Bootstrap inference for fixed-effect models

Ayden Higgins*
University of Oxford

Koen Jochmans†
Toulouse School of Economics

This version: September 21, 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

*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.

Funded by the European Union (ERC-2016-STG-715787 and ERC-NETWORK-101044319) and by the French Government and the French National Research Agency under the Investissements d' Avenir program (ANR-17-EURE-0010). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

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.

where the number of individuals, n, grows large while the number of time periods, m, is 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 when m grows large with n. Under rectangular-array asymptotics, where n and m grow at the same rate, it is asymptotically biased, however, in general. Therefore, inference based on a naive normal approximation to the distribution of the maximum-likelihood estimator is incorrect, even in large samples.

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 a discussion and many references. 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 (see the simulations in Dhaene and Jochmans 2015, for example).

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 centered percentile bootstrap or the reverse-percentile bootstrap; we follow the terminology of Davison and Hinkley 1997, p. 194) or its studentized version have correct coverage and size, respectively, in large samples. Thus, bias correction is not needed. The same conclusion is true for averages over the fixed effects, such as average marginal effects, and for the size of the likelihood-ratio test and the score test. We stress that our bootstrap schemes are different from Efron's (1982, p. 87) percentile bootstrap (Davison and Hinkley 1997, p. 203). The latter bootstrap uses quantiles of uncentered statistics, 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 give an empirical illustration and simulation results. 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. The Appendix contains the proofs. The Online Supplement gathers auxiliary results used in the proofs.

1 Maximum-likelihood estimation

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

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

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

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

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

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

In sufficiently-regular models we have, as $n, m \to \infty$ with $n/m \to \gamma^2$ 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 $\Sigma := (\lim_{n,m\to\infty} \Omega_{nm})^{-1}$ for

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

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

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

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 asymptotic bias and variance of $\hat{\Delta}$ are complicated and are 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

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

and the likelihood-ratio statistic equals

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

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 $y_i^* \coloneqq (y_{i1}^*, \ldots, y_{im}^*)$ is generated recursively by drawing y_{it}^* from the fitted transition density obtained from the original data,

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

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

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

with $z_{it}^* \coloneqq (y_{it-p}^*, \ldots, y_{it-1}^*, y_{it}^*, x_{it})$. 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,

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

as $n, m \to \infty$ with $n/m \to \gamma^2$. Throughout, we use $\overset{L^*}{\to}$ 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

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

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

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

be the implied quantile function. Then

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

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). We reiterate that working with the α quantile of the distribution of the centered $c'(\hat{\varphi}^* - \hat{\varphi})$ instead of the $(1-\alpha)$ quantile of the uncentered $c'\hat{\varphi}^*$, as in Efron's (1982) original proposal, is important for this to be the case.

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 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 $y_i^{**} := (y_{i1}^{**}, \ldots, y_{im}^{**})$ using the density $f(y_{it}^{**}|y_{it-p}^{**}, \ldots, y_{it-1}^{**}, x_{it}; \hat{\varphi}^*, \hat{\eta}_i^*)$ 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. In several examples (not reported on here) we found that iterating the bootstrap typically yielded confidence intervals with improved coverage, although we do not formally establish that iterating yields any asymptotic refinements in this paper.

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 $\sqrt{nm}(\hat{\mathit{\Delta}}^* - \hat{\mathit{\Delta}})$ mimics the distribution of $\sqrt{nm}(\hat{\mathit{\Delta}} - \mathit{\Delta})$ in large samples. The construction of confidence intervals for $\mathit{\Delta}$ 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, ${}^{c'\hat{\mathit{\Sigma}}\,c}/_{nm}$ 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

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

exceeds the $(1-\alpha)$ quantile of the distribution of $(c'\hat{\mathit{\Sigma}}^*\,c)^{-1/2}\,c'(\hat{\varphi}^* - \hat{\varphi})$ 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

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

The corresponding likelihood-ratio statistic is

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

Redefine $F^*(a) \coloneqq \mathbb{P}^*(\hat{w}^* \leq a)$. 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 $p^* \coloneqq 1 - F^*(\hat{w})$ 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 $\hat{\eta}_1^*, \ldots, \hat{\eta}_n^*$ and let Q^{**} be the quantile function of $F^{**}(a) \coloneqq \mathbb{P}^{**}(\hat{w}^{**} \leq a)$, 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 $\hat{\alpha}^*(a) \coloneqq 1 - F^*(Q^{**}(1-a))$.

3 Numerical illustrations

Empirical example We look at the determinants of labor-force participation decisions of married woman using data from the Panel Study of Income Dynamics (PSID). We follow Hyslop (1999) and specify the participation decision as a first-order dynamic probit model

Table 1: Female labor-force participation: Empirical results.

	Coefficient	Standard error	Confidence interval
Maximum likelihood	0.756	0.043	[0.672, 0.840]
Hahn and Kuersteiner (2011)	0.992	0.043	[0.908, 1.075]
Fernández-Val (2009)	1.031	0.043	[0.948, 1.115]
Bootstrap-based bias correction	1.163	0.045	[1.075, 1.251]
Basic bootstrap	—	—	[1.077, 1.250]
Double basic bootstrap	—	—	[1.105, 1.262]
Studentized bootstrap	—	—	[1.053, 1.210]
Double studentized bootstrap	—	—	[1.105, 1.279]

The corrections of Hahn and Kuersteiner (2011) and Fernández-Val (2009) were implemented with a bandwidth set to one, and the bootstrap with 9,999 replications and no winsorization. The bootstrap-based confidence intervals are based on 9,999 replications in the outer loop and (for the iterated bootstrap) 999 replications in the inner loop, following Booth and Hall (1994).

with unit-specific intercepts. As covariates we include the number of children of at most two years of age, between 3 and 5 years of age, and between 6 and 17 years of age, as well as the log of the husband's earnings and a quadratic function of age. Carro (2007), Fernández-Val (2009), and Dhaene and Jochmans (2015) have previously estimated the same specification. For comparability we use the same data, which concern 1461 women and span the period 1979–1988 (available from Fernández-Val 2022).

We focus on the autoregressive coefficient of the participation decision. In the top panel of Table 1 we first report point estimates and standard errors, along with the implied 95% confidence intervals based on a conventional normal approximation, for four different approaches. The first approach is standard maximum likelihood. The second and third approaches are the proposals of Hahn and Kuersteiner (2011) and Fernández-Val (2009) for bias correction, which consist of shifting the maximum-likelihood estimate by an estimate of its leading bias. For each of these the standard error is calculated from the Hessian of the log-likelihood evaluated at the maximum-likelihood estimates. The final approach reported on uses the bootstrap distribution of the maximum-likelihood estimator to correct for bias and to estimate the asymptotic variance, as described above. Here, the point estimate is calculated by subtracting the median of the bootstrap distribution from the maximum-likelihood estimate (using the mean instead gives essentially the same result) and the standard error as the standard deviation of the same bootstrap distribution (without winsorization). Adjusting the maximum-likelihood point estimate for bias leads to an upward revision in the amount of state dependence for all approaches considered. The bootstrap-based revision is larger than those of the analytical corrections. The plug-in estimate of the standard error is very similar to the standard deviation of the bootstrap distribution. It is apparent that the maximum-likelihood point estimate is well outside of any of the bias-corrected confidence intervals.

In the second panel of Table 1 we provide 95% confidence intervals obtained via the basic bootstrap and via the studentized bootstrap applied to the maximum-likelihood estimator, and also via their iterated versions. The basic bootstrap yields a confidence interval that is close to the one obtained by fitting a normal approximation to the bootstrap bias-corrected estimator. Iterating the basic bootstrap narrows the interval by adjusting its left endpoint upward. The interval based on the studentized bootstrap is located more to the left; iterating yields an interval that is very similar to that of the (iterated) basic bootstrap. We remark that these intervals are not symmetric around the bootstrap-based point estimate.

Simulation We now provide results from a simulation exercise built around our data. We fix initial conditions and covariates to their observed values and generate outcomes from a dynamic probit specification. For the autoregressive coefficient we consider four different values; $0, 1/2, 1$, and $3/2$. The remaining parameters, including the fixed effects, are set to their maximum-likelihood estimates. In Table 2 we report bias, standard deviation, and coverage rates for 95% confidence intervals (computed over 2,500 Monte Carlo replications). We consider the same estimators as before and use a normal approximation to construct confidence intervals for all but the bootstrap, where we rely on the basic bootstrap instead.

The corrections of Hahn and Kuersteiner (2011) and Fernández-Val (2009) reduce bias relative to maximum likelihood in all designs. The latter is more successful in doing so; this can be explained by the fact that it exploits the model structure to implement a refined correction of the former by replacing certain sample averages by expected quantities. The extent to which bias is reduced decreases as the state-dependence parameter gets larger. In all but one case, and then only for the approach of Fernández-Val (2009), the bias remains important relative to the standard deviation, implying that the coverage rates for the analytical corrections are dramatically lower than the desired level of 95%. As such, these methods are not effective in our setup.

Use of the (median of the) bootstrap distribution leads to considerably more reduction of the bias in all four cases. The state-dependence coefficient is, on average, estimated to be larger compared to the other corrections. This is fully in line with what was observed in the estimation results in Table 1. The (basic) bootstrap-based confidence intervals also display far better coverage. Only in the most challenging design does the coverage rate fall short in a substantial manner. The coverage properties for the studentized bootstrap (not reported) are inferior to those of the basic bootstrap, although this can be alleviated to a certain extent by iterating the bootstrap. This is a pattern we observed in a further set of simulation experiments.

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

Table 2: Female labor-force participation: Simulation results

	Bias	Standard deviation	Coverage rate
true parameter value 0			
Maximum likelihood	-0.351	0.042	0.000
Hahn and Kuersteiner (2011)	-0.067	0.040	0.675
Fernández-Val (2009)	0.006	0.039	0.972
(Basic) bootstrap	0.007	0.042	0.978
true parameter value $1/2$			
Maximum likelihood	-0.381	0.044	0.000
Hahn and Kuersteiner (2011)	-0.123	0.041	0.194
Fernández-Val (2009)	-0.056	0.039	0.786
(Basic) bootstrap	-0.010	0.044	0.963
true parameter value 1			
Maximum likelihood	-0.439	0.047	0.000
Hahn and Kuersteiner (2011)	-0.233	0.045	0.002
Fernández-Val (2009)	-0.149	0.041	0.089
(Basic) bootstrap	-0.039	0.049	0.914
true parameter value $3/2$			
Maximum likelihood	-0.502	0.053	0.000
Hahn and Kuersteiner (2011)	-0.385	0.057	0.000
Fernández-Val (2009)	-0.271	0.044	0.000
(Basic) bootstrap	-0.078	0.054	0.782

The corrections of Hahn and Kuersteiner (2011) and Fernández-Val (2009) were implemented with a bandwidth set to one, and the bootstrap with 999 replications. Results are based on 2,500 Monte Carlo replications.

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

Let

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

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 $n, m \to \infty$, ${}^n/_m \to \gamma^2$ 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)

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

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

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*

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

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

A consistency result for $\hat{\Sigma}$ and $\hat{\Sigma}^*$ is given in the Supplement. 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 $n, m \to \infty$. 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.

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. Devising a bootstrap procedure that applies outside the likelihood setting appears possible and is the topic of ongoing work.

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) 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.

Appendix

Proof of Theorem 1. Note that

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

is bounded from above by

$$\sup_{\theta\in\Theta_0} \mathbb{P}_\theta\left(\sup_a \left|\mathbb{P}_{\hat{\theta}}(\sqrt{nm}(\hat{\varphi}^* - \hat{\varphi}) \leq a) - \mathbb{P}_\theta(\sqrt{nm}(\hat{\varphi} - \varphi) \leq a)\right| > \varepsilon\right)$$

which, in turn, is below

$$\begin{aligned} &\sup_{\theta\in\Theta_0} \mathbb{P}_\theta\left(\sup_a \left|\mathbb{P}_\theta(\sqrt{nm}(\hat{\varphi} - \varphi) \leq a) - \mathbb{P}_\theta(v_\theta \leq a)\right| > \frac{\varepsilon}{2}\right) \\ &+ \sup_{\theta\in\Theta_0} \mathbb{P}_\theta\left(\sup_a \left|\mathbb{P}_{\hat{\theta}}(\sqrt{nm}(\hat{\varphi}^* - \hat{\varphi}) \leq a) - \mathbb{P}_\theta(v_\theta \leq a)\right| > \frac{\varepsilon}{2}\right). \end{aligned} \tag{A.1}$$

Here and later, we let

$$v_\theta \sim N(\gamma\beta_\theta, \Sigma_\theta)$$

for β_θ and Σ_θ the bias and asymptotic variance of the maximum-likelihood estimator for data generated with parameter θ. Therefore, it suffices to show that each of the terms in (A.1) is $o(1)$.

Theorem S.2 in the Supplement shows that

$$\sup_{\theta\in\Theta_1} \left|\mathbb{P}_\theta(\sqrt{nm}(\hat{\varphi} - \varphi) \leq a) - \mathbb{P}_\theta(v_\theta \leq a)\right| = o(1)$$

for any a. Further, because the normal distribution is a continuous function, we have that

$$\sup_{\theta\in\Theta_1} \left(\sup_a \left|\mathbb{P}_\theta(\sqrt{nm}(\hat{\varphi} - \varphi) \leq a) - \mathbb{P}_\theta(v_\theta \leq a)\right|\right) = o(1) \tag{A.2}$$

by Polya's theorem. This allows us to invoke Lemma A.1 of Andrews (2005) to establish that

$$\sup_{\theta\in\Theta_0} \mathbb{P}_\theta\left(\sup_a \left|\mathbb{P}_\theta(\sqrt{nm}(\hat{\varphi} - \varphi) \leq a) - \mathbb{P}_\theta(v_\theta \leq a)\right| > \frac{\varepsilon}{2}\right) = o(1).$$

This handles the first term in (A.1).

Moving on to the second term in (A.1), note that

$$\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat\theta}(\sqrt{nm}(\hat\varphi^*-\hat\varphi)\le a)-\mathbb{P}_\theta(v_\theta\le a)\right|>\frac{\varepsilon}{2}\right)$$
$$\le\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat\theta}(\sqrt{nm}(\hat\varphi^*-\hat\varphi)\le a)-\mathbb{P}_{\hat\theta}(v_{\hat\theta}\le a)\right|>\frac{\varepsilon}{4}\right)$$
$$+\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat\theta}(v_{\hat\theta}\le a)-\mathbb{P}_\theta(v_\theta\le a)\right|>\frac{\varepsilon}{4}\right).$$

Here, using (A.2), coupled with the consistency result

$$\sup_{\theta\in\Theta_1}\mathbb{P}_\theta(\|\hat\varphi-\varphi\|_2>\epsilon)=o(m^{-1}),\qquad \sup_{\theta\in\Theta_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\|\hat\eta_i-\eta_i\|_2>\epsilon\right)=o(m^{-1}),\tag{A.3}$$

for any $\epsilon>0$ (which follows from a minor modification to Theorem 1 of Kim and Sun 2016), by another application of Lemma A.1 of Andrews (2005),

$$\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat\theta}(\sqrt{nm}(\hat\varphi^*-\hat\varphi)\le a)-\mathbb{P}_{\hat\theta}(v_{\hat\theta}\le a)\right|>\frac{\varepsilon}{4}\right)=o(1)$$

while, again using (A.3),

$$\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat\theta}(v_{\hat\theta}\le a)-\mathbb{P}_\theta(v_\theta\le a)\right|>\frac{\varepsilon}{4}\right)=o(1)$$

follows from the continuous mapping theorem. This takes care of the second term in (A.1) and completes the proof of the theorem. □

Proof of Corollary 1. Given Theorem 1 the proof follows from an application of Lemma 23.3 in van der Vaart (2000). □

Proof of Corollary 2. From Theorems S.1 and S.3 in the Supplement it immediately follows that

$$\sup_{\theta\in\Theta_1}\mathbb{P}_\theta\left(\left\|\sqrt{nm}\,\hat\Sigma_\theta^{-1/2}(\hat\varphi-\varphi)-\sqrt{nm}\,\Sigma_\theta^{-1/2}(\hat\varphi-\varphi)\right\|_2>\varepsilon\right)=o(1)$$

for all $\varepsilon>0$. Also, from Theorem S.2 in the Supplement and the continuous mapping theorem,

$$\sup_{\theta\in\Theta_1}\left|\mathbb{P}_\theta(\sqrt{nm}\,\Sigma_\theta^{-1/2}(\hat\varphi\ -\varphi)\le a)-\mathbb{P}_\theta(\Sigma_\theta^{-1/2}v_\theta\le a)\right|=o(1)$$

for any a. Because $\Sigma_\theta^{-1/2} v_\theta \sim N(\gamma\, \Sigma_\theta^{-1/2} \beta_\theta, I)$ and the normal distribution is continuous we can apply Lemma S.4 from the Supplement to obtain

$$\sup_{\theta \in \Theta_1} \left| \mathbb{P}_\theta(\sqrt{nm}\, \hat{\Sigma}_\theta^{-1/2}(\hat{\varphi} \; - \varphi) \leq a) - \mathbb{P}_\theta(\Sigma_\theta^{-1/2} v_\theta \leq a) \right| = o(1).$$

With this result in hand we may proceed in precisely the same way as we did in the proof of Theorem 1 to establish that

$$\mathbb{P}\left(\sup_a \left| \mathbb{P}^*(\sqrt{nm}\, \hat{\Sigma}^{*^{-1/2}}(\hat{\varphi}^* - \hat{\varphi}) \leq a) - \mathbb{P}(\sqrt{nm}\, \hat{\Sigma}^{-1/2}(\hat{\varphi} - \varphi_0) \leq a) \right| > \varepsilon \right) = o(1)$$

for all $\varepsilon > 0$. An application of Lemma 23.3 in van der Vaart (2000) then yields the result. □

Proof of Theorem 2. Without loss of generality we take ϕ to be scalar valued. We let

$$\Phi_{\bar{\varphi}} := \left. \frac{\partial \phi(\varphi)}{\partial \varphi'} \right|_{\varphi = \bar{\varphi}}.$$

As in the proof of Theorem 1 it suffices to show that

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(\sup_a \left| \mathbb{P}_{\hat{\theta}}(\sqrt{nm}(\phi(\hat{\varphi}^*) - \phi(\hat{\varphi})) \leq a) - \mathbb{P}_\theta(\sqrt{nm}(\phi(\hat{\varphi}) - \phi(\varphi)) \leq a) \right| > \varepsilon \right) = o(1)$$

for all $\varepsilon > 0$. An upper bound on this probability is

$$\begin{aligned} &\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(\sup_a \left| \mathbb{P}_\theta(\sqrt{nm}(\phi(\hat{\varphi}) \; - \phi(\varphi)) \leq a) - \mathbb{P}_\theta(\Phi_\varphi v_\theta \leq a) \right| > \frac{\varepsilon}{2} \right) \\ &+ \sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(\sup_a \left| \mathbb{P}_{\hat{\theta}}(\sqrt{nm}(\phi(\hat{\varphi}^*) - \phi(\hat{\varphi})) \leq a) - \mathbb{P}_\theta(\Phi_\varphi v_\theta \leq a) \right| > \frac{\varepsilon}{2} \right), \end{aligned} \tag{A.4}$$

where v_θ is as in the proof of Theorem 1. We handle each of these terms, in turn, showing that they are $o(1)$ for all $\varepsilon > 0$.

Following the proof of Theorem 3.8 in van der Vaart (2000) by using Theorem S.2 in the Supplement gives

$$\sup_{\theta \in \Theta_1} \mathbb{P}_\theta(\| \sqrt{nm}(\phi(\hat{\varphi}) - \phi(\varphi)) - \Phi_\varphi \sqrt{nm}(\hat{\varphi} - \varphi) \|_2 > \varepsilon) = o(1)$$

for all $\varepsilon > 0$. The continuous mapping theorem, again combined with Theorem S.2 in the Supplement, yields

$$\sup_{\theta\in\Theta_1}|\mathbb{P}_\theta(\Phi_\varphi\sqrt{nm}(\hat{\varphi}-\varphi)\leq a)-\mathbb{P}_\theta(\Phi_\varphi v_\theta\leq a)|=o(1)$$

for any a. Because v_θ is a normal random variable its distribution function is continuous. We may thus apply Lemma S.4 from the Supplement to obtain

$$\sup_{\theta\in\Theta_1}|\mathbb{P}_\theta(\sqrt{nm}(\phi(\hat{\varphi})-\phi(\varphi))\leq a)-\mathbb{P}_\theta(\Phi_\varphi v_\theta\leq a)|=o(1),$$

and, next, invoke Polya's theorem to establish that

$$\sup_{\theta\in\Theta_1}\left(\sup_a|\mathbb{P}_\theta(\sqrt{nm}(\phi(\hat{\varphi})-\phi(\varphi))\leq a)-\mathbb{P}_\theta(\Phi_\varphi v_\theta\leq a)|\right)=o(1). \tag{A.5}$$

This allows us to apply Lemma A.1 of Andrews (2005) and conclude that

$$\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_\theta(\sqrt{nm}(\phi(\hat{\varphi})\ -\phi(\varphi))\leq a)-\mathbb{P}_\theta(\Phi_\varphi v_\theta\leq a)\right|>\frac{\varepsilon}{2}\right)=o(1)$$

for all $\varepsilon > 0$. This handles the first term in (A.4).

Moving on, to see that

$$\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat{\theta}}(\sqrt{nm}(\phi(\hat{\varphi}^*)-\phi(\hat{\varphi}))\leq a)-\mathbb{P}_\theta(\Phi_\varphi v_\theta\leq a)\right|>\frac{\varepsilon}{2}\right)=o(1)$$

for all $\varepsilon > 0$, we first note that it is bounded from above by

$$\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a\left|\mathbb{P}_{\hat{\theta}}(\sqrt{nm}(\phi(\hat{\varphi}^*)-\phi(\hat{\varphi}))\leq a)-\mathbb{P}_\theta(\Phi_{\hat{\varphi}} v_\theta\leq a)\right|>\frac{\varepsilon}{4}\right)$$
$$+\sup_{\theta\in\Theta_0}\mathbb{P}_\theta\left(\sup_a|\mathbb{P}_\theta(\Phi_{\hat{\varphi}} v_\theta\leq a)-\mathbb{P}_\theta(\Phi_\varphi v_\theta\leq a)|>\frac{\varepsilon}{4}\right).$$

The first term converges to zero by an application of Lemma A.1 of Andrews (2005), using (A.3) and (A.5). The second term converges to zero by an application of the continuous mapping theorem. This completes the proof. □

Proof of Theorem 3. In Theorem S.4 in the Supplement we show that

$$\sup_{\theta\in\Theta_1}|\mathbb{P}_\theta(\hat{w}\leq a)-\mathbb{P}_\theta(w_\theta\leq a)|=o(1)$$

where the random variable w_θ has a non-central χ^2-distribution with $\dim \phi$ degrees of freedom and non-centrality parameter

$$\gamma\, \beta_\theta' \Phi_\varphi' (\Phi_\varphi \Sigma_\theta \Phi_\varphi')^{-1} \Phi_\varphi \beta_\theta.$$

With this result in hand we may proceed in exactly the same way as in the proof of Theorem 1 to establish the result. □

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. (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.

Fernández-Val, I. (2022). Datasets for Fernández-Val (2009), Journal of Econometrics 150(1), pp. 71-85. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-06-10. https://doi.org/10.3886/E172521V1.

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.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

PSID. Panel Study of Income Dynamics, public use dataset. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann

Arbor, MI.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Wang, W. and M. Kaffo (2016). Bootstrap inference for instrumental variable models with many weak instruments. *Journal of Econometrics 192*, 231–268.