Limited Mobility Bias: More Corrections But No More Costs*

Miren Azkarate-Askasua† Miguel Zerecero‡

February, 2024

Abstract

Variance components of linear models with many covariates exhibit small-sample bias. This is known as limited mobility bias in the context of variance decompositions in AKM regressions. The direct computation for a bias correction is not feasible when the number of covariates is large. We propose a bootstrap method for correcting this bias that accommodates general heteroskedasticity and serial correlation of the errors. Our approach is suited to correct variance decompositions and the bias of multiple quadratic forms of the same linear model without increasing the computational cost. We show with Monte Carlo simulations that our bootstrap procedure is effective in correcting the bias and find that is faster than other methods in the literature. Using administrative data for France, we correct variance decompositions overall and per groups such as commuting zones. We find that the correlation between worker and firm fixed effects becomes positive depending on the specification and that the positive correlation between the covariance of worker and firm effects per commuting zone and its population is stable to the corrections.

JEL Codes: C13, C23, C55, J30, J31
Keywords: Limited mobility bias, bias correction, variance components, many regressors, fixed effects.

*This paper previously circulated as *Correcting Small Sample Bias in Linear Models with Many Covariates*. We thank our advisor Christian Hellwig for his guidance throughout this project. We thank Fabrice Collard, Thomas Crossley, Patrick Fève, Simen Gaure, Silvia Goncalves, Cristina Gualdani, Koen Jochmans, Elia Lapenta, Tim Lee, Thierry Magnac, Nour Meddahi, Jean-Marc Robin and Uta Schönberg for helpful comments. We gratefully acknowledge financial support from the Toulouse School of Economics to get access to the administrative data provided by CASD. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'avenir' program (reference: ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD). Zerecero acknowledges the financial support from CONACYT (reference: 329103/383874). Azkarate-Askasua acknowledges the support from the German Research Foundation (through the CRC-TR-224 project B06). All errors are our own. First version: May 2018.

†azkarate-askasua@uni-mannheim.de; University of Mannheim
‡mzerecer@uci.edu; University of California, Irvine

1 Introduction

Researchers have used employer-employee matched datasets to study the sorting patterns of workers into firms. Various papers have estimated a linear model of log wages with person and firm fixed effects, following the seminal work of Abowd, Kramarz, and Margolis (1999) (AKM henceforth). These studies compute the correlation between the person and firm fixed effects to determine the degree of sorting in the labor market. Most studies have found zero or negative correlations, casting doubt on whether there is sorting in the labor market. However, as noted by Abowd, Kramarz, Lengermann, and Pérez-Duarte (2004) this correlation is likely to suffer from small-sample bias, dubbed *limited mobility* bias in their paper. Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2023) show that the limited mobility bias is substantial when performing a variance decomposition of log wages for several countries.

The elements of a variance decomposition of a linear model are quadratic objects in the parameters. As long as the parameters are estimated with noise, these quadratic objects are subject to small-sample bias. This bias can be substantial in typical applications and can even change the sign of estimated covariances. Moreover, in most applications this bias does not fade away by increasing the sample size. This is the case when using panel data, as the number of parameters to estimate grows with the sample size.

Andrews, Gill, Schank, and Upward (2008) derive formulas for correcting the bias when the errors are homoskedastic. Gaure (2014) provides formulas for more general variance structures. Unfortunately, the direct implementation of these corrections in high dimensional models is infeasible. The reason is that the corrections entail computing the inverse of an impractically large matrix, which has prevented the direct application of the correction formulas.[1]

In this paper we propose a bootstrap method to correct limited mobility bias of variance components. The main advantage of our bootstrap method is that it allows the computation of many corrections at the same time without increasing the computational cost. On top of being scalable in the number of corrections, out method is easy to implement, fast, and it accommodates different assumptions on the error structure.

To illustrate the advantages of our method, consider a researcher who is interested in understanding how much the different components of an AKM model explain the variance of log wages for different subgroups of the population. This can be done, for example, by estimating doing separate variance decompositions for workers by race and gender (Gerard, Lagos, Severnini, and Card, 2021), or by city (Dauth, Findeisen, Moretti, and Suedekum, 2022). The computational cost of correcting for the variance components with other alternative methods scales linearly with the number of subgroups. The increasing cost has prevented researchers from analyzing variance components at increasingly finer partitions of the data.[2] Our method overcomes this limitation. The computational cost of doing an arbitrary number of corrections with our method is practically the same cost of doing one correction.

Our method consists on re-estimating the same quadratic forms of a linear model on boot-

[1] Some examples of papers doing a variance decomposition of log wages into worker and firm fixed effects without correcting the limited mobility bias are: Song, Price, Guvenen, Bloom, and Von Wachter (2019), Sorkin (2018), Card, Cardoso, Heining, and Kline (2018), Alvarez, Benguria, Engbom, and Moser (2018) (who focus on changes over time and assume the bias is constant) and Arellano-Bover and San (2023).

[2] One could think of conditioning for different occupations, industries, commuting zones, education groups etc. in the AKM model.

strapped data. The sample means of these quadratic forms are our bias correction terms. Using Monte Carlo simulations we show that our method successfully corrects the bias of quadratic forms for multiple assumptions on the variance structure of the error term, such as heteroskedasticity, serial correlation or clustering. In practice, under the assumption of a diagonal covariance matrix, we use a wild bootstrap. When the covariance matrix is assumed non-diagonal, we use a wild block bootstrap (Cameron, Gelbach, and Miller, 2008) that is valid for unrestricted dependence of the error terms within group. The wild block bootstrap is flexible in the definition of the group and therefore allows, for example, the clustering of the errors depending on geographical area or serial correlation within the worker-firm match.[3]

Our bootstrap approach is similar to the ones proposed by Gaure (2014) and by Kline, Saggio, and Sølvsten (2020). The bias is equal to the trace of a matrix. When the number of covariates of the linear model is large, the explicit computation of this trace is not feasible. Like ours, both of their methods rely on iterative procedures to compute an estimate of the trace term. Gaure exploits the fact that the trace can be represented as the expectation of a more manageable quadratic form in a random vector. This expectation can in turn be approximated by estimating a sample mean after simulating different random vectors.[4]

Kline, Saggio, and Sølvsten (2020) (KSS henceforth) follow a similar approach to Gaure (2014). In their large-scale computation procedure, they estimate the trace term leading to the bias and implement a bias correction assuming either heteroskedasticity or serial correlation of the errors. An important point of their paper is that their leave-one-out covariance-matrix estimate is unbiased. Our approach differs in the way we estimate the trace term which allows us to be more flexible and to increase the number of corrections without increasing the computational cost. We adapt our method to use their unbiased estimator of the variance of the error terms.

The computational cost in Gaure and KSS comes from estimating a bias correction for each interested quadratic form, as it requires solving a large system of linear equations in each iteration that are particular to each quadratic form. In contrast, we re-estimate the model with bootstrapped data and show that a sample mean of the *bootstrapped* moment estimates is an unbiased and consistent estimator of the direct bias correction term. In our method, the computational cost comes from estimating the linear model in each bootstrap but does not increase depending on the number of moments to correct. We need to solve from one to two systems of linear equations per bootstrap regardless of the number of moments to correct, while with the Gaure and KSS methods, one needs to solve as many systems of equations per iteration as needed corrections.[5] They implement corrections of the second order moments of the two leading fixed effects while we can directly perform a full variance decomposition, which is therefore suited for corrections on multi-way fixed effect regressions.

Our method is easy to implement as it only requires estimating linear models. While of course

[3]Other examples include errors correlated within firms, workers or occupations.

[4]In particular, the way Gaure estimates the trace is known as the Hutchinson method. Denote a random vector $x \in \mathbb{R}^n$, where each individual entry is independently distributed Rademacher (entries can take values of 1 or -1 with probability 1/2). Then, for a square matrix $A \in \mathbb{R}^{n \times n}$ we have that $tr(A) = \mathbb{E}(x'Ax)$. The Hutchinson estimator of the trace of matrix A is $\frac{1}{M}\sum_{i=1}^{M} x_i' A x_i$, where x_i is the i-th draw of the random vector x; see Hutchinson (1989) and Avron and Toledo (2011). Gaure's R package *lfe* implements the correction when the error terms are assumed homoskedastic. The function applying the correction is *bccorr*; see Gaure (2013). Gaure (2014) sketches the procedure to correct for the bias when the error terms are heteroskedastic, but to the best of our knowledge he does not implement it in his R package.

[5]Consider the linear model $y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + \varepsilon_t$ where one is interested in doing a variance decomposition for each period t. This would yield three quadratic objects to correct ($Var(X_1\widehat{\beta}_1)$, $Var(X_2\widehat{\beta}_2)$, $Cov(X_1\widehat{\beta}_1, X_2\widehat{\beta}_2)$) *per period*.

this requires solving a system of linear equations, like in the case of Gaure and KSS, these systems are more ubiquitous. Therefore, there is a wide range of algorithms that estimate linear models for different softwares. We provide codes in Matlab, but the user can easily implement the correction method by taking profit of other algorithms in alternative softwares.

We apply our method to French administrative data and perform a variance decomposition of an estimated AKM type model. Consistent with the Andrews et al. (2008) formulation, we find that sample variances of person and firm effects are reduced and their covariance increased after the correction. The estimated correlation at the connected set passes from -0.10 to almost zero under the assumption of serial correlation of the error terms within the match.[6] We show the usefulness of our method by correcting variance decompositions per commuting zone and per commuting zone-occupation combinations. In the latter case, that entails doing about 4000 corrections. Performing decompositions per commuting zone, we find that the positive gradient between the correlation of worker and firm fixed effects and population is robust flattened after correcting second order moments.

Labor economists have been aware of the small-sample bias problem with quadratic forms in the parameters and the difficulty in estimating a correction at least since Andrews et al. (2008). There have been several attempts to correct this bias when performing variance decompositions of estimated linear models. Some methods are based on variations of the jacknife, such as the split-panel jacknife estimator by Dhaene and Jochmans (2015) or the leave-one-out estimator by KSS mentioned above. Bonhomme, Lamadon, and Manresa (2019) relax the exogenous mobility assumption from the AKM model and mitigate the small-sample bias by reducing the dimensionality of the estimated parameters. Borovičková and Shimer (2017) propose an alternative method to estimate the correlation between worker and firm types.

2 The Bias

For clarity of exposition we layout the source of the bias. Consider the following linear model:

$$Y = X\beta + u, \tag{1}$$

where Y is a $n \times 1$ vector representing the endogenous variable, X is a matrix of covariates of size $n \times k$, and β is a vector of parameters.[7] The error term u satisfies mean independence $\mathbb{E}(u|X) = 0$. The OLS estimate of β is,

$$\widehat{\beta} = \beta + Qu,$$

where $Q = (X'X)^{-1} X'$.

We are interested in estimating the following quadratic form $\varphi = \beta' A \beta$ for some matrix A of dimensions $k \times k$, where $\mathbb{E}(A|X) = A$. From the expression for $\widehat{\beta}$ we can decompose the plug-in estimator $\widehat{\varphi}_{PI} = \widehat{\beta}' A \widehat{\beta}$ as,

$$\widehat{\varphi}_{PI} = \beta' A \beta + u' Q' A Q u + 2u' Q' A \beta. \tag{2}$$

[6] Abowd et al. (2004), also using French data but a different sample, found a correlation of -0.28.

[7] We follow loosely the notation in Kline et al. (2020) for the interested reader to compare the papers.

Using the general formula for the expectation of quadratic forms, the exclusion restriction $\mathbb{E}(u|X) = 0$, and $\mathbb{E}(A|X) = A$ we obtain,[8]

$$\mathbb{E}\left(\widehat{\varphi}_{PI}|X\right) = \beta' A \beta + \text{trace}\left(Q' A Q \mathbb{V}(u|X)\right) = \varphi + \delta, \tag{3}$$

where the bias $\delta \equiv \text{trace}\left(Q' A Q \mathbb{V}(u|X)\right)$ comes from the fact that $\widehat{\beta}$ is estimated with noise.

To get a bias correction one needs an estimate of the trace term δ. One option is to just plug-in the estimate for the conditional covariance matrix $\widehat{\mathbb{V}}(u|X)$. We define $\widehat{\delta}$ as the direct bias correction term:

$$\widehat{\delta} \equiv \text{trace}\left(Q' A Q \widehat{\mathbb{V}}(u|X)\right). \tag{4}$$

Computing $\widehat{\delta}$ is difficult when the number of covariates is large because it requires to calculate first the matrix Q, which is itself a function of the inverse of a very large matrix.[9] In the next section we propose a methodology to apply a computationally feasible correction.

We define the following bias-corrected estimate of the quadratic form φ as:

$$\widehat{\varphi} = \widehat{\beta}' A \widehat{\beta} - \widehat{\delta}.$$

As long as $\mathbb{E}(\widehat{\delta}|X) = \delta$, then it follows that $\mathbb{E}(\widehat{\varphi}|X) = \varphi$.

Proposition 1. *The direct bias correction $\widehat{\delta}$ is an unbiased estimate of the bias term δ if and only if $\widehat{\mathbb{V}}(u|X)$ is an unbiased estimator of $\mathbb{V}(u|X)$.*

Thus, it is necessary to have an unbiased estimate of the covariance matrix $\mathbb{V}(u|X)$ to have an unbiased estimate of the quadratic form φ.

2.1 Components of a variance decomposition as quadratic objects

When performing a variance decomposition of a linear model, one can think of each element as a particular form of $\widehat{\beta}' A \widehat{\beta}$ with the appropriate choice of A. To see this, we can rewrite (1) as:

$$Y = X_1 \beta_1 + X_2 \beta_2 + u, \tag{5}$$

where X_1 and X_2 are matrices of covariates of size $n \times k_1$ and $n \times k_2$, $k = k_1 + k_2$ with $X = [X_1\ X_2]$ and $\beta' = [\beta_1'\ \beta_2']$.

We are interested in the sample variances $(\widehat{\text{var}}(X_1\beta_1), \widehat{\text{var}}(X_2\beta_2))$ and covariance $(\widehat{\text{cov}}(X_1\beta_1, X_2\beta_2))$, denoted, respectively, as σ_1^2, σ_2^2 and σ_{12}.[10] Define $\mathbf{1}$ as a vector of ones with appropriate length. Then, denote the demeaning operator as $M_1 = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$. We can then write the sample variances and covariances in matrix notation as:

$$\sigma_j^2 = \beta' A_j \beta, \quad \text{for } j = \{1,2\} \text{ and}$$

$$\sigma_{12} = \beta' A_{12} \beta,$$

[8] Given a random vector x and a symmetric matrix B we have that $\mathbb{E}(x'Bx) = \mathbb{E}(x')B\mathbb{E}(x) + \text{trace}(B\mathbb{V}(x))$.

[9] The dimension of this matrix is related to the number of covariates that are estimated in the linear model. In a typical AKM type model the data will comprise of hundreds of thousands or millions of workers and tens of thousands of firms, each representing a covariate in the model.

[10] The sample variance for a vector $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ is $\widehat{\text{var}}(\mathbf{x}) = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \overline{\mathbf{x}})^2$, where $\overline{\mathbf{x}}$ is the sample mean. Similarly, the sample covariance for vectors $\mathbf{x}$ and $\mathbf{y}$ is $\widehat{\text{cov}}(\mathbf{x}, \mathbf{y}) = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \overline{\mathbf{x}})(y_i - \overline{\mathbf{y}})$.

where the symmetric matrices A_1, A_2 and A_{12} are equal to:

$$A_1 = \frac{1}{n-1}\begin{pmatrix} X_1'M_1X_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \qquad A_2 = \frac{1}{n-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & X_2'M_1X_2 \end{pmatrix}, \qquad A_{12} = \frac{1}{2(n-1)}\begin{pmatrix} \mathbf{0} & X_1'M_1X_2 \\ X_2'M_1X_1 & \mathbf{0} \end{pmatrix}.$$

The plug-in estimators of σ_1^2, σ_2^2 and σ_{12}, obtained by substituting β with the OLS estimate $\widehat{\beta}$, are just particular examples of $\widehat{\varphi}_{PI}$. Therefore, these estimates are biased.

3 Bootstrap Correction

The bootstrap correction estimates the direct bias correction (4) by replicating the bias structure of the plug-in estimates (2). In this section we present the bootstrap correction and discuss different implementations depending on the choice of the covariance matrix estimate.

Suppose that we have the residuals of our original regression $\widehat{u} = Y - X\widehat{\beta}$. Using these residuals we can construct an estimate of the covariance matrix, $\widehat{\mathbb{V}}(u|X)$. We generate a new dependent variable for the bootstrap Y^* as:

$$Y^* = v^*,$$

where v^* is a vector containing the bootstrapped residuals. This is equivalent to performing a linear regression on bootstrapped data, while setting $\widehat{\beta} = \mathbf{0}$. The construction of v^* will depend ultimately in the assumption that we are making about the error term. In particular, we need that the variance of the bootstrapped errors $\mathbb{V}(v^*|X,u)$ to be equal to $\widehat{\mathbb{V}}(u|X)$. The following proposition states the main result of the paper and all the proofs are left to the Appendix.

Proposition 2. *Suppose the regression model (1) is correctly specified. Let p denote the number of bootstraps and $v^{*(j)}$ the vector of bootstrapped residuals for the j-th bootstrap iteration. Define $\beta^{*(j)}$ as the OLS estimate of regressing $v^{*(j)}$ over X for the j-th bootstrap iteration. If the conditional variance-covariance matrix of the bootstrapped residuals $\mathbb{V}(v^{*(j)}|X,u)$ is equal to $\widehat{\mathbb{V}}(u|X)$, and $\mathbb{E}(v^{*(j)}|X,u) = 0$, then*

$$\delta^* \equiv \frac{1}{p}\sum_{j=1}^{p} \beta^{*(j)\prime} A \beta^{*(j)}$$

is an unbiased and consistent estimator of the direct bias correction $\widehat{\delta}$.

The proposition tells us that instead of computing directly the direct bias correction term $\widehat{\delta}$, we can estimate it using a sample average of estimated quadratic forms.

The intuition behind our bias estimator is that in every bootstrap iteration we are replicating the source of the bias, which is the noise embedded in the estimated parameters. The computational burden of our method comes from estimating $\beta^{*(j)}$ for each bootstrap.[11] The main advantage of our method is that we can correct several moments simultaneously, without increasing the computational time. Therefore, our method is easily scalable to the estimation of many moments. Assume we are interested in doing a variance decomposition per groups, for example for each city or commuting zone as Dauth et al. (2022). Then, we would need to do a correction for the variances and the covariances for *every* group, for example city or commuting zone, but do the bootstrap only

[11]Current softwares avoid the inversion of the $X'X$ matrix to estimate linear models and are therefore able to estimate linear models even when the number of covariates is large.

one time.[12] Other advantages of our method are its flexibility and implementability. We further discuss them after presenting the implementation of the method for different assumptions on the covariance matrix structure of the error terms.

The key for the bootstrap correction to work is that $\mathbb{V}(v^*|X,u)$ is equal to the sample variance-covariance matrix $\widehat{\mathbb{V}}(u|X)$, so the bootstrap correction δ^* is an unbiased and consistent estimator of the direct bias correction term $\widehat{\delta}$. Therefore, the bootstrap procedure has to be compatible with the underlying assumption on the structure of the error term.

The small sample properties of the bootstrap estimate δ^* would depend ultimately on the choice of estimate for the covariance matrix $\mathbb{V}(u|X)$. In particular, we have the following corollary for the bias which is just a consequence of Propositions 1 and 2.

Corollary 1. *Conditioning on X, if $\widehat{\mathbb{V}}(u|X)$ is an unbiased estimator of $\mathbb{V}(u|X)$, then the bootstrap correction δ^* is an unbiased estimator of the bias δ.*

In what follows we provide examples for some popular choices for estimators of the covariance matrix and how to implement the bootstrap correction. We complement the examples in the Online Appendix for non-block-diagonal covariance matrices.

Example 1: Homoskedasticity. When the errors are homoskedastic, we can use the well-known unbiased estimate of the covariance matrix $\widehat{\sigma}^2\mathbf{I}$, where $\widehat{\sigma}^2 = n/(n-k)\sum_{i=1}^{n}\widehat{u}_i^2$.[13] A suitable bootstrap could be a residual bootstrap with a degrees of freedom correction. This would mean resampling with replacement the estimated residuals and multiplying them by $\sqrt{n/(n-k)}$. Thus the variance of the bootstrapped errors would be equal to the estimated covariance-matrix $\widehat{\sigma}^2\mathbf{I}$. Another possibility could be to simulate errors from a normal distribution with zero mean and variance $\widehat{\sigma}^2$. In the case of homoskedastic errors, the proposed bootstraps can replicate the variance of an unbiased estimate of the covariance matrix. Thus, the bootstrap bias correction δ^* is an unbiased estimate of the bias term δ; see Corollary 1.

Example 2. Heteroskedasticity I. Assume the covariance matrix is diagonal, with non-zero ith diagonal element equal to ψ_i. Let $\widehat{\psi}_i$ be the estimate of the variance for the ith observation error term. MacKinnon and White (1985) explore different consistent variance estimates $\widehat{\psi}_i$. These include:

$$HC_0 : \widehat{\psi}_i = \widehat{u}_i^2, \quad HC_1 : \widehat{\psi}_i = \frac{n}{n-k}\widehat{u}_i^2 \quad \text{and} \quad HC_2 : \widehat{\psi}_i = \frac{\widehat{u}_i^2}{1-h_{ii}},$$

where h_{ii} is the ith diagonal element of the projection matrix $H = X\left(X'X\right)^{-1}X'$. The term h_{ii} is sometimes known as the *leverage* of observation i, because, as explained by Angrist and Pischke (2008), it tell us how much *pull* a particular observation exerts over the regression line.

A suitable bootstrap for the different covariance matrix estimators is the Wild bootstrap. In our exercises below, we implement this bootstrap by first generating i.i.d. Rademacher random

[12] Dauth et al. (2022) correct their variance component decompositions city-by-city to implement the KSS correction. Beyond the increase in computational time, other drawbacks are that one cannot compare the estimates across cities due to the normalization and that the set of firms in the per-city connected set might be different as it excludes firm effects identified by across-city mobility. As they note in footnote 21 *"This implies that the plant effects are now only identified from within-city mobility. Their first moments are therefore not comparable across cities. This prevents us from using this procedure as our baseline model. The procedure is computationally very demanding, both in terms of speed and memory"*.

[13] The origin of the bias is again a trace term that under homoskedasticity is equal to $n-k$. For a textbook explanation see Proposition 1.2. in Hayashi (2000).

variables that take values of 1 or -1 with probability $1/2$. Then we multiply $\sqrt{\widehat{\psi}_i}$ to the ith Rademacher entry r_i. This would constitute the ith bootstrapped residual. For example, if we use HC_0, the bootstrap residual would be $v_i^* = \widehat{u}r_i$. The Online Appendix outlines the algorithms to implement this procedure.

When using the HC_2 estimates, we need first to calculate the leverage h_{ii}. When the number of covariates is large, a direct computation of the leverage is unfeasible. In the Online Appendix we show how to estimate this leverages by means of averaging the squared fitted values of linear regressions. We also provide a diagnostic and correction method to ensure that the estimated leverages are bounded above by 1.

In general, the three alternatives of covariance matrix estimates (HC_0, HC_1 and HC_2) are biased.[14] For example, for HC_0 we have:

$$\mathbb{E}(\widehat{u}_i^2|X) = \psi_i - 2\psi_i h_{ii} + h_i' \mathbb{V}(u|X) h_i,$$

where h_i is the ith column of the projection matrix H.[15] Thus, while δ^* is an unbiased estimate of $\widehat{\delta}$ (Proposition 2), it would be biased with respect to δ (Proposition 1).

Example 3. Heteroskedasticity II: Unbiased Estimator. Recently, Kline et al. (2020) and Jochmans (2018) have proposed the following unbiased estimator of the conditional variance of the ith error term:[16]

$$HC_U : \widehat{\psi}_i = \frac{Y_i \widehat{u}_i}{1 - h_{ii}}. \tag{6}$$

In practice, we sometimes obtain negative estimates of the conditional variance when using HC_U. Therefore, taking the square root of $\widehat{\psi}_i$, as in Example 2, gives imaginary numbers. As a result, the bootstrap residual vector v^*—which is equal to the left-hand-side bootstrap variable Y^*—may contain complex numbers. Thus, our bootstrap correction procedure needs to be adapted to account for the presence of complex numbers.

Let $Y^* \equiv Y_{\mathbb{R}}^* + \mathbf{i}Y_{\mathbb{Z}}^*$, where $Y_{\mathbb{R}}^*$ and $Y_{\mathbb{Z}}^*$ represent the real and imaginary parts of Y^*, and $\mathbf{i} \equiv \sqrt{-1}$. The resulting OLS estimator β^* is also a complex number given by the standard OLS formula:

$$\beta^* = \left(X'X\right)^{-1} X'Y^* = \left(X'X\right)^{-1} X'Y_{\mathbb{R}}^* + \mathbf{i}\left(X'X\right)^{-1} X'Y_{\mathbb{Z}}^* = \beta_{\mathbb{R}}^* + \mathbf{i}\beta_{\mathbb{Z}}^*,$$

where $\beta_{\mathbb{R}}^*$ and $\beta_{\mathbb{Z}}^*$ are the real and imaginary parts of β^*.

For every bootstrap iteration, we can then estimate β^* and obtain the quadratic object:

$$\begin{aligned}\beta^{*\prime} A \beta^* &= \left(\beta_{\mathbb{R}}^{*\prime} + \mathbf{i}\beta_{\mathbb{Z}}^{*\prime}\right) A \left(\beta_{\mathbb{R}}^* + \mathbf{i}\beta_{\mathbb{Z}}^*\right) \\ &= \underbrace{\beta_{\mathbb{R}}^{*\prime} A \beta_{\mathbb{R}}^* - \beta_{\mathbb{Z}}^{*\prime} A \beta_{\mathbb{Z}}^*}_{\text{Real part}} + \mathbf{i}\,\underbrace{2\beta_{\mathbb{R}}^{*\prime} A \beta_{\mathbb{Z}}^*}_{\text{Imaginary part}}.\end{aligned}$$

This quadratic object is also complex number. It turns out that we can construct a bias correction by taking the average of only the real part of the above quadratic object. This is because, in expectation, the imaginary part of the quadratic object is equal to zero. As we show in Proposition 3 below, the

[14] A particular case where the estimate is unbiased is when using HC_1 and the error terms are homoskedastic.
[15] A textbook exposition of these issues can be found in Chapter 8 of Angrist and Pischke (2008).
[16] See page 1,862 of Kline et al. (2020).

real part of the quadratic object is an unbiased and consistent estimator of the direct bias correction when using HC_U to estimate the covariance matrix. But before introducing the Proposition, let us introduce some notation.

Note that whenever the ith element of $Y^*_{\mathbb{R}}$ is equal to zero, then the ith element of $Y^*_{\mathbb{Z}}$ is different from zero, and viceversa. This is entirely determined by the sign of the $\widehat{\psi}$ estimate. It would then be convenient to treat the observations separately depending on the sign of $\widehat{\psi}$. Define for each observation i the following:

$$\widehat{\psi}_{\mathbb{R},i} = \begin{cases} \widehat{\psi}_i & \text{if } \widehat{\psi}_i \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$\widehat{\psi}_{\mathbb{Z},i} = \begin{cases} |\widehat{\psi}_i| & \text{if } \widehat{\psi}_i < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then define two bootstrapped residuals for each observation i as follows:

$$v^*_{\mathbb{R},i} = \sqrt{\widehat{\psi}_{\mathbb{R},i}} \times r_i, \text{ and} \tag{7}$$

$$v^*_{\mathbb{Z},i} = \sqrt{\widehat{\psi}_{\mathbb{Z},i}} \times r_i, \tag{8}$$

where r_i is an independent Rademacher entry. The following proposition establishes a result to apply a bootstrap correction when the variance estimate is HC_U.

Proposition 3. *Suppose the regression model (1) is correctly specified. Let p denote the number of bootstraps, and $v^{*(j)}_{\mathbb{R}}$ and $v^{*(j)}_{\mathbb{Z}}$ are vectors of bootstrapped residuals for the j-th bootstrap iteration given by (7) and (8). Let the conditional variance estimate $\widehat{\mathbb{V}}(u|X)$ be diagonal, with non zero ith diagonal element given according to (6). Define $\beta^{*(j)}_{\mathbb{R}}$ and $\beta^{*(j)}_{\mathbb{Z}}$ as the OLS estimate of regressing $v^{*(j)}_{\mathbb{R}}$ and $v^{*(j)}_{\mathbb{Z}}$, respectively, over X for the j-th bootstrap iteration. Then,*

$$\delta^*_{HC_U} \equiv \frac{1}{p}\sum_{j=1}^{p}\left(\beta^{*(j)\prime}_{\mathbb{R}} A \beta^{*(j)}_{\mathbb{R}}\right) - \frac{1}{p}\sum_{j=1}^{p} \beta^{*(j)\prime}_{\mathbb{Z}} A \beta^{*(j)}_{\mathbb{Z}}$$

is an unbiased and consistent estimator of the direct bias correction $\widehat{\delta}$ when using a variance estimate $\widehat{\mathbb{V}}(u|X)$ according to equation (6).

Corollary 2. *Let $\mathbb{V}(u|X)$ be diagonal. Then the bootstrap correction $\delta^*_{HC_U}$ is an unbiased estimator of the bias δ.*

However, even though HC_U is unbiased and HC_2 is not, it is not clear that minimizes the mean squared error compared to other variance estimates. For example, let $\widehat{Y}_i = h_i' Y$ be the fitted value for observation i. Then,

$$HC_U = \frac{Y_i \widehat{u}_i}{1-h_{ii}} = \frac{\left(\widehat{Y}_i + \widehat{u}_i\right)\widehat{u}_i}{1-h_{ii}} = \frac{\widehat{Y}_i \widehat{u}_i}{1-h_{ii}} + HC_2.$$

While the expectation of HC_U is equal to ψ_i, it can be the case that its variance is larger than the one of HC_2. Thus, it is not clear that using the correction with HC_U would yield a more efficient bias corrected estimate of the quadratic forms compared to HC_2. In fact, we show in the simulation exercises below that our method is in general more efficient in terms of mean squared errors when using the HC_2 estimator instead of HC_U. However, as MacKinnon and White (1985) note, the HC_2 is biased and Cattaneo, Jansson, and Newey (2018) found that this bias does not vanish asymptotically when the number of covariates increases with the sample size. This is the case in most applications that use panel data.

Example 3: Clustered errors and serial correlation. When the error terms are clustered or present serial correlation within group, the covariance matrix is no longer diagonal. We restrict our attention to dependence of the errors only within a given group. Thus, we restrict to the case where the variance covariance matrix is block diagonal, as there are non zero elements around the diagonal corresponding to the dependence of the errors within the group g, but not across groups.[17] One particular example is when the group is a worker-firm match and errors are autocorrelated within match. Following Roodman, Nielsen, MacKinnon, and Webb (2019) we estimate the variance of observation i, $\widehat{\psi}_i$, with a variant of HC_1 from Example 2 that takes into account the number of groups G: $\widehat{\psi}_i = \frac{G}{G-1}\frac{n}{n-k}\widehat{u}_i^2$.

When the errors present dependence within the group we use a wild block bootstrap as proposed by Cameron et al. (2008). This consists of a wild bootstrap that takes into account the group dependence of the data. It has the benefit of accommodating any structure of the dependence within group. The Online Appendix describes the algorithm to implement our bias correction that keeps the dependence structure among groups through a wild block bootstrap.

In the following we describe the main features of our correction method and the source of efficiency gains from generating the dependent variable directly from bootstrapped residuals.

Scalability, flexibility, and easy implementation. Our method stands out for its: (i) scalability to correct many moments at once, (ii) flexibility on the assumption of the covariance matrix of the error terms, and (iii) it is easy to implement.

First, the method is suited to correct for additional quadratic objects besides the variances and covariances of the two leading fixed effects, say worker and firm fixed effects, as well as to perform corrections per subgroups, without increasing the computational cost. For example, on top of correcting the correlation of workers and firms fixed effects, one could correct for other moments that reflect labor market sorting, like the correlation between the worker fixed effect and the average fixed effect of the coworkers, as proposed by Lopes de Melo (2018). Also, one could do the correction of a quadratic object for the entire sample or for subsamples, like different periods,

[17] Assume that the the errors have a first order autocorrelation within group g and the true innovations are i.i.d. and therefore homoskedastic. We consider that the error term u of worker i at group g at time t in (1) is:

$$u_{i,g,t} = \rho u_{i,g,t-1} + \varepsilon_{i,g,t}, \quad \varepsilon_{i,g,t} \text{ i.i.d.}$$

We denote the variance of the innovation ε as σ_ε^2. Ordering the data by group, suppose the first group has three observations and the second group two observations. Then, $\mathbb{V}(u|X)$ is:

$$\mathbb{V}(u|X) = \frac{\sigma_\varepsilon^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & 0 & & \cdots & & 0 \\ & & & \vdots & & \ddots & & \vdots \\ \rho & 1 & \rho & \vdots & & & & \vdots \\ \rho^2 & \rho & 1 & 0 & & \cdots & & 0 \\ 0 & \cdots & 0 & 1 & \rho & 0 & \cdots & 0 \\ & & & \rho & 1 & 0 & \ddots & \\ \vdots & \ddots & \vdots & & & \ddots & & 0 \\ 0 & & 0 & & & & & 1 \end{pmatrix}.$$

The covariance matrix under clustering of the errors is similar but with al non-zero elements out of the diagonal equal to ρ.

genders, occupations, locations, etc. Similar to the previous case, this does not increase the number of bootstraps needed to estimate the corrections nor entail any computational burden in terms of methods available.

Second, our method is also flexible because it allows for a wide range of assumptions on the error's covariance matrix. It is only limited by the capacity of the bootstrap to replicate the assumed covariance matrix.

Finally, our method is easy to implement as it relies only on estimating linear regressions in every iteration. Thus, our method can profit from the development of any fast estimation procedure handling linear regressions with many covariates. The estimated coefficients of a linear regression $\widehat{\beta}$ are defined as the unknowns that solve for the normal equations, $X'X\widehat{\beta} = X'Y$. Standard algorithms solve the normal equations, without explicitely computing the inverse of $X'X$. In fact, if we could compute that inverse, we could compute directly the bias correction term $\widehat{\delta}$ as defined in equation (4). There has been significant progress in the development of efficient algorithms to estimate linear models with a large number of covariates. Especially when that large number stems from fixed effects. For example, when we compare our method with existing alternatives in Section 4, we use the preconditioned conjugate gradient method in Matlab to solve for the normal equations. However, the choice of algorithm for solving the normal equations is up to the user.

Efficiency gains. Using the bootstrap to correct for biases is ubiquitous in the literature. MacKinnon and Smith Jr (1998) (MS, henceforth) propose a similar bootstrap to correct for flat biases like the one considered here.[18] MS propose building the bootstrapped dependent variable by using the original estimate of β, $Y^* = X\widehat{\beta} + v^*$. In the context of our application, that would mean to compute the quadratic objects ${\beta^*_{MS}}^{(j)'} A {\beta^*_{MS}}^{(j)}$ for each bootstrap j and use them to create a bias correction:

$$\delta^*_{MS} = \frac{1}{p}\sum_{j=1}^{p} {\beta^*_{MS}}^{(j)'} A {\beta^*_{MS}}^{(j)} - \widehat{\beta}' A \widehat{\beta}.$$

MS already note that one can estimate a flat bias correction by using any $\widehat{\beta}$ to generate Y^*. For example, one can use $\widehat{\beta} = \mathbf{0}$, as we do in Proposition 2.

Analogously to equation (2) we have that in bootstrap j, ${\beta^*_{MS}}^{(j)'} A {\beta^*_{MS}}^{(j)} = \widehat{\beta}' A \widehat{\beta} + {v^*}^{(j)'} Q' A Q {v^*}^{(j)} + 2{v^*}^{(j)'} Q' A \widehat{\beta}$. When the errors are independent and the third moment is zero, it can be shown that the covariance of the last two terms conditional on X and u is equal to zero.[19] Thus we have that the variance of their bias correction conditional on X and u is:

$$\mathbb{V}(\delta^*_{MS}|X,u) = \frac{1}{p}\mathbb{V}(v^{*\prime} Q' A Q v^*|X,u) + \frac{4}{p}\mathbb{V}(v^{*\prime} Q' A \widehat{\beta}|X,u).$$

The expression above can be rewritten as:

$$\mathbb{V}(\delta^*_{MS}|X,u) = \mathbb{V}(\delta^*|X,u) + \frac{4}{p}\mathbb{V}(v^{*\prime} Q' A \widehat{\beta}|X,u), \tag{9}$$

which is larger than the variance of our estimator, $\mathbb{V}(\delta^*|X,u)$, attributable to the presence of the last

[18] A flat bias is one that does not depend on the levels of the original estimates. The bias from the quadratic forms is flat because the trace term in (3) is independent of $\widehat{\beta}$.

[19] These conditions would be satisfied, for example, if the error terms would be distributed normal, or, as in our applications, when we use the Rademacher errors for the bootstrap. A formal proposition of the statement and its proof can be found in the Online Appendix.

term, similarly to equation (2). While both methods yield an unbiased and consistent estimate of the direct bias correction $\widehat{\delta}$, δ^* is more efficient. The Online Appendix shows an illustrative example that δ^* is an unbiased estimate of $\widehat{\delta}$ and that it yields lower Mean Squared Errors (MSE) than δ^*_{MS}.

4 Comparison of Methods

In this section we first compare our method to Gaure (2014) and Kline et al. (2020). Both methods aim to estimate the trace term in equation (3). In the Online Appendix we also compare our method with Borovičková and Shimer (2017) who propose an alternative method to estimate directly some quadratic forms without first estimating a linear model.

The differences between Gaure, KSS and our method are on the scope of error structures allowed, the covariance matrix estimation and how easy they are to apply. All three methods are in principle suited to perform corrections with homoskedastic and heteroskedastic errors. Nevertheless, Gaure implemented his bias correction method on the R package *lfe* only under the assumption of homoskedastic errors. In contrast, KSS and ourselves provide corrections under heteroskedasticity and serial correlation or clustering of the errors.

Our method is the only one capable of doing multiple corrections at a time, for more covariates or corrections per groups, without increasing the computational cost. KSS and Gaure, on the contrary, need to solve new sets of equations in order to approximate each trace term that corresponds to any additional correction. Finally, our method is flexible in specifying serial correlation at different levels than the match.

4.1 Labor market simulations

An important application of two-way fixed effect models are the AKM type log wage regressions with worker and firm fixed effects. We closely follow Card, Heining, and Kline (2013) to implement the estimation of the following regression model for the log of the wage of worker i at time t:

$$w_{it} = \theta_i + \psi_{J(i,t)} + q_{it}\gamma + \varepsilon_{it}, \tag{10}$$

where the function $J(i,t)$ gives the identity of the unique firm that employs worker i at time t, θ_i is a worker fixed effect, $\psi_{J(i,t)}$ is the firm $J(i,t)$ fixed effect, q_{it} are time varying observables (age and education interacted with year effects), and ε_{it} is the error term.

Equation (10) can be estimated by OLS where the person and firm fixed effect estimators have the same structure as the ones in Section 2. Thus, the second order moments exhibit a similar bias and the implementation of the correction is analogous.

We compare the correction methods by simulating many labor markets under different assumptions on the error terms and evaluate them in terms of computation time and mean squared errors. In the Online Appendix, we also explore differences between the covariance estimation methods described in Section 3.

We first compare the bootstrap correction to Gaure and KSS under conditional homoskedasticity of the errors. We explore the HC_2 and HC_U estimators of the covariance matrix in our correction. Results are in Table 1. All the methods reduce the initial bias of the plug-in estimate. Gaure,

Table 1: Monte Carlo simulations. Homoskedastic errors.

	Time	Mean Squared Error (MSE×10^2)			
		$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.637	0.341	0.114	2.364
Gaure	17.3	0.050	0.109	0.015	0.058
Boot HC2	0.9	0.050	0.106	0.014	0.057
Boot HCU	1.0	0.050	0.106	0.014	0.057
KSS	0.9	0.050	0.106	0.014	0.057

Notes: *Plug-in*: naive plug-in estimator, *Gaure*: Gaure (2014) method implemented through the R package *lfe*, *Boot HC*2: bootstrap correction with HC_2 covariance matrix estimator of the error term, *Boot HCU*: bootstrap correction with HC_U, *KSS*: Kline et al. (2020) method leaving the observation out, *Time*: computing time in seconds. True moments are computed at the final sample for each method, i.e. largest connected set for *Gaure* and the largest leave-one-out connected set for *Boot HC*2, *Boot HCU* and *KSS*. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors (MSE) of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled).

KSS and our method are very similar in terms of MSE, and even look identical after rounding the numbers, with Gaure doing slightly worse.[20] Gaure is the slowest method and the bootstrap correction is similar to KSS in time.

Table 2 presents the comparison of our method using HC_2 and HC_U to KSS under conditional heteroskedasticity for different degrees of worker mobility.[21] Both methods reduce by more than 97% the MSE compared to the plug-in estimates in the low mobility case.[22] Our method with the HC_2 estimator is slightly more efficient for both mobility cases, and it also outperforms KSS in terms of time. The bootstrap correction using the HC_U estimator for the covariance matrix of the errors is identical to KSS in terms of MSE and time.[23] In the Online Appendix, we show that HC_2 estimate for the covariance matrix outperforms HC_0, HC_1 and HC_U measured by MSE when doing the bias correction with heteroskedastic errors.

Table 3 presents the results from a simulation with a non diagonal covariance matrix. In particular, we assume that there is serial correlation of the wages within a given match and the true innovation is homoskedastic. The table compares the plug-in estimate to our bootstrap correction and to KSS. *Boot* is the best performing correction method in terms of accuracy and time. We show in the Online Appendix that the differences in performance are amplified when we have heteroskedastic innovations at the match level.

Why can our method be faster? In the above simulations, our method is faster when using the HC_2 covariance matrix estimator and similar to the method proposed by KSS when using HC_U. The underlying reason is that our method with HC_2 needs to do *at most* two iterative procedures regardless of the number of corrections: one for estimating the leverage—for example, if one uses HC_2 for the covariance matrix estimator—and one for the bootstrap. When using the HC_U estimator we need to estimate at most three iterative procedures. On the other hand, KSS method needs to

[20]We use the *bccor* command of Gaure's *lfe* R package with 300 maximum samples and tolerance of 1e-6. We run Version 3 of KSS Matlab code eliminating observations (instead of matches) for the leave-one-out estimation with 300 simulations to estimate the leverage and corrections at once. We run our corrections in Matlab with tolerance of 1e-6 and 300 simulations.

[21]When workers are more mobile, the firm fixed effect estimates are less noisy. As this noise is the source of the bias of the quadratic objects, more precise estimates will yield a smaller bias as one can see from the *Plugin* estimates in Table 2.

[22]Table 1 in Kline et al. (2020) shows that their connected set is similar to our low mobility scenario with 2.7 movers per firm and average firm size of 12.

[23]In the Online Appendix, we compare the densities of the bias for the different methods. The densities show that both methods (KSS and bootstrap) are similar but the bootstrap method with HC_2 has smaller variance for the reasons suggested in Section 3. We also show in the Online Appendix that the results are similar even when using a more realistic sample size of roughly 5 million observations.

Table 2: Monte Carlo simulations. Heteroskedastic errors.

			Mean Squared Error (MSE×10^2)			
Mov/firm	Model	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Low Mobility						
3	Plug-in		22.885	7.702	6.451	12.346
3	Boot HC2	0.6	0.225	0.666	0.192	0.361
3	Boot HCU	0.7	0.265	0.711	0.233	0.403
3	KSS	0.7	0.268	0.709	0.233	0.404
Mid Mobility						
5	Plug-in		10.518	1.670	1.070	4.419
5	Boot HC2	0.7	0.085	0.256	0.048	0.130
5	Boot HCU	0.8	0.086	0.258	0.049	0.131
5	KSS	0.8	0.087	0.258	0.049	0.131

Notes: *Plug-in*: naive plug-in estimator, *Boot HC2*: bootstrap correction with HC_2 covariance matrix estimator of the error term, *Boot HCU*: bootstrap correction with HC_U, *KSS*: Kline et al. (2020) method leaving the observation out. True moments are computed at the leave-one-out connected set. *Mov/firm*: number of movers per firm and the average firm has 12 employees, *Time*: computing time in seconds. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled).

Table 3: Monte Carlo simulations. Serial correlation with homoskedasticity.

		Mean Squared Error (MSE×10^2)			
	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		94.352	1.670	0.603	32.208
Boot	0.5	9.674	0.263	0.053	3.330
KSS	1.1	21.572	0.254	0.052	7.293

Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method with a wild block bootstrap where each match defines a block and we skip the pruning of the data. *KSS* is the Kline et al. (2020) method leaving a match out. The average firm has 10 movers and 12 employees. *Time* is the computing time in seconds. True moments are computed at the largest connected set for *Boot* and at the largest leave-one-out connected set for *KSS*. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors (MSE) multiplied by 100 of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. *Average* is the average MSE (also scaled).

do, in general, the same number of iterative procedures as number of corrections plus the iteration for the leverage estimation. When interested in the two leading fixed effects, they can reduce the minimum number of iterative procedures to three: one for the leverage and two extra for the variance of the worker fixed effects, the variance of firm fixed effects, and their covariance.[24] With HC_U, having the same number of iterative procedures than KSS, the computational times are also similar. Table **??** in the Online Appendix shows in a Monte Carlo simulation with a larger sample size that the bootstrap correction with HC_2 is about a third faster than with HC_U.

5 Application

In the application we use a panel data from the French statistical agency (INSEE) from 2002 to 2019.[25] Our dependent variable is (log) gross daily wage of full time employees with ages between

[24] See section 2.3.2 in the computational appendix of KSS.

[25] In particular we use *Panel tous salariés-EDP* that consists of a random subsample of workers with firm identifiers and socio-demographic variables. The sample consists of workers born in October on certain days. The sample size was increased in 2002 so we took this as the starting year.

Table 4: Application. Plug-in vs corrected decomposition.

	Plug-in	Boot *HCU*		Plug-in	Boot *HCU*
$Var(y)$	0.2929	0.2929	$2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.0107	0.0054
$Var(\widehat{\theta}_i)$	0.1791	0.1595	$2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$	-0.0125	-0.0123
$Var(\widehat{\psi}_j)$	0.0425	0.0320	$2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$	0.0002	0.0002
$Var(\mathbf{q}\widehat{\gamma})$	0.0173	0.0173	$Corr(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.0615	0.0375
$Var(\widehat{\epsilon})$	0.0770	0.0909	Obs.	4652631	4652631

Notes: *Plug-in* refers to the uncorrected estimates of each of the variance components at the largest connected set and *Boot HCU* refers to the estimates after our bootstrapped correction using the HC_U estimator of the variance of the error terms. $Var(y)$ is the variance of log wages, $Var(\widehat{\theta}_i)$ the variance of worker fixed effects (naive $\hat{\sigma}^2_\theta$ or corrected $\tilde{\sigma}^2_\theta$), $Var(\widehat{\psi}_j)$ is the variance of firm fixed effects, $Var(\mathbf{q}\widehat{\gamma})$ is the variance of other covariates and $Var(\widehat{\epsilon})$ is the variance of the error term. The other terms of the decomposition are twice the covariances between the fixed effects and the covariates ($2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$, $2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$ and $2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$). Finally, $Corr(\widehat{\theta}_i, \widehat{\psi}_j)$ is the estimated correlation between worker and firm fixed effects and *Obs.* is the number of observations.

20 and 60 working at private firms.

The goal is to use our bootstrap method to do a bias corrected variance decomposition of log wages. Table 4 shows the variance decomposition of log wages as well as the correlation between firm and worker fixed effects using the plug-in moments and the corrected ones under the assumption of diagonal covariance matrix using the HC_U estimator. The variance of the person and firm effects are both reduced and they explain a lower share of the total variance after the correction. The correlation becomes positive but still close to zero. Naturally, the variances and covariance of the person and firm effects are the moments that change the most after the correction. The reason is that the underlying estimates of the person and fixed effects are very noisy. In contrast, when the underlying estimates of a particular moment are estimated with precision, as it is in the case of the parameters $\widehat{\gamma}$ associated with the common covariates $\mathbf{q}$, the change between the plug-in and corrected moments is negligible.

We explore the benefits of our method by computing the correlation between the worker fixed effect and the average fixed effect of coworkers (Lopes de Melo, 2018). We find in the application that the estimate of the plugin correlation between workers and coworkers is slightly upward biased. The estimated correlation goes from 0.263 with the plugin to 0.243 after the correction.

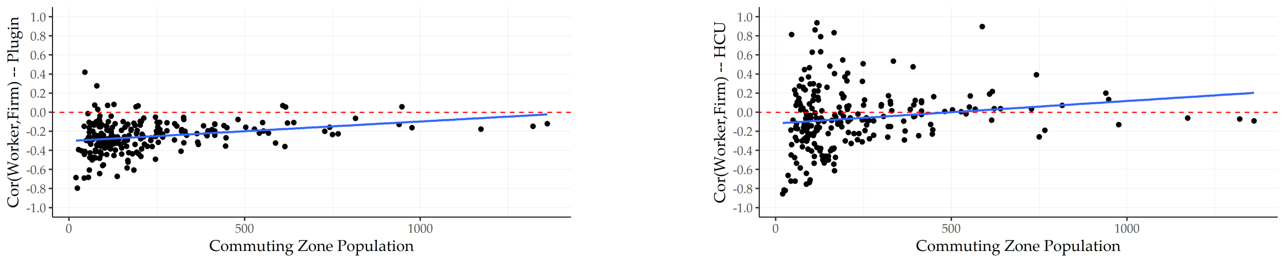
To fully exploit the benefit of our bootstrap correction method we also perform several variance decompositions per commuting zone. Figure 1 presents the relationship between the correlation of worker and firm fixed effects with the commuting zone population. Taking the plugin estimates on Panel (a), most of the commuting zone sorting estimates are negative and there seems to be a positive gradient between sorting and commuting zone size. Similarly, Panel (b) shows the estimates after correcting for limited mobility bias with the *HCU* estimator of the variance of the error terms. We find that many correlations turn to be positive and the sorting-population gradient is stable.

Finally, we compare our method with HC_2 and HC_U estimators against KSS using the French data in the Online Appendix.

6 Conclusion

In this paper, we propose a computationally feasible bootstrap method to correct for the small-sample bias found in all quadratic forms in the parameters of linear models with a very large

Figure 1: Application. Correlation of Worker and Firm Fixed Effects.



Notes: These figures present on the x axis commuting zone population. On the y axis, Panel (a) shows on the estimated correlation between worker and firm fixed effects with the plugin estimates and Panel (b) shows the analogous with the *HCU* estimator of the variance of the error terms. We remove the commuting zone with the highest population for readability of the figure.

number of covariates. We show using Monte Carlo simulations that the method is effective at reducing the bias. The application to French labor market data shows that the correction increases the correlation between firm and worker fixed effects. Depending on the sample and on the specification, our bias correction method changes the sign of that correlation and in all cases it changes the relative importance of the different components in explaining the variance of log wages.

The only requirements to implement our correction is to have a bootstrap procedure that is consistent with the assumption on the variance-covariance matrix of the error term and to estimate the model several times. The correction can thus be applied easily to any study running an AKM type regression or two-way fixed effects regressions. Our method is similar in time to Kline et al. (2020) and as accurate in the simulations. The main advantage of our approach is that it allows to increase the number of moments to correct without increasing the computational costs.

Appendix for Publication

A Proofs

Proposition 1

Proof. By the linearity of the trace and expectation operators we have that

$$\mathbb{E}(\widehat{\delta}|X) = \mathbb{E}\left(\text{trace}\left(Q'AQ\widehat{\mathbb{V}}(u|X)\right)|X\right) = \text{trace}\left(Q'AQ\mathbb{E}\left(\widehat{\mathbb{V}}(u|X)|X\right)\right) = \text{trace}\left(Q'AQ\mathbb{V}(u|X)\right) = \delta$$

□

Proposition 2

Proof. First, note that for any bootstrap estimate j of the quadratic form $\left(\beta^{*(j)}\right)' A\beta^{*(j)}$ we have that

$$\left(\beta^{*(j)}\right)' A\beta^{*(j)} = (v^{*(j)})'Q'AQv^{*(j)}.$$

Under the bootstrap, i.e. conditional on X and u, the only source of randomness is $v^{*(j)}$. Taking expectations under the bootstrap of $\left(\beta^{*(j)}\right)' A\beta^{*(j)}$, conditioning on X and u, and using the assumption $\mathbb{E}(v^{*(j)}|X,u) = 0$, we get

$$\mathbb{E}_{v^*}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)} \middle| X,u\right) = \text{trace}\left(Q'AQ\mathbb{V}(v^{*(j)}|X,u)\right).$$

By assumption $\mathbb{V}(v^{*(j)}|X,u) = \widehat{\mathbb{V}}(u|x)$, then $\mathbb{E}_{v^*}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)} \middle| X,u\right) = \widehat{\delta}$.

Unbiased. Taking expectations over $\delta^* \equiv \frac{1}{p}\sum_{j=1}^{p}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)}\right)$, conditioning on X and u, we obtain

$$\mathbb{E}_{v^*}(\delta^*|X,u) = \frac{1}{p}\sum_{j=1}^{p}\mathbb{E}_{v^*}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)} \middle| X,u\right) = \frac{1}{p}\sum_{j=1}^{p}\widehat{\delta} = \widehat{\delta}.$$

Consistent. From the definition of $\delta^* \equiv \frac{1}{p}\sum_{j=1}^{p}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)}\right)$, we have that

$$\frac{1}{p}\sum_{j=1}^{p}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)}\right) \xrightarrow{p} \mathbb{E}_{v^*}\left(\left(\beta^{*(j)}\right)' A\beta^{*(j)} \middle| X,u\right) = \widehat{\delta}.$$

□

Corollary 1

Proof. Using the Law of Iterated Expectations we get

$$\mathbb{E}(\delta^*|X) = \mathbb{E}_u\left(\mathbb{E}_{v^*}(\delta^*|X,u) \mid X\right) = \mathbb{E}_u(\widehat{\delta}|X) = \delta.$$

□

Proposition 3

Proof. Let the vector of complex bootstrap residuals v^* be equal to

$$v^* = v^*_{\mathbb{R}} + \mathbf{i}v^*_{\mathbb{Z}}.$$

Let β^* be the complex bootstrap estimate when using v^* as the left-hand-side variable in the regression. Then, the conditional expectation of the quadratic object is equal to:

$$\mathbb{E}_{v^*}\left(\beta^{*\prime}A\beta^* \,\middle|X,u\right) = \text{trace}\left(Q'AQ\mathbb{V}(v^*|X,u)\right).$$

The variance of the complex residuals is

$$\begin{aligned}\mathbb{V}(v^*|X,u) &= \mathbb{E}_{v^*}\left(v^*v^{*\prime} \,\middle|X,u\right)\\ &= \mathbb{E}_{v^*}\left(v^*_{\mathbb{R}}v^{*\prime}_{\mathbb{R}} \,\middle|X,u\right) - \mathbb{E}_{v^*}\left(v^*_{\mathbb{Z}}v^{*\prime}_{\mathbb{Z}} \,\middle|X,u\right) + \mathbf{i}\mathbb{E}_{v^*}\left(v^*_{\mathbb{R}}v^{*\prime}_{\mathbb{Z}} + v^*_{\mathbb{Z}}v^{*\prime}_{\mathbb{R}} \,\middle|X,u\right).\end{aligned}$$

Now, $\mathbb{E}\left(v^*_{\mathbb{R}}v^{*\prime}_{\mathbb{Z}} \,\middle|X,u\right)$ is equal to zero. This is because whenever $v^*_{\mathbb{R}}$ is different than zero then $v^*_{\mathbb{Z}}$ is zero, and viceversa (see equations (7) and (8)). So all of the diagonal terms of the matrix $v^*_{\mathbb{R}}v^{*\prime}_{\mathbb{Z}}$ are equal to zero. The expectation of all the off-diagonal terms are also equal to zero by the independence of the Rademacher entries for different observations. Clearly, also the conditional expectation of the transpose matrix is equal to zero. Therefore, we have that

$$\mathbb{V}(v^*|X,u) = \mathbb{E}_{v^*}\left(v^*_{\mathbb{R}}v^{*\prime}_{\mathbb{R}} \,\middle|X,u\right) - \mathbb{E}_{v^*}\left(v^*_{\mathbb{Z}}v^{*\prime}_{\mathbb{Z}} \,\middle|X,u\right) = \widehat{\Psi}_{\mathbb{R}} - \widehat{\Psi}_{\mathbb{Z}},$$

where $\widehat{\Psi}_{\mathbb{R}}$ and $\widehat{\Psi}_{\mathbb{Z}}$ are matrices where their ith diagonal terms are $\widehat{\psi}_{\mathbb{R},i}$ and $\widehat{\psi}_{\mathbb{Z},i}$, respectively, and their off diagonal terms are equal to zero. Then, $\mathbb{V}(v^*|X,u) = \widehat{\mathbb{V}}(u|x)$, where the covariance estimate $\widehat{\mathbb{V}}(u|x)$ is a diagonal matrix, with diagonal elements given by equation (6).

Now that we have established that the conditional variance $\mathbb{V}(v^*|X,u)$ is equal to $\widehat{\mathbb{V}}(u|x)$, then we have that $\mathbb{E}_{v^*}\left(\beta^{*\prime}A\beta^* \,|X,u\right)$ is equal to the direct bias correction $\widehat{\delta}$ when using HC_U as the covariance estimator. Now let establish that the bootstrap correction $\delta^*_{HC_U}$ is an unbiased and consistent estimator of $\widehat{\delta}$.

Unbiased. Let the OLS estimates for the real and imaginary terms of the jth bootstrapped estimate be equal to:

$$\begin{aligned}\beta_{\mathbb{R}^*}{}^{(j)} &= \left(X'X\right)^{-1}X'v_{\mathbb{R}^*}{}^{(j)}\\ \beta_{\mathbb{Z}^*}{}^{(j)} &= \left(X'X\right)^{-1}X'v_{\mathbb{Z}^*}{}^{(j)}.\end{aligned}$$

Taking expectations over $\delta^*_{HC_U} \equiv \frac{1}{p}\sum_{j=1}^{p}\left(\beta_{\mathbb{R}^*}{}^{(j)\prime}A\beta_{\mathbb{R}^*}{}^{(j)}\right) - \frac{1}{p}\sum_{j=1}^{p}\left(\beta_{\mathbb{Z}^*}{}^{(j)\prime}A\beta_{\mathbb{Z}^*}{}^{(j)}\right)$, conditioning on X and u, we obtain

$$\begin{aligned}\mathbb{E}_{v^*}(\delta^*_{HC_U}|X,u) &= \frac{1}{p}\sum_{j=1}^{p}\mathbb{E}_{v^*}\left(\beta_{\mathbb{R}^*}{}^{(j)\prime}A\beta_{\mathbb{R}^*}{}^{(j)} \,\middle|\, X,u\right) - \frac{1}{p}\sum_{j=1}^{p}\mathbb{E}_{v^*}\left(\beta_{\mathbb{Z}^*}{}^{(j)\prime}A\beta_{\mathbb{Z}^*}{}^{(j)} \,\middle|\, X,u\right)\\ &= \mathbb{E}_{v^*}\left(\beta^{*\prime}_{\mathbb{R}}A\beta^*_{\mathbb{R}} \,\middle|\, X,u\right) - \mathbb{E}_{v^*}\left(\beta^{*\prime}_{\mathbb{Z}}A\beta^*_{\mathbb{Z}} \,\middle|\, X,u\right),\end{aligned}$$

where we omitted the j terms in the last equality to ease on notation. Let $B \equiv Q'AQ$. Then,

$$\begin{aligned}\mathbb{E}_{v^*}(\delta^*_{HC_U}|X,u) &= \text{trace}\left(B\mathbb{V}(v^*_{\mathbb{R}}|X,u)\right) - \text{trace}\left(B\mathbb{V}(v^*_{\mathbb{Z}}|X,u)\right) \\ &= \sum_{i=1} B_{ii}\widehat{\psi}_{\mathbb{R},i} - \sum_{i=1} B_{ii}\widehat{\psi}_{\mathbb{Z},i} = \sum_{i=1} B_{ii}\left(\widehat{\psi}_{\mathbb{R},i} - \widehat{\psi}_{\mathbb{Z},i}\right) \\ &= \sum_{i=1} B_{ii}\widehat{\psi}_i = \text{trace}\left(B\widehat{\mathbb{V}}(u|x)\right) = \widehat{\delta}.\end{aligned}$$

Consistent. From the definition of $\delta^*_{HC_U} \equiv \frac{1}{p}\sum_{j=1}^{p}\left(\beta_{\mathbb{R}^*}{}^{(j)\prime}A\beta_{\mathbb{R}^*}{}^{(j)}\right) - \frac{1}{p}\sum_{j=1}^{p}\left(\beta_{\mathbb{Z}^*}{}^{(j)\prime}A\beta_{\mathbb{Z}^*}{}^{(j)}\right)$, we have that:

$$\begin{aligned}&\frac{1}{p}\sum_{j=1}^{p}\left(\beta_{\mathbb{R}^*}{}^{(j)\prime}A\beta_{\mathbb{R}^*}{}^{(j)}\right) - \frac{1}{p}\sum_{j=1}^{p}\left(\beta_{\mathbb{Z}^*}{}^{(j)\prime}A\beta_{\mathbb{Z}^*}{}^{(j)}\right) \\ &\xrightarrow{p} \mathbb{E}_{v^*}\left(\beta_{\mathbb{R}^*}{}^{(j)\prime}A\beta_{\mathbb{R}^*}{}^{(j)} \;\middle|\; X,u\right) - \mathbb{E}_{v^*}\left(\beta_{\mathbb{Z}^*}{}^{(j)\prime}A\beta_{\mathbb{Z}^*}{}^{(j)} \;\middle|\; X,u\right) = \widehat{\delta}.\end{aligned}$$

□

Corollary 2

Proof. If the covariance matrix of the error terms $\mathbb{V}(u|x)$ is diagonal, then the HC_U covariance estimate $\widehat{\mathbb{V}}(u|x) = \Psi$, whose diagonal terms are given by equation (6) and off-diagonal terms are zero, is an unbiased estimate of the covariance matrix $\mathbb{V}(u|x)$. Then, just apply Proposition 1 and Corollary 1.

Although this is not our result, for completeness, we write down the proof that the HC_U covariance estimate is an unbiased estimator if $\mathbb{V}(u|x)$ is diagonal.

Following Angrist and Pischke (2008), note that the vector of estimated residuals $\widehat{u}$ is equal to:

$$\widehat{u} = Y - X\left(X'X\right)^{-1}X'Y = \left[I - X\left(X'X\right)^{-1}X'\right](X\beta + u) = Mu,$$

where $M \equiv I - X\left(X'X\right)^{-1}X'$ is a symmetric non-sthocastic matrix and u is the error term. Let m_i be the ith column of M. Then $\widehat{u}_i = u'm_i$.

Assume the true covariance matrix Ψ is diagonal with non-zero off diagonal ith term equal to ψ_i. Let $\widehat{\psi}_i$ be the HC_U estimate of ψ_i. This estimate is equal to $\widehat{\psi}_i = \frac{y_i\widehat{u}_i}{1-h_{ii}}$, where $h_{ii} = X\left(X_i'X\right)^{-1}X_i'$. To further simplify let $m_i = \ell_i - h_i$ where ℓ_i is the ith column of the identity matrix I and $h_i = X\left(X'X\right)^{-1}X_i'$ is the ith column of the H matrix. Then the expectation of $\widehat{\psi}_i$ is equal to:

$$\begin{aligned}\mathbb{E}\left(\widehat{\psi}_i \mid X\right) &= \mathbb{E}\left(\frac{y_i\widehat{u}_i}{1-h_{ii}} \;\middle|\; X\right) = \frac{1}{1-h_{ii}}\mathbb{E}\left(y_i(u'm_i) \mid X\right) \\ &= \frac{1}{1-h_{ii}}\mathbb{E}\left(\ell_i'Y(u'm_i) \mid X\right) = \frac{1}{1-h_{ii}}\mathbb{E}\left(\ell_i'Y(u'm_i) \mid X\right) \\ &= \frac{1}{1-h_{ii}}\mathbb{E}\left(\ell_i'(X\beta+u)(u'm_i) \mid X\right) = \frac{1}{1-h_{ii}}\left[\ell_i'\mathbb{E}\left(uu' \mid X\right)(\ell_i - h_i)\right] \\ &= \frac{1}{1-h_{ii}}\left[\ell_i'\Psi(\ell_i - h_i)\right] = \frac{1}{1-h_{ii}}\left[\psi_i - \ell_i'\Psi h_i\right] = \frac{1}{1-h_{ii}}\left[\psi_i - \psi_i h_{ii}\right] = \psi_i.\end{aligned}$$

□

References

Abowd, J., F. Kramarz, P. Lengermann, and S. Pérez-Duarte (2004): "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States," *Unpublished Manuscript*.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018): "Firms and the decline in earnings inequality in Brazil," *American Economic Journal: Macroeconomics*, 10, 149–189.

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

Angrist, J. D. and J.-S. Pischke (2008): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Arellano-Bover, J. and S. San (2023): "The Role of Firms and Job Mobility in the Assimilation of Immigrants: Former Soviet Union Jews in Israel 1990-2019," .

Avron, H. and S. Toledo (2011): "Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix," *Journal of the ACM (JACM)*, 58, 8.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023): "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, 41.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

Borovičková, K. and R. Shimer (2017): "High wage workers work for high wage firms," Tech. rep., National Bureau of Economic Research.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 90, 414–427.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018): "Inference in linear regression models with many covariates and heteroscedasticity," *Journal of the American Statistical Association*, 113, 1350–1361.

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum (2022): "Matching in cities," *Journal of the European Economic Association*, 20, 1478–1521.

Dhaene, G. and K. Jochmans (2015): "Split-panel jackknife estimation of fixed-effect models," *The Review of Economic Studies*, 82, 991–1030.

Gaure, S. (2013): "lfe: Linear group fixed effects," *The R Journal*, 5, 104–117.

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

Gerard, F., L. Lagos, E. Severnini, and D. Card (2021): "Assortative matching or exclusionary hiring? the impact of employment and pay policies on racial wage differences in brazil," *American Economic Review*, 111, 3418–3457.

Hayashi, F. (2000): "Econometrics. 2000," *Princeton University Press. Section*, 1, 60–69.

Hutchinson, M. (1989): "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines," *Communications in Statistics-Simulation and Computation*, 18, 1059–1076.

Jochmans, K. (2018): "Heteroskedasticity-robust inference in linear regression models," *arXiv preprint arXiv:1809.06136*.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

Lachowska, M., A. Mas, R. Saggio, and S. A. Woodbury (2022): "Do firm effects drift? Evidence from Washington administrative data," *Journal of Econometrics*.

Lopes de Melo, R. (2018): "Firm wage differentials and labor market sorting: Reconciling theory and evidence," *Journal of Political Economy*, 126, 000–000.

MacKinnon, J. G. and A. A. Smith Jr (1998): "Approximate bias correction in econometrics," *Journal of Econometrics*, 85, 205–230.

MacKinnon, J. G. and H. White (1985): "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305–325.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019): "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 19, 4–60.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019): "Firming up inequality," *The Quarterly Journal of Economics*, 134, 1–50.

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

Online Appendix

Limited Mobility Bias: More Corrections But No More Cost

Miren Azkarate-Askasua and Miguel Zerecero

In this Appendix we first illustrate our bootstrap correction method in a simple example. Second, we provide details on how we construct the simulated labor markets that we use to test and compare our bootstrap correction. Third, we explain how to estimate the leverage of an observation in a linear regression model. This is useful when one uses covariance matrix estimators that require the leverage, and when the direct computation of the leverage is computationally costly. Fourth, we briefly explain how to choose the number of bootstraps based in Chebyshev's inequality. Sixth, we discuss additional structures of the covariance matrix and explain the algorithms used in the paper. Seventh, we compare our method to KSS in the application and to Borovičková and Shimer (2017), both with simulated labor market data and the French data. Eighth, we present a formal proposition that yields as a corollary that our bias correction is more efficient than the one proposed by MacKinnon and Smith Jr (1998). Finally, we present tables and figures that correspond to additional exercises that complement the analysis in the main text.

OA-1 Simple illustration

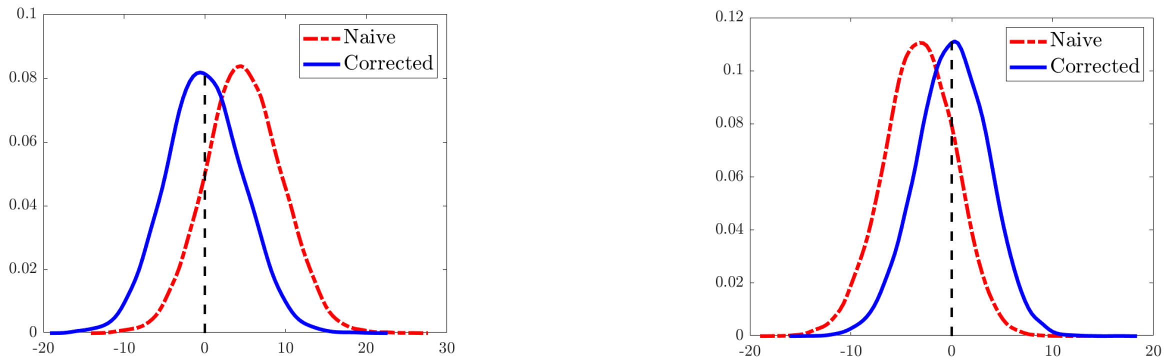
We illustrate the effectiveness of our bias correction method with some simple Monte Carlo simulations. The model design is the same as in equation (5) with homoskedastic errors and sample size $n = 500$. The number of covariates is $k_1 = k_2 = 200$. We keep this number relatively low to be able to compute what we dubbed previously the direct bias correction $\widehat{\delta}$. We do 10,000 simulations in total. In each simulation, keeping X fixed, we draw new error terms to form the dependent variable. We estimate $\widehat{\beta}$ and compute the direct bias correction terms. After the estimation, we perform $p = 100$ bootstraps and use them to compute the estimation of the bootstrap correction. We do a Wild bootstrap consistent with using HC_1 as the covariance estimator.[1]

Figures OA-1a and OA-1b show the effectiveness of our method. Figure OA-1a plots the distribution of the difference between the plug-in estimate of the variance ($\widehat{\sigma}^2_{1,PI}$) and the true variance ($\sigma^2_{1,PI}$). The figure also plots the difference between the bootstrap corrected variance ($\widehat{\sigma}^2_{1,b}$) and the true variance. Figure OA-1b shows analogous distributions of differences between estimates and true moments but for the covariance ($\sigma_{12,PI}$). The figures show that the distribution of the differences between the plug-in estimates the true moment are not centered at zero, reflecting the bias. On the other hand the distribution of difference between the bootstrap corrected moments and the true ones are centered at zero, suggesting our method is effective in reducing the bias.

In terms of efficiency, our methods is very close to the direct correction—which is the best one can do—but outperforms more traditional bootstrap methods. Table OA-1 presents the mean and variance of the differences of our bootstrap method δ^* and the bootstrap following MacKinnon and

[1] In other words, for each observation and bootstrap iteration, we sample a Rademacher random variable and multiply it to each observation's residual times $\sqrt{N/(N-K)}$.

Figure OA-1: Differences between the true moments with plug-in and bootstrap estimators



(a) Density of $\widehat{\sigma}^2_{1,PI} - \sigma^2_1$ and $\widehat{\sigma}^2_{1,b} - \sigma^2_1$

(b) Density of $\widehat{\sigma}_{12,PI} - \sigma_{12}$ and $\widehat{\sigma}_{12,b} - \sigma_{12}$

Notes: The left figure presents the distributions of the differences between the true variance σ^2_1 and both, the naive plug-in estimated variance $\widehat{\sigma}^2_{1,PI}$ and the bias corrected estimated variance $\widehat{\sigma}^2_{1,b}$. The distribution of the difference between the true moment and the bias corrected estimated covariance is centered at zero. The right figure does the same but for the covariance σ_{12}.

Table OA-1: Comparison Bootstrap and Direct Estimations.

	$\widehat{\delta} - \delta^*$		$\widehat{\delta} - \delta^*_{MS}$		Mean Squared Error			
	Mean	Variance	Mean	Variance	Plug-In	Direct	Boot	Boot MS
$\widehat{\text{var}}(X_1\beta_1)$	-0.00015	0.0015	-0.0037	0.08	47.78	24.34	24.34	24.44
$\widehat{\text{var}}(X_2\beta_2)$	-1.2×10^{-5}	0.0015	0.0054	0.19	79.00	55.55	55.55	55.74
$\widehat{\text{cov}}\,(X_1\beta_1, X_2\beta_2)$	9.3×10^{-5}	0.0014	-0.0014	0.05	25.83	15.18	15.18	15.22

The first two columns represent, respectively, the mean and the variance of the difference between the direct correction $\widehat{\delta}$ and the bootstrap correction δ^*. Columns 3 and 4 are analogous but using the bootstrap correction proposed by MacKinnon and Smith Jr (1998), δ^*_{MS}. Columns 5 to 8 compute the Mean Squared Error between the different estimated moments and the true ones. *Plug-In* refers to the non-corrected estimated moments using the estimates of the linear regression. *Direct* uses the estimated moments with the direct bias correction. *Boot* and *Boot MS* refer, respectively, to the moments with our bootstrap correction and with the bootstrap correction proposed by MacKinnon and Smith Jr (1998).

Smith Jr (1998) δ^*_{MS} with respect to the direct correction $\widehat{\delta}$.[2] The mean differences of our method are very small as well as the variances, meaning that the estimated bootstrap correction is performing almost as well as the direct correction in almost all simulations. The alternative bootstrap correction δ^*_{MS} in Columns 3 and 4 performs worse in terms of bias and variance.

Table OA-1 also shows the Mean Squared Error (MSE) of the different estimated moments. The MSE of naive plug-in estimators is larger than the one obtained with the directly corrected and bootstrap corrected moments. As our estimator is a noisy estimate of the direct correction, it is expected that the MSE of the corrected moments using our estimator to be larger than the directly corrected moments, although very close. In fact, to the level of rounding presented in the table, the two are indistinguishable. Also, as expected, our bootstrap has lower MSE than the alternative bootstrap corrected moments which follows MacKinnon and Smith Jr (1998).

[2] As previously stated, MacKinnon and Smith Jr (1998) propose to generate the bootstrap dependent variable as $Y^* = X\widehat{\beta} + v^*$. Their correction is: $\delta^*_{MS} = \frac{1}{p}\sum_{j=1}^{p}\left(\beta^{*\prime}_{j,MS} A \beta^*_{j,MS}\right) - \widehat{\beta}' A \widehat{\beta}$, where the last term is the plug-in estimate.

OA-2 Construction of Simulated Labor Market Data

We construct several simulated labor markets depending on the number of movers per firm, and type of error term. Here, we briefly describe the construction of the simulated labor markets.[3]

We start by determining the size of the labor market. We have 5000 unique workers and 400 unique firms at the beginning of the sample. This gives an average firm size of 12 workers which is similar to the average firm size in the data used by Kline, Saggio, and Sølvsten (2020).[4] Their connected set with an average of 2.7 movers per firm is similar to our low mobility simulations with 3 movers per firm. The sample runs for 7 periods (years) but we allow that workers randomly drop from the sample with a minimum of 2 observations per worker. This leads to a total sample size of roughly 22,000 observations.

Worker and firm fixed effects are random draws from normal distributions. We assume that there is sorting depending on the permanent types, which leads to non negative correlations between worker and firm fixed effects while fulfilling exogenous mobility. That is, a low type worker is more likely to match with a low type firm if we assume positive sorting but sorting does not depend on match specific shocks. This preserves the exclusion restriction necessary for OLS. Matches are formed either at the beginning of the sample or afterwards for the movers. Errors are i.i.d. and normally distributed in the baseline simulation with homoscedastic errors. When we use heteroscedastic errors, these are also normally distributed with an observation (worker-year) specific variance that is randomly drawn from a uniform distribution. Finally, when we use serially correlated errors, these are simulated from a first order autoregressive process with persistence of 0.7 and homoscedastic or heteroscedastic innovations. The simulated log wage is like equation (10) in the main text with only the firm and worker fixed effects

$$w_{it} = \theta_i + \psi_{J(i,t)} + \varepsilon_{it}. \tag{OA-1}$$

OA-3 Leverage Estimation

The direct computation of the leverage, by using the diagonal of the projection matrix $H \equiv X\left(X'X\right)^{-1}X'$, is computationally infeasible when the number of covariates is large. Again, the problem is the computation of $\left(X'X\right)^{-1}$.

Here we follow a way to estimate the leverage first proposed by Kline, Saggio, and Sølvsten (2021).[5] This procedure is very similar to our bias estimator. We simulate repeatedly random variables and use the fitted values of the projection into X to estimate the leverage. The procedure starts by generating the endogenous variable ω where each entry is i.i.d. with (conditional) mean equal to zero and (conditional) variance equal to 1. Projecting it into X, we have that the expectation of the squared of the fitted value $\widehat{\omega}$ is

$$\mathbb{E}\left(\widehat{\omega}_i^2|X\right) = x_i\left(X'X\right)^{-1}X'\mathbb{E}\left(\omega\omega'|X\right)X\left(X'X\right)^{-1}x_i' = x_i\left(X'X\right)^{-1}x_i' = h_{ii},$$

where x_i' is the ith row of matrix of covariates X. Let n_h be the number of simulations for the vector

[3] We thank Simen Gaure for sharing with us a piece of code that we used as a base for the simulations.

[4] See Table 1 in Kline et al. (2020) where each worker is observed twice.

[5] The reference for Kline et al. (2021) which contains the details on the derivations of the leverage estimator can be found here.

ω used to estimate the leverages $\widehat{h}_{ii}$. Similarly to what we do to estimate the bias correction, we simulate different vectors of the dependent variable ω, compute the fitted values for each simulation j and then take a sample mean across all the simulations $j = \{1, ..., n_h\}$ of ω.

Additionally, and following Kline et al. (2021), we can also estimate a value for one minus the leverage, $m_{ii} = 1 - h_{ii}$ by averaging the squared residuals of the same regressions we run above. So the ith residual is equal to $\omega_i - \widehat{\omega}_i$. Then, defining $\mathbf{1}_i$ as a vector of zeros except for the ith entry which is equal to one we have that

$$\begin{aligned}
\mathbb{E}\left((\omega_i - \widehat{\omega}_i)^2 | X\right) &= \mathbb{E}\left(\omega_i^2 - 2\widehat{\omega}_i\omega_i + \widehat{\omega}_i^2 | X\right) \\
&= \mathbb{E}\left(\omega_i^2 | X\right) - 2x_i \left(X'X\right)^{-1} X'\mathbb{E}\left(\omega\omega_i | X\right) + \mathbb{E}\left(\widehat{\omega}_i^2 | X\right) \\
&= 1 - 2x_i \left(X'X\right)^{-1} X'\mathbf{1}_i + h_{ii} \\
&= 1 - 2h_{ii} + h_{ii} \\
&= 1 - h_{ii}.
\end{aligned}$$

So we can take also a sample mean of the squared residuals to get an estimate for m_{ii}. Let us define the estimated values with their corresponding hat variables, $\widehat{h}_{ii}$, $\widehat{m}_{ii}$. Thus, we have two estimates for the one minus the leverage, $1 - \widehat{h}_{ii}$ and $\widehat{m}_{ii}$. As Kline et al. (2021) mention, the infeasible variance minimizing unbiased linear combination of both estimators is

$$\frac{h_{ii}}{m_{ii} + h_{ii}} \widehat{m}_{ii} + \frac{m_{ii}}{m_{ii} + h_{ii}} \left(1 - \widehat{h}_{ii}\right).$$

The feasible estimator of m_{ii} would then be equal to

$$\bar{m}_{ii} \equiv \frac{\widehat{m}_{ii}}{\widehat{m}_{ii} + \widehat{h}_{ii}},$$

and $\bar{h}_{ii} \equiv 1 - \bar{m}_{ii}$. We then use $\bar{m}_{ii}$ to construct the covariance matrix estimate when using HC_2. We do this by multiplying $1/\bar{m}_{ii}$ to the squared residual of observation i. We also correct for a bias coming from the non-linear estimation of $1/\bar{m}_{ii}$ up to a second order. The expected value of the second-order approximation of $1/m_{ii}$ is

$$\mathbb{E}\left(\frac{1}{\bar{m}_{ii}}\right) \approx \frac{1}{m_{ii}} + \frac{h_{ii}}{m_{ii}^3} \mathbb{E}\left(\widehat{m}_{ii} - m_{ii}\right)^2 - \frac{1}{m_{ii}^2} \left(\mathbb{E}\left((\widehat{h}_{ii} - h_{ii})(\widehat{m}_{ii} - m_{ii})\right)\right).$$

Thus, the final estimate of $1/m_{ii}$ would be

$$\frac{1}{\bar{m}_{ii}} \left(1 - \frac{\bar{h}_{ii}}{\bar{m}_{ii}^2} \widehat{\text{var}}(\widehat{m}_{ii}) + \frac{1}{\bar{m}_{ii}} \widehat{\text{cov}}\left(\widehat{h}_{ii}, \widehat{m}_{ii}\right)\right),$$

where $\widehat{\text{var}}$ and $\widehat{\text{cov}}$ are sample variance and covariance estimates.[6]

Direct computation. Alternatively, an exact computation of the leverage is possible by using the definition of fitted values $\widehat{Y} = HY$ and a regression-intensive procedure. We have that the leverage

[6]The sample variance of $\widehat{m}_{ii}$ is $\frac{1}{n_h - 1}\left(\frac{1}{n_h}\sum_{j=1}^{n_h}\left(\omega_{i,j} - \widehat{\omega}_{i,j}\right)^2 - \widehat{m}_{ii}^2\right)$. The sample covariance is $\frac{1}{n_h - 1}\left(\frac{1}{n_h}\sum_{j=1}^{n_h}\left(\omega_{i,j} - \widehat{\omega}_{i,j}\right)^2 \widehat{\omega}_{i,j}^2 - \widehat{m}_{ii}\widehat{h}_{ii}\right)$.

of observation i is equal to

$$h_{ii} = \frac{\partial \widehat{y}_i}{\partial y_i}.$$

The following remark shows how to compute these leverages without computing the projection matrix H using only linear regressions.

Proposition OA-1. *Let $\tilde{Y}(i)$ be a vector of length n where every entry is equal to zero, except the ith entry that is equal to one. The leverage of observation i is equal to the fitted value $\widehat{y}_i$ of a linear regression of $\tilde{Y}(i)$ on X.*

Proof. Let h_i be the ith row of the projection matrix H. Then, for any vector Y we have that the ith fitted value $\widehat{y}_i$ is equal to $\widehat{y}_i = h_i Y = \sum_j h_{ij} y_j$. Let $Y = \tilde{Y}(i)$. Then $\widehat{y}_i = h_{ii}$. □

Recovering the estimates of a linear regression is very efficient nowadays and in principle we could compute the leverages one by one in what would involve n regressions. When the data set is large, this is clearly not plausible and we leave the exact computation for the problematic cases identified by the following diagnostic.

Diagnostic and adjustment. Although, as mentioned by Kline et al. (2021), the above estimate of m_{ii} rules out nonsensical estimates outside the $[0, 1]$ interval, the estimates for $1/m_{ii}$, could still violate some theoretical bounds. We detect problematic estimations of $1/m_{ii}$ by checking that they are within some bounds that are consistent with the theoretical bounds for the leverages $h_{ii} \in [1/n, 1]$. These bounds are derived from the following proposition, which might be well known for some readers.

Proposition OA-2. *Let X be a full rank matrix of dimensions $n \times k$, where a vector of ones can be obtained through column operations. Let $H = X\left(X'X\right)^{-1}X'$, with ith diagonal element h_{ii}. Then $1/n \leq h_{ii} \leq 1$ for all i.*

Proof. As H is idempotent then $h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2$. Then $h_{ii} \leq h_{ii}^2 \implies h_{ii} \leq 1$.

Now, let $\tilde{X}$ be the full rank matrix of dimensions $n \times k$ that contains a vector of ones after doing column operations on X. Then define $\tilde{H} = \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'$ with diagonal elements $\tilde{h}_{ii}$. It is well known that $1/n \leq \tilde{h}_{ii}$ (see for example Lemma 2.2 in Mohammadi (2016)). As X and $\tilde{X}$ have the same column space, then $H = \tilde{H}$. Thus, $1/n \leq h_{ii}$. □

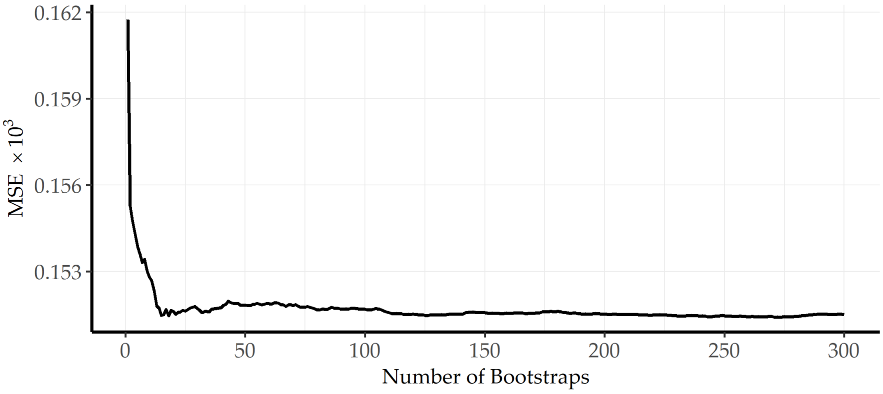
The corollary of the proposition above is that $1/m_{ii} \geq n/(n-1)$. Thus, we check if our estimates of $1/m_{ii}$ satisfy this bound.[7] We directly compute leverages corresponding to the estimates of $1/mii$ that fall outside those bounds by using the result of Remark OA-1.

Algorithm 4 in Section OA-6 of this Online Appendix takes as inputs the covariates X and gives output a combination of actual and estimates for $1/m_{ii}$.

Leave-one-out connected set. Two-way fixed effect models are only identified within a connected set. In typical applications on the labor market or teacher evaluations, firm (school) fixed effects are

[7]When we use any estimate of the covariance matrix that requires calculating $1/(1 - h_{ii})$, we prune the data such that observations with $h_{ii} = 1$ are not in the sample.

Figure OA-2: MSE of corrected $\widehat{\sigma}(\theta, \psi)$ by number of bootstraps.



Notes: This figure presents the mean squared error (MSE) of the covariance between worker-firm fixed effects $\widehat{\sigma}(\theta, \psi)$ across 1000 homoskedastic error simulations. The bootstrap correction assumes a diagonal covariance matrix and we use the HC_1 covariance matrix estimator.

only identified within the connected set that is generated by moving workers (teachers). Movers therefore determine the connected set of firms (schools) whose fixed effect can be identified. When the estimator of a the covariance matrix requires to compute $1/(1-h_{ii})$, as is the case with the HC_2 estimator, then we need to have $h_{ii} < 1$ for all i. In practice a leverage h_{ii} equal to 1 usually means that one single observation identifies a particular fixed effect. For example, when one firm has only one mover, then that worker is key to identify the firm fixed effect and will have a leverage of 1. The leave-one-out connected set requires that no single observation is necessary to estimate a particular fixed effect. That is, after eliminating any observation the set of fixed effects in the connected set needs to remain the same. We achieve this by first pruning the data to get the leave-one-out connected set without critical movers identifying a given firm fixed effect, and eliminating unique observations. The pruning is the same as the one used by Kline et al. (2020). Algorithm 3 in Section OA-6 describes the details.

OA-4 Choosing the number of bootstraps

Some readers might feel uneasy with the arbitrary number of bootstraps necessary to correct the bias. To guide our choice of number of bootstraps, we perform some simulations with a fixed set of covariates with low mobility and simulate a thousand samples by simulating the error. With each dataset we perform corrections from one to 300 bootstraps as in the Monte Carlo simulations of Section 4. Figure OA-2 shows the mean squared error between the true covariance of person and firm fixed effects and the corrected one for different number of bootstraps.[8] The figure shows that with the first 25 bootstraps the MSE reduces significantly and around 100 it flattens. This suggests that few bootstraps are enough to gain accuracy.[9]

In this section we show a way to discipline the choice of the number of bootstraps. We exploit the fact that our estimator δ^* is a sample mean estimate of the direct bias correction term $\widehat{\delta}$. This allows us to exploit the information given by Chebyshev's inequality.

[8] For all the samples we take the corrections obtained with different bootstraps and take the mean squared error against the true moment.

[9] Throughout the application corrections we run corrections with 1000 iterations to estimate the leverage and 1000 bootstraps to estimate the corrections of second order moments.

Let $\delta_j^* \equiv \beta_j^{*\prime} A \beta_j^*$ be the quadratic form for bootstrap j. In the proof for Proposition 2 we show that $\mathbb{E}_{v^*}\left(\delta_j^*|X,u\right) = \widehat{\delta}$. Now assume that $\mathbb{V}(\delta_j^*|X,u) = \eta^2 < \infty$. As δ^* is a sample mean over a sequence of $\{\widehat{\delta}_j^*\}_{j=1}^p$, we have that $\mathbb{E}_{v^*}(\widehat{\delta}^*|X,u) = \widehat{\delta}$ (as shown in the proof of Proposition 2) and $\mathbb{V}(\delta^*|X,u) = \frac{1}{p}\eta^2$.[10] Then, by Chebyshev's inequality we have

$$\mathbf{P}\left(\left|\widehat{\delta}^* - \widehat{\delta}\right| \geq k\frac{\eta}{\sqrt{p}} \;\middle|\; X,u\right) \leq \frac{1}{k^2}.$$

Next one can choose the number of bootstraps p such that the distance between the bootstrap estimate $\widehat{\delta}^*$ and the direct bias correction term $\widehat{\delta}$ is greater or equal than λ standard deviations with probability smaller than α. So, for arbitrary $\alpha > 0$ and $\lambda > 0$ we have

$$\frac{1}{k^2} = \alpha, \quad \frac{k}{\sqrt{p}} = \lambda.$$

Solving for p we get $p = \frac{1}{\alpha\lambda^2}$. So if, for example, we set $\alpha = 0.05$ and $\lambda = 1/2$ we get that the number of bootstraps such that the distance between the bootstrap estimate and the unfeasible correction term is greater than half a standard deviation is an event with a probability smaller than 5 percent is $p = \frac{1}{0.05\times(1/2)^2} = 20 \times 4 = 80$. One could be more conservative and set $\lambda = 0.1$. In that case, we would obtain $p = 20 \times 1000 = 2000$ bootstraps.

Admittedly, the number of bootstraps suggested by inequality for any α and λ can be quite conservative. But this just reflects the generality of the result. Indeed, this criteria would work regardless the distribution of v^*, therefore regardless the choice of bootstrap.

OA-5 Additional error structures

While the method proposed by KSS can also be adapted to *"settings where the data are organized into mutually and independent 'clusters'"* (page 1863 of Kline et al. (2020)), our method can also accomodate more general settings as the example below explains.

Example: Non-block-diagonal covariance matrices. Sometimes, the assumption on the error dependence do not yield a diagonal or block-diagonal covariance matrix. This can happen when there are two (or more) dimensions of dependency. For example, when there are temporal and spatial dependencies, as in Driscoll and Kraay (1998). In the AKM context, for example, there would be a non-block-diagonal covariance matrix if there is temporal dependence at the person *and* firm dimensions. With workers changing firms, the resulting dependence across observations would break any block-diagonal structure in the covariance matrix.

OA-6 Algorithms

In this Section we detail the implementation algorithms of our method. Algorithm 1 and 2 describe, respectively, the estimation of the bias correction for diagonal and non diagonal covariance matrices.

[10] We have that $\mathbb{V}(\delta^*|X,u) = \frac{1}{p^2}\mathbb{V}(\sum_j^p \widehat{\delta}_j^*|X,u) = \frac{1}{p^2}\sum_j^p \mathbb{V}(\widehat{\delta}_j^*|X,u) = \frac{1}{p}\eta^2$ where we used the independence of different $\widehat{\delta}_j^*$ conditional on X and u.

Algorithm 3 describes how to prune the data to ensure that the maximum leverage is below 1 and Algorithm 4 details how to estimate the leverage.

To speed the solution of the normal equations, we use the preconditioner developed by Koutis, Miller, and Tolliver (2011) when the matrix of covariates accepts a Laplacian representation. In our setup, the Laplacian representation is possible when we only have two leading fixed effects or the rest of the covariates are residualized previously.

Notation. For a number of moments to correct M (for example a variance decomposition of a two-way fixed effect model has at least three corrections: the two variances of the fixed effects and their covariance), the bias correction of the mth moment $m \in \{1, ..., M\}$ is denoted as $\widehat{\delta}^*_m$.

Algorithm 1 Estimate $\{\widehat{\delta}^*_m\}_{m=1}^M$ when the covariance matrix is diagonal

---

for $j = 1, ..., p$ **do**
  Simulate a vector r^* of length n of mutually independent Rademacher entries.
  Generate a vector of residuals v^* of length n whose ith entry is equal to $\sqrt{\widehat{\psi}_i} \times r^*_i$.
  Compute β^* as the estimate of a regression of v^* on X.
  Compute $\widehat{\delta}^{(j)}_{aux,m} = (\beta^*)' A_m \beta^*$ for all $m \in \{1, ..., M\}$.
end for
Compute $\widehat{\delta}^*_m = \dfrac{\sum_{j=1}^p \widehat{\delta}^{(j)}_{aux,m}}{p}$ for all $m \in \{1, ..., M\}$.

Algorithm 2 Estimate $\{\widehat{\delta}^*_m\}_{m=1}^M$ when covariance matrix is non diagonal

---

Let $\mathbb{G} = \{1, ..., G\}$ be the set of groups g each with length n_g.
for $j = 1, ..., p$ **do**
  Simulate a vector r^*_g of length G of mutually independent Rademacher entries. All the observations withing the group will have the same Rademacher entry.
  Generate a vector of residuals v^* of length n whose ith entry belonging to group g is equal to $\sqrt{\widehat{\psi}_i} \times r^*_g$.
  Compute β^* as the estimate of a regression of v^* on X.
  Compute $\widehat{\delta}^{(j)}_{aux,m} = (\beta^*)' A_m \beta^*$ for all $m \in \{1, ..., M\}$.
end for
Compute $\widehat{\delta}^*_m = \dfrac{\sum_{j=1}^p \widehat{\delta}^{(j)}_{aux,m}}{p}$ for all $m \in \{1, ..., M\}$.

Algorithm 3 Leave-one-out connected set

Let Λ be the connected set.
$a = 1$.
while $a > 0$ **do**
  Compute the articulation points a.
  Eliminate articulation points a.
  Compute the new connected set Λ_1.
end while

Algorithm 4 Estimate leverages, diagnosis and compute those out of bounds

$z_h^{(0)} = \mathbf{0}$, $z_m^{(0)} = \mathbf{0}$, $z_2^{(0)} = \mathbf{0}$, and $z_{hm}^{(0)} = \mathbf{0}$ are vectors of length n.
for $j = 1, ..., p$ **do**
    Simulate a vector ω^* of length n of mutually independent Rademacher entries.
    Compute fitted values $\widehat{\omega^*}$ from a regression of ω^* on X.
    Compute $z_h^{(j)} = z_h^{(j-1)} + \left(\widehat{\omega^*}\right)^2$ and $z_m^{(j)} = z_m^{(j-1)} + \left(\widehat{\omega^*} - \omega^*\right)^2$.
    Compute $z_2^{(j)} = z_2^{(j-1)} + \left(\widehat{\omega^*} - \omega^*\right)^4$ and $z_{hm}^{(j)} = z_{hm}^{(j-1)} + \left(\widehat{\omega^*} - \omega^*\right)^2 \left(\widehat{\omega^*}\right)^2$
end for
Compute $\widehat{h}_{ii} = z_{h,i}^{(p)}/p$ and $\widehat{m}_{ii} = z_{m,i}^{(p)}/p$ for all $i \in \{1, ..., n\}$.
Compute $\widehat{\text{var}}(\widehat{m}_{ii}) = \frac{1}{p-1}\left(\frac{z_{m,i}^{(p)}}{p} - \widehat{m}_{ii}^2\right)$ for all $i \in \{1, ..., n\}$.
Compute $\widehat{\text{cov}}(\widehat{h}_{ii}, \widehat{m}_{ii}) = \frac{1}{p-1}\left(\frac{z_{hm,i}^{(p)}}{p} - \widehat{h}_{ii}\widehat{m}_{ii}\right)$ for all $i \in \{1, ..., n\}$.
Compute $\bar{m}_{ii} = \frac{\widehat{m}_{ii}}{\widehat{m}_{ii} + \widehat{h}_{ii}}$ for all $i \in \{1, ..., n\}$.
for $i = 1, ..., n$ **do**
    if $\frac{1}{\bar{m}_{ii}}\left(1 - \frac{\bar{h}_{ii}}{\bar{m}_{ii}^2}\widehat{\text{var}}(\widehat{m}_{ii}) + \frac{1}{\bar{m}_{ii}}\widehat{\text{cov}}\left(\widehat{h}_{ii}, \widehat{m}_{ii}\right)\right) \leq \frac{n}{n-1}$ **then**
        Generate $\tilde{Y}(i) \in \mathbb{R}^n$, where $\tilde{Y}(i)_{j \neq i} = 0$, $\tilde{Y}(i)_i = 1$.
        Compute the fitted values $\widehat{\tilde{Y}}(i)$ of a regression of $\tilde{Y}(i)$ on X.
        Get actual leverage $h_{ii} = \widehat{\tilde{Y}}(i)_i$.
        Get actual $1/m_{ii} = 1/(1 - h_{ii})$.
    end if
end for

OA-7 Comparison with Borovičková and Shimer (2017)

Borovičková and Shimer (2017) (henceforth BS) provide an alternative method to compute the correlation of firm types and workers, which has the advantage of not requiring estimates of all the worker and firm fixed effects and directly computing the correlation. Their method completely bypasses the need to estimate a linear model and therefore avoids using noisy estimates—which are the source of the bias—to compute the correlation.

As explained by BS, the worker and firm types that they define are different to the types defined in the AKM model. In BS, a worker's type, denoted λ_i, is defined to be the expected log wage of the worker, while a firm's type, denoted $\mu_{J(i,t)}$, is defined to be the expected log wage that a firm pays. In contrast, in the AKM model, a worker and firm types $(\theta_i, \psi_{J(i,t)})$ are defined as such that a change in type will change the expected log wage while holding fixed the partner's type.[11]

BS show that their correlation and the AKM correlation, ρ, will be the same if the joint distribution of θ and ψ is elliptical (e.g. a bivariate normal) and $(\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) > 0$, where σ_λ and σ_μ are, respectively, the standard deviations of worker and firm types. With these assumptions, there is also a direct correspondance between the standard deviation of AKM types and BS types:[12]

$$\sigma_\theta = \frac{\sigma_\lambda - \rho\sigma_\mu}{1 - \rho^2}, \quad \sigma_\psi = \frac{\sigma_\mu - \rho\sigma_\lambda}{1 - \rho^2}.$$

[11] We refer to an old version of the Borovičková and Shimer from 2017 where they provide a way to translate the variances and covariances of their worker and firm types to the ones in AKM. In the latest version, they slightly changed their estimator and no longer provide this link.

[12] See Proposition 1 in Borovičková and Shimer (2017).

Table OA-2: Monte Carlo simulations. Homoscedastic errors.

	Time	Mean Squared Error (MSE×10^2)			
		$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.637	0.341	0.114	2.364
BS	0.1	1.580	0.615	0.040	0.745
Gaure	17.3	0.050	0.109	0.015	0.058
Boot HC2	0.9	0.050	0.106	0.014	0.057
Boot HCU	1.0	0.050	0.106	0.014	0.057
KSS	0.9	0.050	0.106	0.014	0.057

Notes: *Plug-in* is the naive plug-in estimator, *BS* refers to Borovičková and Shimer (2017), *Gaure* refers to the method Gaure (2014) implemented through the R package *lfe*, *Boot* is our method with HC_2 covariance matrix estimator, and *KSS* is the Kline et al. (2020) method. The results of Borovičková and Shimer correspond to the AKM worker and firm types present in the cited version of the paper. The average firm has 10 movers and 12 employees. *Time* is the computing time in seconds. True moments are computed at the final sample for each method, i.e. largest connected set for *Gaure* and the largest leave-one-out connected set for *Boot* and *KSS*. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors (MSE) of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled).

The key identifying assumption in the BS method is that for each worker and firm they have two or more observations of the wage which are independent and identically distributed conditional on the types. In AKM, the identifying assumption is a standard exclusion restriction, i.e. that the error term is mean zero conditional on the types (and other covariates) with the underlying assumption of exogenous mobility.

OA-7.1 Comparison of Methods

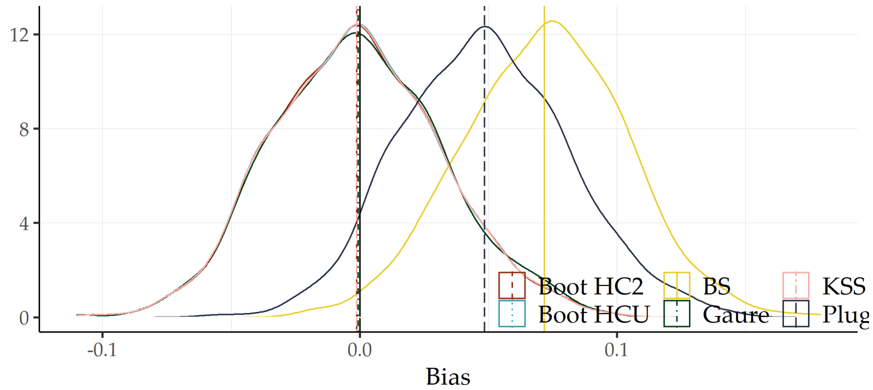
We perform two exercises to compare our method with BS. First, we simulate labor market data that fulfills the key identifying assumptions of the AKM linear model and of BS. We find that both methods correct the bias but ours outperforms theirs in terms of accuracy of the estimation of each of the elements of the correlation, but is naturally more time consuming. Second, we apply BS method to the French data which requires some changes to the original dataset in the sample selection, which we explain in more detail below.

The results of the comparison using simulated data are in Table OA-2. For completeness we also include Gaure and KSS's methods in the comparison. The table shows that the least accurate method is BS reducing by 56% the MSE of the naive estimates whereas the other three methods reduce it by 98%.[13] The objective of BS is to provide an estimate of the correlation but they base their estimation in different worker and firm types (λ and μ respectively). Table 1 presents their estimates of the corresponding AKM moments. Figure OA-3 shows the distribution of the difference of the firm variances for the plug-in estimate and the true variance ($\hat{\sigma}^2_{\psi,PI} - \sigma^2_\psi$), as well as the the distributions of the differences using the different correction methods. The figure shows that our method is very similar to KSS and both are the ones with lowest biases. Even if the bias of Gaure is higher, his method has lower variance and outperforms KSS and ours in terms of MSE. Regarding the computation time, BS is the fastest one with computation time of less than a second. Our method is the one performing the fastest among the AKM based competitors (Gaure, KSS and our method).[14]

[13] We wrote the code for BS following Borovičková and Shimer (2017) and converting the data to the match level.

[14] KSS and our method do not incorporate the simplifications that come from having homoscedastic errors. In particular, under homoscedasticity of the errors, one could gain speed by using the covariance estimate HC_1 which is unbiased, and therefore skip the pruning of the data and the leverage estimation.

Figure OA-3: Model Comparison: Homoscedastic Errors.



Notes: This figure presents the distributions of the bias of $\widehat{\sigma}^2_{\psi}$ for the naive plug-in estimate and the corrected moments for the different methods. Simulated errors are homoscedastic and labor mobility is high.

Table OA-3: Application. Comparison of the Methods.

	Plug-in	Boot HC_2	Boot HC_U	KSS	Boot Serial	KSS Serial
$\widehat{\sigma}^2_{\theta}$	0.179	0.157	0.159	0.157	0.130	0.137
$\widehat{\sigma}^2_{\psi}$	0.042	0.029	0.032	0.030	0.048	0.021
$\widehat{\sigma}_{\theta,\psi}$	-0.005	0.005	0.003	0.005	-0.005	0.011
$\widehat{\rho}_{\theta,\psi}$	-0.061	0.077	0.039	0.069	-0.061	0.202
Obs.	4652631	4652631	4652631	4652631	8075514	4652631
Time (min)		24	35	36	38	29

Notes: *Plug-in* are the plug-in estimates at the leave-one-out connected set, *Boot* HC_2 are the results of our method under diagonal covariance matrix estimator HC_2 at the leave-one-out connected set, *Boot Serial (Connected)* are the results using a Wild block bootstrap at the connected set, *Boot Serial* are the results using a Wild block bootstrap at the leave-one-out connected set like KSS, *KSS* are the results corrected with Kline et al. at the leave-one-out connected set similarly to *Boot* HC_2 and *KSS Serial* are the results at the leave-one-out connected set when leaving a match out. $\widehat{\sigma}^2_{\theta}$ and $\widehat{\sigma}^2_{\psi}$ are respectively the estimates of the variance of worker and firm fixed effects. $\widehat{\sigma}_{\psi,\theta}$ is the covariance, $\widehat{\rho}_{\psi,\theta}$ the correlation between worker and firm fixed effects, *Obs.* is the number of observations and *Time (min)* is the correction time in minutes.

Now, we compare BS method using the French data with our method as well as with KSS's method. In order to do so, we need to deviate in two aspects from the original sample used in our main application: first, we need to restrict the sample to workers that have at least two jobs and firms that have at least two workers; second, we need to take averages of every match between firm and workers.[15] The first restriction implies that the sample used for BS is about half of the original sample of private firms.[16] Suggestive of the potential sample selection issues is that the plug-in estimate of the correlation between worker and firm fixed effects is -0.10 under the original data whereas is -0.06 under the connected set generated from BS data.

We compare our method to in the application by first residualizing other covariates and we then estimate the variances of worker and firm fixed effects together with their covariance and correlation. The results are in Table OA-3.

In the application, our bootstrap method is slightly faster than the KSS correction when assuming a diagonal covariance matrix but is slower when assuming serial correlation. Both methods yield slightly different estimates of second order moments with HC_2 and HC_U that also lead to

[15]More precisely this would mean that if we observe one worker employed for a certain firms for several years, we would take the average wage of that worker in that firm as one observation.

[16]The original data of private firms has 5.8 million observations while after filtering of two job and worker restrictions the sample has only 2.5 million observations.

Table OA-4: Application. Extended Comparison of the Methods (BS Data).

	BS	Plug-in	Boot HC_2	Boot HC_U	KSS
$\widehat{\sigma}^2_\theta$	0.037	0.111	0.058	0.058	0.058
$\widehat{\sigma}^2_\psi$	0.014	0.070	0.040	0.041	0.041
$\widehat{\sigma}_{\theta,\psi}$	0.013	-0.008	0.004	0.005	0.005
$\widehat{\rho}_{\theta,\psi}$	0.570	-0.093	0.092	0.093	0.094
Obs.	1623985	1610912	1610912	1610912	1610912

Notes: The results of *BS* correspond to the AKM worker and firm types of Borovičková and Shimer. *Plug-in* are the plug-in estimates at the connected set originated from BS data, *Boot* HC_1 are the results of our method under diagonal covariance matrix estimator HC_1 at the connected set originated in the BS data, *Boot* HC_2 are the results of our method under diagonal covariance matrix estimator HC_2 at the leave-one-out connected set in the BS data and *KSS* are the results corrected with the method of Kline et al. at the same sample as for *Boot* HC_2. $\widehat{\sigma}^2_\theta$ and $\widehat{\sigma}^2_\psi$ are respectively the estimates of the variance of worker and firm fixed effects. $\widehat{\sigma}_{\psi,\theta}$ is the covariance, $\widehat{\rho}_{\psi,\theta}$ the correlation between worker and firm fixed effects and *Obs.* is the number of observations.

differences in the estimate of the correlation between firm and worker fixed effects. When assuming a diagonal covariance matrix with heteroskedastic errors (*Boot* HC_2, *Boot* HC_U and *KSS*), both methods yield a positive estimate for the correlation between firm and worker fixed effects. The estimate of this correlation with the bootstrap correction is 0.077 with HC_2 and 0.039 with HC_U while the one of KSS is 0.069.[17]

When assuming that the error terms are correlated at the match level, the estimated correlation between worker and firm fixed effects is positive with KSS under the leave-one-out connected set. On the contrary, *Boot Serial* is estimated in the connected set and yields a negative estimate for the worker-firm correlation of -0.061. We can conclude that, even after correcting for the limited-mobility bias, the estimates of the correlation between workers and firms fixed effects are sensitive to sample selection. The leave-one-out connected set is comprised of more mobile workers who could have a different sorting pattern than the rest of the labor force. Thus, it could be that the suggested small, yet positive correlation, is driven by those workers who change jobs more frequently.

We also compare our method to BS first residualizing for other covariates. We use the averaged match-level residual wage as the dependent variable to compute the moments for all the methods. We estimate the bootstrap corrected moments at the connected set or leave-one-out-corrected set of the BS final sample.

Table OA-4 compares the estimated moments using the BS method and the bootstrap correction method on French data. Both columns report the moments using the AKM defined worker and firm types. In contrast to the Monte Carlo simulations that satisfied the assumptions for both methods, estimates differ by a large amount when using French labor market data. The bootstrap corrected estimated correlation is 0.16 (0.09) under HC_2 (HC_1) covariance matrix estimation, well below the estimated one using BS method, 0.56.[18] Looking at each of the components of the correlation, both variances are larger and the covariance is smaller when using the bootstrap corrected method instead of BS method.

There are different reasons why BS estimates might differ from ours. To begin with, the types defined by BS are fundamentally different from the ones defined in the AKM model. They are related only when the assumptions stated at the beginning of this section are satisfied. It might be

[17]The results from *Boot* HC_U are analogous to the ones in Table 4 but computed with residualized log wages.
[18]The BS estimates are obtained by using the formulas of Section 5.2. in Borovičková and Shimer (2017).

Table OA-5: Application. Summary Statistics.

BS Data	Obs.	Mean Wage	Mean Age	Mean Education
0	4096250	4.40	41.73	4.29
1	4708713	4.39	37.24	4.57

Notes: *BS Data* is an indicator if the observation belongs to the final sample of Borovičková and Shimer (2017), *Obs.* is the number of observations before taking match level averages in the original data and before computing the connected set, *Mean Wage* is the average log daily wage, *Mean Age* is the average age in years and *Mean Education* is the average education where education is a discrete variable between 1 (no education) and 8 (university degree).

that the two correlations are not comparable because, even if the log-linear AKM model is correctly specified, these assumptions are violated, in particular, if the joint distribution of AKM types is not elliptical. Second, it might be that the identification assumption of at least one of the methods fail. It is easy to think of examples where *both* identification assumptions are violated. For example, whenever there is selection of workers via the error term, some matches will be formed whenever this idiosyncratic component is high. This endogenous mobility would violate both the AKM and BS identification assumptions.

Results in Table OA-4 under our method also differ from the ones previously reported in Table 4 in the main text. Table OA-5 presents some summary statistics of the original data differentiated by being in the final BS data or not.[19] The Table shows that the requirements to use the Borovičková and Shimer (2017) method are more demanding as only 77% of the original observations are included in their final sample. Furthermore, Table OA-5 shows that their data requirements lead to a sample with similar average wage but almost 5 years younger on average and slightly more educated. The applied user might be worried by sample selection when using the BS method to estimate worker and firm correlation as Lentz, Piyapromdee, and Robin (2018) document that most of the worker-firm sorting happens early in the career which would lead to higher correlations for younger workers.

OA-8 Additional Results and Proofs

The following proposition gives conditions under where our bootstrap estimate is more efficient than the one proposed by MacKinnon and Smith Jr (1998) (MS). The proposition proofs that a covariance is zero. When that is the case, the variance of the bias correction of MS is strictly larger than the one from our bias correction as shown by equation 9 in the main text.

Proposition OA-3. *Let X and u be the exogenous covariates and the error term in the original model. Let v_i^* be the bootstrap residual for observation i. These are independent across observations with $\mathbb{E}\left(v_i^* \mid X, u\right) = 0$, $\mathbb{E}\left((v_i^*)^2 \mid X, u\right) = \psi_i$, and $\mathbb{E}\left((v_i^*)^3 \mid X, u\right) = 0$. Let $Q = \left(X'X\right)^{-1} X'$ and A independent of v^*, conditional on X and u. Then,*

$$cov\left(v_i^{*\prime} Q' A Q v_i^*, 2 v_i^{*\prime} Q' A \widehat{\beta} | X, u\right) = 0.$$

Proof. Let the matrix $Q'AQ \equiv R$, with elements (i, j) equal to $r_{i,j}$. Also, let the vector $Q'A\widehat{\beta} \equiv S$

[19] The original data constitutes of almost 5.9 million observations that translate into a connected set of 5.1 million observations as in Table 4.

Table OA-6: Monte Carlo simulations with a larger sample. Heteroscedastic errors.

		Mean Squared Error (MSE$\times 10^3$)			
Model	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		20.366	6.059	5.280	10.569
Boot HC2	398.2	0.001	0.003	0.001	0.002
Boot HCU	582.6	0.001	0.003	0.001	0.002
KSS	595.9	0.001	0.003	0.001	0.002

Notes: We simulate a labor market with a connected set similar to the one we use in the application with more than 5 million observations. *Plug-in* is the naive plug-in estimator, *Boot HC2* refers to our method with HC_2 covariance matrix estimator, *Boot HCU* implements it with HC_U, and *KSS* is the Kline et al. (2020) method. True moments are computed at the leave-one-out connected set. In all the exercises the number of movers per firm is 3 and the average firm has 12 employees. *Time* is the computing time in seconds. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 1000 due to high accuracy of the corrections. *Average* is the average MSE (also scaled).

with element k equal to s_k. Then,

$$cov\left((v_i^{*\prime} R v_j^*, 2v_i^{*\prime} S | X, u\right) = \mathbb{E}\left(\left(\sum_{i=1}^n \sum_{j=1}^n r_{i,j} v_i^* v_j^*\right)\left(\sum_{k=1}^n s_k v_k^*\right) \,\middle|\, X, u\right),$$

where we use the fact that $\mathbb{E}\left(v_i^* \mid X, u\right) = 0$. Then,

$$\mathbb{E}\left(\left(\sum_{i=1}^n \sum_{j=1}^n r_{i,j} v_i^* v_j^*\right)\left(\sum_{k=1}^n s_k v_k^*\right) \,\middle|\, X, u\right) = \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n r_{i,j} s_k \mathbb{E}\left(v_i^* v_j^* v_k^* \,\middle|\, X, u\right)\right) = 0,$$

where we use that the bootstrap errors are independent across observations and the fact that $\mathbb{E}\left((v_i^*)^3 \mid X, u\right) = 0$. □

OA-9 Additional Tables and Figures

Table OA-6 does the same exercise as the *Low Mobility* part of Table 2 in the main text in a more realistic sample size of roughly 5 million observations. Table OA-7 compares the MSE for the different moments when using different assumptions on the covariance matrix estimators applicable with our bootstrap method. The original error terms in the simulation were heteroscedastic. As expected, all the corrections effectively reduce the MSE compared to the baseline regardless of the covariance matrix estimator. However, HC_2 outperforms the rest.

Table OA-8 present the Monte Carlo simulation results for serially correlated error terms when the true innovation is heteroscedastic. Figure OA-4 complements Table 2 from the main text and shows the distribution of the corrections in Monte Carlo simulations when the error terms are heteroscedastic. Table OA-3 compares the bootstrap correction to the KSS correction in the French application. Finally, Table OA-9 shows a correction of the variance decomposition similar when using a Wild block bootstrap. Different to Table 4 in the main text, the estimated correlation with the plugin estimates is -0.24 (as opposed to -0.06 estimated with the plugin in the leave-one-out connected set) and the correlation after the correction is estimated to be -0.06.

Table OA-7: Comparison of variance estimators. Heteroscedastic errors

	Mean Squared Error (MSE×10^2)			
Model	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in	25.199	2.922	9.674	12.598
Boot HC_0	3.397	0.740	2.334	2.157
Boot HC_1	0.801	1.300	1.104	1.068
Boot HC_2	0.220	0.681	0.211	0.371
Boot HC_U	0.278	0.733	0.262	0.424

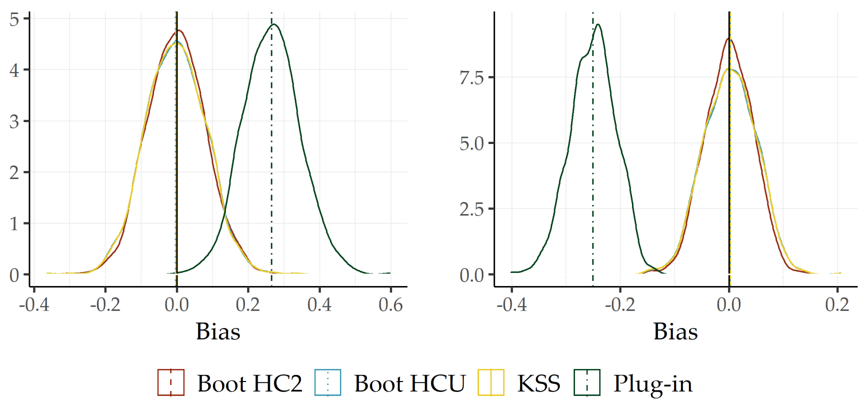
Notes: The original errors in the simulation exhibit heteroscedastic errors. *Plug-in* is the naive plug-in estimator, *Boot* refers to our method. True moments are computed at the largest leave-one-out connected set to make results comparable. *Model* is the model and type of variance estimator. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors of the estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled). Simulated data exhibits low mobility like in the top panel of Table 2 and all the estimations are done using the leave-one-out sample.

Table OA-8: Monte Carlo simulations. Serial correlation with heteroscedasticity.

		Mean Squared Error (MSE×10^2)			
	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		93.900	1.658	0.603	32.053
Boot	0.5	9.528	0.259	0.047	3.278
KSS	1.1	21.350	0.251	0.045	7.215

Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method with a wild block bootstrap where each match defines a block and we skip the pruning of the data. *KSS* is the Kline et al. (2020) method leaving a match out. The average firm has 10 movers and 12 employees. *Time* is the computing time in seconds. True moments are computed at the largest connected set for *Boot* and at the largest leave-one-out connected set for *KSS*. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors (MSE) multiplied by 100 of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. *Average* is the average MSE (also scaled).

Figure OA-4: Model Comparison: Heteroscedastic Errors. Bias of $\widehat{\sigma}^2_\psi$ and $\widehat{\sigma}_{\theta,\psi}$



Notes: These figures present the distributions of the bias for the naive plug-in estimate and the bias of corrected moments for *KSS* and our method. Simulated errors are heteroscedastic and labor mobility is low.

Table OA-9: Application. Plug-in vs corrected decomposition.

	Plug-in	Boot Serial		Plug-in	Boot Serial
$Var(y)$	0.3003	0.3003	$2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.0622	-0.0096
$Var(\widehat{\theta}_i)$	0.1880	0.1298	$2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$	-0.0095	-0.0095
$Var(\widehat{\psi}_j)$	0.0856	0.0477	$2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$	-0.0000	-0.0000
$Var(\mathbf{q}\widehat{\gamma})$	0.0171	0.0171	$Corr(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.2453	-0.0609
$Var(\widehat{\epsilon})$	0.0814	0.1248	Obs.	8075514	8075514

Notes: *Plug-in* refers to the uncorrected estimates of each of the variance components at the largest connected set and *Boot Serial* refers to the estimates after our bootstrapped correction using a wild block bootstrap. $Var(y)$ is the variance of log wages, $Var(\widehat{\theta}_i)$ the variance of worker fixed effects (naive $\hat{\sigma}^2_\theta$ or corrected $\tilde{\sigma}^2_\theta$), $Var(\widehat{\psi}_j)$ is the variance of firm fixed effects, $Var(\mathbf{q}\widehat{\gamma})$ is the variance of other covariates and $Var(\widehat{\epsilon})$ is the variance of the error term. The other terms of the decomposition are twice the covariances between the fixed effects and the covariates ($2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$, $2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$ and $2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$). Finally, $Corr(\widehat{\theta}_i, \widehat{\psi}_j)$ is the estimated correlation between worker and firm fixed effects and *Obs*. is the number of observations.

References

Borovičková, K. and R. Shimer (2017): "High wage workers work for high wage firms," Tech. rep., National Bureau of Economic Research.

Driscoll, J. C. and A. C. Kraay (1998): "Consistent covariance matrix estimation with spatially dependent panel data," *Review of Economics and Statistics*, 80, 549–560.

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

Hayashi, F. (2000): "Econometrics. 2000," *Princeton University Press. Section*, 1, 60–69.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

——— (2021): "Improved stochastic approximation of regression leverages for bias correction of variance components," Tech. rep.

Koutis, I., G. L. Miller, and D. Tolliver (2011): "Combinatorial preconditioners and multilevel solvers for problems in computer vision and image processing," *Computer Vision and Image Understanding*, 115, 1638–1646.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2018): "On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data," *PIER Discussion Papers*, 91.

MacKinnon, J. G. and A. A. Smith Jr (1998): "Approximate bias correction in econometrics," *Journal of Econometrics*, 85, 205–230.

Mohammadi, M. (2016): "On the bounds for diagonal and off-diagonal elements of the hat matrix in the linear regression model," *REVSTAT Statistical Journal*, 14, 75–87.