

k -step Bootstrap Bias Correction for Fixed Effects Estimators in Nonlinear Panel Models

Min Seong Kim and Yixiao Sun*

Department of Economics, UC San Diego

April 2009

Abstract

Fixed effects estimators in nonlinear panel models with fixed T usually suffer from inconsistency because of the incidental parameters problem first noted by Neyman and Scott (1948). Moreover, even though T grows at the same rate as n , they are asymptotically biased and therefore the associated confidence interval has a large coverage error. This paper proposes a k -step parametric bootstrap bias corrected estimator. We prove that our estimator is asymptotically normal and is centered at the true parameter if T grows faster than $\sqrt[3]{n}$. In addition to bias correction, we construct a confidence interval with a double bootstrap procedure, and Monte Carlo experiments confirm that the error in coverage probability of our CI's is smaller than those of the alternatives. We also propose bias correction for average marginal effects.

Keywords: Bias Correction, Fixed-Effect Estimator, Incidental Parameters Problem, k -step Bootstrap, Nonlinear Panel Data Models.

JEL Classification Numbers: C13, C33

*Email: msk003@ucsd.edu and yisun@ucsd.edu. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. We thank Graham Elliott Dimitris Politis and Hal White for helpful comments.

1 Introduction

Panel data consists of repeated observations from different individuals across time. One virtue of this data structure is that we can control for unobserved time-invariant individual heterogeneity in an econometric model. When individual effects are correlated with explanatory variables, we may use the fixed effects estimator which treats each unobserved individual effect as a parameter to be estimated. However, this approach usually suffers from inconsistency when the time series sample size (T) is short. This is known as the *incidental parameters problem*, first noted by Neyman and Scott (1948). Furthermore, even though T grows at the same rate as n , the fixed effects estimators are asymptotically biased so that the inference drawn from them may give misleading results.

This paper proposes a k -step parametric bootstrap bias corrected maximum likelihood (ML) estimator of nonlinear static panel models. In the k -step bootstrap procedure, we approximate the standard bootstrap estimator by taking k -steps of a Newton-Raphson (NR) iterative scheme. We employ the original estimate as the starting point for the NR steps. We estimate the asymptotic bias using the k -step bootstrap method and subtract this from the original biased estimator. We prove that the standard and k -step bootstrap bias corrected estimators are asymptotically normal and centered at the true parameter if T grows faster than $\sqrt[3]{n}$. This condition is important in practice because many economic data sets nowadays are composed of small T and large n and therefore the usefulness of the bias corrected estimation particularly depends on how much of the bias is corrected in small T . Our Monte Carlo experiments show that in finite samples, the k -step bootstrap bias corrected estimators reduce the bias remarkably even for small T . This bias correction does not increase the asymptotic variance and thus bias correction substantially improves statistical inference. In addition to bias correcting the parameter estimators, we also apply the k -step bootstrap bias correction to the average marginal effect estimation.

The substantial advantage of our approach over alternatives is that our method enables us not only to correct the asymptotic bias but also to improve the coverage accuracy of the associated confidence intervals (CI). We construct the CI's using a double k -step bootstrap procedure. In Monte Carlo experiments we find that in finite samples the error in coverage probability of our CI's is smaller than those of the other standard alternatives especially when T is small. This is true for the estimators of the model parameters as well as the estimators of the average marginal effects.

Another clear advantage is ease of computation. Standard bootstrap methods in nonlinear models are usually very time-intensive because it is required to solve R nonlinear optimization problems to obtain R bootstrap estimates. R usually needs to be fairly large for the bootstrap method to be reliable. Unless the optimization problem is simple, this would be a very time-intensive task. Particularly, as the fixed effects approach treats the individual effects as parameters, there are many parameters to be estimated and computational intensiveness can be particularly serious in this type of models. For example, in our empirical application (not reported here), there are 1461 individuals, which means there are more than 1461 parameters to be estimated. In addition, the double bootstrap procedure which is used for constructing CI's in this method also increases computational intensiveness substantially. In order to overcome this problem, we introduce the k -step bootstrap estimation which only involves computing the Hessian and the score functions. We show that when $n \rightarrow \infty$ the stochastic difference between the standard and k -step bootstrap estimators is $O_p(T^{-2^{k-1}})$. When $k \geq 2$, this difference is of smaller order than the bias

term we intend to remove. As a result, we can use the k -step bootstrap in place of the standard bootstrap to achieve bias reduction.

Several papers have discussed the difficulties involved in controlling for this incidental parameters problem in nonlinear panel models and have suggested bias correction methods. Lancaster (2000) and Arellano and Hahn (2006) give an overview on the subject. Anderson (1970) and Honoré and Kyriazidou (2000) propose estimators which do not depend on individual effects in some specific cases. However, their approaches are the exception rather than the rule and so usually provide no guidance in general cases to eliminate the bias from nonlinear panel models. More generally, Hahn and Newey(2004) (denoted HN hereinafter) and Fernández-Val (2009) propose jackknife and analytic procedures for nonlinear static models, while Hahn and Kuersteiner (2004) propose analytic estimators in nonlinear dynamic models. Both expand the estimator in orders of T and estimate the leading bias term using the sample analogue. Bester and Hansen (2008) propose a penalized objective function approach to solve this problem.

There is also a large literature on bootstrap bias corrected estimation. Hall (1992) introduces general bootstrap algorithms for bias correction and for the construction of CI's which we adapt in this paper. Hahn, Kuersteiner and Newey (2004) analyze the asymptotic properties of a bootstrap bias corrected ML estimator in cross sectional data and show that it is higher order efficient. Pace and Salvan (2006) suggest a bootstrap bias corrected estimator when there are nuisance parameters, but their algorithm is different from ours. While we estimate the asymptotic bias of the fixed effects estimator directly by bootstrap, they use the bootstrap procedure to adjust the profile likelihood function from which they obtain their bias corrected estimator. The k -step bootstrap procedure first appears in Davidson and Mackinnon (1999) and Andrews (2002, 2005), in which they prove its higher-order equivalence to the standard bootstrap for extremum estimators.

The paper is organized as follows. Section 2 reviews the incidental parameters problem in nonlinear panel models and introduces the analytic form of the bias term which is demonstrated in HN. Section 3 describes the bootstrap bias correction procedure. Section 4 explains the k -step bootstrap bias correction. Section 5 establishes the asymptotic properties of our estimators. Section 6 discusses bias correction for average marginal effects. The Monte Carlo simulation results are reported in Section 7. The last section concludes.

2 Incidental Parameters Problem

In this section, we introduce the incidental parameters problem and present the asymptotic bias of the fixed effects estimator in nonlinear panel models.

Consider a nonlinear panel data model:

$$z_{it} \sim f(z; \theta, \alpha_i)$$

where θ is ($L_\theta \times 1$) vector of parameters of interest and α_i is a scalar individual heterogeneity and f is a probability density function with parameters θ and α_i . For any given parameter value, $\{z_{it}\}$ are independently distributed across $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$. The model includes discrete choice models and censored and truncation models as special cases.

Denote the true values of θ and $\alpha \equiv (\alpha_1, \dots, \alpha_i, \dots, \alpha_N)$ by θ_0 and $\alpha_0 = (\alpha_{10}, \dots, \alpha_{i0}, \dots, \alpha_{N0})$ respectively and let $l(\theta, \alpha_i; z_{it}) \equiv \log f(z_{it}; \theta, \alpha_i)$. The objective function for the fixed effects estimator, $\hat{\theta}_{nT}$, is the concentrated log-likelihood function based on the preliminary

estimator $\hat{\alpha}_i$. That is, we obtain $\hat{\theta}_{nT}$ by solving

$$\hat{\theta}_{nT} = \arg \max_{\theta} \sum_{i=1}^n \sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); z_{it}), \quad (1)$$

where

$$\begin{aligned} \hat{\alpha}_i(\theta) &= \arg \max_{\alpha_i} \sum_{t=1}^T l(\theta, \alpha_i; z_{it}) \\ \hat{\alpha}(\theta) &\equiv (\hat{\alpha}_1(\theta), \dots, \hat{\alpha}_i(\theta), \dots, \hat{\alpha}_N(\theta)) \end{aligned} \quad (2)$$

and the maximization is taken over a compact set.

Equation (2) implies that the estimation of α_i uses only T time series observations (z_{i1}, \dots, z_{iT}) . Therefore, given that T is fixed, $\hat{\alpha}_i$ does not converge to α_{i0} even though $n \rightarrow \infty$. This estimation error of $\hat{\alpha}_i$ causes $\hat{\theta}_{nT}$ to be inconsistent, which means $\text{Plim}_{n \rightarrow \infty} \hat{\theta}_{nT} \neq \theta_0$. This is known as the incidental parameters problem first noted by Neyman and Scott (1948). From the asymptotic properties of extremum estimators (e.g. Amemiya (1985)), as $n \rightarrow \infty$ with T fixed

$$\hat{\theta}_{nT} \xrightarrow{p} \theta_T, \quad \theta_T \equiv \arg \max_{\theta} \bar{E} \left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); z_{it}) \right] \quad (3)$$

where

$$\bar{E} \left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); z_{it}) \right] \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); z_{it}) \right].$$

Since θ_0 maximizes $E[\sum_{i=1}^n \sum_{t=1}^T l(\theta, \alpha_{i0}; z_{it})]$, usually $\theta_T \neq \theta_0$. If the likelihood function is smooth enough, we can show by stochastic expansion that

$$\theta_T = \theta_0 + \frac{B}{T} + O\left(\frac{1}{T^2}\right) \quad (4)$$

for some B . This implies that $\theta_T \rightarrow \theta_0$ as $T \rightarrow \infty$. However, it is still asymptotically biased if T grows at the same rate as n . That is, as $n, T \rightarrow \infty$ and $n/T \rightarrow \rho$,

$$\begin{aligned} \sqrt{nT}(\hat{\theta}_{nT} - \theta_0) &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + \sqrt{nT} \frac{B}{T} + O_p\left(\sqrt{\frac{n}{T^3}}\right) \\ &\xrightarrow{d} N(0, \Omega) + B\sqrt{\rho} \stackrel{d}{=} N(B\sqrt{\rho}, \Omega), \end{aligned} \quad (5)$$

for some variance matrix Ω . $\sqrt{nT}(\hat{\theta}_{nT} - \theta_T)$ converges to a normal distribution centered at zero, since θ_T is the probability limit of $\hat{\theta}_{nT}$. However, the second term, $\sqrt{nT}B/T$, does not vanish but converges to $B\sqrt{\rho}$. Hence, statistical inference drawn from this will result in misleading conclusions even when T is as large as n .

HN establish the analytic form of the leading bias of $\hat{\theta}_{nT}$ using stochastic expansion. For notational convenience, we define

$$u_{it}(\theta, \alpha_i) \equiv \frac{\partial}{\partial \theta} l(\theta, \alpha_i; z_{it}) \quad \text{and} \quad v_{it}(\theta, \alpha_i) \equiv \frac{\partial}{\partial \alpha_i} l(\theta, \alpha_i; z_{it})$$

and let additional subscripts denote partial derivatives, i.e. $v_{it\alpha}(\theta, \alpha_i) \equiv \frac{\partial^2}{\partial \alpha_i^2} l(\theta, \alpha_i; z_{it})$. We suppress the arguments of the functions such as u_{it} , when they are evaluated at the true value (θ_0, α_{i0}) . HN show that in equation (5)

$$B = -H^{-1}(\theta_0, \alpha_0)b(\theta_0, \alpha_0), \quad (6)$$

where

$$H(\theta_0, \alpha_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[U_{it}U'_{it}],$$

$$b(\theta_0, \alpha_0) = \frac{1}{2} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{E[V_{2it}U_{it}]}{E[v_{it}^2]},$$

and

$$U_{it} \equiv u_{it} - \frac{E[u_{it}v_{it}]}{E[v_{it}^2]} \cdot v_{it}, \quad V_{2it} \equiv v_{it}^2 + v_{it\alpha}.$$

3 Bootstrap Bias Correction

In this section, we provide the bias corrected estimator using a parametric bootstrap procedure. The parametric bootstrap is different from nonparametric bootstrap in that the former utilizes the parametric structure of the DGP by replacing the original parameters with their estimators to generate bootstrap samples, while the latter generates them from the empirical distribution function.

Let $F^* \equiv F_{\hat{\theta}_{nT}, \hat{\alpha}}$ denote the distribution function of bootstrap samples. We obtain F^* from F by replacing θ_0 and α_0 , with $\hat{\theta}_{nT}$ and $\hat{\alpha} \equiv \hat{\alpha}(\hat{\theta}_{nT})$. Therefore, in the bootstrap world, $\hat{\theta}_{nT}$ and $\hat{\alpha}$ are the true parameters. Let $\{z_{it}^*\}$ denote the bootstrap sample drawn at random from F^* . Based on $\{z_{it}^*\}$, we can obtain the bootstrap estimators, $\hat{\theta}_{nT}^*$ and $\hat{\alpha}_i^*$ by ML estimation. That is,

$$\hat{\theta}_{nT}^* = \arg \max_{\theta} \sum_{i=1}^n \sum_{t=1}^T l(\theta, \hat{\alpha}_i^*(\theta); z_{it}^*), \quad (7)$$

where

$$\hat{\alpha}_i^*(\theta) = \arg \max_{\alpha_i} \sum_{t=1}^T l(\theta, \alpha_i; z_{it}^*), \quad (8)$$

and as before the maximization is taken over a compact set.

The intuition behind the bootstrap bias correction is that the bias of a bootstrap estimator is a good approximation to that of a true parameter estimator. Under some regularity conditions, as $n \rightarrow \infty$ and $T \rightarrow \infty$,

$$E^*(\hat{\theta}_{nT}^*) - \hat{\theta}_{nT} = \frac{B}{T} + O_p\left(\frac{1}{T^2}\right)$$

where¹

$$E^*(\hat{\theta}_{nT}^*) = \lim_{R \rightarrow \infty} \frac{1}{R} \sum_{r=1}^R \hat{\theta}_{nT}^{*(r)},$$

and E^* is the expectation operator with respect to F^* . Therefore, the bootstrap bias corrected estimator can be defined as

$$\tilde{\theta}_{nT} = 2\hat{\theta}_{nT} - E^*(\hat{\theta}_{nT}^*). \quad (9)$$

The above estimator reduces the order of the magnitude of a bias from $O_p(T^{-1})$ to $O_p(T^{-2})$. To show this, we employ the same definitions of HN in the bootstrap world, i.e.

$$u_{it}^* \equiv \frac{\partial}{\partial \theta} l(\hat{\theta}_{nT}, \hat{\alpha}_i; z_{it}^*), v_{it}^* \equiv \frac{\partial}{\partial \alpha_i} l(\hat{\theta}_{nT}, \hat{\alpha}_i; z_{it}^*) \text{ and } v_{it\alpha}^* \equiv \frac{\partial^2}{\partial \alpha_i^2} l(\hat{\theta}_{nT}, \hat{\alpha}_i; z_{it}^*).$$

Then,

$$P_{n \rightarrow \infty}^* \lim \left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT} - \frac{B^*}{T} \right) = O_p \left(\frac{1}{T^2} \right), \quad (10)$$

and

$$B^* = - \left[H^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) \right]^{-1} b^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))$$

where

$$H^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E^* [U_{it}^* U_{it}^{*'}],$$

$$b^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) = \frac{1}{2} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{E^* [U_{it}^* V_{2it}^*]}{E^* [(v_{it}^*)^2]}$$

and

$$U_{it}^* \equiv u_{it}^* - \frac{E^* [u_{it}^* v_{it}^*]}{E^* [(v_{it}^*)^2]} \cdot v_{it}^*, \quad V_{2it}^* \equiv v_{it}^{*2} + v_{it\alpha}^*.$$

The conditional distribution of the bootstrap sample given the data or $(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))$ is the same as the distribution of the original sample except that the former uses $(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))$ rather than (θ_0, α_0) as true parameters². Therefore, we have

$$H^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) = H(\theta, \alpha)|_{(\theta, \alpha) = (\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))}$$

$$b^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) = b(\theta, \alpha)|_{(\theta, \alpha) = (\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))}$$

¹Note that $E^*(\hat{\theta}_{nT}^*)$ may not be finite, in which case we can introduce truncation to prevent it from going to infinity. See equation (17) for such a modification. When the truncation threshold goes to infinity at an appropriate rate, the introduction of truncation does not affect the limiting distribution. For simplicity of exposition, we do not explicitly incorporate truncation here but do so for the k -step bootstrap estimator.

²When $(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))$ is regarded as a vector, it is understood to be $(\hat{\theta}'_{nT}, \hat{\alpha}'(\hat{\theta}_{nT}))'$. For notational simplicity, we omit the transpose notation if confusion is unlikely.

By stochastic expansion, we show that in Appendix I that

$$\begin{aligned} H(\theta, \alpha)|_{(\theta, \alpha) = (\hat{\theta}_{nT}, \hat{\alpha}_{nT})} &= H(\theta_0, \alpha_0) \left[1 + O_p\left(\frac{1}{T}\right) \right], \\ b(\theta, \alpha)|_{(\theta, \alpha) = (\hat{\theta}_{nT}, \hat{\alpha}_{nT})} &= b(\theta_0, \alpha_0) \left[1 + O_p\left(\frac{1}{T}\right) \right]. \end{aligned}$$

Combining the above two equations with the definition of B^* , we obtain

$$B^* = B + O_p\left(\frac{1}{T}\right). \quad (11)$$

As a final step, we show that under the assumptions given in Section 5

$$\sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - E^*(\hat{\theta}_{nT}^*) \right) = o_p(1). \quad (12)$$

From (10), (11) and (12), for $T/\sqrt[3]{n} \rightarrow \infty$,

$$\begin{aligned} \sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + \sqrt{nT} \left[(\theta_T - \theta_0) - (P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT}) \right] \\ &\quad + \sqrt{nT} \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - E^*(\hat{\theta}_{nT}^*) \right] \\ &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + o_p(1) \xrightarrow{d} N(0, \Omega). \end{aligned} \quad (13)$$

This implies that the bootstrap bias corrected estimator removes the dominant bias and is asymptotically unbiased.

4 k -step Bootstrap Bias Correction

In this section, we define the k -step bootstrap bias corrected estimator and demonstrate its higher order equivalence to the standard bootstrap estimator.

The k -step procedure approximates $\hat{\theta}_{nT}^*$ by the NR iterative procedure. Let $\hat{\theta}_{nT,k}^*$ and $\hat{\alpha}_{i,k}^*$ denote the k -step bootstrap estimator. We define $\hat{\theta}_{nT,k}^*$ and $\hat{\alpha}_k^*$ recursively in the following way:

$$\begin{pmatrix} \hat{\theta}_{nT,k}^* \\ \hat{\alpha}_k^* \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{nT,k-1}^* \\ \hat{\alpha}_{k-1}^* \end{pmatrix} - H_{k-1}^{-1} S_{k-1} \quad (14)$$

where³

$$H_{k-1} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{\partial^2 \log l(\theta, \alpha_i; z_{it}^*)}{\partial(\theta', \alpha') \partial(\theta', \alpha')'} \Big|_{\theta = \hat{\theta}_{nT,k-1}^*, \alpha = \hat{\alpha}_{k-1}^*} \quad (15)$$

$$S_{k-1} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{\partial \log l(\theta, \alpha_i; z_{it}^*)}{\partial(\theta', \alpha')'} \Big|_{\theta = \hat{\theta}_{nT,k-1}^*, \alpha = \hat{\alpha}_{k-1}^*}$$

and the start-up estimator $\hat{\theta}_{nT,0}^* = \hat{\theta}_{nT}$, $\hat{\alpha}_0^* = \hat{\alpha}$.

³The Hessian matrix we used is called the observed Hessian. We note that some terms in $\partial^2 \log l(\theta, \alpha_i; z_{it}^*) / \partial(\theta', \alpha') \partial(\theta', \alpha')'$ have zero expectation. Dropping these terms in equation (15), we obtain the expected Hessian. Our asymptotic results remain valid for the expected Hessian, as the dropped terms are of smaller order.

In the appendix of proofs, we show that

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* = P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* + O_p \left(\frac{1}{T^{2k-1}} \right). \quad (16)$$

This implies the quadratic convergence of $\hat{\theta}_{nT,k}^*$ to $\hat{\theta}_{nT}^*$ as k increases. In particular, when $k \geq 2$, $P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* = P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* + O_p(1/T^2)$. So in large samples, the approximation error in using the k -step bootstrap instead of the standard bootstrap is of smaller order than the bias term that we intend to remove. Therefore, condition $k \geq 2$ is necessary for the k -step bootstrap to achieve effective bias reduction.

To implement the k -step bootstrap, we have to invert the Hessian matrix. Depending on the observations we sample, H_{j-1} may be close to be singular in practice, in which case $\hat{\theta}_{nT,j}^*$ goes to infinity. As a result, the mean of $\hat{\theta}_{nT,k}^*$ may not be finite. To circumvent the undue influence of the second derivative of the objective function on our estimator, we introduce the truncated version, $\check{\theta}_{nT,k}^*$. The truncated estimator is defined as

$$\check{\theta}_{nT,k}^* \equiv \hat{\theta}_{nT} + \left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right) 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| \leq M_{nT} \right). \quad (17)$$

$\check{\theta}_{nT,k}^*$ yields the same value as $\hat{\theta}_{nT,k}^*$ when the difference of $\hat{\theta}_{nT,k}^*$ from $\hat{\theta}_{nT}$ is bounded by M_{nT}/\sqrt{nT} , but does not blow up when it has an infinite value. Similarly, we define

$$\check{\alpha}_{i,k}^* \equiv \hat{\alpha}_i + \left(\hat{\alpha}_{i,k}^* - \hat{\alpha}_i \right) 1 \left(\sqrt{nT} \left| \hat{\alpha}_{i,k}^* - \hat{\alpha}_i \right| \leq M_{nT} \right). \quad (18)$$

We can set M_{nT} large enough that this truncation does not affect the asymptotic properties. We show that when $M_{nT} \rightarrow \infty$ such that $\sqrt{n/T} = o(M_{nT})$, we have:

$$P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* = P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* + O_p \left(\frac{1}{T^2} \right).$$

As a final step, we show that

$$\sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^*(\check{\theta}_{nT,k}^*) \right) = o_p(1). \quad (19)$$

From (16) and (19) the limiting distribution of the bootstrap bias corrected estimator, $\tilde{\theta}_{nT}$, which is defined in (9), will be invariant even though we replace $\hat{\theta}_{nT}^*$ with $\check{\theta}_{nT,k}^*$. Hence, we can define our truncated k -step bootstrap biased corrected estimator as

$$\tilde{\theta}_{nT,k} \equiv 2\hat{\theta}_{nT} - E^*(\check{\theta}_{nT,k}^*). \quad (20)$$

Then for $T/\sqrt[3]{n} \rightarrow \infty$ and all $k \geq 2$,

$$\begin{aligned} \sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + \sqrt{nT} \left[(\theta_T - \theta_0) - (P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - \hat{\theta}_{nT}) \right] \\ &\quad + \sqrt{nT} \left(E^*(\check{\theta}_{nT,k}^*) - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* \right) \\ &= \sqrt{nT}(\hat{\theta}_{nT} - \theta_T) + o_p(1) \xrightarrow{d} N(0, \Omega). \end{aligned} \quad (21)$$

5 Asymptotic Properties

In this section, we state the assumptions and rigorously establish the asymptotic properties of the standard bootstrap and k -step bootstrap estimators.

For easy of exposition, we write $l(\theta, \alpha; z_{it}) \equiv l(\theta, \alpha_i; z_{it})$ so that $l(\cdot, \cdot; z_{it})$ is regarded as a function of θ and α . We maintain the following assumptions:

Assumption 1 $n, T \rightarrow \infty$ such that $n = o(T^3)$ and $T = O(n)$.

Assumption 2 (i) $l(\theta, \alpha; z_{it})$ is continuous in $(\theta, \alpha) \in \Theta$; (ii) the parameter space Θ is compact; (iii) (θ_0, α_0) is an interior point in Θ .

Assumption 3 For each $\eta > 0$, there exists $\delta > 0$ such that

$$\inf_i \left[G_i(\theta_0, \alpha_0) - \sup_{\{(\theta, \alpha): |(\theta, \alpha) - (\theta_0, \alpha_0)| > \eta\}} G_i(\theta, \alpha) \right] \geq \delta > 0,$$

where

$$G_i(\theta, \alpha) \equiv E \frac{1}{T} \sum_{t=1}^T l(\theta, \alpha; z_{it}).$$

Assumption 4 (i) $l(\theta, \alpha; z_{it})$ is continuously differentiable to six orders; (ii) there exists some $M(z_{it})$ such that

$$\left| \frac{\partial^{m_1+m_2} l(\theta, \alpha; z_{it})}{\partial \theta_j^{m_1} \partial \alpha_i^{m_2}} \right| \leq M(z_{it}), \quad 0 \leq m_1 + m_2 \leq 6$$

(iii) For some $Q > 64$, $E [M(z_{it})^Q] < C$ for a constant C and all $i = 1, 2, \dots, N$.

Assumption 5 (i) $\bar{E}[\mathcal{I}_i] \equiv \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E[U_{it}U'_{it}]$ exists and is positive definite (ii) $\min_i E[v_{it}^2] > 0$.

Assumption 6 Let $B_n = (n^{-1} \sum_{i=1}^n E[U_{it}U'_{it}])^{-1} \left(n^{-1} \sum_{i=1}^n \frac{E[V_{2it}U_{it}]}{E[v_{it}^2]} \right)$, then $B_n = B + O(1/\sqrt{n})$.

Assumption 1 shows that our estimator is applicable as long as T grows faster than $\sqrt[3]{n}$. This implies that our asymptotic theory is valid with relatively small T and large n , which is often the case in micro panel data sets. Assumption 2 is a standard regularity assumption. Assumption 3 is the identification assumption for extremum estimators. Assumption 4 is the same as Condition 4 in Newey and Hahn (2004). It is stronger than the moment assumption for extremum estimators and under this assumption the asymptotic bias depends on the second order expansion and higher order terms go to 0 under Assumption 1. Assumption 5 allows us to invoke the central limit theorem. Assumption 6 ensures that the limiting bias term B is close to its finite sample analogue B_n . This assumption holds trivially if z_{it} are iid across i .

Theorem 1 Under Assumptions 1-6,

$$\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, (\bar{E}[\mathcal{I}_i])^{-1}).$$

For the proof see Appendix I.

Proposition 2 Under Assumptions 1-6, for all $k \geq 1$

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* = P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* + O_p \left(\frac{1}{T^{2^{k-1}}} \right).$$

For the proof see Appendix II.

Theorem 3 Under Assumptions 1-6, for all $k \geq 2$

$$\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) \xrightarrow{d} N(0, (\bar{E}[\mathcal{I}_i])^{-1}).$$

For the proof see Appendix III.

6 Bias Correction for Average Marginal Effects

In this section, we suggest bias corrected estimators of the average marginal effects using the k -step bootstrap procedure. In nonlinear models, the average marginal effect may be as interesting as the model parameters because it summarizes the effect over certain sub-population, which is often the quantity of interest in empirical studies.

The first average marginal effect, which we refer to as “the fixed effect average” or simply the average marginal effect, is the marginal effect averaged over α_i . It is defined as

$$\mu(w) = \frac{1}{n} \sum_{i=1}^n m(w, \theta_0, \alpha_{i0})$$

where w is the value of the covariate vector where the average effect is desired. For example, in a probit model, $m(w, \theta_0, \alpha_{i0}) = \theta_{0(j)} \phi(x' \theta_0 + \alpha_{i0})$ where $\theta_{0(j)}$ and $\phi(\cdot)$ are the coefficient on the j -th regressor of interest and the standard normal density function respectively.

The bias uncorrected estimator of $\mu(w)$ is

$$\hat{\mu}_{nT}(w) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w, \hat{\theta}_{nT}, \hat{\alpha}_i). \quad (22)$$

As in the case for the estimation of model parameters, we can construct a k -step bootstrap bias corrected estimator of the fixed effect average by estimating the bias with the difference between $\hat{\mu}_{nT}(w)$ and its bootstrap estimator. Our k -step bootstrap bias corrected estimator of the fixed effects average is

$$\tilde{\mu}_{nT,k}(w) = \frac{2}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w, \hat{\theta}_{nT}, \hat{\alpha}_i) - E^* \left[\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w, \check{\theta}_{nT,k}^*, \check{\alpha}_{i,k}^*) \right]. \quad (23)$$

The second average marginal effect, which we refer to as “the overall average marginal effect”, is the marginal effect averaged over both α_i and the covariates. It is defined as

$$\nu = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w_{it}, \theta_0, \alpha_{i0}).$$

See also Fernández-Val (2009). Similarly to equations (22) and (23), the original and bias corrected estimators of ν are

$$\begin{aligned} \hat{\nu}_{nT} &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w_{it}, \hat{\theta}_{nT}, \hat{\alpha}_i), \\ \tilde{\nu}_{nT,k}(w) &= \frac{2}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w_{it}, \hat{\theta}_{nT}, \hat{\alpha}_i) - E^* \left[\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T m(w_{it}, \check{\theta}_{nT,k}^*, \check{\alpha}_{i,k}^*) \right]. \end{aligned}$$

7 Monte Carlo Study

In this section, we report our Monte Carlo experiment results, which show that k -step bootstrap bias correction reduces the bias significantly in finite samples and also improves the coverage accuracy of CIs.

For our Monte Carlo experiment, we employ the design used in Heckman (1981), Greene (2004), HN, and Fernández-Val(2009). It is based on the following probit model:

$$\begin{aligned} Y_{it} &= 1\{X_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\}; & \epsilon_{it} &\sim N(0, 1), \quad \alpha_i \sim N(0, 1), \\ X_{it} &= t/10 + X_{i,t-1}/2 + u_{it}; & u_{it} &\sim U(-1/2, 1/2), \\ n &= 100; \quad T = 4, 8, 12; \quad \theta_0 = 1. \end{aligned}$$

As discussed in HN, this model does not fit completely within our framework. First, X_{it} is correlated overtime. The correlation does not cause any problem as we can use the conditional MLE approach and all the asymptotic results remain valid. Second, there is no correlation between X_{it} and α_i . This is different from the usual condition under which the fixed effects estimator is used. However the incidental parameters problem is still present as it has nothing to do with whether there is a correlation between X_{it} and α_i . The bias of the fixed effects estimator can be severe for fixed effects models as well as for random effects models. The effectiveness of different bias reduction methods can be well evaluated with our data generating process. Another reason to use this design is that it is widely cited and used in other simulation studies, which helps us compare our estimator with the alternatives.

The uncorrected estimator of model parameters is

$$(\hat{\theta}_{nT}, \hat{\alpha}) = \arg \max_{\theta, \alpha} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T [y_{it} \log \Phi(x_{it}\theta + \alpha_i) + (1 - y_{it}) \log(1 - \Phi(x_{it}\theta + \alpha_i))]$$

and the estimators of the average marginal effects are

$$\hat{\mu}_{nT}(\bar{x}) = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{nT} \phi(\bar{x} \hat{\theta}_{nT} + \hat{\alpha}_i), \quad \hat{\nu}_{nT} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{\theta}_{nT} \phi(x_{it} \hat{\theta}_{nT} + \hat{\alpha}_i).$$

where $\Phi(\cdot)$ is the standard normal distribution function and \bar{x} is the sample mean of $\{x_{it}, i = 1, 2, \dots, N, t = 1, \dots, T\}$.

For the k -step bootstrap, we generate bootstrap samples based on $\hat{\theta}_{nT}$ and $\{\hat{\alpha}_i\}_{i=1}^n$ and estimate $\hat{\theta}_{nT,k}^*$ using (14) with the bootstrap samples. We repeat this procedure 1000 times ($R = 1000$). Then, we obtain the bias corrected k -step bootstrap estimator from (20). As discussed before, for each k value, we can use either observed Hessian or expected Hessian in the NR step, leading to two versions of the k -step procedure. Each simulation is repeated 1000 times.

We compare the performance of our bias-corrected estimator with four alternative bias correction estimators: the jackknife and the analytic bias corrected estimators by Hahn and Newey (2004) and the analytical bias-corrected estimator by Fernández-Val (2009). The jackknife bias-corrected estimator is denoted ‘Jackknife’. For HN analytic estimators, there are two versions: the analytic bias-corrected estimator using Bartlett equalities, denoted ‘BC1’; the analytic bias-corrected estimator based on general estimating equations, denoted ‘BC2’. Fernández-Val’s estimator is denoted as ‘BC3’.

For each estimator, we report its mean, median, standard deviation, root mean squared errors, and the empirical sizes of two-sided nominal 5% and 10% tests. The tests are based on symmetric CI’s, that is, we reject the null hypothesis if the parameter value under the null falls outside the CI’s. For the jackknife and analytical bias correction procedures, the interval estimator or the testing method are the same as that given in the respective papers. For the k -step procedure, the CI’s are based the double bootstrap procedure.

To describe the double bootstrap procedure, we focus on an element of θ . Hence, without loss of generality, we can consider the case that θ_0 is a scalar. By iterating the bootstrap procedure, we define:

$$\tilde{\theta}_{nT,k}^* \equiv 2\check{\theta}_{nT,k}^* - E^{**}(\check{\theta}_{nT,k}^{**}),$$

where $E^{**}(\check{\theta}_{nT,k}^{**})$ is defined on the double bootstrap, that is, the k -step bootstrap using $(\check{\theta}_{nT,k}^*, \check{\alpha}_{i,k}^*)$ as the true model parameters. Similarly

$$\tilde{\alpha}_{i,k}^* \equiv 2\check{\alpha}_{i,k}^* - E^{**}(\check{\alpha}_{i,k}^{**}).$$

Let

$$t\text{-stat} = \frac{\sqrt{nT} |\tilde{\theta}_{nT,k} - \theta_0|}{SE(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k^*)}$$

be the t-statistic for θ_0 where

$$\left[SE(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k^*) \right]^2 = \left(-\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T U_{it\theta}(\tilde{\theta}_{nT,k}, \tilde{\alpha}_{i,k}^*) \right)^{-1}, \quad \tilde{\alpha}_{i,k}^* \equiv 2\hat{\alpha}_i - E^*(\check{\alpha}_{i,k}^*).$$

Let

$$t^*\text{-stat} = \frac{\sqrt{nT} |\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}|}{SE(\hat{\theta}_{nT,k}^*, \hat{\alpha}_k^*)}$$

be the corresponding t-statistic in the bootstrap world where

$$\left[SE(\hat{\theta}_{nT,k}^*, \hat{\alpha}_k^*) \right]^2 = \left(-\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T U_{it\theta}^*(\hat{\theta}_{nT,k}^*, \hat{\alpha}_{i,k}^*) \right)^{-1}, \quad \hat{\alpha}_{i,k}^* \equiv 2\check{\alpha}_{i,k}^* - E^{**}(\check{\alpha}_{i,k}^{**}).$$

Then the bootstrap CI is :

$$\left[\tilde{\theta}_{nT,k} - T_{1-\alpha/2}^* \frac{1}{\sqrt{nT}} SE(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k), \quad \tilde{\theta}_{nT,k} + T_{1-\alpha/2}^* \frac{1}{\sqrt{nT}} SE(\tilde{\theta}_{nT,k}, \tilde{\alpha}_k) \right]$$

where $T_{1-\alpha/2}^*$ is the $(1 - \alpha/2) \times 100\%$ percentile of t^* -stat. Our double bootstrap two-sided test is based on the above CI. We can use the same procedure to construct CIs for the average marginal effect and the overall average marginal effect. In our simulation experiment, we set the number of double bootstrap samples to be 100. We do so in order to reduce the computational burden. In empirical applications, we should use a larger number.

Table 1 shows the performance of the k -step bootstrap for different values of k . According to this result, the k -step bootstrap procedure reduces the bias significantly when $k \geq 2$. Results not reported here show that the one-step procedure is not effective in bias reduction. This result is consistent with our Theorem 2, which demonstrates the order of bias is reduced from $O_p(1/T)$ to $O_p(1/T^2)$ when $k \geq 2$. In terms of the MSE, the 2-step bootstrap with observed Hessian, the 3-step bootstrap with observed Hessian and the 3-step bootstrap with expected Hessian are efficient in general.

Table 2 compares different bias correction methods. We choose the 2-step bootstrap with observed Hessian as our benchmark. First, we see that the estimator without bias correction is severely biased when T is small. As T gets larger, the bias gets smaller, but there is still no improvement in the coverage accuracy of CI's. When $T = 4$, the bias of the uncorrected estimator is 42%. When $T = 12$, the bias is reduced to 13%. But the rejection probability is still 29% for the 5% two-sided test. Second, the k -step bootstrap performs better in finite samples than the other methods regardless of the size of T . In particular when $T = 4$, the outperforming of the k -step bootstrap procedure is remarkable, while as T increases other estimators become as accurate as ours. When $T = 4$ the bias of our estimator is 6% and RMSE is 0.249, while the bias of the jackknife method is 25% and its RMSE is 0.373. The analytic method by Fernández-Val (2009) also has the bias of 6% but its RMSE is 0.281 which implies that its variance is larger than ours. The k -step procedure achieves the smallest RMSE among all bias correction procedures. Third, in term of coverage accuracy, the CI's based on the double k -step bootstrap outperform other CI's in an overall sense.

Table 3 shows the ratio of the estimator of the average marginal effect to the true value. As HN and Fernández-Val (2009) show, the bias of the uncorrected estimator is negligible, even when $T = 4$. Its bias is less than 2% and in terms of RMSE, it performs as good as the bias corrected ones. However its CI's are not accurate especially when we have small T . When $T = 4$, its error in coverage probability for the 95% CI is about 5%. Inaccurate CI's are not just the problem of the bias uncorrected estimator. Jackknife and the analytic bias correction do not reduce the coverage error either. When $T = 4$, the errors in coverage probability for the 95% CI from jackknife and analytic estimators are 11% and 4-6% respectively. In contrast, the coverage error of the 95% CI constructed from the k -step double bootstrap is only 1.5%. We find that the our estimator improves the accuracy of the CI's which is not the case in the other standard alternatives.

Table 4 gives the Monte Carlo results for the ratio of the estimator of the overall average marginal effect to the true value. This is similar to the fixed effect average in Table 3 except that the average is taken over both the fixed effects and the covariate. As in the previous case, we find little evidence that bias correction is necessary in terms of RMSE. Actually,

the RMSE of the bias uncorrected estimator is smaller than that of the jackknife estimator in general. It also shows that in contrast to other estimators our double k -step bootstrap procedure improves the coverage accuracy of the CI's particularly when T is small.

8 Conclusion

In this paper, we propose the k -step bootstrap bias correction for the fixed effects estimator in nonlinear static panel models and establish the asymptotic properties of our bias corrected estimator. In simulation experiments, we show that the k -step bias correction procedure is often more effective than the alternatives. When T is small, the procedure achieves substantial bias reduction and has the smallest RMSE among the competing procedures. The confidence interval based on the double k -step bootstrap has a smaller coverage error than other CI's. This is true for both model parameters and average marginal effects. The asymptotic properties of our CIs and the possible higher order refinement are not studied here. It is an interesting topic for future research.

Table 1: Finite Sample Performances of Different k -step Bootstrap Estimators (cross section sample size $n = 100$ and the true value $\theta_0 = 1$)

Estimator	Mean	Median	SD	RMSE
$T = 4$				
$k=2, O$	0.94	0.94	0.242	0.249
$k=3, O$	0.84	0.84	0.192	0.253
$k=3, E$	0.83	0.84	0.194	0.256
$k=2, E$	1.02	1.01	0.270	0.271
$T = 8$				
$k=2, O$	0.98	0.98	0.114	0.115
$k=3, E$	0.97	0.97	0.113	0.117
$k=2, E$	1.02	1.02	0.122	0.123
$k=3, O$	0.96	0.96	0.118	0.124
$T = 12$				
$k=3, E$	0.99	0.99	0.078	0.078
$k=3, O$	0.97	0.97	0.077	0.080
$k=2, O$	0.99	0.99	0.081	0.082
$k=2, E$	1.01	1.00	0.087	0.087

Notes: We use “E” to indicate the use of the expected Hessian in the k -step bootstrap while we use “O” to indicate the use of the observed Hessian in the k -step bootstrap. The estimators are ordered according to their RMSE.

Table 2: Finite Sample Performance of Different Bias Corrected Estimators of θ (cross section sample size $n = 100$ and the true value $\theta_0 = 1$)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
$T = 4$						
Probit	1.42	1.40	0.385	0.30	0.40	0.569
2-step Bootstrap	0.94	0.94	0.242	0.04	0.06	0.249
Jackknife	0.75	0.75	0.277	0.11	0.19	0.373
BC1	1.11	1.10	0.304	0.04	0.11	0.323
BC2	1.20	1.19	0.333	0.09	0.16	0.388
BC3	1.06	1.06	0.275	0.02	0.06	0.281
$T = 8$						
Probit	1.18	1.18	0.132	0.28	0.39	0.238
2-step Bootstrap	0.98	0.98	0.114	0.04	0.10	0.115
Jackknife	0.95	0.96	0.118	0.05	0.11	0.128
BC1	1.05	1.05	0.134	0.05	0.11	0.143
BC2	1.05	1.05	0.132	0.05	0.10	0.141
BC3	1.02	1.02	0.124	0.03	0.07	0.126
$T = 12$						
Probit	1.13	1.12	0.090	0.29	0.40	0.161
2-step Bootstrap	0.99	0.99	0.081	0.05	0.11	0.082
Jackknife	0.98	0.98	0.080	0.05	0.10	0.083
BC1	1.04	1.04	0.087	0.07	0.13	0.096
BC2	1.03	1.03	0.085	0.06	0.11	0.090
BC3	1.01	1.01	0.082	0.04	0.09	0.083

Notes: Jackknife denotes HN Jackknife bias corrected estimator; BC1 denotes HN bias corrected estimator based on Bartlett equalities; BC2 denotes HN bias corrected estimator based on general estimating equations; BC3 denotes Fernández-Val (2009) bias corrected estimator which uses expected quantities in the estimation of the bias.

Table 3: Finite Sample Performance of Different Bias Corrected Estimators of the Average Marginal Effect μ (cross section sample size $n = 100$ and the true value $\mu_0 = 1$)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
$T = 4$						
Probit	0.98	0.98	0.256	0.097	0.157	0.256
2-step bootstrap	0.98	0.98	0.256	0.065	0.122	0.256
Jackknife	1.06	1.05	0.307	0.159	0.224	0.313
BC1	1.00	0.99	0.265	0.113	0.178	0.265
BC2	1.05	1.05	0.266	0.117	0.185	0.271
BC3	0.94	0.94	0.240	0.090	0.155	0.247
$T = 8$						
Probit	1.02	1.01	0.116	0.065	0.122	0.117
2-step bootstrap	1.00	1.00	0.113	0.043	0.096	0.117
Jackknife	1.00	0.99	0.130	0.086	0.153	0.130
BC1	1.02	1.02	0.133	0.090	0.153	0.134
BC2	1.02	1.02	0.131	0.087	0.154	0.133
BC3	1.00	1.00	0.117	0.058	0.107	0.117
$T = 12$						
Probit	1.02	1.01	0.072	0.072	0.119	0.074
2-step bootstrap	1.00	1.00	0.070	0.052	0.094	0.070
Jackknife	1.00	1.00	0.074	0.05	0.093	0.074
BC1	1.02	1.02	0.075	0.061	0.122	0.078
BC2	1.02	1.02	0.074	0.059	0.112	0.077
BC3	1.01	1.01	0.074	0.049	0.096	0.075

Notes: See notes to table 2.

Table 4: Finite Sample Performance of Different Bias Corrected Estimators of the Overall Average Marginal Effect ν (cross section sample size $n = 100$ and the true value $\nu_0 = 1$)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
<i>T = 4</i>						
Probit	0.99	1.00	0.248	0.11	0.17	0.249
2-step Bootstrap	0.99	0.98	0.248	0.04	0.09	0.248
Jackknife	1.02	1.02	0.285	0.12	0.19	0.286
BC1	1.00	1.00	0.261	0.12	0.18	0.261
BC2	1.04	1.04	0.255	0.12	0.19	0.258
BC3	0.94	0.94	0.226	0.08	0.13	0.234
<i>T = 8</i>						
Probit	0.99	0.99	0.100	0.07	0.13	0.100
2-step Bootstrap	0.99	0.99	0.100	0.04	0.08	0.100
Jackknife	1.01	1.01	0.107	0.07	0.14	0.107
BC1	1.01	1.01	0.110	0.09	0.15	0.110
BC2	1.00	1.00	0.105	0.07	0.13	0.105
BC3	0.97	0.97	0.103	0.08	0.13	0.107
<i>T = 12</i>						
Probit	0.99	0.99	0.063	0.09	0.16	0.065
2-step Bootstrap	0.99	0.99	0.064	0.06	0.12	0.065
Jackknife	1.00	1.00	0.064	0.05	0.11	0.064
BC1	1.00	1.00	0.065	0.06	0.11	0.065
BC2	0.99	0.99	0.062	0.05	0.10	0.063
BC3	0.98	0.98	0.062	0.05	0.11	0.065

Notes: See notes to table 2.

Appendix

I. Proof of Theorem 1

Throughout the proof, we assume that we have truncated $\hat{\theta}_{nT}^*$ so that $\hat{\theta}_{nT}^* - \hat{\theta}_{nT}$ is bounded in absolute value by M_{nT}/\sqrt{nT} . The technical details for showing that truncation has a negligible effect on the asymptotic properties of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$ are the same as those for $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$. We present the details for the latter in Appendix III and omit them here. Truncation allows us to convert probability orders into moment orders.

The bootstrap bias corrected estimator is defined as

$$\tilde{\theta}_{nT} = 2\hat{\theta}_{nT} - E^* \left(\hat{\theta}_{nT}^* \right).$$

Therefore,

$$\sqrt{nT} \left(\tilde{\theta}_{nT} - \theta_0 \right) = \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[E^* \left(\hat{\theta}_{nT}^* \right) - \hat{\theta}_{nT} \right] \right).$$

HN has shown that

$$\sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) \xrightarrow{d} N(0, \bar{E}[\mathcal{I}_i]^{-1}),$$

where

$$\mathcal{I}_i \equiv E[U_{it}U'_{it}].$$

Therefore, in order to prove Theorem 1, it suffices to show

$$\sqrt{nT} \left(\theta_T - \theta_0 - \left[E^* \left(\hat{\theta}_{nT}^* \right) - \hat{\theta}_{nT} \right] \right) = o_p(1). \quad (24)$$

The LHS of (24) can be decomposed into two parts :

$$\begin{aligned} & \sqrt{nT} \left(\theta_T - \theta_0 - \left[E^* \left(\hat{\theta}_{nT}^* \right) - \hat{\theta}_{nT} \right] \right) \\ &= \sqrt{nT} \left(\theta_T - \theta_0 - \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right] \right) + \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - E^* \left(\hat{\theta}_{nT}^* \right) \right). \end{aligned}$$

Hence, it suffices to show

$$\theta_T - \theta_0 - \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right] = O_p \left(\frac{1}{T^2} \right), \quad (25)$$

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - E^* \left(\hat{\theta}_{nT}^* \right) = o_p \left(\frac{1}{\sqrt{nT}} \right). \quad (26)$$

1. Prove $\theta_T - \theta_0 - \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right] = O_p \left(\frac{1}{T^2} \right)$

Let $F \equiv (F_1, \dots, F_n)$ and $\hat{F} \equiv (\hat{F}_1, \dots, \hat{F}_n)$, where $F_i := F_{\theta_0, \alpha_{i0}}$ is the distribution function of stratum i and \hat{F}_i is its empirical distribution function. Define $F(\epsilon) \equiv F + \epsilon\sqrt{T} \left(\hat{F} - F \right)$ for $\epsilon \in [0, 1/\sqrt{T}]$. For each fixed θ and ϵ , let $\alpha_i(\theta, F_i(\epsilon))$ and $\theta(F(\epsilon))$ be the solutions to the estimating equations

$$\begin{aligned} 0 &= \int V_i(\theta, \alpha_i(\theta, F_i(\epsilon))) dF_i(\epsilon), \\ 0 &= \sum_{i=1}^n \int U_i(\theta(F(\epsilon)), \alpha_i(\theta(F(\epsilon)), F_i(\epsilon))) dF_i(\epsilon). \end{aligned}$$

A Taylor series expansion gives

$$\hat{\theta}_{nT} - \theta_0 = \frac{1}{\sqrt{T}}\theta^\epsilon(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}) \quad (27)$$

where $\theta^\epsilon(\epsilon) \equiv d\theta(F(\epsilon))/d\epsilon$, $\theta^{\epsilon\epsilon}(\epsilon) \equiv d^2\theta(F(\epsilon))/d\epsilon^2, \dots$, and $\tilde{\epsilon}$ is between 0 and $1/\sqrt{T}$. HN show that

$$\begin{aligned} \sqrt{nT}\frac{1}{\sqrt{T}}\theta^\epsilon(0) &\xrightarrow{d} N\left(0, \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathcal{I}_i\right)^{-1}\right), \\ \left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}(0) &= \frac{1}{T}\left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i\right)^{-1} \left(-\frac{1}{n} \sum_{i=1}^n \frac{E[V_{2it}U_{it}]}{E[v_{it}^2]}\right) + O_p\left(\frac{1}{T^2}\right), \\ \left(\frac{1}{\sqrt{T}}\right)^3\theta^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}) &= o_p\left(\frac{1}{T^2}\right). \end{aligned}$$

Therefore,

$$\theta_T - \theta_0 = \frac{B}{T} + O\left(\frac{1}{T^2}\right), \quad (28)$$

where

$$B = \lim_{n \rightarrow \infty} \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i\right)^{-1} \left(-\frac{1}{n} \sum_{i=1}^n \frac{E[V_{2it}U_{it}]}{E[v_{it}^2]}\right).$$

Similarly, in the bootstrap world, for each fixed θ and ϵ , let $\alpha_i(\theta, F_i^*(\epsilon))$ and $\theta(F^*(\epsilon))$ be the solutions to the estimating equations

$$\begin{aligned} 0 &= \int V_i^*(\theta, \alpha_i(\theta, F_i^*(\epsilon))) dF_i^*(\epsilon), \\ 0 &= \sum_{i=1}^n \int U_i^*(\theta(F^*(\epsilon)), \alpha_i(\theta(F^*(\epsilon)), F_i^*(\epsilon))) dF_i^*(\epsilon), \end{aligned}$$

where $F_i^*(\epsilon) = F_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} + \epsilon\sqrt{T}\left(\hat{F}_i^* - F_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})}\right)$, $F_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})}$ is the distribution function of stratum i in the bootstrap sample and \hat{F}_i^* is the corresponding empirical distribution. Note that $F_i^*(0)$ is the same as $F_i = F_{\theta_0, \alpha_{i0}}$ except that the true parameter is $(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT}))$ rather than (θ_0, α_{i0}) .

Similar to equation (27), in the bootstrap world, we have

$$\hat{\theta}_{nT}^* - \hat{\theta}_{nT} = \frac{1}{\sqrt{T}}\hat{\theta}^\epsilon(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\hat{\theta}^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\hat{\theta}^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}^*)$$

where $\hat{\theta}^\epsilon(\epsilon) \equiv d\hat{\theta}(F^*(\epsilon))/d\epsilon$, $\hat{\theta}^{\epsilon\epsilon}(\epsilon) \equiv d^2\hat{\theta}(F^*(\epsilon))/d\epsilon^2, \dots$, and $\tilde{\epsilon}^*$ is between 0 and $1/\sqrt{T}$.

Also,

$$\sqrt{nT} \frac{1}{\sqrt{T}} \hat{\theta}^\epsilon(0) \rightarrow N \left(0, \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \right), \quad (29)$$

$$\left(\frac{1}{\sqrt{T}} \right)^2 \hat{\theta}^{\epsilon\epsilon}(0) = \frac{1}{T} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \left(-\frac{1}{n} \sum_{i=1}^n \frac{E^* [V_{2it}^* U_{it}^*]}{E^* [v_{it}^{*2}]} \right) + O_{p^*} \left(\frac{1}{T^2} \right), \quad (30)$$

$$\left(\frac{1}{\sqrt{T}} \right)^3 \hat{\theta}^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}^*) = o_{p^*} \left(\frac{1}{T^2} \right). \quad (31)$$

Using the same argument as in HN and with some calculations, we have, under Assumption 4(ii):

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} = \frac{B^*}{T} + O_p \left(\frac{1}{T^2} \right) \quad (32)$$

where

$$B^* = \lim_{n \rightarrow \infty} \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \left[-\frac{1}{n} \sum_{i=1}^n \frac{E^* [V_{2it}^* U_{it}^*]}{E^* [v_{it}^{*2}]} \right].$$

We proceed to show that $B^* = B + O_p(1/T)$. First, we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* - \frac{1}{n} \sum_{i=1}^n \mathcal{I}_i \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n \int U_{it} \left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT}) \right) U_{it} \left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT}) \right)' dF_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} - \frac{1}{n} \sum_{i=1}^n E(U_{it} U_{it}') \right| \\ &\leq A_{nT} + B_{nT} \end{aligned} \quad (33)$$

where

$$\begin{aligned} A_{nT} &= \left| \frac{1}{n} \sum_{i=1}^n \int \left(H_{it} \left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT}) \right) - H_{it}(\theta_0, \alpha_{i0}) \right) dF_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} \right| \\ B_{nT} &= \left| \frac{1}{n} \sum_{i=1}^n \int H_{it}(\theta_0, \alpha_{i0}) \left(dF_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} - dF_i \right) \right| \end{aligned}$$

and

$$H_{it}(\theta, \alpha_i) = U_{it}(\theta, \alpha_i) U_{it}(\theta, \alpha_i)'$$

To evaluate the stochastic order of A_{nT} and B_{nT} , we use the following result:

$$\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0} = \frac{\beta_i}{T} + \frac{1}{T} \sum_{t=1}^T \psi_{it} + O_p \left(\frac{1}{T^2} \right)$$

uniformly over $i = 1, 2, \dots, n$ where

$$\psi_{it} = - (E[v_{it\alpha}])^{-1} v_{it}, \quad \beta_i = - (E[v_{it\alpha}])^{-1} \left\{ E[v_{it\alpha} \psi_{it}] + \frac{1}{2} E[v_{it\alpha\alpha}] E[\psi_{it}^2] \right\}.$$

This result is given in HN and follows from the standard higher order expansion.

For A_{nT} , suppose that $\bar{\theta}_1$ is a parameter value between θ_0 and $\hat{\theta}_{nT}$ and $\bar{\alpha}_{i1}$ is a value between α_{i0} and $\hat{\alpha}_i$. Then

$$\begin{aligned}
& A_{nT} \\
&= \left| \frac{1}{n} \sum_{i=1}^n \int \left(\frac{\partial H_{it}(\bar{\theta}_1, \bar{\alpha}_{i1})}{\partial \theta} (\hat{\theta}_{nT} - \theta_0) + \frac{\partial H_{it}(\bar{\theta}_1, \bar{\alpha}_{i1})}{\partial \alpha_i} (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \right) dF_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} \right| \\
&\leq \left| \hat{\theta}_{nT} - \theta_0 \right| \left| \frac{1}{n} \sum_{i=1}^n \int \frac{\partial H_{it}(\bar{\theta}_1, \bar{\alpha}_{i1})}{\partial \theta} dF_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} \right| \\
&+ \left| \frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \int \frac{\partial H_{it}(\bar{\theta}_1, \bar{\alpha}_{i1})}{\partial \alpha_i} dF_{\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})} \right| \\
&\leq \left| \hat{\theta}_{nT} - \theta_0 \right| \left| \frac{1}{n} \sum_{i=1}^n \int M^2(z_{it}) dF_{\theta, \alpha_i} \right|_{\theta=\hat{\theta}_{nT}, \alpha=\alpha(\hat{\theta}_{nT})} \\
&+ \left| \frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \left[\int \frac{\partial H_{it}(\theta, \alpha)}{\partial \alpha_i} dF_{\theta, \alpha_i} \right]_{\theta=\hat{\theta}_{nT}, \alpha=\alpha(\hat{\theta}_{nT})} \right| (1 + o(1)) \\
&= O_p\left(\frac{1}{T}\right) + \left| \frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \left[\int \frac{\partial H_{it}(\theta, \alpha)}{\partial \alpha_i} dF_{\theta, \alpha_i} \right]_{\theta=\hat{\theta}_{nT}, \alpha=\alpha(\hat{\theta}_{nT})} \right| (1 + o(1)) \quad (34)
\end{aligned}$$

using Assumptions 1 and 4. Next

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \left[\int \frac{\partial H_{it}(\theta, \alpha)}{\partial \alpha_i} dF_{\theta, \alpha_i} \right]_{\theta=\hat{\theta}_{nT}, \alpha=\alpha(\hat{\theta}_{nT})} \right| \\
&\leq \left| \frac{1}{nT} \sum_{i=1}^n \beta_i E \left[\frac{\partial H_{it}(\theta_0, \alpha_{i0})}{\partial \alpha_i} \right] (1 + o(1)) \right| \\
&+ \left| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \psi_{it} E \left[\frac{\partial H_{it}(\theta_0, \alpha_{i0})}{\partial \alpha_i} \right] (1 + o(1)) \right| + O_p\left(\frac{1}{T^2}\right) \\
&= O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{1}{T^2}\right) = O_p\left(\frac{1}{T}\right) \quad (35)
\end{aligned}$$

using LLN and CLT. We have thus proved

$$A_{nT} = O_p\left(\frac{1}{T}\right). \quad (36)$$

For B_{nT} , suppose that $\bar{\theta}_2$ is between θ_0 and $\hat{\theta}_{nT}$ and that $\bar{\alpha}_{i2}$ is between α_{i0} and $\hat{\alpha}_i$. Then

$$\begin{aligned}
& B_{nT} = \left| \frac{1}{n} \sum_{i=1}^n \int H_{it}(\theta_0, \alpha_{i0}) \left(f(z; \hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})) - f(z; \theta_0, \alpha_{i0}) \right) dz \right| \\
&= \left| \frac{1}{n} \sum_{i=1}^n \int H_{it}(\theta_0, \alpha_{i0}) \left(\frac{\partial f(z; \bar{\theta}_2, \bar{\alpha}_{i2})}{\partial \theta} (\hat{\theta}_{nT} - \theta_0) + \frac{\partial f(z; \bar{\theta}_2, \bar{\alpha}_{i2})}{\partial \alpha_i} (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \right) dz \right| \\
&:= I_1 + I_2
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \left| \frac{1}{n} \sum_{i=1}^n \left[(\hat{\theta}_{nT} - \theta_0) \int H_{it}(\theta_0, \alpha_{i0}) u_{it}(\bar{\theta}_2, \bar{\alpha}_{i2}) dF_{\bar{\theta}_2, \bar{\alpha}_{i2}} \right] \right| \\
&= \left| (\hat{\theta}_{nT} - \theta_0) \left| \frac{1}{n} \sum_{i=1}^n \left(\int H_{it}(\theta_0, \alpha_{i0}) u_{it} dF_i + o_p(1) \right) \right| \right| \\
&\leq \sup_i E [M(z_{it})^3] \left| (\hat{\theta}_{nT} - \theta_0) \right| = O_p \left(\frac{1}{T} \right),
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
I_2 &= \left| \frac{1}{n} \sum_{i=1}^n \left[(\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \int H_{it}(\theta_0, \alpha_{i0}) v_{it}(\bar{\theta}_2, \bar{\alpha}_{i2}) dF_{\bar{\theta}_2, \bar{\alpha}_{i2}} \right] \right| \\
&= \left| \frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i(\hat{\theta}_{nT}) - \alpha_{i0}) \left[\int H_{it}(\theta_0, \alpha_{i0}) v_{it}(\theta_0, \alpha_{i0}) dF_i + o_p(1) \right] \right| = O_p \left(\frac{1}{T} \right)
\end{aligned} \tag{38}$$

using the argument similar to (35). Therefore,

$$B_{nT} = O_p \left(\frac{1}{T} \right). \tag{39}$$

Combining (36) and (39) yields:

$$\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* = \frac{1}{n} \sum_{i=1}^n \mathcal{I}_i + O_p \left(\frac{1}{T} \right). \tag{40}$$

Using the same procedure, we can show that

$$\frac{1}{n} \sum_{i=1}^n \frac{E^* [V_{2it}^* U_{it}^*]}{E^* [v_{it}^{*2}]} = \frac{1}{n} \sum_{i=1}^n \frac{E [V_{2it} U_{it}]}{E [v_{it}^2]} + O_p \left(\frac{1}{T} \right). \tag{41}$$

Therefore, from (40) and (41),

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \left(-\frac{1}{n} \sum_{i=1}^n \frac{E^* [V_{2it}^* U_{it}^*]}{E^* [v_{it}^{*2}]} \right) \\
&= \lim_{n \rightarrow \infty} \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i \right)^{-1} \left(-\frac{1}{n} \sum_{i=1}^n \frac{E [V_{2it} U_{it}]}{E [v_{it}^2]} \right) + O_p \left(\frac{1}{T} \right).
\end{aligned} \tag{42}$$

That is

$$B^* = B + O_p \left(\frac{1}{T} \right)$$

completing the proof of (25).

2. Prove $\sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - E^* \left(\hat{\theta}_{nT}^* \right) \right) \xrightarrow{p} 0$

We write

$$\begin{aligned}
& \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - E^* \left(\hat{\theta}_{nT}^* \right) \right) \\
&= \sqrt{nT} \left(E^* \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* \right) - E^* \left(\hat{\theta}_{nT}^* \right) \right) = \sqrt{nT} \left(\lim_{n \rightarrow \infty} E^* \left(\hat{\theta}_{nT}^* \right) - E^* \left(\hat{\theta}_{nT}^* \right) \right) \\
&= \sqrt{nT} \left(\lim_{n \rightarrow \infty} E^* \left(\frac{1}{\sqrt{T}} \hat{\theta}^\epsilon(0) \right) - E^* \left(\frac{1}{\sqrt{T}} \hat{\theta}^\epsilon(0) \right) \right) \\
&+ \frac{\sqrt{nT}}{2} \left(\lim_{n \rightarrow \infty} E^* \left(\frac{1}{T} \hat{\theta}^{\epsilon\epsilon}(0) \right) - E^* \left(\frac{1}{T} \hat{\theta}^{\epsilon\epsilon}(0) \right) \right) + o_p(1).
\end{aligned}$$

The second equality holds by the dominated convergence theorem and the last equality follows from an argument similar to HN. Therefore, it suffices to show that

$$\sqrt{nT} \left(\frac{1}{\sqrt{T}} \lim_{n \rightarrow \infty} E^* \left(\hat{\theta}^\epsilon(0) \right) - \frac{1}{\sqrt{T}} E^* \left(\hat{\theta}^\epsilon(0) \right) \right) = 0, \quad (43)$$

$$\sqrt{nT} \left(\frac{1}{T} \lim_{n \rightarrow \infty} E^* \left(\hat{\theta}^{\epsilon\epsilon}(0) \right) - \frac{1}{T} E^* \left(\hat{\theta}^{\epsilon\epsilon}(0) \right) \right) = o_p(1). \quad (44)$$

Equation (43) holds because

$$\begin{aligned}
& \sqrt{nT} \left(\frac{1}{\sqrt{T}} \lim_{n \rightarrow \infty} E^* \left(\hat{\theta}^\epsilon(0) \right) - \frac{1}{\sqrt{T}} E^* \left(\hat{\theta}^\epsilon(0) \right) \right) \\
&= \sqrt{n} \left(\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \frac{1}{n} \sum_{i=1}^n E^* \left(U_i^* \right) - \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \frac{1}{n} \sum_{i=1}^n E^* \left(U_i^* \right) \right) \\
&= 0,
\end{aligned}$$

using $E^* \left(U_i^* \right) = 0$.

Equation (44) holds because

$$\begin{aligned}
& \sqrt{nT} \left(\frac{1}{T} \lim_{n \rightarrow \infty} E^* \left(\hat{\theta}^{\epsilon\epsilon}(0) \right) - \frac{1}{T} E^* \left(\hat{\theta}^{\epsilon\epsilon}(0) \right) \right) \\
&= \sqrt{\frac{n}{T}} \left[\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \frac{E^* \left(V_{2it}^* U_{it}^* \right)}{E^* \left(v_{it}^* \right)} \right) \right. \\
&\quad \left. - \left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}_i^* \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \frac{E^* \left(V_{2it}^* U_{it}^* \right)}{E^* \left(v_{it}^{*2} \right)} \right) \right] + O_p \left(\sqrt{\frac{n}{T^3}} \right) \quad (45)
\end{aligned}$$

where the $O_p(\cdot)$ terms come from $O_{p^*}(\cdot)$ in (30) and the leading term in (45) is

$$\begin{aligned} & \sqrt{\frac{n}{T}} \left[\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n E_{\theta, \alpha_i} U_{it}(\theta, \alpha_i) U'_{it}(\theta, \alpha_i) \right)^{-1} \left(\sum_{i=1}^n \frac{E_{\theta, \alpha_i} (V_{2it}(\theta, \alpha_i) U_{it}(\theta, \alpha_i))}{E_{\theta, \alpha_i} (v_{it}^2(\theta, \alpha_i))} \right) \right. \\ & \left. - \left(\sum_{i=1}^n E_{\theta, \alpha_i} U_{it}(\theta, \alpha_i) U'_{it}(\theta, \alpha_i) \right)^{-1} \left(\sum_{i=1}^n \frac{E_{\theta, \alpha_i} (V_{2it}(\theta, \alpha_i) U_{it}(\theta, \alpha_i))}{E_{\theta, \alpha_i} (v_{it}^2(\theta, \alpha_i))} \right) \right] \Bigg|_{(\theta, \alpha) = (\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT}))} \\ & = \sqrt{\frac{n}{T}} \left(O_p \left(\frac{1}{\sqrt{n}} \right) \right) = o_p(1) \end{aligned}$$

using Assumption 6. In the above equation, we use the subscript θ, α_i on E_{θ, α_i} to emphasize that it is the expectation under F_{θ, α_i} . This completes the proof of Theorem 1. ■

II. Proof of Proposition 2

It is easy to show that

$$P^* \lim_{n \rightarrow \infty} \left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right)^2 = O_p \left(\frac{1}{nT} \right) = O_p \left(\frac{1}{T^2} \right), \quad (46)$$

and

$$P^* \lim_{n \rightarrow \infty} \left(\hat{\alpha}_i^* - \hat{\alpha}_i \right)^2 = O_p \left(\frac{1}{T} \right). \quad (47)$$

By the definition of the k -step bootstrap estimator:

$$\begin{pmatrix} \hat{\theta}_{nT, k}^* \\ \hat{\alpha}_k^* \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{nT, k-1}^* \\ \hat{\alpha}_{k-1}^* \end{pmatrix} - H_{k-1}^{-1} S_{k-1}.$$

For notational compactness, let $\beta = (\theta', \alpha')'$, $\hat{\beta}^* = (\hat{\theta}_{nT}^*, \hat{\alpha}^*)'$, $\hat{\beta} = (\hat{\theta}_{nT}, \hat{\alpha}')'$. Then

$$\hat{\beta}_k^* = \hat{\beta}_{k-1}^* - \left[H(\hat{\beta}_{k-1}^*; z_{it}^*) \right]^{-1} S \left(\hat{\beta}_{k-1}^*; z_{it}^* \right)$$

for $\hat{\beta}_0^* = \hat{\beta}$. Using a Taylor expansion and the first order condition:

$$S \left(\hat{\beta}^*; z_{it} \right) = 0,$$

we have

$$\begin{aligned} \hat{\beta}_k^* - \hat{\beta}^* &= \hat{\beta}_{k-1}^* - \left[H(\hat{\beta}_{k-1}^*; z_{it}^*) \right]^{-1} S \left(\hat{\beta}_{k-1}^*; z_{it}^* \right) - \hat{\beta}^* \\ &= \left[H(\hat{\beta}_{k-1}^*; z_{it}^*) \right]^{-1} \left[S \left(\hat{\beta}^*; z_{it} \right) - S \left(\hat{\beta}_{k-1}^*; z_{it}^* \right) - H(\hat{\beta}_{k-1}^*; z_{it}^*) \left(\hat{\beta}^* - \hat{\beta}_{k-1}^* \right) \right] \quad (48) \\ &= \frac{1}{2} \left[H(\hat{\beta}_{k-1}^*; z_{it}^*) \right]^{-1} \xi \end{aligned}$$

where $\hat{\beta}_{k-1}^{\dagger}$ lies between $\hat{\beta}^*$ and $\hat{\beta}_{k-1}^*$, $\xi = (\xi_1, \dots, \xi_u, \dots, \xi_{L_\beta})$ is a vector with u -th element

$$\xi_u = \left(\hat{\beta}^* - \hat{\beta}_{k-1}^* \right)' H_{\beta_u}(\hat{\beta}_{k-1}^{\dagger}; z_{it}^*) \left(\hat{\beta}^* - \hat{\beta}_{k-1}^* \right)$$

and

$$\begin{aligned} H_{\beta_u}(\beta; z_{it}^*) &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{\partial}{\partial \beta_u} \frac{\partial^2 \log(\beta; z_{it}^*)}{\partial \beta \partial \beta'} \\ &:= \begin{pmatrix} H_{\theta\theta\beta_u}(\beta; z_{it}^*) & H_{\theta\alpha\beta_u}(\beta; z_{it}^*) \\ H_{\alpha\theta\beta_u}(\beta; z_{it}^*) & H_{\alpha\alpha\beta_u}(\beta; z_{it}^*) \end{pmatrix}. \end{aligned}$$

A more explicit expression for ξ_u is

$$\begin{aligned} \xi_u &= \left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right)' H_{\theta\theta\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right) \\ &\quad + 2 \left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right)' H_{\theta\alpha\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^* \right) \\ &\quad + \left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^* \right)' H_{\alpha\alpha\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^* \right). \end{aligned}$$

Hence

$$\begin{aligned} \|\xi_u\| &\leq \left\| 2 \left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right)' H_{\theta\theta\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \left(\hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right) \right\| \\ &\quad + \left\| 2 \left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^* \right)' H_{\alpha\alpha\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \left(\hat{\alpha}_{k-1}^* - \hat{\alpha}^* \right) \right\| \\ &\leq 2 \left\| H_{\theta\theta\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right\| \left\| \hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right\|^2 \\ &\quad + \left\| \frac{2}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \left| H_{\alpha_i\alpha_i\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left(\hat{\alpha}_{i,k-1}^* - \hat{\alpha}_i^* \right)^2 \right\| \end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm, that is, for a symmetric matrix A , $\|A\|^2 = \text{trace}(AA')$. So

$$\left\| \hat{\beta}_k^* - \hat{\beta}^* \right\| \leq \zeta_{\theta,k-1}^* \left\| \hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right\|^2 + \zeta_{\alpha,k-1}^*, \quad (49)$$

where

$$\zeta_{\theta,k-1}^* = \left\| \left(H(\beta_{k-1}^*; z_{it}^*) \right)^{-1} \left\| \sum_{u=1}^{L_\beta} \left\| H_{\theta\theta\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right\| \right\|, \quad (50)$$

$$\zeta_{\alpha,k-1}^* = \left\| \left(H(\beta_{k-1}^*; z_{it}^*) \right)^{-1} \left\| \sum_{u=1}^{L_\beta} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left| H_{\alpha_i\alpha_i\beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left(\hat{\alpha}_{i,k-1}^* - \hat{\alpha}_i^* \right)^2 \right\|. \quad (51)$$

For $k = 1$, we have

$$\begin{aligned} &P \left(TP^* \lim_{n \rightarrow \infty} \left\| \hat{\beta}_k^* - \hat{\beta}^* \right\| > 2C_1 \right) \\ &\leq P \left(TP^* \lim_{n \rightarrow \infty} \zeta_{\theta,k-1}^* \left\| \hat{\theta}_{nT,k-1}^* - \hat{\theta}_{nT}^* \right\|^2 > C_1 \right) + P \left(TP^* \lim_{n \rightarrow \infty} \zeta_{\alpha,k-1}^* > C_1 \right) \end{aligned} \quad (52)$$

The first probability in (52) satisfies

$$\begin{aligned}
& P \left(TP^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* \left\| \hat{\theta}_{nT, k-1}^* - \hat{\theta}_{nT}^* \right\|^2 > C_1 \right) \\
& \leq P \left(TP^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* \left\| \hat{\theta}_{nT, k-1}^* - \hat{\theta}_{nT}^* \right\|^2 > C_1, P^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* < C_2 \right) \\
& \quad + P \left(P^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* \geq C_2 \right) \\
& \leq P \left(TP^* \lim_{n \rightarrow \infty} \left\| \hat{\theta}_{nT, k-1}^* - \hat{\theta}_{nT}^* \right\|^2 > C_1 / (C_2) \right) + P \left(P^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* \geq C_2 \right) \\
& = P \left(P^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* \geq C_2 \right) + o(1)
\end{aligned} \tag{53}$$

using (46). We proceed to bound $P \left(P^* \lim_{n \rightarrow \infty} \zeta_{\theta, k-1}^* \geq C_2 \right)$. By definition,

$$\begin{aligned}
& P \left(P^* \lim_{n \rightarrow \infty} \zeta_{nT}^* \geq C_2 \right) \\
& = P \left(\left\| P^* \lim_{n \rightarrow \infty} \left(H(\hat{\beta}_{k-1}^*; z_{it}^*) \right)^{-1} \right\| \left\| P^* \lim_{n \rightarrow \infty} \sum_{u=1}^{L_\beta} \left\| H_{\theta\theta\beta_u}(\hat{\beta}_{k-1}^\dagger; z_{it}^*) \right\| \right\| \geq C_2 \right) \\
& \leq P \left(\left\| P^* \lim_{n \rightarrow \infty} \sum_{u=1}^{L_\beta} \left\| n^{-1} H_{\theta\theta\beta_u}(\hat{\beta}_{k-1}^\dagger; z_{it}^*) \right\| \right\| \geq \sqrt{C_2} \right) + P \left(\left\| P^* \lim_{n \rightarrow \infty} \left(n^{-1} H(\hat{\beta}_{k-1}; z_{it}^*) \right)^{-1} \right\| \geq \sqrt{C_2} \right) \\
& := A + B,
\end{aligned} \tag{54}$$

where

$$A = P \left(\left\| P^* \lim_{n \rightarrow \infty} \sum_{u=1}^{L_\beta} \left\| n^{-1} H_{\theta\theta\beta_u}(\hat{\beta}_{k-1}^\dagger; z_{it}^*) \right\| \right\| \geq \sqrt{C_2} \right), \tag{55}$$

and

$$B = P \left(\left\| P^* \lim_{n \rightarrow \infty} \left(n^{-1} H(\hat{\beta}_{k-1}; z_{it}^*) \right)^{-1} \right\| \geq \sqrt{C_2} \right).$$

Note that $\hat{\beta}_{k-1}^\dagger$ is between $\hat{\beta}$ and $\hat{\beta}^*$ and

$$\hat{\beta}^* = \hat{\beta} + o_{P^*}(1). \tag{56}$$

Using a uniform law of large numbers under the probability measure P^* and the dominated convergence theorem, we have

$$\begin{aligned}
& P^* \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u=1}^{L_\beta} \left\| H_{\theta\theta\beta_u}(\hat{\beta}_{k-1}^\dagger; z_{it}^*) \right\| = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u=1}^{L_\beta} P^* \lim_{n \rightarrow \infty} \left\| \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{\partial}{\partial \beta_u} \frac{\partial^2 \log(\hat{\beta}_{k-1}^\dagger; z_{it}^*)}{\partial \theta \partial \theta'} \right\| \\
& = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u=1}^{L_\beta} \left\| \lim_{n \rightarrow \infty} E^* \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \frac{\partial}{\partial \beta_u} \frac{\partial^2 \log(\hat{\beta}; z_{it}^*)}{\partial \theta \partial \theta'} \right\| \\
& = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u=1}^{L_\beta} \left\| \bar{E}_\beta \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \beta_u} \frac{\partial^2 \log(\beta; z_{it})}{\partial \theta \partial \theta'} \right] \right\|_{\beta=\hat{\beta}}
\end{aligned} \tag{57}$$

where the last equality follows because P^* conditional on the data (i.e. $\hat{\theta}_{nT}, \hat{\alpha}_i$) is the same as P but with different model parameters. Hence,

$$A = P \left(\frac{1}{n} \sum_{u=1}^{L_\beta} \left\| \bar{E}_\beta \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \beta_u} \frac{\partial^2 \log(\beta; z_{it}^*)}{\partial \theta \partial \theta'} \right] \right\|_{\beta=\hat{\beta}} \geq \sqrt{C_2} \right)$$

which can be made arbitrarily small if we choose a large C_2 .

Using the same argument, we can show that, when C_2 is large enough, $B = o(1)$ as n and T go to ∞ . We have therefore proved

$$P \left(P^* \lim_{n \rightarrow \infty} \zeta_{nT}^* \geq C_2 \right) = o(1). \quad (58)$$

when C_2 is large enough.

To show that the second probability in (52) is $o(1)$, it suffices to prove

$$P \left(\frac{T}{n} \sum_{u=1}^{L_\beta} P^* \lim_{n \rightarrow \infty} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left| H_{\alpha_i \alpha_i \beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left(\hat{\alpha}_{i,k-1}^* - \hat{\alpha}_i^* \right)^2 > \sqrt{C_2} \right) = o(1).$$

But

$$\begin{aligned} & \frac{1}{n} \sum_{u=1}^{L_\beta} P^* \lim_{n \rightarrow \infty} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left| H_{\alpha_i \alpha_i \beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left(\hat{\alpha}_{i,k-1}^* - \hat{\alpha}_i^* \right)^2 \\ &= \frac{1}{n} \sum_{u=1}^{L_\beta} P^* \lim_{n \rightarrow \infty} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left| H_{\alpha_i \alpha_i \beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left(\frac{\beta_i^*}{T} + \frac{1}{T} \sum_{t=1}^T \psi_{it}^* \right)^2 (1 + o_p(1)) \\ &= \frac{1}{n} \sum_{u=1}^{L_\beta} P^* \lim_{n \rightarrow \infty} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left| H_{\alpha_i \alpha_i \beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left[\frac{(\beta_i^*)^2}{T^2} + \frac{1}{T} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_{it}^* \right)^2 \right] (1 + o_p(1)) \\ &= \frac{1}{n} \sum_{u=1}^{L_\beta} P^* \lim_{n \rightarrow \infty} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left| H_{\alpha_i \alpha_i \beta_u} \left(\hat{\beta}_{k-1}^\dagger; z_{it}^* \right) \right| \left[\frac{(\beta_i^*)^2}{T^2} + \frac{1}{T^2} \sum_{t=1}^T (\psi_{it}^*)^2 \right] (1 + o_p(1)) \\ &= O_p \left(\frac{1}{T} \right) \end{aligned}$$

where the last equality follows from the ULLN.

Combining (53) with (58), we have, for $k = 1$.

$$P \left(TP^* \lim_{n \rightarrow \infty} \left| \hat{\beta}_k^* - \hat{\beta}^* \right| > C \right) = o(1)$$

when C is large enough. That is, when $k = 1$

$$P^* \lim_{n \rightarrow \infty} \hat{\beta}_k^* = P^* \lim_{n \rightarrow \infty} \hat{\beta}^* + O_p \left(\frac{1}{T} \right). \quad (59)$$

For $k \geq 2$, we note that

$$\left\| \hat{\beta}_k^* - \hat{\beta}^* \right\| \leq \eta_{nT}^* \left\| \hat{\beta}_{k-1}^* - \hat{\beta}^* \right\|^2$$

where

$$\eta_{nT}^* = \frac{1}{2} \max_k \left\| \left[H(\hat{\beta}_{k-1}^*; z_{it}^*) \right]^{-1} H_{\beta_u}(\hat{\beta}_{k-1}^*; z_{it}^*) \right\|$$

Using the recursive relationship repeatedly, we have, for $k \geq 2$,

$$\left\| \hat{\beta}_k^* - \hat{\beta}^* \right\| \leq (\eta_{nT}^*)^\phi \left\| \hat{\beta}_1^* - \hat{\beta}^* \right\|^{2^{k-1}} \quad (60)$$

where $\phi = \sum_{j=2}^k 2^{j-1}$.

Using a similar argument, we can show that $P^* \lim_{n \rightarrow \infty} \eta_{nT}^* = O_p(1)$. Combining this with (59) and (60), we have

$$P^* \lim_{n \rightarrow \infty} \left(\hat{\beta}_k^* - \hat{\beta}^* \right) = O_p \left(\left(\frac{1}{T} \right)^{2^{k-1}} \right),$$

which implies that

$$P^* \lim_{n \rightarrow \infty} \left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}^* \right) = O_p \left(\left(\frac{1}{T} \right)^{2^{k-1}} \right)$$

as desired. ■

III. Proof of Theorem 3

Define our truncated k -step bootstrap bias corrected estimator to be:

$$\check{\theta}_{nT,k} \equiv 2\hat{\theta}_{nT} - E^* \left(\check{\theta}_{nT,k}^* \right).$$

Then

$$\begin{aligned} & \sqrt{nT} \left(\check{\theta}_{nT,k} - \theta_0 \right) \\ &= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[E^* \left(\check{\theta}_{nT,k}^* \right) - \hat{\theta}_{nT} \right] \right) \\ &= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right] \right) + \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* \right) \\ &+ \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* \right) + \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* \left(\check{\theta}_{nT,k}^* \right) \right) \end{aligned}$$

As HN has shown that

$$\sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) \xrightarrow{d} N(0, \bar{E}[\mathcal{I}_i]^{-1})$$

and we have shown that

$$\sqrt{nT} \left(\theta_T - \theta_0 - \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right] \right) = O_p \left(\frac{1}{T^2} \right)$$

and

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* = O_p \left(\frac{1}{T^2} \right)$$

Therefore, it suffices to show

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* = O_p \left(\frac{1}{T^2} \right) \quad (61)$$

$$P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* \left(\check{\theta}_{nT,k}^* \right) = O_p \left(\frac{1}{T^2} \right) \quad (62)$$

1. Prove $P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* = O_p\left(\frac{1}{T^2}\right)$

By definition (c.f. (17)) we have:

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* = P^* \lim_{n \rightarrow \infty} \left(\hat{\theta}_{nT} - \hat{\theta}_{nT,k}^* \right) 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| > M_{nT} \right).$$

As

$$\begin{aligned} & \left| P^* \lim_{n \rightarrow \infty} \left(\hat{\theta}_{nT} - \hat{\theta}_{nT,k}^* \right) \right| \\ & \leq \left| P^* \lim_{n \rightarrow \infty} \left(\hat{\theta}_{nT} - \hat{\theta}_{nT}^* \right) \right| + \left| P^* \lim_{n \rightarrow \infty} \left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT,k}^* \right) \right| = O_p\left(\frac{1}{T}\right), \end{aligned}$$

it suffices to show

$$P^* \lim_{n \rightarrow \infty} 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| > M_{nT} \right) = O_p\left(\frac{1}{T}\right).$$

For any given $\delta > 0$, we have

$$\begin{aligned} & P \left(T \cdot P^* \lim_{n \rightarrow \infty} 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| > M_{nT} \right) > \delta \right) \\ & = P \left(P^* \lim_{n \rightarrow \infty} 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| > M_{nT} \right) = 1 \right) \\ & = P \left(1 \left(P^* \lim_{n \rightarrow \infty} \sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| > M_{nT} \right) = 1 \right) \\ & = P \left(\left(P^* \lim_{n \rightarrow \infty} \sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| \right) > M_{nT} \right) \\ & \leq P \left(\left(P^* \lim_{n \rightarrow \infty} \sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}^* \right| + P^* \lim_{n \rightarrow \infty} \sqrt{nT} \left| \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right| \right) > M_{nT} \right) \\ & = P \left(\sqrt{nT} \left[O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{T^2}\right) \right] > M_{nT} \right) \\ & = o(1) \end{aligned}$$

using the condition that $\sqrt{\frac{n}{T}} = o(M_{nT})$ and $M_{nT} \rightarrow \infty$. Here we have used $P^* \lim_{n \rightarrow \infty} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}^* \right| = O_p\left(\frac{1}{T^2}\right)$ and $P^* \lim_{n \rightarrow \infty} \sqrt{nT} \left| \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right| / M_{nT} = o_p(1)$.

Combining the above results yields

$$P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* = O_p\left(\frac{1}{T^2}\right)$$

as desired.

2. Prove $\sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* \left(\check{\theta}_{nT,k}^* \right) \right) = o_p(1)$

We write $\sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* \left(\check{\theta}_{nT,k}^* \right) \right)$ as

$$\begin{aligned} & \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* \left(\check{\theta}_{nT,k}^* \right) \right) \\ & = \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \left[\left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right) 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| \leq M_{nT} \right) \right] \right. \\ & \quad \left. - E^* \left[\left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right) 1 \left(\sqrt{nT} \left| \hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right| \leq M_{nT} \right) \right] \right) \end{aligned}$$

We consider only the case that $k = 1$ as the proof for other cases is similar. Let

$$D_{nT}(\theta, \alpha) = \left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T U_{it\theta}(\theta, \alpha) \right)^{-1} \left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T U_{it}(\theta, \alpha) \right),$$

then

$$\begin{aligned} & \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* (\check{\theta}_{nT,k}^*) \right) \\ &= -P^* \lim_{n \rightarrow \infty} \left[\sqrt{nT} D_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) 1 \left(\left| \sqrt{nT} D_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) \right| \leq M_{nT} \right) \right] \\ &+ E^* \left[\sqrt{nT} D_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) 1 \left(\left| \sqrt{nT} D_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}(\hat{\theta}_{nT})) \right| \leq M_{nT} \right) \right] \\ &= E \left[\sqrt{nT} D_{nT}(\theta, \alpha) 1 \left(\sqrt{nT} |D_{nT}(\theta, \alpha)| \leq M_{nT} \right) \right] \Big|_{\theta=\hat{\theta}_{nT}, \alpha=\hat{\alpha}(\hat{\theta}_{nT})} \\ &- P \lim_{n \rightarrow \infty} \sqrt{nT} D_{nT}(\theta, \alpha) 1 \left(\sqrt{nT} |D_{nT}(\theta, \alpha)| \leq M_{nT} \right) \Big|_{\theta=\hat{\theta}_{nT}, \alpha=\hat{\alpha}(\hat{\theta}_{nT})} \\ &= o_p(1) \end{aligned}$$

where the last equality follows from the dominated convergence theorem.

Finally,

$$\begin{aligned} & \sqrt{nT} \left(\tilde{\theta}_{nT,k} - \theta_0 \right) \\ &= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[E^* (\check{\theta}_{nT,k}^*) - \hat{\theta}_{nT} \right] \right) \\ &= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + \sqrt{nT} \left(\theta_T - \theta_0 - \left[P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right] \right) + \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT}^* - P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* \right) \\ &+ \sqrt{nT} \left(P^* \lim_{n \rightarrow \infty} \hat{\theta}_{nT,k}^* - P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* \right) + \sqrt{nT} \left[P^* \lim_{n \rightarrow \infty} \check{\theta}_{nT,k}^* - E^* (\check{\theta}_{nT,k}^*) \right] \\ &= \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_T \right) + o_p(1) \xrightarrow{d} N(0, \bar{E} [\mathcal{I}_i]^{-1}) \end{aligned}$$

as $n/T^3 \rightarrow 0$. Thus, Theorem 3 is proved. ■

References

- Andersen, E. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B*, 32(2):283–301.
- Andrews, D. (2002). Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators. *Econometrica*, 70(1):119–162.
- Andrews, D. (2005). Higher-order Improvements of the Parametric Bootstrap for Markov Processes. *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*, ed. by D.W.K. Andrews and J.H. Stock.
- Arellano, M. and Hahn, J. (2006). Understanding Bias in Nonlinear Panel Models: Some Recent Developments. *Advances in Economics and Econometrics, Ninth World Congress, Cambridge University Press*.
- Bester, A. and Hansen, C. (2008). A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects. *Working paper, University of Chicago*.
- Davidson, R. and MacKinnon, J. (1999). Bootstrap Testing in Nonlinear Models. *International Economic Review*, 40(2):487–508.
- Fernandez-Val, I. (2009). Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models. *Journal of Econometrics, Forthcoming*.
- Greene, W. (2004). The Behavior of the Fixed Effects Estimator in Nonlinear Models. *The Econometrics Journal*, 7(1):98–119.
- Hahn, J., Kuersteiner, G., and Newey, W. (2004). Higher order efficiency of bias corrections. *Unpublished Manuscript, MIT*.
- Hahn, J. and Newey, W. (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica*, 72(4):1295–1319.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Verlag.
- Heckman, J. (1981). The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. *Structural Analysis of Discrete Data With Economic Applications*.
- Honore, B. and Kyriazidou, E. (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica*, 68(4):839–874.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413.
- Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.
- Pace, L. and Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, 136(10):3554–3564.