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 $\widehat{\delta}_{aux,m}^{(j)} = (\beta^*)' A_m \beta^*$ for all $m \in \{1, ..., M\}$.
end for
Compute $\widehat{\delta}_m^* = \dfrac{\sum_{j=1}^p \widehat{\delta}_{aux,m}^{(j)}}{p}$ 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 $\widehat{\delta}_{aux,m}^{(j)} = (\beta^*)' A_m \beta^*$ for all $m \in \{1, ..., M\}$.
end for
Compute $\widehat{\delta}_m^* = \dfrac{\sum_{j=1}^p \widehat{\delta}_{aux,m}^{(j)}}{p}$ 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 $j = 1, \ldots, p$ **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 $i \in \{1, \ldots, n\}$.
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 $i \in \{1, \ldots, n\}$.
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 $i \in \{1, \ldots, n\}$.
Compute $\bar{m}_{ii} = \frac{\widehat{m}_{ii}}{\widehat{m}_{ii} + \widehat{h}_{ii}}$ for all $i \in \{1, \ldots, n\}$.
for $i = 1, \ldots, n$ **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$\times 10^2$)			
		$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.637	0.341	0.114	2.364
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

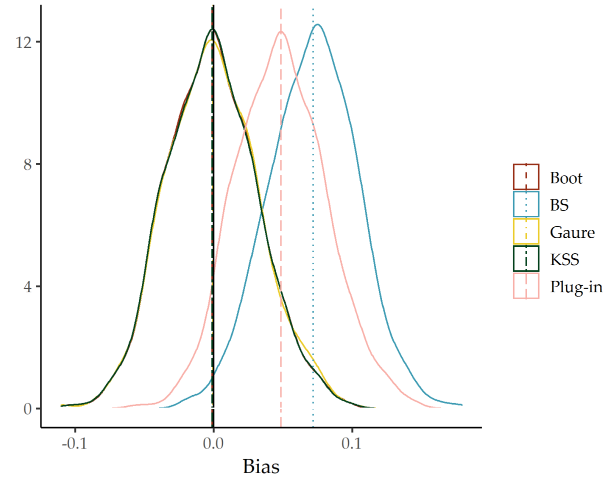
We perform two exercises to compare our method with BS. First, we simulate labor market data that fulfills the key identifying assumptions of the AKM linear model and of BS. We find that both methods correct the bias but ours outperforms theirs in terms of accuracy of the estimation of each of the elements 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,

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

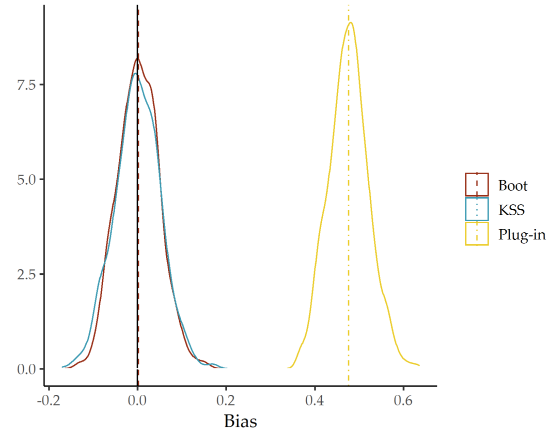
where we use that the bootstrap errors are independent across observations and the fact that $\mathbb{E}\left((v_i^*)^3 \mid X, u\right) = 0$. □

OA-7 Additional Tables and Figures

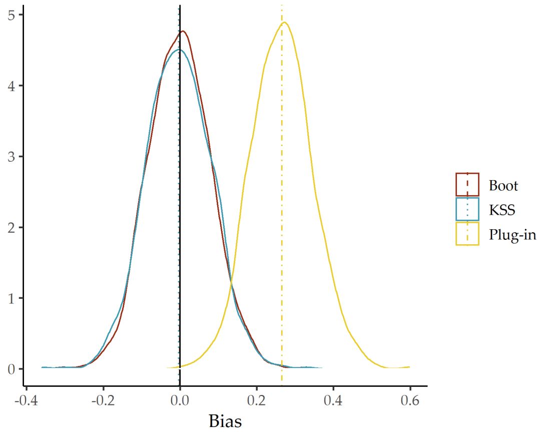
Table OA-4 does the same exercise as the *Low Mobility* part of Table 3 in the main text in a more realistic sample size of roughly 5 million observations. Table OA-5 compares the MSE for the different moments when using different assumptions on the covariance matrix estimators applicable with our bootstrap 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.

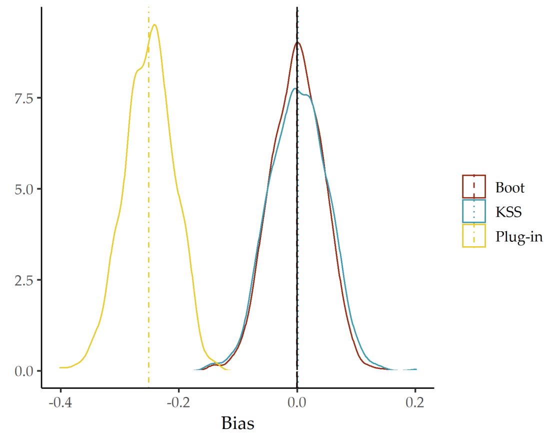
Figure OA-2: Model Comparison: Heteroscedastic Errors.



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



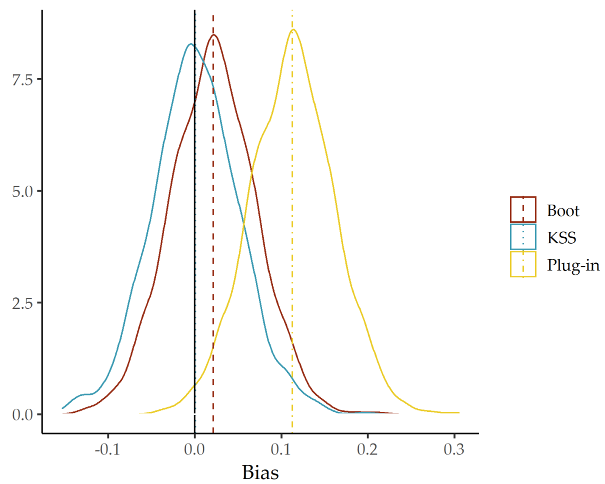
(b) Bias of $\widehat{\sigma}^2_\psi$



(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.