and we have used (B1) to derive (B5). In addition, this simple CRE model implies a number of overidentifying restrictions. Covariance restrictions such as (B4) and (B5) are the basis of our strategy to estimate the CRE model.

**Specification details.** Specifying the random-effects model consists in listing the restrictions that we impose on the vector $\mu(A)$ and the square matrices $\Sigma(A)$ and $\Omega(A)$. $\Omega(A)$ captures the error structure of the residuals across observations and has a number of rows equal to the number of observations. $\mu(A)$ and $\Sigma(A)$ describe the mean and variance of $\gamma$, and have respective length and number of rows equal to the number of workers plus the number of firms.

To be exhaustive, we need to specify how each entry in these matrices and vectors depends on $A$. To do so, we note that the $\gamma$ vector contains three distinct types of elements: workers with only one employer, workers with multiple employers (i.e., movers), and firms. We describe the specification of $\mu(A)$ and $\Sigma(A)$ by listing the elements of $\mu(A)$ and $\Sigma(A)$ for each of these three types of entries. Throughout, we assume the data are in event study format, and hence movers have exactly two employers. We also make use of a firm grouping structure, where $k_j$ denotes the group of firm $j$ and we write $k_{it} = k_{j(i,t)}$ to simplify the notation.

We assume that $\mu(A)$ does not depend on worker and firm identities beyond firm groups. We denote

$$\begin{aligned}
&\mathbb{E}[\alpha_i \,|\, A] = \mathbb{E}[\alpha_i \,|\, k_{i1}] = \mu_\alpha(k_{i1}) \text{ for stayers,} \\
&\mathbb{E}[\alpha_i \,|\, A] = \mathbb{E}[\alpha_i \,|\, k_{i1}, k_{i2}] = \mu_\alpha(k_{i1}, k_{i2}) \text{ for movers,} \\
&\mathbb{E}[\psi_j \,|\, A] = E[\psi_j \,|\, k_j] = \mu_\psi(k_j).
\end{aligned}$$

The matrix $\Sigma(A)$ consists of variances and covariances of worker effects and firm effects. We assume that $\Sigma(A)$ does not depend on worker and firm identities beyond firm groups. We denote, for any firm $j$,

$$\mathrm{Var}[\psi_j \,|\, A] = \mathrm{Var}[\psi_j \,|\, k_j] = \Sigma_{\psi\psi}(k_j).$$

For the off-diagonal terms, we assume that $\text{Cov}[\psi_j, \psi_{j'}|k_j, k_{j'}] = 0$ for $k_j \neq k_{j'}$ and leave the covariance within unrestricted. In estimation we do not estimate within-group covariances. It is important to also note that this does not restrict the covariance at the group level, since the $\mu_\psi(k)$ are unrestricted. Next, for any firm $j$ and any two movers $i$ and $i'$ we denote:

$$\begin{aligned}
\text{Cov}[\psi_j, \alpha_i \,|\, A] &= \text{Cov}[\psi_j, \alpha_i \,|\, j, j(i,1), j(i,2)] \\
&= \mathbf{1}\big[j(i,1){=}j \text{ or } j(i,2){=}j\big]\,\Sigma^m_{\alpha\psi}(k_j), \\
\text{Cov}[\alpha_i, \alpha_{i'} \,|\, A] &= \text{Cov}[\alpha_i, \alpha_{i'}|j(i,1), j(i,2), j(i',1), j(i',2)] \\
&= \mathbf{1}\big[j(i,1){=}j(i',1)\big]\,\Sigma^m_{\alpha\alpha'}(k_{j(i,1)}) + \mathbf{1}\big[j(i,2){=}j(i',2)\big]\,\Sigma^m_{\alpha\alpha'}(k_{j(i,2)}) \\
&\quad + \mathbf{1}\big[j(i,2){=}j(i',1)\big]\,\Sigma^m_{\alpha\alpha'}(k_{j(i,2)}) + \mathbf{1}\big[j(i,1){=}j(i',2)\big]\,\Sigma^m_{\alpha\alpha'}(k_{j(i,1)}).
\end{aligned}$$

For any firm $j$ and any two stayers $i$ and $i'$ we denote

$$\begin{aligned}
\text{Cov}[\psi_j, \alpha_i \,|\, A] &= \text{Cov}[\psi_j, \alpha_i \,|\, j, j(i,1)] = \mathbf{1}\big[j(i,1){=}j\big]\,\Sigma^s_{\alpha\psi}(k_j), \\
\text{Cov}[\alpha_i, \alpha_{i'} \,|\, A] &= \text{Cov}[\alpha_i, \alpha_{i'} \,|\, j(i,1), j(i',1)] = \mathbf{1}\big[j(i,1){=}j(i',1)\big]\,\Sigma^s_{\alpha\alpha'}(k_{j(i,1)}).
\end{aligned}$$

For any given stayer $i$ and any given mover $i'$ we denote:

$$\begin{aligned}
\text{Cov}[\alpha_i, \alpha_{i'} \,|\, A] &= \text{Cov}[\alpha_i, \alpha_{i'} \,|\, j(i,1), j(i',1), j(i',2)] \\
&= \mathbf{1}\big[j(i,1){=}j(i',1)\big]\,\Sigma^{sm}_{\alpha\alpha'}(k_{j(i,1)}) + \mathbf{1}\big[j(i,1){=}j(i',2)\big]\,\Sigma^{sm}_{\alpha\alpha'}(k_{j(i,1)}).
\end{aligned}$$

Finally, we let the diagonal along workers unspecified since our focus is on the variance of firm effects and the covariance between worker and firm effects.[23]

As a reminder, the approach in Woodcock (2008) would set $\mu_\alpha(k){=}\mu_\alpha$, $\mu_\psi(k){=}\mu_\psi$, and $\Sigma_{\psi\psi}(k){=}\Sigma_{\psi\psi}$, as well as $\Sigma^m_{\alpha\psi}(k){=}\Sigma^s_{\alpha\psi}(k){=}\Sigma^m_{\alpha\alpha'}(k){=}\Sigma^s_{\alpha\alpha'}(k){=}\Sigma^{sm}_{\alpha\alpha'}(k){=}0$. Based on this specification, Woodcock focused on posterior estimates.

**Estimation.** Here we describe how we estimate the quantities that we use to reconstruct our two main parameters of interest (that is, the variance of firm effects and the covariance), as presented in equation (7). This involves the vector $\mu(A)$ and

[23] A natural specification would be to allow for the variance of the worker effects of stayers to be group-specific and for the variance of the worker effects of movers to depend on the group pairs.

a subset of the elements in $\Sigma(A)$.

First we estimate all elements in $\mu(A)$ as

$$\min_{\mu_\alpha(k,k'),\mu_\alpha(k),\mu_\psi(k)} \sum_{i:\,stayer} \Big(Y_{i1} - \mu_\psi(k_{i1}) - \mu_\alpha(k_{i1})\Big)^2 + \sum_{i:\,mover} \Big(Y_{i1} - \mu_\psi(k_{i1}) - \mu_\alpha(k_{i1}, k_{i2})\Big)^2 + \sum_{i:\,mover} \Big(Y_{i2} - \mu_\psi(k_{i2}) - \mu_\alpha(k_{i1}, k_{i2})\Big)^2.$$

Next, it turns out that the elements in $\Sigma(A)$ enter equation (7) only through the following group aggregates. Specifically we define for $(t, t', p) \in \{1, 2\}^3$ and compute:

$$C^m_{tt'}(p) = \widehat{\mathbb{E}}_{(i,i')\in S^m_p} \left[ \Big(Y_{it} - \mu_\alpha(k_{i1}, k_{i2}) - \mu_\psi(k_{it}))\Big) \Big(Y_{i't'} - \mu_\alpha(k_{i'1}, k_{i'2}) - \mu_\psi(k_{i't'}))\Big) \right],$$

where the set $S^{\mathrm{m}}_p$ of pairs of workers consists of movers leaving the same firm and moving to a different firm group, or alternatively moving to the same firm and coming from two different firm groups; that is,

$$S^m_p = \{(i, i' {\neq} i) \text{ movers, s.t. } j(i,p){=}j(i',p), k_{i,-p}{\neq}k_{i',-p}, k_{i,-p}{\neq}k_{i,p}, k_{i',-p}{\neq}k_{i',p}\}.$$

Similarly, we define for $(t', p) \in \{1, 2\}^2$ and compute:

$$C^s_{t'}(p) \;=\; \widehat{\mathbb{E}}_{(i,i')\in S^s_p} \left[ \Big(Y_{it} - \mu_\alpha(k_{i1}) - \mu_\psi(k_{i1}))\Big) \Big(Y_{i't'} - \mu_\alpha(k_{i'1}, k_{i'2}) - \mu_\psi(k_{i't'}))\Big) \right],$$

where

$$S^s_p = \{(i, i' {\neq} i), i \text{ stayer}, i' \text{ mover, s.t. } j(i,1){=}j(i',p), k_{i',-p}{\neq}k_{i1}\}.$$

To see the mapping between the sufficient elements of $\Sigma(A)$ in equation (7) and the previously defined group aggregates, note that:

$$\begin{aligned}
C^m_{22}(1) = C^m_{11}(2) &= \widehat{\mathbb{E}}_k\big[\Sigma^m_{\alpha\alpha'}(k)\big],\\
C^m_{12}(1) = C^m_{12}(2) &= \widehat{\mathbb{E}}_k\big[\Sigma^m_{\alpha\alpha'}(k) + \Sigma^m_{\alpha\psi}(k)\big],\\
C^m_{11}(1) = C^m_{22}(2) &= \widehat{\mathbb{E}}_k\big[\Sigma_{\psi\psi}(k) + \Sigma^m_{\alpha\alpha'}(k) + 2\Sigma^m_{\alpha\psi}(k)\big],
\end{aligned}$$

where $\widehat{\mathbb{E}}_k$ denote means, weighted by group sizes. In turn, the covariances based on combinations of stayers and movers give:

$$C_2^s(1) = C_1^s(2) = \widehat{\mathbb{E}}_k\big[\Sigma_{\alpha\alpha'}^{sm}(k) + \Sigma_{\alpha\psi}^{m}(k)\big],$$
$$C_1^s(1) = C_2^s(2) = \widehat{\mathbb{E}}_k\big[\Sigma_{\psi\psi}(k) + \Sigma_{\alpha\alpha'}^{sm}(k) + \Sigma_{\alpha\psi}^{s}(k) + \Sigma_{\alpha\psi}^{m}(k)\big].$$

Lastly, given the estimated $\mu$'s and $C$'s we construct the variance components appearing in equation (7).

**Posterior estimator.** Under an additional joint normality assumption of $\gamma$ and $\varepsilon$ given $A$, a posterior estimator $\widehat{V}_Q^{\mathrm{P}}$ of $V_Q$ is given by the posterior mean of $\gamma' Q \gamma$ in the Gaussian model; that is:

$$(\widehat{\Sigma}(A)^{-1}\widehat{\mu}(A) + A'\widehat{\Omega}(A)^{-1}Y)'\widehat{B}(A)^{-1}Q\widehat{B}(A)^{-1}(\widehat{\Sigma}(A)^{-1}\widehat{\mu}(A) + A'\widehat{\Omega}(A)^{-1}Y) + \mathrm{Trace}(\widehat{B}(A)^{-1}Q),$$

where $\widehat{B}(A) = \widehat{\Sigma}(A)^{-1} + A'\widehat{\Omega}(A)^{-1}A$. Relative to the main CRE estimator, we need all the elements of $\widehat{\Sigma}(A)$, and hence specify those by imposing additional zeros and modeling the entire diagonal. There are two computational challenges. First, $\widehat{\Sigma}(A)$ is a non-sparse matrix since we model covariances between worker effects and firm effects. Second, implementation requires computing the inverse of the matrix in the trace expression. This second challenge is as for the FE-HO estimator. In the paper we focus on the computation of the posterior estimator for the variance of firm effects. This only involves the part of $\widehat{\Sigma}(A)$ between firms, which is diagonal. We approximate the trace using the Hutchinson approach, as we do for FE-HO.

# C Consistency of grouped fixed-effects and correlated random-effects in the AKM model

Consider model (1) without covariates, with $T = 2$ periods:

$$Y_{it} = \alpha_i + \psi_{j(i,t)} + \varepsilon_{it}, \quad t \in \{1, 2\}. \tag{C6}$$

Let $\eta_j = (\psi_j, \xi_j)'$ denote a $d$-dimensional vector of firm heterogeneity. In period 1, $\alpha_i$ are drawn in firm $j(i,1)$ from a distribution that depends on $\eta_{j(i,1)}$. This corresponds to the setup in Bonhomme et al. (2019), except that here $\eta_j$ is continuous and the model is additive in worker and firm effects.

We consider a grouped fixed-effects (GFE) estimator where we cluster firms according to a moment of log-wages in the firm (e.g., a discretized estimate of the log-wage cdf), using $K$ groups. We study the consistency of the GFE estimator of the firm effects $\psi$, relative to the average squared norm.

Let $J$ denote the number of firms, $n$ denote the number of job movers in the sample, and $m$ denote the minimum number of observations per firm (i.e., minimum firm size) in the first period. Let $G$ denote the $J \times K$ matrix of zeros and ones, which maps group parameters to firms, where the group structure is the one estimated using k-means clustering.

By Bonhomme et al. (Forthcoming), as the number of groups $K$ tends to infinity with the minimum firm size $m$, we have, for some constant $A$,

$$\|GA - \eta'\|^2/J = O_p\left(m^{-1}\right) + O_p\left(K^{-\frac{2}{d}}\right),$$

where $\|\cdot\|$ denotes the Euclidean norm. Notice the rate of convergence depends on the dimension $d$ of $\eta_j$. Letting $a$ be the first column of $A$, we thus have

$$\|Ga - \psi\|^2/J = O_p\left(m^{-1}\right) + O_p\left(K^{-\frac{2}{d}}\right). \tag{C7}$$

Next, let us write model (C6) in first differences; that is, stacking all observations in column vectors,

$$\Delta Y = B\psi + U,$$

where $\Delta Y_{it} = Y_{i,t+1} - Y_{it}$. We make the following assumptions, where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of $B'B/n$, respectively.

<u>A1.</u> $\frac{\|B'U\|}{n\sqrt{J}\lambda_{\min}} = o_p(1)$.

<u>A2.</u> $\frac{\lambda_{\max}}{\lambda_{\min}} \times \max\left\{m^{-1}, K^{-\frac{2}{d}}\right\} = o_p(1)$.

For A1 to hold, it is sufficient that $\frac{\mathbb{E}(U'B'BU)}{n^2 J\lambda_{\min}^2} = o(1)$. As an example, if $\mathbb{E}(U) = 0$

and $\mathbb{E}(UU') = \sigma^2 I$, then it suffices that $\frac{\sigma^2 \operatorname{Trace}\left(\frac{B'B}{n}\right)}{nJ\lambda_{\min}^2} = o(1)$. A sufficient condition for this is $\sigma^2 \frac{\lambda_{\max}}{n\lambda_{\min}^2} = o(1)$. This allows $\lambda_{\min}$ to tend to zero, and $\lambda_{\max}$ to tend to infinity, albeit at sufficiently slow rates.

For A2 to hold, it is sufficient that $\frac{\lambda_{\max}}{\lambda_{\min}}$ does not tend to infinity faster than the minimum firm size, and that $K$ tends to infinity sufficiently quickly relative to it. The required rate on $K$ increases with the dimension $d$.

When $B$ represents the first differences of worker-firm employment relationships, $\lambda_{\min}$ is a measure of the connectedness of the worker-firm graph. Jochmans and Weidner (2019) show how measures of graph connectedness influence firms-specific least squares estimates. Moreover, $\lambda_{\max} \leq \operatorname{Trace}(B'B/n) = 1$.

Let us denote the least squares (AKM) estimator as

$$\widehat{\psi} = (B'B)^{-1} B' \Delta Y.$$

In addition, let us denote the GFE estimator as

$$\widetilde{\psi} = G(G'B'BG)^{-1} G'B' \Delta Y.$$

**Proposition C1.**

*If A1 holds, then* $\|\widehat{\psi} - \psi\|^2 / J = o_p(1)$.

*If A1 and A2 hold, then* $\|\widetilde{\psi} - \psi\|^2 / J = o_p(1)$.

By Proposition C1, the GFE estimate of a bounded quadratic form $V_Q = \psi' Q \psi$ is consistent; that is,

$$\widetilde{\psi}' Q \widetilde{\psi} = \psi' Q \psi + o_p(1).$$

In addition, writing model (C6) in vector form, we have

$$Y = A_\alpha \alpha + A_\psi \psi + \varepsilon,$$

and the GFE estimator of $V^R \equiv \alpha' R \psi$ is consistent as well; that is,

$$\left(A_\alpha^\dagger (Y - A_\psi \widetilde{\psi})\right)' R \widetilde{\psi} = \alpha' R \psi + o_p(1).$$

This shows that GFE estimators of the variance of firm effects and the covariance between firm and worker effects are consistent.

The CRE estimators of these variance components will then also be consistent under A1 and A2, since the within-group variance tends to zero as the number of groups tends to infinity. Note that this asymptotic result holds as both $K$ and the minimum firm size tend to infinity. In finite samples, and for fixed $K$, accounting for the within-group variance of firm effects may have a non-negligible effect on the estimates, as illustrated by our empirical findings.

In addition, although under the current assumptions the AKM, GFE and CRE estimators are all consistent, our findings also suggest that, in finite samples, the GFE and CRE estimators of firm effects may be more precise than AKM, resulting in variance components that are less biased.

*Proof.* We have, using A1,

$$\|\widehat{\psi}-\psi\|/\sqrt{J}=\left\|\left(\frac{B'B}{n}\right)^{-1}\frac{B'U}{n}\right\|/\sqrt{J}\leq\lambda_{\min}^{-1}\|B'U\|/(n\sqrt{J})=o_p(1).$$

This shows the first claim.

To show the second claim, let $\widetilde{\mu}=(G'B'BG)^{-1}G'B'\Delta Y$. We have, by the least squares property,

$$\|\Delta Y-B\widetilde{\psi}\|^2/n=\|\Delta Y-BG\widetilde{\mu}\|^2/n\leq\|\Delta Y-BGa\|^2/n.$$

Equivalently, we have

$$\begin{aligned}&\|B\psi-B\widetilde{\psi}\|^2/n+\|U\|^2/n+2U'(B\psi-B\widetilde{\psi})/n\\&\leq\|B\psi-BGa\|^2/n+\|U\|^2/n+2U'(B\psi-BGa)/n.\end{aligned}$$

That is,

$$\|B\psi-B\widetilde{\psi}\|^2/n+2U'(B\psi-B\widetilde{\psi})/n\leq\|B\psi-BGa\|^2/n+2U'(B\psi-BGa)/n.$$

Hence, using the Cauchy Schwartz inequality,

$$\lambda_{\min}\|\psi-\widetilde{\psi}\|^2\leq 2\|B'U\|\|\psi-\widetilde{\psi}\|/n+\lambda_{\max}\|\psi-Ga\|^2+2\|B'U\|\|\psi-Ga\|/n.$$

It follows that

$$\|\psi - \widetilde{\psi}\|^2/J \leq 2\|B'U\|/(n\sqrt{J}\lambda_{\min})\|\psi - \widetilde{\psi}\|/\sqrt{J}$$
$$+ (\lambda_{\max}/\lambda_{\min})\|\psi - Ga\|^2/J + 2\|B'U\|/(n\sqrt{J}\lambda_{\min})\|\psi - Ga\|/\sqrt{J}.$$

Using A1, A2, and (C7), we thus have

$$\|\psi - \widetilde{\psi}\|^2/J \leq o_p(1)\|\psi - \widetilde{\psi}\|/\sqrt{J} + o_p(1).$$

It follows that

$$\|\psi - \widetilde{\psi}\|^2/J = o_p(1).$$

∎

# D Bias due to Estimating the Variance of Firm Effects on a Selected Subsample

When implementing FE estimation, a number of recent studies restrict the population of interest to a subset of firms for which firm effects may be more easily recovered, such as large firms (see for example Song et al. 2019, Sorkin 2018, and Bassier et al. 2021). Similarly, the FE-HE bias-correction method restricts the population to the leave-one-out subsample of strongly connected firms (Kline et al., 2020). Because each of these included subsamples is selected by the researcher on observable differences from the corresponding excluded subsample, the included and excluded subsamples may have very different distributions of firm effects. Thus, even if these approaches recover the true variance of firm effects for the included subsample, it is not obvious that one can extrapolate results from the included subsample to the full population.

In this appendix, we characterize analytically and numerically the bias introduced by approximating the variance of firm effects in the population using estimates for a selected subsample. Let $V_1$ denote the variance of firm effects for the included subsample, $V_0$ denote the variance of firm effects for the excluded subsample, and $\pi$ denote the share of workers employed by the included subsample of firms.[24] By

[24] Throughout this paper, we refer to the largest connected set of firms as the population of interest,

the law of total variance, the variance of firm premiums in the full population, $\bar{V}$, is related to $V_1$ and $V_0$ by the following decomposition:

$$\bar{V} = \underbrace{\pi V_1 + (1-\pi)V_0}_{\text{Within Variance}} + \underbrace{\pi E_1^2 + (1-\pi)E_0^2 - \bar{E}^2}_{\text{Between Variance}}, \tag{D8}$$

where $\bar{E}$, $E_1$, and $E_0$ denote the mean firm effect in the population, included subsample, and excluded subsample, respectively. Normalizing $\bar{E} = 0$ without loss of generality, this expression becomes,

$$\bar{V} = \underbrace{\pi V_1 + (1-\pi)V_0}_{\text{Within Variance}} + \underbrace{\pi(1-\pi)(E_1 - E_0)^2}_{\text{Between Variance}}. \tag{D9}$$

which emphasizes the importance of the difference in mean firm effects between the included and excluded subsamples, $E_1 - E_0$.

The object of interest is the variance of firm effects in the population, $\bar{V}$. Assume the researcher knows $V_1$ and $\pi$, but does not know $V_0$ or $E_1 - E_0$. Using the above decomposition, the bias when using $V_1$ as an approximation to $\bar{V}$ is given by,

$$\underbrace{V_1 - \bar{V}}_{\text{Subsample Bias}} = \underbrace{(1-\pi)(V_1 - V_0)}_{\text{Within Contribution}} - \underbrace{\pi(1-\pi)(E_1 - E_0)^2}_{\text{Between Contribution}}. \tag{D10}$$

This expression provides three results. First, $V_1$ tends to be upward-biased (downward-biased) for $\bar{V}$ if the excluded subsample is relatively less (more) variable. Second, $V_1$ becomes more downward-biased as the mean firm premium difference grows between the included and excluded subsamples. For example, if larger firms have much greater mean firm premiums than smaller firms, then $E_1 - E_0$ is large when the included set only contains large firms, introducing substantial downward-bias. Third, $\lim_{\pi \to 1} \bar{V} = V_1$, so $V_1$ provides a good approximation to $\bar{V}$ when the excluded subsample contains a small share of the population. In the US, 5% of workers are employed by firms that are excluded from the leave-one-out set ($\pi = 0.95$), while 22% of workers are employed by firms that are excluded by the 20 workers per firm sample restriction

as this is traditionally the population under focus in studies based on the AKM model. However, one may be interested in the population inclusive of disconnected firms. The CRE approach can be used to produce variance component estimates for the entire sample, including disconnected firms.

($\pi = 0.78$). Thus, there may be little bias when only using the leave-one-out set to learn about the population but substantial bias when only using large firms.

We now characterize the bias numerically. First, it is useful to parameterize the bias relative to the size of $V_1$ as follows:

$$\underbrace{\frac{V_1 - \bar{V}}{V_1}}_{\text{\% Subsample Bias}} = \underbrace{(1-\pi)V_Z}_{\text{\% Within Contribution}} - \underbrace{\pi(1-\pi)E_Z^2}_{\text{\% Between Contribution}}, \tag{D11}$$

where $V_Z \equiv \frac{V_1 - V_0}{V_1}$ and $E_Z \equiv \frac{E_1 - E_0}{\sqrt{V_1}}$. We can use this parameterization to choose a reasonable range of numerical values over which to evaluate the bias. For the variance, suppose that $V_0$ is in the range from 50% below $V_1$ to 50% above $V_1$, which is equivalent to assuming $V_Z \in [\frac{-1}{2}, \frac{1}{2}]$. For the mean firm effect, suppose $E_0$ is in the range from equal to $E_1$ to a standard deviation different from $E_1$, which is equivalent to assuming $E_Z \in [0, 1]$. Note that we focus on $E_1 \geq E_0$ because the restrictions imposed in the literature favor keeping large firms in the included subsample, and we expect larger firms to have greater firm effects. For now, we choose $\pi = 0.78$, which corresponds to the share of workers in the included sample when imposing a minimum of 20 workers per firm; we consider alternative choices of $\pi$ below.

In Appendix Figure F15(a), we plot the Between contribution across $V_Z$. We find that the Between contribution leads to a downward-bias of about 5% when the mean firm effect differs by one-half of a standard deviation ($E_Z = \frac{1}{2}$). However, this increases to a downward-bias of about 17% when the mean firm effect differs by a full standard deviation ($E_Z = 1$). We see that, because the bias is increasing at an increasing rate in $E_Z$, it can become quite large when the included and excluded subsamples contain firms of different average sizes. In Appendix Figure F15(b), we plot the Within contribution across $V_Z$. We find that the Within contribution leads to a downward-bias of about 10% when the excluded sample is half as variable as the included sample ($V_Z = \frac{-1}{2}$), and a 10% upward-bias when the excluded sample is 50% more variable than the included sample. The bias grows linearly in, and has the same sign as, $V_1 - V_0$. In Appendix Figures F15(c-d), we plot the total bias across combinations of $(E_Z, V_Z)$. We see that the Between and Within contributions to the bias can combine to imply a downward-bias of nearly 30% or an upward-bias of about 10%.

Lastly, in Appendix Figure F16, we examine numerically how the bias depends on $\pi$. We compare the value of $\pi$ in the US when using a 2-year panel ($\pi = 0.47$), a 3-year panel ($\pi = 0.62$), and a 6-year panel ($\pi = 0.93$). In Appendix Figure F16(a), we find that reducing $\pi$ from the value in the 6-year panel to the value in the 3-year panel magnifies the downward-bias substantially, from a maximum of 5% downward bias to a maximum of 25% downward bias. However, reducing $\pi$ from the value in the 3-year panel to the value in the 2-year panel has little impact on the Between contribution. This is because $\pi$ enters the Between contribution as $\pi(1-\pi)$, which is maximized at $\pi = 0.5$ and relatively flat near this value. In Appendix Figure F16(b), we find that reducing $\pi$ from the value in the 6-year panel to the value in the 3-year panel has the effect of rotating the line of bias. The absolute value of the bias rises from a maximum of about 4% in the 6-year panel to a maximum of about 19% in the 3-year panel. Reducing $\pi$ from the value in the 3-year panel to the value in the 2-year panel further rotates the line such that the absolute value of the bias rises to a maximum of about 26%. Combining the Between and Within contributions to bias, we see that using only the included subsample in the estimation can lead to 11% downward bias in the 6-year panel but more than 50% downward bias in the 2-year panel.

# E Comparisons to Existing Work

In this section, we compare the results obtained from the methods we use to those obtained in previous studies.

## E.1 Italian data

We first compare our results on the Italian data to those from the May 2020 version of Kline et al. (2020). Rather than our baseline sample selection (described in Section 2), we use their replication code to construct a sample as similar to theirs as possible. A key difference from our baseline analysis is that we now focus only on the years 1999 and 2001. Comparing descriptive statistics of our replication sample in row 3 of Appendix Table F4 to those reported in Table 1 of Kline et al. (2020), we find that the sample counts for number of observations, movers, and firms are nearly identical, and the estimates of the total variance of daily wages are very close (0.199 compared to 0.206).

In Appendix Table F4, we also apply the FE, FE-HO, and FE-HE estimators to our Kline et al. (2020) replication sample. Our implementation of the estimators differs from Kline et al. (2020) in two ways. First, we collapse yearly data to spell level data as described in Appendix A. Second, as in our main analysis, we use only one spell observation per stayer spell rather than assuming errors are uncorrelated over time within stayer spells. This choice matters for FE-HO, but not for FE-HE.

We find that these differences in implementation do not materially change the estimates when using our replication sample. Using our replication sample, we find similar results as in Kline et al. (2020). Concretely, we compare estimates from our replication sample in row 3 of Appendix Table F4 to Table 2 of Kline et al. (2020). The contribution of firm effects to wage inequality is 19% for FE, 15% for FE-HO, and 14% for FE-HE, while Kline et al. (2020) estimate 19% for FE, 14% for FE-HO, and 13% for FE-HE. We find that the contribution of sorting to wage inequality is 6% for FE, 15% for FE-HO, and 16% for FE-HE, while Kline et al. (2020) estimate 4% for FE, 11% for FE-HO, and 16% for FE-HE.

In sum, we conclude that our implementation of the estimators delivers similar results to Kline et al. (2020) on the Italian data once we use a similar sample.

## E.2 US data

We now compare our results on the US tax data to those from Song et al. (2019) (Table 3, interval 2007-2013) and Sorkin (2018) (Table 1). We differ from their papers in three key dimensions. First, we consider the full sample of W-2 tax records, whereas Sorkin (2018) considers LEHD data (UI records) from 27 states and Song et al. (2019) consider SSA earnings records for men. Second, we use a minimum earnings threshold of 100% of the annualized minimum wage, whereas Sorkin (2018) and Song et al. (2019) set the minimum earnings threshold to 25% of the annualized minimum wage. Third, since we want to include small firms when studying inequality, we do not impose a minimum firm size restriction in the baseline results. By comparison Sorkin (2018) restricts the sample to firms with a minimum of 15 workers in each year (among workers who appear at least twice in the sample) and Song et al. (2019)

restrict the sample to firms with at least 20 workers in each year.

To understand the impact of the restrictions made by Sorkin (2018) and Song et al. (2019), we now consider alternative minimum earnings and minimum firm size thresholds:

**Minimum earnings threshold.** As discussed in Subsection 2.2, we examine how our results change when imposing minimum earnings thresholds ranging from 25% to 100% of the annualized minimum wage. When using the 25% threshold, we find that the variance of log earnings is 0.82 (see Appendix Table F3). This estimate is higher than the estimate of 0.67 reported in Table 1 of Sorkin (2018), and lower than the estimate of 0.92 reported in Table 3 of Song et al. (2019) for years 2007-2013. When increasing the minimum earnings threshold, the variance of log earnings must mechanically decline, and our baseline sample (100% minimum earnings threshold) has a substantially smaller variance of 0.41. However, the between-firm share of variance is nearly constant at about 40% across all minimum earnings thresholds, which is the same number reported in Table 2 of Song et al. (2019). Shifting attention to the AKM estimates, we find that the FE estimate of the share of earnings variation due to firm effects is somewhat decreasing in the minimum earnings threshold while the share due to sorting is strongly decreasing (see Appendix Figure F2).

**Minimum firm size threshold.** As discussed in detail in Section 6, we examine how our results change when imposing minimum firm size thresholds ranging from 0 to 50 workers. Neither the variance of log earnings nor the between-firm share of earnings variation changes materially with the minimum firm size threshold. However, the FE estimate of the share of earnings variation due to firm effects is decreasing in the firm size threshold while the share due to sorting is increasing (see Appendix Figure F7). When imposing a minimum firm size threshold of 20 workers, the FE estimate of the share of earnings variation due to sorting rises to 9% (see Appendix Table F3), which is close to the estimates by Sorkin (2018) and Song et al. (2019) of 10% and 12%, respectively.

Taken together, the results in Appendix Table F3 help explain how our estimates compare to Sorkin (2018) and Song et al. (2019). On the one hand, imposing a higher earnings threshold in the baseline sample tends to decrease our FE estimate of the contribution of firm effects to wage inequality and decrease our FE estimate of the contribution of sorting. On the other hand, imposing a lower firm size threshold in our baseline sample for the US tends to increase our FE estimate of the contribution of firm effects to wage inequality and decrease our FE estimate of the contribution of sorting. These differences partially offset each other for the contribution of firm effects, resulting in a FE estimate of the share of earnings inequality due to firm effects at 12%, in between the estimates of Sorkin (2018) and Song et al. (2019) at 14% and 9%, respectively. However, both tend to decrease our FE estimate of the share due to sorting relative to the estimates of Sorkin (2018) and Song et al. (2019).

# F Additional Tables and Figures

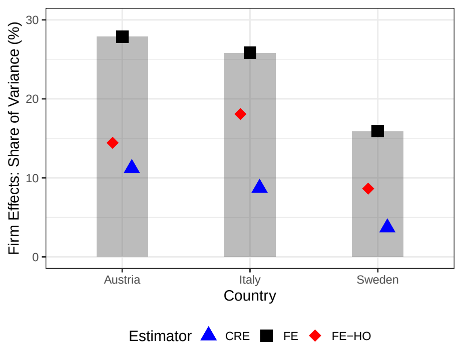
Table F1: Survey of Estimates in the Existing Literature

| Paper | Country | Years | Total Var | Firm Effects | Sorting |
|---|---|---|---|---|---|
| Abowd et al. (1999) | France | 1976-1987 ($\neq$ 1981, 1983) | 0.269 | 87.0% | 23.1% |
| Abowd et al. (2002) | France | 1976-1987 ($\neq$ 1981, 1983) | 0.269 | 30.1% | -13.6% |
| Abowd et al. (2002) | USA, WA | LEHD 1984-1993 | 0.278 | 19.2% | -1.0% |
| Abowd et al. (2004)⋆ | France | 1976-1996 | 0.354 | 61.4% | -15.9% |
| Abowd et al. (2004)⋆ | USA | LEHD 1990-2000 | 0.800 | 16.3% | 2.5% |
| Alvarez et al. (2018) | Brasil | 1988-1992 | 0.750 | 21.3% | 8.7% |
| Alvarez et al. (2018) | Brasil | 1992-1996 | 0.750 | 22.7% | 9.3% |
| Alvarez et al. (2018) | Brasil | 1996-2000 | 0.690 | 23.2% | 10.1% |
| Alvarez et al. (2018) | Brasil | 2000-2004 | 0.620 | 21.0% | 9.7% |
| Alvarez et al. (2018) | Brasil | 2004-2008 | 0.530 | 17.0% | 9.4% |
| Alvarez et al. (2018) | Brasil | 2008-2012 | 0.470 | 14.9% | 9.6% |
| Andrews et al. (2008)⋆ | Germany | LIAB 1993-1997, Bias Corr. | 0.055 | 21.5% | -6.6% |
| Andrews et al. (2008)⋆ | Germany | LIAB 1993-1997, Not Corr. | 0.057 | 23.5% | -9.0% |
| Bagger and Lentz (2019) | Denmark | 1985-2003 | 0.097 | 14.4% | -1.0% |
| Card et al. (2013) | Germany | Universe, 1985-1991 | 0.137 | 18.2% | 1.1% |
| Card et al. (2013) | Germany | Universe, 2002-2009 | 0.249 | 21.3% | 8.2% |
| Card et al. (2018)⋆ | Portugal | 2005-2009 | 0.275 | 22.8% | 6.5% |
| Lopes de Melo (2018)⋆ | Brasil | 1995-2005 | 0.601 | 29.9% | 1.8% |
| Engbom and Moser (2021) | Brasil | 1996-2000 | 0.690 | 23.2% | 10.1% |
| Goldschmidt and Schmieder (2017) | Germany | IEB, 2008 | 0.205 | 26.7% | 10.4% |
| Goldschmidt and Schmieder (2017) | Germany | IEB, 1985 | 0.132 | 21.9% | -1.9% |
| Goux and Maurin (1999)⋆ | France | 1990-1992 | 0.181 | 12.9% | -6.1% |
| Goux and Maurin (1999)⋆ | France | 1991-1993 | 0.157 | 30.2% | -2.5% |
| Goux and Maurin (1999)⋆ | France | 1992-1994 | 0.154 | 65.3% | -24.0% |
| Goux and Maurin (1999)⋆ | France | 1993-1995 | 0.151 | 19.6% | 0.7% |
| Gruetter and Lalive (2009) | Austria | 1990-1997 | 0.224 | 26.6% | -11.3% |
| Iranzo et al. (2008) | Italy | Manufacturing, 1981-1997 | 0.110 | 13.1% | 6.4% |
| Kline et al. (2020) | Italy | 1999-2001, AKM | 0.184 | 19.4% | 2.1% |
| Kline et al. (2020) | Italy | 1999-2001, Homosk. Corr. | 0.184 | 16.0% | 5.3% |
| Kline et al. (2020) | Italy | 1999-2001, Leave-out | 0.184 | 13.0% | 8.0% |
| Song et al. (2019) | USA | 1980-1986 | 0.708 | 11.9% | 2.3% |
| Song et al. (2019) | USA | 1987-1993 | 0.777 | 9.7% | 3.7% |
| Song et al. (2019) | USA | 1994-2000 | 0.828 | 8.1% | 4.6% |
| Song et al. (2019) | USA | 2001-2007 | 0.884 | 8.5% | 5.3% |
| Song et al. (2019) | USA | 2007-2013 | 0.924 | 8.8% | 5.8% |
| Sorkin (2018) | USA | LEHD 2000-2008 | 0.700 | 14.0% | 5.0% |
| Woodcock (2015) | USA | 2007-2013 | 0.410 | 19.5% | -0.5% |

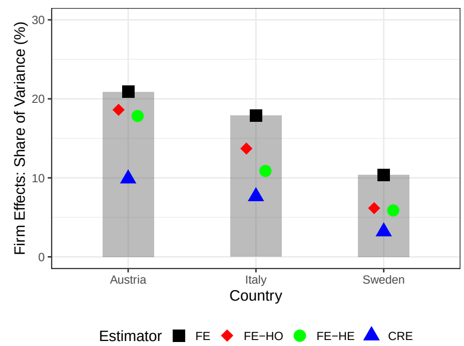
*Notes:* In this table, we provide a survey of estimates from a set of studies that estimate the contribution to earnings or wage inequality of firm effects and the sorting of workers to firms using the FE estimator. $*$ indicates that the total variance is not reported so we estimate it as Var($\psi$)+Var($\alpha$)+2Cov($\psi, \alpha$).

Figure F1: Workers Employed the Full Year by a Single Firm

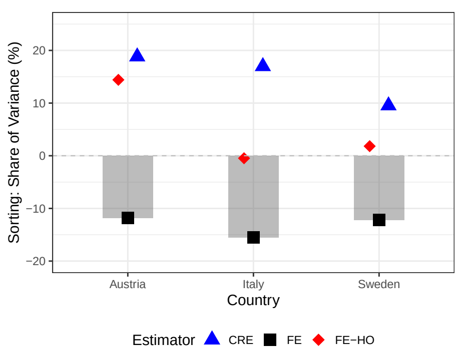
(a) Firm effects (connected set)



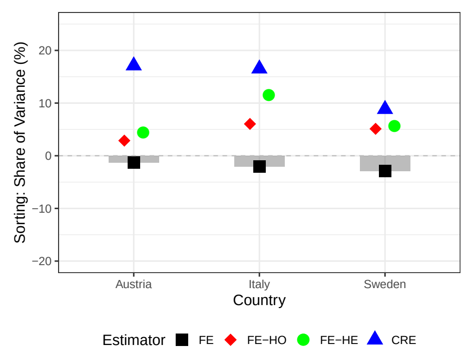

(b) Firm effects (leave-one-out set)




(c) Sorting (connected set)



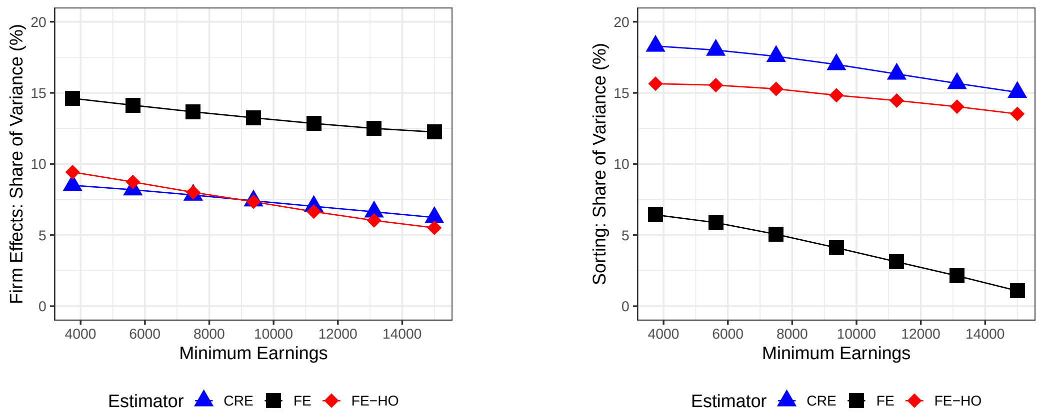

(d) Sorting (leave-one-out set)




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigures a and b) and the sorting of workers to firms (Subfigures c and d) in Austria, Italy, and Sweden. We consider the connected (Subfigures a and c) and leave-one-out (Subfigures b and d) sets of firms. We consider only workers employed in the firm for the full calendar year.

Figure F2: Minimum Earnings Threshold for Defining Full-time Equivalence in the US

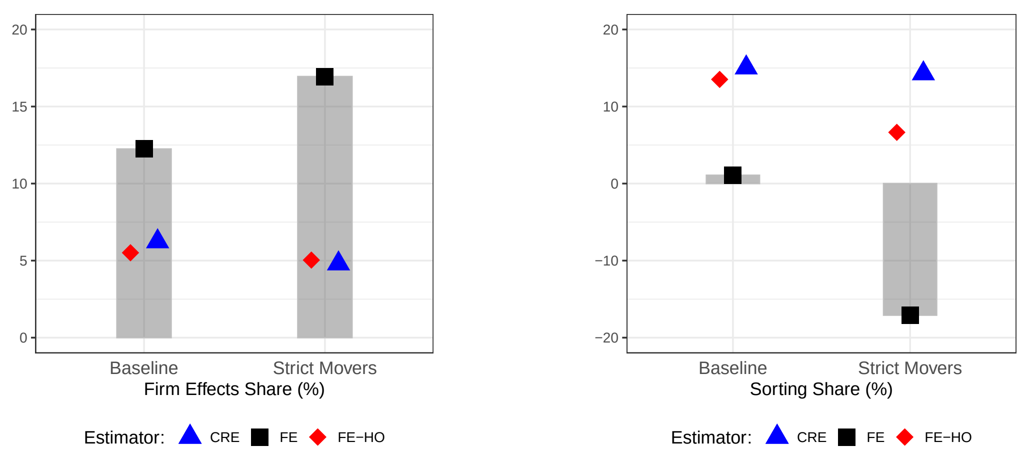
(a) Firm Effects: Share (%)    (b) Sorting: Share (%)




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We restrict the sample to workers with at least the annual earnings (at the highest-paying employer) indicated on the x-axis. We consider the connected set of firms for each restricted sample.

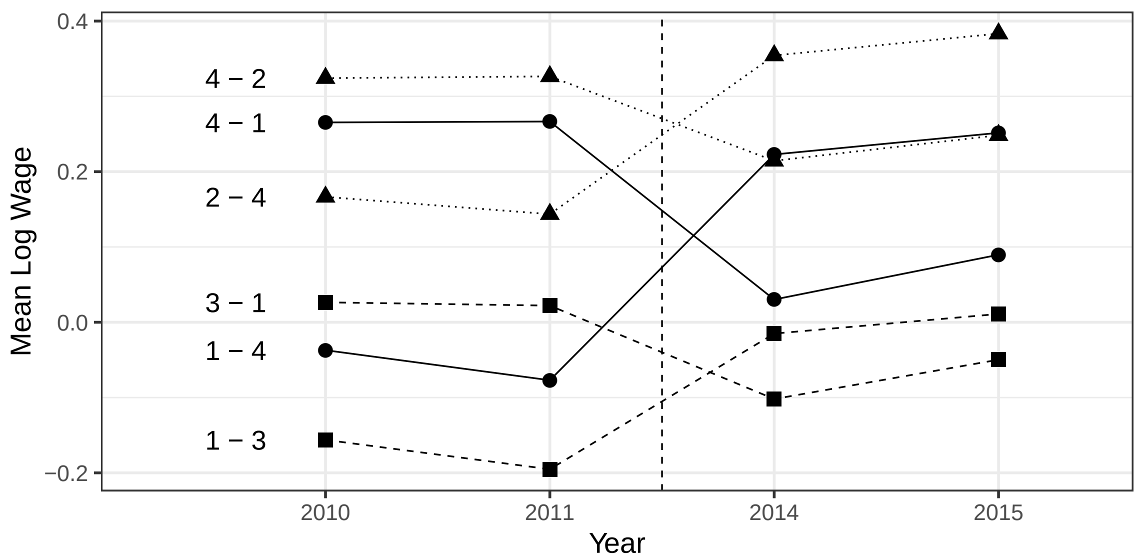
Figure F3: Firm Effects and Sorting in the US over Mover Definitions

(a) Firm Effects    (b) Sorting




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We compare estimates using the baseline definition of movers and the strict definition of movers defined in the text.

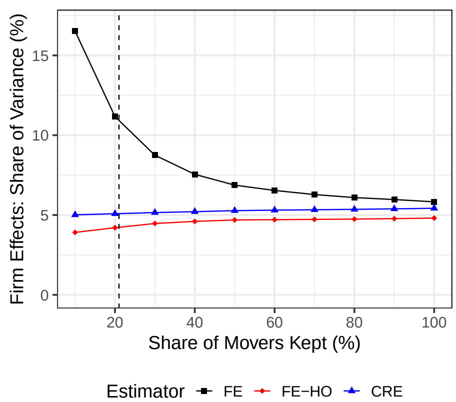
Figure F4: US Sample: Event Study around Moves



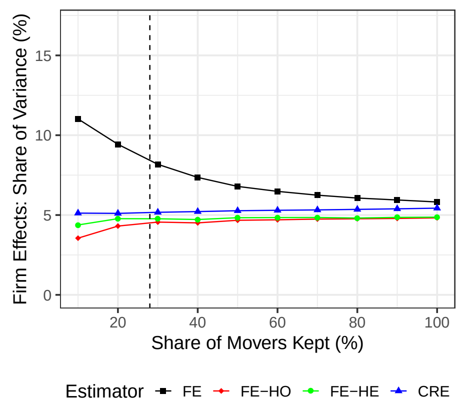

*Notes:* In this figure, we classify firms into four equally sized groups based on the mean earnings of stayers in the firm (with 1 and 4 being the group with the lowest and highest mean earnings, respectively). We compute mean log-earnings for the workers that move firms during 2012-2013. Note that the employer differs between event times 2012 and 2013, but we do not know exactly when the change in employer occurred. To avoid concerns over workers exiting and entering employment during these years, we do not display the transition years.

Figure F5: Evidence on Limited Mobility Bias in the United States

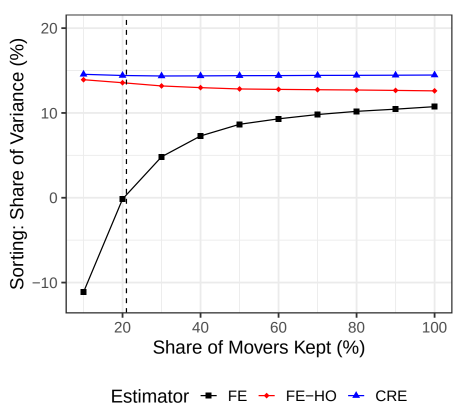
(a) Firm effects (connected set)



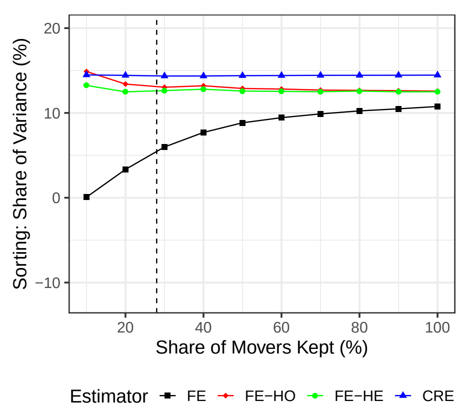

(b) Firm effects (leave-one-out set)




(c) Sorting (connected set)



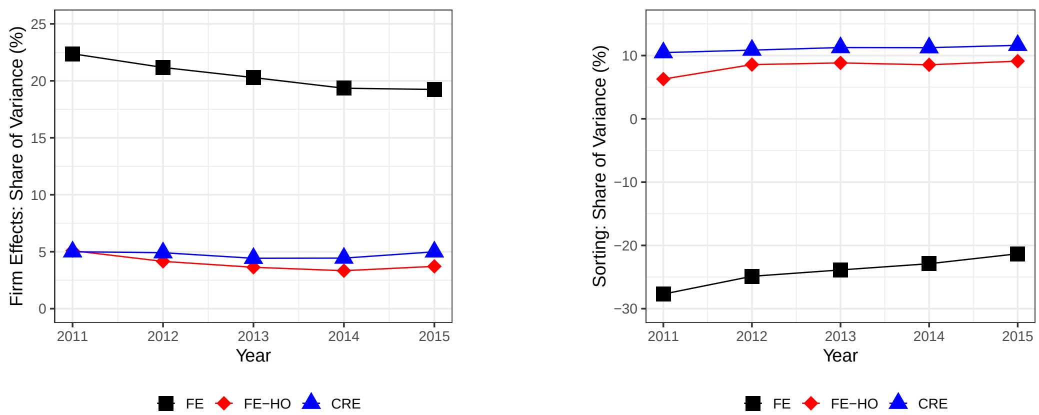

(d) Sorting (leave-one-out set)




*Notes:* In this figure, we consider the subset of firms in the US with at least 15 movers. We randomly remove movers within each firm and re-estimate the variance of firm effects and covariance between firm and worker effects using the various estimators. For each estimator, we repeat this procedure several times then average the estimates across repetitions. The procedure allows us to keep the connected or leave-one-out set of firms the same and examine the bias that results from having fewer movers available in estimation. The vertical dashed line approximates the point at which movers per firm in this sample matches movers per firm in the full sample.

Figure F6: Firm Effects and Sorting in the US: Short-Panel Estimation (Connected Set)

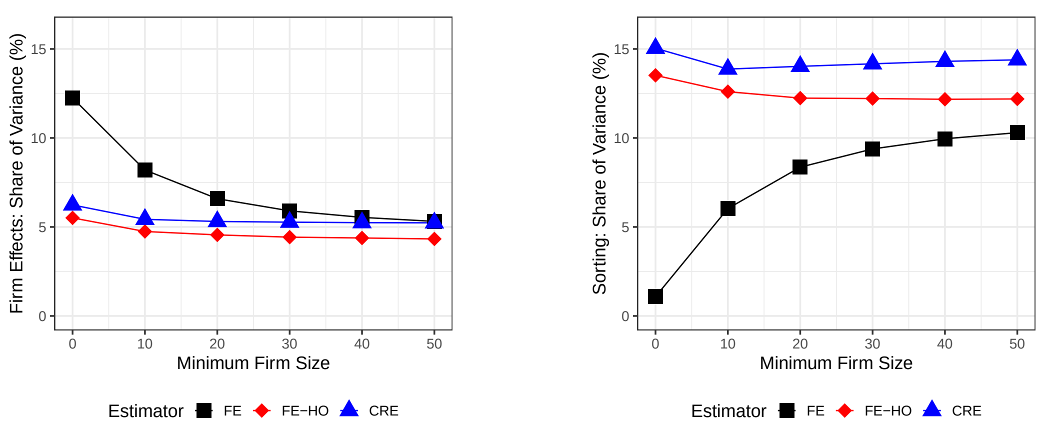
(a) Firm Effects (b) Sorting




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We consider the connected set of firms, and compare estimates on each 2-year panel during 2010-2015 (the latter year of the 2-year panel is indicated on the x-axis).

Figure F7: Firm Size Restrictions in the US (Connected Set)

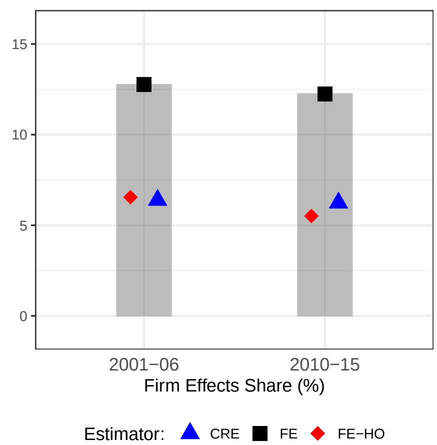
(a) Firm Effects (b) Sorting



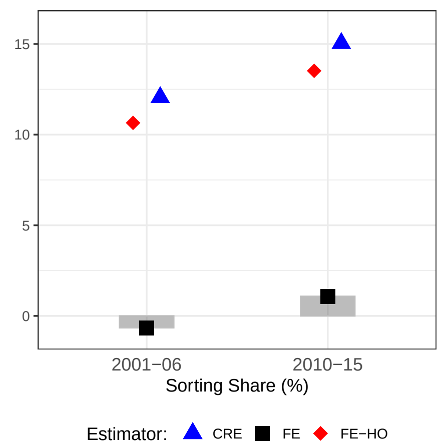

*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings and wage inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We restrict the sample to firms with at least the number of workers indicated on the x-axis. We consider the connected set of firms for each restricted sample.

Figure F8: Firm Effects and Sorting in the United States over Time

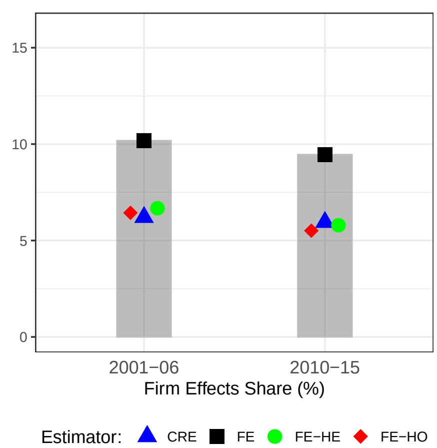
(a) Firm Effects (Connected Set)



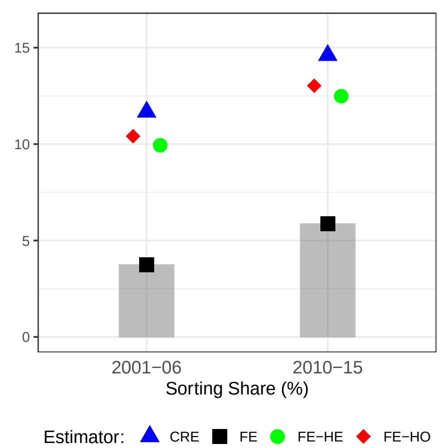

(b) Sorting (Connected Set)




(c) Firm Effects (Leave-one-out Set)

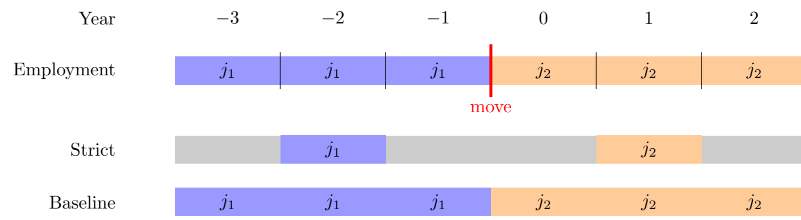



(d) Sorting (Leave-one-out Set)




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We consider the connected set of firms. We compare the 6-year panel during 2001-2006 to the 6-year panel during 2010-2015.

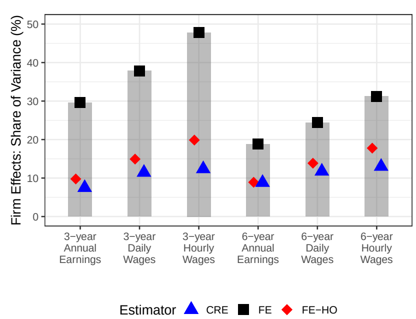
Figure F9: Visualizing Alternative Mover Definitions for the US



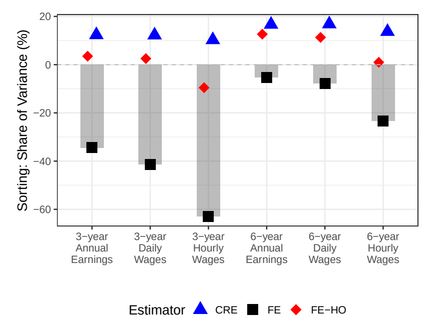

*Notes:* In this figure, we provide a diagram to help visualize the difference between the main definition of a mover ("Baseline") and the mover definition that uses only intermediate years within spells ("Strict").

Figure F10: Norway: Annual Earnings, Daily Wages, and Hourly Wages

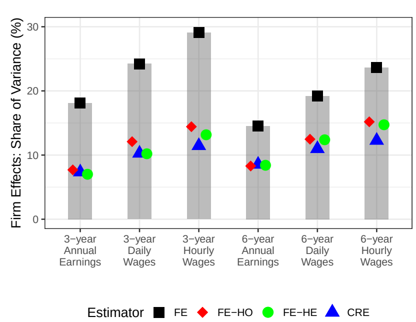
(a) Firm effects (connected set)



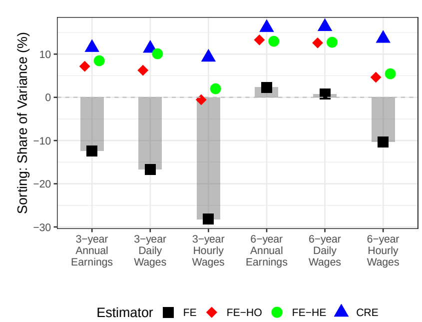

(b) Sorting (connected set)




(c) Firm effects (leave-one-out set)



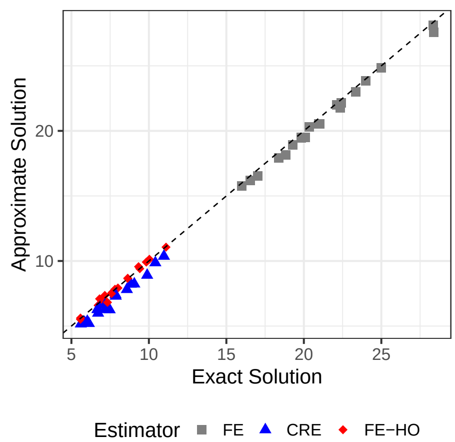

(d) Sorting (leave-one-out set)



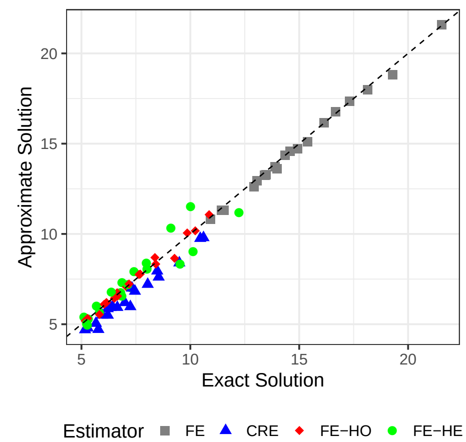

*Notes:* In this figure, we provide FE, FE-HO, FE-HE, and CRE estimates of the contribution to earnings or wage inequality of firm effects (Subfigures a and c) and the sorting of workers to firms (Subfigures b and d) in Norway. We consider the connected set of firms (Subfigures a and b) and the leave-one-out set of firms (Subfigures c and d) for the 6-year panel and the 3-year panel. We compare results for three outcome measures: log annual earnings, log daily wages, and log hourly wages.

Figure F11: Exact and Approximate Solutions: Firm Effects Variance (%) for the Small US States

(a) Connected Set

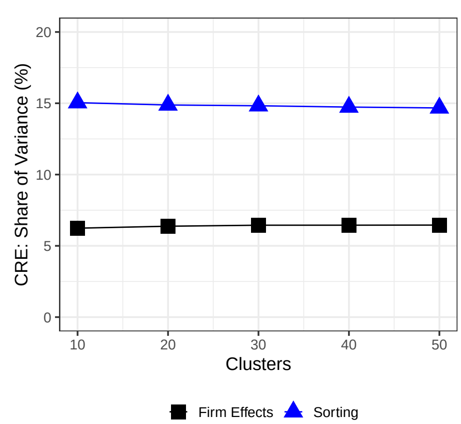



(b) Leave-one-out Set




*Notes:* In this figure, we provide FE, FE-HO, and CRE estimates for the connected set (Subfigure a) and FE, FE-HO, FE-HE, and CRE estimates for the leave-one-out set (Subfigure b) of the contribution to earnings inequality of firm effects in the 20 smallest US states. We compare the exact solution (x-axis) and the approximate solution (y-axis) described in the text, so that the dashed 45-degree line represents equality between the exact and approximate solutions.

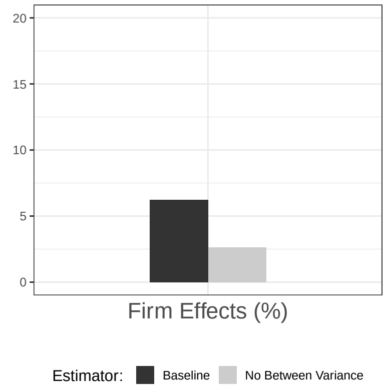
Figure F12: Number of Groups for CRE Estimates in the US (Connected Set)



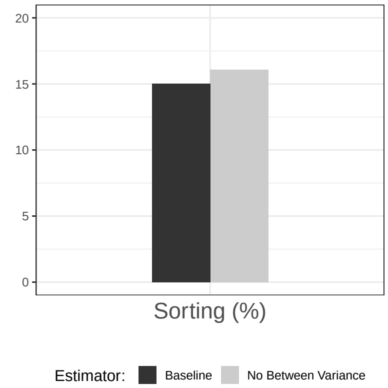

*Notes:* In this figure, we provide CRE estimates of the contribution to earnings inequality of firm effects and the sorting of workers to firms in the US. We consider the connected set of firms, and vary the number of firm groups considered in the CRE estimation procedure (indicated on the x-axis).

Figure F13: Firm Effects and Sorting in the US over Type of CRE Estimator (Connected Set)

(a) Firm Effects



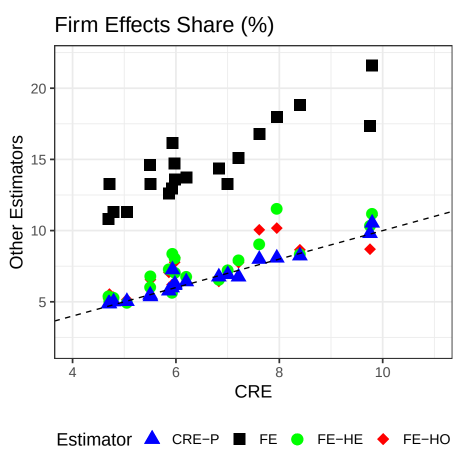

(b) Sorting



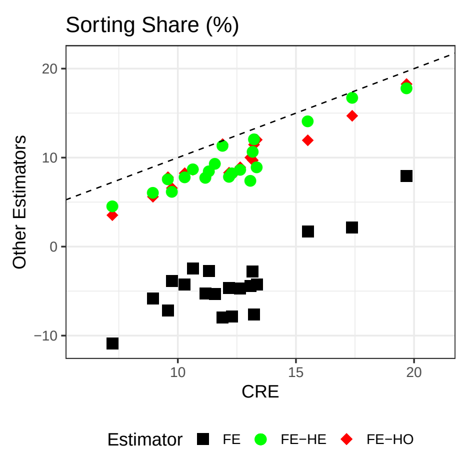

*Notes:* In this figure, we provide CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We compare the baseline CRE estimates to the posterior estimates for a random-effects specification that does not condition on firm groups.

Figure F14: Leave-one-out Set: Small US States

(a) Firm effects



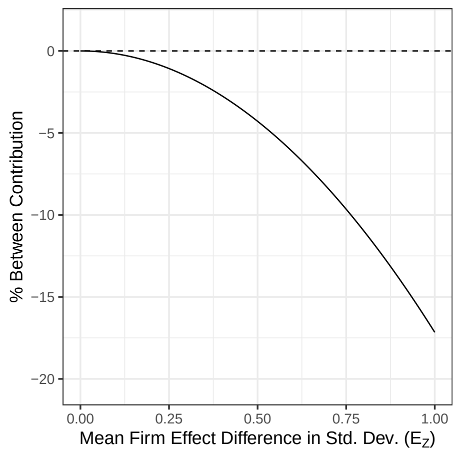

(b) Sorting



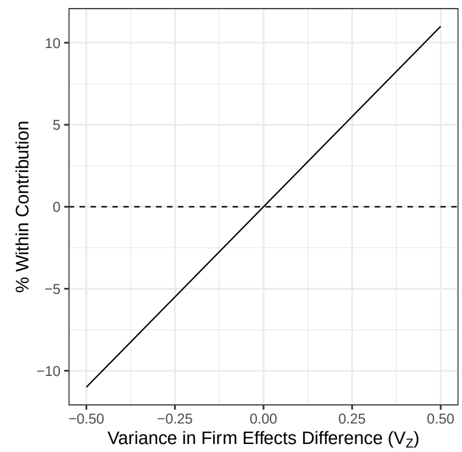

*Notes:* In this figure, we provide FE, FE-HO, FE-HE, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the 20 smallest US states. We consider the leave-one-out set of firms within each state. CRE estimates are displayed on the x-axis, and the dashed 45-degree line represents equality between CRE and the alternate estimators. The posterior CRE estimator (CRE-P) for firm effects is also displayed (Subfigure a).

Figure F15: Bias when using Estimates for a Subsample to Approximate the Variance of Firm Effects in the Full Population (given $\pi = 0.78$)

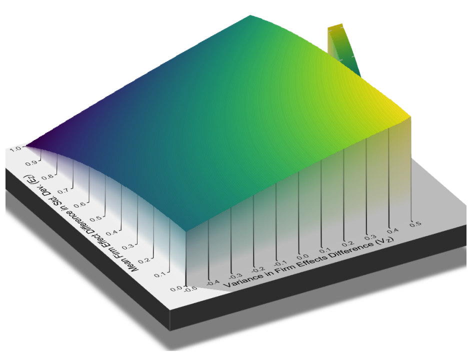
(a) Between Contribution to Bias



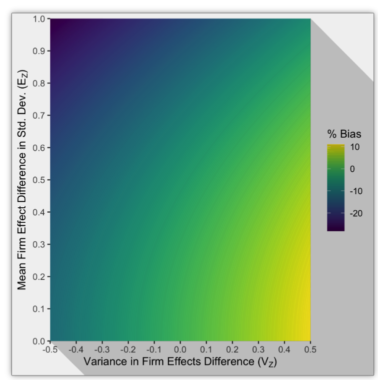

(b) Within Contribution to Bias




(c) Total Bias (side view)

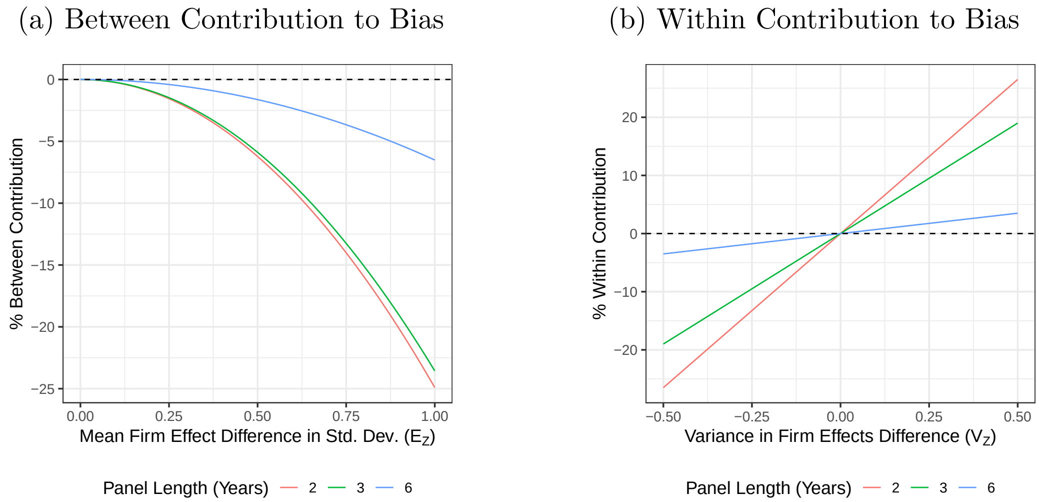



(d) Total Bias (overhead view)




*Notes:* In this figure, we use equation (D11) to visualize the bias that arises from using the variance of firm premiums estimated for a subsample of firms to approximate the variance of firm effects in the full population. We calibrate $\pi = 0.78$, which corresponds to the 20 workers per firm sample restriction in the US data. Subfigure (a) provides the between-firm contribution to the bias, subfigure (b) provides the within-firm contribution to the bias, and subfigures (c-d) provide the joint determination of the total bias by both the between-firm and within-firm components.

Figure F16: Bias when using Estimates for a Subsample to Approximate the Variance of Firm Effects in the Full Population (various choices of $\pi$ based on panel length)




*Notes:* In this figure, we use equation (D11) to visualize the bias that arises from using the variance of firm premiums estimated for a subsample of firms to approximate the variance of firm effects in the full population. We compare the value of $\pi$ in the US when using a 2-year panel ($\pi = 0.47$), a 3-year panel ($\pi = 0.62$), or a 6-year panel ($\pi = 0.93$). Subfigure (a) provides the between-firm contribution to the bias and subfigure (b) provides the within-firm contribution to the bias.

## Table F2: Baseline Results

| | | sample information | | | | | | $Var(\psi)$ | | | | $2 \times Cov(\alpha, \psi)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | set | firms | workers | movers | var(y) | btw firm | FE | FE-HO | FE-HE | CRE | FE | FE-HO | FE-HE | CRE |
| US | 6 | connected | 2,568 | 55,464 | 14,888 | 0.414 | 39.6% | 12.2% | 5.5% | - | 6.2% | 1.1% | 13.5% | - | 15.0% |
| | 6 | leave-out | 1,689 | 52,484 | 13,968 | 0.416 | 38.8% | 9.5% | 5.5% | 5.8% | 5.9% | 5.9% | 13.0% | 12.5% | 14.6% |
| US | 3 | connected | 1,241 | 36,826 | 4,252 | 0.436 | 38.2% | 16.3% | 4.1% | - | 5.3% | -12.0% | 11.7% | - | 12.5% |
| | 3 | leave-out | 670 | 33,031 | 3,645 | 0.440 | 37.6% | 10.4% | 4.3% | 4.5% | 5.0% | -0.8% | 11.0% | 10.6% | 12.1% |
| Austria | 6 | connected | 206 | 3,396 | 1,123 | 0.187 | 45.5% | 18.7% | 15.3% | - | 11.7% | 4.7% | 10.5% | - | 19.6% |
| | 6 | leave-out | 140 | 3,240 | 1,055 | 0.182 | 43.7% | 15.5% | 12.7% | 12.9% | 11.1% | 8.7% | 13.5% | 13.0% | 18.9% |
| Austria | 3 | connected | 117 | 2,845 | 387 | 0.183 | 43.7% | 19.7% | 12.1% | - | 10.1% | -5.3% | 9.3% | - | 17.5% |
| | 3 | leave-out | 68 | 2,604 | 336 | 0.178 | 41.8% | 15.0% | 10.7% | 13.9% | 9.2% | 1.5% | 9.7% | 3.2% | 16.2% |
| Italy | 6 | connected | 92 | 1,111 | 379 | 0.167 | 46.1% | 23.1% | 17.5% | - | 12.7% | -1.3% | 8.7% | - | 20.0% |
| | 6 | leave-out | 61 | 1,034 | 346 | 0.168 | 44.8% | 19.3% | 15.8% | 15.7% | 12.3% | 4.7% | 11.1% | 11.2% | 19.3% |
| Italy | 3 | connected | 54 | 864 | 148 | 0.176 | 44.9% | 24.1% | 15.7% | - | 11.0% | -8.4% | 7.7% | - | 17.7% |
| | 3 | leave-out | 30 | 755 | 121 | 0.181 | 43.5% | 18.5% | 14.6% | 10.9% | 10.2% | 1.3% | 8.8% | 16.1% | 17.2% |
| Norway | 6 | connected | 114 | 1,286 | 556 | 0.239 | 47.2% | 24.4% | 13.9% | - | 11.8% | -7.7% | 11.3% | - | 16.8% |
| | 6 | leave-out | 78 | 1,199 | 519 | 0.236 | 45.8% | 19.2% | 12.5% | 12.3% | 11.0% | 0.8% | 12.6% | 12.8% | 16.3% |
| Norway | 3 | connected | 63 | 986 | 203 | 0.229 | 44.5% | 37.8% | 14.9% | - | 11.5% | -41.3% | 2.5% | - | 12.2% |
| | 3 | leave-out | 37 | 856 | 175 | 0.227 | 42.6% | 24.2% | 12.1% | 10.2% | 10.3% | -16.7% | 6.3% | 10.1% | 11.3% |
| Sweden | 6 | connected | 63 | 1,921 | 608 | 0.164 | 31.6% | 14.6% | 8.2% | - | 5.0% | -8.1% | 3.9% | - | 10.3% |
| | 6 | leave-out | 52 | 1,850 | 596 | 0.164 | 30.9% | 11.6% | 7.8% | 7.1% | 4.7% | -3.2% | 3.7% | 5.0% | 10.0% |
| Sweden | 3 | connected | 42 | 1,497 | 237 | 0.161 | 31.3% | 23.6% | 11.6% | - | 4.6% | -28.5% | -5.4% | - | 9.0% |
| | 3 | leave-out | 29 | 1,377 | 221 | 0.161 | 30.2% | 15.5% | 8.9% | 7.4% | 4.3% | -14.1% | -1.3% | 1.5% | 8.1% |

*Notes:* In this table, we provide FE, FE-HO, FE-HE, and CRE estimates for the main results shown in the paper.

Table F3: Sample comparison for the US

| | T | set | firms | workers | movers | var(y) | btw firm | $Var(\psi)$ | $2 \times Cov(\alpha, \psi)$ |
|---|---|---|---|---|---|---|---|---|---|
| US (baseline results) | 6 | connected | 2,568 | 55,464 | 14,888 | 0.414 | 39.6% | 12.2% | 1.1% |
| Varying earnings threshold | | | | | | | | | |
| US (earnings threshold at $7,500, 50% of min. wage) | 6 | connected | 3,201 | 66,474 | 22,140 | 0.601 | 42.5% | 13.7% | 5.1% |
| US (earnings threshold at $3,750, 25% of min. wage) | 6 | connected | 3,489 | 72,228 | 27,325 | 0.823 | 42.9% | 14.6% | 6.4% |
| US (earnings threshold at $1,875 12.5% of min. wage ) | 6 | connected | 3,624 | 75,412 | 30,587 | 1.056 | 42.6% | 15.2% | 6.6% |
| Varying minimum firm size | | | | | | | | | |
| US (firm size $\geq$ 10) | 6 | connected | 748 | 48,810 | 10,704 | 0.425 | 38.0% | 8.2% | 6.0% |
| US (firm size $\geq$ 20) | 6 | connected | 360 | 43,140 | 8,361 | 0.434 | 38.1% | 6.6% | 8.4% |
| US (firm size $\geq$ 30) | 6 | connected | 227 | 39,581 | 7,015 | 0.440 | 38.3% | 5.9% | 9.4% |

*Notes:* In this table, we provide FE estimates for a range of sample construction rules. This focuses on 6 years of data and the connected set. See Appendix E for more details.

Table F4: Sample comparison for Italy

| | sample information | | | | | | | $Var(\psi)$ | | | $2 \times Cov(\alpha, \psi)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | set | firms | workers | movers | var(y) | btw firm | FE | FE-HO | FE-HE | FE | FE-HO | FE-HE |
| Italy (baseline results) | 6 | leave-out | 61 | 1,034 | 346 | 0.168 | 44.8% | 19.3% | 15.8% | 15.7% | 4.7% | 11.1% | 11.2% |
| Italy (baseline sample selection) | 3 | leave-out | 30 | 755 | 121 | 0.181 | 43.5% | 18.5% | 14.6% | 10.9% | 1.3% | 8.8% | 16.1% |
| Italy (sample construction of Kline et al. 2020) | 2 | leave-out | 42 | 985 | 164 | 0.206 | 46.1% | 19.4% | 14.8% | 14.4% | 6.1% | 14.9% | 15.7% |

*Notes:* In this table, we provide FE, FE-HO and FE-HE estimates for an alternative sample construction based on Kline et al. (2020) for comparison. We report results for the leave-out set. See Appendix E for more details.