Quantitative Economics 15 (2024), 783–816 1759-7331/20240783

Efficient bias correction for cross-section and panel data

Jinyong Hahn
Department of Economics, U.C.L.A.

David W. Hughes
Department of Economics, Boston College

Guido Kuersteiner
Department of Economics, University of Maryland

Whitney K. Newey
Department of Economics, M.I.T. and NBER

Bias correction can often improve the finite sample performance of estimators. We show that the choice of bias correction method has no effect on the higher-order variance of semiparametrically efficient parametric estimators, so long as the estimate of the bias is asymptotically linear. It is also shown that bootstrap, jackknife, and analytical bias estimates are asymptotically linear for estimators with higher-order expansions of a standard form. In particular, we find that for a variety of estimators the straightforward bootstrap bias correction gives the same higher-order variance as more complicated analytical or jackknife bias corrections. In contrast, bias corrections that do not estimate the bias at the parametric rate, such as the split-sample jackknife, result in larger higher-order variances in the i.i.d. setting we focus on. For both a cross-sectional MLE and a panel model with individual fixed effects, we show that the split-sample jackknife has a higher-order variance term that is twice as large as that of the "leave-one-out" jackknife.

Keywords. Bias correction, higher-order variance, bootstrap, jackknife.

JEL classification. C13.

1. Introduction

Asymptotic bias corrections can be useful for centering estimators nearer to the truth. One approach is to use analytical corrections such as the standard textbook expansion for functions of sample means and the more complicated formulas required for other estimators. Alternatively, we may use jackknife and bootstrap bias corrections. To help

Jinyong Hahn: hahn@econ.ucla.edu
David W. Hughes: dw.hughes@bc.edu
Guido Kuersteiner: kuersteiner@econ.umd.edu
Whitney K. Newey: wnewey@mit.edu
We thank T. Rothenberg for providing helpful comments, and William Liu for valuable research assistance. Guido Kuersteiner gratefully acknowledges financial support from NSF Grants SES-0095132 and SES-0523186. Whitney Newey gratefully acknowledges financial support from NSF Grant 1757140.

Available at http://qeconomics.org. https://doi.org/10.3982/QE2350

choose a bias correction method it would be useful to know which, if any, is preferable on asymptotic efficiency grounds. Although the bias correction does not affect the first-order asymptotic variance, it can affect the higher-order variance. We show that the method of bias correction does not affect the higher-order variance of any parametric estimator that is efficient in a semiparametric model, as long as the bias estimator is asymptotically linear. Thus, one can choose a bias correction for an efficient estimator based on computational convenience, or some other criteria, without affecting its higher-order efficiency. We give a formal expansion showing this property for a parametric estimator in a general semiparametric model, that is, a model with a parametric component in which some other components are left unspecified. We also prove that the bootstrap, jackknife, and analytical bias estimates are asymptotically linear when the estimator of the parameters of interest has a standard form of stochastic expansion (which is known to exist for a large class of models). Derivations in the case of the MLE, show that the jackknife, bootstrap, and one type of analytical bias correction deliver estimators that have identical stochastic expansions to third order, and so have an even stronger equivalence property.

There are many implications of this higher-order efficiency result. One is that bias corrections that are not asymptotically linear may not have the same higher-order variance as those that are. For example, split-sample jackknife bias corrections are not asymptotically linear in cross-section or panel data and have a larger higher-order variance than other bias-corrected estimators. We find that the higher-order variance term is twice the size of that for the leave-one-out jackknife bias correction. On the other hand, the split-sample jackknife is useful in time series or panel data when the observations are not independent over time because the leave-one-out panel jackknife does not work in this case.

Another implication of the result is that it allows researchers to choose the bias correction method that is computationally convenient. For example, Newey and Smith (2004) showed that the empirical likelihood estimator is higher-order efficient in moment condition models when certain analytical bias corrections are used. Asymptotic linearity of the bootstrap means that calculation of the bias formula can be avoided by using the bootstrap bias correction instead. As another example, Cattaneo, Jansson, and Ma (2019) showed that a jackknife bias correction can be important when a first step regression with many regressors is plugged into a second step regression. Although our current results do not include asymptotics in which the number of regressors increases with the sample size, we conjecture that the bootstrap bias correction has similar properties to the jackknife bias correction when the number of regressors increases slowly enough.

The higher-order variance concept that we consider is the $O(n^{-1})$ variance of a third-order stochastic expansion of the estimator. Its use for comparison of estimators was pioneered by Nagar (1959). As shown in Pfanzagl and Wefelmeyer (1978) and Ghosh, Sinha, and Wieand (1980), and discussed in Rothenberg (1984), under appropriate regularity conditions rankings based on this higher-order variance correspond to rankings based on the variance of an Edgeworth approximation. Thus, the bias and variance of leading terms in a stochastic expansion are also the leading terms of an expansion of the bias and variance of an approximating distribution. Furthermore, as noted by Rothenberg (1984), Akahira and Takeuchi (1981) have shown that all well-behaved asymptotically efficient estimators of a parametric likelihood model necessarily have the same skewness and kurtosis to an order n^{-1} approximation, and that to compare the dispersion of second-order approximations to the distribution of efficient estimators, it suffices to compare their higher-order variances. This motivates our focus here on the higher-order variance of bias-corrected estimators, which serves to quantify their higher-order efficiency. In having this focus, we follow much of the more recent literature, such as Rilstone, Srivastava, and Ullah (1996) and Newey and Smith (2004).

We also derive asymptotic higher-order variance expressions for panel data models with unobserved, individual fixed effects. These estimators are known to suffer from asymptotic bias under asymptotic sequences in which n and T grow at the same rate. Two common methods of bias correction in this setting are the "leave-one-out" panel jackknife (Hahn and Newey (2004)), and the split-sample jackknife (Dhaene and Jochmans (2015)). Both deliver estimates that are asymptotically normal and centered at the truth when n and T grow at the same rate, with equal first-order (asymptotic) variances. Our analysis makes it possible to compare the two bias corrections in terms of their higher-order variance. We find that with i.i.d. data the split-sample correction has a higher-order variance that is twice the size of the "leave-one-out" jackknife. Although we focus on the maximum likelihood setting for panel data, the results are applicable to a broader set of moment condition estimators under suitable assumptions on the moment functions. Numerical comparisons in recent papers confirm that the difference in higher-order variance can be meaningful in practice, with the split-sample jackknife having larger dispersion and lower coverage than analytical or leave-one-out jackknife corrections in a variety of settings; see, for example, Alexander and Breunig (2016), Fernández-Val and Weidner (2018), Czarnowske and Stammann (2019). Our own simulations of a panel probit model with individual fixed effects also support this result. This comparison is also true of estimates of a marginal effect parameter.

1.1 *Related literature*

Higher-order efficiency of the MLE was analyzed by Pfanzagl and Wefelmeyer (1978) in terms of risk functions or Akahira and Takeuchi (1981) in terms of concentration probabilities. Ghosh (1994), and Taniguchi and Kakizawa (2000) for time-series models, contain surveys of this literature. A common theme is that a higher-order bias-corrected version of the MLE is higher-order efficient in the case of higher-order squared risk. Similar results obtain for median bias-corrected MLEs in the case of concentration probabilities. The bias correction in this literature is typically of a known parametric form that only depends on the estimated parameters. A plug-in estimator is then a regular estimator for the bias term. Amari (1982) obtains similar results for curved exponential families using differential geometry that characterizes the MLE in terms of tangent spaces. Akahira (1983, 1989) shows that when, instead of using parametric bias correction, one relies on the jackknife, the same higher-order efficiency results remain true for the jackknifed MLE. We add to this literature by analyzing the effects of bias correction for first-order, but not necessarily higher-order, semiparametrically efficient estimators.

We demonstrate that any bias-corrected semiparametrically efficient estimator using an asymptotically linear bias estimator has a higher-order variance that does not depend on the nature of the bias estimate. This result applies to cases where the higher-order bias may not be known in closed form. We show that efficient bias correction may be based on sample averages, bootstrap, or jackknife methods as long as the bias estimates are asymptotically linear.

The use of a jackknife bias estimator goes back to Quenouille (1949, 1956) and Tukey (1958). Bootstrap bias estimation was discussed by Parr (1983), Shao (1988b), Hall (1992), and Horowitz (1998) in the context of nonlinear transformations of OLS estimators of linear models and nonlinear functions of the mean. Akahira (1983) showed that the jackknife bias-corrected maximum likelihood estimator is higher-order efficient. Our work extends the literature on the bootstrap and jackknife bias-corrected estimators by analyzing any semiparametrically efficient parametric estimator rather than nonlinear transformations of linear estimators as in Shao (1988a,b).

In the panel data literature, methods to control for unobserved heterogeneity are well established (early literature includes Rasch (1960, 1961) and Andersen (1970)); see, for example, Chamberlain (1984), Arellano and Honoré (2001), and Arellano and Hahn (2010) for reviews. Because of the incidental parameters problem, the best that can be achieved in a fixed-T setting is partial identification in general; this is especially true for policy relevant parameters such as average marginal effects (see Chernozhukov, Newey, Hahn, and Fernández-Val (2013)). Even under sequences in which T grows at the same rate as n, fixed effects estimators may be asymptotically biased, as discussed in Hahn and Kuersteiner (2002), Hahn and Newey (2004). Given the typical size of panel data sets, in which n is much larger than T, it is desirable to find estimators that have biases of order $O(T^{-2})$ or smaller, rather than the typical $O(T^{-1})$ of fixed effects estimators.

For a static model, Hahn and Newey (2004) show that a "leave-one-out" jackknife estimator is asymptotically normal and centered at the truth when n and T grow at the same rate. Other styles of jackknife bias correction are also possible. For example, Dhaene and Jochmans (2015) suggest a split-sample bias correction that, in its simplest form, is constructed by splitting the sample into two half-panels of length $T/2$. It should be understood that the split-sample jackknife provides valid bias corrections with autocorrelated data where the "leave-one out" jackknife does not. Thus, the split-sample jackknife is preferred to the leave-one-out jackknife with autocorrelated data. Our results show that the leave-one-out jackknife is preferred to the split-sample jackknife in i.i.d. data in the sense that the higher-order variance is smaller and small sample performance is better.

The remainder of the paper is set out as follows. In Section 2, we discuss the higher-order bias and variance of parametric estimators, and provide our main results on the higher-order efficiency of bias corrections for semiparametric efficient estimators and asymptotic linearity of analytical, bootstrap, and jackknife bias corrections. In Section 3, we provide expressions for the higher-order expansions and variances of various bias-corrections in a cross-sectional MLE. We extend the results to panel settings in Section 4, by deriving the higher-order variances of the leave-one-out and split-sample jackknifes in a model with individual fixed effects. Section 5 provides Monte Carlo evidence to support the theory in the panel setting. Section 6 concludes. Proofs and additional derivations can be found in Supplementary Appendices contained in the working paper Hahn, Hughes, Kuersteiner, and Newey (2024).

2. Higher-order efficiency of bias-corrected estimators

2.1 *Higher-order bias and variance*

We begin with a discussion of higher-order expansions for parametric estimators, and make precise our definition of higher-order variance, following closely the exposition in Rothenberg (1984). We focus on semiparametric models in which the data $\{Z_i\}_{i=1}^n$ are i.i.d. with some distribution F_0 contained in a class of distributions that is the model.[1] Let $\widehat{\theta}$ denote an estimator of a parameter θ_0. When $\widehat{\theta}$ is an estimator based on parametric moment conditions, such as GMM, and the moment conditions are smooth enough in the parameter θ and possibly other parameters, there will be a stochastic expansion of the form

$$\begin{gathered}\sqrt{n}(\widehat{\theta}-\theta_0)=A_1+\frac{A_2}{\sqrt{n}}+\frac{A_3}{n}+o_p\big(n^{-1}\big),\\ A_1=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi(Z_i),\quad E\big[\psi(Z_i)\big]=0,\end{gathered}\tag{1}$$

where A_2 and A_3 are second- and third-order products of sample averages of mean-zero random variables, each multiplied by $\sqrt{n}$; see, for example, Rilstone, Srivastava, and Ullah (1996). The terms A_1, A_2, and A_3 are bounded in probability so that equation (1) is an expansion where the stochastic order of terms is smaller as one moves to the right in the expansion.

We define the higher-order bias and variance of estimators using the first and second moments of the leading three terms in this expansion. We also compare higher-order efficiency of bias-corrected estimators by comparing their higher-order variance, as discussed in the Introduction. The higher-order bias of $\sqrt{n}(\widehat{\theta}-\theta_0)$ is given by $E[A_2]/\sqrt{n}$. This follows from the fact that $E[A_1]=0$, while the expectation of A_3/n is generally of smaller order than $E[A_2]/\sqrt{n}$. Dividing through by $\sqrt{n}$, the higher-order bias of $\widehat{\theta}$ is

$$\text{Bias}(\widehat{\theta})\approx\frac{b_0}{n},\quad b_0=E[A_2].$$

In general $b_0=E[B(Z)]$, where the function $B(z)$ captures the "own observation" term in A_2, that is, where the observation indices coincide in the product of sample means that make up A_2. The term $B(z)$ may depend on the distribution of the data through the parameter θ or in other ways. We let this dependence be implicit for notational convenience.

[1] By "semiparametric model," we refer to a model which has a parametric component, but leaves the functional form of some other components unspecified. See Newey (1990) for further discussion.

Similarly, the higher-order variance of $\sqrt{n}(\widehat{\theta}-\theta_0)$ can be obtained from the variance of the first three expansion terms, up to order $1/n$. This gives

$$\begin{aligned} \operatorname{Var}\big(\sqrt{n}(\widehat{\theta}-\theta_0)\big) &\approx \operatorname{Var}(A_1)+\frac{v}{n}, \\ v &:= \operatorname{Var}(A_2)+2\sqrt{n}E[A_1A_2]+2E[A_1A_3], \end{aligned}$$

where $\operatorname{Var}(A_1)$ is the asymptotic (first-order) variance of the estimator and we refer to v as the higher-order variance.

Example 1. In order to demonstrate ideas, we will follow a simple example. Suppose that we observe a sample of observations $Z_i \sim N(\sqrt{\theta}, 1)$. The MLE provides an efficient estimate for θ, and is given by $\widehat{\theta}=(\frac{1}{n}\sum_{i=1}^n Z_i)^2$. It can be shown that the asymptotic expansion for this estimator is of the form in (1), with

$$A_1 = 2\sqrt{\theta}\frac{1}{\sqrt{n}}\sum_i (Z_i-\sqrt{\theta}), \qquad A_2 = \left(\frac{1}{\sqrt{n}}\sum_i (Z_i-\sqrt{\theta})\right)^2, \tag{2}$$

and $A_3 = 0$.[2] In this case, since the estimator is quadratic, the expansion is exact. We can conclude that the bias of the estimator is $E[A_2]/n = 1/n$. Since $E[A_1A_2]=0$ in this example, the higher-order variance is

$$\operatorname{Var}\big(\sqrt{n}(\widehat{\theta}-\theta)\big) = \operatorname{Var}(A_1)+\operatorname{Var}(A_2)/n = 4\theta+2/n.$$

2.2 *The effect of bias correction*

A bias-corrected estimator can be formed by subtracting off an estimator $\widehat{b}$ of b_0 from $\widehat{\theta}$,

$$\tilde{\theta} = \widehat{\theta} - \frac{\widehat{b}}{n}.$$

The focus of this paper is on the effect of the choice of $\widehat{b}$ on the variance of $\tilde{\theta}$. Generally, $\widehat{b}$ has no effect on the asymptotic variance (i.e., the first-order variance) of $\tilde{\theta}$, as long as $\widehat{b}$ is bounded in probability, because then $\sqrt{n}(\widehat{b}/n) = \widehat{b}/\sqrt{n} = o_p(1)$. The choice of $\widehat{b}$ can have an effect on the higher-order variance of $\tilde{\theta}$. To describe this effect note that the asymptotic expansion for $\widehat{\theta}$ implies that, by adding and subtracting $b_0/\sqrt{n}$,

$$\begin{aligned} \sqrt{n}(\tilde{\theta}-\theta_0) &= A_1 + \frac{A_2 - b_0}{\sqrt{n}} + \frac{A_3 - \sqrt{n}(\widehat{b}-b_0)}{n} + o_p\big(n^{-1}\big) \\ &= A_1 + \frac{\tilde{A}_2}{\sqrt{n}} + \frac{\tilde{A}_3}{n} + o_p\big(n^{-1}\big), \end{aligned} \tag{3}$$

where $\tilde{A}_2 = A_2 - b_0$, and $\tilde{A}_3 = A_3 - \sqrt{n}(\widehat{b}-b_0)$. This is again an expansion whose terms are of decreasing stochastic order, so long as $\widehat{b}$ is $\sqrt{n}$-consistent so that the third term

[2] See Section H in the Supplementary Appendix for details of this example.

is smaller order than the second term. Importantly, the second-order term $\tilde{A}_2/\sqrt{n}$ has expectation zero, so that the bias-corrected estimator $\tilde{\theta}$ is higher-order unbiased.

To analyze the higher-order variance of the bias-corrected estimator, it is helpful to be more specific about $\widehat{b}$. For now, we assume $\widehat{b}$ is asymptotically equivalent to a sample average, that is, that there exists a $\phi(z)$ with $E[\phi(Z)] = 0$, $\text{Var}(\phi(Z)) < \infty$, such that

$$\sqrt{n}(\widehat{b} - b_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(Z_i) + o_p(1).$$

Here, we are assuming that $\widehat{b}$ is asymptotically linear with influence function $\phi(z)$. Bias corrections that are based on directly estimating $b_0 = E[B(Z)]$ will often have this property and we find that other bias corrections, like the jackknife and bootstrap also have this property. In this case, the expansion in equation (3) continues to hold with $\tilde{A}_3 = A_3 - \Delta$ for $\Delta := \sum_{i=1}^{n} \phi(Z_i)/\sqrt{n}$. The variance of the third-order approximation for $\tilde{\theta}$ is $\text{Var}(A_1) + \tilde{v}/n$, with higher-order variance:

$$\begin{aligned} \tilde{v} &:= \text{Var}(\tilde{A}_2) + 2\sqrt{n}E[A_1\tilde{A}_2] + 2E[A_1\tilde{A}_3] \\ &= \text{Var}(A_2 - b_0) + 2\sqrt{n}E\big[A_1(A_2 - b_0)\big] + 2E\big[A_1(A_3 - \Delta)\big] \\ &= v - 2E[A_1\Delta]. \end{aligned} \tag{4}$$

Higher-order efficiency of the bias-corrected estimator refers to the size of $\tilde{v}$. One bias-corrected estimator is higher-order more efficient than another if it has smaller $\tilde{v}$.

The contribution of the bias correction to the higher-order variance $\tilde{v}$ is through the term

$$-2E[A_1\Delta] = -2E\big[\psi(Z)\phi(Z)\big]. \tag{5}$$

Thus, the bias correction $\widehat{b}$ affects the higher-order variance only through the covariance of its influence function with that of $\widehat{\theta}$. In the following theorem, we show that when $\tilde{\theta}$ is an efficient estimator, its higher-order variance does not depend on the choice of $\widehat{b}$, by demonstrating that the covariance in (5) is the same for any $\phi(Z)$. We adopt the notation and terminology of Newey (1990, pp. 104–106) for the statement and proof of this result. Here, $\mathcal{T}$ denotes the tangent set for the semiparametric model, which is the mean-square closure of the set of all scores for regular parametric submodels. A parameter is differentiable if there is a random variable $d(Z)$ such that the derivative of the parameter with respect to parameters of any regular parametric submodel exists and equals the expected product of $d(Z)$ with the submodel score. Under additional regularity conditions, the influence function of any asymptotically linear estimator will be equal to such a $d(Z)$.

Theorem 1. *If $\mathcal{T}$ is linear, θ_0 and b_0 are differentiable parameters of the semiparametric model, $\widehat{\theta}$ is asymptotically linear and efficient with influence function $\psi(Z)$, and $\widehat{b}$ is asymptotically linear with influence function $d(Z)$, then the higher-order variance $\tilde{v}$ does not depend on $\widehat{b}$.*

Proof. It follows from asymptotic linearity and efficiency of $\widehat{\theta}$ that its influence function $\psi(Z)$ is an element of $\mathcal{T}$. By Theorem 3.1 of Newey (1990), the efficient influence function $\delta(Z)$ for b_0 is equal to the projection of $d(Z)$ on $\mathcal{T}$. This allows us to decompose $d(Z)$ as

$$d(Z) = \delta(Z) + U(Z), \qquad E\big[U(Z)t(Z)\big] = 0 \quad \text{for all } t \in \mathcal{T}.$$

As shown above in (5), the higher-order variance of $\tilde{\theta} = \widehat{\theta} - \widehat{b}/n$, depends on $\widehat{b}$ only through the covariance of $\psi(Z)$ with $d(Z)$. By the above decomposition, since $\psi(Z)$ is in the tangent set, this covariance is

$$E\big[\psi(Z)\, d(Z)\big] = E\big[\psi(Z)\delta(Z)\big] + E\big[\psi(Z)U(Z)\big] = E\big[\psi(Z)\delta(Z)\big].$$

Since the efficient influence function $\delta(Z)$ is unique, it follows that $E[\psi(Z)\, d(Z)]$ does not vary with $d(Z)$, and hence $\tilde{v}$ does not vary with $\widehat{b}$. □

The proof uses geometry associated with semiparametric models, and relies on the efficiency of $\widehat{\theta}$. There is also some relatively simple intuition for this result. Consider any two semiparametric estimators $\widehat{b}_1$ and $\widehat{b}_2$ of b_0. The random variable $\widehat{b}_1 - \widehat{b}_2$ is a semiparametric estimator of 0. If the asymptotic covariance of $\widehat{\theta}$ with $\widehat{b}_1 - \widehat{b}_2$ were nonzero then for some fixed constant C the asymptotic variance of $\bar{\theta} = \widehat{\theta} + C(\widehat{b}_1 - \widehat{b}_2)$ would be less than the asymptotic variance of $\widehat{\theta}$, so that $\widehat{\theta}$ could not be semiparametrically efficient. That is, semiparametric efficiency of $\widehat{\theta}$ implies zero asymptotic covariance of $\widehat{\theta}$ with $\widehat{b}_1 - \widehat{b}_2$, which is equivalent to the asymptotic covariance of $\widehat{\theta}$ with $\widehat{b}_1$ being equal to the asymptotic covariance of $\widehat{\theta}$ with $\widehat{b}_2$. The asymptotic covariance of $\widehat{\theta}$ with any asymptotically linear $\widehat{b}$ is $E[\psi(Z)\phi(Z)]$, so semiparametric efficiency of $\widehat{\theta}$ implies that this covariance does not depend on $\phi(Z)$. This intuition extends the Hausman (1978) result that the covariance of any estimator of a parameter of interest with an efficient estimator of that parameter is equal to the variance of the efficient estimator, that is, the covariance does not depend on the estimator. The intuition and result here show that the covariance of an efficient estimator of a parameter of interest with an estimator of any object does not depend on the estimator of that object.

2.3 *Asymptotically linear bias correction*

There are many examples of bias corrections to semiparametric estimators that have different influence functions. Consider a parametric model of the conditional pdf of an outcome variable Y given regressors X, where the pdf of Y conditional on X has a parametric form $f(y|x, \beta)$ and θ is some function of β. This is a familiar semiparametric model where the marginal distribution of X is unspecified. A special case is a parametric likelihood model where X is a constant and $f(y|x, \beta)$ specifies the unconditional pdf of observation Y. Since generally $b_0 = E[B(Z)]$, for $Z = (Y, X)$ and for some function $B(z)$ that may also depend on β, one could devise a number of estimates for b_0. For example, an analytical bias correction could be obtained by plugging in estimates $\widehat{\beta}$ and replacing the expectation operators with sample averages. Alternative forms of analytical bias correction could make use of the structure imposed by $f(y|x, \beta)$ to integrate over y in $B(z)$, or use restrictions implied by $\int f(y|x, \beta)\, dy = 1$ to estimate b_0 (e.g., applying the information equality).

One could also use a nonparametric method like the bootstrap or jackknife to estimate b_0. The bootstrap bias correction estimates the bias by $\frac{1}{n}\widehat{b}_B \equiv E^*[\widehat{\theta}^*] - \widehat{\theta}$, where $\widehat{\theta}^*$ are estimates obtained using bootstrap samples from the empirical distribution of Z. Alternatively, the jackknife estimates the bias by taking the difference between $\widehat{\theta}$ and the average of all "leave-one-out" estimates, that is, estimates formed by excluding a single observation. The jackknife bias estimator is given by $\frac{1}{n}\widehat{b}_J = (n-1)(\frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}_{(i)} - \widehat{\theta})$ where $\widehat{\theta}_{(i)}$ is the estimate of the parameter that excludes observation i. In Section 3, we describe these bias correction techniques in detail in the context of MLE.

Example 1 (Continued). Continuing our earlier example, we can compute the jackknife and analytical bias corrections for $\widehat{\theta}$. From the asymptotic expansion in (2), we have that

$$E[A_2] = E\big[(Z_i - \sqrt{\theta})^2\big]$$

which suggests the analytical bias estimate $\widehat{b_a} = \frac{1}{n}\sum_i (Z_i - \bar{Z})^2$. Alternatively, using the known variance, we could simply use $b_a = 1$. To construct a jackknife bias estimator, we use

$$\frac{1}{n}\widehat{b}_J = (n-1)\left(\frac{1}{n}\sum_i\left(\frac{1}{n-1}\sum_{j\neq i} Z_j\right)^2 - \left(\frac{1}{n}\sum_i Z_i\right)^2\right) = \frac{1}{n(n-1)}\sum_i (Z_i - \bar{Z})^2$$

which is the same as the first analytical bias estimate, up to the factor $n/(n-1)$. It is straightforward to show that both bias estimates can be written as

$$\sqrt{n}(\widehat{b} - 1) = \frac{1}{n}\sum_i\big((Z_i - \sqrt{\theta})^2 - 1\big) + o_p(1)$$

so that they are both asymptotically linear, in this case with the same influence function $d(Z) = (Z - \sqrt{\theta})^2 - 1$. Given that the third moment of Z is equal to zero, since it is normally distributed, we have $E[A_1(\widehat{b} - 1)] = 0$, so that the higher-order variance of the bias-corrected estimators that use either of the analytical or the jackknife bias estimates, are the same as that of $\widehat{\theta}$, that is, $4\theta + 2/n$.

Theorem 1 implies that the various analytical bias corrections as well as both nonparametric bias correction methods lead to the same higher-order variance as any other method when the $\widehat{b}$ is asymptotically linear. In the example above, both the analytical and jackknife bias estimates are asymptotically linear. This is not a coincidence. It turns out that for a fairly general class of models, the analytical, jackknife, and bootstrap bias estimates are asymptotically linear, so that Theorem 1 can be applied.[3]

[3]The theorem is stated for a scalar parameter θ, but straightforwardly extends to a vector-valued parameter, at the cost of more complicated notation.

THEOREM 2. *Let $\widehat{\theta}$ be an estimator with a stochastic expansion*

$$\widehat{\theta} - \theta_0 = \frac{1}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + \frac{1}{n^{3/2}} A_3 + \frac{1}{n^2} A_4 + \frac{1}{n^{5/2}} A_5 + o_p\big(n^{5/2}\big), \tag{6}$$

where $n^{-k/2} A_k$ is a kth order V-statistic, that is,

$$\frac{1}{n^{k/2}} A_k = \frac{1}{n^k} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} g_{k,1}(z_{i_1}, \theta_0) \cdots g_{k,k}(z_{i_k}, \theta_0) \tag{7}$$

and $g_{k,j}$ are functions of the data that are continuously differentiable in θ at θ_0, mean zero, that is, $E[g_{j,k}(z_i, \theta_0)] = 0$, and with $E[g_{j,k}(z_i, \theta_0)^{10}] \le C < \infty$.

Then the jackknife, bootstrap, and (sample average) analytical[4] bias estimates of the higher-order bias $b_0 = E[A_2] = E[g_{2,1}(z_i, \theta_0) g_{2,2}(z_i, \theta_0)]$, are asymptotically linear.

PROOF. See Appendix A. □

The V-statistic structure assumed in Theorem 2 is shown to hold for the case of MLE under standard regularity conditions on the likelihood function (see Supplementary Appendix I A). In Section 4, we derive such expansions for a maximum likelihood estimator. The derivation is based on an expansion of the first-order condition of the MLE and so applies to any other estimator that can be characterized by a moment equation satisfying similar regularity conditions. Therefore, the structure can be shown to appear in the expansion of any M-estimator, that is, an estimator that maximizes the sample average of some function, which need not be a log-likelihood.

An important semiparametric model is a model of unconditional moment restrictions that motivates GMM estimation. Newey and Smith (2004) derive the asymptotic expansion for generalized empirical likelihood (GEL) estimators up to third order and examination of that expansion shows the same V-statistic structure. Under sufficient regularity conditions, the structure should also exist in the fifth-order expansion used in Theorem 2, so that the equivalence result of Theorem 1 would imply that jackknife and bootstrap bias corrections lead to the same higher-order variance for GEL estimators. This extends the result in Newey and Smith (2004), who found that averaging over an efficient estimator of the distribution of the data to obtain $\widehat{b}$ does not affect the higher-order variance of bias-corrected GMM and generalized empirical likelihood (GEL) estimators.

Another interesting example is a nonparametric model where no restrictions are placed on the distribution of the data. In this case, b_0 will also be a nonparametric object and there will exist only one influence function corresponding to b_0; see van der Vaart (1991) and Newey (1994). Here, all asymptotically linear estimators of b_0 will have the same influence function so that Theorem 1 holds trivially. In particular, when general misspecification is allowed for in debiasing, so that the distribution of the data is unrestricted, there will only be one influence function for b_0. Then any bias correction

[4]See Section 3 for exact definitions of these three bias corrections, $\widehat{\theta}_J$, $\widehat{\theta}_B$, and $\tilde{\theta}_a$.

method, such as the analytical, bootstrap, or jackknife, must have the same higher-order variance, as long as they are asymptotically linear.

We emphasize that Theorem 1 only applies to comparisons between bias corrections for a given semiparametrically efficient estimator. The stochastic expansion term A_2 may differ across semiparametrically efficient estimators so that their higher-order variances need not be the same. For example, Newey and Smith (2004) showed that GMM will tend to have larger second-order bias than GEL when there are many instrumental variables, so that A_2 is different for GMM and GEL. Theorem 1 has nothing to say about how the higher-order variance of bias-corrected GMM compares with that of bias-corrected GEL. It only implies that the form of $\widehat{b}$ does not affect the higher-order variance of bias-corrected GMM and separately for bias-corrected GEL.

Note that while Theorem 2 implies that bootstrap, jackknife, and analytical bias estimates are asymptotically linear in general, this is not true of all bias estimates. An example of this is the split-sample jackknife, which estimates the bias using $\frac{1}{n}\widehat{b}_{\text{SS}} = \frac{1}{2}(\widehat{\theta}_{(1)} + \widehat{\theta}_{(2)}) - \widehat{\theta}$, where $\widehat{\theta}_{(1)}$ and $\widehat{\theta}_{(2)}$ are estimates using two separate halves of the data set.

Example 1 (Continued). Let $N = 2m$, and define $\xi^{(1)} = \frac{1}{\sqrt{m}}\sum_{i\le m}(Z_i - \sqrt{\theta})$ and $\xi^{(2)} = \frac{1}{\sqrt{m}}\sum_{i>m}(Z_i - \sqrt{\theta})$. Some algebra gives the expression

$$\widehat{b}_{\text{SS}} = \frac{1}{2}\big(\xi^{(1)} - \xi^{(2)}\big)^2.$$

Since $\xi^{(1)}$ and $\xi^{(2)}$ are independent standard normal variables, we have $E[\widehat{b}_{\text{SS}}] = 1$, so that the split-sample jackknife gives an unbiased estimator for b_0. However, $\widehat{b}_{\text{SS}} = O_p(1)$, so that the bias estimator is also inconsistent, and hence not asymptotically linear. In this case, the higher-order variance of the bias-corrected estimator can be shown to be $\text{Var}(\sqrt{n}(\widehat{\theta}_{\text{SS}})) = 4\theta + 4/n$, which is larger than that of the analytical and jackknife estimators. In Section 3.3, we give a formal result that shows that this larger higher-order variance is true in general for MLE.

3. Bias-corrected MLE

In this section, we demonstrate the higher-order equivalence of bias corrections in a fully parametric model by deriving higher-order expansions and variance expressions for several bias-corrected maximum likelihood estimators. We consider the bootstrap bias correction, jackknife bias correction, as well as three different versions of analytical bias correction, including the bias-corrected MLE of Pfanzagl and Wefelmeyer (1978), which was shown to be third-order optimal. We show that all of these corrections lead to estimators with the same higher-order variance. Further, we show that the bootstrap, jackknife, and a particular analytical bias-correction, result in identical third-order asymptotic expansions.

To describe the parametric model, let $\{Z_i\}_{i=1}^n$ be an i.i.d. sample $Z_i \sim f(z,\theta_0)$, such that $f(z,\theta)$ satisfies sufficient smoothness conditions.[5] The density $f(z,\theta)$ is a member

[5] The results in Section 3 are predicated on regularity conditions, including Conditions 1, 2, and 3, which are presented in Appendix B.1.

of a parametric family of distributions P_θ indexed by $\theta \in \Theta$ with $\Theta \in \mathbb{R}$ a compact set.[6] We consider properties of the MLE $\widehat{\theta}$, where

$$\widehat{\theta} \equiv \arg\sup_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} \log f(Z_i, \theta).$$

We adopt the following set of notation in this section of the paper. We let $\ell(\cdot, \theta) \equiv .\partial \log f(\cdot, \theta)/\partial\theta$, $\ell^\theta(\cdot, \theta) \equiv .\partial^2 \log f(\cdot, \theta)/\partial\theta^2$, $\ell^{\theta\theta}(\cdot, \theta) \equiv .\partial^3 \log f(\cdot, \theta)/\partial\theta^3$, etc. We also define $\mathcal{I} \equiv -E[\ell^\theta(Z_i, \theta_0)]$, $\mathcal{Q}_1(\theta) \equiv E[\ell^{\theta\theta}(Z_i, \theta)]$, and $\mathcal{Q}_2(\theta) \equiv E[\ell^{\theta\theta\theta}(Z_i, \theta)]$. Furthermore, we let $U_i(\theta) \equiv \ell(Z_i, \theta)$, $V_i(\theta) \equiv \ell^\theta(Z_i, \theta) - E[\ell^\theta(Z_i, \theta)]$, $W_i(\theta) \equiv \ell^{\theta\theta}(Z_i, \theta) - E[\ell^{\theta\theta}(Z_i, \theta)]$. Finally, we define $U(\theta) \equiv n^{-1/2} \sum_{i=1}^n U_i(\theta)$, $V(\theta) \equiv n^{-1/2} \sum_{i=1}^n V_i(\theta)$, and $W(\theta) \equiv n^{-1/2} \sum_{i=1}^n W_i(\theta)$. We next describe the bias corrections and derive higher-order properties of the bias-corrected estimators.

3.1 *The bootstrap and jackknife bias corrections*

The bootstrap estimator constructs an estimate of the higher-order bias as the difference between the average of bootstrap replicates $\widehat{\theta}^*$ and the MLE $\widehat{\theta}$. In particular, we first obtain bootstrapped estimates $\widehat{\theta}^*$ by sampling $Z_1^*, \dots, Z_n^*$ identically and independently from the empirical distribution $\widehat{F}(z) = \frac{1}{n} \sum 1\{Z_i \leq z\}$. Let E^* be the expectation operator with respect to $\widehat{F}$. The idea behind the bootstrap bias correction is to estimate $E[\widehat{\theta}] - \theta_0$, if it exists, by $\frac{1}{n}\widehat{b}_B \equiv E^*[\widehat{\theta}^*] - \widehat{\theta}$. This in turn will allow us to construct the bias-corrected estimate

$$\widehat{\theta}_B \equiv \widehat{\theta} - \frac{\widehat{b}_B}{n} = 2\widehat{\theta} - E^*\big[\widehat{\theta}^*\big].$$

An alternative nonparametric bias correction is a jackknife bias-corrected estimator. The jackknife estimates the higher-order bias by taking the difference between the MLE and the average of all "leave-one-out" estimates, that is, estimates formed by excluding a single observation. The jackknife estimate is given by

$$\widehat{\theta}_J = \widehat{\theta} - \frac{\widehat{b}_J}{n} = n\widehat{\theta} - \frac{n-1}{n} \sum_{i=1}^{n} \widehat{\theta}_{(i)},$$

where $\widehat{\theta}_{(i)}$ is the estimate of the parameter that excludes observation i. The jackknife uses the bias estimate $\frac{1}{n}\widehat{b}_J = (n-1)(\frac{1}{n}\sum_{i=1}^n \widehat{\theta}_{(i)} - \widehat{\theta})$.

The following proposition establishes the higher-order properties of the bootstrap and jackknife bias-corrected MLEs.

Proposition 1. *Let $\tilde{b}$ be either the bootstrap bias estimate $\widehat{b}_B$, or the jackknife bias estimate $\widehat{b}_J$. Then, under regularity conditions stated in Appendix B.1, $\tilde{b}$ satisfies*

$$\sqrt{n}(\tilde{b} - b_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{B}_i + o_p(1),$$

[6]For simplicity of notation, we will assume that $p = 1$, where $p = \dim(\theta)$. The result is expected to hold for any finite p.

where

$$\mathbb{B}_i = \left(\frac{1}{2}\mathcal{I}^{-3}\mathcal{Q}_2 + \frac{3}{2}\mathcal{I}^{-4}\mathcal{Q}_1^2 + 3\mathcal{I}^{-4}\mathcal{Q}_1 E[U_iV_i] + \mathcal{I}^{-3}\big(E[U_iW_i] + E\big[\big(\ell^\theta\big)^2\big]\big)\right)U_i$$
$$+ \left(\frac{3}{2}\mathcal{I}^{-3}\mathcal{Q}_1 + 2\mathcal{I}^{-3}E[U_iV_i]\right)V_i + \frac{1}{2}\mathcal{I}^{-2}W_i$$
$$+ \frac{1}{2}\mathcal{I}^{-3}\mathcal{Q}_1\big(\ell(Z_i,\theta_0)^2 - E\big[\ell^2\big]\big) + \mathcal{I}^{-2}\big(\ell(Z_i,\theta_0)\ell^\theta(Z_i,\theta_0) - E\big[\ell\ell^\theta\big]\big),$$

and hence, for $\tilde{\theta}$ either the bootstrap or jackknife bias-corrected MLE,

$$\sqrt{n}(\tilde{\theta} - \theta_0) = A_1 + \frac{1}{\sqrt{n}}\big(A_2 - b(\theta_0)\big) + \frac{1}{n}(A_3 - \mathbb{B}) + o_p\big(n^{-1}\big),$$

where $\mathbb{B} = \sum_i \mathbb{B}_i/\sqrt{n}$.

Here, we see that the jackknife and bootstrap bias estimates are $\sqrt{n}$-consistent and asymptotically linear (note that $E[\mathbb{B}_i] = 0$), which from Theorem 1 implies that their higher-order variances are the same as any other estimator that uses an asymptotically linear estimator of the bias. In fact, the jackknife and bootstrap share an even stronger equivalence, in the sense that their bias estimates have identical influence functions, which implies that their asymptotic expansions are identical up to the third order.

3.2 *Analytical bias corrections*

We next introduce and discuss three forms of analytical bias correction. An analytical expression for the higher-order bias is given by

$$b(\theta_0) \equiv E[A_2] = \frac{1}{2\mathcal{I}^3}E\big[\ell^{\theta\theta}\big]E\big[\ell^2\big] + \frac{1}{\mathcal{I}^2}E\big[\ell\ell^\theta\big]. \tag{8}$$

The first bias correction is based on the bias formula (8), replacing expectations with sample averages. This gives the estimator $\widetilde{\theta}_a \equiv \widehat{\theta} - \frac{\widetilde{b}(\widehat{\theta})}{n}$, where

$$\widetilde{b}(\widehat{\theta}) = -\frac{\left(\frac{1}{n}\sum_i \ell^{\theta\theta}(Z_i,\widehat{\theta})\right)\left(\frac{1}{n}\sum_i \ell(Z_i,\widehat{\theta})^2\right)}{2\left(\frac{1}{n}\sum_i \ell^\theta(Z_i,\widehat{\theta})\right)^3} + \frac{\left(\frac{1}{n}\sum_i \ell(Z_i,\widehat{\theta})\ell^\theta(Z_i,\widehat{\theta})\right)}{\left(\frac{1}{n}\sum_i \ell^\theta(Z_i,\widehat{\theta})\right)^2}. \tag{9}$$

It can be shown that this form of analytical bias estimate shares the same influence function as the bootstrap and jackknife bias estimators. It follows that it has an identical third-order asymptotic expansion.

Proposition 2. *Let the regularity conditions stated in Appendix B.1 hold. The estimator $\widetilde{\theta}_a$ has the higher-order expansion*

$$\sqrt{n}(\widetilde{\theta}_a - \theta_0) = A_1 + \frac{1}{\sqrt{n}}\big(A_2 - b(\theta_0)\big) + \frac{1}{n}(A_3 - \mathbb{B}) + o_p\big(n^{-1}\big).$$

Alternative analytical corrections can be constructed by imposing the information equality, $E[\ell^2]=\mathcal{I}$ in the characterization of the bias, that is, using the bias formula

$$b(\theta_0)\equiv E[A_2]=\mathcal{I}^{-2}\left(\frac{1}{2}E\left[\ell^{\theta\theta}\right]+E\left[\ell\ell^{\theta}\right]\right). \tag{10}$$

An analytical bias correction that uses the information equality is given by

$$\widehat{\theta}_a\equiv\widehat{\theta}-\frac{\widehat{b}(\widehat{\theta})}{n}\equiv\widehat{\theta}-\frac{1}{n}\left(\frac{\left(n^{-1}\sum_i\ell^{\theta\theta}(Z_i,\widehat{\theta})\right)}{2\left(n^{-1}\sum_i\ell^{\theta}(Z_i,\widehat{\theta})\right)^2}+\frac{\left(n^{-1}\sum_i\ell(Z_i,\widehat{\theta})\ell^{\theta}(Z_i,\widehat{\theta})\right)}{\left(n^{-1}\sum_i\ell^{\theta}(Z_i,\widehat{\theta})\right)^2}\right).$$

It is also possible to construct a bias correction that computes the expectations in (10) in their integral form. This is the analytical bias-corrected estimator of Pfanzagl and Wefelmeyer (1978), which they showed is higher-order efficient:

$$\widehat{\theta}_c\equiv\widehat{\theta}-\frac{b(\widehat{\theta})}{n}=\widehat{\theta}-\frac{1}{n}\left(\frac{\int\ell^{\theta\theta}(z,\widehat{\theta})f(z,\widehat{\theta})\,dz}{2\left(\int\ell^{\theta}(z,\widehat{\theta})f(z,\widehat{\theta})\,dz\right)^2}+\frac{\int\ell(z,\widehat{\theta})\ell^{\theta}(z,\theta)f(z,\widehat{\theta})\,dz}{\left(\int\ell^{\theta}(z,\widehat{\theta})f(z,\widehat{\theta})\,dz\right)^2}\right).$$

The analytical bias-corrected estimators that impose the information equality do not have identical higher-order expressions to the bias-corrected estimators that do not rely on the information equality. Instead, they differ in their third expansion terms, as can be seen in the following proposition.

PROPOSITION 3. *Let the regularity conditions stated in Appendix B.1 hold. The analytical bias estimates satisfy*

$$\sqrt{n}\big(\widehat{b}(\widehat{\theta})-b_0\big)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{A}_i+o_p(1),$$

$$\sqrt{n}\big(b(\widehat{\theta})-b_0\big)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{C}_i+o_p(1),$$

where $\mathbb{A}_i$ and $\mathbb{C}_i$ are as defined in Supplementary Appendix I A. Hence, the analytical bias-corrected estimators $\widehat{\theta}_c$ and $\widehat{\theta}_a$ have higher-order expansions

$$\sqrt{n}(\widehat{\theta}_c-\theta_0)=A_1+\frac{1}{\sqrt{n}}\big(A_2-b(\theta_0)\big)+\frac{1}{n}(A_3-\mathbb{C})+o_p\big(n^{-1}\big),$$

$$\sqrt{n}(\widehat{\theta}_a-\theta_0)=A_1+\frac{1}{\sqrt{n}}\big(A_2-b(\theta_0)\big)+\frac{1}{n}(A_3-\mathbb{A})+o_p\big(n^{-1}\big),$$

where $\mathbb{A}=\sum_i\mathbb{A}_i/\sqrt{n}$ and $\mathbb{C}=\sum_i\mathbb{C}_i/\sqrt{n}$.

Propositions 1, 2, and 3 imply that the higher-order variances of $\widehat{\theta}_B$, $\widehat{\theta}_J$, $\widetilde{\theta}_a$ are equal to $\mathrm{Var}(A_2) + 2n^{1/2}E[A_1A_2] + 2E[A_1(A_3 - \mathbb{B})]$, while the higher-order variances of $\widehat{\theta}_a$, and $\widehat{\theta}_c$ are $\mathrm{Var}(A_2) + 2n^{1/2}E[A_1A_2] + 2E[A_1(A_3 - \mathbb{A})]$ and $\mathrm{Var}(A_2) + 2n^{1/2}E[A_1A_2] + 2E[A_1(A_3 - \mathbb{C})]$, respectively. Therefore, it is natural to conjecture that the higher-order variances of $\widehat{\theta}_B$, $\widehat{\theta}_J$, and $\widetilde{\theta}_a$ are different from that of $\widehat{\theta}_c$ or $\widehat{\theta}_a$. However, the propositions also state that each estimator uses an estimate of the higher-order bias b_0 that is asymptotically linear, and hence a corollary of Theorem 1 is that the higher-order variances of the bias-corrected estimators are all equal.

Corollary 1. *Let the regularity conditions stated in Appendix B.1 hold. Then*

$$E[\mathbb{B}A_1] = E[\mathbb{A}A_1] = E[\mathbb{C}A_1].$$

By Corollary 1, and the expression for the higher-order variance (4), we conclude that all five estimators are higher-order efficient. This result follows because (i) they have identical higher-order variance and (ii) $\widehat{\theta}_c$ was shown to be higher-order efficient by Pfanzagl and Wefelmeyer (1978).

The next proposition provides an expression for the higher-order variance of these five bias-corrected estimators.

Proposition 4. *Let the regularity conditions stated in Appendix B.1 hold, and define*

$$\begin{aligned} X_i &= \mathcal{I}^{-1}U_i, \\ Y_i &= \frac{1}{2}\mathcal{I}^{-2}\mathcal{Q}_1U_i + \mathcal{I}^{-1}V_i, \\ \Upsilon &= \left(E\left[X_i^2\right]E\left[Y_i^2\right] + E[X_1Y_i]^2\right). \end{aligned}$$

The higher-order variance of $\widehat{\theta}_B$, $\widehat{\theta}_J$, $\widetilde{\theta}_a$, $\widehat{\theta}_a$, *and* $\widehat{\theta}_c$ *is*

$$\mathrm{Var}(A_1) + \frac{\tilde{v}}{n} = \mathcal{I}^{-1} + \frac{1}{n}\Upsilon. \tag{11}$$

The above higher-order variance expression is useful since the higher-order variance formula in (4) includes two covariance terms $2E[A_1A_3] + 2n^{1/2}E[A_1A_2]$, and as such, it is not obvious that (4) is positive. On the other hand, because $\Upsilon > 0$, the expression in the proposition consists of positive terms and eliminates this concern. According to Theorem 1, (11) is also the higher-order variance of any other asymptotically linear bias correction.

3.3 *Inefficient bias correction*

Our discussion above shows that many common methods of bias correction use estimates of the bias that are asymptotically linear, and hence estimate the bias at the $n^{-1/2}$-rate. These bias-corrected estimators have equivalent higher-order variances. We now show by an example that the equivalence result does not hold in general if this rate assumption is violated. We showed in Example 1 (see the end of Section 2.3) that the split-sample jackknife provided an inconsistent estimate of the higher-order bias, so that its higher-order variance was larger than that of the leave-one-out jackknife. In the Appendix, we show that this result holds more generally, for estimators with the expansion structure used in Theorem 2. Here, we derive expressions for the higher-order variance in the MLE case.[7]

Recall that the split-sample jackknife estimator is given by

$$\widehat{\theta}_{\mathrm{SS}} \equiv 2\widehat{\theta} - (\widehat{\theta}_{(1)} + \widehat{\theta}_{(2)})/2,$$

where $\widehat{\theta}_{(1)}$ and $\widehat{\theta}_{(2)}$ denote the MLEs based on separate halves of the sample. The implicit bias estimate used by the split-sample jackknife is given by $\frac{1}{n}\widehat{b}_{\mathrm{SS}} = (\widehat{\theta}_{(1)} + \widehat{\theta}_{(2)})/2 - \widehat{\theta}$.

Proposition 5. *Let the regularity conditions stated in Appendix B.1 hold. The bias estimate used by the split-sample estimator can be written as*

$$\begin{aligned}\widehat{b}_{\mathrm{SS}} - b_0 &= \frac{1}{2}\big(X_{(1)}Y_{(1)} + X_{(2)}Y_{(2)} - 2b_0 - (X_{(1)}Y_{(2)} + X_{(2)}Y_{(1)})\big)\\ &= O_p(1),\end{aligned}$$

where, for $n = 2m$, $X_{(1)} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} X_i$ is the scaled sum over X_i in the first half of the sample, and $X_{(2)}$, $Y_{(1)}$, and $Y_{(2)}$ are defined similarly (for X_i and Y_i as defined in Proposition 4).

The higher-order variance of $\widehat{\theta}_{\mathrm{SS}}$ is

$$\mathrm{Var}(A_1) + \frac{\tilde{v}_{\mathrm{SS}}}{n} = \mathcal{I}^{-1} + \frac{2}{n}\Upsilon.$$

That is, the higher-order variance of $\widehat{\theta}_{\mathrm{SS}}$ is strictly larger than that of $\widehat{\theta}_J$. As shown in Proposition 5, the split-sample bias correction provides an inconsistent estimate of the bias term b_0, which results in a bias correction that impacts the second-order term A_2 in the expansion, and hence affects the higher-order variance of the estimator.

4. Bias correction for panel data

We now extend some of these ideas to the panel data setting by comparing the jackknife and split-sample bias correction methods for a model with individual fixed effects. We begin with a description of the fixed effects maximum likelihood estimator. Let Z_{it}, for $i = 1, \dots, n$ and $t = 1, \dots, T$, be a vector of observed data. Denote θ a $p \times 1$ parameter vector and α_i a scalar unobserved individual effect.[8] The data have density function

[7]Derivation of the expressions used below are available in Section A7 of Supplementary Appendix I.

[8]As before, we will assume that $p = 1$ for notational simplicity, but results should be expected to hold for any finite $p > 1$. The analysis in this section assumes that time effects are not present.

$f(z|\theta,\alpha)$ with respect to some measure, and so (treating the α_i as parameters to be estimated) we may estimate θ via maximum likelihood. Assuming that the Z_{it} are independent across both i and t, the MLE solves

$$\widehat{\theta}_T \equiv \arg\max_{\theta} \sum_{i=1}^{n}\sum_{t=1}^{T} \ln f\big(Z_{it}|\theta, \widehat{\alpha}_i(\theta)\big), \qquad \widehat{\alpha}_i(\theta) \equiv \arg\max_{\alpha} \sum_{t=1}^{T} \ln f(Z_{it}|\theta,\alpha).$$

In contrast to the higher-order bias discussed earlier for the cross-sectional model, the fixed-effects panel data estimator suffers from the well-known incidental parameters problem (Neyman and Scott (1948)). For fixed T, the probability limit of the estimator $\theta_T \equiv \text{plim}_{n\to\infty} \widehat{\theta}_T$ generally differs from θ_0.[9] Even when the number of time periods grow, so that $n/T \to \rho$ (which will be assumed in this section), the estimator $\widehat{\theta}$ remains asymptotically biased, that is, $\sqrt{nT}(\widehat{\theta}-\theta_0) \Rightarrow N(\sqrt{\rho}\mathbf{B}, \Omega)$ for some bias term $\mathbf{B}$. The bias is of order $O(T^{-1})$, and can be substantial if T is not sufficiently large.

It is useful to think about θ_0 and α_i as solutions to a set of moment equations given by the score functions

$$0 = \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\ln f(z_{it}|\theta_0,\alpha_i)\right], \qquad 0 = E\left[\frac{\partial}{\partial\alpha_i}\ln f(z_{it}|\theta_0,\alpha_i)\right].$$

As earlier, we can expand these first-order conditions to produce asymptotic expansions. We may also consider other quantities of interest that can be defined via some moment condition, for example, an average effect parameter μ_0, defined as the solution to

$$0 = \mu_0 - \frac{1}{n}\sum_{i=1}^{n} E\big[m(z_{it},\theta_0,\alpha_i)\big],$$

where m is some function of interest, for example, the partial derivative of a conditional expectation function. Stacking these moment conditions, we can define the common parameter to be (θ,μ), so that the results presented below will also apply to these types of parameters.

We use the following notation for this section. Let $u_{it}(\theta,\alpha) \equiv \frac{\partial}{\partial\theta}\ln f(z_{it}|\theta,\alpha)$ and $V_{it}(\theta,\alpha) \equiv \frac{\partial}{\partial\alpha_i}\ln f(z_{it}|\theta,\alpha)$ be the score functions. When evaluating functions at the true value of parameters, arguments will be dropped, for example, $u_{it} = u_{it}(\theta_0,\alpha_i)$. Further, let $U_{it}(\theta,\alpha) = u_{it}(\theta,\alpha) - \delta V_{it}(\theta,\alpha)$, for $\delta = E[u_{it}V_{it}]/E[V_{it}^2]$, be the efficient score for θ. All expectations are taken with respect to the distribution for an individual i, that is, $E[h(z_{it})] = \int h(z)f(z|\theta,\alpha_i)\,dz$. Also define $\mathcal{I}_n = \frac{1}{n}\sum_{i=1}^{n} E[U_{it}^2]$. As in the cross-sectional case, we denote partial derivatives of these functions with superscripts, for example, $\partial U_{it}/\partial\theta = U_{it}^{\theta}$.

4.1 *Higher-order comparison of jackknife estimators*

In this section, we derive the higher-order properties of both the leave-one-out jackknife and split-sample jackknife estimators for the panel data model.

[9] θ_T is given by $\text{argmax}_\theta \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E[\sum_{t=1}^{T} \ln f(z_{it}|\theta,\widehat{\alpha}_i(\theta))]$.

It has be shown previously (e.g., Hahn and Kuersteiner (2002), Hahn and Newey (2004)) that the MLE $\widehat{\theta}$ has an expansion of the form

$$\sqrt{nT}(\widehat{\theta}-\theta_0)=A_1+\frac{\sqrt{n}}{\sqrt{T}}A_2+\frac{\sqrt{n}}{T}A_3+O_p\big(T^{-1}\big), \tag{12}$$

where the expansion terms A_1, A_2, A_3 are each $O_p(1)$.[10] This expansion is similar in style to the asymptotic expansion for the cross-sectional model developed earlier, except that the asymptotic order of the leading terms in the expansion are in terms of $T^{-1/2}$ due to the presence of the individual fixed effects. Here, the asymptotic bias is given by $\sqrt{\frac{n}{T}}\mathbf{B}=\sqrt{\frac{n}{T}}E[A_2]$.

The “leave-one-out” jackknife estimator is

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

where $\widehat{\theta}_{(t)}$ is the estimator formed from the subsample that excludes time period t.[11] Dhaene and Jochmans (2015) propose the use of split-panel jackknifes that only make use of subpanels that contain consecutive time periods. The split-sample jackknife estimator is

$$\widehat{\theta}_{\mathrm{SS}}=2\widehat{\theta}-\bar{\theta}_{\mathrm{SS}},$$

where $\bar{\theta}_{\mathrm{SS}}=\frac{1}{2}(\widehat{\theta}_1+\widehat{\theta}_2)$, with $\widehat{\theta}_1$ being the estimate using observations from the first half of time periods, and $\widehat{\theta}_2$ the estimate that uses the second half of time periods. Other choices of split-sample jackknife are available; however, the results in Dhaene and Jochmans (2015) show that nonoverlapping subpanels in general have lower asymptotic variance, and that among the nonoverlapping options, splitting in two leads to the smallest inflation of higher-order bias. Hence, in this paper we focus on this half sample version, and simply refer to it as the split-sample jackknife.

Both the jackknife and split-sample corrections have no impact on the first-order term in (12) so that both estimators have the same asymptotic variance, equal to $\lim_{n\to\infty}\mathrm{Var}(A_1)=\lim_{n\to\infty}\mathcal{I}_n^{-1}$. To compute the higher-order variance for the estimators, we take the variance of the first three expansions terms, retaining terms up to $O(T^{-1})$, similar to what was done in Section 3. The following proposition establishes the higher-order variance expressions for the two estimators, from which we can conclude that the higher-order variance of the split-sample estimator is larger than that of the jackknife.

[10] In Supplementary Appendix IV, we provide an even higher-order expansion than is available in Hahn and Newey (2004).

[11] Jackknife estimators that drop k time periods, rather than just one, could also be used. However, averaging over all $\binom{T}{k}$ leave-k-out estimates would be computationally demanding, and so we do not pursue this idea here.

Proposition 6. *Let the regularity conditions stated in Appendix B.2 hold. The higher-order variances of the jackknife and split-sample bias-corrected estimators are given by*

$$\operatorname{Var}(\widehat{\theta}_J) \approx \operatorname{Var}(A_1) + \frac{1}{T-1}\tilde{v}, \qquad \operatorname{Var}(\widehat{\theta}_{\mathrm{SS}}) \approx \operatorname{Var}(A_1) + \frac{2}{T}\tilde{v},$$

where

$$\operatorname{Var}(A_1) = \mathcal{I}_n^{-1},$$

$$\tilde{v} = \mathcal{I}_n^{-2}\frac{1}{n}\sum_i \frac{\frac{1}{2}E\big[U_{it}^{\alpha\alpha}\big]^2 + 2E\big[U_{it}^{\alpha\alpha}\big]E\big[V_{it}U_{it}^{\alpha}\big] + E\big[V_{it}^2\big]E\big[\big(U_{it}^{\alpha}\big)^2\big] + E\big[V_{it}U_{it}^{\alpha}\big]^2}{E\big[V_{it}^2\big]^2}.$$

4.2 *Accuracy in estimating the bias*

In Section 3.3, it was demonstrated that while the jackknife uses a consistent estimate of the bias term, the split-sample jackknife uses an unbiased, but inconsistent estimate. A similar situation arises in the panel model. For the "leave-one-out" jackknife, the implicit bias estimate is equal to $\frac{1}{T}\widehat{b}_J = (T-1)(\frac{1}{T}\sum_{t=1}^T \widehat{\theta}_{(t)} - \widehat{\theta})$, while the split-sample jackknife uses the bias estimate $\frac{1}{T}\widehat{b}_{\mathrm{SS}} = (\bar{\theta}_{\mathrm{SS}} - \widehat{\theta})$. The following proposition establishes the accuracy of $\widehat{b}_J$ and $\widehat{b}_{\mathrm{SS}}$ as estimators for $\mathbf{B} = \lim_{n\to\infty} E[A_2]$.

Proposition 7. *Let the regularity conditions stated in Appendix B.2 hold. Let $\frac{1}{T}\widehat{b}_J = (T-1)(\frac{1}{T}\sum_{t=1}^T \widehat{\theta}_{(t)} - \widehat{\theta})$ and $\frac{1}{T}\widehat{b}_{\mathrm{SS}} = (\bar{\theta}_{\mathrm{SS}} - \widehat{\theta})$ be the jackknife and split-sample estimators for the bias term* $\mathbf{B}$. *Then*

$$\sqrt{nT}\frac{1}{T}(\widehat{b}_J - \mathbf{B}) = O_p\big(T^{-1}\big),$$

$$\sqrt{nT}\frac{1}{T}(\widehat{b}_{\mathrm{SS}} - \mathbf{B}) = O_p\big(T^{-1/2}\big).$$

Similar to the cross-sectional analysis, the jackknife estimates the bias term at a faster rate than the split-sample estimator. In contrast, in the panel setting the split-sample bias estimate is in fact consistent; it is an average over n unbiased, but inconsistent, estimates of individual-level bias terms. Nonetheless, Proposition 7 shows that the jackknife bias correction affects the third order, $O_p(T^{-1})$, part of the expansion, while the split-sample bias correction appears as a second order, $O_p(T^{-1/2})$, term. This implies that the jackknife bias estimate will only impact the higher-order variance through its covariance with the first-order term A_1, that is, through the term $\operatorname{Cov}(\sqrt{n}A_1, \sqrt{nT}(\widehat{b}_J - \mathbf{B}))$. In contrast, the split-sample bias estimate appears in the higher-order variance both through its covariance with A_1, as well as through its own variance, $\operatorname{Var}(\sqrt{n}(\widehat{b}_{\mathrm{SS}} - \mathbf{B}))$.

It should be noted that the results in Propositions 6 and 7 are derived under i.i.d. sampling over both individuals and time. In this setting, the results mirror those for the cross-sectional analysis. In settings where serial correlation exists, the leave-one-out style jackknife does not provide a valid estimate of the bias, while the split-sample estimator continues to produce a consistent estimate of the bias, albeit at a slower rate. This is an important advantage of the method in Dhaene and Jochmans (2015). Although a general result on higher-order properties under serial correlation is not yet available, it may be reasonable to speculate that the convergence rate of the estimator of the bias will continue to play an important role. This would suggest that analytical bias corrections involving parametric estimators, when available, are likely to be superior to nonparametric estimators that require estimates of long-run variances.

Example 2. A simple example may help to highlight the results. Suppose that $z_{it} \sim N(\alpha_i, \theta)$ are independent across i and t. This model was studied by Neyman and Scott (1948). The MLE is given by $\widehat{\theta} = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T (z_{it} - \bar{z}_i)^2$, and a standard calculation gives that $E[\widehat{\theta}] = \frac{T-1}{T}\theta$ so that the MLE has a bias of $\frac{1}{T}\mathbf{B} = -\frac{1}{T}\theta$. It can be shown that the jackknife bias-corrected estimator has the form

$$\tilde{\theta}_J = \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T (z_{it} - \bar{z}_i)^2$$

while the split-sample estimator is

$$\tilde{\theta}_{\text{SS}} = 2\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T (z_{it} - \bar{z}_i)^2 - \frac{1}{2}\left(\frac{1}{nM}\sum_{i=1}^n \sum_{t=1}^M (z_{it} - \bar{z}_{i,1})^2 + \frac{1}{nM}\sum_{i=1}^n \sum_{t=M+1}^T (z_{it} - \bar{z}_{i,2})^2\right),$$

where $T = 2M$, and $\bar{z}_{i,1}$ and $\bar{z}_{i,2}$ are the sample means in the first and second halves of the sample time period. In this simple model, the formula given in Proposition 6 gives $V_J = \frac{2T}{T-1}\theta^2$ and $V_{\text{SS}} = \frac{2T+4}{T}\theta^2$, which we can easily confirm are also the exact finite sample variances of the two estimators in this case. Considering the estimation of the bias itself, we can see that the jackknife bias-correction estimates the bias as $\frac{1}{T}\widehat{b}_J = -\frac{1}{T}\frac{1}{n(T-1)}\sum_{i=1}^n \sum_{t=1}^T (z_{it} - \bar{z}_i)^2$, while the split-sample correction estimates the bias using $\frac{1}{T}\widehat{b}_{\text{SS}} = -\frac{1}{2n}\sum_{i=1}^n ((\bar{z}_{i,1} - \bar{z}_i)^2 + (\bar{z}_{i,2} - \bar{z}_i)^2)$. Both are unbiased estimators of the bias term in this case, that is, $E[\widehat{b}_J] = E[\widehat{b}_{\text{SS}}] = -\theta$. It is straightforward to show $\text{Var}(\sqrt{\frac{n}{T}}(\widehat{b}_J - \mathbf{B})) = \frac{2\theta^2}{T(T-1)}$, whereas $\text{Var}(\sqrt{\frac{n}{T}}(\widehat{b}_{\text{SS}} - \mathbf{B})) = \frac{2\theta^2}{T}$, so that the variance of the split-sample bias estimate is larger by a factor T, as predicted by Proposition 7.

4.3 *Extension to time-series data*

In Proposition 4, we noted that the higher-order variance of the bias corrected MLE can be expressed in terms of Υ in (11). This characterization is useful because it is intuitively positive, which is not obvious from the definition in (4). A natural question is whether an analytical bias correction in a general time-series environment would lead to a higher-order variance of the form $E[A_1^2] + \frac{1}{n}(\Upsilon + o(n^{-1}))$, where Υ is appropriately replaced by the long-run counterpart of

$$E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T X_t\right)^2\right] E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T Y_t\right)^2\right] + \left(E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T X_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T Y_t\right)\right]\right)^2.$$

Although we are unable to answer the question in its general form, it can be shown to be valid for a strictly stationary AR(1) model with normally distributed errors. To be more precise, we go through the formal expansion of the MLE of the AR(1) model $y_t = \theta y_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$ and $y_0 \sim N(0, .\sigma^2/(1-\theta^2))$.[12] The higher-order bias based on the asymptotic expansion for the MLE yields $\lim_{T\to\infty} E[A_2] = -2\theta$, so the analytical bias-corrected estimator takes the form $\widehat{\theta} + .2\widehat{\theta}/T$. This means that the higher-order variance of the bias-corrected estimator is $\mathrm{Var}(A_1) + \frac{1}{T}\mathrm{Var}(A_2) + \frac{2}{\sqrt{T}}E[A_1A_2] + \frac{2}{T}E[A_1(A_3 + 2A_1)]$, which is not obviously positive. We show that this higher-order variance is equal to $\mathrm{Var}(A_1) + \frac{1}{T}(\mathrm{Var}(\mathcal{X})\mathrm{Var}(\mathcal{Y}) + (E[\mathcal{X}\mathcal{Y}])^2)$ for $\mathcal{X} = (1-\theta^2)\frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{y_{t-1}\varepsilon_t}{\sigma^2}$ and $\mathcal{Y} = -(1-\theta^2)\frac{1}{\sqrt{T}}\sum_{t=1}^T(\frac{y_{t-1}^2}{\sigma^2} - \frac{1}{1-\theta^2})$. Analogously to the result in Proposition 5, we show that the split-sample jackknife bias-corrected estimator for the AR(1) model has higher-order variance that is given by $\mathrm{Var}(A_1) + \frac{2}{T}(\mathrm{Var}(\mathcal{X})\mathrm{Var}(\mathcal{Y}) + (E[\mathcal{X}\mathcal{Y}])^2)$, and so the higher-order part of its variance is larger than that of the analytical bias-corrected estimator by a factor of two. See Propositions 5 and 6 in Supplementary Appendix III. These results are admittedly specific to the AR(1) model. We conjecture that they carry over to more general parametric time-series models as long as the bias has a closed-form parametric expression that can be estimated at parametic rates. On the other hand, estimators with nonparametrically estimated bias corrections that typically involve estimated long run variances may not share the same efficiency properties. We conjecture the same is true for nonparametric block bootstrap procedures and other subsampling techniques used to estimate the higher-order bias.

5. Monte Carlo analysis

To highlight the relevance of the results in a more practical setting, we conduct two Monte Carlo exercises. In the cross-sectional setting, we estimate a marginal treatment effect (MTE) model as in Heckman and Vytlacil (2005). The simulation design follows that used in Cattaneo, Jansson, and Ma (2019) and is a simplified model with no covariates. The treatment T_i is assigned according to $T_i = 1\{P(Z_i) \geq V_i\}$, where $P(Z_i)$ is a propensity score (which is a function of observed instrumental variables Z_i), and $V_i \sim \mathrm{Uniform}[0, 1]$ is an unobservable shock that is correlated with potential outcomes and generates selection. Potential outcomes under treated and control states are generated according to

$$Y_i(0) = U_{0i}, \qquad U_{0i}|Z_i, V_i \sim \mathrm{Uniform}[-1, 1],$$
$$Y_i(1) = 0.5 + U_{1i}, \qquad U_{1i}|Z_i, V_i \sim \mathrm{Uniform}[-0.5, 1.5 - 2V_i].$$

The observed outcome variable is given by $Y_i = T_iY_i(1) + (1-T_i)Y_i(0)$. We create a set of 19 potential instruments, $Z_{j,i} \sim \mathrm{Uniform}[0, 1]$ for $j = 1, \ldots, 19$, to be used in estimation, although the true propensity score only depends on the first four of these (in addition to the constant) $P(Z_i) = 0.1 + Z_{1,i} + Z_{2,i} + Z_{3,i} + Z_{4,i}$.

[12]The result is available in Supplementary Appendix III.

The MTE function is defined as $\mathrm{MTE}(v) = E[Y_1 - Y_0 | V = v]$, and measures the average treatment effect for individuals with a given level of unobserved resistance to treatment $V_i = v$. Many objects of interest can be represented as weighted averages of the MTE function, although for the purposes of this simulation we follow Cattaneo, Jansson, and Ma (2019) and focus on $\theta = \mathrm{MTE}(0.5)$. The MTE function can be identified as $\mathrm{MTE}(v) = \partial E[Y_i | P(Z_i) = p]/\partial p|_{p=v}$.

To estimate the marginal treatment effect parameter, the propensity score $P(z)$ is first estimated via regression of the treatment dummy variable T_i on $(1, Z_{1,i}, \dots, Z_{k-1,i})$, where k ranges from 5 to 20. The second step then regresses the outcome Y_i on a quadratic in the propensity score, $Y_i = \beta_1 + \beta_2 \widehat{P}(z) + \beta_3 \widehat{P}(z)^2$. The scalar parameter of interest θ is the derivative of this function at $p = 0.5$, that is, $\widehat{\theta} = \widehat{\beta}_2 + \widehat{\beta}_3$.

Table 1 reports simulation results for the bootstrap, jackknife, and split-sample bias corrections. As predicted by the theory, the standard deviation of the bootstrap and

TABLE 1. Simulation of marginal treatment effect estimates.

		Conventional			Bootstrap		
n	k	Bias	SD	RMSE	Bias	SD	RMSE
1000	5	0.441	4.922	4.941	0.105	5.146	5.146
	10	1.059	4.712	4.828	0.254	5.320	5.325
	15	1.570	4.460	4.727	0.511	5.274	5.297
	20	1.952	4.231	4.659	0.713	5.167	5.215
2000	5	0.317	4.730	4.740	0.075	4.840	4.839
	10	0.798	4.636	4.703	0.154	4.953	4.954
	15	1.226	4.505	4.668	0.275	4.972	4.979
	20	1.604	4.421	4.702	0.423	5.014	5.031
3000	5	0.155	4.756	4.757	−0.044	4.827	4.826
	10	0.582	4.666	4.701	0.025	4.879	4.877
	15	0.971	4.574	4.675	0.122	4.899	4.899
	20	1.300	4.517	4.700	0.202	4.951	4.953

		Jackknife			Split-Sample		
n	k	Bias	SD	RMSE	Bias	SD	RMSE
1000	5	0.099	5.135	5.135	0.092	5.426	5.426
	10	0.184	5.381	5.383	0.243	5.683	5.686
	15	0.326	5.447	5.456	0.459	5.611	5.628
	20	0.425	5.435	5.450	0.716	5.563	5.607
2000	5	0.074	4.830	4.830	0.087	5.012	5.012
	10	0.122	4.964	4.964	0.175	5.185	5.187
	15	0.191	5.021	5.024	0.260	5.204	5.210
	20	0.264	5.105	5.110	0.395	5.267	5.281
3000	5	−0.047	4.822	4.821	−0.068	4.924	4.923
	10	0.008	4.881	4.880	−0.024	5.006	5.005
	15	0.069	4.924	4.923	0.068	5.093	5.092
	20	0.103	4.995	4.994	0.243	5.180	5.184

Note: Results of estimators over 2000 simulations. *Conventional* denotes the standard two-step estimator described in the text; *Jackknife* denotes the leave-one-out jackknife bias-corrected estimator; *Bootstrap* denotes the bootstrap bias-corrected estimator based on $n/2$ bootstrap draws.

jackknife bias-corrected estimators are very similar across the different simulation settings. This is particularly true as the sample size grows large; for $n = 3000$, the jackknife and bootstrap bias corrections are very close in terms of standard deviation. The split-sample estimator has larger standard deviation than the other two bias corrections, as expected given its larger higher-order variance. As the sample size grows, the difference decreases; this is expected given that the estimators all have the same asymptotic variance.

As a panel data example, we estimate a probit model with strictly exogenous covariates and individual fixed effects:

$$y_{it} = 1\{\theta_0' x_{it} + \alpha_i + \varepsilon_{it} > 0\}, \quad \varepsilon_{it} \sim N(0,1).$$

The simulation is calibrated to the female labor force participation application of Fernández-Val (2009), and is the same as that used in Fernández-Val and Weidner (2018). Here, the outcome is an indicator for participation in the labor force, and the covariates include three measures of fertility, the number of children aged 0–2, 3–5, and 6–17 years, as well as the log of husband's income, and a quadratic in age. We focus on the coefficients on the three fertility variables. Below we report the results of simulations drawn from a sample of $n = 500$ individuals and $T = \{4, 8\}$ time periods.[13]

Table 2 reports the results for the biased MLE fixed effects estimator as well as three bias corrections: the analytical correction, leave-one-out jackknife, and the split-sample

Table 2. Simulation of probit model with individual fixed effects.

		MLE				Analytical			
		Bias	SD	RMSE	Rej 5%	Bias	SD	RMSE	Rej 5%
$T = 4$	Ages 0–2	−41.9	24.7	48.6	0.54	−8.2	19.0	20.7	0.06
	Ages 3–5	−42.4	47.8	63.9	0.24	−8.2	37.4	38.3	0.05
	Ages 6–17	−42.7	132.1	138.8	0.12	−1.1	102.9	102.9	0.03
$T = 8$	Ages 0–2	−16.9	11.2	20.3	0.36	−3.9	10.0	10.7	0.05
	Ages 3–5	−16.8	18.6	25.1	0.18	−3.7	16.6	17.0	0.05
	Ages 6–17	−18.6	50.7	54.0	0.09	−4.8	45.2	45.5	0.05
		Jackknife				**Split-Sample**			
		Bias	SD	RMSE	Rej 5%	Bias	SD	RMSE	Rej 5%
$T = 4$	Ages 0–2	21.2	16.7	27.0	0.14	37.7	49.5	62.2	0.53
	Ages 3–5	20.9	30.6	37.1	0.03	35.2	99.5	105.5	0.46
	Ages 6–17	22.4	88.8	91.6	0.02	31.2	303.7	305.3	0.44
$T = 8$	Ages 0–2	4.3	9.3	10.2	0.04	7.1	16.3	17.8	0.25
	Ages 3–5	4.4	15.6	16.2	0.04	7.1	30.6	31.4	0.27
	Ages 6–17	3.1	42.2	42.3	0.03	8.6	82.1	82.5	0.28

Note: Results of estimators over 1000 simulations. Bias, SD, and RMSE are percentages of the true parameter values.

[13] Fernández-Val and Weidner (2018) report results using the $n = 664$ and $T = 9$, which matches the sample size in the PSID data set, and find similar results.

jackknife. We report the bias, standard deviation, and root mean-squared error as a percentage of the true coefficient values, and the rejection rate for a test with 5% significance. The MLE fixed-effects estimator has large bias and rejection rates as large as 54%. As is evident from the theory, the size of the bias is decreasing with the number of time periods. Both the jackknife and analytical bias corrections lead to significant reductions in the bias with no cost in precision; in fact, both bias corrected estimators have smaller standard deviations than the MLE. In contrast, while the split-sample jackknife also reduces bias (although to a lesser degree than the other corrections), it has substantially larger variance and mean-squared error. It is evident that, even for $T = 8$, the impact of higher-order differences in the bias corrections remains important for the finite sample properties of the estimator.

6. Summary

We show that the choice of bias correction method does not affect the higher-order variance of any parametric estimator that is semiparametrically efficient, as long as the bias estimator is asymptotically linear, that is, asymptotically equivalent to a sample average. We give a formal expansion showing this property in a general semiparametric model. We also prove that the bootstrap, jackknife, and a version of the analytical bias estimates are asymptotically linear, when the estimator of the parameters of interest has a standard form of stochastic expansion (which is known to exist for a large class of models). The result implies that a researcher may choose a bias correction for an efficient estimator based on computational convenience, or some other criteria, without affecting its higher-order efficiency.

We have verified this result using derivations of the asymptotic expansion and higher-order variance for maximum likelihood estimation of a parametric model, using a bootstrap, jackknife, and three forms of analytical bias corrections. Furthermore, we found that the third-order stochastic expansion of the bootstrap, jackknife, and one type of analytical bias-corrected MLE are identical, and hence have an even stronger higher-order equivalence property.

These results show that the higher-order efficiency of bias-corrected efficient estimators does not depend on the form of the bias correction, as long as the estimate of the bias term is asymptotically linear (and hence $\sqrt{n}$-consistent). Thus, in practice one might use whatever bias correction method is most convenient. An important caveat is that the split-sample jackknife estimator does not estimate the bias at the $\sqrt{n}$-rate, and so is not asymptotically linear, and we show the resultant higher-order variance to be strictly larger in an i.i.d. setting, suggesting the importance of the accuracy in estimating the bias.

We generalized the result to the analysis of a panel data model with fixed effects, and established that the split-sample bias-corrected estimator has larger higher-order variance than the jackknife estimator, confirming the importance of the accuracy in estimation of the bias even in panel settings. In non-i.i.d. settings, the standard jackknife cannot be used. Comparison of the split-sample correction with alternatives, such as the analytical correction given in Hahn and Kuersteiner (2002, 2011), is a topic that we leave for future research.

Appendix A: Proof of Theorem 2

We assume that $\widehat{\theta}$ is an estimator with a stochastic expansion

$$\sqrt{n}(\widehat{\theta}-\theta_0) = A_1 + \frac{1}{\sqrt{n}}A_2 + \frac{1}{n}A_3 \tag{13}$$

$$+ \frac{1}{n^{3/2}}A_4 + \frac{1}{n^2}A_5 + o_p(n^{-2}), \tag{14}$$

where $n^{-k/2}A_k$ is a kth order V-statistic, that is,

$$\frac{1}{n^{k/2}}A_k = \left(\frac{1}{n}\sum_{i=1}^n g_{k,1}(z_i)\right)\cdots\left(\frac{1}{n}\sum_{i=1}^n g_{k,k}(z_i)\right) \tag{15}$$

and $g_{k,j}$ are functions of the data, evaluated at θ_0 with expectation zero, that is, $E[g_{j,k}(z_i)] = 0$.

The following lemma derives the first-order expansions of the jackknife, bootstrap, and (sample-average) analytical bias estimators, from which asymptotic linearity follows.

Lemma 1. *Define* $b_0 = E[g_{2,1}(z_i)g_{2,2}(z_i)]$ *as the higher-order bias of* $\widehat{\theta}$*. Then the jackknife bias estimate* $\widehat{b}_J$ *satisfies*

$$\begin{aligned}\sqrt{n}(\widehat{b}_J - b_0) &= \frac{1}{\sqrt{n}}\sum_i \bigl(g_{2,1}(z_i)g_{2,2}(z_i) - E\bigl[g_{2,1}(z_i)g_{2,2}(z_i)\bigr]\bigr)\\ &\quad + E\bigl[g_{3,1}(z_i)g_{3,2}(z_i)\bigr]\frac{1}{\sqrt{n}}\sum_i g_{3,3}(z_i) + E\bigl[g_{3,1}(z_i)g_{3,3}(z_i)\bigr]\frac{1}{\sqrt{n}}\sum_i g_{3,2}(z_i)\\ &\quad + E\bigl[g_{3,2}(z_i)g_{3,3}(z_i)\bigr]\frac{1}{\sqrt{n}}\sum_i g_{3,1}(z_i) + o_p(1)\end{aligned}$$

and the bootstrap and analytical bias corrections satisfy

$$\begin{aligned}\sqrt{n}(\widehat{b}_B - b_0) &= \frac{1}{\sqrt{n}}\sum_i \bigl(g_{2,1}(z_i)g_{2,2}(z_i) - E\bigl[g_{2,1}(z_i)g_{2,2}(z_i)\bigr]\bigr)\\ &\quad + \bigl(E\bigl[g_{2,1}(z_i)h_2(z_i)\bigr] + E\bigl[g_{2,2}(z_i)h_1(z_i)\bigr]\bigr)\frac{1}{\sqrt{n}}\sum_i g_{1,1}(z_i) + o_p(1),\end{aligned}$$

where h_1 *and* h_2 *are derivatives of* $g_{2,1}$ *and* $g_{2,2}$*, respectively. That is, they are all asymptotically linear estimators for* b_0*.*

Proof. *Jackknife bias correction*

A jackknife estimate of the higher-order bias of $\widehat{\theta}$ is given by

$$\frac{\widehat{b}_J}{n} = (n-1)\left(\frac{1}{n}\sum_{i=1}^n \widehat{\theta}_{(i)} - \widehat{\theta}\right),$$

where $\widehat{\theta}_{(i)}$ is the estimator that excludes observation i. Note that $\widehat{\theta}_{(i)}$ has an equivalent expansion to $\widehat{\theta}$, with terms $A_{k,(i)}$ that are the same as the terms in the original expansion, simply dropping observation i.

We may write

$$\begin{aligned}\frac{\widehat{b}_J}{n} &= (n-1)\Bigg(\Bigg(\frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{n-1}}A_{1,(i)} - \frac{1}{\sqrt{n}}A_1\Bigg) + \Bigg(\frac{1}{n}\sum_{i=1}^n \frac{1}{n-1}A_{2,(i)} - \frac{1}{n}A_2\Bigg) \\ &\quad + \Bigg(\frac{1}{n}\sum_{i=1}^n \frac{1}{(n-1)^{3/2}}A_{3,(i)} - \frac{1}{n^{3/2}}A_3\Bigg)\Bigg) + \Bigg(\frac{1}{n}\sum_{i=1}^n \frac{1}{(n-1)^2}A_{4,(i)} - \frac{1}{n^2}A_4\Bigg)) \\ &\quad + \Bigg(\frac{1}{n}\sum_{i=1}^n \frac{1}{(n-1)^{5/2}}A_{5,(i)} - \frac{1}{n^{5/2}}A_5\Bigg)) + o_p\big(n^{-5/2}\big) \\ &= \frac{1}{n}\Bigg(\tilde{B}_1 + \frac{1}{n^{1/2}}\tilde{B}_2 + \frac{1}{n}\tilde{B}_3 + \frac{1}{n^{3/2}}\tilde{B}_4 + \frac{1}{n^2}\tilde{B}_5\Bigg) + o_p\big(n^{-5/2}\big).\end{aligned}$$

By Lemma 21 in Supplementary Appendix II F, we have $\tilde{B}_1 = 0$, while Lemma 22 gives

$$\begin{aligned}\tilde{B}_2 &= \frac{1}{\sqrt{n}}\sum_i g_{2,1}(z_i)g_{2,2}(z_i) - \frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i} g_{2,1}(z_i)g_{2,2}(z_j) \\ &= \frac{1}{\sqrt{n}}\sum_i g_{2,1}(z_i)g_{2,2}(z_i) + o_p(1).\end{aligned}$$

For the next term, using Lemma 23 from Supplementary Appendix II F, it is straightforward to show that

$$\begin{aligned}\frac{1}{\sqrt{n}}\tilde{B}_3 &= \frac{2n-1}{n^{3/2}(n-1)}\sum_i g_{3,1}(z_i)g_{3,2}(z_i)g_{3,3}(z_i) \\ &\quad + \frac{n^2-3n+1}{n^{3/2}(n-1)^2}\sum_i\sum_{j\neq i}\big(g_{3,1}(z_i)g_{3,2}(z_i)g_{3,3}(z_j) + g_{3,1}(z_i)g_{3,2}(z_j)g_{3,3}(z_i) \\ &\quad + g_{3,1}(z_j)g_{3,2}(z_i)g_{3,3}(z_i)\big) \\ &\quad - \frac{3n-1}{n^{3/2}(n-1)^2}\sum_i\sum_{j\neq i}\sum_{k\neq\{i,j\}} g_{3,1}(z_i)g_{3,2}(z_j)g_{3,3}(z_k) \\ &= \frac{1}{n^{3/2}}\sum_i\sum_{j\neq i}\big(g_{3,1}(z_i)g_{3,2}(z_i)g_{3,3}(z_j) \\ &\quad + g_{3,1}(z_i)g_{3,2}(z_j)g_{3,3}(z_i) + g_{3,1}(z_j)g_{3,2}(z_i)g_{3,3}(z_i)\big) + o_p(1) \\ &= E\big[g_{3,1}(z_i)g_{3,2}(z_i)\big]\frac{1}{\sqrt{n}}\sum_i g_{3,3}(z_i) + E\big[g_{3,1}(z_i)g_{3,3}(z_i)\big]\frac{1}{\sqrt{n}}\sum_i g_{3,2}(z_i) \\ &\quad + E\big[g_{3,2}(z_i)g_{3,3}(z_i)\big]\frac{1}{\sqrt{n}}\sum_i g_{3,1}(z_i) + o_p(1).\end{aligned}$$

Similar results show that $\tilde{B}_4$ and $\tilde{B}_5$ are also $O_P(1)$ (see, e.g., Section A2 in Supplementary Appendix IV for results on jackknife V-statistics up to sixth order).

Using these results, we may then write

$$\begin{aligned}\sqrt{n}(\widehat{b}_J - b_0) &= \tilde{B}_2 - \sqrt{n}b_0 + \frac{1}{\sqrt{n}}\tilde{B}_3 + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_i \bigl(g_{2,1}(z_i)g_{2,2}(z_i) - E\bigl[g_{2,1}(z_i)g_{2,2}(z_i)\bigr]\bigr) \\ &\quad + E\bigl[g_{3,1}(z_i)g_{3,2}(z_i)\bigr]\frac{1}{\sqrt{n}}\sum_i g_{3,3}(z_i) + E\bigl[g_{3,1}(z_i)g_{3,3}(z_i)\bigr]\frac{1}{\sqrt{n}}\sum_i g_{3,2}(z_i) \\ &\quad + E\bigl[g_{3,2}(z_i)g_{3,3}(z_i)\bigr]\frac{1}{\sqrt{n}}\sum_i g_{3,1}(z_i) + o_p(1),\end{aligned}$$

which gives the result in the proposition.

Bootstrap bias correction

Given the expansion in (13), the bootstrap estimate $\widehat{\theta}^*$ has an equivalent expansion with reference to the empirical distribution $\widehat{F}_n$, rather than the population distribution F_0:

$$\sqrt{n}\bigl(\widehat{\theta}^* - \widehat{\theta}\bigr) = \widehat{A}_1 + \frac{1}{\sqrt{n}}\widehat{A}_2 + \frac{1}{n}\widehat{A}_3 + o_p\bigl(n^{-1}\bigr), \tag{16}$$

where

$$\frac{1}{n^{k/2}}\widehat{A}_k = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_{k,1}\bigl(z_i^*\bigr)\right)\cdots\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_{k,k}\bigl(z_i^*\bigr)\right) \tag{17}$$

with $\widehat{g}_{k,j}$ the same function as in (15), evaluated at $\widehat{\theta}$ and $\widehat{F}_n$ rather than θ_0 and F_0. Note that in this expansion terms are zero mean with respect to $\widehat{F}_n$, that is, $\sum_i \widehat{g}_{k,j}(z_i) = 0$.

We can define a bootstrap bias estimate as $\widehat{b}_B/n = E^*[\widehat{\theta}^* - \widehat{\theta}]$, where E^* is expectation over the bootstrap distribution (i.e., $\widehat{F}_n$). We may write

$$\frac{\widehat{b}_B}{n} = \frac{1}{\sqrt{n}}E^*[\widehat{A}_1] + \frac{1}{n}E^*[\widehat{A}_2] + \frac{1}{n^{3/2}}E^*[\widehat{A}_3] + o_p\bigl(n^{-3/2}\bigr).$$

From Lemma 17 in Supplementary Appendix II F, we have that

$$E^*[\widehat{A}_1] = E^*\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{g}_{1,1}\bigl(z_i^*\bigr)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{g}_{1,1}(z_i) = 0.$$

Lemma 18 gives

$$\begin{aligned}E^*[\widehat{A}_2] &= E^*\left[\frac{1}{n}\sum_i\sum_j \widehat{g}_{2,1}\bigl(z_i^*\bigr)\widehat{g}_{2,2}\bigl(z_j^*\bigr)\right] \\ &= \frac{1}{n}\sum_i \widehat{g}_{2,1}(z_i)\widehat{g}_{2,2}(z_i) + \frac{n-1}{n^2}\sum_i \widehat{g}_{2,1}(z_i)\sum_j \widehat{g}_{2,2}(z_j)\end{aligned}$$

$$= \frac{1}{n}\sum_i \widehat{g}_{2,1}(z_i)\widehat{g}_{2,2}(z_i)$$

and similarly, Lemma 19 gives

$$E^*[\widehat{A}_3] = E^*\left[\frac{1}{n^{3/2}}\sum_i\sum_j\sum_k g_{3,1}(z_i^*)g_{3,2}(z_j^*)g_{3,3}(z_k^*)\right]$$
$$= \frac{1}{n^{3/2}}\sum_i \widehat{g}_{3,1}(z_i)\widehat{g}_{3,2}(z_i)\widehat{g}_{3,3}(z_i).$$

Using these results, we can write

$$\sqrt{n}(\widehat{b}_B - b_0) = \frac{1}{\sqrt{n}}\sum_i\big(\widehat{g}_{2,1}(z_i)\widehat{g}_{2,2}(z_i) - E\big[g_{2,1}(z_i)g_{2,2}(z_i)\big]\big) + o_p(1).$$

Next, note that

$$\frac{1}{\sqrt{n}}\sum_i\big(\widehat{g}_{2,1}(z_i)\widehat{g}_{2,2}(z_i) - g_{2,1}(z_i)g_{2,2}(z_i)\big)$$
$$= \frac{1}{\sqrt{n}}\sum_i g_{2,1}(z_i)\big(\widehat{g}_{2,2}(z_i) - g_{2,2}(z_i)\big)$$
$$+ \frac{1}{\sqrt{n}}\sum_i\big(\widehat{g}_{2,1}(z_i) - g_{2,1}(z_i)\big)g_{2,2}(z_i)$$
$$+ \frac{1}{\sqrt{n}}\sum_i\big(\widehat{g}_{2,1}(z_i) - g_{2,1}(z_i)\big)\big(\widehat{g}_{2,2}(z_i) - g_{2,2}(z_i)\big).$$

Let $h_1(z_i)$ and $h_2(z_i)$ be first derivatives of $g_{2,1}(z)$ and $g_{2,2}(z_i)$ with respect to θ (evaluated at θ_0), so that first-order expansions of $\widehat{g}_{2,1}$ and $\widehat{g}_{2,2}$ are given by

$$\sqrt{n}\big(\widehat{g}_{2,1}(z_i) - g_{2,1}(z_i)\big) = h_1(z_i)\sqrt{n}(\widehat{\theta} - \theta_0) + o_p(1),$$
$$\sqrt{n}\big(\widehat{g}_{2,2}(z_i) - g_{2,2}(z_i)\big) = h_2(z_i)\sqrt{n}(\widehat{\theta} - \theta_0) + o_p(1).$$

Then we can write

$$\frac{1}{n}\sum_i g_{2,1}(z_i)\big(\widehat{g}_{2,2}(z_i) - g_{2,2}(z_i)\big) = \frac{1}{n}\sum_i g_{2,1}(z_i)h_2(z_i)\sqrt{n}(\widehat{\theta} - \theta_0) + o_p(1)$$
$$= E\big[g_{2,1}(z_i)h_2(z_i)\big]\frac{1}{\sqrt{n}}\sum_i g_{1,1}(z_i) + o_p(1)$$

and similarly for $\frac{1}{\sqrt{n}}\sum_i(\widehat{g}_{2,1}(z_i) - g_{2,1}(z_i))g_{2,2}(z_i)$ so that

$$\sqrt{n}(\widehat{b}_B - b_0) = \frac{1}{\sqrt{n}}\sum_i\big(g_{2,1}(z_i)g_{2,2}(z_i) - E\big[g_{2,1}(z_i)g_{2,2}(z_i)\big]\big)$$

$$+ \left(E\left[g_{2,1}(z_i)h_2(z_i)\right] + E\left[g_{2,2}(z_i)h_1(z_i)\right]\right)\frac{1}{\sqrt{n}}\sum_i g_{1,1}(z_i) + o_p(1)$$

giving the result of the proposition.

Analytical bias correction

Under (13) and (15), the bias term has the form $b_0 = E[g_{2,1}(z_i)g_{2,2}(z_i)]$, for some functions $g_{2,1}$ and $g_{2,2}$. Assume that we can construct consistent estimators of these functions, $\widehat{g}_{2,1}$ and $\widehat{g}_{2,2}$ by plugging in $\widehat{\theta}$ in place of θ_0 and replacing expectations with sample means. This implies $\widehat{g}_{2,1}$ and $\widehat{g}_{2,2}$ are the same functions as in the bootstrap expansion above. We can form a bias estimate using

$$\widehat{b}_a = \frac{1}{n}\sum_i \widehat{g}_{2,1}(z_i)\widehat{g}_{2,2}(z_i).$$

This then gives

$$\sqrt{n}(\widehat{b}_a - b_0) = \frac{1}{\sqrt{n}}\sum_i\left(\widehat{g}_{2,1}(z_i)\widehat{g}_{2,2}(z_i) - E\left[g_{2,1}(z_i)g_{2,2}(z_i)\right]\right)$$

and the result follows from the bootstrap result above. □

A.1 *Split-sample bias correction*

Here, we show that, under the same asymptotic expansion structure used above, the split-sample bias estimate is not asymptotically linear. The split-sample bias estimate is given by $\frac{\widehat{b}_{\text{ss}}}{n} = \frac{1}{2}(\widehat{\theta}_1 + \widehat{\theta}_2) - \widehat{\theta}$. Again, we can construct an expansion for this bias estimate from the expansion of $\widehat{\theta}$:

$$\frac{\widehat{b}_{\text{ss}}}{n} = \frac{1}{\sqrt{n}}\tilde{B}_1 + \frac{1}{n}\tilde{B}_2 + o_p\left(n^{-1}\right).$$

Let $m = n/2$. We have

$$\begin{aligned}
\tilde{B}_1 &= \frac{1}{2}(\widehat{A}_{1,1} + \widehat{A}_{1,2}) - \widehat{A}_1 = 0 \\
\tilde{B}_2 &= \frac{1}{2}(\widehat{A}_{2,1} + \widehat{A}_{2,2}) - \widehat{A}_2 \\
&= \frac{1}{2}\left(\frac{4}{n}\sum_{i=1}^{m}\sum_{j=1}^{m} g_{2,1}(z_i)g_{2,2}(z_j) + \frac{4}{n}\sum_{i=m+1}^{n}\sum_{j=m+1}^{n} g_{2,1}(z_i)g_{2,2}(z_j)\right) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{2,1}(z_i)g_{2,2}(z_j) \\
&= \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{m} g_{2,1}(z_i)g_{2,2}(z_j) + \frac{1}{n}\sum_{i=m+1}^{n}\sum_{j=m+1}^{n} g_{2,1}(z_i)g_{2,2}(z_j)
\end{aligned}$$

$$- \frac{1}{n}\sum_{i=1}^{m}\sum_{j=m+1}^{n} g_{2,1}(z_i)g_{2,2}(z_j) - \frac{1}{n}\sum_{i=m+1}^{n}\sum_{j=1}^{m} g_{2,1}(z_i)g_{2,2}(z_j).$$

This then gives

$$\begin{aligned}\widehat{b}_{ss} - b_0 = {} & \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1, j\neq i}^{m} g_{2,1}(z_i)g_{2,2}(z_j) + \frac{1}{n}\sum_{i=m+1}^{n}\sum_{j=m+1, j\neq i}^{n} g_{2,1}(z_i)g_{2,2}(z_j) \\ & - \frac{1}{n}\sum_{i=1}^{m}\sum_{j=m+1}^{n} g_{2,1}(z_i)g_{2,2}(z_j) - \frac{1}{n}\sum_{i=m+1}^{n}\sum_{j=1}^{m} g_{2,1}(z_i)g_{2,2}(z_j) \\ & + o_p(1).\end{aligned}$$

The terms on the RHS are each $O_p(1)$ so that the split-sample bias estimator is inconsistent, and has a U-statistic structure to first order, and hence cannot be asymptotically linear.

Appendix B: Regularity conditions

B.1 *Conditions for cross-sectional models*

Condition 1 (i). *The function* $\log f(z, \theta)$ *is* 7 *times continuously differentiable on* Θ *for each* z; (*ii*) *The parameter space* $\Theta \subset \mathbb{R}$ *is a compact set,* $\theta_0 \in \text{int}(\Theta)$; (*iii*) *There exists a function* $M(z)$ *such that for all* $\theta \in \Theta$,

$$\left|\frac{\partial^m \log f(z, \theta)}{\partial \theta^m}\right| \leq M(z) \quad 0 \leq m \leq 7$$

and $E[M(Z_i)^Q] < \infty$ *for some* $Q > 16$; (*iv*) *If* $\theta \neq \theta_0$, *then* $f(Z_i, \theta) \neq f(Z_i, \theta_0)$.

Condition 2. *For each* $\theta \in \Theta$ *and for* $m \leq 7$, $\partial^m \log f(z, \theta)/\partial \theta^m$ *is a* P*-measurable function of* z.

Condition 3. *Let* $\mathfrak{F}$ *be the class of functions* $\partial^m \log f(z, \theta)/\partial \theta^m$ *indexed by* $\theta \in \Theta$ *for* $m = 1, \dots, 7$ *with envelope* $M(z)$. *Then*

$$\int_0^1 \sup_{\mathcal{Q}\in\mathfrak{P}} \sqrt{\log N\left(\varepsilon\left(\int M^2\, d\mathcal{Q}\right)^{1/2}, \mathfrak{F}, L_2(\mathcal{Q})\right)}\, d\varepsilon < \infty, \tag{18}$$

where $\mathfrak{P}$ *is the class of probability measures on* $\mathbb{R}$ *that concentrate on a finite set and* N *is the cover number defined in* van der Vaart and Wellner (1996, *p.* 90).

Condition 1 is a standard condition guaranteeing identification of the model and imposing sufficient smoothness conditions as well as existence of higher moments to allow for a higher-order stochastic expansion of the estimator. Condition 2, together with separability of the parameter space, guarantees measurability of suprema of our empirical processes. As is well known from the probability literature, measurability conditions could be relaxed somewhat at the expense of more refined convergence arguments. We are abstracting from such refinements for the purpose of this paper.

B.2 *Conditions for panel models*

The assumptions for the panel results follow those used in Hahn and Newey (2004).

Condition 4. *$n, T \to \infty$, with $n/T \to \rho$ for $0 < \rho < \infty$.*

Condition 5 (i). *The data z_{it} are independent over i and t and identically distributed over t according to the density $f(z|\theta, \alpha)$; (ii) the log density $\ln f(z|\theta, \alpha)$ is continuous in both θ and α; (iii) there exists a function $M(z_{it})$ such that $|\ln f(z_{it}|\theta, \alpha_i)| \leq M(z_{it})$, $|.\partial \ln f(z_{it}|\theta, \alpha_i)/\partial(\theta, \alpha_i)| \leq M(z_{it})$ and $\sup_i E[M(z_{it})^{33}] < \infty$.*

Condition 6. *For each $\eta > 0$, $\inf_i [G_i(\theta_0, \alpha_i) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_i)|>\eta\}} G_i(\theta, \alpha)] > 0$ where $G_i(\theta, \alpha) \equiv E[\ln f(z_{it}|\theta, \alpha)]$.*

Condition 7 (i). *There exists some $M(z_{it})$ such that $|.\partial^{m_1+m_2} \ln f(z_{it}|\theta, \alpha)/\partial\theta^{m_1}\partial\alpha^{m_2}| \leq M(z_{it})$ for $0 \leq m_1 + m_2 \leq 7$, and $\sup_i E[M(z_{it})^Q] < \infty$ for some $Q > 64$; (ii) $\lim_{n\to\infty} \mathcal{I}_n > 0$, where $\mathcal{I}_n \equiv \frac{1}{n}\sum_i E[U_{it}^2]$; (iii) $\min_i E[V_{it}^2] > 0$.*

References

Akahira, Masafumi (1983), "Asymptotic deficiency of the jackknife estimator." *Australian Journal of Statistics*, 25, 123–129. [785, 786]

Akahira, Masafumi (1989), "Behaviour of jackknife estimators in terms of asymptotic deficiency under true and assumed models." *Journal of the Japan Statistical Society*, 19, 179–196. [785]

Akahira, Masafumi and Kei Takeuchi (1981), *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*. Lecture Notes in Statistics, Vol. 7. Springer Verlag. [785]

Alexander, Blair and Robert Breunig (2016), "A Monte Carlo study of bias corrections for panel probit models." *Journal of Statistical Computation and Simulation*, 86, 74–90. [785]

Amari, Shun-Ichi (1982), "Differential geometry of curved exponential families—curvatures and information loss." *Annals of Statistics*, 10, 357–385. [785]

Andersen, Erling B. (1970), "Asymptotic properties of conditional maximum likelihood estimators." *Journal of the Royal Statistical Society, Series B*, 32, 283–301. [786]

Arellano, Manuel and Jinyong Hahn (2010), "Understanding bias in nonlinear panel models: Some recent developments." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Vol. 3, 381–409. [786]

Arellano, Manuel and Bo Honoré (2001), "Panel data models: Some recent developments." In *Handbook of Econometrics*, Vol. 5 (James J. Heckman and Edward Leamer, eds.), North-Holland. [786]

Cattaneo, Matias, Michael Jansson, and Xinwei Ma (2019), “Two-step estimation and inference with possibly many included covariates.” *Review of Economic Studies*, 86, 1095–1122. [784, 803, 804]

Chamberlain, Gary (1984), “Panel data.” In *Handbook of Econometrics*, Vol. 2 (Zvi Griliches and Michael D. Intriligator, eds.), North-Holland. [786]

Chernozhukov, Victor, Whitney K. Newey, Jinyong Hahn, and Ivan Fernández-Val (2013), “Average and quantile effects in nonseparable panel models.” *Econometrica*, 81, 535–580. [786]

Czarnowske, Daniel and Amrei Stammann (2019), “Fixed effects binary choice models: Estimation and inference with long panels.” arXiv preprint arXiv:1904.04217. [785]

Dhaene, Geert and Koen Jochmans (2015), “Split-panel jackknife estimation of fixed-effect models.” *Review of Economic Studies*, 82, 991–1030. [785, 786, 800, 802]

Fernández-Val, Ivan (2009), “Fixed effects estimation of structural parameters and marginal effects in panel probit models.” *Journal of Econometrics*, 150 (1), 71–85. [805]

Fernández-Val, Ivan and Martin Weidner (2018), “Fixed effects estimation of large-t panel data models.” *Annual Review of Economics*, 10 (1), 109–138. [785, 805]

Ghosh, Jayanta K. (1994), *Higher Order Asymptotics*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 4. Institute of Mathematical Statistics. [785]

Ghosh, Jayanta K., Bimal K. Sinha, and Harry S. Wieand (1980), “Second order efficiency of the MLE with respect to any bowl shaped loss function.” *Annals of Statistics*, 8, 506–521. [784]

Hahn, Jinyong, David W. Hughes, Guido Kuersteiner, and Whitney K. Newey (2024), “Efficient bias correction for cross-section and panel data.” arXiv preprint arXiv:2207.09943. [787]

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, 1639–1657. [786, 800, 806]

Hahn, Jinyong and Guido Kuersteiner (2011), “Bias reduction for dynamic nonlinear panel models with fixed effects.” *Econometric Theory*, 27, 1152–1191. [806]

Hahn, Jinyong and Whitney K. Newey (2004), “Jackknife and analytical bias reduction for nonlinear panel models.” *Econometrica*, 72, 1295–1319. [785, 786, 800, 813]

Hall, Peter (1992), *The Bootstrap and Edgeworth Expansion*. Springer Verlag, New York. [786]

Hausman, Jerry A. (1978), “Specification tests in econometrics.” *Econometrica*, 46, 1251–1271. [790]

Heckman, James J. and Edward J. Vytlacil (2005), “Structural equations, treatment effects and econometric policy evaluation.” *Econometrica*, 73, 669–738. [803]

Horowitz, Joel L. (1998), "Bootstrap methods for covariance structures." *The Journal of Human Resources*, 33, 39–61. [786]

Nagar, Anirudh L. (1959), "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations." *Econometrica*, 27, 575–595. [784]

Newey, Whitney K. (1990), "Semiparametric efficiency bounds." *Journal of Applied Econometrics*, 5, 99–135. [787, 789, 790]

Newey, Whitney K. (1994), "The asymptotic variance of semiparametric estimators." *Econometrica*, 62, 1349–1382. [792]

Newey, Whitney K. and Richard J. Smith (2004), "Higher order properties of GMM and generalized empirical likelihood estimators." *Econometrica*, 72, 219–255. [784, 785, 792, 793]

Neyman, Jerzy and Elizabeth L. Scott (1948), "Consistent estimates based on partially consistent observations." *Econometrica*, 16, 1–32. [799, 802]

Parr, William C. (1983), "A note on the jackknife, the bootstrap and the delta method estimators of bias and variance." *Biometrika*, 70, 719–722. [786]

Pfanzagl, Johann and Wolfgang Wefelmeyer (1978), "A third-order optimum property of the maximum likelihood estimator." *Journal of Multivariate Analysis*, 8, 1–29. [784, 785, 793, 796, 797]

Quenouille, Maurice H. (1949), "Problems in plane sampling." *Annals of Mathematical Statistics*, 20, 355–375. [786]

Quenouille, Maurice H. (1956), "Notes on bias in estimation." *Biometrika*, 43, 353–360. [786]

Rasch, Georg (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*. Denmark Paedagogiske Institute, Copenhagen. [786]

Rasch, Georg (1961), *On General Laws and the Meaning of Measurement in Psychology*. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4. University of California Press, Berkeley. [786]

Rilstone, Paul, Virendra K. Srivastava, and Aaman Ullah (1996), "The second-order bias and mean squared error of nonlinear estimators." *Journal of Econometrics*, 75, 369–395. [785, 787]

Rothenberg, Thomas J. (1984), "Approximating the distributions of econometric estimators and test statistics." In *Handbook of Econometrics*, Vol. 2 (Zvi Griliches and Michael D. Intriligator, eds.), North-Holland. [784, 785, 787]

Shao, Jun (1988a), "On resampling methods for variance and bias estimation in linear models." *Annals of Statistics*, 16, 986–1008. [786]

Shao, Jun (1988b), "Bootstrap variance and bias estimation in linear models." *The Canadian Journal of Statistics*, 16, 371–382. [786]

Taniguchi, Masanobu and Yoshihide Kakizawa (2000), *Asymptotic Theory of Statistical Inference for Time Series*. Springer New York. [785]

Tukey, John W. (1958), "Bias and confidence in not-quite large samples." *Annals of Mathematical Statistics*, 29, 614. [786]

van der Vaart, Aad W. (1991), "On differentiable functionals." *Annals of Statistics*, 19, 178–204. [792]

van der Vaart, Aad W. and Jon A. Wellner (1996), *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer New York. [812]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 7 March, 2023; final version accepted 18 April, 2024; available online 22 April, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.10957107. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.