Online Appendix to:
How Much Should we Trust Estimates of Firm Effects and Worker Sorting?

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

March 4, 2022

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.

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 σ^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

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

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 $\widehat{\sigma}^2_{it}$ 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 σ^2_{it} for stayers is equal to the average among movers at the same firm $j(i,t)$; that is,[1]

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

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 $\widehat{\Omega}(A) = \text{diag}[\widehat{\sigma}^2_{it}]$. 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 $\widehat{\sigma}^2_{it}$ as inverses. This is a difficult computational problem that is actively 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

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

[1] 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.

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 ε_{it}:

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

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

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

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 ψ_j and $\psi_{j'}$ because they share the same suppliers. In our implementation, we group firms and we only assume that ψ_j and $\psi_{j'}$ are uncorrelated conditional on j and j' belonging to different firm groups.[2] Likewise, we only assume that α_i and ψ_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 ε_{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 α_i, ψ_j, and ε_{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

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

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

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

To derive both (B4) and (B5) we have used the model in the first line, (2) and (B3)

[2] 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''.

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 γ, 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 γ 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

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

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,

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

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.[3]

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

[3] 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:

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

where the set S^{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:

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

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:

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

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

Posterior estimator. Under an additional joint normality assumption of γ and ε 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:

$$
\begin{aligned}
(\widehat{\Sigma}(A)^{-1}\widehat{\mu}(A) + A'\widehat{\Omega}(A)^{-1}Y)'\widehat{B}(A)^{-1}Q\widehat{B}(A)^{-1}(\widehat{\Sigma}(A)^{-1}\widehat{\mu}(A) + A'\widehat{\Omega}(A)^{-1}Y)&\\
+ \operatorname{Trace}(\widehat{B}(A)^{-1}Q)&,
\end{aligned}
$$

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, α_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 η_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 ψ, 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 η_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.

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

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

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 π denote the share of workers employed by the included subsample of firms.[4] By the law

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

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 π, 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 restric-

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.

tion ($\pi = 0.78$); see Appendix Tables F2 and F3, respectively. 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:

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

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 π 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 π when treating the full set of workers and firms in the 6-year panel (inclusive of disconnected firms) as the population of interest; see Table 1. We compare the value of π in the US connected set 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 π 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 π 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 π 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 π 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 π 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.

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 between 8% and 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

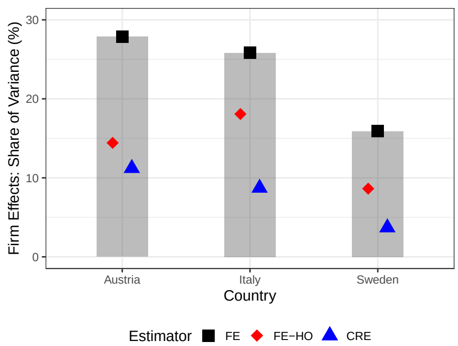
Table F1: Survey of Estimates in the Existing Literature

Paper	Country	Years	Total Var	Firm Effects	Sorting
Abowd et al. (1999)	France	1976-1987 ($\neq$ 1981, 1983)	0.269	87.0%	46.2%
Abowd et al. (1999)	France	1976-1987 ($\neq$ 1981, 1983)	0.269	82.8%	20.3%
Abowd et al. (2002)	France	1976-1987 ($\neq$ 1981, 1983)	0.269	30.1%	-27.2%
Abowd et al. (2002)	USA, WA	1984-1993	0.278	19.2%	-2.0%
Abowd et al. (2004)⋆	France	1976-1996	0.354	61.4%	-31.7%
Abowd et al. (2004)⋆	USA	LEHD, 1990-2000	0.771	16.9%	1.5%
Alvarez et al. (2018)	Brasil	1988-1992	0.750	21.3%	17.3%
Alvarez et al. (2018)	Brasil	1992-1996	0.750	22.7%	18.7%
Alvarez et al. (2018)	Brasil	1996-2000	0.690	23.2%	20.3%
Alvarez et al. (2018)	Brasil	2000-2004	0.620	21.0%	19.4%
Alvarez et al. (2018)	Brasil	2004-2008	0.530	17.0%	18.9%
Alvarez et al. (2018)	Brasil	2008-2012	0.470	14.9%	19.1%
Andrews et al. (2008)⋆	Germany	LIAB 1993-1997, Bias Corr.	0.055	21.5%	-13.1%
Andrews et al. (2008)⋆	Germany	LIAB 1993-1997, Not Corr.	0.057	23.5%	-18.0%
Bagger and Lentz (2019)	Denmark	1985-2003	0.097	14.4%	-2.1%
Card et al. (2013)	Germany	Universe, 1985-1991	0.137	18.2%	2.2%
Card et al. (2013)	Germany	Universe, 2002-2009	0.249	21.3%	16.5%
Card et al. (2018)⋆	Portugal	2005-2009	0.275	22.8%	13.0%
Lopes de Melo (2018)⋆	Brasil	1995-2005	0.601	30.0%	3.6%
Engbom and Moser (2021)	Brasil	2010-2014	0.453	19.4%	19.9%
Engbom and Moser (2021)	Brasil	1994-1998	0.709	29.9%	19.7%
Goldschmidt and Schmieder (2017)	Germany	IEB, 2008	0.205	26.7%	20.8%
Goldschmidt and Schmieder (2017)	Germany	IEB, 1985	0.132	21.9%	-3.8%
Goux and Maurin (1999)⋆	France	1990-1992	0.181	12.9%	-12.1%
Goux and Maurin (1999)⋆	France	1991-1993	0.157	30.2%	-5.1%
Goux and Maurin (1999)⋆	France	1992-1994	0.154	65.3%	-48.1%
Goux and Maurin (1999)⋆	France	1993-1995	0.151	19.6%	1.3%
Gruetter and Lalive (2009)	Austria	1990-1997	0.224	26.6%	-22.5%
Iranzo et al. (2008)	Italy	Manufacturing, 1981-1997	0.110	13.1%	2.1%
Kline et al. (2020)⋆	Italy	1999-2001, AKM	0.198	18.0%	3.9%
Kline et al. (2020)⋆	Italy	1999-2001, Homosk. Corr.	0.198	14.9%	9.8%
Kline et al. (2020)⋆	Italy	1999-2001, Leave-out	0.184	13.0%	16.0%
Song et al. (2019)	USA	1980-1986	0.708	11.9%	4.7%
Song et al. (2019)	USA	1987-1993	0.776	9.7%	7.3%
Song et al. (2019)	USA	1994-2000	0.828	8.1%	9.2%
Song et al. (2019)	USA	2001-2007	0.884	8.5%	10.6%
Song et al. (2019)	USA	2007-2013	0.924	8.7%	11.7%
Sorkin (2018)	USA	LEHD, 2000-2008	0.670	14.0%	10.0%
Woodcock (2015)	USA	LEHD, 1990-2000	0.410	19.5%	-1.0%

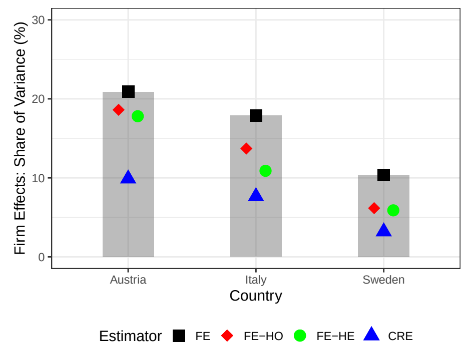
Notes: In this table, we provide a survey of estimates from a set of studies that estimate the contribution to earnings or wage inequality of firm effects and the sorting of workers to firms using the FE estimator. "Firm Effects" refers to $\text{Var}(\psi)/\text{Var}(Y)$ and "Sorting" refers to $2\text{Cov}(\alpha,\psi)/\text{Var}(Y)$,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

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

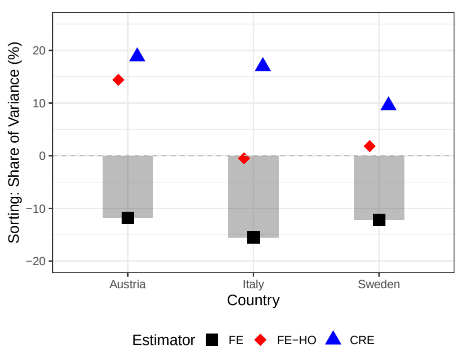
(a) Firm effects (connected set)



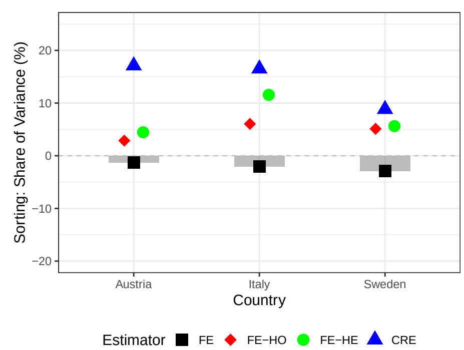
(b) Firm effects (leave-one-out set)



(c) Sorting (connected set)



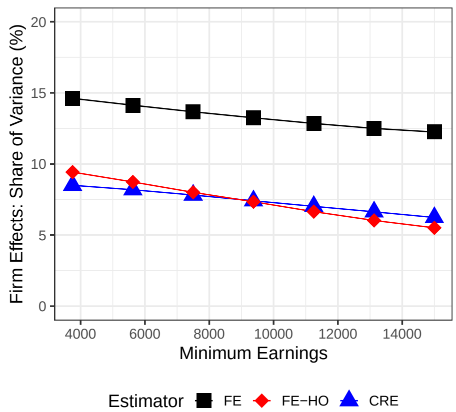
(d) Sorting (leave-one-out set)



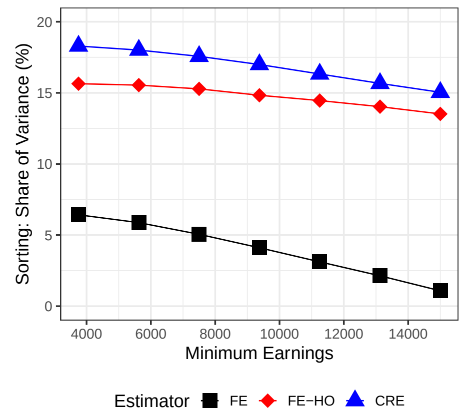
Notes: In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigures a and b) and the sorting of workers to firms (Subfigures c and d) in Austria, Italy, and Sweden. We consider the connected (Subfigures a and c) and leave-one-out (Subfigures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



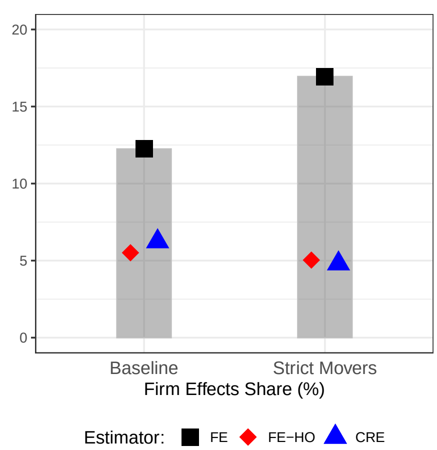
(b) Sorting: Share (%)



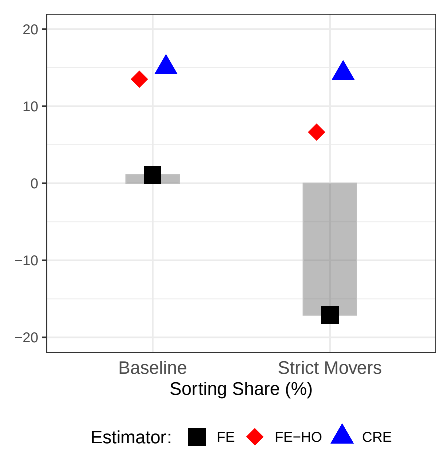
Notes: In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We restrict the sample to workers with at least the annual earnings (at the highest-paying employer) indicated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

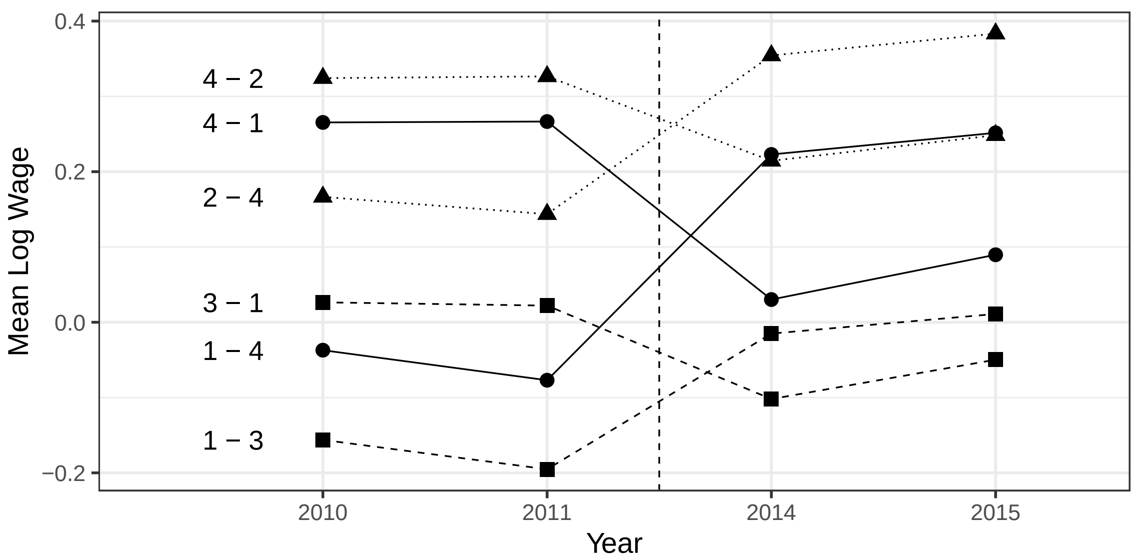


(b) Sorting



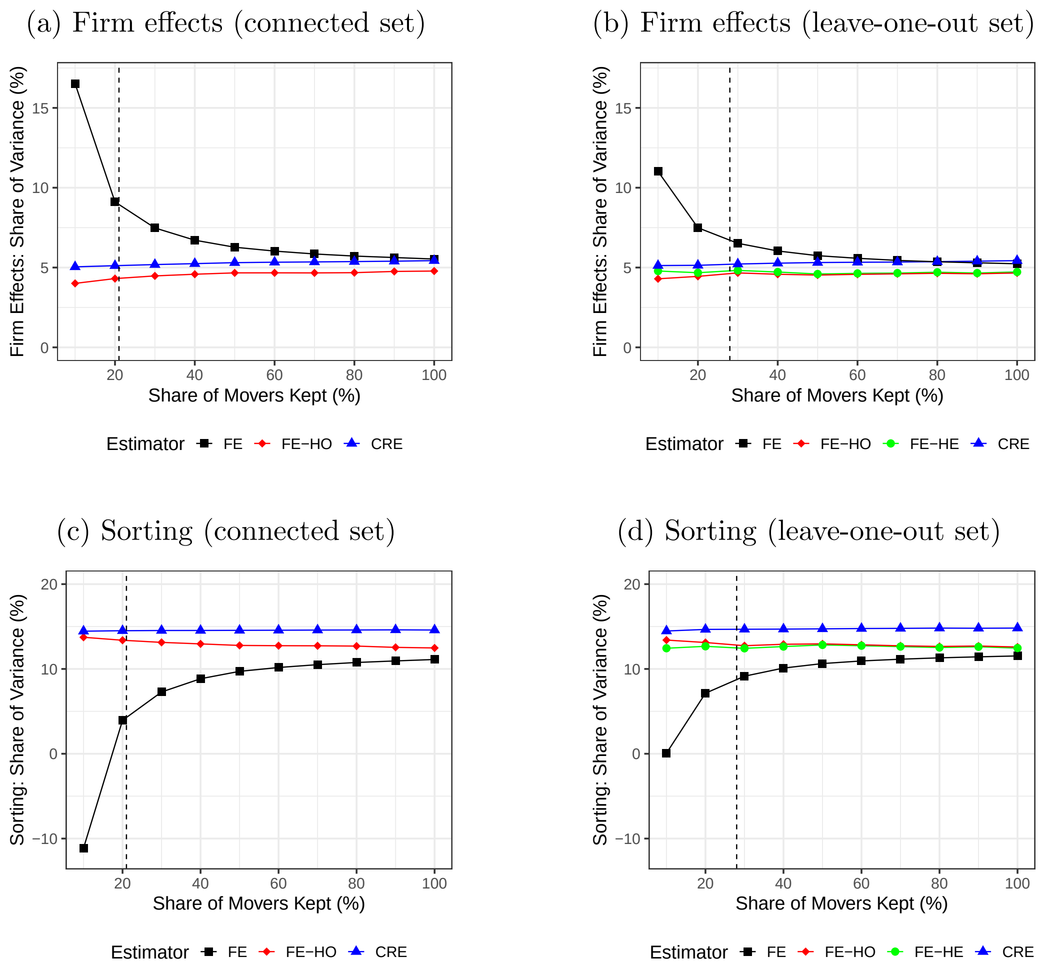
Notes: In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We compare estimates using the baseline definition of movers and the strict definition of movers defined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

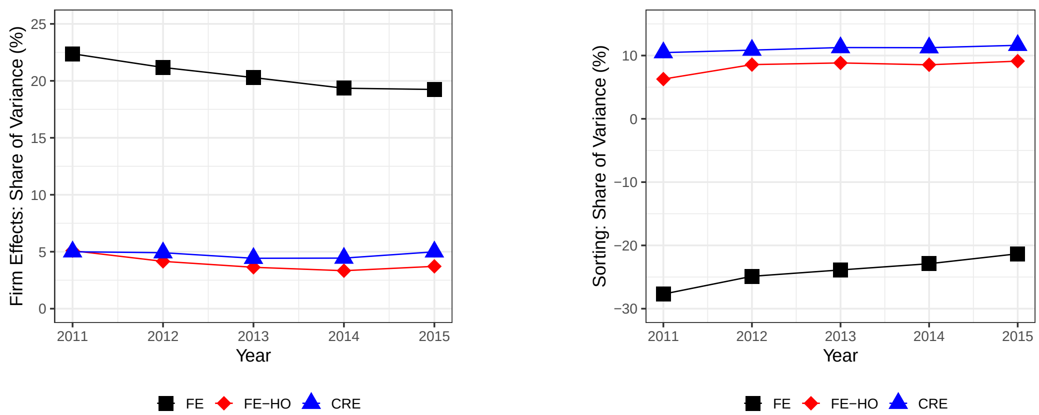
Figure F5: Evidence on Limited Mobility Bias in the United States



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)

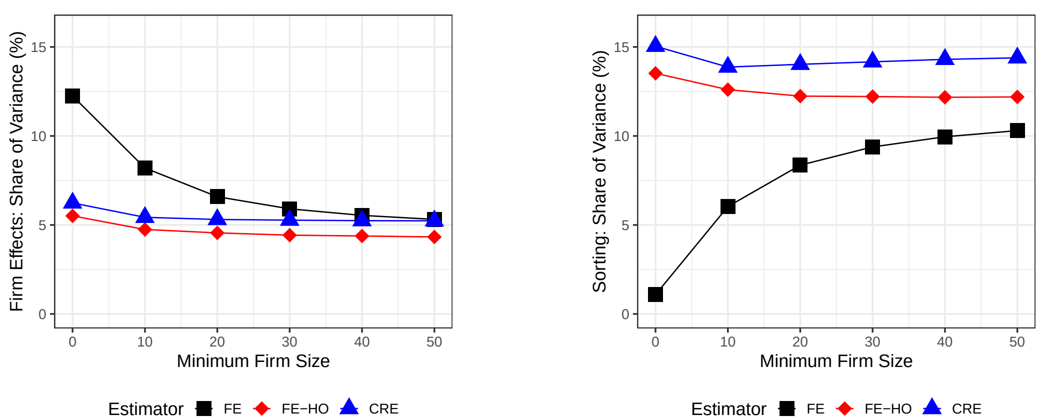
(a) Firm Effects (b) Sorting



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)

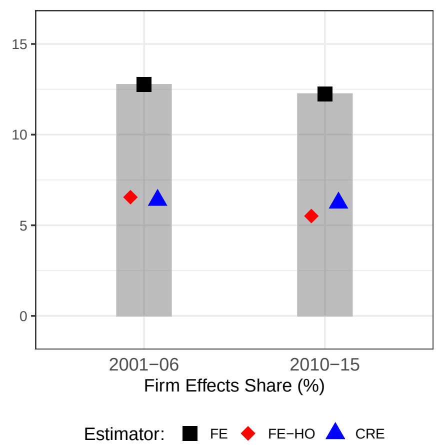
(a) Firm Effects (b) Sorting



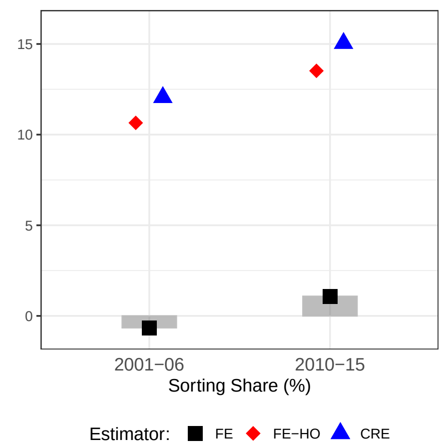
Notes: In this figure, we provide FE, FE-HO, and CRE estimates of the contribution to earnings and wage inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We restrict the sample to firms with at least the number of workers indicated on the x-axis. We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

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

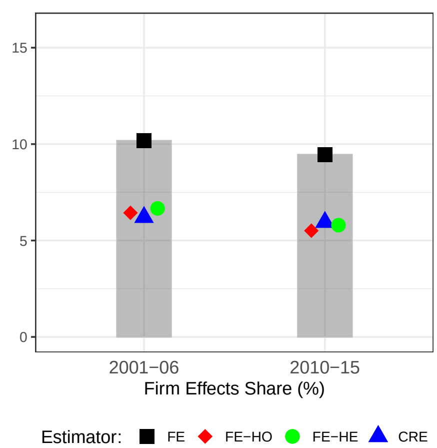
(a) Firm Effects (Connected Set)



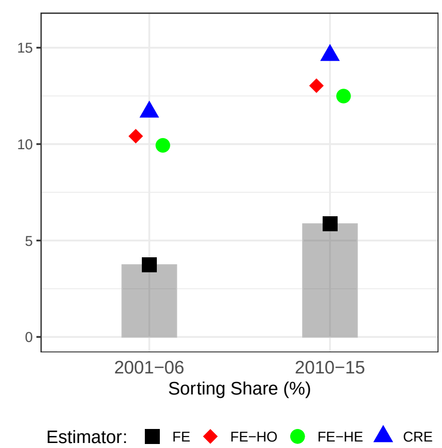
(b) Sorting (Connected Set)



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

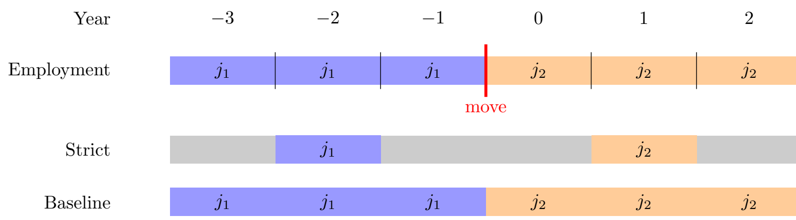


(d) Sorting (Leave-one-out Set)



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.

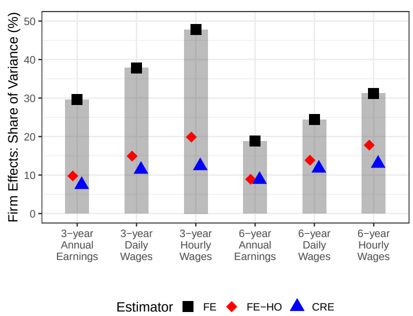
Figure F9: Visualizing Alternative Mover Definitions for the US



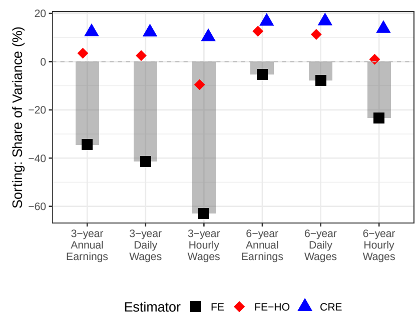
Notes: In this figure, we provide a diagram to help visualize the difference between the main definition of a mover ("Baseline") and the mover definition that uses only intermediate years within spells ("Strict").

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

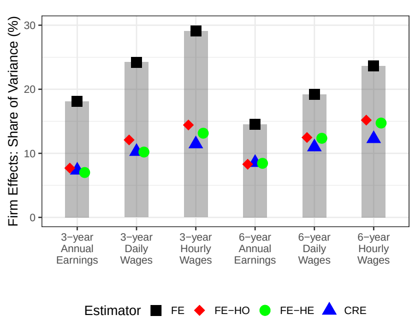
(a) Firm effects (connected set)



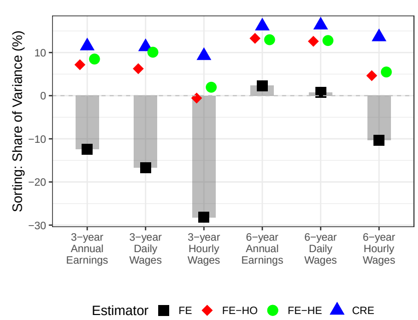
(b) Sorting (connected set)



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

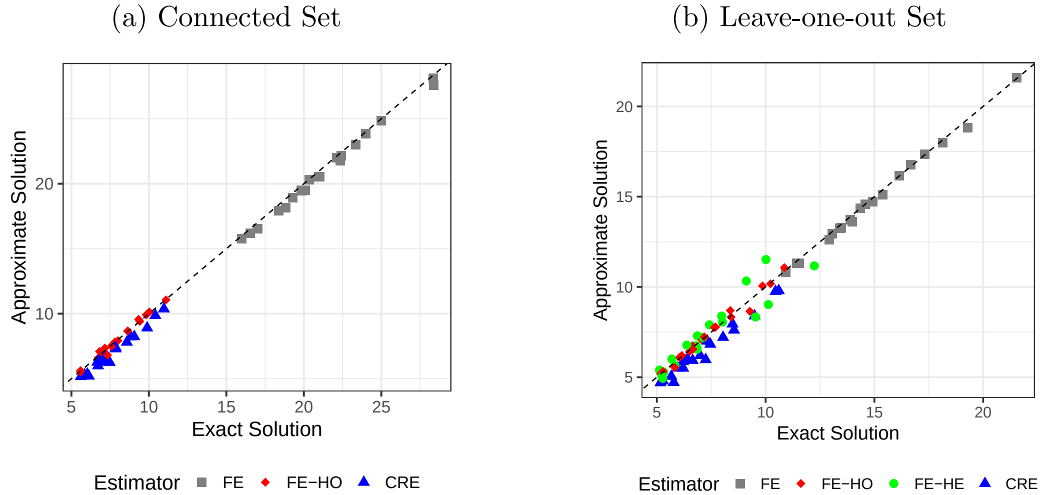


(d) Sorting (leave-one-out set)



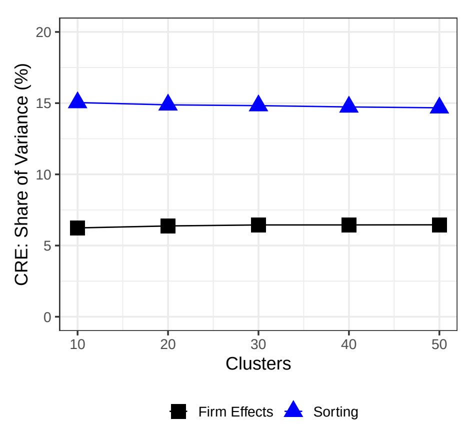
Notes: In this figure, we provide FE, FE-HO, FE-HE, and CRE estimates of the contribution to earnings or wage inequality of firm effects (Subfigures a and c) and the sorting of workers to firms (Subfigures b and d) in Norway. We consider the connected set of firms (Subfigures a and b) and the leave-one-out 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



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.

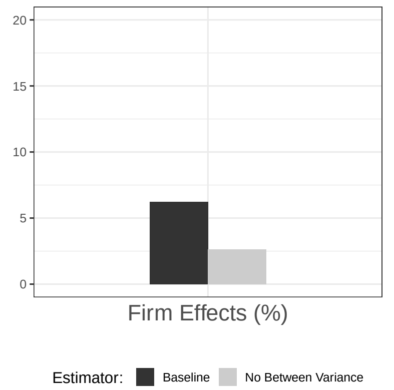
Figure F12: Number of Groups for CRE Estimates in the US (Connected Set)



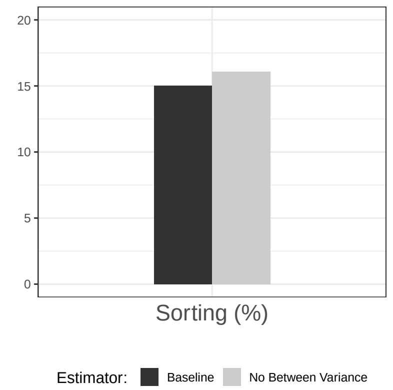
Notes: In this figure, we provide CRE estimates of the contribution to earnings inequality of firm effects and the sorting of workers to firms in the US. We consider the connected set of firms, and vary the number of firm groups considered in the CRE estimation procedure (indicated on the x-axis).

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

(a) Firm Effects



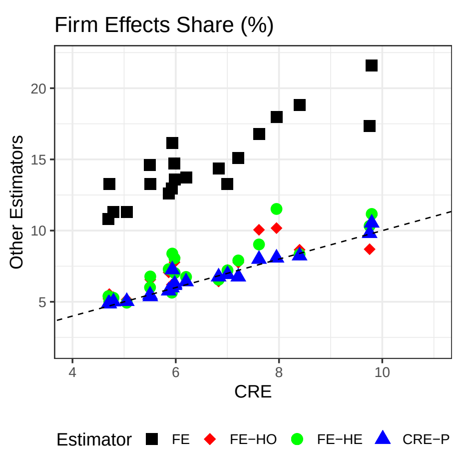
(b) Sorting



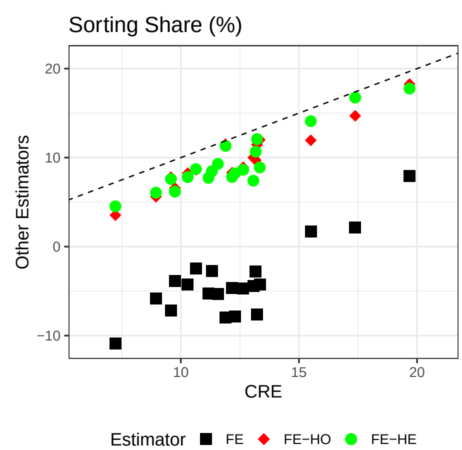
Notes: In this figure, we provide CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the US. We compare the baseline CRE estimates to the posterior estimates for a random-effects specification that does not condition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

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

(a) Firm effects



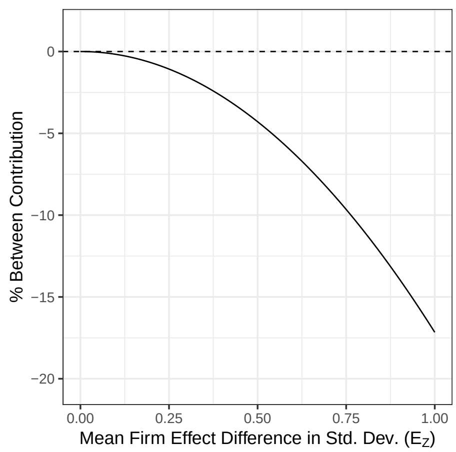
(b) Sorting



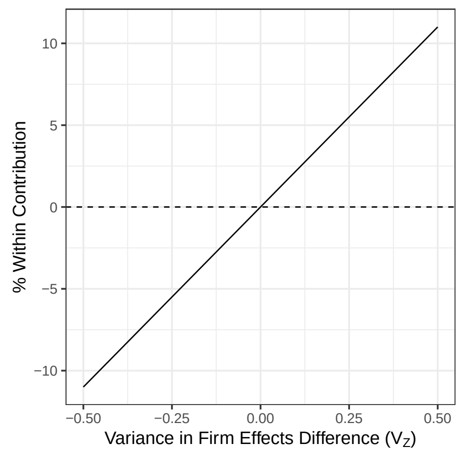
Notes: In this figure, we provide FE, FE-HO, FE-HE, and CRE estimates of the contribution to earnings inequality of firm effects (Subfigure a) and the sorting of workers to firms (Subfigure b) in the 20 smallest US states. We consider the leave-one-out set of firms within each state. CRE estimates are 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 $\pi = 0.78$)

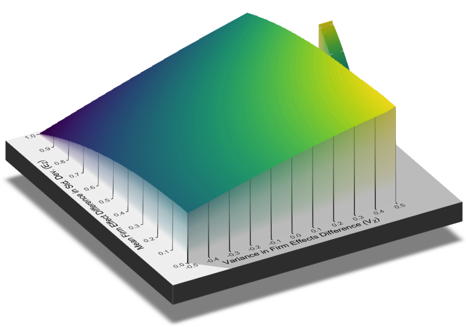
(a) Between Contribution to Bias



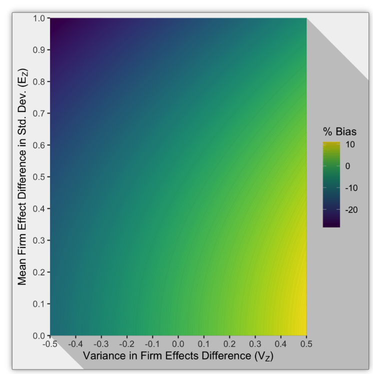
(b) Within Contribution to Bias



(c) Total Bias (side view)



(d) Total Bias (overhead view)

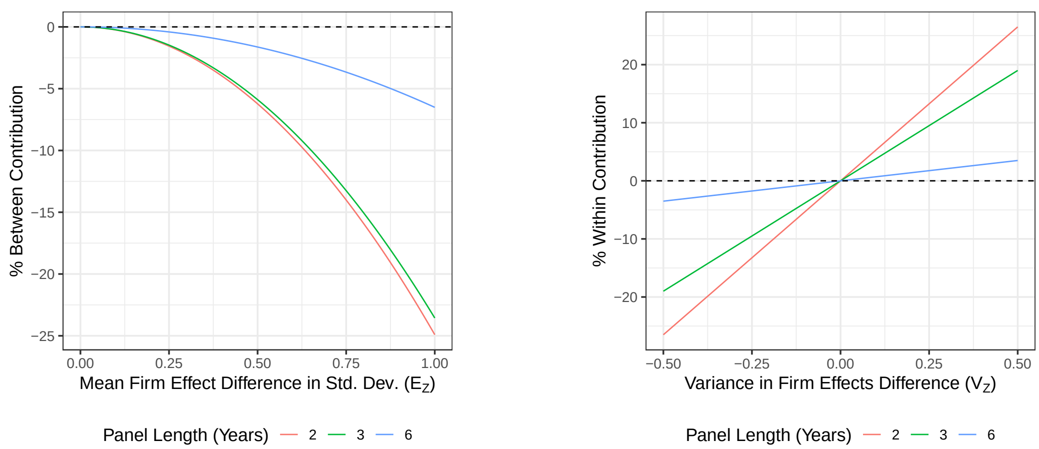


Notes: In this figure, we use equation (D11) to visualize the bias that arises from using the variance of firm premiums estimated for a subsample of firms to approximate the variance of firm effects in the full population. We calibrate $\pi = 0.78$, which corresponds to the 20 workers per firm sample 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 π based on panel length)

(a) Between Contribution to Bias

(b) Within Contribution to Bias



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 π 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					Firm Effects: $Var(\psi)$				Sorting: $2 \times Cov(\alpha, \psi)$			
	T	Set	Firms	Workers	Movers	Total Var	Btw Var	FE	FE-HO	FE-HE	CRE	FE	FE-HO	FE-HE	CRE
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%	-	12.8%	-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%	-	11.7%	-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%	-	11.4%	-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%
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%	-	5.2%	-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%

Notes: In this table, we provide FE, FE-HO, FE-HE, and CRE estimates for the main samples considered in the paper. “Total Var” refers to the total variance in log wages or earnings, while “Btw Var” refers to the share of variance that is between firms.

Table F3: Sample comparison for the US

US Samples:	T	Set	Firms	Workers	Movers	Total Var	Firm Effects	Sorting
Baseline Sample	6	Connected	2,568	55,464	14,888	0.414	12.2%	1.1%
Varying the earnings threshold:								
Earnings $\geq$ \$11,250	6	Connected	2,886	60,916	18,077	0.485	12.8%	3.1%
Earnings $\geq$ \$7,500	6	Connected	3,201	66,474	22,140	0.601	13.7%	5.1%
Earnings $\geq$ \$3,750	6	Connected	3,489	72,228	27,325	0.823	14.6%	6.4%
Varying the minimum firm size:								
Firm Size $\geq$10	6	Connected	748	48,810	10,704	0.425	8.2%	6.0%
Firm Size $\geq$20	6	Connected	360	43,140	8,361	0.434	6.6%	8.4%
Firm Size $\geq$30	6	Connected	227	39,581	7,015	0.440	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					Firm Effects: $Var(\psi)$			Sorting: $2 \times Cov(\alpha, \psi)$		
Italy Samples:	T	Set	Firms	Workers	Movers	Total Var	Btw Var	FE	FE-HO	FE-HE	FE	FE-HO	FE-HE
Baseline Sample	6	Leave-out	61	1,034	346	0.168	44.8%	19.3%	15.8%	15.7%	4.7%	11.1%	11.2%
Baseline Sample, Select Years	3	Leave-out	30	755	121	0.181	43.5%	18.5%	14.6%	10.9%	1.3%	8.8%	16.1%
Kline et al. (2020) Sample Construction	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.

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.

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.

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.

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, 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.*

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, 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.

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

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

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.

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.

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

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.