Bootstrap and k-step Bootstrap Bias Corrections for the Fixed Effects Estimator in Nonlinear Panel Data Models

Min Seong Kim
Department of Economics
Ryerson University

Yixiao Sun*
Department of Economics
UC San Diego

July 2015

*Email: minseong.kim@ryerson.ca 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, David Kaplan, Peter Phillips, Dimitris Politis, Hal White, Guido Kuersteiner, the Co-Editor, and three anonymous referees for helpful comments. We also thank Iván Fernández-Val for generously sharing his Matlab codes. Kim gratefully acknowledges financial support from the Office of the Dean of Arts, Ryerson University.

Abstract

Because of the incidental parameters problem, the fixed effects maximum likelihood estimator in a nonlinear panel data model is in general inconsistent when the time series length T is short and fixed. Even if T approaches infinity but at a rate not faster than the cross sectional sample size n, the fixed effects estimator is still asymptotically biased. This paper proposes using the standard bootstrap and k-step bootstrap to correct the bias. We establish the asymptotic validity of the bootstrap bias corrections for both model parameters and average marginal effects. Our results apply to static models as well as some dynamic Markov models. Monte Carlo simulations show that our procedures are effective in reducing the bias of the fixed effects estimator and improving the coverage accuracy of the associated confidence interval.

- Running head: Bootstrap Bias Correction for the Fixed Effects Estimator

- Corresponding Author: Yixiao Sun (yisun@ucsd.edu)

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 if T approaches ∞, the fixed effects estimator can still have an asymptotic bias that is comparable to the asymptotic standard error. Statistical inference that ignores the asymptotic bias may give misleading results.

This paper proposes using the standard parametric bootstrap and the k-step parametric bootstrap to correct for the asymptotic bias. We consider a nonlinear fixed effects panel data model with each time series following a finite order Markov process. We allow for static panel data models as well as some dynamic panel data models. We employ the maximum likelihood (ML) approach. Under some rate conditions on the time series and cross sectional sample sizes, we show that the standard bootstrap-bias-corrected (BBC) estimators are asymptotically normal and centered at the true parameter. The regularity conditions for BBC estimators are essentially the same as those for the analytic bias corrections in the literature. We also establish the asymptotic equivalence of the k-step BBC estimator to the standard BBC estimator. Like other bias correction procedures, bootstrap bias correction reduces the asymptotic bias without inflating the first order asymptotic variance[1]. So inferences based on the BBC estimators are expected to be more accurate and reliable than the ones based on the original fixed effects estimators.

An advantage of the standard bootstrap bias correction is that the method is automatic to a great extent. There is no need to derive the analytic bias correction formulae, which can be quite complicated. The automatic nature is especially appealing in applied research. A drawback of the standard bootstrap method is that it is computationally intensive, as it involves solving 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 consuming task. Particularly, as the fixed effects approach treats the individual effects as parameters, there are many parameters to be estimated, and the computational cost can be excessive. For example, in our empirical application (not reported here), there are 1461 individuals, which means that there are more than 1461 parameters to be estimated.

We propose using the k-step bootstrap method to alleviate the computational cost of the standard bootstrap. In the k-step bootstrap 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. Compared to the standard bootstrap method, the k-step bootstrap procedure is computationally attractive because it involves only closed-form expressions. The main computational cost of the k-step bootstrap is to compute the score and Hessian functions, which is a relatively easy computing task. We show that when $k \geq 2$, the stochastic difference between the standard and k-step bootstrap estimators 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.

In addition to model parameters, we apply the standard and k-step parametric bootstrap bias corrections to the average marginal effect estimation. We also develop a double bootstrap procedure for confidence interval (CI) construction. To rigorously justify our bootstrap procedures, we use the following strategy repeatedly. For an asymptotic result of interest, we first show that it holds uniformly over an open set in the parameter space. Using the fact that the ML estimator, which is the true parameter in the bootstrap world, lies in this set with probability approaching one uniformly, we show that the asymptotic result also holds in the bootstrap world with probability approaching one uniformly. More details of this type of argument are given in the appendix of proofs.

Several papers have discussed the difficulties involved in controlling for the incidental parameters problem in nonlinear panel data 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 do not provide guidance in general cases to eliminate the bias. More generally, Hahn and Newey (2004, denoted HN hereinafter) propose jackknife and analytic procedures for nonlinear static models, and Hahn and Kuersteiner (2011, denoted HK hereinafter) 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 (2009) propose a penalized objective function approach and Fernández-Val (2009) develops bias correction for parametric panel binary choice models. Dhaene and Jochmans (2012) consider split-panel jackknife estimation for the dynamic case. Compared to the analytic and jackknife methods, the bootstrap procedure has gained relatively little attention in this setting even though the latter is a natural method to bias estimation. From this point of view, this paper makes an important contribution to the literature, providing the BBC estimators for nonlinear panel models and establishing their validity.[2]

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) examine the asymptotic properties of a bootstrap bias corrected ML estimator with cross sectional data and show that it is higher order efficient. The k-step bootstrap procedure first appears in Davidson and MacKinnon (1999) and Andrews (2002, 2005), who prove its higher order equivalence to the standard bootstrap. Our paper in particular builds on Andrews (2005), who considers the standard and k-step parametric bootstrap methods for Markov processes. For papers that aim to reduce the computational cost of bootstrap procedures and simulation experiments that involve bootstrap estimators, see, for example, the fast double bootstrap by Davidson and MacKinnon (2002, 2007) and the warp-speed bootstrap by Giacomini, Politis, and White (2013).

The rest of the paper is organized as follows. Section 2 discusses the incidental parameters problem in a nonlinear panel data model with fixed effects. Section 3 describes the standard bootstrap bias correction procedure and the k-step bootstrap bias correction procedure. Section 4 establishes the asymptotic properties of our estimators. Some extensions are given in Section 5. Section 5.1 discusses bias correction for average marginal effect estimation, and Section 5.2 introduces the double bootstrap procedure for CI construction. Monte Carlo simulation results are reported in Section 6. The last section concludes. For easy reference, we provide a list of selected notations before presenting technical proofs.

2 Incidental Parameters Problem

In this section, we introduce the incidental parameters problem and discuss the asymptotic bias of the fixed effects estimator in a nonlinear panel data model.

Throughout the paper, we maintain the assumption of cross sectional independence. To specify the nonlinear panel data model, we only need to describe the data generating process for each time series $\{W_i : W_{i1}, \ldots, W_{iT}\}$ where $W_{it} \in \mathbb{R}^{d_W}$, $i = 1, 2, \ldots, n$ and d_W is the dimension of W. We partition W_{it} into $W_{it} = (Y_{it}', X_{it}')'$ where $\{X_{it}\}$ are strictly exogenous, and hence can be conditioned on or taken as fixed, and $\{Y_{it}\}$ form a κ-th order Markov process for some finite integer κ. To simplify the presentation, we ignore X_{it} so that $W_{it} = Y_{it}$. In the presence of $\{X_{it}\}$, our results hold conditional on $\{X_{it}\}$.

We assume that the density and distribution functions of Y_{it} given $Y_{it-1}, \ldots, Y_{i1}$ are

$$f(\cdot|Y_{it-1}, \ldots, Y_{it-\kappa}; \theta, \alpha_i) \text{ and } F(\cdot|Y_{it-1}, \ldots, Y_{it-\kappa}; \theta, \alpha_i)$$

respectively, where $\theta \in \mathbb{R}^{d_\theta}$ is a vector of parameters of interest and α_i is a scalar capturing individual heterogeneity. We assume that the initial observations $\{Y_{i0}, \ldots, Y_{i1-\kappa}\}$ are available in which case there are $T + \kappa$ time series observations. But our inference will be conditioned on the initial observations, so the effective time series sample size is T.

Let $\gamma_i = \left(\theta', \alpha_i\right)'$ be the parameter that governs individual time series, and denote the parameter space for γ_i by $\Gamma \equiv \Gamma_\theta \times \Gamma_\alpha$.[3] The true parameter $\gamma_{i0} = \left(\theta_0', \alpha_{i0}\right)'$ belongs to a subset Γ_0 of Γ. Let $Z_{it} = (Y_{it}, Y_{it-1}, \ldots, Y_{it-\kappa})'$ and $l(\theta, \alpha_i; Z_{it}) \equiv \log f(Y_{it}|Y_{it-1}, \ldots Y_{it-\kappa}; \theta, \alpha_i)$. When we need to emphasize the dependence of Z_{it} on the true parameter γ_{i0}, we write $Z_{it} = Z_{it}\left(\gamma_{i0}\right)$. The objective function for the fixed effects estimator, $\hat{\theta}_{nT}$, is the concentrated log-likelihood function based on $\hat{\alpha}_i(\theta)$. That is, we obtain $\hat{\theta}_{nT}$ by solving

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

where

$$\hat{\alpha}_i(\theta) = \arg\max_{\alpha_i \in \Gamma_\alpha} \sum_{t=1}^{T} l(\theta, \alpha_i; Z_{it}), \tag{2}$$

and Γ_θ and Γ_α are assumed to be compact. We denote $\hat{\alpha}\left(\theta\right) \equiv (\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_i(\theta), \ldots, \hat{\alpha}_n(\theta))'$, $\alpha_0 \equiv (\alpha_{10}, \ldots, \alpha_{i0}, \ldots, \alpha_{n0})'$ and $\hat{\alpha} \equiv \hat{\alpha}(\hat{\theta}_{nT})$.

It follows from equation (2) that $\hat{\alpha}_i(\theta)$ is based on only T time series observations $(Z_{i1}, \ldots, Z_{iT})$. Therefore, if T is fixed, $\hat{\alpha}_i$ does not converge to α_{i0} even though $n \to \infty$. The estimation error in $\hat{\alpha}_i$ leads to the inconsistency of $\hat{\theta}_{nT}$, i.e. $\text{plim}_{n\to\infty} \hat{\theta}_{nT} \neq \theta_0$. This is known as the incidental parameters problem first noted by Neyman and Scott (1948). From the standard asymptotic theory for extremum estimators (e.g. Amemiya, 1985), we have $\hat{\theta}_{nT} \to^p \theta_T$ as $n \to \infty$ with T

fixed, where $\theta_T \equiv \arg\max_\theta \overline{\mathbb{E}}\left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); Z_{it})\right]$ and[4]

$$\overline{\mathbb{E}}\left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); Z_{it})\right] \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); Z_{it})\right]. \tag{3}$$

Since θ_0 maximizes $\overline{\mathbb{E}}[\sum_{t=1}^T l(\theta, \alpha_{i0}; Z_{it})]$, which is different from $\overline{\mathbb{E}}\left[\sum_{t=1}^T l(\theta, \hat{\alpha}_i(\theta); Z_{it})\right]$, it is usually the case that $\theta_T \neq \theta_0$. As a result, $\hat{\theta}_{nT}$ is inconsistent as $n \to \infty$ for a fixed T.

As $T \to \infty$, we can show by stochastic expansion that

$$\theta_T = \theta_0 + \frac{B(\theta_0, \alpha_0)}{T} + o\left(\frac{1}{T}\right) \tag{4}$$

for some $B(\theta_0, \alpha_0)$ that is not zero in general. So the asymptotic bias is of order $O(1/T)$ and approaches zero as T increases.

The above asymptotic analysis, which is first discussed in HN, is conducted under the sequential asymptotics under which $n \to \infty$ for a fixed T followed by letting $T \to \infty$. This approach is also employed by Hahn and Kuersteiner (2002) who consider dynamic linear panel data models. The basic intuition on the incidental parameters bias holds under the joint asymptotics where n and T go to ∞ simultaneously. When n and T grow at the same rate, the asymptotic bias is of the same order of magnitude as the asymptotic standard error, which is of order $O(1/\sqrt{nT})$. In this case, the asymptotic distribution of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$ will not be centered at zero. More specifically, when $n \to \infty$ and $T \to \infty$ simultaneously such that $n/T \to \rho \in (0, \infty)$, we can write:

$$\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = A_{nT}(\theta_0, \alpha_0) + \sqrt{\frac{n}{T}} B_{nT}(\theta_0, \alpha_0) + o_p(1) \tag{5}$$

for some $A_{nT}(\theta_0, \alpha_0)$ and $B_{nT}(\theta_0, \alpha_0)$ that satisfy

$$A_{nT}(\theta_0, \alpha_0) \xrightarrow{d} N(0, \Omega) \text{ and } B_{nT}(\theta_0, \alpha_0) \to^p B(\theta_0, \alpha_0) \tag{6}$$

where $\Omega \equiv \Omega(\theta_0, \alpha_0)$ is the asymptotic variance matrix. The expressions for A_{nT} and B_{nT} are given in Section 4. So the asymptotic distribution of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$ is centered at $\sqrt{\rho}B(\theta_0, \alpha_0)$, which is in general not zero. Statistical inference that ignores the nonzero center may result in misleading conclusions.

Since the mean of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$ may not exist, $\sqrt{n/T}B_{nT}(\theta_0, \alpha_0)$ is not necessarily equal to $\mathbb{E}[\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)]$. For any $M>0$, define the truncation function $\tilde{g}_M(x)$ on $\mathbb{R}$ by

$$\tilde{g}_M(x)=\begin{cases} M, & \text{if } x>M; \\ x, & \text{if } |x|\leq M; \\ -M, & \text{if } x<-M. \end{cases} \tag{7}$$

Then $\tilde{g}_M(x)$ is bounded and Lipschitz-continuous. With some abuse of notation, for any $x\in\mathbb{R}^{d_x}$, let $g_M(x)=(\tilde{g}_M(x_1),\ldots,\tilde{g}_M(x_{d_x}))'\in\mathbb{R}^{d_x}$. By construction, $\mathbb{E}g_M[\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)]$ always exists. Under some rate conditions on M, we can show that

$$\mathbb{E}g_M\left[\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)\right]=\sqrt{n/T}B_{nT}(\theta_0,\alpha_0)+o(1). \tag{8}$$

To correct the asymptotic bias of $\hat{\theta}_{nT}$, we need to estimate $\mathbb{E}g_M[\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)]$.

3 Bootstrap Bias Correction

In this section, we introduce the parametric bootstrap bias correction procedures. We maintain cross sectional independence in the bootstrap world and generate each time series $\{Y_{it}^*\}_{t=1-\kappa}^T$ according to the conditional probability density function $f(\cdot|Y_{it-1}^*,\ldots,Y_{it-\kappa}^*;\hat{\theta}_{nT},\hat{\alpha}_i)$, which is the same as the conditional probability density function for the original sample but with $\gamma_{i0}=\left(\theta_0',\alpha_{i0}\right)'$ replaced by $\hat{\gamma}_i=(\hat{\theta}_{nT}',\hat{\alpha}_i)'$. For the bootstrap sample, we use the same initial values and other conditioning variables, if any, as for the original sample.

Let $Z_{it}^*=\left(Y_{it}^*,Y_{it-1}^*,\ldots,Y_{it-\kappa}^*\right)'$. Based on $\{Z_{i1}^*,\ldots,Z_{iT}^*\}_{i=1}^n$, we can obtain the bootstrap estimators $\hat{\theta}_{nT}^*$ and $\hat{\alpha}_i^*(\hat{\theta}_{nT}^*)$ by conditional ML estimation. That is,

$$\hat{\theta}_{nT}^*=\arg\max_{\theta\in\Gamma_\theta}\sum_{i=1}^n\sum_{t=1}^T l(\theta,\hat{\alpha}_i^*(\theta);Z_{it}^*), \tag{9}$$

where

$$\hat{\alpha}_i^*(\theta)=\arg\max_{\alpha_i\in\Gamma_\alpha}\sum_{t=1}^T l(\theta,\alpha_i;Z_{it}^*). \tag{10}$$

Under some regularity conditions, as $n\to\infty$ and $T\to\infty$ such that $n/T\to\rho\in(0,\infty)$, we have

$$\sqrt{nT}(\hat{\theta}_{nT}^*-\hat{\theta}_{nT})=A_{nT}^*(\hat{\theta}_{nT},\hat{\alpha})+\sqrt{\frac{n}{T}}B_{nT}^*(\hat{\theta}_{nT},\hat{\alpha})+o_{p^*}(1) \tag{11}$$

conditional on a set with probability approaching 1 (see Lemma A.4). Here $A^*_{nT}(\theta,\alpha)$ and $B^*_{nT}(\theta,\alpha)$ are defined in the same manner as $A_{nT}(\theta,\alpha)$ and $B_{nT}(\theta,\alpha)$ but the former are based on the bootstrap sample rather than the original sample. We have

$$\sqrt{\frac{n}{T}}B^*_{nT}(\hat{\theta}_{nT},\hat{\alpha}) = \mathbb{E}^* g_M\left[\sqrt{nT}(\hat{\theta}^*_{nT}-\hat{\theta}_{nT})\right] + o_p\left(1\right), \tag{12}$$

where $\mathbb{E}^*$ is the expectation operator under the bootstrap probability distribution P^* conditional on the original sample.

Using $\sqrt{n/T}B^*_{nT}(\hat{\theta}_{nT},\hat{\alpha})$ or its asymptotically equivalent form $\mathbb{E}^* g_M\left[\sqrt{nT}(\hat{\theta}^*_{nT}-\hat{\theta}_{nT})\right]$ as an estimator of $\sqrt{n/T}B_{nT}(\theta_0,\alpha_0)$, we can define the BBC estimator $\tilde{\theta}_{nT}$ by

$$\sqrt{nT}(\tilde{\theta}_{nT}-\theta_0) = \sqrt{nT}(\hat{\theta}_{nT}-\theta_0) - \mathbb{E}^* g_M\left[\sqrt{nT}(\hat{\theta}^*_{nT}-\hat{\theta}_{nT})\right]. \tag{13}$$

Equivalently,

$$\tilde{\theta}_{nT} = \hat{\theta}_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^* g_M\left[\sqrt{nT}(\hat{\theta}^*_{nT}-\hat{\theta}_{nT})\right]. \tag{14}$$

To compute $\mathbb{E}^* g_M[\sqrt{nT}(\hat{\theta}^*_{nT}-\hat{\theta}_{nT})]$, we generate R bootstrap samples and obtain R bootstrap estimates $\{\hat{\theta}^{*(r)}_{nT},\, r=1,\ldots,R\}$ and then let

$$\tilde{\mathbb{E}}^* g_M\left[\sqrt{nT}(\hat{\theta}^*_{nT}-\hat{\theta}_{nT})\right] = \frac{1}{R}\sum_{r=1}^{R} g_M\left[\sqrt{nT}(\hat{\theta}^{*(r)}_{nT}-\hat{\theta}_{nT})\right]. \tag{15}$$

We choose a large enough R so that the difference between $\tilde{\mathbb{E}}^* g_M[\cdot]$ and $\mathbb{E}^* g_M[\cdot]$ can be made as small as possible.

To reduce the computational cost, we use the k-step NR iterative procedure to approximate $\hat{\theta}^*_{nT}$. Let $\hat{\theta}^*_{nT,k}$ and $\hat{\alpha}^*_k = \left(\hat{\alpha}^*_{1,k},\ldots,\hat{\alpha}^*_{n,k}\right)'$ denote the k-step bootstrap estimators. We define $\hat{\theta}^*_{nT,k}$ and $\hat{\alpha}^*_k$ recursively in the following way:

$$\begin{pmatrix} \hat{\theta}^*_{nT,k} \\ \hat{\alpha}^*_k \end{pmatrix} = \begin{pmatrix} \hat{\theta}^*_{nT,k-1} \\ \hat{\alpha}^*_{k-1} \end{pmatrix} - \left(\mathcal{H}_{k-1}\right)^{-1}\mathcal{S}_{k-1} \tag{16}$$

where[5]

$$\mathcal{H}_{k-1} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left.\frac{\partial^2 l\left(\theta,\alpha_i;Z^*_{it}\right)}{\partial\left(\theta',\alpha'\right)'\partial\left(\theta',\alpha'\right)}\right|_{\theta=\hat{\theta}^*_{nT,k-1},\alpha=\hat{\alpha}^*_{k-1}} \tag{17}$$

$$\mathcal{S}_{k-1} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\partial l\left(\theta, \alpha_i; Z_{it}^*\right)}{\partial \left(\theta', \alpha'\right)'} \Bigg|_{\theta=\hat{\theta}_{nT,k-1}^*, \alpha=\hat{\alpha}_{k-1}^*} \tag{18}$$

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

Define the k-step BBC estimator as

$$\tilde{\theta}_{nT,k} \equiv \hat{\theta}_{nT} - \frac{1}{\sqrt{nT}} \mathbb{E}^* g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right) \right], \tag{19}$$

which is analogous to the BBC estimator defined in (14). We will show that the k-step BBC estimator is asymptotically equivalent to the standard BBC estimator when $k \geq 2$.

4 Asymptotic Properties

In this section, we show that the standard and k-step bootstrap bias corrections remove the nonzero center of the asymptotic distribution of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$.

Let $\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)'$ and $\gamma_0 = \left(\theta_0', \alpha_{10}, \ldots, \alpha_{n0}\right)'$. To emphasize their dependence on the parameter value, we may use P_{γ_0} and $\mathbb{E}_{\gamma_0}$ to denote the probability measure and its expectation under γ_0. Similarly, we use $\hat{\theta}_{nT}\left(\gamma_0\right)$ and $\hat{\alpha}_i\left(\gamma_0\right)$ to highlight their dependence on the observations $\left\{Z_{it}\left(\gamma_{i0}\right), t = 1, \ldots, T\right\}_{i=1}^n$ generated under the parameter value γ_0. When needed, we use similar notations in the bootstrap world, for example, $\mathbb{E}_{\hat{\gamma}}^*$, $\left\{Z_{it}^*\left(\hat{\gamma}_i\right)\right\}$, etc.

Following HN (2004), we define

$$u_{it}(\gamma_i) \equiv \frac{\partial}{\partial \theta} l(\theta, \alpha_i; Z_{it}) \quad \text{and} \quad v_{it}(\gamma_i) \equiv \frac{\partial}{\partial \alpha_i} l(\theta, \alpha_i; Z_{it}) \tag{20}$$

and let additional subscripts denote partial derivatives, e.g., $v_{it\alpha_i}(\gamma_i) \equiv \frac{\partial^2}{\partial \alpha_i^2} l(\theta, \alpha_i; Z_{it})$. Let

$$U_{it}(\gamma_i) = u_{it}\left(\gamma_i\right) - \rho_{i0} v_{it}\left(\gamma_i\right) \tag{21}$$

where

$$\rho_{i0} = \frac{\mathbb{E} \sum_{t=1}^T u_{it\alpha_i}\left(\gamma_{i0}\right)}{\mathbb{E} \sum_{t=1}^T v_{it\alpha_i}\left(\gamma_{i0}\right)}. \tag{22}$$

Define

$$\psi_{it}\left(\gamma_{i0}\right) = -\frac{v_{it}\left(\gamma_{i0}\right)}{\mathbb{E}\left[T^{-1} \sum_{t=1}^T v_{it\alpha_i}\left(\gamma_{i0}\right)\right]}. \tag{23}$$

We suppress the arguments of the functions such as u_{it}, v_{it}, and ψ_{it} when they are evaluated at the true value γ_{i0}. For any function $h(\cdot; Z_{it})$ and $p = 1, 2$, we let $\partial^p h(\gamma_i; Z_{it})$ be the matrix of p-th order derivatives of $h(\gamma_i; Z_{it})$ with respect to γ_i. Denote

$$G(\gamma_{i0}, \gamma_i) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\gamma_{i0}} l(\gamma_i; Z_{it}), \tag{24}$$

assuming that the limit exists for each γ_{i0} and $\gamma_i \in \Gamma$.

Let $\mathcal{A}_t^i(\gamma_{i0}) = \sigma\left(Z_{it}(\gamma_{i0}), Z_{it-1}(\gamma_{i0}), \ldots\right)$ and $\mathcal{B}_t^i(\gamma_{i0}) = \sigma\left(Z_{it}(\gamma_{i0}), Z_{it+1}(\gamma_{i0}), \ldots\right)$ be the σ-algebras generated by the respective sequences. Define the strong mixing coefficients

$$\boldsymbol{\alpha}_i(\gamma_{i0}, m) = \sup_t \sup_{\mathbb{A}\in\mathcal{A}_t^i(\gamma_{i0})} \sup_{\mathbb{B}\in\mathcal{B}_{t+m}^i(\gamma_{i0})} \left| P_{\gamma_{i0}}(\mathbb{A}\cap\mathbb{B}) - P_{\gamma_{i0}}(\mathbb{A}) P_{\gamma_{i0}}(\mathbb{B}) \right|. \tag{25}$$

For some $\nu > 0$, let $\Gamma_1 = int\{\gamma_i \in \Gamma : \|\gamma_i - \Gamma_0\| \le \nu\}$ be a slightly larger set than Γ_0, where $\|\gamma_i - \Gamma_0\|$ is the usual Euclidean distance between a point and a set, and $int(S)$ is the interior of set S.

To establish the consistency of $\hat{\alpha}_i$ and $\hat{\theta}_{nT}$, we maintain the following assumptions:

Assumption 1 *(i) $l(\gamma_i; Z_{it})$ is continuous in $\gamma_i \in \Gamma$; (ii) Γ is compact.*

Assumption 2 *For each i, $\{Z_{it}(\gamma_{i0})\}$ is a strong mixing sequence with strong mixing coefficients satisfying*

$$\sup_i \sup_{\gamma_{i0}\in\Gamma_1} \boldsymbol{\alpha}_i(\gamma_{i0}, m) \le C_1 \exp(-C_2 m)$$

for some constants $C_1 \in (0, \infty)$ and $C_2 \in (0, \infty)$.

Assumption 3 *(i) As a function of γ_i, $l(\gamma_i; Z_{it})$ is four-times continuously differentiable;*

(ii) There exists some function $M(Z_{it}) \equiv M(Z_{it}(\gamma_{i0}))$ such that

$$\left| \frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it})}{\partial\gamma_{i,\tau_1}^{m_1} \partial\gamma_{i,\tau_2}^{m_2}} \right| \le M(Z_{it}),$$

for all $\gamma_i \in \Gamma$, $\tau_1, \tau_2 = 1, 2, \ldots, d_\gamma$ and for $m_1, m_2 = 0, \ldots, 4$, and $m_1 + m_2 \le 4$ where γ_{i,τ_1} and γ_{i,τ_2} are the τ_1-th and τ_2-th elements of γ_i, respectively.

(iii) For some $Q > d_\theta + 12$, $\sup_{i,t} \sup_{\gamma_{i0}\in\Gamma_1} \mathbb{E}_{\gamma_{i0}}\{[M(Z_{it}(\gamma_{i0}))]^Q\} < \infty$.

(iv) As $T \to \infty, T^{-1}\sum_{t=1}^{T}\mathbb{E}_{\gamma_{i0}}M(Z_{it})$ *converges to* $\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\mathbb{E}_{\gamma_{i0}}M(Z_{it})$ *uniformly over* $i = 1, 2, \ldots, n$ *and* $\gamma_{i0} \in \Gamma_1$.

Assumption 4 *For each* $\eta > 0$*, there exists* $\delta > 0$ *such that*

$$\inf_{i}\inf_{\gamma_{i0}\in\Gamma_1}\left[G\left(\gamma_{i0},\gamma_{i0}\right) - \sup_{\{\gamma_i:\|\gamma_i-\gamma_{i0}\|>\eta\}} G(\gamma_{i0},\gamma_i)\right] \geq \delta.$$

Assumption 5 $n \to \infty$ *and* $T \to \infty$ *jointly such that* $n/T \to \rho \in (0,\infty)$.

Assumption 1 ensures that the maximization problem is well defined. The mixing and moment conditions in Assumptions 2 and 3 ensure that the ULLN and/or CLT hold for $l(\gamma_i; Z_{it})$ and its derivatives. These conditions are similar to Conditions 3 and 4 in HK who have verified them for some nonlinear panel data models. Assumption 3(iii) maintains that $M\left(Z_{it}\left(\gamma_{i0}\right)\right)$ has enough moments. We note that the existence of higher moments of $M\left(Z_{it}\left(\gamma_{i0}\right)\right)$ is typically assumed in the literature. For example, HN assume that $\mathbb{E}_{\gamma_{i0}}\{\left[M\left(Z_{it}\left(\gamma_{i0}\right)\right)\right]^{64}\} < \infty$ and HK assume that $\mathbb{E}_{\gamma_{i0}}\left(\{\left[M\left(Z_{it}\left(\gamma_{i0}\right)\right)\right]\}^{Q}\right) < \infty$ for $Q > 42$ in the case when $d_{\gamma_i} = 2$. The minimum value of Q in HK grows with the dimension of γ_i with 42 as the lower bound. Assumption 4 is the identification assumption for extremum estimators. It is similar to Condition 3 in HN and Condition 1 in HK. Assumption 5 specifies the asymptotic sequence we consider.

In Assumptions 2–4, we have assumed that the conditions hold uniformly over $\gamma_{i0} \in \Gamma_1$ and i. The reason for the uniformity requirement over $\gamma_{i0} \in \Gamma_1$ is that in the bootstrap world the true parameter value is $\hat{\gamma}_i$, which is random but falls in a shrinking neighborhood around Γ_0 with probability approaching one at a certain rate. So $\hat{\gamma}_i \in \Gamma_1$ with probability approaching one at this rate. We use the uniformity assumption to ensure that the approximation errors in the bootstrap world are small. For more discussion on the uniformity requirement, see Andrews (2005). The uniformity requirement over $i = 1, 2, \ldots, n$ controls the degree of cross sectional heterogeneity. If there is no heterogeneity beyond what is reflected in γ_{i0}, then this uniformity requirement is satisfied trivially.

Theorem 1 *Let Assumptions 1–5 hold. Then for any* $\eta > 0$,

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}\left(\gamma_0\right) - \theta_0\right\| > \eta\right) = o\left(\frac{1}{T}\right) \text{ and } \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i\left|\hat{\alpha}_i\left(\gamma_0\right) - \alpha_{i0}\right| > \eta\right) = o\left(\frac{1}{T}\right),$$

where $\Gamma_1^{\otimes} \equiv \left\{\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)' : \left(\theta', \alpha_i\right)' \in \Gamma_1\right\}$.

The proof of the theorem is based on a modification of standard arguments for the consistency of extremum estimators. The modification is needed as the dimension of the parameter space increases with the cross sectional sample size n. We also need the uniform consistency of $\hat{\alpha}_i$. Pointwise consistency of $\hat{\alpha}_i$ for each i is not sufficient for our stochastic expansion. It is clear from the proof that the theorem still holds if $Q > d_\theta + 12$ in Assumption 3(iii) is replaced by $Q > d_\theta + 5$ but we require the stronger condition to control the remainder term in the stochastic expansion given in Theorem 2.

Let $\bar{u}_{i\cdot\theta}, \bar{v}_{i\cdot\theta}$ and $\bar{v}_{i\cdot\alpha_i}$ be the time series averages of $u_{it\theta}, v_{it\theta}$ and $v_{it\alpha_i}$, respectively. We make additional assumptions to establish a stochastic expansion of our estimator.

Assumption 6 $\gamma_0 = (\theta_0', \alpha_0')'$ *is an interior point in* $\Gamma^{\otimes} = \left\{\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)' : \theta \in \Gamma_\theta, \alpha_i \in \Gamma_\alpha\right\}$.

Assumption 7 *There exist constants* $\delta > 0$ *and* $C > 0$ *such that when* n *and* T *are large enough:*

(i) $\inf_{i,\gamma_{i0} \in \Gamma_1} \left|\mathbb{E}\bar{v}_{i\cdot\alpha_i}(\gamma_{i0})\right| \geq \delta$ *and* $\sup_{i,\gamma_{i0} \in \Gamma_1} \left\|\mathbb{E}\bar{v}_{i\cdot\theta}\left(\gamma_{i0}\right)\right\| \leq C$;

(ii) $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \lambda_{\min}\left\{n^{-1}\sum_{i=1}^n \left(\mathbb{E}\bar{u}_{i\cdot\theta}(\gamma_{i0}) - \left[\mathbb{E}\bar{v}_{i\cdot\theta}\left(\gamma_{i0}\right)\right]' \left[\mathbb{E}\bar{v}_{i\cdot\alpha_i}(\gamma_{i0})\right]^{-1} \left[\mathbb{E}\bar{v}_{i\cdot\theta}\left(\gamma_{i0}\right)\right]\right)\right\} \geq \delta$ *where* $\lambda_{\min}(A)$ *denotes the smallest eigenvalue of* A.

Assumption 6 is standard. Assumption 7 maintains the full rank condition on the information matrix. Under Assumption 7, we can obtain a rate of convergence result for $\hat{\theta}_{nT}$ and $\hat{\alpha}$.

To simplify the presentation, we introduce some new terms and notations. For a random variable $e_{nT}(\gamma_0)$, we say that $e_{nT}(\gamma_0)$ is of order $o_p^U(1)$ and write $e_{nT}(\gamma_0) = o_p^U(1)$ if for any $\eta > 0$, $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\|e_{nT}(\gamma_0)\| > \eta\right) = o(1)$. We say that $e_{nT}(\gamma_0)$ is of order $O_p^U(1)$ and write $e_{nT}(\gamma_0) = O_p^U(1)$ if for any $\varepsilon > 0$, there exists a $K > 0$ such that $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\|e_{nT}(\gamma_0)\| > K\right) \leq \varepsilon$ when n and T are large enough.

Let

$$
\begin{aligned}
H_{nT} &\equiv H_{nT}(\gamma_0) = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \mathbb{E}U_{it,\theta}(\gamma_0), \\
S_{nT} &\equiv S_{nT}(\gamma_0) = \frac{1}{\sqrt{nT}}\sum_{i=1}^n \sum_{t=1}^T U_{it}(\gamma_0),
\end{aligned} \tag{26}
$$

and

$$b_{nT} \equiv b_{nT}(\gamma_0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t}\mathbb{E}\psi_{is}U_{it\alpha_i} + \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}U_{it\alpha_i\alpha_i}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\psi_{it}^2\right). \tag{27}$$

Define

$$A_{nT}(\gamma_0) = -H_{nT}^{-1}(\gamma_0)\,S_{nT}(\gamma_0) \text{ and } B_{nT}(\gamma_0) = -H_{nT}^{-1}(\gamma_0)\,b_{nT}(\gamma_0). \tag{28}$$

Theorem 2 *Let Assumptions 1–7 hold. Then*

$$\begin{aligned}
\hat{\theta}_{nT} - \theta_0 &= \frac{1}{\sqrt{nT}}A_{nT}(\gamma_0) + \frac{1}{T}B_{nT}(\gamma_0) + \frac{1}{T}e_{nT}^{\theta}(\gamma_0), \\
\hat{\alpha}_i - \alpha_{i0} &= \frac{1}{\sqrt{nT}}C_i(\gamma_0) + \frac{1}{\sqrt{T}}D_i(\gamma_0) + \frac{1}{T}E_i(\gamma_0) + \frac{1}{T}e_{nT}^{\alpha_i}(\gamma_0)
\end{aligned}$$

where $e_{nT}^{\theta}(\gamma_0)$ *and* $\max_i \left|e_{nT}^{\alpha_i}(\gamma_0)\right|$ *are of order* $o_p^U(1)$ *and* $A_{nT}(\gamma_0), B_{nT}(\gamma_0),\ C_i(\gamma_0), D_i(\gamma_0)$ *and* $E_i(\gamma_0)$ *are of order* $O_p^U(1)$.

The expressions for $C_i(\gamma_0), D_i(\gamma_0)$ and $E_i(\gamma_0)$ are not important here. They are given in the appendix of proofs. See (A.21). In the special case when $\{Z_{it}\}$ are iid across t, we have

$$b_{nT} = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}\left(\psi_{it}U_{it\alpha_i}\right) + \frac{1}{2}\left(\mathbb{E}U_{it\alpha_i\alpha_i}\right)\left(\mathbb{E}\psi_{it}^2\right)\right]. \tag{29}$$

Let $V_{2it} \equiv v_{it}^2 + v_{it\alpha}$. Using the Bartlett identities, we can show that $b_{nT} = -\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}\left(V_{2it}U_{it}\right)/\mathbb{E}(v_{it}^2)$, which is the same as what HN obtain for iid data.

Theorem 3 *Let Assumptions 1–7 hold. Assume that* $M \to \infty$ *such that* $M = O(T)$. *Then*

(i) $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0} g_M[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)] - \sqrt{n/T}B_{nT}(\gamma_0)\right\| = o(1)$;

(ii) $\left\|\mathbb{E}_{\hat{\gamma}}^{*} g_M[\sqrt{nT}(\hat{\theta}_{nT}^{*} - \hat{\theta}_{nT})] - \sqrt{n/T}B_{nT}(\hat{\gamma})\right\| = o_p(1)$ *uniformly over* $\gamma_0 \in \Gamma_0^{\otimes}$;

(iii) $\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) \xrightarrow{d} N[0, \Omega(\gamma_0)]$ *for each* $\gamma_0 \in \Gamma_0^{\otimes}$ *where*

$$\Gamma_0^{\otimes} := \left\{\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)' : \left(\theta', \alpha_i\right)' \in \Gamma_0\right\} \text{ and } \Omega(\gamma_0) = -\left[\lim_{(n,T)\to\infty} H_{nT}(\gamma_0)\right]^{-1}.$$

Theorem 3(iii) gives the pointwise normal approximation for each $\gamma_0 \in \Gamma_0^{\otimes}$. In the proof, we obtain the following stronger and uniform result: for any $c \in \mathbb{R}^{d_\theta}$ and $\tau \in \mathbb{R}$,

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} \left|P_{\gamma_0}(c'\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0)/\sqrt{c'\Omega(\gamma_0)\,c} < \tau) - \Phi(\tau)\right| = o(1)$$

where $\Phi(\cdot)$ is the standard normal CDF. The above result implies that the asymptotic size of a normal or chi-square test is well controlled.

Theorem 4 *Let the assumptions in Theorem 3 hold. Then for all $k \geq 2$,*

$$\mathbb{E}^*_{\hat{\gamma}} \left\{ g_M[\sqrt{nT}(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT})] - g_M[\sqrt{nT}(\hat{\theta}^*_{nT} - \hat{\theta}_{nT})] \right\} = o_p(1)$$

uniformly over $\gamma_0 \in \Gamma_0^{\otimes}$.

When $k = 1$, $\mathbb{E}^*_{\hat{\gamma}} \left\{ g_M[\sqrt{nT}(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT})] - g_M[\sqrt{nT}(\hat{\theta}^*_{nT} - \hat{\theta}_{nT})] \right\} = O_p(1)$. In this case, the difference between the two bias estimators is large enough to affect the asymptotic distribution. Therefore, condition $k \geq 2$ is necessary for the k-step bootstrap method to achieve effective bias reduction.

Combining Theorems 3 and 4, we get the following theorem immediately.

Theorem 5 *Let the assumptions in Theorem 3 hold. Then for all $k \geq 2$,*

$$\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) \xrightarrow{d} N\left[0, \Omega(\gamma_0)\right].$$

As in Theorem 3(iii), Theorem 5 holds uniformly over $\gamma_0 \in \Gamma_0^{\otimes}$. It provides the usual basis for asymptotic inference. Both asymptotic normal and chi-square inferences can be conducted. As an example, suppose that we are interested in testing $\mathbb{H}_0 : c'\theta_0 = r$ against $\mathbb{H}_1 : c'\theta_0 \neq r$ for some $c \in \mathbb{R}^{d_\theta}$. We construct the t-statistic as follows:

$$t_{nT} \equiv t_{nT}(\gamma_0) = \frac{\sqrt{nT}\left(c'\tilde{\theta}_{nT,k} - r\right)}{\sqrt{c'\hat{\Omega}_{nT}c}} = \frac{\sqrt{nT}c'\left(\tilde{\theta}_{nT,k} - \theta_0\right)}{\sqrt{c'\hat{\Omega}_{nT}c}} \tag{30}$$

where

$$\hat{\Omega}_{nT} \equiv \hat{\Omega}_{nT}\left(\tilde{\theta}_{nT,k}, \tilde{\alpha}_{i,k}\right) = \left(-\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} U_{it\theta}(\tilde{\theta}_{nT,k}, \tilde{\alpha}_{i,k})\right)^{-1}$$

and

$$\tilde{\alpha}_k \equiv \hat{\alpha} - \frac{1}{\sqrt{T}}\mathbb{E}^* g_M\left[\sqrt{T}\left(\hat{\alpha}^*_k - \hat{\alpha}\right)\right] \tag{31}$$

is the k-step BBC estimator of α_0. The consistency of $\tilde{\alpha}_k$ is proved in Lemma A.5 in the Appendix. Using the standard arguments, we can show that $\hat{\Omega}_{nT}$ is a consistent estimator of $\Omega(\gamma_0)$ under Assumptions 2 and 3. So $t_{nT} \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$.

5 Some Extensions

5.1 Bias Correction for Average Marginal Effects

In this subsection, we suggest bias correction for the estimator of an average marginal effect using the k-step bootstrap procedure. In nonlinear models, the average marginal effect may be as interesting as a model parameter because it summarizes the effect over a certain sub-population, which is often the quantity of interest in empirical studies.

There are several different average marginal effects. 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_{i0}. It is defined as[6] $\mu_1(w) = n^{-1}\sum_{i=1}^{n}\Delta(w,\theta_0,\alpha_{i0})$, where w is the value of the covariate vector where the average effect is desired. For example, in a probit model, $\Delta(w,\theta_0,\alpha_{i0}) = \theta_{0(j)}\phi(w'\theta_0+\alpha_{i0})$ 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 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 $\mu_2 = (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(Z_{it},\theta_0,\alpha_{i0})$. See also Fernández-Val (2009). The third average marginal effect, which bridges the first two definitions, can be defined by

$$\mu(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w,\tilde{Z}_{it},\theta_0,\alpha_{i0}), \tag{32}$$

where we set some covariates at the fixed value w and take an average over the rest of covariates $\tilde{Z}_{it}$ and the fixed effects. The third definition includes the first two as special cases. So without loss of generality, we can focus on the third definition.

A natural estimator of $\mu(w)$ is

$$\hat{\mu}(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w,\tilde{Z}_{it},\hat{\theta}_{nT},\hat{\alpha}_i). \tag{33}$$

As in the case for the estimation of model parameters, we construct a BBC estimator of $\mu(w)$ as follows

$$\tilde{\mu}(w) = \hat{\mu}(w) - \frac{1}{\sqrt{nT}}\mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\mu}^*(w)-\hat{\mu}(w)\right)\right] \tag{34}$$

where

$$\hat{\mu}^*(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w,\tilde{Z}^*_{it},\hat{\theta}^*_{nT},\hat{\alpha}^*_i) \tag{35}$$

and the expectation $\mathbb{E}^* g_M(\cdot)$ can be computed by simulation. Similarly, our k-step BBC estimator of $\mu(w)$ is

$$\tilde{\mu}_k(w) = \hat{\mu}(w) - \frac{1}{\sqrt{nT}} \mathbb{E}^* g_M \left[\sqrt{nT} \left(\hat{\mu}_k^*(w) - \hat{\mu}(w) \right) \right] \tag{36}$$

where

$$\hat{\mu}_k^*(w) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta(w, \tilde{Z}_{it}^*, \hat{\theta}_{nT,k}^*, \hat{\alpha}_{i,k}^*). \tag{37}$$

Let

$$L_{nT}(\gamma_0) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E} \left[\Delta_\theta(w, \tilde{Z}_{it}, \gamma_{i0}) - \rho_{i0} \Delta_{\alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0}) \right]. \tag{38}$$

Define

$$A_{nT}^{\mu}(\gamma_0) = L_{nT}'(\gamma_0) A_{nT}(\gamma_0) \tag{39}$$

and

$$\begin{aligned} B_{nT}^{\mu}(\gamma_0) &= \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E} \Delta_\theta'(w, \tilde{Z}_{it}, \gamma_{i0}) B_{nT}(\gamma_0) + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E} \left[\Delta_{\alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0}) E_i(\gamma_0) \right] \\ &\quad + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t,s} \mathbb{E} \left[\Delta_{\alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0}) \psi_{is} \right] + \frac{1}{2nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathbb{E} \left[\Delta_{\alpha_i \alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0}) D_i^2(\gamma_0) \right] \end{aligned} \tag{40}$$

where $\Delta_\theta(\cdot), \Delta_{\alpha_i}(\cdot)$ and $\Delta_{\alpha_i\alpha_i}(\cdot)$ are the partial derivatives of $\Delta(\cdot)$.

Theorem 6 *Let the assumptions in Theorem 3 hold. In addition, assume that (i) $\Delta(w, \tilde{Z}_{it}, \gamma_i)$ is a twice continuously differentiable function in γ_i; (ii) there exists some function $M_\Delta(Z_{it})$ such that $\left| \partial^{m_1+m_2} \Delta(w, \tilde{Z}_{it}, \gamma_i) / \partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2} \right| \le M_\Delta(Z_{it})$, for $m_1, m_2 = 0, \ldots, 3$ and $m_1 + m_2 \le 3$; (iii) $\sup_{i,t} \sup_{\gamma_{i0} \in \Gamma_1} \mathbb{E}_{\gamma_{i0}} \{ [M_\Delta(Z_{it}(\gamma_{i0}))]^6 \} < \infty$. Then*

(i) $\sqrt{nT} [\hat{\mu}(w) - \mu(w)] \xrightarrow{d} N \left[\sqrt{\rho} B^\mu(\gamma_0), \Omega^\mu(\gamma_0) \right]$;

(ii) $\sqrt{nT} [\tilde{\mu}(w) - \mu(w)] \xrightarrow{d} N [0, \Omega^\mu(\gamma_0)]$;

(iii) $\sqrt{nT} [\tilde{\mu}_k(w) - \mu(w)] \xrightarrow{d} N [0, \Omega^\mu(\gamma_0)]$ *for* $k \ge 2$,

where $B^\mu(\gamma_0) = \lim_{(n,T)\to\infty} B_{nT}^\mu(\gamma_0)$ *and* $\Omega^\mu(\gamma_0) = -\lim_{(n,T)\to\infty} L_{nT}'(\gamma_0) H_{nT}^{-1}(\gamma_0) L_{nT}(\gamma_0)$.

The asymptotic bias of $\hat{\mu}(w)$ comes from three different sources. The first term in (40) comes from the asymptotic bias of $\hat{\theta}_{nT}$. The second term in (40) comes from the asymptotic bias of $\hat{\alpha}_i$. The last two terms in (40) capture the nonlinear bias originated from the fact that $\hat{\mu}(w)$ is nonlinear in $\hat{\alpha}$ and that $\hat{\alpha}$ is random.

5.2 Distributional Approximation by Double Bootstrap

The normal approximations in Theorems 5 and 6 may not work very well in finite samples. As an alternative, we can approximate the distributions of $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$ and $\sqrt{nT}\left[\tilde{\mu}_k(w) - \mu(w)\right]$ using double bootstrap. We focus on $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$ and the results can be easily extended to $\sqrt{nT}\left[\tilde{\mu}_k(w) - \mu(w)\right]$.

The double bootstrap sample is generated under the same DGP but using $\hat{\gamma}^* = (\hat{\theta}^{*\prime}_{nT}, \hat{\alpha}^{*\prime})'$ as the model parameter. Define

$$\tilde{\theta}^*_{nT,k} = \hat{\theta}^*_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^{**}_{\hat{\gamma}^*} g_M\left[\sqrt{nT}\left(\hat{\theta}^{**}_{nT,k} - \hat{\theta}^*_{nT}\right)\right] \tag{41}$$

where $\hat{\theta}^{**}_{nT,k}$ is the k-step estimator based on the double bootstrap sample and using $\hat{\gamma}^*$ as the starting point, and $\mathbb{E}^{**}_{\hat{\gamma}^*}$ is the expectation operator with respect to the randomness in the double bootstrap sample, conditional on $\hat{\gamma}^*$. The above definition is entirely analogous to the k-step BBC estimator given in (19). The mechanics behind $\tilde{\theta}_{nT,k}$ and $\tilde{\theta}^*_{nT,k}$ can be illustrated by the following chart:

$$\begin{aligned}
\theta_0 \quad &\mapsto \quad \hat{\theta}_{nT} \mapsto \hat{\theta}^*_{nT,k} \rightarrowtail \tilde{\theta}_{nT,k} \equiv \hat{\theta}_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^*_{\hat{\gamma}} g_M\left[\sqrt{nT}\left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right)\right], \\
\hat{\theta}_{nT} \quad &\mapsto \quad \hat{\theta}^*_{nT} \mapsto \hat{\theta}^{**}_{nT,k} \rightarrowtail \tilde{\theta}^*_{nT,k} \equiv \hat{\theta}^*_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^{**}_{\hat{\gamma}^*} g_M\left[\sqrt{nT}\left(\hat{\theta}^{**}_{nT,k} - \hat{\theta}^*_{nT}\right)\right].
\end{aligned} \tag{42}$$

where, for example, "$\theta_0 \mapsto \hat{\theta}_{nT}$" signifies that $\hat{\theta}_{nT}$ is the MLE based on the sample governed by the model parameter θ_0 (and α_0). We will show in Theorem 7 below that the distribution of $\sqrt{nT}(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT})$ is consistent for that of $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$.

While $\tilde{\theta}^*_{nT,k}$ is the bootstrap analogue of $\tilde{\theta}_{nT,k}$, it involves the standard bootstrap estimator $\hat{\gamma}^*$ which is computationally intensive. To alleviate the computational burden, we use the k-step estimate $\hat{\gamma}^*_k = (\hat{\theta}^*_{nT,k}, \hat{\alpha}^*_k)$ as the model parameter to generate the double bootstrap sample. We call this sample the double k-step bootstrap sample. We define our double k-step BBC estimator as

$$\tilde{\theta}^*_{nT,kk} = \hat{\theta}^*_{nT,k} - \frac{1}{\sqrt{nT}}\mathbb{E}^{**}_{\hat{\gamma}^*_k} g_M\left[\sqrt{nT}\left(\hat{\theta}^{**}_{nT,kk} - \hat{\theta}^*_{nT,k}\right)\right] \tag{43}$$

where $\hat{\theta}^{**}_{nT,kk}$ is the k-step estimator based on the double k-step bootstrap sample and using $\hat{\gamma}^*_k$ as the starting point. Similar to the chart in (42), we can use the chart below to illustrate the mechanics behind $\tilde{\theta}^*_{nT,kk}$ and the corresponding quantity $\tilde{\theta}_{nT,kk}$:

$$\begin{aligned} \theta_0 \quad &\mapsto \quad \hat{\theta}_{nT,k} \mapsto \hat{\theta}^*_{nT,kk} \rightarrowtail \tilde{\theta}_{nT,kk} \equiv \hat{\theta}_{nT,k} - \frac{1}{\sqrt{nT}} \mathbb{E}^*_{\hat{\gamma}_k} g_M \left[\sqrt{nT} \left(\hat{\theta}^*_{nT,kk} - \hat{\theta}_{nT,k} \right) \right], \\ \hat{\theta}_{nT} \quad &\rightarrow \quad \hat{\theta}^*_{nT,k} \rightarrow \hat{\theta}^{**}_{nT,kk} \rightarrowtail \tilde{\theta}^*_{nT,kk} \equiv \hat{\theta}^*_{nT,kk} - \frac{1}{\sqrt{nT}} \mathbb{E}^{**}_{\hat{\gamma}^*_k} g_M \left[\sqrt{nT} \left(\hat{\theta}^{**}_{nT,kk} - \hat{\theta}^*_{nT,k} \right) \right]. \end{aligned}$$

Here $\hat{\theta}_{nT,k}$ is the k-step 'estimator' starting with θ_0 and α_0. It is a theoretical and infeasible object. For this reason, $\hat{\theta}^*_{nT,kk}$ and hence $\tilde{\theta}_{nT,kk}$ are infeasible, but they are useful in analyzing the asymptotic properties of $\tilde{\theta}^*_{nT,kk}$.

The following theorem establishes the consistency of the bootstrap approximation.

Theorem 7 *Let the assumptions in Theorem 3 hold. Let ϑ be a vector in $\mathbb{R}^{d_\theta}$, then for any $\delta > 0$,*
(i) $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0} \left\{ \sup_{\vartheta} \left| P^*_{\hat{\gamma}} \left[\sqrt{nT} (\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT}) < \vartheta \right] - P_{\gamma_0} \left[\sqrt{nT} (\tilde{\theta}_{nT,k} - \theta_0) < \vartheta \right] \right| \geq \delta \right\} = o(1);$
(ii) $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0} \left\{ \sup_{\vartheta} \left| P^*_{\hat{\gamma}} \left[\sqrt{nT} (\tilde{\theta}^*_{nT,kk} - \hat{\theta}_{nT}) < \vartheta \right] - P_{\gamma_0} \left[\sqrt{nT} (\tilde{\theta}_{nT,kk} - \theta_0) < \vartheta \right] \right| \geq \delta \right\} = o(1).$

The proof of the theorem shows that $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$ and $\sqrt{nT}(\tilde{\theta}_{nT,kk} - \theta_0)$ have the same limiting distribution. Therefore the distribution of $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0)$ can be approximated by the distribution of either $\sqrt{nT}(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT})$ or $\sqrt{nT}(\tilde{\theta}^*_{nT,kk} - \hat{\theta}_{nT})$.

Theorem 7 can be used to construct confidence intervals or regions. As an example, let c be a vector in $\mathbb{R}^{d_\theta}$ and suppose that $c'\theta_0$ is the parameter of interest. Let $q_{c,1-\alpha}$ be $1-\alpha$ quantile of $\sqrt{nT}c'(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT})$, i.e., $P^*_{\hat{\gamma}} \left[\sqrt{nT}c'(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT}) \geq q_{c,1-\alpha} \right] = \alpha$. In the appendix, we show that

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0} \left[\sqrt{nT}c'(\tilde{\theta}_{nT,k} - \theta_0) \geq q_{c,1-\alpha} \right] - \alpha = o(1). \tag{44}$$

Similar results hold for the approximation based on the distribution of $\sqrt{nT}(\tilde{\theta}^*_{nT,kk} - \hat{\theta}_{nT})$.

We can also approximate the distribution of t_{nT} defined in (30) using double bootstrap. We define

$$t^*_{nT,k} = \frac{\sqrt{nT}c' \left(\tilde{\theta}^*_{nT,kk} - \hat{\theta}_{nT} \right)}{\sqrt{c'\hat{\Omega}^*_{nT} \left(\tilde{\theta}^*_{nT,kk}, \tilde{\alpha}^*_{kk} \right) c}}. \tag{45}$$

Then under the assumptions in Theorem 3, we have for $\tau \in \mathbb{R}$,

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0} \left\{ \sup_{\tau \in \mathbb{R}} \left| P^*_{\hat{\gamma}} \left(t^*_{nT,k} \left(\hat{\gamma} \right) < \tau \right) - P_{\gamma_0} \left(t_{nT} \left(\gamma_0 \right) < \tau \right) \right| \geq \delta \right\} = o(1).$$

The proof is similar to Theorem 7 and is omitted.

6 Monte Carlo Study

In this section, we report our Monte Carlo results. We consider both the static model and dynamic model. The number of simulation replications is 1000.

6.1 Static Model

We consider the following probit model:

$$\begin{aligned}
&Y_{it} = 1\{X_{it}\theta_0 + \alpha_i - \epsilon_{it} \geq 0\}; \quad \epsilon_{it} \sim N(0,1),\ \alpha_i \sim N(0, 1/10^2), \\
&X_{it} = \beta t + X_{i,t-1}/2 + \nu_{it}; \qquad X_{i0} = \nu_{i0},\ \nu_{it} \sim U(-1/2, 1/2), \\
&n = 100;\ T = 4, 8, 12;\ \theta_0 = 1.
\end{aligned}$$

This design is based on the DGP used in Heckman (1981), Greene (2004), HN, and Fernández-Val (2009). The only difference is that we simulate α_i using $N(0, 1/10^2)$ instead of $N(0,1)$ to reduce the chance of each time series Y_{it} $(t = 1, \ldots, T)$ being constant over time.

If there is no trend component in X_{it}, i.e., $\beta = 0$, the above probit model fits within our framework. While X_{it} is correlated over time, our framework accommodates the conditional MLE with strictly exogenous regressors. On the contrary, even if $\beta = 0$, it does not fit within the theoretical framework employed by HN where (X_{it}, Y_{it}) is assumed to be independent across i and t. As we discuss in Section 4, the bias expression in the presence of serial dependence is different from that with iid data. Thus, their analytic bias corrections are not valid in this case. However, we still employ their estimators here for the purpose of comparison. We note that there is no correlation between X_{it} and α_i in the model, and 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 or not. As an empirical model, Heckman (1981), Greene (2004), HN, and Fernández-Val (2009) consider $\beta = 1/10$. For comparison purposes, we consider $\beta = 1/10$ as well as $\beta = 0$. We report the results for $\beta = 1/10$ only as the results for $\beta = 0$ are qualitatively similar.

The original ML estimator of model parameters is

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

and the estimators of the average marginal effect and overall average marginal effect are

$$\hat{\mu}_1(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{nT} \phi(\bar{X}\hat{\theta}_{nT} + \hat{\alpha}_i), \ \hat{\mu}_2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\theta}_{nT} \phi(X_{it}\hat{\theta}_{nT} + \hat{\alpha}_i),$$

where $\bar{X}$ is the sample mean of $\{X_{it}\}$.

For the k-step bootstrap, we use the fixed regressor bootstrap and use $\hat{\theta}_{nT}$ and $\{\hat{\alpha}_i\}_{i=1}^{n}$ as the true parameters to generate bootstrap samples. We obtain $\hat{\theta}^*_{nT,k}$ using the NR iterative procedure given in (16). We repeat this procedure 999 times ($R = 999$). Then, we construct the bias corrected k-step bootstrap estimator using (19). To implement the bias correction, we need to choose the truncation parameter M. While the rate $M = O(T)$ is theoretically interesting, it does not give us any practical guidance. In practice, we suggest selecting a large M to avoid imposing a too tight restriction on bias estimation. For example, we can set M equal to $C\sqrt{nT}$ with C being an integer multiple of the typical value of the individual model parameters. The multiple can be set equal to the (rounded) average standard error of $\hat{\theta}_{nT}$. In the simulations, we set $M = 1000$. We examine the sensitivity of bootstrap bias estimation to the choice of M. The simulation results, which are not reported for brevity, show that the performances of the BBC estimators do not change across the values of $M = 50, 100, 500, 1000,\ 2000$ in all cases.

We compare the performance of our bias corrected estimator with four alternative bias corrected estimators: the jackknife and analytic bias corrected estimators by HN and the analytic bias corrected estimator by Fernández-Val (2009). The jackknife bias corrected estimator is denoted 'JK'. For the analytic estimators by HN, 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 'BC3'. In this static setting, we ignore the serial dependence and choose the bandwidth to be zero in implementing BC3.

For each estimator, we report its mean, median, standard deviation, root mean squared error, and the empirical size 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 analytic bias correction procedures, the interval estimators or the testing methods are based on normal approximation using the standard errors evaluated at the bias corrected parameter estimators as those given in the respective papers. For the k-step procedure, the CI's are constructed based on the double bootstrap approximation as well as the normal approximation. The double bootstrap CI's are

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

where $T^*_{1-\alpha/2}$ is the $(1-\alpha/2)\times 100\%$ percentile of $|t^*_{nT,k}|$ defined in (45). We also employ the double bootstrap procedure to construct CI's for the average marginal effects. The standard errors for the bias corrected estimators of the average marginal effects and the corresponding bias corrected estimators in the bootstrap world are evaluated at the bias uncorrected parameter estimator $(\hat{\theta}_{nT}, \hat{\alpha})$ and its bootstrap version $(\hat{\theta}^*_{nT,k}, \hat{\alpha}^*_k)$ respectively. In the simulation experiment, we set the number of double bootstrap iterations to be 100. We do so in order to reduce the computational burden. In empirical applications, we should use a larger number.

Table 1 reports the performance of the standard and k-step BBC estimators. From this table, we see that there is no sacrifice of accuracy by using the k-step bootstrap procedure in this setting. When $k \geq 2$, the k-step bootstrap bias correction tends to reduce the bias significantly as the standard bootstrap does. Results not reported here show that the one-step procedure is not effective in bias reduction. This result is consistent with our Theorem 5, which shows that the order of bias is reduced from $O(1/T)$ to $o(1/T)$ when $k \geq 2$. As discussed before, for each k value, we can use either the observed Hessian or expected Hessian in the NR steps, leading to two versions of the k-step procedure. In terms of the MSE, the 2-step bootstrap estimators with expected and observed Hessians are the best when $T = 4$. When $T = 8$ and 12, there is little difference in performances among the standard and k-step BBC estimators.

Table 2 compares different bias correction methods. We choose the 2-step bootstrap with observed Hessian as our benchmark. First, we see that the original 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 original estimator is 35%. When $T = 12$, the bias is reduced to 14%. But the empirical null rejection probability is 35.5% for the two-sided 5% test, which is even more inaccurate than the one with $T = 4$. Second, the k-step bootstrap performs well in finite samples compared to the other methods. In particular, when $T = 4$, the bias of our estimator is 9% and RMSE is 0.213, while the bias of the jackknife method is 24% and its RMSE is 0.319. The analytic method by Fernández-Val (2009) has a bias of 4%, but its RMSE is 0.240, which implies that its variance is larger than ours. As T becomes large, all the bias corrected estimators we consider tend to have similar RMSEs. Third, in terms of coverage accuracy, we see that our bootstrap bias corrections (double k-step bootstrap approximation and normal approximation) and BC3 yield more accurate CI's than the other methods.

Table 3 presents the ratio of the estimators of the average marginal effect to the true value. As HN and Fernández-Val (2009) show, the bias of the original estimator is negligible, even when $T = 4$. Its bias is less than 1%, and in terms of RMSE, it performs as well as the bias corrected ones. However, the CI's based on the bias uncorrected estimator are inaccurate when we have a 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. The CI's using normal approximations with the bias corrected estimators still suffer from large coverage errors. When $T = 4$, the errors in coverage probability for the 95% CI from jackknife and analytic estimators are 13% and 4–7% respectively. It is 4.6% for the k-step BBC estimator with normal approximation. In contrast, the coverage error of the 95% CI constructed using the double k-step bootstrap is 2.7%. Compared with the normal approximation, the improvement from the double k-step bootstrap approximation is not large but exceeds the margin of simulation error.

To save space, we do not report the table for the overall average marginal effect, but we note that the qualitative observations for Table 3 remain valid.

6.2 Dynamic Model

We consider panel logit and probit models for the dynamic case, but we report only the results for the panel logit model here. Using the simulation design of HK, we have

$$
\begin{aligned}
&Y_{i0} = 1\{\alpha_i + X_{i0}'\theta_X - \epsilon_{it} \geq 0\}, Y_{it} = 1\{\alpha_i + X_{it}'\theta_X + Y_{it-1}\theta_Y - \epsilon_{it} \geq 0\},\\
&X_{it} \sim N\left(0, (\pi^2/3) \cdot I_{\dim(X_{it})}\right), \epsilon_{it} \overset{iid}{\sim} \mathcal{L}(0, \pi^2/3),\\
&\dim(X_{it}) = 1, 2, \quad \alpha_i = (X_{i0}^{(1)} + X_{i1}^{(1)} + X_{i2}^{(1)} + X_{i3}^{(1)})/4,\\
&i = 1, \ldots, n, \quad t = 1, \ldots, T-1; \quad n = 250; \quad T = 8, 16; \quad \theta_Y = 0.5, \quad \theta_X = 1, [1, 1]',
\end{aligned}
$$

where $\mathcal{L}$ denotes the standardized logistic distribution and $X_{it}^{(1)}$ is the first component of X_{it}. We report the bias and RMSE of θ_Y and θ_X for each estimator.

Table 4 presents the simulation results. Here we employ the analytic bias corrections by Hahn and Kuersteiner (2011, HK) and Fernández-Val (2009), which are denoted by 'BC1' and 'BC3' respectively. We also consider the bias correction by Honoré-Kyriazidou (2000). For BC1 and Honoré-Kyriazidou, we replicate the Monte Carlo results in Hahn and Kuersteiner (2011). In contrast to our bootstrap bias correction, the alternative methods depend on bandwidth choices. HK use the bandwidth 8 for Honoré-Kyriazidou, and use a bandwidth around 1 for BC1. In our simulations, we set the bandwidth to 1 to implement BC3. Table 4 shows that our bias correction performs best in an overall sense. Honoré-Kyriazidou performs well only when the dimension of X_{it} is low. The performance of BC3 is comparable to our approach for the two sample sizes given in the table. Simulations results not reported here show that BC3 does not perform as well as our approach when the time series sample size reduces to $T = 4$.

7 Conclusion

In this paper, we establish the asymptotic validity of parametric bootstrap bias correction for fixed effects estimators in nonlinear panel data models. In particular, we propose using the k-step bootstrap procedure to alleviate the computational cost of implementing the standard bootstrap and show that it is asymptotically equivalent to the standard parametric BBC estimator when $k \geq 2$. We also apply the k-step bootstrap procedure to average marginal effect estimation and to the double bootstrap for CI construction. In the simulation, we show that the k-step bootstrap bias correction achieves substantial bias reduction. The CI's based on the double k-step bootstrap tend to have smaller coverage errors than the other CI's especially when T is small.

The possible higher order refinement of double bootstrap CI is not studied here, which is an interesting topic for future research. Also, the fact that we employ the parametric bootstrap limits the applicability of our bias correction procedures. It will be interesting in future to consider a nonparametric approach to bootstrap bias correction.

List of Some Notations

$\gamma_i = \left(\theta', \alpha_i\right)' \in \Gamma \equiv \Gamma_\theta \times \Gamma_\alpha$

$\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)'$

$\gamma_{i0} = \left(\theta_0', \alpha_{i0}\right)' \in \Gamma_0$

$\gamma_0 = \left(\theta_0', \alpha_{10}, \ldots, \alpha_{n0}\right)'$

$\hat{\alpha}\left(\theta\right) = \left(\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_i(\theta), \ldots, \hat{\alpha}_n(\theta)\right)'$

$\hat{\alpha} = \left(\hat{\alpha}_1, \ldots, \hat{\alpha}_i, \ldots, \hat{\alpha}_n\right)'$

$\hat{\gamma}_i = (\hat{\theta}_{nT}', \hat{\alpha}_i)'$, $\hat{\gamma} = (\hat{\theta}_{nT}', \hat{\alpha}')'$

$\hat{\alpha}_i = \hat{\alpha}_i(\hat{\theta}_{nT})$

$\Gamma_1 = int\left\{\gamma_i \in \Gamma : \|\gamma_i - \Gamma_0\| \leq \nu\right\}$

$\Gamma_1^{\otimes} = \left\{\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)' : \left(\theta', \alpha_i\right)' \in \Gamma_1\right\}$

$\Gamma^{\otimes} = \left\{\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)' : \left(\theta', \alpha_i\right)' \in \Gamma\right\}$

$\Gamma_0^{\otimes} = \left\{\gamma = \left(\theta', \alpha_1, \ldots, \alpha_n\right)' : \left(\theta', \alpha_i\right)' \in \Gamma_0\right\}$

Table 1: Finite Sample Performance of Standard and k-step BBC Estimators ($n = 100$; $\theta_0 = 1$)

Estimator	Mean	Median	SD	RMSE
		$T = 4$		
k=2, E	0.97	0.97	0.206	0.208
k=2, O	0.91	0.91	0.193	0.213
k=3, E	0.84	0.84	0.164	0.232
k=3, O	0.82	0.82	0.158	0.241
Standard	0.81	0.81	0.157	0.249
		$T = 8$		
k=3, E	0.96	0.96	0.099	0.106
Standard	0.96	0.96	0.099	0.107
k=2, E	1.00	1.00	0.107	0.107
k=2, O	0.98	0.97	0.106	0.109
k=3, O	0.96	0.95	0.101	0.110
		$T = 12$		
Standard	0.98	0.97	0.073	0.076
k=3, O	0.98	0.97	0.073	0.077
k=3, E	0.98	0.98	0.075	0.077
k=2, O	1.00	1.00	0.079	0.079
k=2, E	1.01	1.01	0.078	0.079

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 ($n = 100$; $\theta_0 = 1$)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	1.35	1.33	0.311	0.265	0.363	0.468
2-step-BBC	0.91	0.91	0.193	0.038	0.073	0.213
				(0.032)	(0.071)	
JK	0.76	0.76	0.215	0.112	0.195	0.319
BC1	1.09	1.07	0.262	0.054	0.103	0.276
BC2	1.17	1.15	0.284	0.094	0.166	0.331
BC3	1.04	1.03	0.237	0.030	0.068	0.240
			$T = 8$			
Probit	1.16	1.15	0.132	0.260	0.357	0.209
2-step-BBC	0.98	0.97	0.106	0.040	0.083	0.109
				(0.051)	(0.087)	
JK	0.95	0.94	0.108	0.059	0.111	0.120
BC1	1.04	1.04	0.121	0.068	0.126	0.129
BC2	1.05	1.04	0.120	0.065	0.118	0.128
BC3	1.01	1.01	0.113	0.040	0.085	0.114
			$T = 12$			
Probit	1.14	1.14	0.096	0.355	0.477	0.171
2-step-BBC	1.00	1.00	0.079	0.036	0.078	0.079
				(0.040)	(0.091)	
JK	0.97	0.97	0.080	0.074	0.129	0.086
BC1	1.04	1.04	0.089	0.088	0.148	0.098
BC2	1.03	1.03	0.087	0.063	0.126	0.092
BC3	1.01	1.01	0.083	0.044	0.102	0.084

Notes: 'Probit' denotes the bias uncorrected fixed effects MLE based on the probit model; 'JK' 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; p;.05 and p;.10 denote empirical rejection probabilities of two-sided nominal 5% and 10% tests. The numbers in the parentheses indicate empirical rejection probabilities of the 2-step bootstrap bias corrected test using the normal approximation.

Table 3: Finite Sample Performance of Different Bias Corrected Estimators of the Average Marginal Effect ($n = 100$; true value = 1)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	0.99	0.98	0.218	0.103	0.184	0.218
2-step-BBC	0.99	0.98	0.219	0.077	0.132	0.219
				(0.096)	(0.164)	
JK	1.08	1.06	0.273	0.181	0.250	0.283
BC1	0.99	0.99	0.233	0.115	0.184	0.233
BC2	1.03	1.03	0.227	0.104	0.178	0.229
BC3	0.93	0.93	0.202	0.095	0.156	0.214
			$T = 8$			
Probit	1.00	1.00	0.105	0.068	0.122	0.105
2-step-BBC	0.99	0.99	0.104	0.051	0.110	0.105
				(0.051)	(0.103)	
JK	1.00	0.99	0.107	0.071	0.126	0.107
BC1	1.01	1.01	0.108	0.073	0.137	0.108
BC2	1.00	1.00	0.105	0.067	0.124	0.105
BC3	0.98	0.98	0.102	0.075	0.122	0.104
			$T = 12$			
Probit	1.01	1.01	0.070	0.065	0.121	0.070
2-step-BBC	0.99	0.99	0.069	0.064	0.109	0.069
				(0.046)	(0.096)	
JK	0.99	0.98	0.072	0.065	0.128	0.074
BC1	0.99	0.98	0.070	0.058	0.111	0.072
BC2	0.99	0.99	0.070	0.053	0.106	0.070
BC3	0.98	0.98	0.069	0.060	0.120	0.072

Notes: See notes to table 2.

Table 4: Finite Sample Performance of Different Bias Corrected Estimators of the θ_Y and θ_X ($n = 250$; $\theta_Y = 0.5$,$\theta_X = 1, [1, 1]'$)

	$\dim(X_{it}) = 1$				$\dim(X_{it}) = 2$			
	θ_Y		θ_X		θ_Y		$\theta_X^{(1)}$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$T = 8$								
Logit	-0.75	0.77	0.26	0.27	-0.69	0.72	0.33	0.34
2-step-BBC	0.07	0.17	0.01	0.06	0.10	0.20	0.00	0.07
BC1	-0.25	0.29	0.08	0.11	-0.24	0.30	0.09	0.13
BC3	-0.05	0.15	-0.02	0.06	-0.06	0.17	0.10	0.14
Honoré-Kyriazidou	-0.07	0.18	0.02	0.08	-0.29	0.32	-1.09	1.10
$T = 16$								
Logit	-0.30	0.31	0.10	0.11	-0.28	0.31	0.12	0.13
2-step-BBC	0.02	0.09	0.00	0.04	0.02	0.10	-0.01	0.04
BC1	-0.06	0.11	0.01	0.04	-0.06	0.12	0.02	0.05
BC3	-0.01	0.09	0.00	0.04	-0.02	0.10	0.00	0.04
Honoré-Kyriazidou	-0.08	0.13	0.00	0.04	-0.49	0.26	-1.04	1.04

Notes: 'Logit' denotes the bias uncorrected fixed effects MLE based on the logit model; 'BC1' denotes HK bias corrected estimator; 'BC3' denotes Fernández-Val (2009) bias corrected estimator; 'Honoré-Kyriazidou' denotes Honoré-Kyriazidou (2000) bias corrected estimator. $\theta_X^{(1)}$ denotes the first element of θ_X.

Appendix of Proofs

Proof of Theorem 1. We first prove the result that $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}\left(\gamma_0\right)-\theta_0\right\|>\eta\right)=o\left(T^{-1}\right)$. For η and δ given in Assumption 4, we have

$$
\begin{aligned}
& P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}\left(\gamma_0\right)-\theta_0\right\| \geq \eta\right) \leq P_{\gamma_0}\left(\left\|(\hat{\theta}_{nT}^{\prime}, \hat{\alpha}_i)^{\prime}-(\theta_0^{\prime}, \alpha_{i0})^{\prime}\right\|>\eta \text{ for } i=1,2, \ldots, n\right) \\
= \quad & P_{\gamma_0}\left(\left\|\hat{\gamma}_i-\gamma_{i0}\right\|>\eta \text{ for } i=1,2, \ldots, n\right) \\
\leq \quad & P_{\gamma_0}\left(G\left(\gamma_{i0}, \gamma_{i0}\right)-G(\gamma_{i0}, \hat{\gamma}_i) \geq \delta \text{ for } i=1,2, \ldots, n\right) \\
\leq \quad & P_{\gamma_0}\left(\frac{1}{n} \sum_{i=1}^n\left[G\left(\gamma_{i0}, \gamma_{i0}\right)-G\left(\gamma_{i0}, \hat{\gamma}_i\right)\right] \geq \delta\right). \qquad \text{(A.1)}
\end{aligned}
$$

Under Assumptions 1(ii) and 3, we have

$$
\sup_i \sup_{\gamma_i \in \Gamma}\left|G\left(\gamma_{i0}, \gamma_i\right)-\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\gamma_{i0}} l\left(\gamma_i ; Z_{it}\right)\right|=o(1)
$$

as $T \rightarrow \infty$. Using this and the definition of $\hat{\gamma}_i$, we have

$$
\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}-\theta_0\right\| \geq \eta\right) \\
\leq \quad & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\frac{1}{n} \sum_{i=1}^n G\left(\gamma_{i0}, \gamma_{i0}\right)-\frac{1}{nT} \sum_{i,t} l(\gamma_{i0} ; Z_{it})+\frac{1}{nT} \sum_{i,t} l(\hat{\gamma}_i ; Z_{it})-\frac{1}{n} \sum_{i=1}^n G\left(\gamma_{i0}, \hat{\gamma}_i\right) \geq \delta\right) \\
\leq \quad & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma_1 \in \Gamma} \ldots \sup_{\gamma_n \in \Gamma}\left|\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T\left[l(\gamma_i ; Z_{it})-\mathbb{E}_{\gamma_{i0}} l(\gamma_i ; Z_{it}\left(\gamma_{i0}\right))\right]\right| \geq \frac{\delta}{3}\right)+o(\frac{1}{T}) \\
\leq \quad & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_i \sup_{\gamma_i \in \Gamma}\left|\frac{1}{T} \sum_{t=1}^T\left[l(\gamma_i ; Z_{it})-\mathbb{E}_{\gamma_{i0}} l(\gamma_i ; Z_{it}\left(\gamma_{i0}\right))\right]\right| \geq \frac{\delta}{3}\right)+o(\frac{1}{T}) \qquad \text{(A.2)} \\
\leq \quad & \sum_{i=1}^n \sup_{\gamma_{i0} \in \Gamma_1} P_{\gamma_{i0}}\left(\sup_{\gamma_i \in \Gamma}\left|\frac{1}{T} \sum_{t=1}^T \tilde{l}_{it}(\gamma_{i0}, \gamma_i)\right| \geq \frac{\delta}{3}\right)+o(\frac{1}{T})
\end{aligned}
$$

for $\tilde{l}_{it}(\gamma_{i0}, \gamma_i) \equiv l(\gamma_i ; Z_{it})-\mathbb{E}_{\gamma_{i0}} l\left(\gamma_i ; Z_{it}\left(\gamma_{i0}\right)\right)$. Since Γ is compact, it can be covered by a union of balls $\cup_{j=1,\ldots,J(\delta)}\{\mathcal{B}(\gamma_i^{(j)}, \delta/\sqrt{T})\}$ where the number of balls is $J\left(\delta\right)=O((\sqrt{T})^{d_\theta+1})$. We have

$$
\begin{aligned}
& P_{\gamma_{i0}}\left(\sup_{\gamma_i\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0},\gamma_i)\right|\geq\frac{\delta}{3}\right) \\
\leq\ & P_{\gamma_{i0}}\left(\max_j \sup_{\gamma_i\in\mathcal{B}(\gamma_i^{(j)},\delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right]\right|\geq\frac{\delta}{6}\right) \\
& +\sum_{j=1}^{J(\delta)}P_{\gamma_{i0}}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right|\geq\frac{\delta}{6}\right). \qquad \text{(A.3)}
\end{aligned}
$$

Under Assumption 3, we have

$$
\begin{aligned}
\left|\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right| &\leq C\left[M\left(Z_{it}(\gamma_{i0})\right)+\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\left\|\gamma_i-\gamma_i^{(j)}\right\| \\
&\leq C\delta\left[M\left(Z_{it}(\gamma_{i0})\right)+\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]/\sqrt{T},
\end{aligned}
$$

and so

$$
\max_j \sup_{\gamma_i\in\mathcal{B}(\gamma_i^{(j)},\delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right]\right|\leq\frac{C\delta}{T\sqrt{T}}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)+\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right].
$$

As a result,

$$
\begin{aligned}
& P_{\gamma_{i0}}\left(\max_j \sup_{\gamma_i\in\mathcal{B}(\gamma_i^{(j)},\delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right]\right|\geq\frac{\delta}{6}\right) \\
\leq\ & P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\frac{1}{T}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)+\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\geq\frac{\delta}{6}\right) \\
\leq\ & P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)-\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\right|\geq\frac{\delta}{12}\right) \\
& +P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\left|\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right|\geq\frac{\delta}{12}\right). \qquad \text{(A.4)}
\end{aligned}
$$

Since $\frac{C\delta}{\sqrt{T}}\left|\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right|$ is deterministic and approaches zero, the second term in the above equation goes to zero at an arbitrarily slow rate. For the first term, we use Assumptions 2, 3, and the strong mixing moment inequality of Yokoyama (1980) and Doukhan (1995, Theorem

2 and Remark 2, pp. 25–30) to obtain

$$\mathbb{E}_{\gamma_{i0}}\left|\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)-\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\right|^{p}\leq CT^{p/2}$$

where C is a constant that does not depend on γ_{i0} and $p\geq 2$. Invoking the Markov's inequality, we have

$$\sup_{\gamma_{i0}\in\Gamma_1}P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)-\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\right|\geq\frac{\delta}{12}\right)=O\left(\frac{T^{p/2}}{T^{3p/2}}\right)=O\left(T^{-p}\right).$$

Hence

$$\sup_{\gamma_{i0}\in\Gamma_1}P_{\gamma_{i0}}\left(\max_{j}\sup_{\gamma_i\in\mathcal{B}(\gamma_i^{(j)},\delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right]\right|\geq\frac{\delta}{6}\right)=O\left(T^{-p}\right). \quad \text{(A.5)}$$

Using the strong mixing moment inequality again, we have

$$\sup_{\gamma_{i0}\in\Gamma_1}\sum_{j=1}^{J(\delta)}P_{\gamma_{i0}}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right|\geq\frac{\delta}{6}\right)=O\left(T^{-\frac{p}{2}}\right)O\left(T^{\frac{d_\theta+1}{2}}\right)=O\left(T^{-\frac{p-(d_\theta+1)}{2}}\right).$$

This, combined with (A.3) and (A.5), yields

$$\sup_{\gamma_{i0}\in\Gamma_1}P_{\gamma_{i0}}\left(\sup_{\gamma_i\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0},\gamma_i)\right|\geq\frac{\delta}{3}\right)=O\left(T^{-\frac{p-(d_\theta+1)}{2}}\right). \quad \text{(A.6)}$$

Therefore $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}\left(\gamma_0\right)-\theta_0\right\|\geq\eta\right)=O(nT^{-\frac{p-(d_\theta+1)}{2}})+o(T^{-1})$. Letting $p\in(d_\theta+5,Q)$, we have $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}\left(\gamma_0\right)-\theta_0\right\|\geq\eta\right)=o\left(T^{-1}\right)$.

Next, we prove the second part of the result. For any $i=1,2,\ldots,n$, we have

$$\begin{aligned}
& P_{\gamma_0}(\max_i|\hat{\alpha}_i\left(\gamma_0\right)-\alpha_{i0}|>\eta)\\
\leq\ & P_{\gamma_0}\left(\max_i\left\{G\left(\gamma_{i0},\gamma_{i0}\right)-G(\gamma_{i0},\hat{\gamma}_i)\right\}\geq\delta\right)\\
\leq\ & P_{\gamma_0}\left(\max_i\left\{G\left(\gamma_{i0},\gamma_{i0}\right)-G\left(\gamma_{i0},(\hat{\theta}_{nT}',\alpha_{i0})'\right)\right\}\geq\delta/2\right)\\
& +P_{\gamma_0}\left(\max_i\left\{G\left(\gamma_{i0},(\hat{\theta}_{nT}',\alpha_{i0})'\right)-G(\gamma_{i0},\hat{\gamma}_i)\right\}\geq\delta/2\right). \quad \text{(A.7)}
\end{aligned}$$

But for $\check{\theta}_{nT}$ between θ_0 and $\hat{\theta}_{nT}$,

$$
\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \left\{G\left(\gamma_{i0}, \gamma_{i0}\right) - G\left(\gamma_{i0}, (\hat{\theta}_{nT}', \alpha_{i0})'\right)\right\} \geq \delta/2\right) \\
\leq\ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \left\| \frac{\partial G(\gamma_{i0}, (\check{\theta}_{nT}', \alpha_{i0})')}{\partial \theta} \right\| \left\| \hat{\theta}_{nT} - \theta_0 \right\| \geq \delta/2\right) \\
\leq\ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \sup_{\gamma_{i0} \in \Gamma_1} C \mathbb{E}_{\gamma_{i0}} M(Z_{it}\left(\gamma_{i0}\right)) \left\| \hat{\theta}_{nT} - \theta_0 \right\| \geq \delta/4\right) + o\left(1/T\right) \\
=\ & o(1/T).
\end{aligned}
$$

In addition, using (A.6), we obtain

$$
\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \left\{G\left(\gamma_{i0}, (\hat{\theta}_{nT}', \alpha_{i0})'\right) - G(\gamma_{i0}, \hat{\gamma}_i)\right\} \geq \delta/2\right). \\
=\ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \left\{G\left(\gamma_{i0}, (\hat{\theta}_{nT}', \alpha_{i0})'\right) - \frac{1}{T}\sum_{t=1}^{T} l(\hat{\gamma}_i; Z_{it}) + \frac{1}{T}\sum_{t=1}^{T} l(\hat{\gamma}_i; Z_{it}) - G(\gamma_{i0}, \hat{\gamma}_i)\right\} \geq \delta/2\right) \\
\leq\ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \left\{G\left(\gamma_{i0}, (\hat{\theta}_{nT}', \alpha_{i0})'\right) - \frac{1}{T}\sum_{t=1}^{T} l(\hat{\theta}_{nT}, \alpha_{i0}; Z_{it})\right.\right. \\
& \left.\left. + \frac{1}{T}\sum_{t=1}^{T} l(\hat{\theta}_{nT}, \hat{\alpha}_i; Z_{it}) - G(\gamma_{i0}, \hat{\gamma}_i)\right\} \geq \delta/2\right) \\
\leq\ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \sup_{\gamma_i \in \Gamma} \left| \frac{1}{T}\sum_{t=1}^{T} l(\gamma_i; Z_{it}) - G(\gamma_{i0}, \gamma_i) \right| \geq \delta/4\right) \\
=\ & \sum_{i=1}^{n} \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma_i \in \Gamma} \left| \frac{1}{T}\sum_{t=1}^{T} \tilde{l}_{it}(\gamma_{i0}, \gamma_i) \right| \geq \delta/8\right) + o\left(\frac{1}{T}\right) = O\left(nT^{-(\frac{p}{2} - \frac{d_\theta + 1}{2})}\right) + o\left(\frac{1}{T}\right). \qquad (A.8)
\end{aligned}
$$

If we choose $p \in (d_\theta + 5, Q)$, then $nT^{-(\frac{p}{2} - \frac{d_\theta + 1}{2})} = o(T^{-1})$, and as a result

$$
\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}(\max_i |\hat{\alpha}_i - \alpha_{i0}| > \eta) = o(1/T) \qquad (A.9)
$$

for any $\eta > 0$. ■

To prove Theorem 2, we first establish a lemma, which employs the convention that

$$\begin{aligned}
\frac{\partial^{m_1+m_2} l(\tilde{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} &= \frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} |_{\gamma_i = \tilde{\gamma}_i}, \\
\mathbb{E} \frac{\partial^{m_1+m_2} l(\tilde{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} &= \left[\mathbb{E} \frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right]_{\gamma_i = \tilde{\gamma}_i}.
\end{aligned}$$

Lemma A.1 *Let Assumptions 1–7 hold. In addition, let $m_1, m_2 = 0, \ldots, 4$ such that $m_1 + m_2 \leq 4$, and let $\breve{\gamma} = (\breve{\theta}', \breve{\alpha}')'$ satisfy*

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\left\| \breve{\theta} - \theta_0 \right\| > \eta \right) = o(1/T), \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\max_i |\breve{\alpha}_i - \alpha_{i0}| > \eta \right) = o(1/T)$$

for any $\eta > 0$. Then

(i) for any $\delta > 0$,

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\left| \sup_i \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial^{m_1+m_2} l(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} - \mathbb{E} \frac{\partial^{m_1+m_2} l(\gamma_{i0}; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right] \right| > \delta \right) = o(1/T).$$

(ii) there exists $K < \infty$ such that

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\sup_i \frac{1}{T} \sum_{t=1}^T \left| \frac{\partial^{m_1+m_2} l(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right| > K \right) = o(1/T).$$

Proof of Lemma A.1. Part (i). Given the assumptions on $\breve{\gamma}$, we have $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P(\breve{\gamma}_i \notin \Gamma_1) = o(1/T)$. It thus suffices to focus on the event that $\breve{\gamma}_i \in \Gamma_1$. Note that for $\gamma_i^{\dagger}$ between $\breve{\gamma}_i$ and γ_{i0},

$$\begin{aligned}
& \left| \sup_i \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial^{m_1+m_2} l(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} - \mathbb{E} \frac{\partial^{m_1+m_2} l(\gamma_{i0}; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right] \right| \\
\leq\ & \sup_i \sup_{\gamma_i \in \Gamma_1} \left| \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} - \mathbb{E} \frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right] \right| \\
& + \sup_i \frac{1}{T} \sum_{t=1}^T \sum_{\tau_3=1}^{d_{\gamma_i}} \mathbb{E} \left| \frac{\partial^{m_1+m_2+1} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2} \partial \gamma_{i,\tau_3}} \right|_{\gamma_i = \gamma_i^{\dagger}} \left| \breve{\gamma}_{i,\tau_3} - \gamma_{i0,\tau_3} \right| \\
\leq\ & \sup_i \sup_{\gamma_i \in \Gamma_1} \left| \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} - \mathbb{E} \frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right] \right| \\
& + d_{\gamma_i} \sup_{i,t} \sup_{\gamma_{i0} \in \Gamma_1} \mathbb{E} M(Z_{it}(\gamma_{i0})) \max_{\tau_3} \left| \breve{\gamma}_{i,\tau_3} - \gamma_{i0,\tau_3} \right| \equiv \mathcal{T}_1 + \mathcal{T}_2.
\end{aligned}$$

Under Assumption 3, the second term $\mathcal{T}_2$ satisfies $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\mathcal{T}_2 > \frac{\delta}{2}\right) = o\left(1/T\right)$. It remains to show that

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{\max_i \sup_{\gamma_i\in\Gamma_1} |\mathcal{T}_1| > \frac{\delta}{2}\right\} = o\left(\frac{1}{T}\right).$$

But this can be proved using exactly the same argument in showing

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_i \sup_{\gamma_i\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}\left[l(\gamma_i;Z_{it}\left(\gamma_{i0}\right)) - \mathbb{E}l(\gamma_i;Z_{it}\left(\gamma_{i0}\right))\right]\right| \geq \frac{\delta}{2}\right) = o(\frac{1}{T})$$

in the proof of Theorem 1. See (A.2). The differences are (i) $l(\gamma_i;Z_{it}\left(\gamma_{i0}\right))$ and $\mathbb{E}l(\gamma_i;Z_{it}\left(\gamma_{i0}\right)$ are now replaced by $\partial^{m_1+m_2}l(\gamma_i;Z_{it}\left(\gamma_{i0}\right))/\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}$ and its expectation (ii) the sup inside the probability is taken over a smaller set $\Gamma_1\subset\Gamma$. In view of Assumptions 2 and 3, this replacement does not cause any problem in our argument.

Part (ii). Again, we focus on the event that $\breve{\gamma}_i\in\Gamma_1$, which happens with probability $1-o\left(1/T\right)$. Observing that

$$\frac{1}{T}\sum_{t=1}^{T}\left|\frac{\partial^{m_1+m_2}l(\breve{\gamma}_i;Z_{it}\left(\gamma_{i0}\right))}{\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}}\right| \leq \frac{1}{T}\sum_{t=1}^{T} M(Z_{it}\left(\gamma_{i0}\right)),$$

we have, for $K > \lim_{T\to\infty}\max_i T^{-1}\sum_{t=1}^{T}\mathbb{E}M(Z_{it}\left(\gamma_{i0}\right))$,

$$\begin{aligned}
&\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\max_i \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\partial^{m_1+m_2}l(\breve{\gamma}_i;Z_{it}\left(\gamma_{i0}\right))}{\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}}\right| > K\right)\\
\leq\ & \sum_{i=1}^{n}\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\frac{1}{T}\sum_{t=1}^{T}\left[M(Z_{it}\left(\gamma_{i0}\right)) - \mathbb{E}M(Z_{it}\left(\gamma_{i0}\right))\right] > K - \max_i\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}M(Z_{it}\left(\gamma_{i0}\right))\right)\\
\leq\ & \sum_{i=1}^{n}\sup_{\gamma_0\in\Gamma_1^{\otimes}}\frac{\mathbb{E}\left\{\sum_{t=1}^{T}\left[M(Z_{it}\left(\gamma_{i0}\right)) - \mathbb{E}M(Z_{it}\left(\gamma_{i0}\right))\right]\right\}^6}{T^6\left[K-\max_i\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}M(Z_{it}\left(\gamma_{i0}\right)\right]^6}\\
=\ & O\left(\frac{n}{T^3}\right) = O\left(\frac{1}{T^2}\right) = o\left(\frac{1}{T}\right).
\end{aligned}\tag{A.10}$$

■

Proof of Theorem 2. Following the arguments similar to HK(2011), we can establish the following representations

$$\begin{aligned}\hat{\theta}_{nT}-\theta_0 &= \frac{1}{\sqrt{nT}}\tilde{A}_{nT}+\frac{1}{T}\tilde{B}_{nT}+o_p^U\left(\frac{1}{T}\right)\\ \hat{\alpha}_i-\alpha_{i0} &= \frac{1}{\sqrt{nT}}\tilde{C}_i+\frac{1}{\sqrt{T}}\tilde{D}_i+\frac{1}{T}\tilde{E}_i+o_p^U\left(\frac{1}{T}\right)\end{aligned} \tag{A.11}$$

where $\tilde{A}_{nT}$, $\tilde{B}_{nT}$, $\tilde{C}_i, \tilde{D}_i$ and $\tilde{E}_i$ are matrices of order $O_p^U(1)$. To determine these matrices, we first take a second order Taylor expansion of $\sum_{t=1}^T v_{it}(\hat{\theta}_{nT}, \hat{\alpha}_i) = 0$, leading to

$$\begin{aligned}o_p^U\left(\frac{1}{T}\right) &= \bar{v}_{i\cdot}+\bar{v}_{i\cdot\theta}\left[\frac{1}{\sqrt{nT}}\tilde{A}_{nT}+\frac{1}{T}\tilde{B}_{nT}\right]+\bar{v}_{i\cdot\alpha_i}\left[\frac{1}{\sqrt{nT}}\tilde{C}_i+\frac{1}{\sqrt{T}}\tilde{D}_i+\frac{1}{T}\tilde{E}_i\right]\\ &\quad +\frac{1}{2}\bar{v}_{i\cdot\alpha_i\alpha_i}\left[\frac{1}{\sqrt{nT}}\tilde{C}_i+\frac{1}{\sqrt{T}}\tilde{D}_i+\frac{1}{T}\tilde{E}_i\right]^2\\ &= \left[\sqrt{T}\bar{v}_{i\cdot}+\tilde{D}_i\bar{v}_{i\cdot\alpha_i}\right]\frac{1}{\sqrt{T}}+\left[\bar{v}_{i\cdot\theta}\tilde{A}_{nT}+\bar{v}_{i\cdot\alpha_i}\tilde{C}_i\right]\frac{1}{\sqrt{nT}}\\ &\quad +\left[\bar{v}_{i\cdot\theta}\tilde{B}_{nT}+\bar{v}_{i\cdot\alpha_i}\tilde{E}_i+\frac{1}{2}\bar{v}_{i\cdot\alpha_i\alpha_i}\tilde{D}_i^2\right]\frac{1}{T},\end{aligned}$$

where $\bar{v}_{i\cdot}$ is the time series average of v_{it} and other notations are similarly defined.

Since the above holds for all n and T, the coefficients for $1/\sqrt{T}, 1/\sqrt{nT}$ and $1/T$ must be zero or approach zero at a certain rate. So

$$\begin{aligned}\tilde{D}_i &= -\frac{\sqrt{T}\bar{v}_{i\cdot}}{\bar{v}_{i\cdot\alpha_i}}+o_p^U\left(\frac{1}{\sqrt{T}}\right),\ \tilde{C}_i=-\frac{\bar{v}_{i\cdot\theta}}{\bar{v}_{i\cdot\alpha_i}}\tilde{A}_{nT}+o_p^U\left(\sqrt{\frac{n}{T}}\right),\\ \tilde{E}_i &= -\frac{\bar{v}_{i\cdot\theta}\tilde{B}_{nT}+\frac{1}{2}\bar{v}_{i\cdot\alpha_i\alpha_i}\tilde{D}_i^2}{\bar{v}_{i\cdot\alpha_i}}+o_p^U(1).\end{aligned} \tag{A.12}$$

The $o_p^U(\cdot)$ terms above can be absorbed into the remainder term of order $o_p^U(1/T)$ in (A.11). We drop them from now on without loss of generality and redefine $\tilde{D}_i, \tilde{C}_i$ and $\tilde{E}_i$ as

$$\tilde{D}_i=-\frac{\sqrt{T}\bar{v}_{i\cdot}}{\bar{v}_{i\cdot\alpha_i}},\ \tilde{C}_i=-\frac{\bar{v}_{i\cdot\theta}}{\bar{v}_{i\cdot\alpha_i}}\tilde{A}_{nT},\ \tilde{E}_i=-\frac{\bar{v}_{i\cdot\theta}\tilde{B}_{nT}+\frac{1}{2}\bar{v}_{i\cdot\alpha_i\alpha_i}\tilde{D}_i^2}{\bar{v}_{i\cdot\alpha_i}}. \tag{A.13}$$

We use the above definitions hereafter. Let

$$D_i=-\frac{\sum_{t=1}^T v_{it}}{\sqrt{T}\mathbb{E}\left[\bar{v}_{i\cdot\alpha_i}\right]}=\frac{1}{\sqrt{T}}\sum_{t=1}^T\psi_{it}.$$

We can expand $\tilde{D}_i$ as follows:

$$\begin{aligned}
\tilde{D}_i &= D_i \left[1 + \frac{T^{-1}\sum_{t=1}^T (v_{it\alpha_i} - \mathbb{E}v_{it\alpha_i})}{\mathbb{E}\left[\bar{v}_{i\cdot\alpha_i}\right]}\right]^{-1} \qquad (A.14)\\
&= D_i + \frac{1}{\sqrt{T}} \frac{T^{-1}\left[\sum_{t=1}^T v_{it}\right]\left[\sum_{t=1}^T (v_{it\alpha_i} - \mathbb{E}v_{it\alpha_i})\right]}{\left\{\mathbb{E}\left[\bar{v}_{i\cdot\alpha_i}\right]\right\}^2} + O_p^U\left(\frac{1}{T}\right).
\end{aligned}$$

Next, we take a second order Taylor expansion of $\sum_{i=1}^n \sum_{t=1}^T u_{it}(\hat{\theta}_{nT}, \hat{\alpha}_i) = 0$, leading to

$$\begin{aligned}
0 &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it} + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\theta}(\hat{\theta}_{nT} - \theta_0)\\
&+ \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\alpha_i}\left[\hat{\alpha}_i - \alpha_{i0}\right] + \frac{1}{2nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\alpha_i\alpha_i}(\breve{\theta}, \breve{\alpha}_i)\left[\hat{\alpha}_i - \alpha_{i0}\right]^2\\
&+ \frac{1}{2nT}\sum_{i=1}^n\sum_{t=1}^T \xi_{it}(\breve{\theta}, \breve{\alpha}_i) + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left[\hat{\alpha}_i - \alpha_{i0}\right] u_{it\theta\alpha_i}(\breve{\theta}, \breve{\alpha}_i)(\hat{\theta}_{nT} - \theta_0), \qquad (A.15)
\end{aligned}$$

where $\xi_{it}(\breve{\theta}, \breve{\alpha}_i) = \left(\xi_{it,1}, \ldots, \xi_{it,r}, \ldots, \xi_{it,d_\theta}\right)'$ is a vector with r-th element

$$\xi_{it,r} \equiv \xi_{it,r}(\breve{\theta}, \breve{\alpha}_i) = (\hat{\theta}_{nT} - \theta_0)' u_{it,\theta\theta_r}(\breve{\theta}, \breve{\alpha}_i)(\hat{\theta}_{nT} - \theta_0).$$

Plugging the definitions of $\tilde{C}_i, \tilde{D}_i$ and $\tilde{E}_i$ into (A.15), we have

$$\begin{aligned}
&O_p^U\left(\frac{1}{T\sqrt{T}} + \frac{1}{\sqrt{n}T}\right)\\
=& \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it} + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\theta}\left(\frac{1}{\sqrt{nT}}\tilde{A}_{nT} + \frac{1}{T}\tilde{B}_{nT}\right)\\
&+ \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\alpha_i}\left[\frac{1}{\sqrt{nT}}\tilde{C}_i + \frac{1}{\sqrt{T}}\tilde{D}_i + \frac{1}{T}\tilde{E}_i\right] + \frac{1}{2nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\alpha_i\alpha_i}\frac{1}{T}\tilde{D}_i^2\\
=& \left(\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it} + \frac{1}{n}\sum_{i=1}^n \bar{u}_{i\cdot\alpha_i}\frac{1}{\sqrt{T}}D_i\right) + \frac{1}{\sqrt{nT}}\left(\frac{1}{n}\sum_{i=1}^n \bar{u}_{i\cdot\theta} - \frac{1}{n}\sum_{i=1}^n \bar{u}_{i\cdot\alpha_i}\frac{\bar{v}_{i\cdot\theta}}{\bar{v}_{i\cdot\alpha_i}}\right)\tilde{A}_{nT}\\
&+ \frac{1}{nT}\sum_{i=1}^n\left(\bar{u}_{i\cdot\theta}\tilde{B}_{nT} + \bar{u}_{i\cdot\alpha_i}\tilde{E}_i + \frac{1}{2}\bar{u}_{i\cdot\alpha_i\alpha_i}\tilde{D}_i^2\right)\\
&+ \frac{1}{nT}\sum_{i=1}^n\left[\frac{\bar{u}_{i\cdot\alpha_i}\left(T^{-1}\sum_{t=1}^T\sum_{s=1}^T v_{it}v_{is\alpha_i}\right)}{\left[\mathbb{E}\left(\bar{v}_{i\cdot\alpha_i}\right)\right]^2} - \frac{T^{-1}\sum_{t=1}^T\sum_{s=1}^T u_{it\alpha_i}v_{is}}{\mathbb{E}\left(\bar{v}_{i\cdot\alpha_i}\right)}\right],
\end{aligned}$$

where

$$
\begin{aligned}
& \frac{1}{nT}\sum_{i=1}^{n}\left(\bar{u}_{i\cdot\theta}\tilde{B}_{nT}+\bar{u}_{i\cdot\alpha_i}\tilde{E}_i+\frac{1}{2}\bar{u}_{i\cdot\alpha_i\alpha_i}\tilde{D}_i^2\right) \\
= & \frac{1}{nT}\sum_{i=1}^{n}\left(\bar{u}_{i\cdot\theta}-\frac{\bar{u}_{i\cdot\alpha_i}}{\bar{v}_{i\cdot\alpha_i}}\bar{v}_{i\cdot\theta}\right)\tilde{B}_{nT}+\frac{1}{nT}\sum_{i=1}^{n}\left(\frac{1}{2}\bar{u}_{i\cdot\alpha_i\alpha_i}-\frac{1}{2}\frac{\bar{u}_{i\cdot\alpha_i}\bar{v}_{i\cdot\alpha_i\alpha_i}}{\bar{v}_{i\cdot\alpha_i}}\right)\tilde{D}_i^2 \\
= & \frac{1}{nT}\sum_{i=1}^{n}\left(\bar{u}_{i\cdot\theta}-\frac{\bar{u}_{i\cdot\alpha_i}}{\bar{v}_{i\cdot\alpha_i}}\bar{v}_{i\cdot\theta}\right)\tilde{B}_{nT}+\frac{1}{2nT}\sum_{i=1}^{n}\left[\left(\bar{u}_{i\cdot\alpha_i\alpha_i}-\frac{\bar{u}_{i\cdot\alpha_i}\bar{v}_{i\cdot\alpha_i\alpha_i}}{\bar{v}_{i\cdot\alpha_i}}\right)\right]\frac{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}}{\left(\bar{v}_{i\cdot\alpha_i}\right)^2}.
\end{aligned}
$$

Define

$$
\tilde{H}_{nT}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(u_{it\theta}-\frac{\bar{u}_{i\cdot\alpha_i}}{\bar{v}_{i\cdot\alpha_i}}v_{it\theta}\right), \tag{A.16}
$$

$$
\tilde{S}_{nT}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(u_{it}-\frac{\bar{u}_{i\cdot\alpha_i}}{\mathbb{E}\bar{v}_{i\cdot\alpha_i}}v_{it}\right), \tag{A.17}
$$

$$
\begin{aligned}
\tilde{b}_{nT,1} &= \frac{1}{2n}\sum_{i=1}^{n}\left(\bar{u}_{i\cdot\alpha_i\alpha_i}-\frac{\bar{u}_{i\cdot\alpha_i}\bar{v}_{i\cdot\alpha_i\alpha_i}}{\bar{v}_{i\cdot\alpha_i}}\right)\frac{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}}{\left(\bar{v}_{i\cdot\alpha_i}\right)^2}, \\
\tilde{b}_{nT,2} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\bar{u}_{i\cdot\alpha_i}\left(T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is\alpha_i}\right)}{\left[\mathbb{E}\left(\bar{v}_{i\cdot\alpha_i}\right)\right]^2}-\frac{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it\alpha_i}v_{is}}{\mathbb{E}\left(\bar{v}_{i\cdot\alpha_i}\right)}\right\}.
\end{aligned}
$$

Then $\tilde{A}_{nT}=-\tilde{H}_{nT}^{-1}\tilde{S}_{nT}$ and $\tilde{B}_{nT}=-\tilde{H}_{nT}^{-1}\left(\tilde{b}_{nT,1}+\tilde{b}_{nT,2}\right)$.

It follows from Lemma A.1 that

$$
\tilde{H}_{nT}=\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}U_{it\theta}\right)\left[1+o_p^U(1)\right]=H_{nT}+o_p^U(1),
$$

$$
\tilde{S}_{nT}=\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\left[1+o_p^U(1)\right]=S_{nT}+o_p^U(1).
$$

In addition,

$$
\tilde{b}_{nT,1}=-\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{\mathbb{E}T^{-1}\sum_{t=1}^{T}U_{it\alpha_i\alpha_i}}{\mathbb{E}T^{-1}\sum_{t=1}^{T}v_{it\alpha_i}}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}\psi_{is}\right)+o_p^U(1), \tag{A.18}
$$

$$
\begin{aligned}
\tilde{b}_{nT,2} &= -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\bar{u}_{i\cdot\alpha_i}}{\mathbb{E}\left(\bar{v}_{i\cdot\alpha_i}\right)}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\psi_{is}v_{it\alpha_i}+\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it\alpha_i}\psi_{is}\right)\\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\psi_{is}\left(u_{it\alpha_i}-\frac{\bar{u}_{i\cdot\alpha_i}}{\mathbb{E}\left(\bar{v}_{i\cdot\alpha_i}\right)}v_{it\alpha_i}\right)\\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}U_{it\alpha_i}\psi_{is}+o_p^U(1), \qquad \text{(A.19)}
\end{aligned}
$$

and so

$$
\begin{aligned}
\tilde{b}_{nT,1}+\tilde{b}_{nT,2} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}U_{it\alpha_i}\psi_{is}+o_p^U(1),\\
&\quad -\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{\mathbb{E}T^{-1}\sum_{t=1}^{T}U_{it\alpha_i\alpha_i}}{\mathbb{E}T^{-1}\sum_{t=1}^{T}v_{it\alpha_i}}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}\psi_{is}\right)+o_p^U(1)\\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\psi_{is}\left[U_{it\alpha_i}+\left(\frac{1}{2T}\mathbb{E}\sum_{\tau=1}^{T}U_{i\tau\alpha_i\alpha_i}\right)\psi_{it}\right]+o_p^U(1)\\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t}\mathbb{E}\psi_{is}U_{it\alpha_i}+\frac{1}{2n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}U_{it\alpha_i\alpha_i}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\psi_{it}^2\right)+o_p^U(1)\\
&= b_{nT}+o_p^U(1).
\end{aligned}
$$

Combining the above approximations, we have

$$
\hat{\theta}_{nT}-\theta_0=\frac{1}{\sqrt{nT}}A_{nT}+\frac{1}{T}B_{nT}+o_p^U\left(\frac{1}{T}\right) \qquad \text{(A.20)}
$$

for $A_{nT}=-H_{nT}^{-1}S_{nT}$ and $B_{nT}=-H_{nT}^{-1}b_{nT}$. It is easy to see that $A_{nT}=O_p^U(1)$ and $B_{nT}=O_p^U(1)$.

For the stochastic expansion of $\hat{\alpha}_i$, we can take

$$
\begin{aligned}
C_i(\gamma_0) &= -\frac{\mathbb{E}\bar{v}_{i\cdot\theta}}{\mathbb{E}\bar{v}_{i\cdot\alpha_i}}A_{nT},\ D_i(\gamma_0)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{it}\\
E_i(\gamma_0) &= -\frac{\left(\mathbb{E}\bar{v}_{i\cdot\theta}\right)B_{nT}+\frac{1}{2}\left(\bar{v}_{i\cdot\alpha_i\alpha_i}\right)\mathbb{E}\left(D_i^2(\gamma_0)\right)}{\mathbb{E}\bar{v}_{i\cdot\alpha_i}}+\frac{T^{-1}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}v_{it}\left(v_{is\alpha_i}-\mathbb{E}v_{is\alpha_i}\right)\right]}{\left(\mathbb{E}\bar{v}_{i\cdot\alpha_i}\right)^2}. \qquad \text{(A.21)}
\end{aligned}
$$

Then by Lemma A.1,

$$\hat{\alpha}_i - \alpha_{i0} = \frac{1}{\sqrt{nT}} C_i(\gamma_0) + \frac{1}{\sqrt{T}} D_i(\gamma_0) + \frac{1}{T} E_i(\gamma_0) + o_p^U\left(\frac{1}{T}\right)$$

where all of $C_i(\gamma_0)$, $D_i(\gamma_0)$ and $E_i(\gamma_0)$ are of order $O_p^U(1)$. Using the same argument for proving (A.9), we can show that

$$\max_i \left| \hat{\alpha}_i - \alpha_{i0} - \left(\frac{1}{\sqrt{nT}} C_i(\gamma_0) + \frac{1}{\sqrt{T}} D_i(\gamma_0) + \frac{1}{T} E_i(\gamma_0) \right) \right| = o_p^U(1). \tag{A.22}$$

■

To prove Theorem 3 and other results, we will use the following two lemmas repeatedly. Lemma A.2 helps translate the asymptotic results for the original sample into the corresponding ones for the bootstrap sample. Lemma A.3 shows that the effect of nonlinear truncation can be ignored in large samples.

Lemma A.2 *Let $\lambda_{nT}(\gamma)$ be a sequence of real functions on $\Gamma_1^{\otimes}$. If (i) $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} |\lambda_{nT}(\gamma_0)| = o(a_{nT})$ for some sequence a_{nT}; (ii) $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(\|\hat{\gamma} - \gamma_0\| \geq \nu) = o(a_{nT})$, then for any $\varepsilon > 0$,*

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(|\lambda_{nT}(\hat{\gamma})| \geq a_{nT}\varepsilon) = o(a_{nT}).$$

Proof of Lemma A.2. In view of condition (ii), we have $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(\hat{\gamma} \notin \Gamma_1^{\otimes}\right) = o(a_{nT})$. Now

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(|\lambda_{nT}(\hat{\gamma})| \geq a_{nT}\varepsilon) \\
\leq\ & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(|\lambda_{nT}(\hat{\gamma})| \geq a_{nT}\varepsilon, \hat{\gamma} \in \Gamma_1^{\otimes}\right) + \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(\hat{\gamma} \notin \Gamma_1^{\otimes}\right) \\
\leq\ & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma \in \Gamma_1^{\otimes}} |\lambda_{nT}(\gamma)| \geq a_{nT}\varepsilon\right) + o(a_{nT}) \\
=\ & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(o(a_{nT}) \geq a_{nT}\varepsilon) + o(a_{nT}) = o(a_{nT}),
\end{aligned} \tag{A.23}$$

where the last equality holds because $P_{\gamma_0}(o(a_{nT}) \geq a_{nT}\varepsilon) = 0$ when n and T are large enough.

■

Lemma A.3 *Let $\hat{J}_{nT}(\gamma)$ and $\tilde{J}_{nT}(\gamma)$ be two sequences of random vectors in $\mathbb{R}^{d_J}$ indexed by $\gamma \in \Gamma_1^{\otimes}$. Assume (i) $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}(||\hat{J}_{nT}(\gamma_0) - \tilde{J}_{nT}(\gamma_0)|| \geq \delta) = o(c_{nT})$ for any $\delta > 0$ and some $c_{nT} = o(1)$ as $(n,T) \to \infty$; (ii) $\sup_{n,T} \sup_{\gamma_0 \in \Gamma_1^{\otimes}} (\mathbb{E}_{\gamma_0} ||\tilde{J}_{nT}(\gamma_0)||^2) < \infty$. Then as $M \to \infty$ such that $M = O(1/c_{nT})$, we have $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0} g_M[\hat{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0}[\tilde{J}_{nT}(\gamma_0)] \right\| = o(1)$.*

Proof of Lemma A.3. Note that $\tilde{g}_M(\cdot)$ is Lipschitz continuous. It follows from condition (i) that

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\| g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right) \right\| \geq \delta \right) = o(c_{nT}).$$

Using this result, we have, as $M \to \infty$ such that $M = O(1/c_{nT})$:

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0}\left[g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right] \right\| \\
\leq\ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0} \left\| g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right\| 1\left\{ \left\| g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right\| \geq \delta \right\} \\
& + \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0} \left\| g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right\| 1\left\{ \left\| g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right\| < \delta \right\} \\
\leq\ & o(Mc_{nT}) + \delta = \delta + o(1),
\end{aligned}$$

for any $\delta > 0$. This implies that $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0}\left[g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right] \right\| = o(1)$ as $(n,T,M) \to \infty$ such that $M = O(1/c_{nT})$. But

$$\mathbb{E}_{\gamma_0}\left[g_M\left(\tilde{J}_{nT}(\gamma_0)\right) \right] = \mathbb{E}_{\gamma_0}\left[\tilde{J}_{nT}(\gamma_0) \right] + \mathcal{R}_{nT,M}(\gamma_0)$$

where $\mathcal{R}_{nT,M}(\gamma_0) \equiv \mathbb{E}_{\gamma_0}\left[g_M\left(\tilde{J}_{nT}(\gamma_0)\right) - \tilde{J}_{nT}(\gamma_0) \right]$ satisfies

$$\begin{aligned}
\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \|\mathcal{R}_{nT,M}(\gamma_0)\| &\leq \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0} \left\| \tilde{J}_{nT}(\gamma_0) \right\| 1\left\{ \left\| \tilde{J}_{nT}(\gamma_0) \right\| > \sqrt{d_J} M \right\} \\
&\leq \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0} \left\| \tilde{J}_{nT}(\gamma_0) \right\|^2 \right)^{1/2} \left(P_{\gamma_0}\left\{ \left\| \tilde{J}_{nT}(\gamma_0) \right\| > \sqrt{d_J} M \right\} \right)^{1/2} \\
&\leq \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0} \left\| \tilde{J}_{nT}(\gamma_0) \right\|^2 \right)^{1/2} \left(\mathbb{E}_{\gamma_0} \left\| \tilde{J}_{nT}(\gamma_0) \right\|^2 / \left(d_J M^2 \right) \right)^{1/2} \\
&\leq \sup_{n,T} \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0} \left\| \tilde{J}_{nT}(\gamma_0) \right\|^2 \right) / \left(\sqrt{d_J} M \right) \leq C/M
\end{aligned} \tag{A.24}$$

for some constant C that does not depend on n, T and γ_0. Here we have used condition (ii) in the lemma. That is, $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0} g_M[\tilde{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0}[\tilde{J}_{nT}(\gamma_0)] \right\| = o(1)$, as $(n, T, M) \to \infty$ such that $M = O(1/c_{nT})$. Combining this with $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0} \left[g_M(\hat{J}_{nT}(\gamma_0)) - g_M(\tilde{J}_{nT}(\gamma_0)) \right] \right\| = o(1)$ leads to the desired result. ■

Lemma A.4 *Let the assumptions in Theorem 3 hold. For any $\delta > 0$ and $\varepsilon > 0$,*

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0} \left\{ P_{\hat{\gamma}}^* \left[\left\| \sqrt{nT} \left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right) - A_{nT}^*(\hat{\gamma}) - \sqrt{\frac{n}{T}} B_{nT}^*(\hat{\gamma}) \right\| \geq \delta \right] > \varepsilon \right\} = o\left(\frac{1}{T} \right).$$

Proof of Lemma A.4. Let $\hat{\theta}_{nT}(\gamma)$ be the estimator of θ based on the sample $\{Z_{it}(\gamma)\} \equiv \{Z_{i1}(\gamma), \ldots, Z_{iT}(\gamma)\}_{i=1}^n$ generated under the original DGP with parameter value $\gamma = (\theta', \alpha')'$. Similarly let $\hat{\theta}_{nT}^*(\gamma)$ be the estimator of θ based on the bootstrap sample $\{Z_{it}^*(\gamma)\}$ generated under the bootstrap DGP with parameter value $\gamma = (\theta, \alpha)$. To apply Lemma A.2 with $a_{nT} = 1/T$, we define

$$\lambda_{nT}(\gamma) = P_{\gamma} \left\{ \left\| \hat{\theta}_{nT}(\gamma) - \theta - \frac{1}{\sqrt{nT}} A_{nT}(\gamma; \{Z_{it}(\gamma)\}) - \frac{1}{T} B_{nT}(\gamma; \{Z_{it}(\gamma)\}) \right\| \geq \frac{\delta}{T} \right\},$$

and

$$\lambda_{nT}^*(\gamma) = P_{\gamma}^* \left\{ \left\| \hat{\theta}_{nT}^*(\gamma) - \theta - \frac{1}{\sqrt{nT}} A_{nT}^*(\gamma; \{Z_{it}^*(\gamma)\}) - \frac{1}{T} B_{nT}^*(\gamma; \{Z_{it}^*(\gamma)\}) \right\| \geq \frac{\delta}{T} \right\}.$$

Since the original DGP and the bootstrap DGP are the same, we have $A_{nT}(\gamma; \{Z_{it}(\gamma)\}) = A_{nT}^*(\gamma; \{Z_{it}(\gamma)\})$ and $B_{nT}(\gamma; \{Z_{it}(\gamma)\}) = B_{nT}^*(\gamma; \{Z_{it}(\gamma)\})$. That is, A_{nT} and A_{nT}^* have the same function form, so do B_{nT} and B_{nT}^*. In addition, $\lambda_{nT}(\gamma) = \lambda_{nT}^*(\gamma)$.

It follows from Theorem 2 that $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \lambda_{nT}(\gamma_0) = o(1)$ for any $\delta > 0$. Using the same arguments as in the proof of Theorem 1 and with additional calculations, we can strengthen the result to $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \lambda_{nT}(\gamma_0) = o(1/T)$ for any $\delta > 0$. This stronger result requires $Q > d_\theta + 12$. By Theorem 1, $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(\|\hat{\gamma} - \gamma_0\| \geq \nu) = o(1/T)$. Invoking Lemma A.2 yields:

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(\lambda_{nT}^*(\hat{\gamma}) \geq \varepsilon) = \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(\lambda_{nT}(\hat{\gamma}) \geq \varepsilon) = o(1/T).$$

■

Proof of Theorem 3. Part (i). We apply Lemma A.3 with

$$\hat{J}_{nT}(\gamma_0) = \sqrt{nT}\left(\hat{\theta}_{nT}(\gamma_0) - \theta_0\right),\ \tilde{J}_{nT}(\gamma_0) = A_{nT}(\gamma_0) + \sqrt{n/T}B_{nT}(\gamma_0).$$

It is not hard to verify the two conditions in the lemma. So $\left\|\mathbb{E}_{\gamma_0} g_M[\hat{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0}\tilde{J}_{nT}(\gamma_0)\right\| = o(1)$, and $\left\|\mathbb{E}_{\gamma_0} g_M[\hat{J}_{nT}(\gamma_0)] - \sqrt{n/T}B_{nT}(\gamma_0)\right\| = o(1)$ uniformly over $\gamma_0 \in \Gamma_1^{\otimes}$.

Part (ii). Let $\hat{J}^*_{nT}(\gamma) = \sqrt{nT}\left(\hat{\theta}^*_{nT}(\gamma) - \theta\right)$ and define

$$\begin{aligned}
\lambda_{nT}(\gamma) &= \mathbb{E}_{\gamma}\left[g_M(\hat{J}_{nT}(\gamma)) - \sqrt{\frac{n}{T}}B_{nT}(\gamma)\right], \\
\lambda^*_{nT}(\gamma) &= \mathbb{E}^*_{\gamma}\left[g_M(\hat{J}^*_{nT}(\gamma)) - \sqrt{\frac{n}{T}}B^*_{nT}(\gamma)\right].
\end{aligned}$$

For the same reason as given in Lemma A.4, $\lambda_{nT}(\gamma) = \lambda^*_{nT}(\gamma)$. It follows from Part (i) that $\sup_{\gamma_0\in\Gamma_1^{\otimes}} \lambda_{nT}(\gamma_0) = o(1)$. Hence, by Lemma A.2, $\sup_{\gamma_0\in\Gamma_0^{\otimes}} P_{\gamma_0}(\lambda^*_{nT}(\hat{\gamma}) \geq \varepsilon) = o(1)$.

Part (iii). Following the same arguments in HK, we can show that $A_{nT}(\gamma_0) \xrightarrow{d} N[0, \Omega(\gamma_0)]$. Under Assumptions 2, 3 and 7, the result can be strengthened to

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} \left|P_{\gamma_0}(c' A_{nT}(\gamma_0)/\sqrt{c'\Omega(\gamma_0)c} < \tau) - \Phi(\tau)\right| = o(1) \tag{A.25}$$

for any conformable vector c and $\tau \in \mathbb{R}$. In addition, it is easy to see that $B_{nT}(\hat{\gamma}) - B_{nT}(\gamma_0) = o_p(1)$. Using these two results and part (ii), we have

$$\begin{aligned}
& P\left\{\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) < \vartheta\right\} \\
= \;& P\left\{\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) < \vartheta, \hat{\gamma} \in \Gamma_1^{\otimes}\right\} + P\left\{\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) < \vartheta, \hat{\gamma} \notin \Gamma_1^{\otimes}\right\} \\
= \;& P\left\{\sqrt{nT}\left[\hat{\theta}_{nT} - \theta_0\right] - \mathbb{E}^*_{\hat{\gamma}} g_M\left[\sqrt{nT}\left(\hat{\theta}^*_{nT} - \hat{\theta}_{nT}\right)\right] < \vartheta,\ \hat{\gamma} \in \Gamma_1^{\otimes}\right\} + o(1) \\
= \;& P\left[A_{nT}(\gamma_0) - \sqrt{n/T}\left[B_{nT}(\hat{\gamma}) - B_{nT}(\gamma_0)\right] < \vartheta,\ \hat{\gamma} \in \Gamma_1^{\otimes}\right] + o(1) \\
= \;& P\left[A_{nT}(\gamma_0) < \vartheta, \hat{\gamma} \in \Gamma_1^{\otimes}\right] + o(1) \\
= \;& P\left[A_{nT}(\gamma_0) < \vartheta\right] - P\left[A_{nT}(\gamma_0) < \vartheta, \hat{\gamma} \notin \Gamma_1^{\otimes}\right] + o(1) \\
= \;& P\left[A_{nT}(\gamma_0) < \vartheta\right] + o(1) \xrightarrow{d} N[0, \Omega(\gamma_0)].
\end{aligned} \tag{A.26}$$

Because of (A.25) and that the $o(1)$ terms in the above equation hold uniformly over $\gamma_0 \in \Gamma_0^{\otimes}$, we have for $\tau \in \mathbb{R}$, $\sup_{\gamma_0\in\Gamma_0^{\otimes}} \left|P_{\gamma_0}(c'\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0)/\sqrt{c'\Omega(\gamma_0)c} < \tau) - \Phi(\tau)\right| = o(1)$. ■

Proof of Theorem 4. We first consider the k-step 'estimator' for the original sample:

$$\hat{\gamma}_k = \hat{\gamma}_{k-1} - \left[\mathcal{H}\left(\hat{\gamma}_{k-1}; Z\right)\right]^{-1} \mathcal{S}\left(\hat{\gamma}_{k-1}; Z\right) \tag{A.27}$$

where $Z = \{Z_{it}, t = 1, \ldots, T\}_{i=1}^{n}$,

$$\mathcal{H}\left(\hat{\gamma}_{k-1}; Z\right) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \left.\frac{\partial^2\left(\theta, \alpha_i; Z_{it}\right)}{\partial\left(\theta', \alpha'\right)'\partial\left(\theta', \alpha'\right)}\right|_{\theta=\hat{\theta}_{k-1}, \alpha_i=\hat{\alpha}_{i,k-1}} \tag{A.28}$$

$$\mathcal{S}\left(\hat{\gamma}_{k-1}; Z\right) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \left.\frac{\partial l\left(\theta, \alpha_i; Z_{it}\right)}{\partial\left(\theta', \alpha'\right)'}\right|_{\theta=\hat{\theta}_{k-1}, \alpha_i=\hat{\alpha}_{i,k-1}} \tag{A.29}$$

and $\hat{\gamma}_0 = \gamma_0$. While the k-step estimator is feasible for the bootstrap sample, the above k-step estimator for the original sample is not feasible, as we do not know the true value γ_0. Nevertheless, we want to show that there exists a constant $K > 0$ such that

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(T^{2^{k-1}}\left\|\hat{\gamma}_k - \hat{\gamma}\right\| > K\right) = o\left(1/T\right).$$

That is, had we known the true value γ_0, the infeasible k-step estimator $\hat{\gamma}_k$ would be very close to the MLE $\hat{\gamma}$ when k is large enough. This result will be used in establishing the convergence property of the k-step estimator in the bootstrap world.

Using a Taylor expansion and the first order condition $\mathcal{S}\left(\hat{\gamma}; Z\right) = 0$, we have

$$\begin{aligned}
\hat{\gamma}_k - \hat{\gamma} &= \hat{\gamma}_{k-1} - \left[\mathcal{H}(\hat{\gamma}_{k-1}; Z)\right]^{-1} \mathcal{S}\left(\hat{\gamma}_{k-1}; Z\right) - \hat{\gamma} \\
&= \left[\mathcal{H}(\hat{\gamma}_{k-1}; Z)\right]^{-1}\left[\mathcal{S}\left(\hat{\gamma}; Z\right) - \mathcal{S}\left(\hat{\gamma}_{k-1}; Z\right) - \mathcal{H}(\hat{\gamma}_{k-1}; Z)\left(\hat{\gamma} - \hat{\gamma}_{k-1}\right)\right] \\
&= \frac{1}{2}\left[\mathcal{H}(\hat{\gamma}_{k-1}; Z)\right]^{-1}\xi\left(\hat{\gamma}_{k-1}^{\dagger}; Z\right)
\end{aligned} \tag{A.30}$$

where $\hat{\gamma}_{k-1}^{\dagger}$ lies between $\hat{\gamma}$ and $\hat{\gamma}_{k-1}$, $\xi\left(\hat{\gamma}_{k-1}^{\dagger}; Z\right) = \left(\xi_1, \ldots, \xi_u, \ldots, \xi_{d_\gamma}\right)'$ is a vector with the u-th element:

$$\xi_u \equiv \xi_u\left(\hat{\gamma}_{k-1}^{\dagger}; Z\right) = \left(\hat{\gamma}_{k-1} - \hat{\gamma}\right)' \mathcal{H}_{\gamma_u}(\hat{\gamma}_{k-1}^{\dagger}; Z)\left(\hat{\gamma}_{k-1} - \hat{\gamma}\right)$$

and

$$\mathcal{H}_{\gamma_u}(\gamma; Z) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial}{\partial\gamma_u}\frac{\partial^2 l(\gamma; Z_{it})}{\partial\gamma\partial\gamma'} \equiv \begin{pmatrix} \mathcal{H}_{\theta\theta\gamma_u}\left(\gamma; Z\right) & \mathcal{H}_{\theta\alpha\gamma_u}\left(\gamma; Z\right) \\ \mathcal{H}_{\alpha\theta\gamma_u}\left(\gamma; Z\right) & \mathcal{H}_{\alpha\alpha\gamma_u}\left(\gamma; Z\right) \end{pmatrix}. \tag{A.31}$$

Now

$$
\begin{aligned}
\left\|\xi_u\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\right\| &\leq \left\|2\left(\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right)'\mathcal{H}_{\theta\theta\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right)\right\| \\
&+\left\|2\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right)'\mathcal{H}_{\alpha\alpha\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right)\right\| \\
&\leq 2\left\|\mathcal{H}_{\theta\theta\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\right\|\left\|\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right\|^2 \\
&+\left\|\frac{2}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{\alpha_i\alpha_i\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z_{it}\right)\right]\left(\hat{\alpha}_{i,k-1}-\hat{\alpha}_i\right)^2\right\|,
\end{aligned}
$$

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

$$
\|\hat{\gamma}_k-\hat{\gamma}\| \leq \zeta_{\theta,k-1}\left\|\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right\|^2+\zeta_{\alpha,k-1}, \tag{A.32}
$$

where

$$
\begin{aligned}
\zeta_{\theta,k-1} &= \left\|\left[\mathcal{H}\left(\hat{\gamma}_{k-1};Z\right)\right]^{-1}\right\|\sum_{u=1}^{d_\gamma}\left\|\mathcal{H}_{\theta\theta\gamma_u}(\hat{\gamma}_{k-1}^{\dagger};Z)\right\|, \\
\zeta_{\alpha,k-1} &= \left\|\left[\mathcal{H}\left(\hat{\gamma}_{k-1};Z\right)\right]^{-1}\right\|\left|\sum_{u=1}^{d_\gamma}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathcal{H}_{\alpha_i\alpha_i\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z_{it}\right)\left(\hat{\alpha}_{i,k-1}-\hat{\alpha}_i\right)^2\right|.
\end{aligned}
$$

For $k=1$, we have for any $K>0$,

$$
\begin{aligned}
& P_{\gamma_0}\left(T\|\hat{\gamma}_k-\hat{\gamma}\|\geq K\right) \\
\leq\ & P_{\gamma_0}\left(T\zeta_{\theta,k-1}\left\|\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right\|^2+T\zeta_{\alpha,k-1}\geq K\right) \\
\leq\ & P_{\gamma_0}\left(T\zeta_{\theta,k-1}\left\|\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right\|^2\geq 0.5K\right)+P_{\gamma_0}\left(T\zeta_{\alpha,k-1}\geq 0.5K\right) \\
\leq\ & P_{\gamma_0}\left(T\left\|\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right\|^2\geq\sqrt{0.5K}\right)+P_{\gamma_0}\left(\zeta_{\theta,k-1}\geq\sqrt{0.5K}\right)+P_{\gamma_0}\left(T\zeta_{\alpha,k-1}\geq 0.5K\right).
\end{aligned}
$$

By Lemma A.1, it is not hard to show that $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}\left(\zeta_{\theta,k-1}\geq\sqrt{0.5K}\right)=o(1/T)$ for a large enough K. Similarly, using the expansions in Theorem 2 and Lemma A.1, we can show that $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}\left(T\zeta_{\alpha,k-1}\geq 0.5K\right)=o(1/T)$ for a large enough K. Therefore

$$
\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}\left(T\|\hat{\gamma}_k-\hat{\gamma}\|\geq K\right)\leq\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}\left(T\left\|\hat{\theta}_{k-1}-\hat{\theta}_{nT}\right\|^2\geq\sqrt{0.5K}\right)+o(1/T)=o(1/T)
$$

using $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(T \left\| \theta_0 - \hat{\theta}_{nT} \right\|^2 \geq \sqrt{0.5K} \right) = o(1/T)$ and $\hat{\theta}_{k-1} = \theta_0$ for $k = 1$.

Using the same steps, we can show that the above holds for $k \geq 2$. Hence

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(T^{2^{k-1}} \left\| \hat{\gamma}_k (\gamma_0) - \hat{\gamma} (\gamma_0) \right\| > K \right) = o(1/T) \text{ for all } k \geq 1 \tag{A.33}$$

where we have written $\hat{\gamma}_k \equiv \hat{\gamma}_k (\gamma_0)$ and $\hat{\gamma} \equiv \hat{\gamma} (\gamma_0)$ to emphasize their dependence on the true parameter γ_0.

Letting

$$\hat{J}_{nT} (\gamma_0) = \sqrt{nT} \left[\hat{\theta}_{nT,k} (\gamma_0) - \theta_0 \right], \; \tilde{J}_{nT} (\gamma_0) = A_{nT} (\gamma_0) + \sqrt{n/T} B_{nT} (\gamma_0)$$

and invoking Lemma A.3, we have, for $k \geq 2$:

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT,k} (\gamma_0) - \theta_0 \right) \right] - \sqrt{n/T} B_{nT} (\gamma_0) \right\| = o(1).$$

Combining this with Theorem 3(ii) yields

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT,k} (\gamma_0) - \theta_0 \right) \right] - \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT} - \theta_0 \right) \right] \right\| = o(1) \tag{A.34}$$

for $k \geq 2$.

Define $\lambda_{nT} (\gamma_0) = \left\| \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT,k} (\gamma_0) - \theta_0 \right) \right] - \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT} - \theta_0 \right) \right] \right\|$. Then for $k \geq 2$, $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} |\lambda_{nT} (\gamma_0)| = o(1)$. It follows from Lemma A.2 that when $k \geq 2$,

$$\mathbb{E}_{\hat{\gamma}}^* \left\{ g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT} \right) \right] - g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT} \right) \right] \right\} = o_p (1)$$

uniformly over $\gamma_0 \in \Gamma_0^{\otimes}$. ■

Proof of Theorem 5. In view of Theorem 4, we have $\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) = \sqrt{nT} \left(\tilde{\theta}_{nT} - \theta_0 \right) + o_p (1)$. The theorem follows from Theorem 3(iii). ■

Lemma A.5 *Let the assumptions in Theorem 3 hold. Then for any $\eta > 0$ and $k \geq 1$*

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\max_i |\tilde{\alpha}_{ik} - \alpha_{i0}| > \eta \right) = o(1).$$

Proof of Lemma A.5. Note that

$$\max_i |\tilde{\alpha}_{ik} - \alpha_{i0}| \le \max_i |\hat{\alpha}_i - \alpha_{i0}| + \max_i \frac{1}{\sqrt{T}} \mathbb{E}^*_{\hat{\gamma}} g_M \left(\sqrt{T} \left(\hat{\alpha}^*_{ik} - \hat{\alpha}_i \right) \right).$$

Using Theorem 1, we have

$$\begin{aligned} & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\max_i |\tilde{\alpha}_{ik} - \alpha_{i0}| > \eta \right) \\ = & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left\{ \frac{1}{\sqrt{T}} \max_i \mathbb{E}^*_{\hat{\gamma}} g_M \left(\sqrt{T} \left(\hat{\alpha}^*_{ik} - \hat{\alpha}_i \right) \right) > \eta \right\} + o(\frac{1}{T}). \end{aligned}$$

Let $\lambda_{nT}(\gamma_0) = \max_i \mathbb{E}_{\gamma_0} g_M \left(\sqrt{T} \left(\hat{\alpha}_{ik} - \hat{\alpha}_i \right) \right) / \sqrt{T}$, then for a large enough K, we have,

$$\begin{aligned} \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \lambda_{nT}(\gamma_0) &= \frac{1}{\sqrt{T}} \max_i \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0} g_M \left(\sqrt{T} \left(\hat{\alpha}_{ik} - \hat{\alpha}_i \right) \right) 1 \left\{ T \max_i \left(|\hat{\alpha}_{ik} - \hat{\alpha}_i| \right) \ge K \right\} \\ &\quad + \frac{1}{\sqrt{T}} \max_i \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0} g_M \left(\sqrt{T} \left(\hat{\alpha}_{ik} - \hat{\alpha}_i \right) \right) 1 \left\{ T \max_i \left(|\hat{\alpha}_{ik} - \hat{\alpha}_i| \right) \le K \right\} \\ &\le \frac{M}{\sqrt{T}} \max_i \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left\{ T \max_i \left(|\hat{\alpha}_{ik} - \hat{\alpha}_i| \right) \ge K \right\} + \frac{1}{\sqrt{T}} \frac{K}{\sqrt{T}} \\ &= o\left(\frac{M}{\sqrt{T}} \left(\frac{1}{T} \right) + \frac{1}{\sqrt{T}} \frac{K}{\sqrt{T}} \right) = o\left(\frac{1}{\sqrt{T}} \right). \end{aligned}$$

where we have used (A.33). Invoking Lemma A.2 with $a_{nT} = 1$, we obtain

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left\{ \frac{1}{\sqrt{T}} \max_i \mathbb{E}^* g_M \left(\sqrt{T} \left(\hat{\alpha}^*_{ik} - \hat{\alpha}_i \right) \right) > \eta \right\} = o(1).$$

As a result, $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\max_i |\tilde{\alpha}_{ik} - \alpha_{i0}| > \eta \right) = o(1)$. ■

Proof of Theorem 6. Part (i). Under the assumptions on $\Delta(w, \tilde{Z}_{it}, \gamma_i)$, we can use the same argument for proving Lemma A.1 to obtain

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\sup_i \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\partial^{m_1 + m_2} \Delta(w, \tilde{Z}_{it}, \gamma_i)}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right| > K \right) = o(\frac{1}{T}) \tag{A.35}$$

for some large enough constant K.

Now

$$
\begin{aligned}
& \sqrt{nT}\left[\hat{\mu}(w)-\mu(w)\right] \\
= \quad & \frac{\sqrt{nT}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\Delta(w,\tilde{Z}_{it},\hat{\theta}_{nT},\hat{\alpha}_i)-\Delta(w,\tilde{Z}_{it},\theta_0,\alpha_{i0})\right] \\
= \quad & \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\theta}(w,\tilde{Z}_{it},\gamma_{i0})\right]'\sqrt{nT}\left(\hat{\theta}_{nT}-\theta_0\right)+\frac{\sqrt{nT}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\left(\hat{\alpha}_i-\alpha_{i0}\right) \\
& +\frac{\sqrt{nT}}{2nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\left(\hat{\alpha}_i-\alpha_{i0}\right)^2+o_p^U(1). \qquad \text{(A.36)}
\end{aligned}
$$

Using Theorem 2, we have

$$
\begin{aligned}
& \frac{\sqrt{nT}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\left(\hat{\alpha}_i-\alpha_{i0}\right) \\
= \quad & -\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\frac{\mathbb{E}\bar{v}_{i\cdot\theta}}{\mathbb{E}\bar{v}_{i\cdot\alpha_i}}\right]A_{nT}-\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\frac{\sum_{t=1}^{T}v_{it}}{\sqrt{T}\mathbb{E}\bar{v}_{i\cdot\alpha_i}} \\
& +\sqrt{\frac{n}{T}}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})E_i\left(\gamma_0\right)+o_p^U(1) \\
= \quad & -\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\frac{\mathbb{E}\bar{v}_{i\cdot\theta}}{\mathbb{E}\bar{v}_{i\cdot\alpha_i}}\right]A_{nT}-\sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\left[\sum_{t=1}^{T}\frac{\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})}{\mathbb{E}\bar{v}_{i\cdot\alpha_i}}\right]\left[\sum_{t=1}^{T}v_{it}\right] \\
& +\sqrt{\frac{n}{T}}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})E_i\left(\gamma_0\right)+o_p^U(1);
\end{aligned}
$$

$$
\begin{aligned}
& \frac{\sqrt{nT}}{2nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\left(\hat{\alpha}_i-\alpha_{i0}\right)^2 \\
= \quad & \sqrt{\frac{n}{T}}\frac{1}{2nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\alpha_i\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})D_i^2\left(\gamma_0\right)+o_p^U(1);
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\theta}'(w,\tilde{Z}_{it},\gamma_{i0})\sqrt{nT}\left(\hat{\theta}_{nT}-\theta_0\right) \\
= \quad & \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\theta}'(w,\tilde{Z}_{it},\gamma_{i0})\right]\left[A_{nT}\left(\gamma_0\right)+\sqrt{\frac{n}{T}}B_{nT}\left(\gamma_0\right)\right]+o_p^U(1).
\end{aligned}
$$

So

$$\sqrt{nT}\left[\hat{\mu}(w)-\mu(w)\right]=A_{nT}^{\mu}\left(\gamma_0\right)+\sqrt{\frac{n}{T}}B_{nT}^{\mu}\left(\gamma_0\right)+e_{nT}^{\mu}$$

for some e_{nT}^{μ} satisfying $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}(\left|e_{nT}^{\mu}\right|>\delta)=o\left(1/T\right)$ for any $\delta>0$. This holds because $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}(\left|e_{nT}^{\theta}\right|>\delta)=o\left(1/T\right)$, $\sup_{\gamma_0\in\Gamma_1^{\otimes}}P_{\gamma_0}(\max_i\left|e_{nT}^{\alpha_i}\right|>\delta)=o\left(1/T\right)$ and (A.35). As $n,T\to\infty$ such that $n/T\to\rho\in(0,\infty)$, $A_{nT}^{\mu}\left(\gamma_0\right)\xrightarrow{d}N\left[0,\Omega^{\mu}\left(\gamma_0\right)\right]$ uniformly in the sense of (A.25) and $B_{nT}^{\mu}\left(\gamma_0\right)=B^{\mu}\left(\gamma_0\right)+o_p^U\left(1\right)$. Hence $\sqrt{nT}\left[\hat{\mu}(w)-\mu(w)\right]\xrightarrow{d}N\left[\sqrt{\rho}B^{\mu}\left(\gamma_0\right),\Omega^{\mu}\left(\gamma_0\right)\right]$ uniformly over $\gamma_0\in\Gamma_1^{\otimes}$.

Part (ii). Using the same argument for proving Theorem 3(ii), we can show

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}}\left|\mathbb{E}_{\gamma_0}g_M\left[\sqrt{nT}\left(\hat{\mu}(w)-\mu(w)\right)\right]-\sqrt{\frac{n}{T}}B_{nT}^{\mu}\left(\gamma_0\right)\right|=o\left(1\right). \tag{A.37}$$

This, combined with Lemma A.2, implies that

$$\sup_{\gamma_0\in\Gamma_0^{\otimes}}P_{\gamma_0}\left(\left|\mathbb{E}_{\hat{\gamma}}^{*}g_M\left[\sqrt{nT}\left(\hat{\mu}^{*}(w)-\hat{\mu}(w)\right)\right]-\sqrt{\frac{n}{T}}B_{nT}^{\mu*}\left(\hat{\gamma}\right)\right|\geq\delta\right)=o\left(1\right)$$

where

$$\begin{aligned}B_{nT}^{\mu*}\left(\hat{\gamma}\right) &= \mathbb{E}_{\hat{\gamma}}^{*}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta_{\theta}'(w,\tilde{Z}_{it}^{*},\hat{\gamma}_i)B_{nT}^{*}\left(\hat{\gamma}\right)+\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}_{\hat{\gamma}}^{*}\left[\Delta_{\alpha_i}(w,\tilde{Z}_{it}^{*},\hat{\gamma}_i)E_i^{*}\left(\hat{\gamma}\right)\right]\\ &\quad -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t,s}\Delta_{\alpha_i}(w,\tilde{Z}_{it}^{*},\hat{\gamma}_i)\psi_{is}^{*}\left(\hat{\gamma}_i\right)+\frac{1}{2nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}_{\hat{\gamma}}^{*}\Delta_{\alpha_i\alpha_i}(w,\tilde{Z}_{it}^{*},\hat{\gamma}_i)D_i^2\left(\hat{\gamma}\right).\end{aligned}$$

Since the DGP's for the original sample and the bootstrap sample differ only in terms of parameter values, we have $B_{nT}^{\mu*}\left(\hat{\gamma}\right)=B_{nT}^{\mu}\left(\hat{\gamma}\right)$. As a result, $B_{nT}^{\mu*}\left(\hat{\gamma}\right)=B_{nT}^{\mu}\left(\gamma_0\right)+o_p^U\left(1\right)$. Using this, we have

$$\begin{aligned}\sqrt{nT}\left[\tilde{\mu}(w)-\mu(w)\right] &= \sqrt{nT}\left[\hat{\mu}(w)-\mu(w)\right]-\mathbb{E}_{\hat{\gamma}}^{*}g_M\left[\sqrt{nT}\left(\hat{\mu}^{*}(w)-\hat{\mu}(w)\right)\right]\\ &= A_{nT}^{\mu}\left(\gamma_0\right)+\sqrt{\frac{n}{T}}B_{nT}^{\mu}\left(\gamma_0\right)-\sqrt{\frac{n}{T}}B_{nT}^{\mu}\left(\hat{\gamma}\right)+o_p\left(1\right)\\ &= A_{nT}^{\mu}\left(\gamma_0\right)+o_p\left(1\right)\xrightarrow{d}N\left[0,\Omega^{\mu}\left(\gamma_0\right)\right],\end{aligned} \tag{A.38}$$

uniformly in that the underlying probability error converges to zero uniformly over $\gamma_0\in\Gamma_0^{\otimes}$.

Part (iii). It suffices to show that

$$\sup_{\gamma_0\in\Gamma_0^{\otimes}} P_{\gamma_0}\left\{\left|\mathbb{E}^*_{\hat{\gamma}} g_M\left[\sqrt{nT}\left(\hat{\mu}^*(w)-\hat{\mu}(w)\right)\right]-\mathbb{E}^*_{\hat{\gamma}} g_M\left[\sqrt{nT}\left(\hat{\mu}^*_k(w)-\hat{\mu}(w)\right)\right]\right|\geq\delta\right\}=o\left(1\right),$$

for $k\geq 2$. In view of Lemma A.2, this holds if

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}}\left\|\mathbb{E}_{\gamma_0} g_M\left[\sqrt{nT}\left(\hat{\mu}(w)-\mu(w)\right)\right]-\mathbb{E}_{\gamma_0} g_M\left[\sqrt{nT}\left(\hat{\mu}_k(w)-\hat{\mu}(w)\right)\right]\right\|=o\left(1\right) \quad \text{(A.39)}$$

for $k\geq 2$, where $\hat{\mu}_k(w)=\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\Delta(w,\tilde{Z}_{it},\hat{\gamma}_k)$ and $\hat{\gamma}_k$ is the k-step estimator of γ_0 defined in (A.27). Noting that for some $\check{\gamma}_i$ between $\hat{\gamma}_i$ and $\hat{\gamma}_{i,k}$,

$$\sqrt{nT}\left(\hat{\mu}(w)-\mu(w)\right)-\sqrt{nT}\left(\hat{\mu}_k(w)-\mu(w)\right)=\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left[\Delta_\gamma(w,\tilde{Z}_{it},\check{\gamma}_i)\sqrt{nT}\left(\hat{\gamma}_i-\hat{\gamma}_{i,k}\right)\right], \quad \text{(A.40)}$$

we have, for any $\delta>0$ and a large enough $K>0$:

$$\begin{aligned}
&\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left|\sqrt{nT}\left(\hat{\mu}(w)-\mu(w)\right)-\sqrt{nT}\left(\hat{\mu}_k(w)-\mu(w)\right)\right|\geq\delta\right)\\
\leq\;&\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left|\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\Delta_\gamma(w,\tilde{Z}_{it},\check{\gamma}_i)\sqrt{nT}\left(\hat{\gamma}_i-\hat{\gamma}_{i,k}\right)\right|\geq\delta,\left\|T^2\left(\hat{\gamma}_i-\hat{\gamma}_{i,k}\right)\right\|\leq K\right)\\
&+\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|T^2\left(\hat{\gamma}_i-\hat{\gamma}_{i,k}\right)\right\|>K\right)\\
=\;&\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left\|\Delta_\gamma(w,\tilde{Z}_{it},\check{\gamma}_i)\right\|\frac{\sqrt{nT}K}{T^2}\geq\delta\right)+o\left(\frac{1}{T}\right)=o\left(\frac{1}{T}\right)
\end{aligned}$$

for $k\geq 2$. Given this, (A.39) follows from the same argument for proving (A.34). ■

Proof of Theorem 7. Part (i). We first introduce the notations $\Gamma_{1/2}=int\left\{\gamma_i\in\Gamma:\|\gamma_i-\Gamma_0\|\leq\nu/2\right\}$ and $\Gamma_{1/2}^{\otimes}=\left\{\gamma=\left(\theta',\alpha_1,\ldots,\alpha_n\right)':\left(\theta',\alpha_i\right)'\in\Gamma_{1/2}\right\}$. Let $\Phi\left[\cdot;\Omega\right]$ be the CDF of the multivariate normal distribution with variance matrix Ω. Define

$$\lambda_{nT}\left(\gamma_0\right)=\sup_{\vartheta}\left|P_{\gamma_0}\left[\sqrt{nT}(\tilde{\theta}_{nT,k}-\theta_0)<\vartheta\right]-\Phi\left[\vartheta;\Omega(\gamma_0)\right]\right|.$$

It is not hard to show that $\sup_{\gamma_0\in\Gamma_{1/2}^{\otimes}}\left|\lambda_{nT}\left(\gamma_0\right)\right|=o\left(1\right)$ and hence $\sup_{\gamma_0\in\Gamma_0^{\otimes}} P_{\gamma_0}\left(\left|\lambda^*_{nT}\left(\hat{\gamma}\right)\right|\geq\delta/2\right)=$

$o(1)$ by an argument similar to Lemma A.2. But

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(|\lambda_{nT}^*(\hat{\gamma})| \geq \frac{\delta}{2}\right) \\
= & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\hat{\gamma}}^*\left[\sqrt{nT}(\tilde{\theta}_{nT,k}^* - \hat{\theta}_{nT}) < \vartheta\right] - \Phi\left[\vartheta, \Omega(\hat{\gamma})\right]\right| \geq \frac{\delta}{2}\right\} \\
= & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\hat{\gamma}}^*\left[\sqrt{nT}(\tilde{\theta}_{nT,k}^* - \hat{\theta}_{nT}) < \vartheta\right] - \Phi\left[\vartheta, \Omega(\gamma_0)\right]\right| \geq \frac{\delta}{2}\right\} + o(1)
\end{aligned}$$

and so

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\hat{\gamma}}^*\left[\sqrt{nT}(\tilde{\theta}_{nT,k}^* - \hat{\theta}_{nT}) < \vartheta\right] - \Phi\left[\vartheta, \Omega(\gamma_0)\right]\right| \geq \frac{\delta}{2}\right\} = o(1).$$

Combining this with

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\gamma_0}\left[\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) < \vartheta\right] - \Phi\left[\vartheta, \Omega(\gamma_0)\right]\right| \geq \frac{\delta}{2}\right\} = o(1),$$

we obtain part (i).

Part (ii). We use the same argument as in Part (i). We redefine $\lambda_{nT}(\gamma_0)$ to be

$$\lambda_{nT}(\gamma_0) = \sup_{\vartheta}\left|P_{\gamma_0}\left[\sqrt{nT}(\tilde{\theta}_{nT,kk} - \theta_0) < \vartheta\right] - \Phi\left[\vartheta; \Omega(\gamma_0)\right]\right|.$$

Note that $\sqrt{nT}(\tilde{\theta}_{nT,kk} - \theta_0) = \sqrt{nT}\left(\hat{\theta}_{nT,k} - \theta_0\right) - \mathbb{E}_{\hat{\gamma}_k}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT,kk}^* - \hat{\theta}_{nT,k}\right)\right]$. Using exactly the same argument that leads to Theorem 3(ii), we have

$$\left|\mathbb{E}_{\hat{\gamma}_k}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT,kk}^* - \hat{\theta}_{nT,k}\right)\right] - \sqrt{n/T} B_{nT}(\hat{\gamma}_k)\right| = o_p(1)$$

uniformly over $\gamma_0 \in \Gamma^{\otimes}_{1/2}$. Using this and (A.33), we have for $k \geq 2$

$$
\begin{aligned}
& \sup_{\gamma_0 \in \Gamma^{\otimes}_{1/2}} \left| P_{\gamma_0} \left[\sqrt{nT}(\tilde{\theta}_{nT,kk} - \theta_0) < \vartheta \right] - \Phi \left[\vartheta ; \Omega(\gamma_0) \right] \right| \\
= & \sup_{\gamma_0 \in \Gamma^{\otimes}_{1/2}} \left| P_{\gamma_0} \left[\sqrt{nT}(\hat{\theta}_{nT,k} - \theta_0) - \sqrt{n/T} B_{nT}\left(\hat{\gamma}_k\right) < \vartheta \right] - \Phi \left[\vartheta ; \Omega(\gamma_0) \right] \right| + o(1) \\
= & \sup_{\gamma_0 \in \Gamma^{\otimes}_{1/2}} \left| P_{\gamma_0} \left[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) - \sqrt{n/T} B_{nT}\left(\hat{\gamma}_k\right) < \vartheta \right] - \Phi \left[\vartheta ; \Omega(\gamma_0) \right] \right| + o(1) \\
= & \sup_{\gamma_0 \in \Gamma^{\otimes}_{1/2}} \left| P_{\gamma_0} \left[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) - \sqrt{n/T} B_{nT}\left(\hat{\gamma}\right) < \vartheta \right] - \Phi \left[\vartheta ; \Omega(\gamma_0) \right] \right| + o(1) \\
= & \ o(1).
\end{aligned}
$$

Combining this with Poyla's lemma, we have $\sup_{\gamma_0 \in \Gamma^{\otimes}_{1/2}} \left| \lambda_{nT}\left(\gamma_0\right) \right| = o\left(1\right)$. It then follows from an argument similar to Lemma A.2 that $\sup_{\gamma_0 \in \Gamma^{\otimes}_{0}} P_{\gamma_0}\left(\left| \lambda^*_{nT}\left(\hat{\gamma}\right) \right| \geq \delta/2 \right) = o\left(1\right)$. The rest of the proof is the same as that in the proof of Part (i) and is omitted here. ■

Proof of Equation (44). It follows from Theorem 7 (i) that

$$
\sup_{\gamma_0 \in \Gamma^{\otimes}_{0}} P_{\gamma_0} \left\{ \sup_{\tau} \left| P^*_{\hat{\gamma}} \left[\sqrt{nT} c'(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT}) \geq \tau \right] - P_{\gamma_0} \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k} - \theta_0) \geq \tau \right] \right| \geq \delta \right\} = o(1).
$$

So

$$
\sup_{\gamma_0 \in \Gamma^{\otimes}_{0}} P_{\gamma_0} \left\{ \left| P^*_{\hat{\gamma}} \left[\sqrt{nT} c'(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT}) \geq q_{c,1-\alpha} \right] - P_{\gamma_0} \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k} - \theta_0) \geq q_{c,1-\alpha} \right] \right| \geq \delta \right\} = o(1).
$$

By definition $P^*_{\hat{\gamma}} \left[\sqrt{nT} c'(\tilde{\theta}^*_{nT,k} - \hat{\theta}_{nT}) \geq q_{c,1-\alpha} \right] = \alpha$. As a result,

$$
\sup_{\gamma_0 \in \Gamma^{\otimes}_{0}} P_{\gamma_0} \left\{ \left| P_{\gamma_0} \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k} - \theta_0) \geq q_{c,1-\alpha} \right] - \alpha \right| \geq \delta \right\} = o(1).
$$

Since $\left| P_{\gamma_0} \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k} - \theta_0) \geq q_{c,1-\alpha} \right] - \alpha \right|$ is deterministic, we must have

$$
\sup_{\gamma_0 \in \Gamma^{\otimes}_{0}} \left| P_{\gamma_0} \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k} - \theta_0) \geq q_{c,1-\alpha} \right] - \alpha \right| = o(1).
$$

■

Notes

[1]For issues on high order efficiency of bias correction, see Hahn, Kuersteiner and Newey (2004).

[2]One exception is Pace and Salvan (2006). They suggest a bootstrap bias corrected estimator when there are nuisance parameters. 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.

[3]For notational simplicity, we have implicitly assumed that the parameter spaces for α_i's are the same across i. When (θ, α_i) is regarded as a vector, it is understood to be $(\theta', \alpha_i')'$. For notational economy, we sometimes omit the transpose notation if confusion is unlikely.

[4]The expectation in (3) may not exist because it depends on $\hat{\alpha}_i(\theta)$, an estimator that may not have enough moments. For the purpose of exposition, we assume that it exists for now. We later explicitly address this issue using a truncation argument.

[5]The Hessian matrix we used is called the observed Hessian. We note that some terms in $\frac{\partial^2 \log l\left(\theta,\alpha_i;Z_{it}^*\right)}{\partial(\theta',\alpha')'\partial(\theta',\alpha')}$ have zero expectation. Dropping these terms in equation (17), we obtain the expected Hessian. Our asymptotic results remain valid for the expected Hessian, as the dropped terms are of smaller order.

[6]Strictly speaking, we should define $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Delta(w, \theta_0, \alpha_{i0})$ as the population parameter of interest, but the difference between the finite sample version and the limiting version is asymptotically negligible.

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.W.K. (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.

Amemiya, T. (1985) *Advanced Econometrics*, Harvard University Press.

Arellano, M. & J. Hahn (2006) Understanding bias in nonlinear panel models: some recent developments. *Advances in Economics and Econometrics*, Ninth World Congress, Cambridge University Press.

Bester, A. and C. Hansen (2009) A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics* 27(2), 131–148.

Davidson, R. & J. MacKinnon (1999) Bootstrap testing in nonlinear models. *International Economic Review* 40(2), 487–508.

Davidson, R. & J. MacKinnon (2002) Fast double bootstrap tests of nonnested linear regression models. *Econometric Review* 21(4), 419–429.

Davidson, R. & J. MacKinnon (2007) Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics and Data Analysis* 51(7), 3259–3281.

Dhaene, G. & K. Jochmans (2012) Split-panel jackknife estimation of fixed-effect models. Unpublished manuscript.

Doukhan, P. (1995) *Mixing: Properties and Examples*. New York: Springer-Verlag.

Ferández-Val, I. (2009) Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150(1), 71–85.

Giacomini, R., D. Politis, & H. White (2013) A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* 29(3), 567–589.

Greene, W. (2004) The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal* 7(1), 98–119.

Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects. *Econometrica* 70, 1639–1657.

Hahn, J. & G. Kuersteiner (2011) Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27(6), 1152–1191.

Hahn, J., G. Kuersteiner, & W. Newey (2004) Higher order efficiency of bias corrections. Unpublished Manuscript, MIT.

Hahn, J. & W. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72(4), 1295–1319.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion.* New York: 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,* edited by C. F. Manski & D.L. McFadden, Cambridge: MIT Press.

Honoré, B. & E. Kyriazidou (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. & E. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16(1), 1–32.

Pace, L. & A. Salvan (2006) Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference* 136(10), 3554–3564.

Yokoyama, R. (1980) Moment bounds for stationary and mixing sequences. *Z. Wahrsch. Verw. Gebiete* 52, 45–57.