Nonparametric bootstrap correction for incidental parameter bias in maximum likelihood and (G)MM estimation*

Geert Dhaene[‡] Yitian Li[§]

October 20, 2023

Abstract

We show that the nonparametric bootstrap can be used to remove the leading bias term of maximum likelihood, method of moments, and generalized method of moment estimates of panel data models with incidental parameters. The bootstrap can also be iterated to remove higher-order bias terms, and it can be applied to improve estimates of other parameters of interest, such as average marginal effects or the variance of the unobserved heterogeneity. We discuss several examples and also compare the bootstrap with the jackknife.

Keywords: Panel data, incidental parameters, bias correction, GMM, nonparametric bootstrap

JEL classification code: C23

*This research was supported by the Flemish Research Council grant G073620N.
[‡]KU Leuven, `geert.dhaene@kuleuven.be`
[§]KU Leuven, `yitian.li@kuleuven.be`

1 Introduction

In panel data applications, it is customary to seek to control for unobserved cross-sectional heterogeneity, usually by including individual-specific parameters (often called "fixed effects") in the model. However, in many models of interest, standard estimation methods such as maximum likelihood and least squares are inconsistent in short panels when fixed effects are included. Prominent examples are linear autoregressive models (Nickell 1981) and binary-choice models (see, e.g., Chamberlain 1980). The problem is known as the incidental parameter problem (IPP; see Neyman and Scott 1948) and has been studied extensively in the parametric likelihood framework.[1] In a limited number of parametric models, the IPP can be solved in the sense of re-establishing consistency in short panels. A famous example is the panel logit model with fixed effects, where conditioning on a sufficient statistic eliminates the fixed effects from the likelihood function, thereby restoring consistency (Andersen 1970). However, the logit model is the only binary choice model where such an approach is possible.

Given the difficulty of obtaining consistent estimators in short panels, the literature has moved in the direction of asymptotic bias corrections, based on approximation theory as the number of time periods becomes large.[2] Most of this literature uses the parametric likelihood structure to derive approximate bias corrections. However, the IPP also arises in moment-based models, that is, when the model consists only of a set of moment conditions that are imposed on the data. Given the prominence of GMM estimation in economics, it seems surprising that the IPP in the GMM framework has not received more attention. To our knowledge, the paper by Fernández-Val and Lee (2013) is the only one that explicitly focuses on the IPP in GMM.[3]

In this paper, we show that the nonparametric bootstrap can be used in the likelihood

1. The random-effect approach avoids the IPP altogether by specifying the distribution of the individual-specific parameters (conditional on exogenous covariates), but in most models, consistency requires this distribution to be correctly specified.

2. See, e.g., Hahn and Kuersteiner (2002), Woutersen (2002), Hahn and Newey (2004), Arellano and Bonhomme (2009), Bester and Hansen (2009), Bonhomme and Manresa (2015) Dhaene and Jochmans (2015), Arellano and Hahn (2016), Fernández-Val and Weidner (2016), Kim and Sun (2016), Dhaene and Sun (2021), Schumann, Severini, and Tripathi (2021), Jochmans and Higgins (2022), and Schumann (2022).

3. This is not to say that all existing bias correction methods are valid only in the likelihood setting. The analytical bias corrections of Hahn and Newey (2004, p. 1302) and Arellano and Hahn (2016) are also valid for M-estimators. Moreover, jackknife bias corrections also do not require the parametric likelihood assumption.

framework and, more generally, in the (G)MM framework for correcting IPP bias. The nonparametric bootstrap is conceptually very simple to apply, basically substituting analytical work to derive an expression of the approximate bias with bootstrap simulations and computations. In addition, analytical bias corrections have to be made specific to the given model and estimand, while bootstrap bias correction schemes do not need to be tailored to such a degree. A further advantage of the bootstrap is that it can be iterated, though at the cost of increased computation, to yield higher-order bias corrections. By contrast, extending analytical first-order bias corrections to higher order is challenging because the number of terms involved in the expansions increases very rapidly. Similar to the existing analytical bias corrections in the likelihood framework, we show that the nonparametric bootstrap can be applied directly to the estimator and, alternatively, to the profile score function of the (G)MM objective function. The parametric version of the bootstrap has recently been studied by Kim and Sun (2016) and Jochmans and Higgins (2022) in a likelihood framework with incidental parameters. This paper uses the nonparametric version of the bootstrap, thereby also covering models that assume only moment conditions.

Section 2 introduces our framework. Section 3 presents the nonparametric bootstrap corrections for IPP bias. We discuss examples in Section 4 and also compare the bootstrap with the jackknife. Section 5 concludes.

2 Setup

Suppose we observe a random variable z_{it} (possibly a vector) for cross-sectional units $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$. The model is given by a moment condition

$$\mathbb{E}\left[m(z_{it}; \theta_0, \alpha_{i0})\right] = 0, \tag{1}$$

where $m(z_{it}; \theta, \alpha_i)$ is a known function and θ_0 and α_{i0} are the (unknown) true values of the finite-dimensional parameters $\theta \in \Theta$ and $\alpha_i \in \mathcal{A}$. The expectation operator $\mathbb{E}[\cdot]$ has to be understood as the conditional expectation, given α_{i0}. Thus, we treat α_{i0} as the (unobserved) realization of a random variable that captures unobserved cross-sectional heterogeneity in panel data. We further assume that, given α_{i0}, z_{it} is conditionally independent and identically distributed across t, and that (z_{it}, α_{i0}) is independent and identically distributed across i. The parameter of interest is θ, while $\alpha := (\alpha_1, \ldots, \alpha_n)$ is an incidental parameter, whose dimension grows with the cross-sectional sample size, n. For identification, we assume that (1) holds at the true parameter values only and that $\dim(m) \geq \dim(\theta) + \dim(\alpha_i)$.

Our setup is static (by the time series independence assumption) but within the static setting, it covers many models of interest. For example, the moment function $m(z_{it}; \theta, \alpha_i) = w_{it}(y_{it} - r(x_{it}; \theta, \alpha_i))$, for a given function $r(x_{it}; \theta, \alpha_i)$, corresponds to a (possibly nonlinear) IV regression model with outcome variable y_{it}, regressors x_{it}, and instruments w_{it}.

Let $z_i := (z_{i1}, \ldots, z_{iT})'$ and $z := (z_1, \ldots, z_n)$, and define the scalar-valued objective functions

$$
\begin{aligned}
Q_i(\theta, \alpha_i; z_i) &:= \left(\frac{1}{T} \sum_{t=1}^{T} m(z_{it}; \theta, \alpha_i) \right)' \widehat{\Omega}_i^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} m(z_{it}; \theta, \alpha_i) \right), \\
Q(\theta, \alpha; z) &:= \frac{1}{n} \sum_{i=1}^{n} Q_i(\theta, \alpha_i; z_i),
\end{aligned}
$$

for some chosen weight matrix $\widehat{\Omega}_i^{-1}$, where we assume that $\widehat{\Omega}_i$ is nonsingular and converges in probability to a nonsingular matrix Ω_i as $T \to \infty$. The efficient weight matrix would correspond to $\Omega_i = \mathbb{E}\left[m(z_{it}; \theta_0, \alpha_{i0}) m(z_{it}; \theta_0, \alpha_{i0})' \right]$, but we leave the choice of $\widehat{\Omega}_i$ unrestricted (apart from the assumptions made). In fact, in our examples, we shall often simply set $\widehat{\Omega}_i = I$ (the identity matrix) because we focus on bias reduction, not on estimation efficiency (which would make sense only if the bias is already zero).

Now define

$$
\widehat{\alpha}_i(\theta; z_i) := \operatorname*{argmin}_{\alpha_i \in \mathcal{A}} Q_i(\theta, \alpha_i; z_i)
$$

and let $\widehat{\alpha}(\theta; z) := (\widehat{\alpha}_1(\theta; z_1), \ldots, \widehat{\alpha}_n(\theta; z_n))$. Then

$$
\widehat{\theta} := \operatorname*{argmin}_{\theta \in \Theta} Q(\theta, \widehat{\alpha}(\theta; z); z)
$$

is a standard GMM estimator of θ_0. If $T \to \infty$ and $n/T \to 0$, then $\widehat{\theta}$ will be consistent for θ_0 and, if the efficient weight matrix is used, $\widehat{\theta}$ will also be (asymptotically) efficient in the class of GMM estimators based on (1) only. However, if $T \to \infty$ but n/T does not go to zero (i.e., T grows, but not at a faster rate than n), then $\widehat{\theta}$ is no longer efficient (in general), even if the efficient weight matrix is used. In the worst case scenario, when $n \to \infty$ and T is fixed, $\widehat{\theta}$ may be even be inconsistent, i.e., it may converge to a value, say

θ_T, that is different from θ_0. This is because, when T is fixed, $\widehat{\alpha}_i(\theta; z_i)$ remains a random variable, so it does not converge in probability to $\alpha_i(\theta) := \operatorname{argmin}_{\alpha_i \in \mathcal{A}} \mathbb{E}[Q_i(\theta, \alpha_i; z_i)]$ and hence, letting $\alpha(\theta) := (\alpha_1(\theta), \ldots, \alpha_n(\theta))$, the profile objective function $Q(\theta, \widehat{\alpha}(\theta; z); z)$ differs from the "target" objective function $Q(\theta, \alpha(\theta); z)$, a difference that persists in the limit as $n \to \infty$ when T is fixed. Therefore, while the minimizer of $Q(\theta, \alpha(\theta); z)$ is generally consistent for θ_0, that of $Q(\theta, \widehat{\alpha}(\theta; z); z)$ is not. In short, if incidental parameter bias arises, it is due to the presence of $\widehat{\alpha}(\theta; z)$—a parameter estimate whose dimension grows in proportion to the sample size—in the objective function.

Now let $s(\theta, \alpha; z) := \partial_\theta Q(\theta, \alpha; z)$, where ∂_θ is the derivative with respect to θ, and define the "profile score" function $s(\theta, \widehat{\alpha}(\theta; z); z)$ and the "target score" function $s(\theta, \alpha(\theta); z)$. Under suitable regularity conditions, $\widehat{\theta}$ and $\widehat{\alpha}(\theta; z)$ satisfy the first-order condition for minimization of $Q(\theta, \alpha; z)$ and, therefore, $s(\widehat{\theta}, \widehat{\alpha}(\widehat{\theta}; z); z) = 0$. Analogous to the maximum likelihood setting, one may view the incidental parameter problem in GMM as due to the profile score being biased: $\mathbb{E}[s(\theta_0, \widehat{\alpha}(\theta_0; z); z)]$ is not generally zero. By contrast, the target score is unbiased, i.e., $\mathbb{E}\left[s(\theta_0, \alpha(\theta_0); z)\right] = 0$. The target score, however, is infeasible since $\alpha(\theta)$ is not available.

There are various ways to correct for incidental parameter bias, in particular regarding the object that is being corrected. In this paper, we will focus on directly correcting the estimator, $\widehat{\theta}$, and on correcting the profile score, function, $s(\theta, \widehat{\alpha}(\theta; z); z)$.

Remark 1: The special case where $\dim(m) = \dim(\theta) + \dim(\alpha_i)$ corresponds to M-estimation. The case where, in addition,

$$m(z_{it}; \theta, \alpha_i) = \begin{pmatrix} \partial_\theta \log f(z_{it}; \theta, \alpha_i) \\ \partial_{\alpha_i} \log f(z_{it}; \theta, \alpha_i) \end{pmatrix} \tag{2}$$

where $f(z_{it}; \theta_0, \alpha_{i0})$ is the density (or probability mass) function of z_{it} (possibly given strictly exogenous covariates, embedded in z_{it}), corresponds to the likelihood setting. We will not treat these special cases separately, but some of our examples will be in the setting of M-estimation or maximum likelihood estimation.

Remark 2: In some cases the parameter of interest may not be θ but a more general functional of the joint distribution of (z_{it}, α_{i0}), such as the degree of unobserved cross-sectional heterogeneity (measured by the variance of α_{i0}, for example) and average marginal effects. To embed such quantities in our framework, we augment (1) with a parameter μ_0 defined by the moment condition

$$\overline{\mathbb{E}}\left[h(z_{it};\theta_0,\alpha_{i0};\mu_0)\right]=0, \tag{3}$$

where $h(z_{it};\theta,\alpha_i;\mu)$ is a known function and $\overline{\mathbb{E}}[\cdot]$ is the expectation over the joint distribution of (z_{it},α_{i0}). Given estimates $\widehat{\theta}$ and $\widehat{\alpha}_i(\widehat{\theta};z_i)$, a natural plug-in estimator of μ_0 is to solve

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}h(z_{it};\widehat{\theta},\widehat{\alpha}_i(\widehat{\theta};z_i);\mu)=0$$

for μ, giving $\widehat{\mu}$. It is easy to see that the incidental parameter problem (if any) will typically carry over from $\widehat{\theta}$ to $\widehat{\mu}$. In fact, in some simple models no incidental parameter bias arises for $\widehat{\theta}$, but even then the presence of the inconsistent estimates $\widehat{\alpha}_i(\widehat{\theta};z_i)$ (for fixed T) will generally render $\widehat{\mu}$ inconsistent.

3 Nonparametric bootstrap bias correction

Let F_0 denote the true distribution of z, given $\alpha_0=(\alpha_{1,0},\ldots,\alpha_{n,0})$, i.e., $z\sim F_0$. From here onward we will write expectations under F_0, such as those in Section 2, explicitly as $\mathbb{E}_{F_0}$ to distinguish them from expectations obtained in a bootstrap setting, which we now introduce.

Let $\widehat{F}(z)$ be the empirical distribution defined as follows by z. For $z^*\sim\widehat{F}(z)$, z^* has the same dimension, $T\times n$, as z, and z^*_{it} is independent across i and independent and identically distributed across t with

$$\Pr_{\widehat{F}(z)}[z^*_{it}=z_{is}]=\frac{1}{T}\qquad(i=1,\ldots,n;t,s=1,\ldots,T).$$

That is, sampling from $\widehat{F}(z)$ amounts, for each i, to resampling with replacement from the elements of z_i to form $z^*_i:=(z^*_{i1},\ldots,z^*_{iT})'$, and then $z^*:=(z^*_1,\ldots,z^*_n)$. This is the nonparametric bootstrap resampling strategy we use throughout this paper. Note that our notation comprises nested bootstrap sampling: if z^* is a bootstrap sample (i.e., drawn from $\widehat{F}(z)$), then drawing from $\widehat{F}(z^*)$ gives a nested bootstrap sample, say z^{**}, and so on.

3.1 Correcting the estimator

The bias of $\widehat{\theta}$ is

$$\text{bias}(\widehat{\theta}) = \mathbb{E}_{F_0}[\widehat{\theta}(z) - \theta_0],$$

where we write $\widehat{\theta}$ as $\widehat{\theta}(z)$ to indicate that the estimator is based on the data z. By the bootstrap principle, on replacing F_0 with $\widehat{F}(z)$ and z with $z^* \sim \widehat{F}(z)$, we obtain the ideal first-order bootstrap estimate of the bias as

$$\begin{aligned}\widehat{\text{bias}}(\widehat{\theta}) &= \mathbb{E}_{\widehat{F}(z)}[\widehat{\theta}(z^*) - \widehat{\theta}(z)] \\ &= \mathbb{E}_{\widehat{F}(z)}[\widehat{\theta}(z^*)] - \widehat{\theta}.\end{aligned}$$

Hence

$$\begin{aligned}\widehat{\theta}_1^{\text{boot}} &= \widehat{\theta} - \widehat{\text{bias}}(\widehat{\theta}) \\ &= 2\widehat{\theta} - \mathbb{E}_{\widehat{F}(z)}[\widehat{\theta}(z^*)]\end{aligned}$$

is an ideal first-order bootstrap bias-corrected estimator of θ_0.

3.2 Correcting the score function

Another way to correct for incidental parameter bias is to correct the bias of the profile score function relative to the target score function, which is

$$\text{bias}(s(\theta, \widehat{\alpha}(\theta; z); z)) = \mathbb{E}_{F_0}[s(\theta, \widehat{\alpha}(\theta; z); z) - s(\theta, \alpha(\theta); z)].$$

An exactly, but infeasible, bias-corrected score function is

$$\tilde{s}(\theta; z) = s(\theta, \widehat{\alpha}(\theta; z); z) - \mathbb{E}_{F_0}[s(\theta, \widehat{\alpha}(\theta; z); z) - s(\theta, \alpha(\theta); z)]. \tag{4}$$

Here, again, we can estimate the unknown expectation by the nonparametric bootstrap. Replacing F_0 with $\widehat{F}(z)$, z with $z^* \sim \widehat{F}(z)$, and, crucially, $\alpha(\theta)$ with $\widehat{\alpha}(\theta; z)$ gives the ideal first-order bootstrap bias-corrected score function as

$$\tilde{s}_1^{\text{boot}}(\theta; z) = s(\theta, \widehat{\alpha}(\theta; z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\theta, \widehat{\alpha}(\theta; z^*); z^*) - s(\theta, \widehat{\alpha}(\theta, z); z^*)]. \tag{5}$$

Then, solving

$$\tilde{s}_1^{\text{boot}}(\theta; z) = 0$$

for θ gives an ideal first-order bootstrap bias-corrected estimator, $\tilde{\theta}_1^{\text{boot}}$.

3.3 Correcting the score function only at θ_0

Note that (4) is a bias-corrected score function at every possible $\theta \in \Theta$. Another, and somewhat simpler, way to obtain a bias-corrected estimator is to correct the profile score only at θ_0. Recall, indeed, that the incidental parameter problem stems from possible bias of $s(\theta, \widehat{\alpha}(\theta; z); z)$ at θ_0. This bias is

$$\text{bias}(s(\theta_0, \widehat{\alpha}(\theta_0, z); z)) = \mathbb{E}_{F_0}[s(\theta_0, \widehat{\alpha}(\theta_0; z); z)]$$

and it suffices to subtract this from the profile score function to obtain an (infeasible) exactly unbiased score function at θ_0,

$$\check{s}(\theta; z) = s(\theta, \widehat{\alpha}(\theta, z); z) - \mathbb{E}_{F_0}[s(\theta_0, \widehat{\alpha}(\theta_0; z); z)],$$

with ideal first-order bootstrap estimator

$$\check{s}_1^{\text{boot}}(\theta; z) = s(\theta, \widehat{\alpha}(\theta, z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\widehat{\theta}(z), \widehat{\alpha}(\widehat{\theta}(z); z^*); z^*)]. \tag{6}$$

Solving

$$\check{s}_1^{\text{boot}}(\theta; z) = 0$$

for θ gives another ideal first-order bootstrap bias-corrected estimator, $\check{\theta}_1^{\text{boot}}$.

3.4 Discussion

It is not difficult to see why the nonparametric bootstrap should work well as a bias reduction device. Consider bias-correcting $\widehat{\theta}$ and recall that $\text{bias}(\widehat{\theta}) = \mathbb{E}_{F_0}(\widehat{\theta}(z) - \theta_0)$. If we knew F_0 we could simply calculate $\text{bias}(\widehat{\theta})$ exactly and then the bias-corrected estimate $\widehat{\theta} - \text{bias}(\widehat{\theta})$ would have exactly zero bias by construction. Replacing F_0 with $\widehat{F}(z)$ makes the bias calculation feasible but exactness is lost. Yet the calculation will be approximately correct because $\widehat{F}(z)$ (the "empirical distribution") is a reasonably good approximation to F_0. In some sense it is the best we can do because $\widehat{F}(z)$ is the nonparametric maximum likelihood estimator of F_0. Furthermore, since $\widehat{F}(z) \to F_0$ as $T \to \infty$, the approximation improves as T increases. The relative error involved in the bias calculation via the bootstrap is $O(T^{-1})$ in expectation, that is, $\mathbb{E}[\widehat{\text{bias}(\widehat{\theta})}] = \text{bias}(\widehat{\theta})(1+O(T^{-1}))$ (under suitable regularity conditions). The same reasoning applies to the bootstrap bias correction of the profile score function and of the profile score function evaluated at θ_0. The corresponding bias-corrected profile score functions have a relative error of $O(T^{-1})$ in expectation, and the associated estimators $\tilde{\theta}_1^{\text{boot}}$ and $\check{\theta}_1^{\text{boot}}$ inherit this property, i.e., their bias is reduced by a factor $O(T^{-1})$.

Another way to interpret the bias correction via the bootstrap is that it eliminates the leading term in an expansion of the asymptotic bias of $\widehat{\theta}$. Recall that, as $n \to \infty$ with T fixed, we have $\widehat{\theta} \xrightarrow{p} \theta_T$. Here, θ_T can be expanded (again, under suitable regularity conditions) as

$$\theta_T = \theta_0 + \frac{B_1}{T} + \frac{B_2}{T^2} + \ldots + \frac{B_k}{T^k} + o(T^{-k}) \qquad \text{as } T \to \infty, \tag{7}$$

where $B_1, \ldots, B_k$ are constants (i.e., not depending on T) and k is a positive integer. Thus, as T increases, the bias of $\widehat{\theta}$ becomes approximately equal to the leading bias term in this expansion, B_1/T. Now, because the bootstrap estimates the bias with relative error $O(T^{-1})$, the bootstrap can be thought of as eliminating B_1/T. At the same time, the bootstrap also modifies the higher-order terms in the expansion, but without modifying their order of magnitude. Thus, the bootstrap reduces the bias from $O(T^{-1})$ (the bias of $\widehat{\theta}$) down to $O(T^{-2})$ (the bias of $\widehat{\theta}_1^{\text{boot}}$). A similar reasoning applies to the bootstrap bias corrections applied to the score function, both globally and at θ_0 only. As $n \to \infty$ with fixed T, the profile score function converges in probability to a nonstochastic function of θ, say $s_T(\theta)$, by analogy to the convergence $\widehat{\theta} \xrightarrow{p} \theta_T$. The function $s_T(\theta)$ can be expanded as in (7), with $B_1, \ldots, B_k$ now being functions of θ, and the bootstrap eliminates the leading bias term in this expansion (either globally or at θ_0 only). Hence, compared to the uncorrected profile score function, the bias of the bias-corrected profile score function is reduced from $O(T^{-1})$ to $O(T^{-2})$ (globally or at θ_0 only), and the corresponding estimators $\tilde{\theta}_1^{\text{boot}}$ and $\check{\theta}_1^{\text{boot}}$ inherit the bias reduction property.

It is of interest to link the bias reduction property of the bootstrap to that of the jackknife, a nonparametric method for bias reduction and variance estimation originally developed for cross-sectional and time series data by Quenouille (1949, 1956). Hahn and Newey (2004) extended the jackknife to panel data applications with incidental parameters. Their panel jackknife estimator of θ_0 is is a "delete-one" jackknife estimator defined as

$$\widehat{\theta}_1^{\text{jack}} = T\widehat{\theta} - (T-1)\frac{\sum_{t=1}^{T}\widehat{\theta}_{(t)}}{T},$$

where $\widehat{\theta}_{(t)}$ is $\widehat{\theta}$ computed without the observations from t-th time period. It is easy to see that $\widehat{\theta}_1^{\text{jack}}$ reduces the incidental parameter bias of $\widehat{\theta}$. As $n \to \infty$, we have $\widehat{\theta}_{(t)} \xrightarrow{p} \theta_{T-1}$ and, therefore, $\widehat{\theta}_1^{\text{jack}} \xrightarrow{p} \theta_T^{\text{jack}}$ where

$$\begin{aligned}
\theta_T^{\text{jack}} &= T\theta_T - (T-1)\theta_{T-1} \\
&= \theta_0 + T\left(\frac{B_1}{T} + \frac{B_2}{T^2} + \ldots + \frac{B_k}{T^k} + o(T^{-k})\right) \\
&\quad - (T-1)\left(\frac{B_1}{T-1} + \frac{B_2}{(T-1)^2} + \ldots + \frac{B_k}{(T-1)^k} + o((T-1)^{-k})\right) \\
&= \theta_0 + \frac{B_2}{T^2} + o(T^{-2}),
\end{aligned}$$

provided that $k \geq 2$. Hence the delete-one jackknife eliminates the bias term B_1/T, just as the bootstrap does. Given that the bootstrap can be viewed as a generalization of the jackknife (Efron 1982), it should not come as a surprise that the bootstrap and the jackknife have similar bias reduction properties.

3.5 Higher-order corrections, bootstrap simulations, and further discussion

Bootstrap bias corrections can be iterated, as discussed in Hall and Martin (1988).

Second-order correction of the estimator. Starting with $\widehat{\theta}_1^{\text{boot}}$, its bias is

$$\begin{aligned}
\text{bias}(\widehat{\theta}_1^{\text{boot}}) &= \mathbb{E}_{F_0}[\widehat{\theta}_1^{\text{boot}}(z) - \theta_0] \\
&= \mathbb{E}_{F_0}\left[2\widehat{\theta}(z) - \mathbb{E}_{\widehat{F}(z)}[\widehat{\theta}(z^*)] - \theta_0\right].
\end{aligned}$$

Applying the bootstrap principle again, we replace F_0 with $\widehat{F}(z)$, $\widehat{F}(z)$ with $\widehat{F}(z^*)$, z with $z^* \sim \widehat{F}(z)$, and z^* with $z^{**} \sim \widehat{F}(z^*)$ to obtain the ideal second-order bootstrap estimate of the bias as

$$\widehat{\text{bias}}(\widehat{\theta}_1^{\text{boot}}) = \mathbb{E}_{\widehat{F}(z)}\left[2\widehat{\theta}(z^*) - \mathbb{E}_{\widehat{F}(z^*)}[\widehat{\theta}(z^{**})] - \widehat{\theta}(z)\right]$$

$$= \mathbb{E}_{\widehat{F}(z)}\big[2\widehat{\theta}(z^*) - \mathbb{E}_{\widehat{F}(z^*)}[\widehat{\theta}(z^{**})]\big] - \widehat{\theta}.$$

Hence

$$\begin{aligned}
\widehat{\theta}_2^{\text{boot}} &= \widehat{\theta}_1^{\text{boot}} - \widehat{\text{bias}}(\widehat{\theta}_1^{\text{boot}}) \\
&= 3\widehat{\theta} - 3\mathbb{E}_{\widehat{F}(z)}[\widehat{\theta}(z^*)] + \mathbb{E}_{\widehat{F}(z)}\mathbb{E}_{\widehat{F}(z^*)}[\widehat{\theta}(z^{**})]
\end{aligned}$$

is an ideal second-order bootstrap bias-corrected estimator of θ_0.

Second-order correction of the score function. Similarly, the bias of $\tilde{s}_1^{\text{boot}}(\theta; z)$ is

$$\begin{aligned}
\text{bias}(\tilde{s}_1^{\text{boot}}(\theta; z)) &= \mathbb{E}_{F_0}[\tilde{s}_1^{\text{boot}}(\theta; z) - s(\theta, \alpha(\theta); z)] \\
&= \mathbb{E}_{F_0}\big[s(\theta, \widehat{\alpha}(\theta; z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\theta, \widehat{\alpha}(\theta; z^*); z^*) - s(\theta, \widehat{\alpha}(\theta; z); z^*)] \\
&\qquad - s(\theta, \alpha(\theta); z)\big].
\end{aligned}$$

Using the nonparametric bootstrap, we can approximate it as

$$\begin{aligned}
\widehat{\text{bias}}(\tilde{s}_1^{\text{boot}}(\theta; z)) = \mathbb{E}_{\widehat{F}(z)}\big[&s\left(\theta, \widehat{\alpha}\left(\theta; z^*\right); z^*\right) - \mathbb{E}_{\widehat{F}(z^*)}[s(\theta, \widehat{\alpha}(\theta; z^{**}); z^{**}) - s(\theta, \widehat{\alpha}(\theta; z^*); z^{**})] \\
&- s(\theta, \widehat{\alpha}(\theta; z); z^*)\big].
\end{aligned}$$

Therefore, an ideal second-order bias-corrected score function is

$$\begin{aligned}
\tilde{s}_2^{\text{boot}}(\theta; z) &= \tilde{s}_1^{\text{boot}}(\theta; z) - \widehat{\text{bias}}(\tilde{s}_1^{\text{boot}}(\theta; z)) \\
&= s(\theta, \widehat{\alpha}(\theta; z); z) - 2\mathbb{E}_{\widehat{F}(z)}[s(\theta, \widehat{\alpha}(\theta; z^*); z^*) - s(\theta, \widehat{\alpha}(\theta; z); z^*)] \\
&\qquad + \mathbb{E}_{\widehat{F}(z)}\mathbb{E}_{\widehat{F}(z^*)}[s(\theta, \widehat{\alpha}(\theta; z^{**}); z^{**}) - s(\theta, \widehat{\alpha}(\theta; z^*); z^{**})]
\end{aligned}$$

Second-order correction of the score function at θ_0. The bias of $\check{s}_1^{\text{boot}}(\theta; z)$ at θ_0 is

$$\begin{aligned}
\text{bias}(\check{s}_1^{\text{boot}}(\theta_0; z)) &= \mathbb{E}_{F_0}[\check{s}_1^{\text{boot}}(\theta_0; z)] \\
&= \mathbb{E}_{F_0}\big[s(\theta_0, \widehat{\alpha}(\theta_0; z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\widehat{\theta}(z), \widehat{\alpha}(\widehat{\theta}(z); z^*); z^*)]\big],
\end{aligned}$$

which can be approximated using the nonparametric bootstrap as

$$\widehat{\text{bias}}(\check{s}_1^{\text{boot}}(\theta_0; z)) = \mathbb{E}_{\widehat{F}(z)}\big[s(\widehat{\theta}(z), \widehat{\alpha}(\widehat{\theta}(z); z^*); z^*) - \mathbb{E}_{\widehat{F}(z^*)}[s(\widehat{\theta}(z^*), \widehat{\alpha}(\widehat{\theta}(z^*); z^{**}); z^{**})]\big].$$

Subtracting this from $\check{s}_1^{\text{boot}}(\theta; z)$ gives an ideal second-order bias-corrected score function at θ_0,

$$\begin{aligned}\check{s}_2^{\text{boot}}(\theta; z) &= \check{s}_1^{\text{boot}}(\theta; z) - \widehat{\text{bias}}(\check{s}_1^{\text{boot}}(\theta_0; z)) \\ &= s(\theta, \widehat{\alpha}(\theta; z); z) - 2\mathbb{E}_{\widehat{F}(z)}[s(\widehat{\theta}(z), \widehat{\alpha}(\widehat{\theta}(z); z^*); z^*)] \\ &\qquad + \mathbb{E}_{\widehat{F}(z)}\mathbb{E}_{\widehat{F}(z^*)}[s(\widehat{\theta}(z^*), \widehat{\alpha}(\widehat{\theta}(z^*); z^{**}); z^{**})].\end{aligned}$$

Higher-order corrections. For higher order corrections, let $z^{*(0)} = z$ and, for $b = 0, 1, 2, \ldots$, define $z^{*(b+1)} \sim \widehat{F}(z^{*(b)})$ recursively, so that $z^{*(1)} = z^*$, $z^{*(2)} = z^{**}$, and so on. It is easy to show that the K-th order ideal bootstrap bias corrections are

$$\begin{aligned}\widehat{\theta}_K^{\text{boot}} &= (K+1)\widehat{\theta} - \sum_{k=2}^{K+1}(-1)^k \binom{K+1}{k} \mathbb{E}_{\widehat{F}(z)}\mathbb{E}_{\widehat{F}(z^*)} \cdots \mathbb{E}_{\widehat{F}(z^{*(k-1)})}[\widehat{\theta}(z^{*(k-1)})], \\ \tilde{s}_K^{\text{boot}}(\theta; z) &= s(\theta, \widehat{\alpha}(\theta; z); z) \\ &\quad - \sum_{k=1}^{K}(-1)^{k+1}\binom{K}{k}\mathbb{E}_{\widehat{F}(z)}\mathbb{E}_{\widehat{F}(z^*)} \cdots \mathbb{E}_{\widehat{F}(z^{*(k)})}[s(\theta, \widehat{\alpha}(\theta; z^{*(k)}); z^{*(k)}) \\ &\qquad\qquad - s(\theta, \widehat{\alpha}(\theta; z^{*(k-1)}); z^{*(k)})],\end{aligned}$$

and

$$\begin{aligned}\check{s}_K^{\text{boot}}(\theta; z) &= s(\theta, \widehat{\alpha}(\theta; z); z) \\ &\quad - \sum_{k=1}^{K}(-1)^{k+1}\binom{K}{k}\mathbb{E}_{\widehat{F}(z)}\mathbb{E}_{\widehat{F}(z^*)} \cdots \mathbb{E}_{\widehat{F}(z^{*(k)})}[s(\widehat{\theta}(z^{*(k-1)}), \widehat{\alpha}(\widehat{\theta}(z^{*(k-1)}); z^{*(k)}); z^{*(k)})].\end{aligned}$$

Solving $\tilde{s}_K^{\text{boot}}(\theta; z) = 0$ for θ gives $\tilde{\theta}_K^{\text{boot}}$; solving $\check{s}_K^{\text{boot}}(\theta; z) = 0$ gives $\check{\theta}_K^{\text{boot}}$. Every additional order of the bootstrap eliminates an additional term in the expansion (7) and in the corresponding expansion of the profile score function, thereby reducing the bias by an additional factor $O(T^{-1})$. Therefore, the remaining bias of all three estimators $\widehat{\theta}_K^{\text{boot}}$, $\tilde{\theta}_K^{\text{boot}}$, and $\check{\theta}_K^{\text{boot}}$ is $O(T^{-K-1})$.

Bootstrap simulations. Unlike the jackknife, for which there are closed-form expressions (see also the expression below for the second-order jackknife), the bootstrap bias corrections are usually not available in closed form. Except in very simple cases, the expectations have to be computed by simulation. This involves the following steps to draw nested bootstrap samples, all of which are of the same dimension, $T \times n$, as the original data set z.

- Draw B bootstrap samples from $\widehat{F}(z)$ as described in Section 3 and denote these as $z^*_{b_1}$, $b_1 = 1, \dots, B$.
- For each $z^*_{b_1}$, draw B bootstrap samples from $\widehat{F}(z^*_{b_1})$ and denote these as $z^{*(2)}_{b_1 b_2}$, $b_2 = 1, \dots, B$.
- Iterate this procedure to form further bootstrap samples $z^{*(3)}_{b_1 b_2 b_3}, \dots, z^{*(K)}_{b_1 b_2 b_3 \cdots b_K}$, where $b_j = 1, \dots, B$ for each $j = 1, \dots, K$.

Now we can approximate $\widehat{\theta}^{\text{boot}}_K$, $\tilde{s}^{\text{boot}}_K(\theta; z)$, and $\check{s}^{\text{boot}}_K(\theta; z)$ using

$$
\begin{aligned}
\widehat{\theta}^{\text{boot}}_K &\approx (K+1)\widehat{\theta}(z) \\
&\quad - \sum_{k=2}^{K+1} \binom{K+1}{k} (-1)^k \frac{1}{B^{k-1}} \sum_{b_1=1}^{B} \sum_{b_2=1}^{B} \cdots \sum_{b_{k-1}=1}^{B} \widehat{\theta}(z^{*(k-1)}_{b_1 b_2 \cdots b_{k-1}}), \\
\tilde{s}^{\text{boot}}_K(\theta; z) &\approx s(\theta, \widehat{\alpha}(\theta; z); z) \\
&\quad - \sum_{k=1}^{K} \binom{K}{k} (-1)^{k+1} \frac{1}{B^k} \sum_{b_1=1}^{B} \sum_{b_2=1}^{B} \cdots \sum_{b_k=1}^{B} [s(\theta, \widehat{\alpha}(\theta; z^{*(k)}_{b_1 b_2 \cdots b_k}); z^{*(k)}_{b_1 b_2 \cdots b_k}) \\
&\qquad\qquad - s(\theta, \widehat{\alpha}(\theta; z^{*(k-1)}_{b_1 b_2 \cdots b_{k-1}}); z^{*(k)}_{b_1 b_2 \cdots b_{k-1}})], \\
\check{s}^{\text{boot}}_K(\theta; z) &\approx s(\theta, \widehat{\alpha}(\theta; z); z) \\
&\quad - \sum_{k=1}^{K} \binom{K}{k} (-1)^{k+1} \frac{1}{B^k} \sum_{b_1=1}^{B} \sum_{b_2=1}^{B} \cdots \sum_{b_k=1}^{B} s(\widehat{\theta}(z^{*(k-1)}_{b_1 b_2 \cdots b_{k-1}}), \\
&\qquad\qquad \widehat{\alpha}(\widehat{\theta}(z^{*(k-1)}_{b_1 b_2 \cdots b_{k-1}}); z^{*(k)}_{b_1 b_2 \cdots b_k}); z^{*(k)}_{b_1 b_2 \cdots b_k}),
\end{aligned}
$$

where $z^{*(0)} = z$. As $B \to \infty$, the simulation-based approximations converge in probability to the ideal bootstrap estimators.

Further discussion. The jackknife can also be applied to yield higher-order bias corrections, although not literally by iterating the delete-one jackknife. Extending the delete-one panel jackknife of Hahn and Newey (2004) to the second order gives the delete-two panel jackknife estimator

$$
\widehat{\theta}^{\text{jack}}_2 = \frac{1}{2} T^2 \widehat{\theta} - (T-1)^2 \frac{\sum_{t=1}^{T} \widehat{\theta}_{(t)}}{T} + \frac{1}{2}(T-2)^2 \frac{\sum_{t=1}^{T} \sum_{s=t+1}^{T} \widehat{\theta}_{(t,s)}}{T(T-1)/2},
$$

where $\widehat{\theta}_{(t,s)}$ is $\widehat{\theta}$ computed without the observations from the t-th and s-th time periods. It can be shown that, as $n \to \infty$ with T fixed, $\widehat{\theta}_2^{\text{jack}} \xrightarrow{p} \theta_0 + O(T^{-3})$, that is, the second-order jackknife eliminates the terms B_1/T and B_2/T^2 from (7). In our examples, we shall compare bootstrap bias corrections up to the third order and jackknife bias corrections up to the second order.

Note that the bias reduction properties (of the bootstrap and jackknife alike) are of an asymptotic, large-T nature. This means that for large enough T bias correction methods of higher order should give biases that are smaller than the biases of the corresponding bias correction methods of lower order. For fixed T, however, increasing the order of bias correction is not guaranteed to decrease the bias, although very often it does, as our examples will demonstrate.

There is also no general reason to expect a performance ranking between bootstrapping, jackknifing, and the analytical bias correction derived in Fernández-Val and Lee (2013), at least if we limit attention to the first-order versions of the bootstrap and jackknife. The relative performance ranking generally depends on the model, the distribution of the fixed effects, the value θ_0, and the distribution of the covariates (i.e., on the entire data generating process) and on T. Likewise, among the three ways to apply the bootstrap (to the estimator, to the profile score function, to the profile score function at θ_0), there is no general reason either to expect a performance ranking. However, correcting the score everywhere instead of correcting it at θ_0 only, may be less likely to generate a multiplicity of solutions of the corrected profile score equation.

4 Examples

Here we discuss some examples of GMM models where an incidental parameter problem occurs, and what the nonparametric bootstrap delivers in terms of bias reduction.

4.1 Many means

Suppose z_{it} has mean α_{i0} and variance σ_0^2, which is the parameter of interest (so $\theta = \sigma^2$). The moment function is

$$m(z_{it}; \sigma^2, \alpha_i) = \begin{pmatrix} z_{it} - \alpha_i \\ (z_{it} - \alpha_i)^2 - \sigma^2 \end{pmatrix}.$$

The model consists of the moment condition (1) with the moment function m as defined here. The model corresponds to the many normal means example of Neyman and Scott (1948), but without assuming that z_{it} is normally distributed. The many normal means example figures prominently in the literature on the incidental parameter problem because it is a classic example of a (parametric) model where maximum likelihood fails to be consistent and because it provides a simple test case for any general (i.e., model-unspecific) method that seeks to reduce the incidental parameter bias. The discussion here adds two things: the parametric likelihood assumption is replaced by the (weaker) moment condition (1), and the bias reduction is achieved through the nonparametric bootstrap. For simplicity, we set $\widehat{\Omega}_i = I$.[4] Then the analysis with or without assuming normality turns out to be essentially the same (and straightforward) because the normal likelihood involves only the first two moments, just as the moment condition $\mathbb{E}[m(z_{it}; \sigma_0^2, \alpha_{i0})] = 0$.

We have

$$\widehat{\alpha}_i(\sigma^2; z_i) = \overline{z}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} z_{it}, \qquad \widehat{\sigma}^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - \overline{z}_{i\cdot})^2$$

(so GMM with $\widehat{\Omega}_i = I$ coincides with maximum likelihood), and $\mathbb{E}_{F_0}[\widehat{\sigma}^2] = (1 - T^{-1})\sigma_0^2$. The bias of $\widehat{\sigma}^2$ is $-\sigma_0^2/T$. The ideal first-order bootstrap bias-corrected estimator is

$$\begin{aligned}
\widehat{\sigma}_1^{2\text{boot}} &= 2\widehat{\sigma}^2 - \mathbb{E}_{\widehat{F}(z)}(\widehat{\sigma}^2(z^*)) \\
&= 2\widehat{\sigma}^2 - (1 - T^{-1})\widehat{\sigma}^2 \\
&= (1 + T^{-1})\widehat{\sigma}^2,
\end{aligned}$$

with expectation $\mathbb{E}_{F_0}[\widehat{\sigma}_1^{2\text{boot}}] = (1 - T^{-2})\sigma_0^2$. Thus, the bootstrap reduces the bias to $-\sigma_0^2/T^2$.

The profile score function is

$$s(\sigma^2, \widehat{\alpha}(\sigma^2; z); z) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - \overline{z}_{i\cdot})^2 - \sigma^2$$

4. The optimal weight matrix involves μ_{3i} and μ_{4i}, the third and fourth central moments of z_{it}, given α_{i0}. The weight matrix $\widehat{\Omega}_i = I$ is optimal if, and only if, $\mu_{3i} = 0$ and μ_{4i} is a constant.

(omitting an inessential factor -2). We find the first-order bootstrap bias correction as

$$
\begin{aligned}
\tilde{s}_1^{\text{boot}}(\sigma^2; z) &= s(\sigma^2, \widehat{\alpha}(\sigma^2; z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\sigma^2, \widehat{\alpha}(\sigma^2; z^*); z^*) - s(\sigma^2, \widehat{\alpha}(\sigma^2, z); z^*)] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - \overline{z}_{i\cdot})^2 - \sigma^2 \\
&\qquad - \mathbb{E}_{\widehat{F}(z)}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it}^* - \overline{z}_{i\cdot}^*)^2 - \sigma^2 - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it}^* - \overline{z}_{i\cdot})^2 + \sigma^2\right] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 - \sigma^2 \\
&\qquad - \left(1 - \frac{1}{T}\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 \\
&= \left(1 + \frac{1}{T}\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 - \sigma^2 \\
&= \left(1 + \frac{1}{T}\right)\widehat{\sigma}^2 - \sigma^2.
\end{aligned}
$$

Therefore $\tilde{\sigma}_1^{2\text{boot}} = (1 + T^{-1})\widehat{\sigma}^2$.

When we correct the score at θ_0 only, we have

$$
\begin{aligned}
\check{s}_1^{\text{boot}}(\sigma^2; z) &= s(\sigma^2, \widehat{\alpha}(\sigma^2, z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\widehat{\sigma}^2(z), \widehat{\alpha}(\widehat{\sigma}^2(z); z^*); z^*)] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 - \sigma^2 - \mathbb{E}_{\widehat{F}(z)}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it}^* - \overline{z}_{i\cdot}^*\right)^2 - \widehat{\sigma}^2\right] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 - \sigma^2 - \left(1 - \frac{1}{T}\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z_{it} - \overline{z}_{i\cdot}\right)^2 + \widehat{\sigma}^2 \\
&= \left(1 + \frac{1}{T}\right)\widehat{\sigma}^2 - \sigma^2
\end{aligned}
$$

and $\check{\sigma}_1^{2\text{boot}} = (1 + T^{-1})\widehat{\sigma}^2$.

Thus, all three first-order bootstrap bias corrections coincide. The higher-order corrections also coincide. Similar calculations give

$$
\widehat{\sigma}_K^{2\text{boot}} = \tilde{\sigma}_K^{2\text{boot}} = \check{\sigma}_K^{2\text{boot}} = \left(1 + \sum_{k=1}^{K}\frac{1}{T^k}\right)\widehat{\sigma}^2,
$$

with bias $-\sigma_0^2/T^{K+1} \to 0$ as $K \to \infty$. By comparison, the first-order jackknife estimator is already unbiased (Hahn and Newey 2004). This is a fortuitous consequence of the bias of $\widehat{\sigma}^2$ being exactly proportional to $1/T$, so the jackknife removes the bias completely. Note, finally, that this model is unusual in that the ideal bootstrap calculations can be carried out without simulations.

4.2 Variance of the fixed effects

Suppose z_{it} is generated by

$$z_{it} = \alpha_{i0} + \varepsilon_{it},$$

where $\mathbb{E}[\varepsilon_{it}] = 0$ and $\mathbb{E}[\varepsilon_{it}^2] = \sigma_{\varepsilon 0}^2$. With $\theta = \sigma_\varepsilon^2$, the moment function is

$$m(z_{it}; \sigma_\varepsilon^2, \alpha_i) = \begin{pmatrix} z_{it} - \alpha_i \\ (z_{it} - \alpha_i)^2 - \sigma_\varepsilon^2 \end{pmatrix},$$

exactly as in the many-means example discussed above. Here, however, we are not interested in σ_ε^2, but instead in the variance of the fixed effects, $\sigma_{\alpha 0}^2 = \overline{\mathbb{E}}[(\alpha_{i0} - \mu_{\alpha 0})^2]$, where $\mu_{\alpha 0} = \overline{\mathbb{E}}[\alpha_{i0}]$; see also Fernández-Val and Lee (2013), where we borrow this example from. Intuitively, the estimate of σ_α^2 will be based on the estimates $\widehat{\alpha}_i$, whose variance will be larger than the variance of α_{i0} (due to estimation variability), and therefore an incidental parameter problem arises for σ_α^2.

The true values of the parameters σ_α^2 and μ_α are defined by the moment condition

$$\overline{\mathbb{E}}\big[h(z_{it}; \sigma_{\varepsilon 0}^2, \alpha_{i0}; \sigma_{\alpha 0}^2, \mu_{\alpha 0})\big] = 0,$$

where the moment function h is

$$h(z_{it}; \sigma_\varepsilon^2, \alpha_i; \sigma_\alpha^2, \mu_\alpha) = \begin{pmatrix} \alpha_i - \mu_\alpha \\ (\alpha_i - \mu_\alpha)^2 - \sigma_\alpha^2 \end{pmatrix}.$$

The estimates are readily obtained. Setting $\widehat{\Omega}_i = I$ gives

$$\widehat{\alpha}_i(z_i) = \overline{z}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} z_{it}, \qquad \widehat{\sigma}_\varepsilon^2(z) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - \overline{z}_{i\cdot})^2,$$

(as in the many-means example) and hence

$$\begin{aligned}
\widehat{\mu}_\alpha(z) &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\alpha}_i(z_i) = \frac{1}{n}\sum_{i=1}^{n}\overline{z}_{i\cdot} = \overline{z}_{\cdot\cdot}, \\
\widehat{\sigma}^2_\alpha(z) &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{\alpha}_i(z_i) - \widehat{\mu}_\alpha(z))^2 = \frac{1}{n}\sum_{i=1}^{n}(\overline{z}_{i\cdot} - \overline{z}_{\cdot\cdot})^2.
\end{aligned}$$

We have

$$\begin{aligned}
\overline{\mathbb{E}}[\overline{z}_{i\cdot} - \overline{z}_{\cdot\cdot}] &= 0, \\
\overline{\mathbb{E}}[(\overline{z}_{i\cdot} - \overline{z}_{\cdot\cdot})^2] &= (1-n^{-1})\sigma^2_{\alpha 0} + (1-n^{-1})\sigma^2_{\varepsilon 0}/T.
\end{aligned}$$

Therefore, as $n \to \infty$ with T fixed,

$$\mathbb{E}_{F_0}[\widehat{\sigma}^2_\alpha(z)] \to \sigma^2_{\alpha 0} + \sigma^2_{\varepsilon 0}/T.$$

The incidental parameter bias of $\widehat{\sigma}^2_\alpha$ is $\sigma^2_{\varepsilon 0}/T$.

The ideal bootstrap calculations can be carried out analytically also for this model, as they involve only the first two moments. The ideal first-order bootstrap bias-corrected estimator of $\sigma^2_{\alpha 0}$ is

$$\begin{aligned}
\widehat{\sigma}^{2\text{boot}}_{\alpha,1} &= 2\widehat{\sigma}^2_\alpha(z) - \mathbb{E}_{\widehat{F}(z)}[\widehat{\sigma}^2_\alpha(z^*)] \\
&= 2\widehat{\sigma}^2_\alpha(z) - (1-n^{-1})\widehat{\sigma}^2_\alpha(z) - (1-n^{-1})\widehat{\sigma}^2_\varepsilon(z)/T \\
&= (1+n^{-1})\widehat{\sigma}^2_\alpha(z) - (1-n^{-1})\widehat{\sigma}^2_\varepsilon(z)/T.
\end{aligned}$$

We have

$$\mathbb{E}_{F_0}[\widehat{\sigma}^2_\varepsilon(z)] = \sigma^2_{\varepsilon 0} - \sigma^2_{\varepsilon 0}/T$$

and so, as $n \to \infty$ with T fixed,

$$\begin{aligned}
\mathbb{E}_{F_0}[\widehat{\sigma}^{2\text{boot}}_{\alpha,1}] &\to \sigma^2_{\alpha 0} + \sigma^2_{\varepsilon 0}/T - (\sigma^2_{\varepsilon 0} - \sigma^2_{\varepsilon 0}/T)/T \\
&= \sigma^2_{\alpha 0} + \sigma^2_{\varepsilon 0}/T^2.
\end{aligned}$$

The bootstrap reduces the incidental parameter bias by a factor T^{-1}.

Let $h_2(\cdot)$ be the second element of $h(\cdot)$. The profile score function for σ_α^2 is

$$\begin{aligned} s(\sigma_\alpha^2, \widehat{\alpha}(z); z) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} h_2(z_{it}; \widehat{\sigma}_\varepsilon^2(z), \widehat{\alpha}_i(z_i); \sigma_\alpha^2, \widehat{\mu}_\alpha(z)) \\ &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{\alpha}_i(z_i) - \widehat{\mu}_\alpha(z))^2 - \sigma_\alpha^2 \\ &= \widehat{\sigma}_\alpha^2(z) - \sigma_\alpha^2. \end{aligned}$$

The ideal first-order bootstrap bias-corrected profile score function is

$$\begin{aligned} \tilde{s}_1^{\text{boot}}(\sigma_\alpha^2; z) &= s(\sigma_\alpha^2, \widehat{\alpha}(z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\sigma_\alpha^2, \widehat{\alpha}(z^*); z^*) - s(\sigma_\alpha^2, \widehat{\alpha}(z); z^*)] \\ &= \widehat{\sigma}_\alpha^2(z) - \sigma_\alpha^2 - \mathbb{E}_{\widehat{F}(z)}[s(\sigma_\alpha^2, \widehat{\alpha}(z^*); z^*) - s(\sigma_\alpha^2, \widehat{\alpha}(z); z^*)] \\ &= \widehat{\sigma}_\alpha^2(z) - \sigma_\alpha^2 - \mathbb{E}_{\widehat{F}(z)}\left[\widehat{\sigma}_\alpha^2(z^*) - \sigma_\alpha^2 - \frac{1}{n}\sum_{i=1}^{n}(\overline{z}_{i\cdot} - \overline{z}_{\cdot\cdot}^*)^2 + \sigma_\alpha^2\right] \\ &= \widehat{\sigma}_\alpha^2(z) - \sigma_\alpha^2 - \big[(1-n^{-1})\widehat{\sigma}_\alpha^2(z) + (1-n^{-1})\widehat{\sigma}_\varepsilon^2(z)/T \\ &\qquad\qquad - (1-n^{-1})\widehat{\sigma}_\alpha^2(z) - \widehat{\sigma}_\varepsilon^2(z)/(nT)\big] \\ &= \widehat{\sigma}_\alpha^2(z) - (1-2n^{-1})\widehat{\sigma}_\varepsilon^2(z)/T - \sigma_\alpha^2, \end{aligned}$$

where we used

$$\begin{aligned} \mathbb{E}_{\widehat{F}(z)}[\widehat{\sigma}_\alpha^2(z^*)] &= (1-n^{-1})\widehat{\sigma}_\alpha^2(z) + (1-n^{-1})\widehat{\sigma}_\varepsilon^2(z)/T, \\ \mathbb{E}_{\widehat{F}(z)}[(\overline{z}_{i\cdot} - \overline{z}_{\cdot\cdot}^*)^2] &= (1-n^{-1})\widehat{\sigma}_\alpha^2(z) - \widehat{\sigma}_\varepsilon^2(z)/(nT). \end{aligned}$$

Therefore

$$\tilde{\sigma}_{\alpha,1}^{2\text{boot}} = \widehat{\sigma}_\alpha^2(z) - (1-2n^{-1})\widehat{\sigma}_\varepsilon^2(z)/T.$$

Hence, as $n \to \infty$ with T fixed,

$$\mathbb{E}_{F_0}[\tilde{\sigma}_{\alpha,1}^{2\text{boot}}] \to \sigma_{\alpha 0}^2 + \sigma_{\varepsilon 0}^2/T^2.$$

If we correct the profile score function only at $\sigma_{\alpha 0}^2$, we have

$$\begin{aligned} \check{s}_1^{\text{boot}}(\sigma_\alpha^2; z) &= s(\sigma_\alpha^2, \widehat{\alpha}(z); z) - \mathbb{E}_{\widehat{F}(z)}[s(\widehat{\sigma}_\alpha^2(z), \widehat{\alpha}(z); z^*); z^*)] \\ &= \widehat{\sigma}_\alpha^2(z) - \sigma_\alpha^2 - \mathbb{E}_{\widehat{F}(z)}[\widehat{\sigma}_\alpha^2(z^*) - \widehat{\sigma}_\alpha^2(z)] \end{aligned}$$

$$
\begin{aligned}
&= \widehat{\sigma}^2_\alpha(z) - \sigma^2_\alpha - [(1-n^{-1})\widehat{\sigma}^2_\alpha(z) + (1-n^{-1})\widehat{\sigma}^2_\varepsilon(z)/T - \widehat{\sigma}^2_\alpha(z)] \\
&= (1+n^{-1})\widehat{\sigma}^2_\alpha(z) - (1-n^{-1})\widehat{\sigma}^2_\varepsilon(z)/T - \sigma^2_\alpha
\end{aligned}
$$

and, hence,

$$
\begin{aligned}
\tilde{\sigma}^{2\text{boot}}_{\alpha,1} &= (1+n^{-1})\widehat{\sigma}^2_\alpha(z) - (1-n^{-1})\widehat{\sigma}^2_\varepsilon(z)/T \\
&= \widehat{\sigma}^{2\text{boot}}_{\alpha,1}.
\end{aligned}
$$

For the higher-order bias corrections, it can be shown that $\widehat{\sigma}^{2\text{boot}}_{\alpha,K}$, $\tilde{\sigma}^{2\text{boot}}_{\alpha,K}$, and $\check{\sigma}^{2\text{boot}}_{\alpha,K}$ all converge to $\sigma^2_{\alpha,0} + \sigma^2_{\varepsilon,0}/T^{K+1}$ as $n \to \infty$ with T fixed. On the other hand, the first-order jackknife bias-corrected estimator, $\widehat{\sigma}^{2\text{jack}}_1$, is consistent because the incidental parameter bias of $\widehat{\sigma}^2_\alpha$ is exactly proportional to T^{-1}, so the jackknife removes the bias exactly.

4.3 An IV model with incidental slope parameters

Consider the model

$$
y_{it} = x'_{1it}\alpha_{i0} + x'_{2it}\beta_0 + \varepsilon_{it}, \tag{8}
$$

where x_{2it} is a k-vector of endogenous regressors and $w_{it} = (x'_{1it}, w'_{2it})'$ is an l-vector of instruments, with x_{1it} an l_1-vector, w_{2it} an l_2-vector, and $l_2 \geq k$. The common parameter, β, is associated with the endogenous regressors, while the "incidental" regressors, x_{1it}, are exogenous. The incidental parameter problem in this model is also discussed in Fernández-Val and Lee (2013).

With $z_{it} = (y_{it}, x'_{1it}, x'_{2it}, w'_{2it})'$, the moment function is

$$
m(z_{it}; \beta, \alpha_i) = w_{it}(y_{it} - x'_{1it}\alpha_i - x'_{2it}\beta).
$$

For identification it is assumed that $\mathbb{E}[m(z_{it}; \beta, \alpha_i)] = 0$ if and only if $\alpha_i = \alpha_{i0}$ and $\beta = \beta_0$. The optimal weight matrix for GMM estimation is the inverse of $\Omega_i = \mathbb{E}[\varepsilon^2_{it} w_{it} w'_{it}]$. As in Fernández-Val and Lee (2013), we assume homoskedasticity, i.e., $\mathbb{E}[\varepsilon^2_{it}|w_{it}] = \sigma^2_\varepsilon$, so that $\Omega_i = \sigma^2_\varepsilon \mathbb{E}[w_{it}w'_{it}]$. Here we can drop the factor σ^2_ε since it is irrelevant for estimation, and set $\widehat{\Omega}_i = T^{-1}\sum_{t=1}^T w_{it}w'_{it}$, the sample analog to $\mathbb{E}[w_{it}w'_{it}]$.

To derive the estimators, it is useful to introduce matrix notation. Let $y_i = (y_{i1}, \ldots, y_{iT})'$, $x_{1i} = (x'_{1i1}, \ldots, x'_{1iT})'$, $x_{2i} = (x'_{2i1}, \ldots, x'_{2iT})'$, and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, so that $y_i = x'_{1i}\alpha_{i0} + x'_{2i}\beta_0 + \varepsilon_i$. Further, let $w_{2i} = (w'_{2i1}, \ldots, w'_{2iT})'$ and $w_i = (w'_{i1}, \ldots, w'_{iT})' = (x_{1i}, w_{2i})$.

Now

$$\frac{1}{T}\sum_{t=1}^{T} m(z_{it};\beta,\alpha_i) = \frac{1}{T} w_i'(y_i - x_{1i}'\alpha_i - x_{2i}'\beta)$$

and hence

$$Q_i(\beta,\alpha_i;z_i) = \frac{1}{T}(y_i - x_{1i}'\alpha_i - x_{2i}'\beta)' w_i (w_i'w_i)^{-1} w_i'(y_i - x_{1i}'\alpha_i - x_{2i}'\beta).$$

Minimizing $Q_i(\beta,\alpha_i;z_i)$ with respect to α_i for given β gives the OLS estimator

$$\widehat{\alpha}_i(\beta;z_i) = (x_{1i}'x_{1i})^{-1}x_{1i}'(y_i - x_{2i}'\beta).$$

The profile objective function is

$$Q_i(\beta,\widehat{\alpha}_i(\beta;z_i);z_i) = \frac{1}{T}(y_i - x_{2i}'\beta)' M_{1i} w_i (w_i'w_i)^{-1} w_i' M_{1i}(y_i - x_{2i}'\beta),$$

where $M_{1i} = I_T - x_{1i}(x_{1i}'x_{1i})^{-1}x_{1i}'$ and I_T is the $T\times T$ identity matrix. Since $M_{1i}x_{1i} = 0_{T\times l_1}$, we have

$$\begin{aligned}
M_{1i}w_i(w_i'w_i)^{-1}w_i'M_{1i} &= (0_{T\times l_1}, M_{1i}w_{2i})\begin{pmatrix} x_{1i}'x_{1i} & x_{1i}'w_{2i} \\ w_{2i}'x_{1i} & w_{2i}'w_{2i}\end{pmatrix}^{-1}\begin{pmatrix} 0_{l_1\times T} \\ w_{2i}'M_{1i}\end{pmatrix} \\
&= M_{1i}w_{2i}(w_{2i}'w_{2i} - w_{2i}'x_{1i}(x_{1i}'x_{1i})^{-1}w_{2i}'w_{2i})^{-1}w_{2i}'M_{1i} \\
&= M_{1i}w_{2i}(w_{2i}'M_{1i}w_{2i})^{-1}w_{2i}'M_{1i},
\end{aligned}$$

so the profile objective function simplifies to

$$Q_i(\beta,\widehat{\alpha}_i(\beta;z_i);z_i) = \frac{1}{T}(\widetilde{y}_i - \widetilde{x}_{2i}'\beta)'\widetilde{w}_{2i}(\widetilde{w}_{2i}'\widetilde{w}_{2i})^{-1}\widetilde{w}_{2i}'(\widetilde{y}_i - \widetilde{x}_{2i}'\beta),$$

where $\widetilde{y}_i = M_{1i}y_i$, $\widetilde{x}_{2i} = M_{1i}x_{2i}$, and $\widetilde{w}_{2i} = M_{1i}w_{2i}$ are the residuals of the projections of y_i, x_{2i}, and w_{2i} onto x_{i1}. Now $\widehat{\beta}$ follows as

$$\widehat{\beta} = \left(\sum_{i=1}^{n}\widetilde{x}_{2i}'\widetilde{w}_{2i}(\widetilde{w}_{2i}'\widetilde{w}_{2i})^{-1}\widetilde{w}_{2i}'\widetilde{x}_{2i}\right)^{-1}\sum_{i=1}^{n}\widetilde{x}_{2i}'\widetilde{w}_{2i}(\widetilde{w}_{2i}'\widetilde{w}_{2i})^{-1}\widetilde{w}_{2i}'\widetilde{y}_i$$

Hence, as $n\to\infty$ with T fixed,

$$\widehat{\beta} \xrightarrow{p} \beta + A^{-1}b,$$

where

$$A = \overline{\mathbb{E}}[\widetilde{x}'_{2i}\widetilde{w}_{2i}(\widetilde{w}'_{2i}\widetilde{w}_{2i})^{-1}\widetilde{w}'_{2i}\widetilde{x}_{2i}],$$
$$b = \overline{\mathbb{E}}[\widetilde{x}'_{2i}\widetilde{w}_{2i}(\widetilde{w}'_{2i}\widetilde{w}_{2i})^{-1}\widetilde{w}'_{2i}\widetilde{\varepsilon}_{i}],$$

and $\widetilde{\varepsilon}_i = M_{1i}\varepsilon_i$. In general, $b = O(1)$ is nonzero and $A^{-1} = O(T^{-1})$, so $A^{-1}b = O(T^{-1})$. The incidental parameter bias can be understood as arising from the standard IV estimator not being unbiased in finite samples.

There are no simple expressions for the required expectations under the bootstrap law, $\mathbb{E}_{\widehat{F}(z)}[\cdot]$, so we resort to simulations to implement the bootstrap bias corrections, as described in Section 3.5. Bootstrapping $\widehat{\beta}$ directly is now straightforward. To bootstrap the profile score function, as in (5) and (6), all we need is the unprofiled score function,

$$\begin{aligned} s(\beta,\alpha;z) &= \frac{1}{n}\sum_{i=1}^{n}\partial_\beta Q_i(\beta,\alpha_i;z_i) \\ &= \frac{-2}{n}\sum_{i=1}^{n}x'_{2i}w_i(w'_iw_i)^{-1}w'_i(y_i - x'_{1i}\alpha_i - x'_{2i}\beta). \end{aligned}$$

Since $s(\beta,\alpha;z)$ is linear in β and α, and $\widehat{\alpha}_i(\beta;z_i)$ is linear in β, every term needed to compute $\tilde{s}_1^{\text{boot}}(\beta;z)$ and $\check{s}_1^{\text{boot}}(\beta;z)$ is also linear in β: the profile score $s(\beta,\widehat{\alpha}(\beta;z);z)$ itself, its bootstrap version $s(\beta,\widehat{\alpha}(\beta;z^*);z^*)$, and the terms $s(\beta,\widehat{\alpha}(\beta;z);z^*)$ and $s(\widehat{\beta}(z),\widehat{\alpha}(\widehat{\beta}(z);z^*);z^*)$ (the latter, in fact, is constant in β). Hence, to compute $\tilde{s}_1^{\text{boot}}(\beta;z)$ and $\check{s}_1^{\text{boot}}(\beta;z)$, we just need to evaluate $s(\beta,\alpha;z)$ multiple times, at appropriate values of β, α, and z. By linearity, the solution to $\tilde{s}_1^{\text{boot}}(\beta;z) = 0$ then follows as $\tilde{\beta}_1^{\text{boot}} = -\tilde{s}_1^{\text{boot}}(0;z)/(\tilde{s}_1^{\text{boot}}(1;z) - \tilde{s}_1^{\text{boot}}(0;z))$, and similarly for $\check{\beta}_1^{\text{boot}}$.

We explored the bias reduction properties of the bootstrap (and jackknife) in a small-scale numerical example with $k = l_1 = l_2 = 1$ and $(x_{1it}, x_{2it}, w_{2it}, \varepsilon_{it})' \sim \mathcal{N}(0,\Sigma)$, with

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 & 0 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0 \\ 0 & 0.5 & 0 & 1 \end{pmatrix},$$

and y_{it} generated according to (8) with $\beta_0 = 1$ and α_{i0} for all i. We set $n = 10,000$ and $T \in \{3, 4, 5, 10, 25, 100\}$. Table 1 presents the means and the standard deviations, across

1,000 Monte Carlo replications, of the IV estimator $\widehat{\beta}$ and the bootstrap and jackknife estimators, where the bootstrap is based on $B = 10$ bootstrap simulations. The effect of the endogeneity is quite severe, with $\widehat{\beta}$ approaching β_0 rather slowly as T increases. Except for the bootstrap correction of the score function everywhere, the bootstrap and jackknife improve on $\widehat{\beta}$. Surprisingly, however, the effect is not spectacular and it takes T to be relatively large before the bias reduction really kicks in. The jackknife turns out to perform best. As expected, the standard deviations increase with the order of bias correction, both using bootstrap and using this jackknife. However, this increase is mild, unless T is very small. The behavior of the bias correction of the score function everywhere (i.e., $\tilde{\beta}_j^{\text{boot}}$, $j = 1, 2, 3$) in this example is poor and needs further study.

$\beta_0 = 1$		$T = 3$	$T = 4$	$T = 5$	$T = 10$	$T = 25$	$T = 100$
$\widehat{\beta}$	mean	1.5998	1.5461	1.4999	1.3528	1.1877	1.0559
	std	0.0089	0.0089	0.0086	0.0075	0.0053	0.0031
$\widehat{\beta}_1^{\text{boot}}$	mean	1.5470	1.4678	1.4081	1.2375	1.0827	1.0088
	std	0.0127	0.0131	0.0123	0.0105	0.0069	0.0036
$\widehat{\beta}_2^{\text{boot}}$	mean	1.5040	1.4147	1.3549	1.1844	1.0472	1.0021
	std	0.0168	0.0169	0.0155	0.0129	0.0082	0.0041
$\widehat{\beta}_3^{\text{boot}}$	mean	1.4680	1.3774	1.3232	1.1539	1.0311	1.0006
	std	0.0211	0.0207	0.0186	0.0154	0.0096	0.0048
$\tilde{\beta}_1^{\text{boot}}$	mean	1.6234	1.5885	1.5577	1.4471	1.2850	1.1021
	std	0.0083	0.0075	0.0072	0.0063	0.0048	0.0033
$\tilde{\beta}_2^{\text{boot}}$	mean	1.6255	1.5925	1.5637	1.4548	1.2905	1.1032
	std	0.0095	0.0086	0.0082	0.0074	0.0061	0.0043
$\tilde{\beta}_3^{\text{boot}}$	mean	1.6257	1.5958	1.5672	1.4570	1.2917	1.1034
	std	0.0115	0.0102	0.0096	0.0091	0.0081	0.0058
$\check{\beta}_1^{\text{boot}}$	mean	1.5433	1.4473	1.3803	1.2050	1.0634	1.0059
	std	0.0131	0.0144	0.0138	0.0116	0.0074	0.0037
$\check{\beta}_2^{\text{boot}}$	mean	1.4946	1.3790	1.3155	1.1506	1.0326	1.0011
	std	0.0177	0.0195	0.0181	0.0148	0.0089	0.0042
$\check{\beta}_3^{\text{boot}}$	mean	1.4523	1.3318	1.2821	1.1232	1.0207	1.0002
	std	0.0224	0.0245	0.0225	0.0182	0.0107	0.0050
$\widehat{\beta}_1^{\text{jack}}$	mean	1.4664	1.3831	1.3182	1.1544	1.0426	1.0035
	std	0.0192	0.0180	0.0164	0.0126	0.0075	0.0035
$\widehat{\beta}_2^{\text{jack}}$	mean	NA	1.2987	1.2227	1.0722	1.0102	1.0001
	std	NA	0.0266	0.0236	0.0160	0.0083	0.0036

Table 1: Means and standard deviations of IV and bias-corrected estimators in a panel model with incidental slopes and an endogenous regressor with common slope $\beta_0 = 1$. The bootstrap is applied to the estimator ($\widehat{\beta}^{\text{boot}}$), the profile score ($\tilde{\beta}^{\text{boot}}$), and the profile score at β_0 only ($\check{\beta}^{\text{boot}}$). Subscripts indicate the order of bias correction.

4.4 Logit model

The fixed-effect panel logit model is a well-known example of a parametric model that suffers from an incidental parameter problem. Let $z_{it} = (y_{it}, x_{it})$, where y_{it} is a binary outcome variable and x_{it} is a vector of exogenous regressors. The model specifies

$$\Pr[y_{it} = 1 | x_{it}] = \Lambda(\alpha_{i0} + x_{it}'\beta_0)$$

where $\Lambda(w) = (1 - e^{-w})^{-1}$ is the standard logistic distribution function. Hence the probability mass function is

$$f(y_{it}|x_{it};\beta,\alpha_i) = [\Lambda(\alpha_i + x_{it}'\beta)]^{y_{it}}[1 - \Lambda(\alpha_i + x_{it}'\beta)]^{1-y_{it}}, \qquad y_{it} \in \{0,1\},$$

where $\theta = \beta$ is the parameter of interest, and the log-likelihood function, normalized by the number of observations, is

$$\begin{aligned} l(\beta,\alpha;z) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\log f(y_{it}|x_{it};\beta,\alpha_i) \\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\big[y_{it}\log\Lambda(\alpha_i + x_{it}\beta) + (1-y_{it})\log(1-\Lambda(\alpha_i + x_{it}\beta))\big]. \end{aligned}$$

The maximum likelihood estimator of β_0 is inconsistent as $n \to \infty$ with T fixed.[5]

Recalling (2), the moment function is

$$\begin{aligned} m(z_{it};\beta,\alpha_i) &= \begin{pmatrix} \partial_\theta \log f(z_{it};\beta,\alpha_i) \\ \partial_{\alpha_i} \log f(z_{it};\beta,\alpha_i) \end{pmatrix} \\ &= \begin{pmatrix} \big(y_{it} - \Lambda(\alpha_i + x_{it}'\beta)\big)x_{it} \\ \big(y_{it} - \Lambda(\alpha_i + x_{it}'\beta)\big) \end{pmatrix}, \end{aligned}$$

where we used the property that $\partial_w \Lambda(w) = \Lambda(w)(1 - \Lambda(w))$. The profile score function is

$$s(\beta,\widehat{\alpha}(\beta;z);z) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\big(y_{it} - \Lambda(\widehat{\alpha}_i(\beta;z_i) + x_{it}'\beta)\big)x_{it},$$

where $\widehat{\alpha}_i(\beta;z_i)$ solves

$$\frac{1}{T}\sum_{t=1}^{T}\big(y_{it} - \Lambda(\alpha_i + x_{it}'\beta)\big) = 0$$

for α_i. Unfortunately, the solution $\widehat{\alpha}_i(\beta;z_i)$ cannot be obtained in closed form from the latter equation. We shall, therefore, focus on the special case where $x_{it} \in \{0,1\}$ (i.e., x_{it}

5. In the logit model, there is a sufficient statistic for α_i and conditioning on it yields a conditional likelihood that resolves the incidental parameter problem. However, the logit model is the only binary-choice model where such a solution exists. Our interest here is in using the logit model as a test case for bootstrap corrections, which have a much wider scope of application than conditional likelihood.

is a binary scalar) because then $\widehat{\alpha}_i(\beta; z_i)$ can be obtained in closed form. With binary x_{it}, the log-likelihood function simplifies to

$$l(\beta, \alpha; z) = \frac{1}{nT} \sum_{i=1}^{n} (z_{i,10} \log \Lambda(\alpha_i) + z_{i,00} \log \Lambda(-\alpha_i) + z_{i,11} \log \Lambda(\alpha_i + \beta)$$
$$+ z_{i,01} \log \Lambda(-\alpha_i - \beta)) + c,$$

where c is an inessential constant and

$$z_{i,yx} = \sum_{t=1}^{T} \mathbb{1}(Y_{it} = y, X_{it} = x) \qquad \text{for } y, x \in \{0, 1\}.$$

Now $\widehat{\alpha}_i(\beta; z_i)$ is given by the following lemma.

Lemma 1. *Let $x_{it} \in \{0, 1\}$ for all i and t. Then*

$$\widehat{\alpha}_i(\beta; z_i) = \log \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

where

$$a = e^{\beta}(z_{i,00} + z_{i,01}),$$
$$b = e^{\beta} z_{i,01} - e^{\beta} z_{i,10} + z_{i,00} - z_{i,11},$$
$$c = -z_{i,10} - z_{i,11}.$$

The proof of the lemma is given in the appendix.

Now the profile score function simplifies to

$$s(\beta, \widehat{\alpha}(\beta; z); z) = \frac{1}{nT} \sum_{i=1}^{n} \left(\frac{z_{i,11} - z_{i,01} e^{\widehat{\alpha}_i(\beta; z_i) + \beta}}{1 + e^{\widehat{\alpha}_i(\beta; z_i) + \beta}} \right)$$

and it is easy to solve $s(\beta, \widehat{\alpha}(\beta; z); z) = 0$ numerically for β, and similarly for solving the bias-corrected profile score equations.

We evaluated the performance of the bootstrap numerically in a small-scale setup with $n = 10,000$ (so as to get close to the probability limits as $n \to \infty$), $T \in \{3, 4, 5, 10\}$, and $B = 10$. We generated the binary variables x_{it} and y_{it} according to

$$\Pr[x_{it} = 1] = 1/2,$$

$$\Pr[y_{it} = 1|x_{it}] = \Lambda(\alpha_{i0} + x_{it}\beta_0),$$

with $\beta_0 = 1$ and $\alpha_{i0} \sim \mathcal{N}(-0.5, 1)$. Table 2 reports the means and the standard deviations of the estimators, estimated from $1,000$ Monte Carlo replications. Clearly, the bootstrap performs reasonably well as a bias correction method, and is on par with the jackknife. The bias corrections have the effect of shrinking the maximum likelihood estimate and, as a result, also tend to reduce the standard deviation of the estimator, most prominently for the first-order bias corrections and when T is very small.

$\beta_0 = 1$		$T = 3$	$T = 4$	$T = 5$	$T = 10$
$\widehat{\beta}$	mean	1.5272	1.3509	1.2608	1.1152
	std	0.0534	0.0371	0.0294	0.0176
$\widehat{\beta}_1^{\text{boot}}$	mean	1.0998	0.9916	0.9758	0.9890
	std	0.0440	0.0307	0.0254	0.0165
$\widehat{\beta}_2^{\text{boot}}$	mean	0.8706	0.8925	0.9396	0.9935
	std	0.0504	0.0357	0.0304	0.0191
$\widehat{\beta}_3^{\text{boot}}$	mean	0.7640	0.8957	0.9630	0.9988
	std	0.0671	0.0468	0.0391	0.0232
$\tilde{\beta}_1^{\text{boot}}$	mean	1.1099	1.0166	1.0092	1.0138
	std	0.0447	0.0319	0.0263	0.0168
$\tilde{\beta}_2^{\text{boot}}$	mean	0.8895	0.9424	0.9857	1.0089
	std	0.0519	0.0380	0.0308	0.0191
$\tilde{\beta}_3^{\text{boot}}$	mean	0.7873	0.9561	1.0024	1.0060
	std	0.0690	0.0496	0.0384	0.0227
$\check{\beta}_1^{\text{boot}}$	mean	1.3096	1.1433	1.0795	1.0146
	std	0.0472	0.0325	0.0263	0.0166
$\check{\beta}_2^{\text{boot}}$	mean	1.1564	1.0339	1.0058	0.9976
	std	0.0453	0.0321	0.0271	0.0179
$\check{\beta}_3^{\text{boot}}$	mean	1.0483	0.9791	0.9801	0.9968
	std	0.0470	0.0345	0.0302	0.0205
$\widehat{\beta}_1^{\text{jack}}$	mean	0.5810	0.8178	0.9049	0.9845
	std	0.0322	0.0219	0.0203	0.0154
$\widehat{\beta}_2^{\text{jack}}$	mean	NA	1.0534	1.0362	1.0027
	std	NA	0.0245	0.0236	0.0157

Table 2: Means and standard deviations of maximum likelihood and bias-corrected estimators in the panel logit model with fixed effects and a binary regressor. The bootstrap is applied to the estimator ($\widehat{\beta}^{\text{boot}}$), the profile score ($\tilde{\beta}^{\text{boot}}$), and the profile score at β_0 only ($\check{\beta}^{\text{boot}}$). Subscripts indicate the order of bias correction.

4.5 Probit model

The fixed-effect panel probit model is very similar to the logit model, but the incidental parameter problem is far more challenging. The probit model specifies

$$\Pr[y_{it} = 1|x_{it}] = \Phi(\alpha_{i0} + x_{it}'\beta_0)$$

where $\Phi(\cdot)$ is the standard normal distribution function. The probability mass function and the log-likelihood function, $f(y_{it}|x_{it};\beta,\alpha_i)$ and $l(\beta,\alpha;z)$, are as in logit model, but with $\Phi(\cdot)$ instead of $\Lambda(\cdot)$. As in the logit model, the maximum likelihood estimator of β_0 is inconsistent as $n \to \infty$ with T fixed. Unlike the logit model, there is no conditional likelihood that is free of incidental parameters. Moreover, Chamberlain (2010) showed that, when $T = 2$ and x_{it} is binary, β_0 is not point-identified. One may conjecture that this holds for any fixed T and for any x_{it} with bounded support. Therefore, a complete resolution of the incidental parameter problem would require calculating (and estimating) the identified set for β_0. Here, instead, we explore the performance of the bootstrap as an approximate (and simpler) approach toward the incidental parameter problem.

The moment function in the probit model is

$$m(z_{it};\beta,\alpha_i) = \begin{pmatrix} \big(y_{it} - \Phi(\alpha_i + x_{it}'\beta)\big)g(\alpha_i + x_{it}'\beta)x_{it} \\ \big(y_{it} - \Phi(\alpha_i + x_{it}'\beta)\big)g(\alpha_i + x_{it}'\beta) \end{pmatrix}$$

where

$$g(w) = \frac{\phi(w)}{\Phi(w)(1-\Phi(w))}$$

and $\phi(w) = \partial_w\Phi(w)$ is the standard normal density function. The profile score function is

$$s(\beta,\widehat{\alpha}(\beta;z);z) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\big(y_{it} - \Phi(\widehat{\alpha}_i(\beta;z_i) + x_{it}'\beta)\big)g(\widehat{\alpha}_i(\beta;z_i) + x_{it}'\beta)x_{it},$$

where $\widehat{\alpha}_i(\beta;z_i)$ solves

$$\frac{1}{T}\sum_{t=1}^{T}\big(y_{it} - \Phi(\alpha_i + x_{it}'\beta)\big)g(\alpha_i + x_{it}'\beta) = 0$$

for α_i. Again, $\widehat{\alpha}_i(\beta; z_i)$ cannot be obtained in closed form. As in the logit model, we consider the case where x_{it} is binary to simplify the analysis. Then $\widehat{\alpha}_i(\beta; z_i)$ solves the simpler equation

$$\frac{1}{T}\left(z_{i,00}\frac{-\phi(\alpha_i)}{1-\Phi(\alpha_i)} + z_{i,01}\frac{-\phi(\alpha_i+\beta)}{1-\Phi(\alpha_i+\beta)} + z_{i,10}\frac{\phi(\alpha_i)}{1-\Phi(\alpha_i)} + z_{i,11}\frac{\phi(\alpha_i+\beta)}{1-\Phi(\alpha_i+\beta)}\right) = 0$$

for α_i, where $z_{i,yx}$ (for $y, x \in \{0,1\}$) is defined as in the logit model. We solve this equation numerically to obtain $\widehat{\alpha}_i(\beta; z_i)$ and subsequently obtain $\widehat{\beta}$ from solving $s(\beta, \widehat{\alpha}(\beta; z); z) = 0$ for β, and similarly for the bias-corrected estimators. Thus, our numerical procedure to obtain any estimate of β is a nested one, where the inner loop computes $\widehat{\alpha}_i(\beta; z_i)$.

Our setup to evaluate the performance of the bootstrap is similar to that for the logit model: $n = 10,000$, $T \in \{3, 4, 5, 10\}$, $B = 10$, and

$$\Pr[x_{it} = 1] = 1/2, \qquad \Pr[y_{it} = 1 | x_{it}] = \Phi(\alpha_{i0} + x_{it}\beta_0),$$

with $\beta_0 = 1$ and $\alpha_{i0} \sim \mathcal{N}(-0.5, 1)$. Table 3 reports the means and standard deviations of the estimators, estimated from $1,000$ Monte Carlo replications. Again, the bootstrap performs well, broadly similar to the jackknife.

$\beta_0 = 1$		$T = 3$	$T = 4$	$T = 5$	$T = 10$
$\widehat{\beta}$	mean	1.5962	1.3964	1.2946	1.1253
	std	0.0393	0.0272	0.0204	0.0114
$\widehat{\beta}_1^{\text{boot}}$	mean	1.2971	1.1030	1.0446	1.0032
	std	0.0374	0.0239	0.0173	0.0107
$\widehat{\beta}_2^{\text{boot}}$	mean	1.1155	0.9923	0.9856	0.9994
	std	0.0417	0.0267	0.0196	0.0125
$\widehat{\beta}_3^{\text{boot}}$	mean	1.0096	0.9633	0.9865	1.0031
	std	0.0510	0.0340	0.0248	0.0153
$\tilde{\beta}_1^{\text{boot}}$	mean	1.2553	1.1007	1.0617	1.0218
	std	0.0365	0.0244	0.0185	0.0108
$\tilde{\beta}_2^{\text{boot}}$	mean	1.0546	1.0193	1.0269	1.0133
	std	0.0426	0.0284	0.0223	0.0125
$\tilde{\beta}_3^{\text{boot}}$	mean	0.9535	1.0187	1.0341	1.0095
	std	0.0554	0.0361	0.0288	0.0154
$\check{\beta}_1^{\text{boot}}$	mean	1.4270	1.2184	1.1320	1.0284
	std	0.0369	0.0245	0.0180	0.0107
$\check{\beta}_2^{\text{boot}}$	mean	1.3049	1.1182	1.0580	1.0072
	std	0.0367	0.0242	0.0180	0.0116
$\check{\beta}_3^{\text{boot}}$	mean	1.2158	1.0620	1.0257	1.0035
	std	0.0382	0.0257	0.0195	0.0134
$\widehat{\beta}_1^{\text{jack}}$	mean	0.6873	0.8014	0.8844	0.9767
	std	0.0311	0.0161	0.0133	0.0097
$\widehat{\beta}_2^{\text{jack}}$	mean	NA	0.9160	1.0084	1.0014
	std	NA	0.0128	0.0149	0.0101

Table 3: Means and standard deviations of maximum likelihood and bias-corrected estimators in the panel probit model with fixed effects and a binary regressor. The bootstrap is applied to the estimator ($\widehat{\beta}^{\text{boot}}$), the profile score ($\tilde{\beta}^{\text{boot}}$), and the profile score at β_0 only ($\check{\beta}^{\text{boot}}$). Subscripts indicate the order of bias correction.

5 Conclusion

We have shown that the nonparametric bootstrap can be applied in (G)MM panel data models to reduce the estimation bias arising from the inclusion of fixed effects. In many panel data models of interest, including parametric models, this is the dominant source of bias when the number of periods is small. The implementation of the bootstrap is relatively straightforward and uses standard bootstrap bias correction formulas that need no tailoring to the given model. In addition to the common parameters defined by the model's moment conditions, the bootstrap can also be used to correct the bias of estimates of other functionals of the joint distribution of the data and the unobserved heterogeneity, such as average marginal effects or other functionals of their distribution. In our examples, we found that the bootstrap effectively reduces the bias of the uncorrected estimates while only mildly affecting the standard deviation, except when the number of periods is very small. The bootstrap tends to compare similarly to the jackknife, although sometimes the jackknife appears to perform better. This difference was most outspoken in an IV model with random slopes, which needs further study because the bias reduction rates, for the bootstrap and jackknife alike, appear to be slower than expected from the theory. In the other examples we studied, the bootstrap and the jackknife reduced the incidental parameter bias in a manner that aligns with the theory.

A Proof of Lemma 1

The score function for α_i is

$$\partial_{\alpha_i} l(\beta, \alpha; z) = \frac{z_{i,10}}{1 + e^{\alpha_i}} - \frac{z_{i,00} e^{\alpha_i}}{1 + e^{\alpha_i}} + \frac{z_{i11}}{1 + e^{\alpha_i + \beta}} - \frac{z_{i01} e^{\alpha_i + \beta}}{1 + e^{\alpha_i + \beta}}.$$

Solving $\partial_{\alpha_i} l(\beta, \alpha; z) = 0$ for α_i gives $\widehat{\alpha}_i(\beta; z_i)$. We can rearrange $\partial_{\alpha_i} l(\beta, \alpha; z) = 0$ as

$$a e^{2\alpha_i} + b e^{\alpha_i} + c = 0.$$

This is a quadratic equation in e^{α_i} with unique solution

$$e^{\alpha_i} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

since $e^{\alpha_i} > 0$ rules out the solution with the negative root. Taking the logarithm gives $\widehat{\alpha}_i(\beta; z_i)$.

References

Andersen, Erling Bernhard. 1970. "Asymptotic Properties of Conditional Maximum-Likelihood Estimators." *Journal of the Royal Statistical Society. Series B (Methodological)* 32 (2): 283–301.

Arellano, Manuel, and Stéphane Bonhomme. 2009. "Robust Priors in Nonlinear Panel Data Models." *Econometrica* 77 (2): 489–536.

Arellano, Manuel, and Jinyong Hahn. 2016. "A Likelihood-Based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects." *Global Economic Review* 45 (3): 251–274.

Bester, C. Alan, and Christian Hansen. 2009. "A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects." *Journal of Business & Economic Statistics* 27 (2): 131–148.

Bonhomme, Stéphane, and Elena Manresa. 2015. "Grouped patterns of heterogeneity in panel data." *Econometrica* 83 (3): 1147–1184.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *The Review of Economic Studies* 47 (1): 225–238.

———. 2010. "Binary Response Models for Panel Data: Identification and Information." *Econometrica* 78 (1): 159–168.

Dhaene, Geert, and Koen Jochmans. 2015. "Split-Panel Jackknife Estimation of Fixed-effect Models." *The Review of Economic Studies* 82 (3): 991–1030.

Dhaene, Geert, and Yutao Sun. 2021. "Second-Order Corrected Likelihood for Nonlinear Panel Models with Fixed Effects." *Journal of Econometrics,* Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future, 220 (2): 227–252.

Efron, Bradley. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM.

Fernández-Val, Iván, and Joonhwah Lee. 2013. "Panel Data Models with Nonadditive Unobserved Heterogeneity: Estimation and Inference." *Quantitative Economics* 4 (3): 453–481.

Fernández-Val, Iván, and Martin Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large N, T." *Journal of Econometrics* 192 (1): 291–312.

Hahn, Jinyong, and Guido Kuersteiner. 2002. "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large." *Econometrica* 70 (4): 1639–1657.

Hahn, Jinyong, and Whitney Newey. 2004. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models." *Econometrica* 72 (4): 1295–1319.

Hall, Peter, and Michael A. Martin. 1988. "On Bootstrap Resampling and Iteration." *Biometrika* 75 (4): 661–671.

Jochmans, Koen, and Ayden Higgins. 2022. "Bootstrap inference for fixed-effect models."

Kim, Min Seong, and Yixiao Sun. 2016. "Bootstrap and K-Step Bootstrap Bias Correction for Fixed Effects Estimators in Nonlinear Panel Data Models." *Econometric Theory* 32 (6): 1523–1568.

Neyman, J., and Elizabeth L. Scott. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16 (1): 1–32.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (6): 1417–1426.

Quenouille, M. H. 1949. "Approximate Tests of Correlation in Time-Series 3." *Mathematical Proceedings of the Cambridge Philosophical Society* 45 (3): 483–484.

———. 1956. "Notes on Bias in Estimation." *Biometrika* 43 (3-4): 353–360.

Schumann, Martin. 2022. "Second-order bias reduction for nonlinear panel data models with fixed effects based on expected quantities." *Econometric Theory,* 1–44.

Schumann, Martin, Thomas A. Severini, and Gautam Tripathi. 2021. "Integrated Likelihood Based Inference for Nonlinear Panel Data Models with Unobserved Effects." *Journal of Econometrics* 223 (1): 73–95.

Woutersen, Tiemen. 2002. *Robustness against Incidental Parameters.* UWO Department of Economics Working Paper 20028. University of Western Ontario, Department of Economics.