Discussion Paper Series – CRC TR 224

Discussion Paper No. 376
Project B 06

Correcting Small Sample Bias in Linear Models With Many Covariates

Miren Azkarate-Askasua [1]
Miguel Zerecero [2]

November 2022
(First version: May 2018)

[1] University of Mannheim, Email : azkarate-askasua@uni-mannheim.de
[2] University of California, Email : mzerecer@uci.edu

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Correcting Small Sample Bias in Linear Models with Many Covariates

Miren Azkarate-Askasua Miguel Zerecero*

November 10, 2022

Abstract

Estimations of quadratic forms in the parameters of linear models exhibit small-sample bias. 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 different assumptions on the structure of the error term including general heteroscedasticity and serial correlation. 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 apply our method by carrying out a variance decomposition of a linear model of log wages with person and firm fixed effects. We find that the person and firm effects are less important in explaining the variance of log wages after correcting for the bias and depending on the specification their correlation becomes positive after the correction.

JEL Codes: C13, C23, C55, J30, J31

Keywords: Variance components, many regressors, fixed effects, bias correction.

*Azkarate-Askasua: University of Mannheim (azkarate-askasua@uni-mannheim.de); Zerecero: University of California, Irvine (mzerecer@uci.edu). First version: May 2018. 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.

1 Introduction

With the increased availability of large panel data sets, researchers have been interested in understanding to what extent unobserved heterogeneity can explain the variation of an outcome of interest. Usually, econometricians include fixed effects in a standard linear model to control for this unobserved heterogeneity and then perform a variance decomposition. These methods have been used in the context of education to study the importance of classroom effects (e.g. Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011)) and extensively in the labor market context where log-additive models of wages are used to study the determinants of labor income and sources of wage inequality (e.g. Abowd, Kramarz, and Margolis (1999); Card, Heining, and Kline (2013); Iranzo, Schivardi, and Tosetti (2008); Lopes de Melo (2018)).

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 empirical 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, i.e. the number of fixed effects, grows with the sample size.

Focusing on the context of labor economics, 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.

Andrews, Gill, Schank, and Upward (2008) derive formulas for correcting the bias when the errors are homoscedastic. 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][2]

In this paper we propose a bootstrap method to correct for small-sample bias in quadratic forms in the estimated parameters of linear models with a large number of covariates. The main application of the method is the correction of variance decompositions of multi-way fixed effects models. The contribution of the paper is to provide an easy-to-implement and fast alternative to the bias correction methods for quadratic forms present in the literature. We see our paper as an application of the bootstrap in variance decompositions of multi-way fixed effects.

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

[1] By large matrix we mean a matrix with dimension in the order of hundreds of thousands or millions.

[2] Some recent 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).

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 heteroscedasticity, 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 and heteroscedasticity. 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 approach is similar to the ones proposed by Gaure (2014) and by Kline, Saggio, and Sølvsten (2020). The bias appears as the trace of a matrix, but 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 heteroscedasticity 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 faster and more flexible. One drawback is that we can not use their leave-one-out covariance-matrix estimate as its not suitable for bootstraps.[5] However, our bootstrap procedure can accommodate easily more complicated variance structures. In terms of speed, Monte Carlo simulations show that our correction takes between 56% to 70% of the computing time of KSS, and has similar accuracy.

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 one system 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.[6] They implement corrections of the

[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 $\frac{1}{M}\sum_{i=1}^{M} x_i'Ax_i$, 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 homoscedastic. 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 heteroscedastic, but to the best of our knowledge he does not implement it in his R package.

[5]In the heteroscedastic case, when using their leave-one-out procedure, some diagonal elements of the estimated covariance-matrix could be negative. In practice, we can not implement a bootstrap procedure with such covariance matrix estimate as we need to take the squared root of each diagonal element.

[6]For example, 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.*

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 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.[7]

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.[8] 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,

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

Using the general formula for the expectation of quadratic forms, the exclusion restriction $\mathbb{E}(u|X) =$

[7] Abowd et al. (2004), also using French data but a different sample, found a correlation of -0.28.
[8] We follow loosely the notation in Kline et al. (2020) for the interested reader to compare the papers.

0, and $\mathbb{E}(A|X) = A$ we obtain,[9]

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

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, which is equal to

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

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.[10] 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}.[11] 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

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

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

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

[10] 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.

[11] 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

$$A_1 = \frac{1}{n-1}\begin{pmatrix} X_1' M_1 X_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \qquad A_2 = \frac{1}{n-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & X_2' M_1 X_2 \end{pmatrix}, \qquad A_{12} = \frac{1}{2(n-1)}\begin{pmatrix} \mathbf{0} & X_1' M_1 X_2 \\ X_2' M_1 X_1 & \mathbf{0} \end{pmatrix}.$$

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. Define β_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*

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

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 β_j^* for each bootstrap.[12] The advantage of our method is twofold. First, we can correct several moments simultaneously, without increasing the computational time. For example, assume we are interested in doing a variance decomposition exercise for each year in the sample of an estimated linear model. Then, we would need to do a correction for the variances of each group of covariates and the covariance term for *every* year but do the bootstrap only one time. Second, to estimate β_j^* in every iteration one just needs to solve for

[12] 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.

a least squares regression. There are extremely efficient procedures to compute these regressions, especially in cases where the high dimensionality of the covariates is the result of a large number of fixed effects.

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.

Example 1: Homoscedasticity. When the errors are homoscedastic, 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 homoscedastic 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. Heteroscedasticity. Another popular assumption is when 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

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

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 variables, meaning they 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 contains the specific algorithms with the steps for this procedure.

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

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).

Recently, Kline et al. (2020) and Jochmans (2018) have proposed the following unbiased estimator of the ith conditional variance:[16]

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

In practice, when estimating $\widehat{\psi}_i$ with HC_U, some of the estimates are negative. This would prevent us from taking the square root of $\widehat{\psi}_i$, which is needed in the bootstrap algorithm.[17] 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 KSS correction would yield a more efficient bias corrected estimate of the quadratic forms compared to our bootstrap method. In fact, we show in the simulation exercises below that our method is in general more efficient in terms of mean squared errors.

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.[18] One particular example is when the group is a worker-firm match and errors are autocorrelated within

[14] A particular case where the estimate is unbiased is when using HC_1 and the error terms are homoscedastic.

[15] A textbook exposition of these issues can be found in Chapter 8 of Angrist and Pischke (2008).

[16] See page 1862 of Kline et al. (2020).

[17] Negative estimates of individual variances are also prevalent in KSS. However, the way they approximate the bias does not require to take the square root of $\widehat{\psi}_i$.

[18] Assume that the the errors have a first order autocorrelation within group g and the true innovations are i.i.d. and therefore homoscedastic. 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 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 or cluster dependence of the data. It has the benefit of accommodating any structure of the dependence within group and also heteroscedasticity of the true innovations. The Online Appendix describes the algorithm to implement our bias correction that keeps the dependence structure among groups through a wild block bootstrap.

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 4: 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 dependency, 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.

In short, our method can be applied whenever one can estimate a covariance matrix with positive diagonal entries. Then, the dependent variable used in the bootstrap can be generated by simulating an error vector with zero mean and a covariance matrix equal to the estimated one.

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.[19] 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 $\beta^{*\prime}_{p,MS} A \beta^*_{p,MS}$ for each bootstrap p and use them to create a bias correction equal to

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

second group two observations. Then, $\mathbb{V}(u|X)$ is:

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

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

[19] 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 $\widehat{\beta}$.

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, $\beta^{*\prime}_{j,MS} A \beta^*_{j,MS} = \widehat{\beta}' A \widehat{\beta} + (v^*_j)' Q' A Q v^*_j + 2 v^{*\prime}_j Q' A \widehat{\beta}$. 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.[20] Thus we have that the variance of their bias correction conditional on X and u is

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

The expression above can be rewritten as

$$\mathbb{V}(\delta^*_{MS}|X,u) = \mathbb{V}(\delta^*|X,u) + \frac{4}{p}\mathbb{V}(v^{*\prime}Q'A\widehat{\beta}|X,u), \tag{6}$$

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

Why is our method flexible and easy to implement? First, our method is 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. Meaning, as long as there is a bootstrap that allows for a re-sampling where the conditional variance of the bootstrap errors is equal to $\widehat{\mathbb{V}}(u|X)$, then one could implement our correction.

Second, our method is easy to implement as it relies only on estimating linear regressions. Thus, our method can profit from the development of any fast estimation procedure handling linear regressions with many covariates.

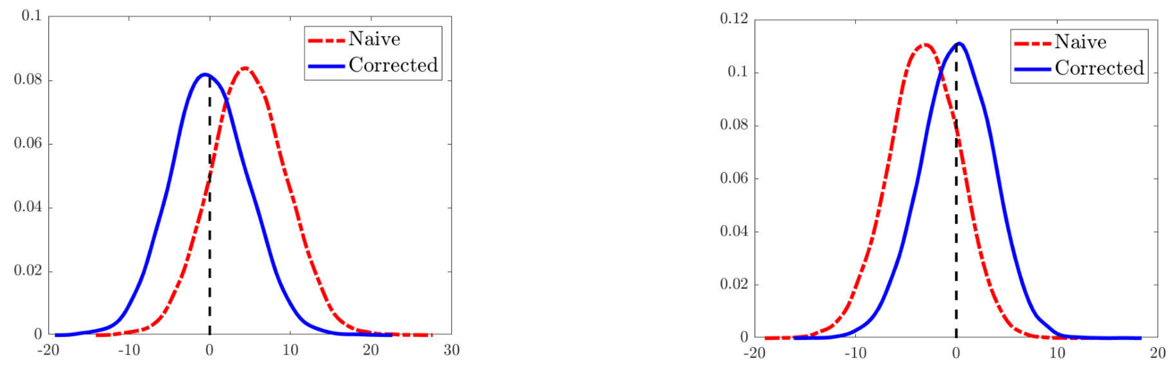
In principle, one could think that both estimating a linear regression and our bias correction deal with the problem of inverting $X'X$. The estimated coefficients of a linear regression $\widehat{\beta}$ are defined as the unknowns that solve for the normal equations, $X'X\widehat{\beta} = X'Y$. One could find this solution by directly inverting the matrix $X'X$, but this is not efficient. Instead, standard algorithms avoid the direct computation of such inverse.[21]

When the number of covariates is large, even standard methods have trouble in doing a linear estimation. But, with the proliferation of large datasets, 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.

[20] 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.

[21] In practice, standard algorithms for estimating linear regressions do a QR decomposition of X, which transforms the normal equations into an upper triangular system. This allows to use a sequential method to solve for $\widehat{\beta}$. While this method directly avoids the computation of the inverse, it is however impractical when the number of covariates is large, as the number of sequential steps is equal to the number of covariates. In addition, the QR decomposition itself is computationally costlier the larger the matrix.

Figure 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}.

3.1 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 homoscedastic 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.[22]

Figures 1a and 1b show the effectiveness of our method. Figure 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 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 1 presents the mean and variance of the differences of our bootstrap method δ^* and the bootstrap following MacKinnon and Smith Jr (1998) δ^*_{MS} with respect to the direct correction $\widehat{\delta}$.[23] 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

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

[23] 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.

Table 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).

δ^*_{MS} in Columns 3 and 4 performs worse in terms of bias and variance.

Table 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).

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 homoscedastic and heteroscedastic errors. Nevertheless, Gaure implemented his bias correction method on the R package *lfe* only under the assumption of homoscedastic errors. In contrast, KSS and ourselves provide corrections under heteroscedasticity and serial correlation or clustering of the errors. Additionally, our method is the only one capable of doing multiple corrections at a time 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 can accommodate more complicated error structures with two or more dimensions of serial dependence, which would yield a non-block-diagonal covariance matrix.

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 et al. (2013) to implement the estimation

Table 2: Monte Carlo simulations. Homoscedastic errors.

	Time	Mean Squared Error (MSE$\times 10^2$)			
		$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.637	0.341	0.114	2.364
Gaure	17.3	0.050	0.109	0.015	0.058
Boot	0.9	0.050	0.105	0.014	0.057
KSS	1.3	0.050	0.106	0.014	0.057

Notes: *Plug-in* is the naive plug-in estimator, *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. *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).

of the following regression model for the log of the wage of worker i at time t:

$$w_{it} = \theta_i + \psi_{J(i,t)} + q_{it}\gamma + \varepsilon_{it}, \tag{7}$$

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 (7) can be estimated by OLS where the person and firm fixed effects 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. We also explore differences between the covariance estimation methods described in Section 3.

We first compare all the methods under conditional homoscedasticity of the errors. Results are in Table 2. All the methods reduce the initial bias of the plug-in estimate. Gaure, KSS and our method are very similar in terms of MSE, and even look identical after rounding the numbers up, with Gaure doing slightly worse.[24] Gaure is the slowest method and the bootstrap correction is also faster than KSS.

Table 3 presents the comparison of our method to KSS under conditional heteroscedasticity for different degrees of worker mobility.[25] Both methods reduce by more than 97% the MSE compared to the plug-in estimates in the low mobility case.[26] Our method is slightly more accurate for both mobility cases, and it also outperforms KSS in terms of time.[27] In the Online Appendix, we show that HC_2 estimate for the covariance matrix outperforms both HC_0 and HC_1 in terms of MSE when doing the bias correction with heteroscedastic errors.

Table 4 presents the results from a simulation with a non diagonal covariance matrix. In par-

[24]We implement Gaure and KSS corrections as follows. 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.

[25]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 for the *Plugin* estimates in Table 3.

[26]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.

[27]In the Online Appendix, we compare the densities of the bias for the different methods. The densities show that both corrections (KSS and our bootstrap) are similar but the bootstrap method 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 3: Monte Carlo simulations. Heteroscedastic errors.

Mov/firm	Model	Time	Mean Squared Error (MSE$\times 10^2$) $\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Low Mobility						
3	Plug-in		22.885	7.702	6.451	12.346
3	Boot	0.7	0.224	0.665	0.192	0.360
3	KSS	1.0	0.268	0.708	0.233	0.403
Mid Mobility						
5	Plug-in		10.518	1.670	1.070	4.419
5	Boot	0.8	0.085	0.255	0.048	0.129
5	KSS	1.1	0.086	0.257	0.048	0.131

Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method with HC_2 covariance matrix estimator, and *KSS* is the Kline et al. (2020) method. True moments are computed at the leave-one-out connected set. *Mov/firm* is the number of movers per firm 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 100. *Average* is the average MSE (also scaled).

ticular, we assume that there is serial correlation of the wages within a given match and the true innovation is homoscedastic. 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 heteroscedastic innovations at the match level.

Why is our method faster? In the above simulations, our method takes between half and two thirds as much time as the method proposed by KSS.[28] The underlying reason is that our method 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. On the other hand, KSS method needs to do, in general, the same number of iterative procedures as number of corrections plus the iteration for the leverage estimation. In some particular cases, like in the AKM model, 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.[29]

5 Application

In the application we use a panel data from the French statistical agency (INSEE) from 2002 to 2014.[30] Our dependent variable is (log) gross daily wage of full time employees with ages between 20 and 60 working at private firms.

The goal is to use our bootstrap method to do a bias corrected variance decomposition of log

[28]This range is computed taking into account the simulations and the application in cases where the bootstrap correction is used with the HC_2 covariance matrix estimator and therefore both methods need to compute the leave-one-out connected set and estimate the leverage of observations. We think this is the fairest comparison. In other cases where we would not need to estimate the leverages our method would be even faster.

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

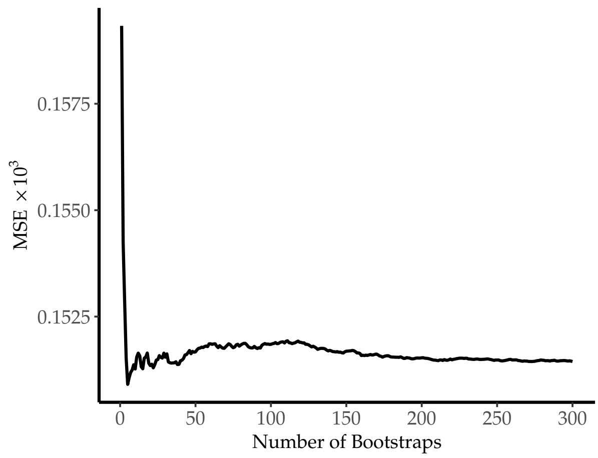
[30]In particular we use *Panel tous salariés-EDP* that consists of a random subsample of workers with firm identifiers and socio-demographic variables. The sample consists of workers born in October on certain days. The sample size was increased in 2002 so we took this as the starting year.

Table 4: Monte Carlo simulations. Serial correlation with homoscedasticity.

	Time	Mean Squared Error (MSE$\times 10^2$)			
		$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		94.352	1.670	0.603	32.208
Boot	0.6	9.674	0.264	0.053	3.330
KSS	1.3	21.571	0.254	0.052	7.292

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 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 homoscedastic error simulations. The bootstrap correction assumes a diagonal covariance matrix and we use the HC_1 covariance matrix estimator.

wages. 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 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.[31] The figure shows that with the first 25 bootstraps the MSE reduces significantly and around 150 it flattens. This suggests that few bootstraps are enough to gain accuracy.[32] In the Online Appendix we show a more formal way of choosing the number of bootstraps based on the Chebyshev's inequality.[33]

Table 5 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 serial correlation within a match. 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 closer to zero and it approaches values that have been found in other countries with a larger number

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

[32] Throughout the application corrections we run corrections with 300 simulation to estimate the leverage and 1000 bootstraps to estimate the corrections of second order moments.

[33] This criterion yields a quite conservative number of bootstraps. It reflects the generality of the result as this criterion would work regardless of the distribution of the error terms.

Table 5: Application. Plug-in vs corrected decomposition.

	Plug-in	Boot Serial		Plug-in	Boot Serial
$Var(y)$	0.2162	0.2162	$2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.0325	-0.0062
$Var(\widehat{\theta}_i)$	0.1688	0.1409	$2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$	-0.0135	-0.0134
$Var(\widehat{\psi}_j)$	0.0493	0.0319	$2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$	-0.0006	-0.0006
$Var(\mathbf{q}\widehat{\gamma})$	0.0146	0.0146	$Corr(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.0964	-0.0004
$Var(\widehat{\epsilon})$	0.0301	0.0490	Obs.	5108399	5108399

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 $\widehat{\sigma}^2_\theta$ 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.

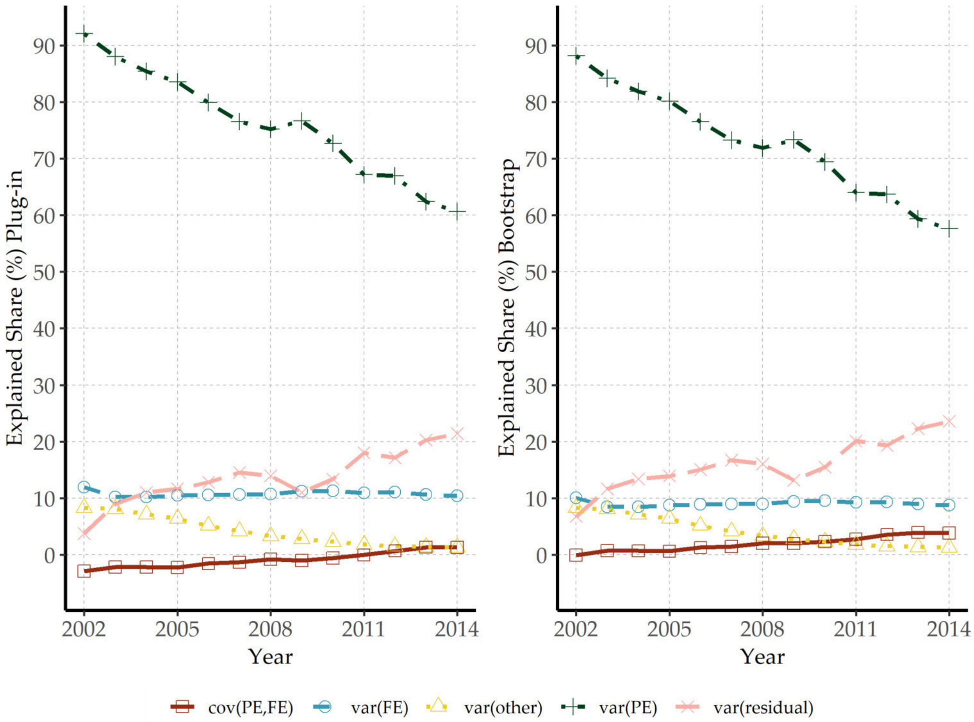
of movers per firm, which should attenuate the bias, as reported by Table 1 of Lopes de Melo (2018). Naturally, the variance and covariance of the person and firm effects are the moments that change the most after the correction. The reason is that the underlying estimates of the person and fixed effects 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.

To fully exploit the benefit of our bootstrap correction method we also perform a yearly variance decomposition. In Figure 3 we compare the year-to-year evolution of the different explained shares using the naive estimated moments and the corrected ones. The main takeaway from this figure is that the correction changes the levels but not the slopes of explained shares. This leads to a change in the relative importance of each component. In particular, the corrected variance of the residuals is relatively more important than the corrected variance of the firm effect in almost every year, while both are similar when considering uncorrected variances. A very interesting trend is the decline in explanatory power of the individual fixed effects for recent years. It might be just a feature of the French data and explanations for this phenomenon are outside the scope of this paper.

The fixed effects we are estimating are constant across the entire sample. The relative importance of each set of fixed effects changes across periods because the composition in the sample changes via the exit and entry of firms and workers. Another potential reason is the variance of the residual changes across periods. The reason for changes in the relative importance of the fixed effects is different than the exercise done by Lachowska, Mas, Saggio, and Woodbury (2022), which allows for the estimated firm effects to change across periods. In their setting, the changes can arise because of sample composition changes and/or changes in the estimated fixed effects for the same individuals across periods. We want to emphasize that the reasons why our method is faster than KSS will also apply in such settings.

Finally, in the Online Appendix we compare our method against KSS using the French data. To make the comparison as fair as possible, we use the HC_2 covariance matrix estimator, which requires estimating the leverage of each observation as well. We also use the leave-one-out connected set sample, which is smaller than the connected set sample used in the baseline. In the application, our bootstrap method takes a little more than a quarter of the time of the KSS correction. Both methods yield slightly different estimates of second order moments that also lead to differences in

Figure 3: Application. Evolution of the explained shares.



Notes: This figure presents the year-to-year evolution of the explained shares of the total log wage variance of the plug-in and corrected estimates of the person and firm fixed effects, their covariance and the variances of other covariates and the residual.

the estimate of the correlation between firm and worker fixed effects. When assuming a diagonal covariance matrix with heteroscedastic errors (*Boot* HC_2 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 0.14 while the one of KSS is 0.11. Also, when assuming that the error terms are correlated at the match level, the estimated correlation between worker and firm fixed effects is positive with both methods when using the leave-one-out connected set.[34] The estimated correlation with the bootstrap is 0.15 and 0.33 with KSS.[35] 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.

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 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 speci-

[34]The results from *Boot Serial (Connected)* are analogous to the ones in Table 5 but computed with residualized log wages.

[35]The column *Boot Serial* in the Online Appendix is computed at the leave-one-out connected set for comparison of the estimates.

fication, 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 multi-way fixed effects regressions. Our method is faster than Kline et al. (2020) and as accurate in the simulations. Besides the speed, another advantage of our approach is that it allows to increase the number of moments to correct without increasing the computational costs. Plus, it allows to consider more elaborated variance structures. In the case of heteroscedastic errors, although KSS estimator would yield to unbiased estimates of the bias correction, we find that using HC_2 for the covariance matrix estimator leads, in general, to more efficient bias correction estimators.

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 $\beta_j^{*\prime}A\beta_j^*$ we have that

$$\beta_j^{*\prime}A\beta_j^* = (v_j^*)'Q'AQv_j^*.$$

Under the bootstrap, i.e. conditional on X and u, the only source of randomness is v_j^*. Taking expectations under the bootstrap of $\beta_j^{*\prime}A\beta_j^*$, conditionally on X and u, and using the assumption $\mathbb{E}(v_j^*|X,u) = 0$, we get

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

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

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

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

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

$$\frac{1}{p}\sum_{j=1}^p\left(\beta_j^{*\prime}A\beta_j^*\right) \xrightarrow{p} \mathbb{E}_{v^*}\left(\beta_i^{*\prime}A\beta_i^*\mid X,u\right) = \widehat{\delta}.$$

□

Corollary 1

Proof. Using the Law of Iterated Expectations we get

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

□

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.

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.

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.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 126, 1593–1660.

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

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. (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.

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.

Iranzo, S., F. Schivardi, and E. Tosetti (2008): "Skill dispersion and firm productivity: An analysis with employer-employee matched data," *Journal of Labor Economics*, 26, 247–285.

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

Correcting Small Sample Bias in Linear Models with Many Covariates

Miren Azkarate-Askasua and Miguel Zerecero

In this Appendix we first provide details on how we construct the simulated labor markets that we use to test and compare our bootstrap correction. Second, 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. Third, we briefly explain how to choose the number of bootstraps based in Chebyshev's inequality. Fourth, we explain the algorithms used in the paper. Fifth, we compare our method to Borovičková and Shimer (2017), both with simulated labor market data and the French data. Sixth, 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 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.[1]

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).[2] 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

[1]We thank Simen Gaure for sharing with us a piece of code that we used as a base for the simulations.
[2]See Table 1 in Kline et al. (2020) where each worker is observed twice.

and homoscedastic or heteroscedastic innovations. The simulated log wage is like equation (7) 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-2 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).[3] 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 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

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

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

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

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

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.[4]

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 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 $h_{ii} \in [1/n, 1]$.

[4] 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 $\frac{1}{n_h - 1}\left(\frac{1}{n_h}\sum_{j=1}^{n_h}\left(\omega_{i,j} - \widehat{\omega}_{i,j}\right)^2 \widehat{\omega}_{i,j}^2 - \widehat{m}_{ii}\widehat{h}_{ii}\right)$.

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.[5] 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-4 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 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-4 describes the details.

OA-3 Choosing the number of bootstraps

Some readers might feel uneasy with the arbitrary number of bootstraps necessary to correct the bias. To choose the number of bootstraps in the main application we first simulated a "similar" labor market and check how many bootstraps were necessary for a significant reduction of the mean squared error. However, this is still an arbitrary procedure and might not be a very efficient way to do for every application. In this section we show a way to discipline the choice of the number

[5] 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.

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.

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 $\mathbb{V}(\delta^*|X,u) = \frac{1}{p}\eta^2$.[6] 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-4 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. 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.

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 $\widehat{\delta}_m^*$.

[6] 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 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-5 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.[7]

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:[8]

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

[7]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.

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

Table OA-1: 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	0.9	0.050	0.105	0.014	0.057
KSS	1.3	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-5.1 Comparison of Methods

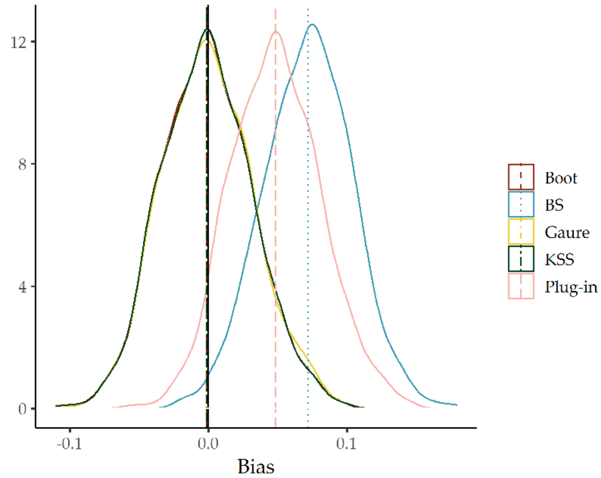
We perform two exercises to compare our method with BS. First, we simulate labor market data that fulfills the key identifying assumptions of the AKM linear model and of BS. We find that both methods correct the bias but ours outperforms theirs in terms of accuracy of the estimation of each of the elements 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-1. 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%.[9] 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 2 presents their estimates of the corresponding AKM moments. Figure OA-1 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).[10]

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

[10] 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-1: 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-2: Application. Extended Comparison of the Methods (BS Data).

	BS	Plug-in	Boot HC_1	Boot HC_2	KSS
$\widehat{\sigma}^2_\theta$	0.061	0.095	0.063	0.063	0.062
$\widehat{\sigma}^2_\psi$	0.005	0.038	0.020	0.019	0.019
$\widehat{\sigma}_{\theta,\psi}$	0.010	-0.004	0.005	0.005	0.006
$\widehat{\rho}_{\theta,\psi}$	0.558	-0.064	0.131	0.157	0.161
Obs.	945356	942235	942235	931925	931925

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.

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.[11] The first restriction implies that the sample used for BS is about half of the original sample of private firms.[12] 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.

In order to accommodate for the extra covariates within the BS method, we first run a linear regression of log wage versus q_{it} (age and education interacted by year effects) and take the residual. We use the averaged match-level residual wage as the dependent variable to compute the moments, both for the BS and our bootstrap method. We estimate the bootstrap corrected moments at the connected set or leave-one-out-corrected set of the BS final sample.

Table OA-2 compares the estimated moments using the BS method and the bootstrap correction

[11] 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.

[12] 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-3: Application. Summary Statistics.

BS Data	Obs.	Mean Wage	Mean Age	Mean Education
No	3311804	4.39	41.43	4.56
Yes	2541773	4.37	36.94	4.95

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).

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.[13] 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 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-2 under our method also differ from the ones previously reported in Table 5 in the main text. Table OA-3 presents some summary statistics of the original data differentiated by being in the final BS data or not.[14] 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-3 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-6 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

[13] The BS estimates are obtained by using the formulas of Section 5.2. in Borovičková and Shimer (2017).

[14] The original data constitutes of almost 5.9 million observations that translate into a connected set of 5.1 million observations as in Table 5.

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 6 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,*

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

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

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

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

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

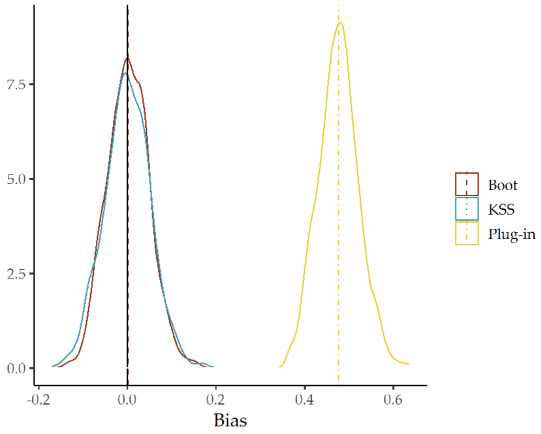
where we use that the bootstrap errors are independent across observations and the fact that $\mathbb{E}\left((v_i^*)^3 \mid X, u\right) = 0$. □

OA-7 Additional Tables and Figures

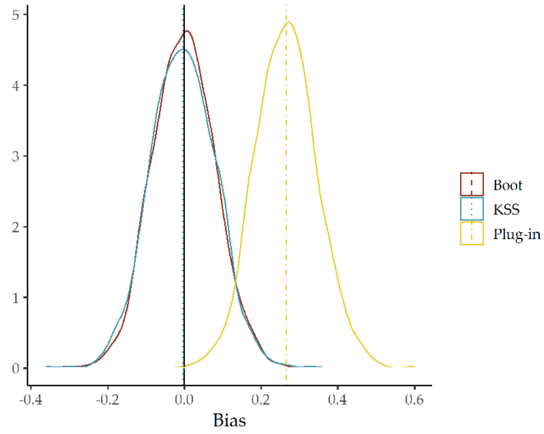
Table OA-4 does the same exercise as the *Low Mobility* part of Table 3 in the main text in a more realistic sample size of roughly 5 million observations. Table OA-5 compares the MSE for the different moments when using different assumptions on the covariance matrix estimators applicable with our bootstrap 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-6 present the Monte Carlo simulation results for serially correlated error terms when the true innovation is heteroscedastic. Figures OA-2 and OA-3 show the distribution of the corrections in Monte Carlo simulations when the error terms are respectively heteroscedastic and when they are serially correlated. Table OA-7 compares the bootstrap correction to the KSS correction in the French application.

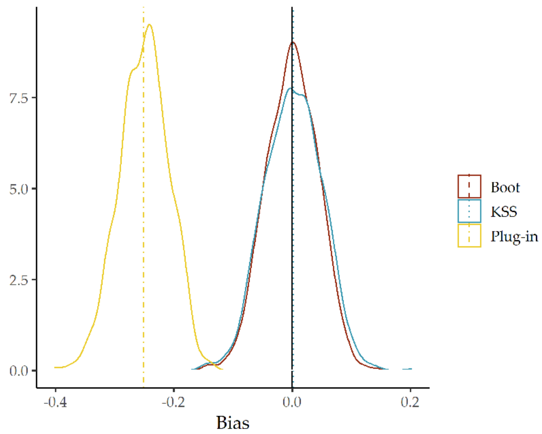
Figure OA-2: Model Comparison: Heteroscedastic Errors.



(a) Bias of $\widehat{\sigma}^2_\theta$



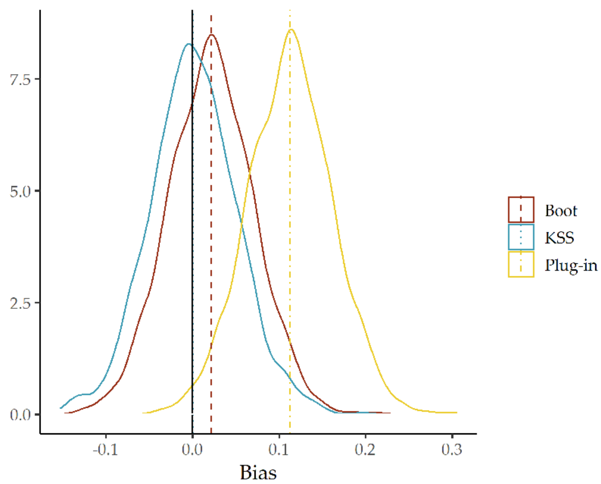
(b) Bias of $\widehat{\sigma}^2_\psi$



(c) Bias of $\widehat{\sigma}_{\theta,\psi}$

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.

Figure OA-3: Model Comparison: Serial Correlation with heteroscedasticity.



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 have serial correlation, true innovations are heteroscedastic and labor mobility is high.

Table OA-4: 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		203.136	60.374	52.419	105.310
Boot	418.1	0.001	0.000	0.001	0.001
KSS	726.1	0.003	0.000	0.002	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* refers to our method with HC_2 covariance matrix estimator, 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).

Table OA-5: Comparison of variance estimators. Heteroscedastic errors

	Mean Squared Error (MSE$\times 10^2$)			
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	3.399	0.740	2.335	2.158
Boot HC_1	0.800	1.301	1.103	1.068
Boot HC_2	0.220	0.679	0.210	0.370

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 3 and all the estimations are done using the leave-one-out sample.

Table OA-6: Monte Carlo simulations. Serial correlation with heteroscedasticity.

		Mean Squared Error (MSE$\times 10^2$)			
	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		88.915	1.510	0.234	30.220
Boot	0.3	0.583	0.284	0.030	0.299
KSS	1.3	21.351	0.251	0.045	7.216

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).

Table OA-7: Application. Comparison of the Methods (KSS Data).

	Plug-in	Boot HC_2	Boot Serial (Connected)	Boot Serial	KSS	KSS Serial
$\widehat{\sigma}^2_\theta$	0.156	0.142	0.130	0.143	0.153	0.140
$\widehat{\sigma}^2_\psi$	0.029	0.020	0.036	0.014	0.016	0.008
$\widehat{\sigma}_{\theta,\psi}$	0.000	0.007	-0.002	0.007	0.006	0.011
$\widehat{\rho}_{\theta,\psi}$	0.000	0.137	-0.032	0.153	0.112	0.327
Obs.	2951415	2951415	5108399	2951415	2951415	2951415
Time (min)		4.86	16.38	4.61	16.34	16.34

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.

References

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

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

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.

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.