

Bootstrap inference for fixed-effect models

Ayden Higgins*

University of Oxford

Koen Jochmans†

Toulouse School of Economics

This version: April 13, 2023

Abstract

The maximum-likelihood estimator of nonlinear panel data models with fixed effects is asymptotically biased under rectangular-array asymptotics. The literature has devoted substantial effort to devising methods that correct for this bias as a means to salvage standard inferential procedures. The chief purpose of this paper is to show that the (recursive, parametric) bootstrap replicates the asymptotic distribution of the (uncorrected) maximum-likelihood estimator and of the likelihood-ratio statistic. This justifies the use of confidence sets and decision rules for hypothesis testing constructed via conventional bootstrap methods. No modification for the presence of bias needs to be made.

JEL Classification: C23

Keywords: bootstrap, fixed effects, incidental parameter problem, inference, panel data

Introduction

The maximum-likelihood estimator of models for panel data is well known to perform poorly when fixed effects are included. The estimator is generally inconsistent under asymptotics where the number of individuals, n , grows large while the number of time periods, m , is

*Address: Department of Economics, University of Oxford, 10 Manor Road, Oxford OX1 3UQ, United Kingdom. E-mail: ayden.higgins@economics.ox.ac.uk.

†Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.

The editor, referees, Russell Davidson, Sílvia Gonçalves, Bo Honoré, Blaise Melly, and Michael Wolf provided much-appreciated comments and suggestions and Jin Hahn kindly shared supplementary material from Hahn and Newey (2004) with us.

Financial support from the European Research Council, grant ERC-2016-STG-715787, and from the French Government and the French National Research Agency under the Investissements d' Avenir program, grant ANR-17-EURE-0010 is gratefully acknowledged.

held fixed (Neyman and Scott 1948). In fact, many parameters of interest are simply not (point) identified in such a setting (see, e.g., Honoré and Tamer 2006). Maximum likelihood is consistent under so-called rectangular-array asymptotics, where n and m grow large at the same rate, but it is asymptotically biased, in general. This implies that confidence sets based on a naive normal approximation to the distribution of the maximum-likelihood estimator have incorrect coverage, even in large samples. As a consequence, inference based on it has been largely abandoned (Lancaster 2002).

Over the last two decades substantial effort has been devoted to devising procedures that remove the asymptotic bias, thereby recentering the limit distribution around zero and restoring the validity of conventional inference procedures based on it; see Arellano and Hahn (2007) for an overview. Theoretical guidelines on which bias-correction method to use and on how to select their respective tuning parameters are largely absent. This is inconvenient because, even though all the proposals lead to estimators with the same (first-order) asymptotic properties, they vary greatly in their ease of implementation and in how effective they are at salvaging inference in finite samples.

The current paper shows that, under rectangular-array asymptotics, the parametric bootstrap consistently estimates the distribution of the (uncorrected) maximum-likelihood estimator, including its asymptotic bias. This implies that confidence sets and hypothesis tests constructed using either the basic bootstrap or its studentized version have correct coverage and size, respectively, in large samples.¹ Thus, bias correction is not needed. The same conclusion is true for averages over the fixed effects, such as average marginal effects (Chamberlain 1984), and for the size of the likelihood-ratio test and the score test. Through several examples we find that simple bootstrapping outperforms inference based on bias correction. Iterating the bootstrap (as proposed by Beran 1988) can yield further improvement.

¹We adopt the terminology of Davison and Hinkley (1997, p. 194). The basic bootstrap is also referred to as the centered percentile bootstrap or the reverse-percentile bootstrap. It is different from Efron's (1982, p. 87) percentile bootstrap (Davison and Hinkley 1997, p. 203), which does not produce confidence sets with correct coverage in our setting.

In Section 1 we present the setting and state our main objectives. In Section 2 we describe our bootstrap procedure and give examples of its use. In Section 3 we investigate the performance of the bootstrap in three examples and provide an empirical illustration. In Section 4 we collect all the formal results that underlie our claims about the validity of the bootstrap in our setting. Concluding remarks end the paper. A supplementary appendix contains all proofs.

1 Maximum-likelihood estimation

Suppose that we have data on n independent stratified observations $\{y_i, y_{i-}, x_i\}$, with $y_i := (y_{i1}, \dots, y_{im})$, $y_{i-} = (y_{i(1-p)}, \dots, y_{i0})$, and $x_i := (x_{i1}, \dots, x_{im})$. We consider models where the conditional density of y_i given y_{i-} and x_i (relative to some dominating measure) is given by

$$\prod_{t=1}^m f(y_{it} | y_{it-1}, \dots, y_{it-p}, x_{it}; \varphi_0, \eta_{i0}),$$

and f is known up to the finite-dimensional parameters φ_0 and η_{i0} . This framework covers autoregressive processes (of order p), for which y_{i-} serves as the initial condition, and allows for exogenous covariates, x_i . In what follows we will treat both the initial condition and the covariates as fixed.

It is convenient to introduce the shorthand

$$\ell(\varphi, \eta_i | z_{it}) := \log f(y_{it} | y_{it-1}, \dots, y_{it-p}, x_{it}; \varphi, \eta_i),$$

where $z_{it} := (y_{it}, y_{it-1}, \dots, y_{it-p}, x_{it})$. The maximum-likelihood estimator is

$$(\hat{\varphi}, \hat{\eta}_1, \dots, \hat{\eta}_n) := \arg \max_{\varphi, \eta_1, \dots, \eta_n} \sum_{i=1}^n \sum_{t=1}^m \ell(\varphi, \eta_i | z_{it}).$$

In sufficiently-regular models we have, as $n, m \rightarrow \infty$ with $n/m \rightarrow \gamma^2$ for some $0 < \gamma < \infty$, that

$$\sqrt{nm}(\hat{\varphi} - \varphi_0) \xrightarrow{L} N(\gamma\beta, \Sigma), \tag{1.1}$$

where β is a non-random (asymptotic) bias term and the variance is $\Sigma := (\lim_{n,m \rightarrow \infty} \Omega_{nm})^{-1}$ for

$$\Omega_{nm} := -\frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m \mathbb{E} \left(\frac{\partial^2 \ell(\varphi_0, \eta_{i0} | z_{it})}{\partial \varphi \partial \varphi'} - \rho_{i,m} \frac{\partial^2 \ell(\varphi_0, \eta_{i0} | z_{it})}{\partial \eta_i \partial \varphi'} \right), \quad (1.2)$$

with

$$\rho_{i,m} := \left(\frac{1}{m} \sum_{t=1}^m \mathbb{E} \left(\frac{\partial^2 \ell(\varphi_0, \eta_{i0} | z_{it})}{\partial \varphi \partial \eta_i'} \right) \right) \left(\frac{1}{m} \sum_{t=1}^m \mathbb{E} \left(\frac{\partial^2 \ell(\varphi_0, \eta_{i0} | z_{it})}{\partial \eta_i \partial \eta_i'} \right) \right)^{-1}.$$

See [Hahn and Newey \(2004\)](#) and [Hahn and Kuersteiner \(2011\)](#) for early derivations of this result in static and dynamic models, respectively.

The presence of asymptotic bias in (1.1) implies that confidence regions and hypothesis tests based on the limit distribution of the maximum-likelihood estimator have to account for it in order to have correct coverage and size unless n/m is close to zero, which is not usually the case. A standard approach to do so is to correct $\hat{\varphi}$ for its (first-order) bias. This amounts to constructing the bias-corrected estimator $\hat{\varphi} - \hat{\beta}/m$, where $\hat{\beta}$ is an estimator of β . Such an approach recenters the estimator's limit distribution around zero, restoring the validity of inference procedures based on the usual normal approximation.

We may also be interested in parameters of the form

$$\Delta := \lim_{n,m \rightarrow \infty} \frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m \mathbb{E}(\mu(z_{it}, \varphi_0, \eta_{i0})),$$

for a chosen function μ . Average marginal effects (as discussed in [Chamberlain 1984](#)) or moments of the fixed effects are typical examples. The maximum-likelihood estimator of Δ is $\hat{\Delta} := 1/nm \sum_{i=1}^n \sum_{t=1}^m \mu(z_{it}, \hat{\varphi}, \hat{\eta}_i)$ which, similar to $\hat{\varphi}$, also suffers from asymptotic bias. This would be true even if $\hat{\varphi}$ would be replaced by a bias-corrected estimator (or, indeed, by φ_0). The form of the asymptotic bias and variance is complicated and not given here. Expressions for them (and estimators of them) can be found in [Hahn and Newey \(2004\)](#) and [Dhaene and Jochmans \(2015\)](#).

Finally, we may wish to test a null hypothesis of the form $\phi(\varphi_0) = 0$ by means of either a likelihood-ratio or a Lagrange-multiplier test. We focus on the former approach here. Let

$$\hat{\eta}_i(\varphi) := \arg \max_{\eta_i} \sum_{t=1}^m \ell(\varphi, \eta_i | z_{it}).$$

Note that $\hat{\varphi} = \arg \max_{\varphi} \sum_{i=1}^n \sum_{t=1}^m \ell(\varphi, \hat{\eta}_i(\varphi) | z_{it})$. The maximum-likelihood estimator of φ_0 under the null is

$$\check{\varphi} := \arg \max_{\varphi: \phi(\varphi)=0} \sum_{i=1}^n \sum_{t=1}^m \ell(\varphi, \hat{\eta}_i(\varphi) | z_{it}),$$

and the likelihood-ratio statistic equals

$$\hat{w} := 2 \sum_{i=1}^n \sum_{t=1}^m (\ell(\hat{\varphi}, \hat{\eta}_i(\hat{\varphi}) | z_{it}) - \ell(\check{\varphi}, \hat{\eta}_i(\check{\varphi}) | z_{it})).$$

The fixed effects introduce bias in the profile likelihood, which implies that (under the null) \hat{w} converges to a non-central χ^2 -distribution. Hence, conventional decision rules for hypothesis testing that are based on comparing \hat{w} to critical values from a χ^2 -distribution do not lead to size-correct inference.

2 Bootstrap inference

The (parametric) bootstrap we consider imposes the data generating process implied by the maximum-likelihood estimator. A bootstrap observation $y_i^* := (y_{i1}^*, \dots, y_{im}^*)$ can be generated recursively by drawing y_{it}^* from the fitted transition density obtained from the original data, i.e.,

$$f(y_{it}^* | y_{it-1}^*, \dots, y_{it-p}^*, x_{it}; \hat{\varphi}, \hat{\eta}_i).$$

The initial condition, like the covariates, is held fixed, i.e., $y_{i-}^* = y_{i-}$. The associated maximum-likelihood estimator is

$$(\hat{\varphi}^*, \hat{\eta}_1^*, \dots, \hat{\eta}_n^*) := \arg \max_{\varphi, \eta_1, \dots, \eta_n} \sum_{i=1}^n \sum_{t=1}^m \ell(\varphi, \eta_i | z_{it}^*),$$

with $z_{it}^* := (y_{it}^*, y_{it-1}^*, \dots, y_{it-p}^*, x_{it})$. We now illustrate how this bootstrap procedure can be used in the construction of confidence intervals, bias-corrected estimators, and hypothesis tests.

Confidence intervals The main observation of this paper is that, in regular situations,

$$\sqrt{nm}(\hat{\varphi}^* - \hat{\varphi}) \xrightarrow{L^*} N(\gamma\beta, \Sigma), \tag{2.3}$$

as $n, m \rightarrow \infty$ with $n/m \rightarrow \gamma^2$. Throughout, we use $\xrightarrow{L^*}$ to denote weak convergence of the bootstrap measure. Equations (1.1) and (2.3) reveal that the bootstrap distribution is consistent for the distribution of the maximum-likelihood estimator. Importantly, the bootstrap mimics the asymptotic bias.

Equation (2.3) implies that asymptotically-valid confidence intervals can be constructed by means of the basic bootstrap without correcting the maximum-likelihood estimator (or, indeed, its bootstrap counterpart) for its bias. As an example, let

$$F^*(a) := \mathbb{P}^*(c'(\hat{\varphi}^* - \hat{\varphi}) \leq a),$$

for a chosen vector of conformable dimension c . The notation \mathbb{P}^* refers to a probability computed with respect to the bootstrap measure, i.e, conditional on the original sample. Let

$$Q^*(\alpha) := \inf \{q : \alpha \leq F^*(q)\}$$

be the implied quantile function. Then

$$\{c'\varphi : c'\hat{\varphi} - Q^*(\alpha) \leq c'\varphi\}, \quad \{c'\varphi : c'\hat{\varphi} - Q^*(1+\alpha/2) \leq c'\varphi \leq c'\hat{\varphi} - Q^*(1-\alpha/2)\}$$

are, respectively, an upper one-sided confidence interval and a two-sided (equal-tailed) confidence interval for the linear combination $c'\varphi_0$ with confidence level equal to α (in large samples).

The conditions under which we establish (1.1) and (2.3) equally imply the consistency of the plug-in estimator $\hat{\Sigma}$ and of its bootstrap counterpart $\hat{\Sigma}^*$ for the inverse Fisher information Σ . This, then, equally validates the construction of confidence intervals by means of the studentized bootstrap. Again for inference on $c'\varphi_0$ we would proceed in the same way as with the basic bootstrap, only now using the quantiles of the distribution of

$$(c'\hat{\Sigma}^*c)^{-1/2}c'(\hat{\varphi}^* - \hat{\varphi}),$$

scaled up by $(c'\hat{\Sigma}c)^{1/2}$, as critical values.

Conventional bootstrap theory advocates the use of the studentized bootstrap over the basic bootstrap when the studentized quantity has a (limit) distribution that is pivotal.

The presence of bias, however, renders the relevant limit distribution non-pivotal even after studentization. As an alternative we can use the double bootstrap (as in [Beran 1988](#)). To describe it, observe that, given $\hat{\varphi}^*$ and $\hat{\eta}_i^*$, we can generate $y_i^{**} := (y_{i1}^{**}, \dots, y_{im}^{**})$ using the density $f(y_{it}^{**} | y_{it-1}^{**}, \dots, y_{it-p}^{**}, x_{it}; \hat{\varphi}^*, \hat{\eta}_i^*)$ for all strata, and subsequently apply maximum likelihood to obtain the estimators $\hat{\varphi}^{**}$ and $\hat{\eta}_i^{**}$ of $\hat{\varphi}^*$ and $\hat{\eta}_i^*$. Consider the quantile function

$$Q^{**}(\alpha) := \inf \{q : \alpha \leq F^{**}(q)\}$$

associated with $F^{**}(a) := \mathbb{P}^{**}(c'(\hat{\varphi}^{**} - \hat{\varphi}^*) \leq a)$ where, now, the notation \mathbb{P}^{**} indicates probabilities taken conditional on both the original sample and the (first layer) bootstrap sample. Suppose we again wish to construct an upper one-sided confidence interval for $c'\varphi_0$ with confidence level α . We can mimic this process via the double bootstrap. For a given $a \in (0, 1)$,

$$\hat{\alpha}^*(a) := \mathbb{P}^*(c'\hat{\varphi} \in \{c'\varphi : c'\hat{\varphi}^* - Q^{**}(a) \leq c'\varphi\})$$

is the (actual) coverage probability of an upper one-sided confidence interval for $c'\hat{\varphi}$ with (theoretical) level a using the bootstrap. Let α^* be such that $\hat{\alpha}^*(\alpha^*) = \alpha$. Then the double bootstrap constructs its one-sided confidence interval with (theoretical) level α for $c'\varphi_0$ as

$$\{c'\varphi : c'\hat{\varphi} - Q^*(\alpha^*) \leq c'\varphi\}.$$

Two-sided confidence intervals and studentized versions can be constructed in a similar manner.

Confidence intervals for Δ are obtained in the same way. Given a bootstrap sample and the associated maximum-likelihood estimator, we construct the implied plug-in estimator

$$\hat{\Delta}^* := \frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m \mu(z_{it}^*, \hat{\varphi}^*, \hat{\eta}_i^*).$$

The bootstrap distribution of $\sqrt{nm}(\hat{\Delta}^* - \hat{\Delta})$ mimics the distribution of $\sqrt{nm}(\hat{\Delta} - \Delta)$ in large samples. The construction of confidence intervals for Δ is then completely analogous to before.

Point estimation It may be of interest to report a bias-corrected point estimator of $c'\varphi_0$, say (the developments for average effects are similar to what follows). Equation (2.3) implies that the median of the bootstrap distribution F^* is a valid estimator of $c'\beta/m$. Hence,

$$c'\hat{\varphi} - Q^*(1/2)$$

is a bias-corrected estimator. [Kim and Sun \(2016\)](#) proposed estimating the bias by a winsorized mean of F^* . The winsorization involves the choice of a cut-off parameter and is needed, in general, because the mean of F^* need not exist. Using the median is a simple alternative that enjoys some robustness.

As correcting for bias leaves the estimator's (first-order) variance unchanged, $c'\hat{\Sigma}c/nm$, is a valid estimator of the variance of $c'\hat{\varphi} - Q^*(1/2)$. An alternative estimator would be the variance of the bootstrap distribution F^* (subjected to suitable winsorization). Our theoretical results below, like most in the literature, concern distributional approximations. They do not imply consistency of the bootstrap variance (see, e.g., [Hahn and Liao 2021](#)). A separate proof is needed that includes, among other things, conditions on the winsorization.

Hypothesis testing Because confidence intervals can be obtained by inverting a test, the validity of confidence intervals based on the studentized bootstrap justifies the use of bootstrap critical values for conventional t -tests. Furthermore, bootstrap p -values for such tests will be asymptotically uniformly distributed. For example, the decision rule that rejects the null that $c'\varphi_0 \leq c'\varphi$ in favor of the alternative hypothesis that $c'\varphi_0 > c'\varphi$ when

$$(c'\hat{\Sigma}c)^{-1/2} c'(\hat{\varphi} - \varphi)$$

exceeds the $(1 - \alpha)$ quantile of the distribution of $(c'\hat{\Sigma}^*c)^{-1/2} c'(\hat{\varphi}^* - \hat{\varphi})$ gives a test of size α in large samples.

We can equally bootstrap the likelihood-ratio statistic. To describe how, consider again the null hypothesis that $\phi(\varphi_0) = 0$ for a chosen function ϕ . For data generated according to our parametric bootstrap the associated constrained maximum-likelihood estimator is

equal to

$$\check{\varphi}^* := \arg \max_{\varphi: \phi(\varphi) = \phi(\hat{\varphi})} \sum_{i=1}^n \sum_{t=1}^m \ell(\varphi, \hat{\eta}_i^*(\varphi) | z_{it}^*), \quad \hat{\eta}_i^*(\varphi) := \arg \max_{\eta_i} \sum_{t=1}^m \ell(\varphi, \eta_i | z_{it}^*).$$

The corresponding likelihood-ratio statistic is

$$\hat{w}^* := 2 \sum_{i=1}^n \sum_{t=1}^m (\ell(\hat{\varphi}^*, \hat{\eta}_i^*(\hat{\varphi}^*) | z_{it}^*) - \ell(\check{\varphi}^*, \hat{\eta}_i^*(\check{\varphi}^*) | z_{it}^*)).$$

Redefine $F^*(a) := \mathbb{P}^*(\hat{w}^* \leq a)$. Then Q^* becomes the bootstrap quantile function of the likelihood-ratio statistic. The decision rule to reject the null if

$$\hat{w} > Q^*(1 - \alpha)$$

yields a test with size α in large samples. In the same way, the use of $p^* := 1 - F^*(\hat{w})$ as p -value is asymptotically justified.

The double bootstrap can equally be used for testing purposes. If we let \hat{w}^{**} denote the likelihood-ratio statistic computed on data generated using parameters $\hat{\varphi}^*$ and $\hat{\eta}_1^*, \dots, \hat{\eta}_n^*$ and let Q^{**} be the quantile function of $F^{**}(a) := \mathbb{P}^{**}(\hat{w}^{**} \leq a)$, then a likelihood-ratio test of theoretical size α based on the double bootstrap is the decision rule to reject the null if

$$\hat{w} > Q^*(1 - \alpha^*),$$

for α^* a solution to $\hat{\alpha}^*(\alpha^*) = \alpha$ with $\hat{\alpha}^*(a) := 1 - F^*(Q^{**}(1 - a))$.

3 Examples

Many normal means In the classic problem of [Neyman and Scott \(1948\)](#) we observe independent variables

$$z_{it} \sim N(\eta_{i0}, \varphi_0).$$

Maximum likelihood estimates the mean parameters by the within-strata sample averages $\bar{z}_i := 1/m \sum_{t=1}^m z_{it}$ and the common variance parameter by

$$\hat{\varphi} = \frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m (z_{it} - \bar{z}_i)^2.$$

It is well-known that, in this case,

$$\sqrt{nm}(\hat{\varphi} - \varphi_0) \xrightarrow{L} N(-\gamma\varphi_0, 2\varphi_0^2), \quad (3.4)$$

under rectangular-array asymptotics. Here, starting from the fact that $nm\hat{\varphi}/\varphi_0 \sim \chi_{n(m-1)}^2$, the exact distribution of the maximum-likelihood estimator can be derived. We find that

$$\sqrt{nm}(\hat{\varphi} - \varphi_0) \sim \text{Gamma}\left(-\sqrt{nm}\varphi_0, \frac{n(m-1)}{2}, \frac{2\varphi_0}{\sqrt{nm}}\right),$$

where $\text{Gamma}(\vartheta_1, \vartheta_2, \vartheta_3)$ refers to the Gamma distribution with shape ϑ_2 and scale ϑ_3 , shifted by ϑ_1 . It is readily verified that the mean and variance of this distribution are equal to

$$-\sqrt{\frac{n}{m}}\varphi_0, \quad 2\varphi_0^2\left(1 - \frac{1}{m}\right),$$

respectively.

In this example, the bootstrap independently samples $z_{it}^* \sim N(\bar{z}_i, \hat{\varphi})$. The associated maximum-likelihood estimators are \bar{z}_i^* and

$$\hat{\varphi}^* = \frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m (z_{it}^* - \bar{z}_i^*)^2.$$

Conditional on the data, the latter estimator follows the same Gamma distribution as above, only with φ_0 replaced by $\hat{\varphi}$. Noting that we can write $\sqrt{nm}(\hat{\varphi} - \varphi_0) = -\sqrt{n/m}\varphi_0 + \epsilon$, for a mean-zero random variable $\epsilon = O_P(1)$, this implies that

$$\sqrt{nm}(\hat{\varphi}^* - \hat{\varphi}) \sim \text{Gamma}\left(-\left(\sqrt{nm}\varphi_0 - \sqrt{\frac{n}{m}}\varphi_0 + \epsilon\right), \frac{n(m-1)}{2}, \frac{2\varphi_0}{\sqrt{nm}}\left(1 - \frac{1}{m}\right) + \frac{2\epsilon}{nm}\right)$$

conditional on the sample. The mean and variance of this distribution are

$$-\sqrt{\frac{n}{m}}\varphi_0 + \frac{1}{m}\left(\sqrt{\frac{n}{m}}\varphi_0 - \epsilon\right), \quad 2\varphi_0^2 + \frac{2}{m}\left(\epsilon^2 + \sqrt{\frac{m}{n}}\varphi_0\epsilon - 3\varphi_0^2\right) + O\left(\frac{1}{m^2}\right)$$

which, to first order, agree with the corresponding moments of the maximum-likelihood estimator.

The studentized maximum-likelihood estimator follows a (translated) inverse-Gamma distribution, mirrored about the origin. Moreover,

$$-\sqrt{nm} \frac{(\hat{\varphi} - \varphi_0)}{\sqrt{2\hat{\varphi}^2}} \sim \text{Inverse-Gamma}\left(-\sqrt{\frac{nm}{2}}, \frac{n(m-1)}{2}, \sqrt{\frac{nm}{2}} \frac{nm}{2}\right),$$

Table 1: Many normal means: bias, coverage, and size for φ_0

n	m	BIAS	STD	COVERAGE					SIZE	
		MLE		MLE	BC	BB	DBB	SB	LR	LR*
10	10	-0.100	0.134	0.827	0.904	0.918	0.950	0.950	0.128	0.050
20	10	-0.100	0.095	0.763	0.903	0.922	0.950	0.950	0.193	0.050
40	10	-0.100	0.067	0.637	0.902	0.926	0.950	0.950	0.323	0.050
100	10	-0.100	0.042	0.330	0.897	0.927	0.950	0.950	0.642	0.050

where Inverse-Gamma($\vartheta_1, \vartheta_2, \vartheta_3$) refers to the Inverse-gamma distribution with shape ϑ_2 and scale ϑ_3 , shifted by ϑ_1 . This distribution is pivotal and the bootstrap replicates it exactly. Thus, at least in this example, the studentized bootstrap yields confidence intervals whose probability of covering φ_0 can be controlled exactly.

A first-order correction to $\hat{\varphi}$ based on a plug-in estimator of its asymptotic bias is

$$\tilde{\varphi} := \hat{\varphi} + \frac{\hat{\varphi}}{m}.$$

It is interesting to compare the performance of confidence intervals for φ_0 based on bias correction with those obtained via the bootstrap. The bias-correction approach uses the large-sample approximation

$$\sqrt{nm} \frac{(\tilde{\varphi} - \varphi_0)}{\sqrt{2\hat{\varphi}^2}} \xrightarrow{L} N(0, 1).$$

Its coverage accuracy can be evaluated for any given sample size from the observation that

$$-\sqrt{nm} \frac{(\tilde{\varphi} - \varphi_0)}{\sqrt{2\hat{\varphi}^2}} \sim \text{Inverse-Gamma} \left(-\sqrt{\frac{nm}{2}} \left(1 + \frac{1}{m} \right), \frac{n(m-1)}{2}, \sqrt{\frac{nm}{2} \frac{nm}{2}} \right).$$

Notice that this distribution coincides with that of the studentized maximum-likelihood estimator up to the location parameter, the current distribution being located closer to zero.

Table 1 contains the bias and standard deviation of the maximum-likelihood estimator for $\varphi_0 = 1$ and gives coverage rates of two-sided 95% confidence intervals for φ_0 . The rates

are invariant to the value of φ_0 . The bias is not small relative to the standard deviation. Consequently, confidence intervals constructed by means of the naive normal approximation to maximum likelihood (MLE) perform poorly but bootstrapping the maximum-likelihood estimator, both using the basic bootstrap (BB) and the studentized bootstrap (SB), yields reliable inference. Here, the latter gives exact coverage. Iterating the basic bootstrap (DBB) also yields exact coverage. Confidence intervals based on bias correcting (BC) the maximum-likelihood estimator improve considerably on MLE but still undercover by about 5 percentage points in all designs considered.

Next, consider testing the null hypothesis that the variance parameter is equal to φ_0 . The likelihood-ratio statistic is

$$nm \hat{\varphi}/\varphi_0 - nm \log(nm \hat{\varphi}/\varphi_0) - nm + nm \log(nm)$$

and depends on the data only through $nm \hat{\varphi}/\varphi_0$. The latter has a pivotal distribution and, hence, so does the test statistic. A small calculation reveals that its limit distribution is a non-central χ_1^2 -distribution with non-centrality parameter $\gamma^2/2$. Consequently, while a decision rule based on critical values from the χ_1^2 -distribution will not yield size control, using the quantiles of the bootstrap distribution will. Furthermore, in this example size is controlled exact in finite samples. The size distortion of the likelihood-ratio (LR) test when using the .95 quantile of the χ_1^2 -distribution and the improvement when working instead with bootstrap critical values (LR*) is illustrated in Table 1.

Dynamic logit For our next example we consider the Markov process

$$y_{it} = \begin{cases} 1 & \text{if } \eta_{i0} + \varphi_0 y_{it-1} > \varepsilon_{it} \\ 0 & \text{if not} \end{cases},$$

where the ε_{it} are independent and identically distributed logistic random variables, i.e., $\mathbb{P}(\varepsilon_{it} \leq a) = (1 + e^{-a})^{-1} =: F(a)$. The initial conditions, y_{i0} , are observed and held fixed throughout.

The maximum-likelihood estimator is not available in closed form. Nonetheless, the log-likelihood function is globally concave and numerical optimization is straightforward,

exploiting the sparsity of the Hessian matrix (see, e.g., [Chamberlain 1980](#)). Further, an excellent starting value for the bootstrap maximum-likelihood estimator comes in the form of the maximum-likelihood estimator based on the original data, as the latter is used to generate the bootstrap samples. Given $\hat{\varphi}$ and $\hat{\eta}_1, \dots, \hat{\eta}_n$ we generate bootstrap samples for the dynamic logit model by recursively drawing y_{it}^* from a Bernoulli distribution with success probability $F(\hat{\eta}_i + \hat{\varphi}y_{it-1}^*)$. Each bootstrap iteration starts at the initial condition y_{i0} .

The exact distribution of $\hat{\varphi}$ is not known so we resort to simulations. We draw y_{i0} from its stationary distribution,

$$\mathbb{P}(y_{i0} = 1) = \frac{F(\eta_{i0})}{1 - F(\eta_{i0} + \varphi_0) + F(\eta_{i0})},$$

set $\eta_{i0} = 0$ for all the strata, and consider autoregressive parameters $\varphi_0 \in \{1/2, 1, 3/2\}$. [Table 2](#) provides the bias and standard deviation of the maximum-likelihood estimator, the coverage rates and average length of various (two-sided) 95% confidence intervals for φ_0 , and the size of the likelihood-ratio test with a theoretical size of 5% for different choices of critical value. We report coverage and length for confidence intervals based on (the naive normal approximation to) the maximum-likelihood estimator (MLE), the basic bootstrap and studentized bootstrap (BB and SB, respectively) and their iterated version (DBB and DSB, respectively), as well as on two procedures that adjust the maximum-likelihood estimator for its bias. The first of these adjustments (BC1) is the analytical correction of [Hahn and Kuersteiner \(2011\)](#). The second adjustment (BC2) is due to [Fernández-Val \(2009\)](#) and exploits the model structure to implement a refined correction that replaces certain sample averages by expected quantities. Both these approaches require a bandwidth choice. We report results for a bandwidth equal to one, which we found was the choice that performed best. For the likelihood-ratio test we report size for the decision rule based on the .95 quantile of the χ_1^2 -distribution (LR), the .95 quantile of the bootstrap distribution (LR*) and quantiles set according to the double bootstrap (LR**). All (single) bootstrap results are based on the use of 999 bootstrap replications. For the double bootstrap, we use 999 replications in the outer iteration and 316 replications in the inner iteration; the

Table 2: Dynamic logit: bias, coverage and length of confidence interval, and size of likelihood-ratio test for φ_0

φ_0	n	m	BIAS	STD	COVERAGE							LENGTH							SIZE		
			MLE	MLE	BC1	BC2	BB	DBB	SB	DSB	MLE	BC1	BC2	BB	DBB	SB	DSB	LR	LR*	LR**	
1/2	100	10	-0.461	0.145	0.111	0.942	0.970	0.964	0.958	0.928	0.940	0.567	0.573	0.575	0.630	0.616	0.543	0.571	0.887	0.064	0.041
1/2	100	20	-0.219	0.095	0.378	0.952	0.962	0.958	0.955	0.943	0.949	0.378	0.380	0.381	0.396	0.394	0.373	0.381	0.640	0.057	0.048
1/2	250	10	-0.459	0.090	0.001	0.895	0.968	0.957	0.949	0.928	0.932	0.358	0.362	0.363	0.397	0.389	0.343	0.364	0.999	0.092	0.051
1/2	250	20	-0.217	0.062	0.054	0.937	0.952	0.958	0.961	0.942	0.945	0.239	0.241	0.241	0.250	0.250	0.236	0.240	0.957	0.062	0.049
1	100	10	-0.514	0.150	0.086	0.880	0.941	0.964	0.949	0.916	0.937	0.605	0.620	0.623	0.657	0.629	0.577	0.614	0.919	0.113	0.065
1	100	20	-0.244	0.103	0.332	0.907	0.921	0.948	0.948	0.931	0.940	0.404	0.410	0.410	0.418	0.416	0.398	0.408	0.627	0.055	0.040
1	250	10	-0.513	0.095	0.000	0.745	0.898	0.970	0.952	0.907	0.952	0.383	0.392	0.394	0.414	0.393	0.366	0.396	0.999	0.144	0.053
1	250	20	-0.244	0.065	0.039	0.881	0.922	0.959	0.950	0.954	0.944	0.256	0.259	0.259	0.264	0.263	0.251	0.258	0.957	0.069	0.053
3/2	100	10	-0.623	0.165	0.034	0.695	0.837	0.942	0.930	0.867	0.926	0.678	0.700	0.706	0.722	0.675	0.642	0.693	0.966	0.178	0.091
3/2	100	20	-0.299	0.113	0.269	0.835	0.883	0.940	0.930	0.932	0.932	0.454	0.462	0.463	0.463	0.460	0.443	0.458	0.712	0.068	0.047
3/2	250	10	-0.624	0.104	0.000	0.304	0.593	0.917	0.913	0.776	0.936	0.428	0.442	0.446	0.453	0.419	0.405	0.451	1.000	0.274	0.109
3/2	250	20	-0.302	0.071	0.010	0.636	0.741	0.953	0.949	0.932	0.948	0.286	0.292	0.292	0.292	0.290	0.280	0.291	0.981	0.101	0.057

latter choice is based on a suggestion of [Booth and Hall \(1994\)](#).

The naive normal approximation to the sampling distribution of the maximum-likelihood estimator again yields unreliable inference in this problem. Bias correction yields a large improvement in coverage rates and comes with only minor increases in the length of the confidence intervals (which is informative about efficiency). Confidence intervals based on the correction underlying BC2 tend to give better coverage than those based on BC1, with the difference sometimes being considerable (up to 30 percentage points in the table). This highlights the sensitivity of bias-corrected inference to how the bias is being estimated. The performance of both BC1 and BC2 also deteriorates substantially as the value of φ_0 increases, highlighting the sensitivity of bias estimators to relatively minor design changes. Both these issues are not accounted for by first-order theory. Confidence interval based on the bootstrap, both in its basic and in its studentized form, are competitive with those based on bias correction and their performance is stable across different values of φ_0 . BB does at least as well as BC2 in terms of coverage, and its iterated version DBB gives very similar coverage. SB and SDB yield somewhat shorter confidence intervals and, especially in the shortest panels, iterating gives improved coverage. For the likelihood-ratio test we observe a similar pattern as for the studentized bootstrap. LR shows large over-rejection

Table 3: Many normal means: bias and coverage and length of confidence interval for $\lim_{n \rightarrow \infty} 1/n \sum_{i=1}^n \eta_{i0}^2$

n	m	BIAS		COVERAGE						LENGTH					
		MLE	STD	MLE	BC	BB	DBB	SB	DSB	MLE	BC	BB	DBB	SB	DSB
50	10	0.100	0.055	0.550	0.953	0.954	0.952	0.917	0.945	0.232	0.232	0.232	0.231	0.210	0.234
50	20	0.050	0.038	0.702	0.958	0.942	0.934	0.915	0.929	0.156	0.156	0.156	0.153	0.145	0.153
50	50	0.020	0.023	0.782	0.936	0.939	0.932	0.916	0.929	0.095	0.095	0.095	0.094	0.091	0.094
100	10	0.100	0.039	0.256	0.951	0.956	0.962	0.922	0.965	0.163	0.163	0.163	0.165	0.147	0.178
100	20	0.050	0.027	0.517	0.955	0.947	0.942	0.918	0.937	0.110	0.110	0.110	0.109	0.101	0.110
100	50	0.020	0.017	0.702	0.941	0.949	0.945	0.935	0.943	0.067	0.067	0.067	0.066	0.064	0.066

rates while LR* and, even more so, LR** yield tests with size close to nominal size.

Many normal means (cont'd) In our third example we reconsider the setup of [Neyman and Scott \(1948\)](#) but change the parameter of interest to

$$\Delta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \eta_{i0}^2,$$

the second moment of the fixed effects. The plug-in estimator is $1/n \sum_{i=1}^n \bar{z}_i^2$. Using the fact that $\bar{z}_i \sim N(\eta_{i0}, \varphi_0/m)$ by normality of the data it is easy to verify that the plug-in bias due to the estimation of the fixed effects is φ_0/m , while the estimator's sampling variance equals

$$\frac{2\varphi_0}{nm} \left(2 \frac{\sum_{i=1}^n \eta_{i0}^2}{n} + \frac{\varphi_0}{m} \right).$$

The second component in the variance expression is of smaller order and asymptotically negligible.

The bootstrap again independently samples $z_{it}^* \sim N(\bar{z}_i, \hat{\varphi})$ and subsequently constructs the estimator $1/n \sum_{i=1}^n \bar{z}_i^{*2}$. The exact distribution of the estimator is a complicated mixture and so we once more resort to simulations to evaluate the performance of the bootstrap. In our simulations we set $\eta_{i0} = i/n$ so that, in large samples, the distribution of the fixed effects is uniform on $[0, 1]$; hence, $\Delta = 1/3$. Data were generated with $\varphi_0 = 1$. We report results for several choices of (n, m) in [Table 3](#). The bootstrap confidence intervals are again

found to yield a large improvement in coverage rates relative to the ones based on the naive plug-in approach and are competitive with those based on bias correction. Again the basic bootstrap does better than the studentized version and has actual coverage very close to theoretical coverage for all designs. Iterating the former does little in terms of coverage rates. Iterating the latter gives further improvement, especially in the shorter panels. The average length of the confidence intervals is very similar across the different methods.

An empirical illustration For our final example we use data from the Panel Study of Income Dynamics to look at determinants of labor-force participation decisions of married woman. We follow [Hyslop \(1999\)](#) and specify the participation decision as a dynamic probit model with unit-specific intercepts. We included the number of children of at most two years of age (# children 0–2), between 3 and 5 years of age (# children 3–5), and between 6 and 17 years of age (# children 6–17), as well as the log of the husband’s earnings (log husband income; expressed in thousands of 1995 U.S. dollars), and a quadratic function of age. [Carro \(2007\)](#), [Fernández-Val \(2009\)](#), and [Dhaene and Jochmans \(2015\)](#) have previously estimated the same specification using various bias-corrected estimators. To ensure comparability with their results we use the same data, which concern the period 1979–1988. The sample consists of 1461 women aged between 18 and 60 in 1985 who, throughout the sampling period, were married to men who were in the active labor force the whole time.

Table 4 contains points estimates, standard errors, and 95% confidence intervals for the coefficients of the probit model. As before, we provide results for maximum likelihood, the bias-corrected estimators of [Hahn and Kuersteiner \(2011\)](#) and [Fernández-Val \(2009\)](#), the basic bootstrap, and the studentized bootstrap. For maximum likelihood and for the two bias-corrected estimators the standard errors and confidence intervals are based on the conventional normal approximation, using the Hessian matrix evaluated at the point estimates to estimate the Fisher information. For the bootstrap we provide a single point estimate, obtained by subtracting the median of the bootstrap distribution of $\hat{\varphi}^* - \hat{\varphi}$ from $\hat{\varphi}$, and a single standard error, calculated as the standard deviation of the bootstrap

Table 4: Female labor-force participation

	MLE	BC1	BC2	BB	SB
Lagged participation	0.756 (0.043) [0.672 0.840]	0.992 (0.043) [0.908 1.076]	1.031 (0.043) [0.947 1.115]	1.162 (0.045) [1.072 1.248]	1.209 (0.045) [1.048 1.209]
# Children 0-2	-0.554 (0.057) [-0.667 -0.442]	-0.477 (0.058) [-0.591 -0.363]	-0.436 (0.058) [-0.550 -0.322]	-0.369 (0.075) [-0.517 -0.220]	-0.265 (0.075) [-0.521 -0.265]
# Children 3-5	-0.279 (0.053) [-0.384 -0.175]	-0.213 (0.054) [-0.319 -0.107]	-0.193 (0.054) [-0.299 -0.087]	-0.146 (0.068) [-0.280 -0.012]	-0.043 (0.068) [-0.280 -0.043]
# Children 6-17	-0.075 (0.042) [-0.158 0.008]	-0.056 (0.043) [-0.140 0.028]	-0.050 (0.043) [-0.134 0.034]	-0.035 (0.054) [-0.143 0.071]	0.055 (0.054) [-0.137 0.055]
Log husband income	-0.246 (0.055) [-0.354 -0.139]	-0.232 (0.055) [-0.340 -0.124]	-0.209 (0.055) [-0.317 -0.101]	-0.187 (0.069) [-0.319 -0.048]	-0.076 (0.069) [-0.312 -0.076]
Age	2.050 (0.387) [1.292 2.809]	1.844 (0.392) [1.076 2.612]	1.616 (0.392) [0.848 2.384]	1.429 (0.497) [0.438 2.397]	2.366 (0.497) [0.612 2.366]
Age squared	-0.250 (0.051) [-0.351 -0.149]	-0.224 (0.052) [-0.326 -0.122]	-0.224 (0.052) [-0.326 -0.122]	-0.175 (0.066) [-0.302 -0.043]	-0.066 (0.066) [-0.297 -0.066]

distribution of $\hat{\varphi}^* - \hat{\varphi}$ (without winsorization). Bias correction using the mean rather than the median (not reported) yielded very similar point estimates.

The difference between standard maximum likelihood and the other approaches is most pronounced in the coefficients that capture state dependence and the impact of having young children. Adjusting the point estimates for bias leads to an upward revision in each of these coefficients for all methods. This revision is especially large (relative to its standard error) for the coefficient on lagged participation. Its maximum-likelihood estimate is well outside any of the theoretically-justified confidence intervals. The bootstrap bias correction is somewhat larger than that of the two analytical corrections (in line with

other corrections; see [Dhaene and Jochmans 2015](#), Table 11) but the differences are not large relative to the standard errors. The bootstrap-based confidence intervals are similarly located somewhat further away from the maximum-likelihood point estimates, as are those obtained by centering a conventional confidence interval around a bias-corrected estimator. Confidence intervals based on the studentized bootstrap are somewhat shorter and more asymmetric than those based on the basic bootstrap. With the exception of the one for lagged participation, the bootstrap confidence intervals are also slightly longer than those for the bias-corrected estimators, which are based on (first-order) asymptotic arguments. All this is in line with what has been observed in the simulation results for the dynamic logit model.

4 Asymptotic theory

Our results hold under a set of assumptions that are standard in the literature. The following formulation is mostly borrowed from [Kim and Sun \(2016\)](#). It differs from [Hahn and Kuersteiner \(2011\)](#) in two respects that are worth noting. The first difference is that the individual time series need not be stationary. This is useful because the requirement that the initial condition is a draw from the steady-state distribution, for example, is often hard to justify. The second difference is that certain requirements are assumed to hold uniformly over a neighborhood of the true parameter value. This is useful for the derivation of our results because it allows us to adopt a technique introduced in [Andrews \(2005\)](#). This technique is to first demonstrate a convergence result for the maximum-likelihood estimator uniformly over a set around the true parameter value. Then, as consistency implies that the maximum-likelihood estimator lies in this set with probability approaching one, this allows us to establish the corresponding property for the bootstrap estimator.

In the assumptions (and in the proofs) it is important to make clear under which data generating process certain expectations and probabilities are being computed. We will write \mathbb{E}_θ and \mathbb{P}_θ for expectations and probabilities involving data that were generated using parameters $\theta := (\varphi, \eta_1, \dots, \eta_n)$. Note that some objects, such as $\mathbb{E}_\theta(z_{it})$, only depend

on a subset of the elements of θ . For simplicity, however, we do not make this explicit in the notation.

Denote by V_φ and V_η the parameter space for φ and η_i , respectively. Then the parameter space for θ is the Cartesian product $\Theta := V_\varphi \times V_\eta \times \cdots \times V_\eta$. We let Θ_0 be a subset of Θ .

Assumption 1.

- (i) *The density f is a continuous function in $\varphi \in V_\varphi$ and $\eta_i \in V_\eta$.*
- (ii) *The true parameter value lies in the interior of Θ_0 , a subset of the compact set Θ .*

For our next assumption, consider the mixing coefficients

$$a_i(\theta, h) := \sup_{1 \leq t \leq m} \sup_{A \in \mathcal{A}_{it}(\theta)} \sup_{B \in \mathcal{B}_{it+h}(\theta)} |\mathbb{P}_\theta(A \cap B) - \mathbb{P}_\theta(A) \mathbb{P}_\theta(B)|,$$

where $\mathcal{A}_{it}(\theta)$ and $\mathcal{B}_{it}(\theta)$ are the sigma algebras generated by the sequences z_{it}, z_{it-1}, \dots and z_{it}, z_{it+1}, \dots when these sequences were generated from our model with the parameter equal to θ .

We will also make use of an open set that covers Θ_0 . This set is of the form

$$\Theta_1 := \{\theta \in \Theta : d(\theta, \Theta_0) < \delta\}$$

for some $\delta > 0$, where $d(\theta, \Theta_0) := \inf\{\|\theta - \vartheta\|_2 : \vartheta \in \Theta_0\}$, i.e., the distance between the point θ and the set Θ_0 .

Assumption 2. $\sup_{1 \leq i \leq n} \sup_{\theta \in \Theta_1} a_i(\theta, h) = O(r^h)$ for some constant $0 < r < 1$.

The next assumption collects smoothness conditions and moment requirements.

Assumption 3.

- (i) *The function $\ell(\varphi, \eta_i | z_{it})$ is almost surely four times continuously-differentiable in φ and η_i .*
- (ii) *The function $\ell(\varphi, \eta_i | z_{it})$ and all its cross-derivatives up to fourth order are almost surely bounded by a function $b(z_{it})$ for which*

$$\sup_{1 \leq i \leq n} \sup_{1 \leq t \leq m} \sup_{\theta \in \Theta_1} \mathbb{E}_\theta(|b(z_{it})|^q) < \infty$$

for some q such that $3 + (\dim(\varphi) + \dim(\eta_i))/2 < qs$ with $0 < s < 1/10$.

(iii) As $m \rightarrow \infty$, $1/m \sum_{t=1}^m \mathbb{E}_\theta(b(z_{it}))$ converges to $\lim_{m \rightarrow \infty} 1/m \sum_{t=1}^m \mathbb{E}_\theta(b(z_{it}))$ uniformly in i and $\theta \in \Theta_1$.

Let

$$H_i(\varphi, \eta_i | \vartheta) := \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{t=1}^m \mathbb{E}_\theta(\ell(\varphi, \eta_i | z_{it})).$$

The next assumption ensures that our parameters are identified from time series variation.

Assumption 4. For each $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that

$$\inf_{1 \leq i \leq n} \inf_{\theta \in \Theta_1} \left(H_i(\varphi, \eta_i | \theta) - \sup_{\{(\bar{\varphi}, \bar{\eta}_i) : \|(\bar{\varphi}, \bar{\eta}_i) - (\varphi, \eta_i)\|_2 > \varepsilon\}} H_i(\bar{\varphi}, \bar{\eta}_i | \theta) \right) > \delta_\varepsilon.$$

Assumption 5 states that we are working under rectangular-array asymptotics.

Assumption 5. As $n, m \rightarrow \infty$, $n/m \rightarrow \gamma^2$ for some $0 < \gamma < \infty$.

The last assumption ensures a well-defined asymptotic variance for $\hat{\varphi}$. We write $\Omega_{nm, \theta}$ for the matrix defined below (1.1) to highlight its dependence on θ , and $\varpi_{\min}(A)$ and $\varpi_{\max}(A)$ for the smallest and largest eigenvalue of a square matrix A .

Assumption 6. There exist positive finite constants ϵ_1, ϵ_2 and $\varepsilon_1, \varepsilon_2$ such that, for n and m large enough,

$$(i) \quad \epsilon_1 \leq \inf_{1 \leq i \leq n} \inf_{\theta \in \Theta_1} \varpi_{\min} \left(\frac{1}{m} \sum_{t=1}^m \mathbb{E}_\theta \left(\frac{\partial^2 \ell(\varphi, \eta_i | z_{it})}{\partial \eta_i \partial \eta_i'} \right) \right) \\ \leq \sup_{1 \leq i \leq n} \sup_{\theta \in \Theta_1} \varpi_{\max} \left(\frac{1}{m} \sum_{t=1}^m \mathbb{E}_\theta \left(\frac{\partial^2 \ell(\varphi, \eta_i | z_{it})}{\partial \eta_i \partial \eta_i'} \right) \right) \leq \epsilon_2,$$

$$(ii) \quad \varepsilon_1 < \inf_{\theta \in \Theta_1} \varpi_{\min}(\Omega_{nm, \theta}) \leq \sup_{\theta \in \Theta_1} \varpi_{\max}(\Omega_{nm, \theta}) < \varepsilon_2.$$

Our first result is stated in the following theorem.

Theorem 1. Let Assumptions 1–6 hold. Then

$$\mathbb{P} \left(\sup_a \left| \mathbb{P}^*(\sqrt{nm}(\hat{\varphi}^* - \hat{\varphi}) \leq a) - \mathbb{P}(\sqrt{nm}(\hat{\varphi} - \varphi_0) \leq a) \right| > \varepsilon \right) = o(1)$$

for any $\varepsilon > 0$.

Theorem 1, through the following corollary, justifies the use of the basic bootstrap to conduct inference on $c'\varphi_0$.

Corollary 1. *Let Assumptions 1–6 hold. Let $Q^*(\alpha)$ be the smallest value Q^* for which $\mathbb{P}^*(c'(\hat{\varphi}^* - \hat{\varphi}) \leq Q^*) \geq \alpha$, where c is a given vector of conformable dimension with $\|c\|_1 < \infty$.*

Then

$$\mathbb{P}(c'\hat{\varphi} - Q^*(\alpha) \leq c'\varphi_0) = \alpha + o(1)$$

for any $\alpha \in (0, 1)$.

A consistency result for $\hat{\Sigma}$ and $\hat{\Sigma}^*$ is given in the supplementary material. This leads to our next corollary.

Corollary 2. *Let Assumptions 1–6 hold. Let $Q^*(\alpha)$ be the smallest value Q^* for which $\mathbb{P}^*((c'\hat{\Sigma}^*c)^{-1/2}c'(\hat{\varphi}^* - \hat{\varphi}) \leq Q^*) \geq \alpha$, where c is a given vector of conformable dimension with $\|c\|_1 < \infty$. Then*

$$\mathbb{P}\left(c'\hat{\varphi} - (c'\hat{\Sigma}c)^{1/2}Q^*(\alpha) \leq c'\varphi_0\right) = \alpha + o(1)$$

for any $\alpha \in (0, 1)$.

Corollary 2 implies that confidence intervals constructed via the studentized bootstrap yield correct coverage in large samples. Another consequence is that a hypothesis test obtained through inversion of a $(1 - \alpha)$ confidence interval so constructed—that is, a conventional t -test—will have size approaching α as $n, m \rightarrow \infty$. Furthermore, p -values for such a test calculated from the bootstrap distribution are asymptotically uniformly distributed on $(0, 1)$.

Our next theorem provides a delta method for the bootstrap.

Theorem 2. *Let Assumptions 1–6 hold. Let ϕ be a non-random vector-valued function that is continuously-differentiable on V_φ . Then*

$$\mathbb{P}\left(\sup_a \left| \mathbb{P}^*(\sqrt{nm}(\phi(\hat{\varphi}^*) - \phi(\hat{\varphi})) \leq a) - \mathbb{P}(\sqrt{nm}(\phi(\hat{\varphi}) - \phi(\varphi_0)) \leq a) \right| > \varepsilon\right) = o(1)$$

for any $\varepsilon > 0$.

Theorem 2 allows to extend Corollaries 1 and 2 to cover inference on nonlinear parameter transformations $\phi(\varphi_0)$.

Our final result concerns the behavior of the likelihood-ratio statistic.

Theorem 3. *Let Assumptions 1–6 hold. Consider testing the null hypothesis that $\phi(\varphi_0) = 0$ for a non-random function ϕ . Suppose that the true parameter value lies in the interior of the set $\Theta \cap \{\varphi \in V_\varphi : \phi(\varphi) = 0\}$ and that ϕ is five times continuously-differentiable on V_φ with bounded derivatives and Jacobian matrix with maximal row rank. Then, under the null,*

$$\mathbb{P} \left(\sup_a |\mathbb{P}^*(\hat{w}^* \leq a) - \mathbb{P}(\hat{w} \leq a)| > \varepsilon \right) = o(1)$$

for any $\varepsilon > 0$.

The chief implication of this result is that bootstrapping the likelihood-ratio statistic yields size control. The same conclusion can be reached for the score statistic; we refer to the supplement for a derivation.

Conclusion

The purpose of this paper has been to show that, in panel data models with fixed effects, inference based on the likelihood remains valid under rectangular-array asymptotics when done by means of the parametric bootstrap. This provides a reevaluation of maximum likelihood in problems with incidental parameters, where its use has been dissuaded.

Our setup covers a broad class of nonlinear models and allows for dynamics in the outcome of interest. Our results do rely on the likelihood being correctly specified. An implication is that any feedback from outcomes to covariates, or any error dependence, must be modelled. Some approaches to bias correction, in contrast, can be applied more generally. Whether a bootstrap procedure that applies to the partial-likelihood setting can be devised is the topic of ongoing work. [Gonçalves and Kaffo \(2015\)](#) have shown that the wild bootstrap replicates the bias in the linear autoregressive setup of [Hahn and](#)

[Kuersteiner \(2002\)](#). Their approach is residual-based but is tailored quite specifically to the linear model.

Although our attention has been devoted to one-way models, we see no reason why our findings would not carry over to models with two-way fixed effects. Such models are useful to capture aggregate time effects and can be applied to estimate dyadic-interaction models. Two-step fixed-effect estimators should also be amenable to bootstrapping.

Our panel data problem is an example of the general challenge to conduct inference when the number of parameters increases with the sample size. The performance of the bootstrap has been investigated for linear regression models with many regressors ([Bickel and Freedman 1983](#), [Mammen 1993](#)) and for linear instrumental-variable estimators with many instruments ([Wang and Kaffo 2016](#)). The bootstrap can be successfully applied there provided that the number of parameters to estimate grows at a certain rate that is slower than the sample size. Recently, [Cattaneo, Jansson and Ma \(2019\)](#) uncovered an asymptotic bias in (possibly nonlinear) two-step estimators when the number of regressors included in the first step grows proportionally to the square-root of the sample size. This is precisely the rate condition implied by our rectangular-array asymptotics. It seems plausible that a version of the bootstrap can be used to sidestep bias correction here in the same way as in the panel problem.

References

- Andrews, D. W. K. (2005). Higher-order improvements of the parametric bootstrap for Markov processes. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models*, Chapter 9, pp. 171–215. Cambridge University Press.
- Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. In R. Blundell, W. K. Newey, and T. Persson (Eds.), *Advances In Economics and Econometrics*, Volume III. Econometric Society: Cambridge University Press.

- Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–697.
- Bickel, P. J. and D. A. Freedman (1983). Bootstrapping regression models with many parameters. In *A Festschrift for Erich L. Lehmann*, pp. 28–48. Wadsworth Statistics/Probability Series.
- Booth, J. G. and P. Hall (1994). Monte Carlo approximation and the iterated bootstrap. *Biometrika* 81, 331–340.
- Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics* 140, 503–528.
- Cattaneo, M. D., M. Jansson, and X. Ma (2019). Two-step estimation and inference with possibly many included covariates. *Review of Economic Studies* 86, 1095–1122.
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies* 47, 225–238.
- Chamberlain, G. (1984). Panel data. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2 of *Handbook of Econometrics*, Chapter 22, pp. 1247–1315. Elsevier.
- Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
- Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82, 991–1030.
- Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*, Volume 38 of *CBMS-NSF Monographs*. Society of Industrial and Applied Mathematics.
- Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150, 71–85.
- Gonçalves, S. and M. Kaffo (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics* 186, 407–426.
- Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.
- Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models

- with fixed effects. *Econometric Theory* 27, 1152–1191.
- Hahn, J. and Z. Liao (2021). Bootstrap standard error estimates and inference. *Econometrica* 89, 1963–1977.
- Hahn, J. and W. K. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.
- Honoré, B. E. and E. Tamer (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74, 611–629.
- Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica* 67, 1255–1294.
- Kim, M. S. and Y. Sun (2016). Bootstrap and k -step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models. *Econometric Theory* 32, 1523–1568.
- Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies* 69, 647–666.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255–285.
- Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–32.
- Wang, W. and M. Kaffo (2016). Bootstrap inference for instrumental variable models with many weak instruments. *Journal of Econometrics* 192, 231–268.