Title

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

Authors
Sun, Yixiao
Kim, Min Seong

Publication Date
2009-11-08

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

Min Seong Kim
Department of Economics
Ryerson University

Yixiao Sun*
Department of Economics
UC San Diego

March 2013

Abstract

Fixed effects estimators in nonlinear panel models with fixed and short time series length T usually suffer from inconsistency because of the incidental parameters problem first noted by Neyman and Scott (1948). Moreover, even if T grows but at a rate not faster than the cross sectional sample size n, they are asymptotically biased, and therefore the associated confidence intervals have a large coverage error. This paper analyzes the properties of the parametric bootstrap bias corrected maximum likelihood (ML) estimators of nonlinear panel models with fixed effects. We assume that each time series follows a finite order Markov process. We show that the bootstrap bias corrected estimators are asymptotically normal and centered at the true parameter. In particular, we propose using the k-step parametric bootstrap procedure to alleviate the computational cost of implementing the standard bootstrap. We also apply the standard and k-step bootstrap bias correction to average marginal effect estimation and to the double bootstrap for confidence interval construction. Our Monte Carlo simulations show that the k-step bootstrap bias corrected estimator reduces the bias remarkably well in finite samples.

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

JEL Classification Numbers: C13, C33

*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, 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 code.

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 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, although many results in the paper also hold for a more general class of estimators. 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. We establish the asymptotic equivalence of the k-step BBC estimator to the standard BBC estimator. Bootstrap bias correction reduces the asymptotic bias without inflating the asymptotic variance. So inferences based on the BBC estimators are more accurate and reliable.

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 can be very appealing in applied research. It is also well known that the standard bootstrap applies to situations so complicated that they lie beyond the power of traditional analysis. 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-intensive task. Particularly, as the fixed effects approach treats the individual effects as parameters, there are many parameters to be estimated, and computational intensiveness can be more serious in this type of models. For example, in our empirical application (not reported here), there are 1461 individuals, which means there are more than 1461 parameters to be estimated.

We propose using the k-step bootstrap method to alleviate the computational cost of the standard bootstrap. Compared to the standard bootstrap method, the k-step bootstrap procedure is computationally attractive because it only involves computing the Hessian and the score functions. It approximates the standard bootstrap estimator by taking k-steps of a Newton-Raphson (NR) iterative scheme. We employ the original estimate as the starting point for the NR steps. We 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 correction to 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 a compact set in the parameter space. Using the fact that the ML estimator, which is the true parameter in the bootstrap world, lies in this compact 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 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 of estimating the bias of estimators. From this point of view, this paper fills an important gap in the literature, providing the bootstrap bias corrected estimators for nonlinear panel models and analyzing their properties.[1]

There is a large literature on bootstrap bias corrected estimation. Hall (1992) introduces general bootstrap algorithms for bias correction and for the construction of CI's, which we adapt in this paper. Hahn, Kuersteiner and Newey (2004) 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), in which they prove its higher order equivalence to the standard bootstrap. Our paper is particularly built on Andrews (2005), which considers the standard and k-step parametric bootstrap methods for Markov processes. Other bootstrap procedures that reduce the computational cost are also available. See, for example, the fast double bootstrap by Davidson and MacKinnon (2002, 2007) and warp-speed bootstrap by Giacomini, Politis, and White (2013).

The paper is organized as follows. Section 2 discusses the incidental parameters problem in nonlinear panel models with fixed effects. Section 3 describes the standard bootstrap bias correction procedure and explains the k-step bootstrap bias correction. Section 4 establishes the asymptotic properties of our estimators. Some extensions are considered 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. The 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.

[1] 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.

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 model.

Throughout the paper, we maintain the assumption of cross sectional independence. To specify the nonlinear panel model, we only need to describe the data generating process for each time series $\{W_i : W_{i1}, \ldots, W_{iT}\}$ where $W_{it} \in \mathbb{R}^{d_w}$ and $i = 1, 2, \ldots, n$. 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 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 $\Gamma := \Gamma_\theta \times \Gamma_\alpha$.[2] 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 the maximization is taken over Γ_θ and Γ_α, which 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 $\hat{\alpha} \equiv \hat{\alpha}(\hat{\theta}_{nT})$.

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

$$\hat{\theta}_{nT} \to^p \theta_T, \tag{3}$$

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

$$\overline{\mathbb{E}}\left[\sum_{t=1}^{T} l(\theta, \hat{\alpha}_i(\theta); Z_{it})\right] \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\sum_{t=1}^{T} l(\theta, \hat{\alpha}_i(\theta); Z_{it})\right]. \tag{4}$$

[2] For notational simplicity, we have implicitly assumed that the parameter spaces for α_i's are the same across i.

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

$$\theta_T = \theta_0 + \frac{B(\theta_0, \alpha_0)}{T} + o\left(\frac{1}{T}\right) \tag{5}$$

for some $B(\theta_0, \alpha_0) \neq 0$. So the asymptotic bias is of order $O(1/T)$ and approaches zero as T increases.

The above asymptotic analysis is conducted under the sequential asymptotics under which $n \to \infty$ for a fixed T followed by letting $T \to \infty$. The basic intuition on the incidental parameter 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:

$$\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = A_{nT}(\theta_0, \alpha_0) + \sqrt{\frac{n}{T}} B_{nT}(\theta_0, \alpha_0) + o_p(1) \tag{6}$$

where $A_{nT}(\theta_0, \alpha_0)$ and $B_{nT}(\theta_0, \alpha_0)$ satisfy

$$A_{nT}(\theta_0, \alpha_0) \xrightarrow{d} N(0, \Omega) \text{ and } B_{nT}(\theta_0, \alpha_0) \to^p B(\theta_0, \alpha_0) \tag{7}$$

for some variance matrix $\Omega := \Omega(\theta_0, \alpha_0)$. So the asymptotic distribution of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$ is centered at $\sqrt{\rho} B(\theta_0, \alpha_0) \neq 0$. Statistical inference that ignores the nonzero center will 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 $\mathbb{E}\left[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)\right]$. For any $M > 0$, define the truncation function $\tilde{g}_M(x)$ on $\mathbb{R}$ by

$$\tilde{g}_M(x) = \begin{cases} M, & \text{if } x > M; \\ x, & \text{if } |x| \leq M; \\ -M, & \text{if } x < -M. \end{cases} \tag{8}$$

Then $\tilde{g}_M(x)$ is bounded and continuous. Let $g_M(\theta) = (\tilde{g}_M(\theta_1), \ldots, \tilde{g}_M(\theta_{d_\theta}))'$. By construction, $\mathbb{E} g_M\left[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)\right]$ always exists. Under some rate conditions on M, we can show that

$$\mathbb{E} g_M\left[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)\right] = \sqrt{n/T} B_{nT}(\theta_0, \alpha_0) + o(1). \tag{9}$$

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

3 Bootstrap Bias Correction

In this section, we introduce parametric bootstrap procedures to estimate $\mathbb{E} g_M\left[\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)\right]$. The parametric bootstrap is different from the nonparametric bootstrap in that the former utilizes the parametric structure of the DGP by replacing the original parameters with their estimators to generate the bootstrap samples, while the latter generates them from the empirical distribution function.

3.1 Standard Bootstrap Bias Correction

In the bootstrap world, we maintain cross sectional independence and generate each time series $\{Y_{it}^*\}$ 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 $\gamma_{i0} = (\theta_0, \alpha_{i0})$ replaced by $\hat{\gamma}_i = (\hat{\theta}_{nT}, \hat{\alpha}_i)$. The initial value for the bootstrap sample is from the stationary distribution of $\left(Y_{it-1}^*, \ldots, Y_{it-\kappa}^*\right)$. In practice, this can be achieved by simulating a long time series using $(Y_{i0}, \ldots, Y_{i1-\kappa})$ as the initial value and keep only the last $T + \kappa$ observations.

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,

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

where

$$\hat{\alpha}_i^*(\theta) = \arg\max_{\alpha_i \in \Gamma_\alpha} \sum_{t=1}^T l(\theta, \alpha_i; Z_{it}^*). \tag{11}$$

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

$$\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right) = A_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}) + \sqrt{\frac{n}{T}} B_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}) + o_{p^*}(1) \tag{12}$$

conditional on a set with probability approaching 1 (see Theorem 3), where

$$\sqrt{\frac{n}{T}} B_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}) = \mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right] + o_p(1) \tag{13}$$

and $\mathbb{E}^*$ is the expectation operator under the bootstrap probability distribution P^* conditional on $\hat{\gamma} \equiv (\hat{\theta}_{nT}', \hat{\alpha}(\hat{\theta}_{nT})')'$.

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

$$\sqrt{nT}\left(\tilde{\theta}_{nT} - \theta_0\right) = \sqrt{nT}\left(\hat{\theta}_{nT} - \theta_0\right) - \mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right]. \tag{14}$$

Equivalently,

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

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

$$\tilde{\mathbb{E}}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right] = \frac{1}{R} \sum_{r=1}^R g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^{*(r)} - \hat{\theta}_{nT}\right)\right]. \tag{16}$$

$\tilde{\mathbb{E}}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right]$ is arbitrarily close to $\mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right]$ when R is large enough.

The bootstrap bias correction eliminates the bias of order $O\left(T^{-1}\right)$ so that $\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0)$ is asymptotically unbiased. To see this, we note that the conditional distribution of the bootstrap sample given the data (or $\hat{\gamma}$) is the same as the distribution of the original sample except that the former uses $\hat{\gamma}$ rather than γ_0 as the true parameter. As a result,

$$B_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}) = B_{nT}\left(\hat{\theta}_{nT}, \hat{\alpha}\right) = B_{nT}\left(\theta_0, \alpha_0\right)\left[1 + o_p\left(1\right)\right], \tag{17}$$

and so if $n/T \to \rho \in (0, \infty)$, we have

$$\begin{aligned}\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) &= \sqrt{nT}\left(\hat{\theta}_{nT} - \theta_0\right) - \mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right] \\ &= \sqrt{nT}\left(\hat{\theta}_{nT} - \theta_0\right) - \sqrt{\frac{n}{T}} B_{nT}^*(\hat{\theta}_{nT}, \hat{\alpha}) + o_p\left(1\right) \\ &= \sqrt{nT} A_{nT}(\theta_0, \alpha_0) - \sqrt{\frac{n}{T}}\left[B_{nT}(\hat{\theta}_{nT}, \hat{\alpha}) - B_{nT}(\theta_0, \alpha_0)\right] + o_p\left(1\right) \\ &= \sqrt{nT} A_{nT}(\theta_0, \alpha_0) + o_p\left(1\right) \xrightarrow{d} N(0, \Omega). \end{aligned} \tag{18}$$

That is, the bootstrap bias correction removes the nonzero center of the asymptotic distribution of $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0)$.

3.2 k-step Bootstrap Bias Correction

In this subsection, we define the k-step BBC estimator and demonstrate its asymptotic equivalence to the standard BBC estimator.

The k-step procedure approximates $\hat{\theta}_{nT}^*$ by the NR iterative procedure. Let $\hat{\theta}_{nT,k}^*$ and $\hat{\alpha}_k^* = \left(\hat{\alpha}_{1,k}^*, ..., \hat{\alpha}_{n,k}^*\right)'$ denote the k-step bootstrap estimators. We define $\hat{\theta}_{nT,k}^*$ and $\hat{\alpha}_k^*$ recursively in the following way:

$$\begin{pmatrix} \hat{\theta}_{nT,k}^* \\ \hat{\alpha}_k^* \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{nT,k-1}^* \\ \hat{\alpha}_{k-1}^* \end{pmatrix} - \left(\mathcal{H}_{k-1}\right)^{-1} \mathcal{S}_{k-1} \tag{19}$$

where[3]

$$\mathcal{H}_{k-1} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \left.\frac{\partial^2 \log l\left(\theta, \alpha_i; Z_{it}^*\right)}{\partial\left(\theta', \alpha'\right)' \partial\left(\theta', \alpha'\right)}\right|_{\theta=\hat{\theta}_{nT,k-1}^*, \alpha=\hat{\alpha}_{k-1}^*} \tag{20}$$

$$\mathcal{S}_{k-1} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \left.\frac{\partial \log l\left(\theta, \alpha_i; Z_{it}^*\right)}{\partial\left(\theta', \alpha'\right)'}\right|_{\theta=\hat{\theta}_{nT,k-1}^*, \alpha=\hat{\alpha}_{k-1}^*} \tag{21}$$

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

We show in Theorem 4 below that

$$\mathbb{E}^*\left\{g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}\right)\right] - g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT}^* - \hat{\theta}_{nT}\right)\right]\right\} = o_p\left(1\right) \tag{22}$$

when $k \geq 2$. Define the k-step BBC estimator as

$$\tilde{\theta}_{nT,k} \equiv \hat{\theta}_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT,k}^* - \hat{\theta}_{nT}\right)\right] \tag{23}$$

[3] The Hessian matrix we used is called the observed Hessian. We note that some terms in $\partial^2 \log l\left(\theta, \alpha_i; Z_{it}^*\right)/\partial\left(\theta', \alpha'\right)' \partial\left(\theta', \alpha'\right)$ have zero expectation. Dropping these terms in equation (20), we obtain the expected Hessian. Our asymptotic results remain valid for the expected Hessian, as the dropped terms are of smaller order.

which is analogous to the BBC estimator defined in (15). Then for $n/T \to \rho \in (0,\infty)$ and all $k \geq 2$, we have

$$\begin{aligned}
&\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) \\
&= \sqrt{nT}\left(\tilde{\theta}_{nT} - \theta_0\right) + \mathbb{E}^*\left\{g_M\left[\sqrt{nT}\left(\hat{\theta}^*_{nT} - \hat{\theta}_{nT}\right)\right] - g_M\left[\sqrt{nT}\left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right)\right]\right\} \\
&= \sqrt{nT}\left(\tilde{\theta}_{nT} - \theta_0\right) + o_p(1) \xrightarrow{d} N(0,\Omega).
\end{aligned} \tag{24}$$

4 Asymptotic Properties

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

Let $\gamma = (\theta, \alpha_1, \ldots, \alpha_n)$ and $\gamma_0 = (\theta_0, \alpha_{10}, \ldots, \alpha_{n0})$. 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 $\hat{\theta}_{nT}(\gamma_0)$ and $\hat{\alpha}_i(\gamma_0)$ to highlight their dependence on the observations $\{Z_{it}(\gamma_{i0}), t = 1, \ldots T\}_{i=1}^n$ generated under the parameter value γ_0. When needed, we use similar notation in the bootstrap world, for example, $\mathbb{E}^*_{\hat{\gamma}}$, $\{Z^*_{it}(\hat{\gamma}_i)\}$, etc.

Following HN (2004), we define

$$u_{it}(\gamma_i) \equiv \frac{\partial}{\partial \theta} l(\theta, \alpha_i; Z_{it}) \quad \text{and} \quad v_{it}(\gamma_i) \equiv \frac{\partial}{\partial \alpha_i} l(\theta, \alpha_i; Z_{it}) \tag{25}$$

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

$$U_{it}(\gamma_i) = u_{it}(\gamma_i) - \rho_{i0} v_{it}(\gamma_i) \tag{26}$$

where

$$\rho_{i0} = \frac{\mathbb{E}\sum_{t=1}^T u_{it\alpha_i}(\gamma_{i0})}{\mathbb{E}\sum_{t=1}^T v_{it\alpha_i}(\gamma_{i0})}. \tag{27}$$

Define

$$\psi_{it}(\gamma_{i0}) = -\frac{v_{it}(\gamma_{i0})}{\mathbb{E}\left[T^{-1}\sum_{t=1}^T v_{it\alpha_i}(\gamma_{i0})\right]}. \tag{28}$$

We suppress the arguments of the functions such as $u_{it}, v_{it}, \psi_{it}$, 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

$$G(\gamma_{i0}, \gamma_i) \equiv \frac{1}{T}\sum_{t=1}^T \mathbb{E}_{\gamma_{i0}} l(\gamma_i; Z_{it}). \tag{29}$$

If Z_{it} is strictly stationary, then $G(\gamma_{i0}, \gamma_i) = \mathbb{E}_{\gamma_{i0}} l(\gamma_i; Z_{it})$.

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

$$\boldsymbol{\alpha}(\gamma_{i0}, m) = \sup_t \sup_{\mathbb{A}\in\mathcal{A}^i_t(\gamma_{i0})} \sup_{\mathbb{B}\in\mathcal{B}^i_{t+m}(\gamma_{i0})} \left|P_{\gamma_{i0}}(\mathbb{A}\cap\mathbb{B}) - P_{\gamma_{i0}}(\mathbb{A}) P_{\gamma_{i0}}(\mathbb{B})\right|. \tag{30}$$

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

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 and Γ_1 is open.*

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

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

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 continuously differentiable to four orders;*
(ii) There exists some function $M(Z_{it}) \equiv M(Z_{it}(\gamma_{i0}))$ such that

$$\left| \frac{\partial^{m_1+m_2} l(\gamma_i; Z_{it})}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right| \leq M(Z_{it}), \quad \text{for } m_1, m_2 = 0,1,2,3 \text{ and } m_1 + m_2 \leq 4$$

where γ_{i,τ_1} and γ_{i,τ_2} are the τ_1-th and τ_2-th elements of γ_i;
(iii) $\sup_{\gamma_{i0}\in\Gamma_1} \mathbb{E}_{\gamma_{i0}} [M(Z_{it}(\gamma_{i0}))]^Q < \infty$ for all $Q \in \mathbb{N}$.

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

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

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 the derivatives of $l(\gamma_i, Z_{it})$. These two assumptions are similar to Conditions 3 and 4 in HK who have verified them for some nonlinear panel data models. Assumption 3 assumes that all moments of $M(Z_{it}(\gamma_{i0}))$ exist. This is stronger than necessary. However, it is typically assumed in the literature that very high moments of $M(Z_{it}(\gamma_{i0}))$ exist. For example, HN assume that $\mathbb{E}_{\gamma_{i0}} [M(Z_{it}(\gamma_{i0}))]^{64} < \infty$ and HK assume that $\mathbb{E}_{\gamma_{i0}} [M(Z_{it}(\gamma_{i0}))]^{Q} < \infty$ for $Q > 42$ in the case when $d_{\gamma_i} = 2$. The required value of Q in HK grows with the dimension of γ_i with 42 as the lower bound. Compared to existing assumptions, our assumption is not overly strong from a practical point of view. Our proof does not require the existence of all moments. It remains valid as long as Q is large enough. Assumption 4 is the identification assumption for extremum estimators. It is similar to Condition 3 in HN and Condition 1 in HK.

In Assumptions 2–4, we have assumed that the bounds hold uniformly over $\gamma_{i0} \in \Gamma_1$. The reason for the uniformity requirement 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. So $\hat{\gamma}_i \in \Gamma_1$ with probability approaching one. 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).

Theorem 1 *Let Assumptions 1–4 hold, then for any $\eta > 0$*

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

as $n \to \infty$ and $T \to \infty$ jointly such that $n = O(T^{\epsilon})$ for some $\epsilon > 0$ where

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

The rate condition $n = O(T^{\epsilon})$ is very mild as ϵ can be a large number. The proof 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.

To establish a stochastic expansion of our estimator, we make additional assumptions.

Assumption 5 *$\gamma_0 = (\theta_0, \alpha_0)$ is an interior point in $\Gamma^{\otimes} = \{\gamma = (\theta, \alpha_1, \ldots, \alpha_n) : \theta \in \Gamma_\theta,\ \alpha_i \in \Gamma_\alpha\}$.*

Assumption 6 *There exist constants $\delta > 0$ and $C > 0$ such that*
(i) $\inf_{\gamma_{i0}\in\Gamma_1} |\mathbb{E}v_{it\alpha_i}(\gamma_{i0})| \geq \delta$ and $\sup_{\gamma_{i0}\in\Gamma_1} \|\mathbb{E}v_{it\theta}(\gamma_{i0})\| \leq C$;
(ii) $\sup_{\gamma_0\in\Gamma_1^{\otimes}} \lambda_{\min}\left\{n^{-1}\sum_{i=1}^{n}\left(\mathbb{E}u_{it\theta}(\gamma_{i0}) - [\mathbb{E}v_{it\theta}(\gamma_{i0})]' [\mathbb{E}v_{it\alpha_i}(\gamma_{i0})]^{-1} [\mathbb{E}v_{it\theta}(\gamma_{i0})]\right)\right\} \geq \delta$ where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A.

Assumption 5 is standard. Assumption 6 maintains the full rank condition on the information matrix.

Let

$$
\begin{aligned}
H_{nT} &\equiv H_{nT}(\gamma_0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}U_{it,\theta}(\gamma_0), \\
S_{nT} &\equiv S_{nT}(\gamma_0) = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}(\gamma_0), \qquad (31)
\end{aligned}
$$

and

$$
b_{nT} \equiv b_{nT}(\gamma_0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\psi_{is}\left(U_{it\alpha_i} + \left[\frac{1}{2T}\mathbb{E}\sum_{t=1}^{T}U_{it\alpha_i\alpha_i}\right]\psi_{it}\right). \qquad (32)
$$

Define

$$
A_{nT}(\gamma_0) = -H_{nT}^{-1}(\gamma_0)\, S_{nT}(\gamma_0) \text{ and } B_{nT}(\gamma_0) = -H_{nT}^{-1}(\gamma_0)\, b_{nT}(\gamma_0). \qquad (33)
$$

For a random variable $e_{nT}(\gamma_0)$, we say that $e_{nT}(\gamma_0)$ is of order $o_p(1)$ uniformly over $\gamma_0 \in \Gamma_1^{\otimes}$ if for any $\eta > 0$, $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(\|e_{nT}(\gamma_0)\| > \eta) = o(1)$. We say that $e_{nT}(\gamma_0)$ is of order $O_p(1)$ uniformly over $\gamma_0 \in \Gamma_1^{\otimes}$ if for any $\varepsilon > 0$, there exists a constant $K > 0$ such that $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(\|e_{nT}(\gamma_0)\| > K) \leq \varepsilon$ when n and T are large enough.

Theorem 2 *Let Assumptions 1–6 hold. If $(n, T) \to \infty$ such that $T = O(n)$, 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 $\max\left|e_{nT}^{\alpha_i}(\gamma_0)\right|$ are of order $o_p(1)$ uniformly over $\gamma_0 \in \Gamma_1^{\otimes}$ 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 $O_p(1)$ uniformly over $\gamma_0 \in \Gamma_1^{\otimes}$.

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 equation (A.27).

The theorem is presented in a general form to accommodate not only the ML estimator. It holds as long as $\hat{\theta}_{nT}$ and $\hat{\alpha}_i$ solve the empirical moment equations

$$\frac{1}{nT}\sum_{i,t} u_{it}\left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})\right) = 0 \text{ and } \frac{1}{T}\sum_{t} v_{it}\left(\hat{\theta}_{nT}, \hat{\alpha}_i(\hat{\theta}_{nT})\right) = 0 \text{ for all } i$$

where u_{it} and v_{it} do not have to be score functions. That is, the theorem remains valid for general method-of-moment estimators or Z estimators.

In the ML framework with a correct specification model, ψ_{it} is a martingale difference sequence. In this case, we have

$$\begin{aligned} b_{nT} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\psi_{it}\left(U_{it\alpha_i} + \left[\frac{1}{2T}\mathbb{E}\sum_{t=1}^{T}U_{it\alpha_i\alpha_i}\right]\psi_{it}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}\left(\psi_{it}U_{it\alpha_i}\right) + \frac{1}{2}\left(\mathbb{E}U_{it\alpha_i\alpha_i}\right)\left(\mathbb{E}\psi_{it}^2\right)\right]. \end{aligned} \tag{34}$$

Let $V_{2it} \equiv v_{it}^2 + v_{it\alpha}$. Using the Bartlett identities, we can show that

$$b_{nT} = -\frac{1}{2n}\sum_{i=1}^{n}\frac{\mathbb{E}\left[V_{2it}U_{it}\right]}{\mathbb{E}\left[v_{it}^2\right]}, \tag{35}$$

which is the same as what HN obtained for iid data.

Theorem 3 *Let Assumptions 1–6 hold and define* $\Gamma_0^{\otimes} := \{\gamma = (\theta, \alpha_1, \ldots, \alpha_n) : (\theta, \alpha_i) \in \Gamma_0\}$. *If* $(n,T) \to \infty$ *such that* $n/T \to \rho \in (0,\infty)$ *and* $M \to \infty$, *then*

(i) for any $\delta > 0$ *and* $\varepsilon > 0$,

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

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

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

(iv) $\sqrt{nT}(\tilde{\theta}_{nT} - \theta_0) \xrightarrow{d} N[0, \Omega(\gamma_0)]$ *for each* $\gamma_0 \in \Gamma_0^{\otimes}$ *where* $\Omega(\gamma_0) = -\left[\lim_{(n,T)\to\infty} H_{nT}(\gamma_0)\right]^{-1}$.

The rate conditions on n and T can be relaxed. The theorem continues to hold if $T = O(n)$ and $T/n^{1/3} \to \infty$. But the proof will be much more tedious, as we have to characterize the exact order of the error terms in the expansions given in Theorem 2. As in HN, we provide the proof only under the stronger rate conditions.

Theorem 3(iv) 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 $c \in \mathbb{R}^{d_\theta}$,

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

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* $k \geq 2$,

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

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

When $k = 1$, $\mathbb{E}^*_{\hat{\gamma}} \left\{ g_M \left[\sqrt{nT} \left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT} \right) \right] - g_M \left[\sqrt{nT} \left(\hat{\theta}^*_{nT} - \hat{\theta}_{nT} \right) \right] \right\} = O_p(1)$. 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(iv), 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 inference 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:

$$t_{nT} := t_{nT}(\gamma_0) = \frac{\sqrt{nT}\left(c'\tilde{\theta}_{nT,k} - r\right)}{\sqrt{c'\hat{\Omega}_{nT}c}} = \frac{\sqrt{nT}c'\left(\tilde{\theta}_{nT,k} - \theta_0\right)}{\sqrt{c'\hat{\Omega}_{nT}c}} \tag{36}$$

where

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

and

$$\tilde{\alpha}_k \equiv \hat{\alpha} - \frac{1}{\sqrt{T}} \mathbb{E}^* g_M \left[\sqrt{T}\left(\hat{\alpha}^*_k - \hat{\alpha}\right) \right] \tag{37}$$

is the k-step BBC estimator of α_0. 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 corrected estimators of the average marginal effects using the k-step bootstrap procedure. In nonlinear models, the average marginal effect may be as interesting as the model parameters because it summarizes the effect over a certain sub-population, which is often the quantity of interest in empirical studies.

The first average marginal effect, which we refer to as "the fixed effect average" or simply the average marginal effect, is the marginal effect averaged over α_{i0}. It is defined as[4]

$$\mu_1(w) = \frac{1}{n} \sum_{i=1}^{n} \Delta(w, \theta_0, \alpha_{i0}) \tag{38}$$

[4] 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.

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

$$\mu_2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(Z_{it}, \theta_0, \alpha_{i0}). \tag{39}$$

See also Fernández-Val (2009). The third average marginal effect, which bridges the first two definitions, can be defined by

$$\mu(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w, \tilde{Z}_{it}, \theta_0, \alpha_{i0}), \tag{40}$$

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

$$\hat{\mu}(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w, \tilde{Z}_{it}, \hat{\theta}_{nT}, \hat{\alpha}_i). \tag{41}$$

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

$$\tilde{\mu}(w) = \hat{\mu}(w) - \frac{1}{\sqrt{nT}}\mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\mu}^*(w) - \hat{\mu}(w)\right)\right] \tag{42}$$

where

$$\hat{\mu}^*(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w, \tilde{Z}_{it}^*, \hat{\theta}_{nT}^*, \hat{\alpha}_i^*) \tag{43}$$

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

$$\tilde{\mu}_k(w) = \hat{\mu}(w) - \frac{1}{\sqrt{nT}}\mathbb{E}^* g_M\left[\sqrt{nT}\left(\hat{\mu}_k^*(w) - \hat{\mu}(w)\right)\right] \tag{44}$$

where

$$\hat{\mu}_k^*(w) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta(w, \tilde{Z}_{it}^*, \hat{\theta}_{nT,k}^*, \hat{\alpha}_{i,k}^*). \tag{45}$$

Let

$$L_{nT}(\gamma_0) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\left[\Delta_\theta(w, \tilde{Z}_{it}, \gamma_{i0}) - \Delta_{\alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0})\frac{\mathbb{E}u_{it\alpha_i}}{\mathbb{E}v_{it\alpha_i}}\right]. \tag{46}$$

Define

$$A_{nT}^{\mu}(\gamma_0) = L_{nT}'(\gamma_0)\, A_{nT}(\gamma_0) \tag{47}$$

and

$$
\begin{aligned}
B^{\mu}_{nT}(\gamma_0) \quad = \quad & \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\Delta'_{\theta}(w,\tilde{Z}_{it},\gamma_{i0})B_{nT}(\gamma_0)+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})E_i(\gamma_0)\right] \\
& +\frac{1}{nT}\sum_{i=1}^{n}\sum_{t,s}\mathbb{E}\left[\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\psi_{is}\right]+\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}\left[\Delta_{\alpha_i\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})D_i^2(\gamma_0)\right].
\end{aligned}
\tag{48}
$$

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 $M(Z_{it})$ such that $\left|\partial^{m_1+m_2}\Delta(w,\tilde{Z}_{it},\gamma_i)/\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}\right| \le M(Z_{it})$, for $m_1,m_2=0,1,2$ and $m_1+m_2\le 2$; (iii) $\sup_{\gamma_{i0}\in\Gamma_1}\mathbb{E}_{\gamma_{i0}}\left[M(Z_{it}(\gamma_{i0}))\right]^Q<\infty$ for all $Q\in\mathbb{N}$. Then*

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

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

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

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

The asymptotic bias of $\hat{\mu}(w)$ comes from three different sources. The first term in (48) comes from the asymptotic bias of $\hat{\theta}_{nT}$. The second term in (48) comes from the asymptotic bias of $\hat{\alpha}_i$. The last two terms in (48) capture the nonlinear bias originated from the fact that $\hat{\mu}(w)$ is nonlinear in $\hat{\alpha}$ and that $\hat{\alpha}$ is random.

As in the case of model parameters, Theorem 6 holds uniformly over $\gamma_0\in\Gamma_0^{\otimes}$. As before, the bias expression for the average marginal effect is presented in a general form and can be simplified in the ML framework. In particular, when the model is correctly specified, $\{\psi_{is}\}$ is a martingale difference sequence. So the third term in (48) becomes $(nT)^{-1}\sum_{i,t}\mathbb{E}\Delta_{\alpha_i}(w,\tilde{Z}_{it},\gamma_{i0})\psi_{it}$. Furthermore, when $\tilde{Z}_{it}$ is not present or is predetermined, this term vanishes.

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 $\hat{\gamma}^*=(\hat{\theta}^*_{nT},\hat{\alpha}^*)$ as the model parameter. Define

$$
\tilde{\theta}^*_{nT,k}=\hat{\theta}^*_{nT}-\frac{1}{\sqrt{nT}}\mathbb{E}^{**}_{\hat{\gamma}^*}g_M\left[\sqrt{nT}\left(\hat{\theta}^{**}_{nT,k}-\hat{\theta}^*_{nT}\right)\right]
\tag{49}
$$

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 (23). The mechanics behind $\tilde{\theta}_{nT,k}$ and $\tilde{\theta}^*_{nT,k}$ can be illustrated by the following chart:

$$\begin{array}{rcl}\theta_0 & \mapsto & \hat{\theta}_{nT} \mapsto \hat{\theta}^*_{nT,k} \rightarrowtail \tilde{\theta}_{nT,k} \equiv \hat{\theta}_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^*_{\hat{\gamma}} g_M\left[\sqrt{nT}\left(\hat{\theta}^*_{nT,k} - \hat{\theta}_{nT}\right)\right] \\ \hat{\theta}_{nT} & \mapsto & \hat{\theta}^*_{nT} \mapsto \hat{\theta}^{**}_{nT,k} \rightarrowtail \tilde{\theta}^*_{nT,k} \equiv \hat{\theta}^*_{nT} - \frac{1}{\sqrt{nT}}\mathbb{E}^{**}_{\hat{\gamma}^*} g_M\left[\sqrt{nT}\left(\hat{\theta}^{**}_{nT,k} - \hat{\theta}^*_{nT}\right)\right]\end{array} \tag{50}$$

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 k-step double bootstrap sample. We define our double k-step BBC estimator as

$$\tilde{\theta}^*_{nT,kk} = \hat{\theta}^*_{nT,k} - \frac{1}{\sqrt{nT}}\mathbb{E}^{**}_{\hat{\gamma}^*_k} g_M\left[\sqrt{nT}\left(\hat{\theta}^{**}_{nT,kk} - \hat{\theta}^*_{nT,k}\right)\right] \tag{51}$$

where $\hat{\theta}^{**}_{nT,kk}$ is the k-step estimator based on the k-step double bootstrap sample and using $\hat{\gamma}^*_k$ as the starting point. Similar to the chart in (50), we can use the chart below to illustrate the mechanics behind $\tilde{\theta}^*_{nT,kk}$ and the corresponding quantity $\tilde{\theta}_{nT,kk}$:

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

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

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left[\sqrt{nT}c'(\tilde{\theta}_{nT,k} - \theta_0) \geq q_{c,1-\alpha}\right] - \alpha = o(1). \tag{52}$$

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 (36) using double bootstrap. We define

$$t^*_{nT,k} = \frac{\sqrt{nT}c'\left(\tilde{\theta}^*_{nT,kk} - \hat{\theta}_{nT}\right)}{\sqrt{c'\hat{\Omega}^*_{nT}\left(\tilde{\theta}^*_{nT,kk}, \tilde{\alpha}^*_{kk}\right)c}}. \tag{53}$$

Then under the assumptions of Theorem 3, we have

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

The proof is similar to Theorem 7 and is omitted.

6 Monte Carlo Study

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

Our Monte Carlo experiment uses the following probit model:

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

This design is based on the one employed 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 for each time series Y_{it} $(t = 1, \ldots T)$ to be constant over time.

The probit model we consider completely fits within our framework. While X_{it} is correlated over time, it does not cause any problem as our framework accommodates the conditional MLE with strictly exogenous regressors. In the model, there is no correlation between X_{it} and α_i, 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. The bias of the fixed effects estimator can be severe for fixed effects models as well as for random effects models.

The uncorrected estimator of model parameters is

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

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 $\hat{\theta}^*_{nT,k}$ by the NR iterative procedure presented in (19). We repeat this procedure 999 times ($R = 999$). Then, we have the bias corrected k-step bootstrap estimator from (23). As discussed before, for each k value, we can use either observed Hessian or expected Hessian in the NR step, leading to two versions of the k-step procedure. Each simulation is repeated 1000 times.

We compare the performance of our bias corrected estimator with four alternative bias correction estimators: the jackknife and analytic bias corrected estimators by HN and the analytic bias corrected estimator by Fernández-Val (2009). The jackknife bias corrected estimator is denoted 'Jackknife'. 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'.

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 the same as that given in the respective papers. For the k-step procedure, the CI's are based the double bootstrap procedure, which are

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

where $T^*_{1-\alpha/2}$ is the $(1-\alpha/2)\times 100\%$ percentile of $\left|t^*_{nT,k}\right|$ defined in (53). We also employ the double bootstrap procedure to construct CI's for the average marginal effects. The standard errors for the original 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 shows the performance of the standard and k-step bootstrap bias corrected estimators. From this table, we see that there is no sacrifice of accuracy by using the k-step bootstrap procedure in this setting. The k-step bootstrap bias correction tends to reduce the bias significantly as the standard bootstrap does, when $k \geq 2$. Results not reported here show that the one-step procedure is not effective in bias reduction. This result is consistent with our Theorem 5, which demonstrates the order of bias is reduced from $O(1/T)$ to $o(1/T)$ when $k \geq 2$. In terms of the MSE, the 2-step bootstrap estimators with expected and observed Hessians are efficient when $T = 4$. When $T = 8$ and 12, there is little difference in performance among the standard and k-step bootstrap bias corrected 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 estimator without bias correction is severely biased when T is small. As T gets larger, the bias gets smaller, but there is still no improvement in the coverage accuracy of CI's. When $T = 4$, the bias of the uncorrected estimator is 35%. When $T = 12$, the bias is reduced to 14%. But the rejection probability is 35.5% for the 5% two-sided 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 double k-step bootstrap procedure and normal approximation with BC3 yield more accurate CI's than 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 good as the bias corrected ones. However, the CI's based on the bias uncorrected estimator are especially not accurate when we have small T. When $T = 4$, its error in coverage probability for the 95% CI is about 5%. Inaccurate CI's are not just the problem of the bias uncorrected estimator. The jackknife and analytic bias corrected estimators do not reduce the coverage error either. When $T = 4$, the errors in coverage probability for the 95% CI from jackknife and analytic estimators are 13% and 4-7% respectively. In contrast, the coverage error of the 95% CI constructed using the k-step double bootstrap is 2.7%.

Table 4 gives the Monte Carlo results for the ratio of the estimators of the overall average marginal effect to the true value. As in the previous case, we find little evidence that bias correction is necessary in terms of RMSE. Actually, the RMSE of the bias uncorrected estimator is smaller than that of the jackknife estimator in general. It also shows that in contrast to other estimators, our double k-step bootstrap procedure tends to improve the coverage accuracy of CI's, particularly when T is small.

7 Conclusion

In this paper, we analyze the properties of parametric bootstrap bias correction for fixed effects estimators in nonlinear panel 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 bootstrap bias corrected 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 k-step double 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.

List of Some Notations

$\gamma_i = (\theta', \alpha_i)' \in \Gamma := \Gamma_\theta \times \Gamma_\alpha$	$\gamma = (\theta, \alpha_1, \ldots, \alpha_n)$
$\gamma_{i0} = (\theta_0', \alpha_{i0})' \in \Gamma_0$	$\gamma_0 = (\theta_0, \alpha_{10}, \ldots, \alpha_{n0})$
$\hat{\alpha}(\theta) = (\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_i(\theta), \ldots, \hat{\alpha}_n(\theta))$	$\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_i, \ldots, \hat{\alpha}_n)$
$\hat{\gamma}_i = (\hat{\theta}_{nT}, \hat{\alpha}_i)$, $\hat{\gamma} = (\hat{\theta}_{nT}', \hat{\alpha}')'$	$\hat{\alpha}_i = \hat{\alpha}_i(\hat{\theta}_{nT})$
$\Gamma_1 = \{\gamma_i \in \Gamma : \lVert\gamma_i - \Gamma_0\rVert < \nu\}$	$\Gamma_1^{\otimes} = \{\gamma = (\theta, \alpha_1, \ldots, \alpha_n) : (\theta, \alpha_i) \in \Gamma_1\}$
$\Gamma^{\otimes} = \{\gamma = (\theta, \alpha_1, \ldots, \alpha_n) : (\theta, \alpha_i) \in \Gamma\}$	$\Gamma_0^{\otimes} = \{\gamma = (\theta, \alpha_1, \ldots, \alpha_n) : (\theta, \alpha_i) \in \Gamma_0\}$

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

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 (cross section sample size $n = 100$ and the true value $\theta_0 = 1$)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	1.35	1.33	0.311	0.265	0.363	0.468
2-step Bootstrap	0.91	0.91	0.193	0.038	0.073	0.213
Jackknife	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 Bootstrap	0.98	0.97	0.106	0.040	0.083	0.109
Jackknife	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 Bootstrap	1.00	1.00	0.079	0.036	0.078	0.079
Jackknife	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: Jackknife denotes HN Jackknife bias corrected estimator; BC1 denotes HN bias corrected estimator based on Bartlett equalities; BC2 denotes HN bias corrected estimator based on general estimating equations; BC3 denotes Fernández-Val (2009) bias corrected estimator which uses expected quantities in the estimation of the bias. p;.05 and p;.10 denote empirical rejection probabilities of two-sided nominal 5% and 10% tests.

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

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	0.99	0.98	0.218	0.098	0.184	0.218
2-step bootstrap	0.99	0.98	0.219	0.077	0.132	0.219
Jackknife	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 bootstrap	0.99	0.99	0.104	0.051	0.110	0.105
Jackknife	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 bootstrap	0.99	0.99	0.069	0.064	0.109	0.069
Jackknife	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 Overall Average Marginal Effect (cross section sample size $n = 100$ and the true value = 1)

Estimator	Mean	Median	SD	p;.05	p;.10	RMSE
			$T = 4$			
Probit	1.00	0.99	0.204	0.112	0.188	0.204
2-step Bootstrap	1.00	0.99	0.204	0.065	0.116	0.204
Jackknife	1.03	1.03	0.237	0.116	0.180	0.240
BC1	1.02	1.02	0.234	0.126	0.203	0.235
BC2	1.05	1.05	0.227	0.116	0.200	0.234
BC3	0.95	0.96	0.203	0.091	0.140	0.207
			$T = 8$			
Probit	0.99	0.99	0.092	0.068	0.116	0.092
2-step Bootstrap	0.99	0.99	0.092	0.054	0.084	0.093
Jackknife	1.01	1.00	0.096	0.068	0.133	0.097
BC1	1.02	1.02	0.099	0.083	0.155	0.101
BC2	1.01	1.01	0.095	0.069	0.126	0.095
BC3	0.99	0.99	0.093	0.066	0.120	0.094
			$T = 12$			
Probit	0.99	0.99	0.614	0.075	0.129	0.062
2-step Bootstrap	1.00	1.00	0.062	0.054	0.107	0.062
Jackknife	1.01	1.01	0.069	0.078	0.140	0.070
BC1	1.01	1.01	0.070	0.090	0.143	0.070
BC2	1.00	1.00	0.068	0.073	0.113	0.068
BC3	0.99	0.99	0.067	0.068	0.125	0.067

Notes: See notes to table 2.

Appendix of Proofs

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

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

Using the definition of $\hat{\gamma}_i$, we have

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

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

$$\begin{aligned} & P_{\gamma_{i0}}\left(\sup_{\gamma_i \in \Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0}, \gamma_i)\right| \geq \frac{\delta}{2}\right) \\ \leq \; & P_{\gamma_{i0}}\left(\max_{j}\sup_{\gamma_i \in \mathcal{B}(\gamma_i^{(j)}, \delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0}, \gamma_i) - \tilde{l}_{it}(\gamma_{i0}, \gamma_i^{(j)})\right]\right| \geq \frac{\delta}{4}\right) \\ & + \sum_{j=1}^{J(\delta)} P_{\gamma_{i0}}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0}, \gamma_i^{(j)})\right| \geq \frac{\delta}{4}\right). \end{aligned} \tag{A.3}$$

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,

$$\begin{aligned} & P_{\gamma_{i0}}\left(\max_j \sup_{\gamma_i\in\mathcal{B}(\gamma_i^{(j)},\delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right]\right|\geq\frac{\delta}{4}\right) \\ \leq\; & P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\frac{1}{T}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)+\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\geq\frac{\delta}{4}\right) \\ \leq\; & P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)-\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\right|\geq\frac{\delta}{8}\right) \\ & +P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\left|\frac{2}{T}\sum_{t=1}^{T}\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right|\geq\frac{\delta}{8}\right). \end{aligned} \tag{A.4}$$

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 Assumption 2 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

$$\sup_{\gamma_{i0}\in\Gamma_1} P_{\gamma_{i0}}\left(\frac{C\delta}{\sqrt{T}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[M\left(Z_{it}(\gamma_{i0})\right)-\mathbb{E}M\left(Z_{it}(\gamma_{i0})\right)\right]\right|\geq\frac{\delta}{8}\right)=O\left(\frac{T^{p/2}}{T^{3p/2}}\right)=O\left(T^{-p}\right).$$

Hence

$$\sup_{\gamma_{i0}\in\Gamma_1} P_{\gamma_{i0}}\left(\max_j \sup_{\gamma_i\in\mathcal{B}(\gamma_i^{(j)},\delta/\sqrt{T})}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{l}_{it}(\gamma_{i0},\gamma_i)-\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right]\right|\geq\frac{\delta}{4}\right)=O\left(T^{-p}\right). \tag{A.5}$$

Using the strong mixing moment inequality again, we have

$$\sup_{\gamma_{i0}\in\Gamma_1}\sum_{j=1}^{J(\delta)} P_{\gamma_{i0}}\left(\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0},\gamma_i^{(j)})\right|\geq\frac{\delta}{4}\right)=O\left(T^{-p}\right)O\left(T^{\frac{d_\theta+1}{2}}\right)=O\left(T^{-(p-\frac{d_\theta+1}{2})}\right).$$

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

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

Therefore

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}(\gamma_0)-\theta_0\right\| \geq \eta\right) = O\left(nT^{-(p-\frac{d_\theta+1}{2})}\right).$$

When $n = O(T^{\epsilon})$, $O\left(nT^{-(p-\frac{d_\theta+1}{2})}\right) = o\left(T^{-1}\right)$, provided that $p > 1+\epsilon+\frac{d_\theta+1}{2}$. So

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\theta}_{nT}(\gamma_0)-\theta_0\right\| \geq \eta\right) = o\left(T^{-1}\right). \tag{A.7}$$

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

$$\begin{aligned}
& P_{\gamma_0}(|\hat{\alpha}_i(\gamma_0)-\alpha_{i0}| > \eta) \\
\leq\ & P_{\gamma_0}(G(\gamma_{i0},\gamma_{i0}) - G(\gamma_{i0},\hat{\gamma}_i) \geq \delta) \\
=\ & P_{\gamma_0}\left\{G(\gamma_{i0},\gamma_{i0}) - G\left(\gamma_{i0},(\hat{\theta}'_{nT},\alpha_{i0})'\right) + G\left(\gamma_{i0},(\hat{\theta}'_{nT},\alpha_{i0})'\right) - G(\gamma_{i0},\hat{\gamma}_i) \geq \delta\right\} \\
\leq\ & P_{\gamma_0}\left\{G(\gamma_{i0},\gamma_{i0}) - G\left(\gamma_{i0},(\hat{\theta}'_{nT},\alpha_{i0})'\right) \geq \delta/2\right\} \\
& + P_{\gamma_0}\left\{G\left(\gamma_{i0},(\hat{\theta}'_{nT},\alpha_{i0})'\right) - G(\gamma_{i0},\hat{\gamma}_i) \geq \delta/2\right\}.
\end{aligned} \tag{A.8}$$

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

$$\begin{aligned}
& \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{G(\gamma_{i0},\gamma_{i0}) - G\left(\gamma_{i0},(\hat{\theta}'_{nT},\alpha_{i0})'\right) \geq \delta/2\right\} \\
\leq\ & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\frac{\partial G(\gamma_{i0},(\check{\theta}'_{nT},\alpha_{i0})')}{\partial\theta}\right\| \left\|\hat{\theta}_{nT}-\theta_0\right\| \geq \delta/2\right) \\
\leq\ & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma_{i0}\in\Gamma_1}\mathbb{E}_{\gamma_{i0}}M(Z_{it}(\gamma_{i0}))\left\|\hat{\theta}_{nT}-\theta_0\right\| \geq \delta/2\right) = o(1/T),
\end{aligned} \tag{A.9}$$

where we have used

$$\begin{aligned}
\frac{\partial G(\gamma_{i0},(\check{\theta}'_{nT},\alpha_{i0})')}{\partial\theta} &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}\mathbb{E}_{\gamma_{i0}}l(\gamma_i;Z_{it}(\gamma_{i0}))\big|_{\gamma_i=(\check{\theta}'_{nT},\alpha_{i0})'} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\gamma_{i0}}\frac{\partial}{\partial\theta}l(\gamma_i;Z_{it}(\gamma_{i0}))\big|_{\gamma_i=(\check{\theta}'_{nT},\alpha_{i0})'} \leq \mathbb{E}_{\gamma_{i0}}M(Z_{it}(\gamma_{i0})).
\end{aligned}$$

The exchange of differentiation with expectation holds under Assumption 3. In addition, using

(A.6), we obtain

$$
\begin{aligned}
& \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{G\left(\gamma_{i0},(\hat{\theta}_{nT}',\alpha_{i0})'\right)-G(\gamma_{i0},\hat{\gamma}_i)\geq\delta/2\right\}. \\
= & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{G\left(\gamma_{i0},(\hat{\theta}_{nT}',\alpha_{i0})'\right)-\frac{1}{T}\sum_{t=1}^{T}l(\hat{\gamma}_i;Z_{it})+\frac{1}{T}\sum_{t=1}^{T}l(\hat{\gamma}_i;Z_{it})-G(\gamma_{i0},\hat{\gamma}_i)\geq\delta/2\right\} \\
\leq & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{G\left(\gamma_{i0},(\hat{\theta}_{nT}',\alpha_{i0})'\right)-\frac{1}{T}\sum_{t=1}^{T}l(\hat{\theta}_{nT},\alpha_{i0};Z_{it})+\frac{1}{T}\sum_{t=1}^{T}l(\hat{\theta}_{nT},\hat{\alpha}_i;Z_{it})-G(\gamma_{i0},\hat{\gamma}_i)\geq\delta/2\right\} \\
\leq & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma_i\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}l(\gamma_i;Z_{it})-G(\gamma_{i0},\gamma_i)\right|\geq\delta/4\right) \\
= & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma_i\in\Gamma}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_{it}(\gamma_{i0},\gamma_i)\right|\geq\delta/4\right)=O\left(T^{-(p-\frac{d_\theta+1}{2})}\right). \qquad \text{(A.10)}
\end{aligned}
$$

Therefore for any $\eta>0$ and $p>(d_\theta+1)/2+1$, $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(|\hat{\alpha}_i-\alpha_{i0}|>\eta)=o(1/T)$. This implies that $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(\max_i|\hat{\alpha}_i-\alpha_{i0}|>\eta)=o(1/T)$, as for some $\bar{\imath}$:

$$
\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(\max_i|\hat{\alpha}_i-\alpha_{i0}|>\eta)=\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(|\hat{\alpha}_{\bar{\imath}}-\alpha_{\bar{\imath}0}|>\eta)=o(1/T).
$$

■

To prove Theorem 2, we first establish the following two lemmas, which employ 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 1–6 hold. Let* $m_1,m_2=0,1,2$ *such that* $m_1+m_2\leq 3$*. Let* $\breve{\gamma}=(\breve{\theta}',\breve{\alpha})'$ *satisfy* $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\breve{\theta}-\theta_0\right\|>\eta\right)=o(1)$ *and* $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}(\max_i|\breve{\alpha}_i-\alpha_{i0}|>\eta)=o(1)$ *for any* $\eta>0$*, then*

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

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

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

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

Proof of Lemma A.1. Part (i). Given the assumptions on $\breve{\gamma}$, we have $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P\left(\breve{\gamma}_i \notin \Gamma_1\right) = o(1)$. 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| \frac{1}{nT} \sum_{i=1}^{n} \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 & \left| \frac{1}{nT} \sum_{i=1}^{n} \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(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}} \right] \right| \\
& + \left| \frac{1}{n} \sum_{i=1}^{n} \left[\mathbb{E} \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 & \max_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| \\
& + \frac{1}{n} \sum_{i=1}^{n} \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 & \max_i \sup_{\gamma_i \in \Gamma_1} \left| T^{-1} \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} \frac{1}{n} \sum_{i=1}^{n} \left[\sup_{\gamma_{i0} \in \Gamma_1} \mathbb{E} M(Z_{it}(\gamma_{i0})) \right] \max_{\tau_3} \left| \breve{\gamma}_{i,\tau_3} - \gamma_{i0,\tau_3} \right| \\
: & = \mathcal{T}_1 + \mathcal{T}_2.
\end{aligned}$$

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

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

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

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

in the proof of Theorem 1. See equation (A.2). The differences are (i) $l(\gamma_i; Z_{it}(\gamma_{i0}))$ and $G(\gamma_{i0}, \gamma_i)$ are now replaced by $\partial^{m_1+m_2} l(\gamma_i; Z_{it}(\gamma_{i0})) / \partial \gamma_{i,\tau_1}^{m_1} \partial \gamma_{i,\tau_2}^{m_2}$ 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 approaching one. Observing that

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

we have, for $K > \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}M(Z_{it}(\gamma_{i0}))$,

$$\begin{aligned}
&\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\frac{\partial^{m_1+m_2} l(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}}\right| > K\right) \\
\leq\ & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[M(Z_{it}(\gamma_{i0})) - \mathbb{E}M(Z_{it}(\gamma_{i0}))\right] > K - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}M(Z_{it}(\gamma_{i0}))\right) \\
\leq\ & \sup_{\gamma_0\in\Gamma_1^{\otimes}} \frac{var\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(Z_{it}(\gamma_{i0}))\right\}}{\left[K - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}M(Z_{it}(\gamma_{i0}))\right]^2} \\
=\ & o(1). \qquad \text{(A.11)}
\end{aligned}$$

■

Lemma A.2 *Let Assumptions 1–6 hold. Then for any $\varepsilon > 0$, there exists a $K > 0$ such that*

(*i*) $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{\left\|\left(\sqrt{nT}\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}(\gamma_{i0})\right\| > K\right\} \leq \varepsilon,$

(*ii*) $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{\left|\left(\sqrt{T}\right)^{-1}\sum_{t=1}^{T} v_{it}(\gamma_{i0})\right| > K\right\} \leq \varepsilon.$

Proof of Lemma A.2. We prove the first part only as the proof for the second part is similar. Since each time series u_{it} is a strong mixing sequence and the time series are independent from each other, $\{u_{11}, \ldots, u_{1T}, u_{21}, \ldots, u_{2T}, \ldots, u_{n1}, \ldots, u_{nT}\}$ is a strong mixing sequence with an exponentially decaying strong mixing numbers. Using the strong mixing moment inequality as before, we have

$$\mathbb{E}_{\gamma_0}\left\|\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}(\gamma_{i0})\right\|^p \leq C(nT)^{p/2}$$

for some constant C that does not depend on γ_0. Therefore

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left\{\left\|\left(\sqrt{nT}\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}(\gamma_{i0})\right\| > K\right\} \leq \frac{C}{K^p} \leq \varepsilon$$

when K is large enough. ■

To keep track of the approximation errors, we introduce a definition for a random variable $\hat{\mathcal{L}}(\gamma_0)$ that depends on γ_0. We write $\hat{\mathcal{L}}(\gamma_0) = o_p^U(1)$ if for any $\delta > 0$, $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\mathcal{L}}(\gamma_0)\right\| \geq \delta\right) = o(1)$. We write $\hat{\mathcal{L}}(\gamma_0) = O_p^U(1)$ if for any $\varepsilon > 0$, there exists a $K > 0$ such that

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|\hat{\mathcal{L}}(\gamma_0)\right\| \geq K\right) \leq \varepsilon.$$

Using this definition, we can write the results in Lemma A.1 as

$$\begin{aligned}
\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\frac{\partial^{m_1+m_2} l(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}} - \mathbb{E}\frac{\partial^{m_1+m_2} l(\gamma_{i0}; Z_{it}(\gamma_{i0}))}{\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}}\right] &= o_p^U(1), \\
\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\frac{\partial^{m_1+m_2} l(\breve{\gamma}_i; Z_{it}(\gamma_{i0}))}{\partial\gamma_{i,\tau_1}^{m_1}\partial\gamma_{i,\tau_2}^{m_2}}\right| &= O_p^U(1).
\end{aligned}$$

Similarly, when Lemma A.2 holds, we write

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}\left(\gamma_{i0}\right)=O_p^U\left(1\right),\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{it}\left(\gamma_{i0}\right)=O_p^U\left(1\right).$$

Proof of Theorem 2. We first establish a preliminary rate of convergence for $\hat{\theta}_{nT}$ and $\hat{\alpha}_i$. Since $l\left(\cdot,\cdot;Z_{it}\right)$ is continuously differentiable with respect to (θ,α_i), (θ_0,α_0) is an interior point of Γ, and $(\hat{\theta},\hat{\alpha})$ is consistent, the following first order conditions (FOCs) hold:

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}(\hat{\theta},\hat{\alpha}_i)=0 \text{ and } \frac{1}{T}\sum_{t=1}^{T}v_{it}(\hat{\theta},\hat{\alpha}_i)=0 \tag{A.12}$$

where for notational economy we write $\hat{\theta}=\hat{\theta}_{nT}$. Taking a first order Taylor expansion of each equation in the above FOCs, we have

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

where $(\breve{\theta},\breve{\alpha})$ may be different across different equations but for notational simplicity we use the generic notation $(\breve{\theta},\breve{\alpha})$. It then follows from Lemma A.1, Lemma A.2 and Theorem 1 that

$$\begin{aligned} O_p^U\left(\frac{1}{\sqrt{nT}}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}u_{it\theta}(\gamma_{i0})\right](\hat{\theta}-\theta_0) \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}v_{it\theta}\left(\gamma_{i0}\right)\right]'\left(\hat{\alpha}_i-\alpha_{i0}\right)+o_p^U\left(1\right)\left(\left\|\hat{\theta}-\theta_0\right\|+\max_i\left|\hat{\alpha}_i-\alpha_{i0}\right|\right) \end{aligned} \tag{A.13}$$

$$\begin{aligned} O_p^U\left(\frac{1}{\sqrt{T}}\right) &= \left[\mathbb{E}v_{it\theta}\left(\gamma_{i0}\right)\right](\hat{\theta}-\theta_0) \\ &\quad +\left[\mathbb{E}v_{it\alpha_i}(\gamma_{i0})\right]\left(\hat{\alpha}_i-\alpha_{i0}\right)+o_p^U\left(1\right)\left\|\hat{\theta}-\theta_0\right\|+o_p^U\left(1\right)\left|\hat{\alpha}_i-\alpha_{i0}\right| \end{aligned} \tag{A.14}$$

Equation (A.14) and Assumption 6(i) imply that

$$\hat{\alpha}_i-\alpha_{i0}=-\left[\mathbb{E}v_{it\alpha_i}(\gamma_{i0})\right]^{-1}\left[\mathbb{E}v_{it\theta}\left(\gamma_{i0}\right)\right](\hat{\theta}-\theta_0)+O_p^U\left(\frac{1}{\sqrt{T}}\right)+o_p^U\left(1\right)\left\|\hat{\theta}-\theta_0\right\|.$$

This, combined with (A.13), gives:

$$\begin{aligned} O_p^U\left(\frac{1}{\sqrt{nT}}\right) &= n^{-1}\sum_{i=1}^{n}\left\{\left[\mathbb{E}u_{it\theta}(\gamma_{i0})\right]-\left[\mathbb{E}v_{it\theta}\left(\gamma_{i0}\right)\right]'\left[\mathbb{E}v_{it\alpha_i}(\gamma_{i0})\right]^{-1}\left[\mathbb{E}v_{it\theta}\left(\gamma_{i0}\right)\right]\right\}(\hat{\theta}-\theta_0) \\ &\quad +O_p^U\left(\frac{1}{\sqrt{T}}\right)+o_p^U\left(1\right)\left\|\hat{\theta}-\theta_0\right\|. \end{aligned}$$

Using Assumption 6(ii), we can then deduce that

$$\hat{\theta} = \theta_0 + O_p^U\left(\frac{1}{\sqrt{T}}\right) \text{ and } \max_i |\hat{\alpha}_i - \alpha_{i0}| = O_p^U\left(\frac{1}{\sqrt{T}}\right).$$

We now proceed to establish the stochastic representation. Taking a second order Taylor expansion of $\sum_{i=1}^n \sum_{t=1}^T u_{it}(\hat{\theta}, \hat{\alpha}_i) = 0$, we have

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

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

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

Using Lemma A.1(ii), we can simplify the above equation as

$$0 = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it} + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\theta}(\hat{\theta}-\theta_0) + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\alpha_i}\left[\hat{\alpha}_i-\alpha_{i0}\right] + O_p^U\left(\frac{1}{T}\right). \tag{A.16}$$

Plugging

$$\hat{\alpha}_i - \alpha_{i0} = -\left[\frac{1}{T}\sum_{t=1}^T v_{it\alpha_i}(\breve{\theta},\breve{\alpha}_i)\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T v_{it} + \frac{1}{T}\sum_{t=1}^T v_{it\theta}(\breve{\theta},\breve{\alpha}_i)(\hat{\theta}-\theta_0)\right]$$

into (A.16) yields:

$$\begin{aligned} O_p^U\left(\frac{1}{T}\right) &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it} + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it\theta}(\hat{\theta}-\theta_0) \\ &- \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left[\frac{1}{T}\sum_{\tau=1}^T u_{i\tau\alpha_i}\right]\left[\frac{1}{T}\sum_{\tau=1}^T v_{i\tau\alpha_i}(\breve{\theta},\breve{\alpha}_i)\right]^{-1} v_{it} \\ &- \frac{1}{n}\sum_{i=1}^n\left[\frac{1}{T}\sum_{\tau=1}^T u_{i\tau\alpha_i}\right]\left[\frac{1}{T}\sum_{\tau=1}^T v_{i\tau\alpha_i}(\breve{\theta},\breve{\alpha}_i)\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T v_{it\theta}(\breve{\theta},\breve{\alpha}_i)\right](\hat{\theta}-\theta_0). \end{aligned}$$

Therefore

$$
\begin{aligned}
&\hat{\theta}-\theta_0 \\
= \quad & -\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it\theta}-\frac{1}{nT}\sum_{i=1}^{n}\left[\sum_{\tau=1}^{T}u_{i\tau\alpha_i}\right]\left[\sum_{\tau=1}^{T}v_{i\tau\alpha_i}(\breve{\theta},\breve{\alpha}_i)\right]^{-1}\left[\sum_{t=1}^{T}v_{it\theta}(\breve{\theta},\breve{\alpha}_i)\right]\right\}^{-1} \\
&\times\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it}-\left[\sum_{\tau=1}^{T}u_{i\tau\alpha_i}\right]\left[\sum_{\tau=1}^{T}v_{i\tau\alpha_i}(\breve{\theta},\breve{\alpha}_i)\right]^{-1}v_{it}\right\}+O_p^U\left(\frac{1}{T}\right). \qquad \text{(A.17)}
\end{aligned}
$$

It then follows that $\hat{\theta}-\theta_0$ can be represented as

$$
\hat{\theta}-\theta_0=\frac{1}{\sqrt{nT}}\tilde{A}_{nT}+\frac{1}{T}\tilde{B}_{nT}+o_p^U\left(\frac{1}{T}\right)
$$

for some matrices $\tilde{A}_{nT}=O_p^U(1)$ and $\tilde{B}_{nT}=O_p^U(1)$. As a consequence, $\hat{\alpha}_i-\alpha_{i0}$ can be represented as

$$
\hat{\alpha}_i-\alpha_{i0}=\frac{1}{\sqrt{nT}}\tilde{C}_i+\frac{1}{\sqrt{T}}\tilde{D}_i+\frac{1}{T}\tilde{E}_i+o_p^U\left(\frac{1}{T}\right) \qquad \text{(A.18)}
$$

for some matrices $\tilde{C}_i, \tilde{D}_i$ and $\tilde{E}_i$ that are all $O_p^U(1)$ and depend on (n,T).

We proceed to find $\tilde{A}_{nT}$ and $\tilde{B}_{nT}$. Using the above stochastic representations and taking a second order Taylor expansion of $\sum_{t=1}^{T}v_{it}(\hat{\theta},\hat{\alpha}_i)=0$, we have

$$
\begin{aligned}
o_p^U\left(\frac{1}{T}\right) \quad = \quad & \frac{1}{T}\sum_{t=1}^{T}v_{it}+\frac{1}{T}\sum_{t=1}^{T}v_{it\theta}\left[\frac{1}{\sqrt{nT}}\tilde{A}_{nT}+\frac{1}{T}\tilde{B}_{nT}\right]+\frac{1}{T}\sum_{t=1}^{T}v_{it\alpha_i}\left[\frac{1}{\sqrt{nT}}\tilde{C}_i+\frac{1}{\sqrt{T}}\tilde{D}_i+\frac{1}{T}\tilde{E}_i\right] \\
&+\frac{1}{2T}\sum_{t=1}^{T}v_{it\alpha_i\alpha_i}\left[\frac{1}{\sqrt{nT}}\tilde{C}_i+\frac{1}{\sqrt{T}}\tilde{D}_i+\frac{1}{T}\tilde{E}_i\right]^2 \\
= \quad & \left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{it}+\tilde{D}_i\frac{1}{T}\sum_{t=1}^{T}v_{it\alpha_i}\right\}\frac{1}{\sqrt{T}}+\left\{\left[\frac{1}{T}\sum_{t=1}^{T}v_{it\theta}\right]\tilde{A}_{nT}+\left[\frac{1}{T}\sum_{t=1}^{T}v_{it\alpha_i}\right]\tilde{C}_i\right\}\frac{1}{\sqrt{nT}} \\
&+\left\{\frac{1}{T}\sum_{t=1}^{T}v_{it\theta}\tilde{B}_{nT}+\tilde{E}_i\frac{1}{T}\sum_{t=1}^{T}v_{it\alpha_i}+\frac{1}{2T}\sum_{t=1}^{T}v_{it\alpha_i\alpha_i}\tilde{D}_i^2\right\}\frac{1}{T}.
\end{aligned}
$$

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

$$
\begin{aligned}
\tilde{D}_i \quad &= \quad -\sqrt{T}\frac{\sum_{t=1}^{T}v_{it}}{\sum_{t=1}^{T}v_{it\alpha_i}}+o_p^U\left(\frac{1}{\sqrt{T}}\right),\ \tilde{C}_i=-\frac{\left(\sum_{t=1}^{T}v_{it\theta}\right)A_{nT}}{\sum_{t=1}^{T}v_{it\alpha_i}}+o_p^U\left(\sqrt{\frac{n}{T}}\right), \\
\tilde{E}_i \quad &= \quad -\frac{\left(\sum_{t=1}^{T}v_{it\theta}\right)\tilde{B}_{nT}+\frac{1}{2}\left(\sum_{t=1}^{T}v_{it\alpha_i\alpha_i}\right)\tilde{D}_i^2}{\sum_{t=1}^{T}v_{it\alpha_i}}+o_p^U(1). \qquad \text{(A.19)}
\end{aligned}
$$

The $o_p^U(\cdot)$ terms above can be absorbed into the remainder term of order $o_p^U(1/T)$ in (A.18). We drop them from now on without loss of generality and redefine $\tilde{D}_i, \tilde{C}_i$ and $\tilde{E}_i$ as

$$\begin{aligned} \tilde{D}_i &= -\sqrt{T}\frac{\sum_{t=1}^T v_{it}}{\sum_{t=1}^T v_{it\alpha_i}},\ \tilde{C}_i = -\frac{\left(\sum_{t=1}^T v_{it\theta}\right) A_{nT}}{\sum_{t=1}^T v_{it\alpha_i}}, \\ \tilde{E}_i &= -\frac{\left(\sum_{t=1}^T v_{it\theta}\right) \tilde{B}_{nT} + \frac{1}{2}\left(\sum_{t=1}^T v_{it\alpha_i\alpha_i}\right) \tilde{D}_i^2}{\sum_{t=1}^T v_{it\alpha_i}}. \end{aligned} \tag{A.20}$$

Let

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

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

$$\begin{aligned} \tilde{D}_i &= D_i\left[1 + \frac{T^{-1}\sum_{t=1}^T (v_{it\alpha_i} - \mathbb{E}v_{it\alpha_i})}{\mathbb{E}\left[T^{-1}\sum_{t=1}^T v_{it\alpha_i}\right]}\right]^{-1} \\ &= D_i\left(1 - \frac{T^{-1}\sum_{t=1}^T (v_{it\alpha_i} - \mathbb{E}v_{it\alpha_i})}{\mathbb{E}\left[T^{-1}\sum_{t=1}^T v_{it\alpha_i}\right]} + O_p^U\left(\frac{1}{T}\right)\right) \\ &= D_i + \frac{1}{\sqrt{T}}\frac{T^{-1}\left[\sum_{t=1}^T v_{it}\right]\left[\sum_{t=1}^T (v_{it\alpha_i} - \mathbb{E}v_{it\alpha_i})\right]}{\left(\mathbb{E}\left[T^{-1}\sum_{t=1}^T v_{it\alpha_i}\right]\right)^2} + O_p^U\left(\frac{1}{T}\right). \end{aligned} \tag{A.21}$$

Next, using the above results and the second order Taylor expansion in (A.15), we have

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

Define

$$\tilde{H}_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(u_{it\theta} - \frac{\sum_{t=1}^{T} u_{it\alpha_i}}{\sum_{t=1}^{T} v_{it\alpha_i}} v_{it\theta}\right), \tag{A.22}$$

$$\tilde{S}_{nT} = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[u_{it} - \frac{\sum_{t=1}^{T} u_{it\alpha_i}}{\mathbb{E}\sum_{t=1}^{T} v_{it\alpha_i}} v_{it}\right], \tag{A.23}$$

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

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

Note 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,

$$\tilde{b}_{nT,1} = -\frac{1}{2n}\sum_{i=1}^{n}\left[\frac{\mathbb{E}T^{-1}\sum_{t=1}^{T} U_{it\alpha_i\alpha_i}}{\mathbb{E}T^{-1}\sum_{t=1}^{T} v_{it\alpha_i}}\right]\left[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} v_{it}\psi_{is}\right] + o_p^U(1), \tag{A.24}$$

$$\begin{aligned}
\tilde{b}_{nT,2} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\rho_{i0}\frac{T^{-1}\left[\sum_{t=1}^{T}\sum_{s=1}^{T} v_{it\alpha_i}v_{is}\right]}{\mathbb{E}\left[T^{-1}\sum_{t=1}^{T} v_{it\alpha_i}\right]} - \frac{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} u_{it\alpha_i}v_{is}}{\mathbb{E}\left[T^{-1}\sum_{t=1}^{T} v_{it\alpha_i}\right]}\right\} + o_p^U(1), \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} U_{it\alpha_i}\psi_{is}\right\} + o_p^U(1),
\end{aligned} \tag{A.25}$$

and so

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

Combining the above approximations, we have

$$\hat{\theta}_{nT} - \theta_0 = \frac{1}{\sqrt{nT}} A_{nT} + \frac{1}{T} B_{nT} + o_p^U\left(\frac{1}{T}\right) \tag{A.26}$$

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

$$\begin{aligned} C_i(\gamma_0) &= -\frac{\mathbb{E}v_{it\theta}}{\mathbb{E}v_{it\alpha_i}} A_{nT},\ D_i(\gamma_0) = -\frac{\sum_{t=1}^T v_{it}}{\sqrt{T}\mathbb{E}v_{it\alpha_i}} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_{it} \\ E_i(\gamma_0) &= -\frac{\mathbb{E}v_{it\theta} B_{nT} + \frac{1}{2}\left(\mathbb{E}v_{it\alpha_i\alpha_i}\right)\mathbb{E}\left(D_i^2(\gamma_0)\right)}{\mathbb{E}v_{it\alpha_i}} + \frac{T^{-1}\left[\sum_{t=1}^T \sum_{s=1}^T \mathbb{E}v_{it}\left(v_{is\alpha_i} - \mathbb{E}v_{is\alpha_i}\right)\right]}{\left(\mathbb{E}\left[v_{it\alpha_i}\right]\right)^2}, \end{aligned} \tag{A.27}$$

then

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

where all of $C_i(\gamma_0)$, $D_i(\gamma_0)$ and $E_i(\gamma_0)$ are of order $O_p^U(1)$. Now for any $\eta > 0$, we have

$$\begin{aligned} & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left\{\max_i \left|\hat{\alpha}_i - \alpha_{i0} - \left(\frac{1}{\sqrt{nT}} C_i(\gamma_0) + \frac{1}{\sqrt{T}} D_i(\gamma_0) + \frac{1}{T} E_i(\gamma_0)\right)\right| > \frac{\eta}{T}\right\} \\ = & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left\{\left|\hat{\alpha}_{\tilde{i}} - \alpha_{\tilde{i}0} - \left(\frac{1}{\sqrt{nT}} C_{\tilde{i}}(\gamma_0) + \frac{1}{\sqrt{T}} D_{\tilde{i}}(\gamma_0) + \frac{1}{T} E_{\tilde{i}}(\gamma_0)\right)\right| > \frac{\eta}{T}\right\} \\ = & o(1). \end{aligned} \tag{A.28}$$

Note that in the statement of the theorem, we do not use the nonstandard notation $O_p^U(\cdot)$ and $o_p^U(\cdot)$ but (A.26) and (A.28) are just alternative ways to present the same results. ■

To prove Theorem 3 and other results, we will use the following two lemmas repeatedly. Lemma A.3 helps translate the asymptotic results for the original sample into the corresponding ones for the bootstrap sample. Lemma A.4 shows that the effect of nonlinear truncation can be ignored in large samples.

Lemma A.3 *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.3. Since $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(\|\hat{\gamma} - \gamma_0\| \geq \nu) = o(a_{nT})$, we have $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(\hat{\gamma} \notin \Gamma_1^{\otimes}\right) =$

$o(a_{nT})$. Now

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(|\lambda_{nT}(\hat{\gamma})| \geq a_{nT}\varepsilon\right) \\
\leq & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(|\lambda_{nT}(\hat{\gamma})| \geq a_{nT}\varepsilon, \hat{\gamma} \in \Gamma_1^{\otimes}\right) + \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(\hat{\gamma} \notin \Gamma_1^{\otimes}\right) \\
\leq & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(\sup_{\gamma \in \Gamma_1^{\otimes}} |\lambda_{nT}(\gamma)| \geq a_{nT}\varepsilon\right) + o(a_{nT}) \\
= & \sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}\left(o(a_{nT}) \geq a_{nT}\varepsilon\right) + o(a_{nT}) = o(a_{nT}) \qquad \text{(A.29)}
\end{aligned}$$

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.4 *Let $\hat{J}_{nT}(\gamma)$ and $\tilde{J}_{nT}(\gamma)$ be two sequences of random vectors in $\mathbb{R}^{d_J}$ and indexed by $\gamma \in \Gamma_1^{\otimes}$. If (i) $\hat{J}_{nT}(\gamma_0) - \tilde{J}_{nT}(\gamma_0) = o_p^U(1)$; (ii) $\sup_{n,T} \sup_{\gamma_0 \in \Gamma_1^{\otimes}} (\mathbb{E}_{\gamma_0} \left\|\tilde{J}_{nT}(\gamma_0)\right\|^2) < \infty$, then as $M \to \infty$, $\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0} g_M[\hat{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0}\left[\tilde{J}_{nT}(\gamma_0)\right]\right\| = o(1)$.*

Proof of Lemma A.4. Since $\hat{J}_{nT}(\gamma_0) - \tilde{J}_{nT}(\gamma_0) = o_p^U(1)$, we deduce that

$$g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right) = o_p^U(1), \qquad \text{(A.30)}$$

because $g_M(\cdot)$ is continuous. That is, for any $\delta > 0$,

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0}\left(\left\|g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right\| \geq \delta\right) = o(1).$$

Using this result, we have

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0}\left[g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right]\right\| \\
\leq & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0}\left\|g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right\| 1\left\{\left\|g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right\| \geq \delta\right\} \\
& + \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0}\left\|g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right\| 1\left\{\left\|g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right\| < \delta\right\} \\
\leq & 2M\sqrt{d_J} \times o(1) + \delta \leq \delta + o(1),
\end{aligned}$$

for any $\delta > 0$. This implies that

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0}\left[g_M(\hat{J}_{nT}(\gamma_0)) - g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right]\right\| = o(1) \qquad \text{(A.31)}$$

for any fixed M. 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 $\mathcal{R}_{nT,M}(\gamma_0) := \mathbb{E}_{\gamma_0}\left[g_M\left(\tilde{J}_{nT}(\gamma_0)\right) - \tilde{J}_{nT}(\gamma_0)\right]$ satisfies

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \|\mathcal{R}_{nT,M}(\gamma_0)\| \\
\leq \ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\| 1\left\{\left\|\tilde{J}_{nT}(\gamma_0)\right\| > \sqrt{d_J}M\right\} \\
\leq \ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\|^2\right)^{1/2} \left(P_{\gamma_0}\left\{\left\|\tilde{J}_{nT}(\gamma_0)\right\| > \sqrt{d_J}M\right\}\right)^{1/2} \\
\leq \ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\|^2\right)^{1/2} \left(\mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\|^2 / \left(d_J M^2\right)\right)^{1/2} \\
\leq \ & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\|^2\right) / \left(\sqrt{d_J}M\right) \\
\leq \ & \sup_{n,T} \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left(\mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\|^2\right) / \left(\sqrt{d_J}M\right) \leq C/M \qquad \text{(A.32)}
\end{aligned}$$

for some constant C that does not depend on n, T or γ_0. That is,

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0}\left[g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right] - \mathbb{E}_{\gamma_0}\left[\tilde{J}_{nT}(\gamma_0)\right]\right\| \leq C/M$$

As a consequence, for any $\varepsilon > 0$, there exists an M_0 that does not depend on (n, T) such that

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0}\left[g_M\left(\tilde{J}_{nT}(\gamma_0)\right)\right] - \mathbb{E}_{\gamma_0}\left[\tilde{J}_{nT}(\gamma_0)\right]\right\| \leq \varepsilon$$

for $M \geq M_0$ and for all n and T. Fix an $M^+ \geq M_0$ and any $\varepsilon > 0$, we have

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0} g_{M^+}[\hat{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0} g_{M^+}\left[\tilde{J}_{nT}(\gamma_0)\right]\right\| \leq \varepsilon$$

when (n, T) are large enough. So for any $\varepsilon > 0$,

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0} g_{M^+}[\hat{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0}\left[\tilde{J}_{nT}(\gamma_0)\right]\right\| \leq 2\varepsilon.$$

This implies that for some $M \to \infty$,

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\|\mathbb{E}_{\gamma_0} g_M[\hat{J}_{nT}(\gamma_0)] - \mathbb{E}_{\gamma_0}\left[\tilde{J}_{nT}(\gamma_0)\right]\right\| = o(1).$$

■

Proof of Theorem 3. **Part (i).** Let $\hat{\theta}_{nT}(\gamma)$ be the estimator of θ based on the sample $\{Z_{i1}(\gamma)\} := \{Z_{i1}(\gamma), ..., Z_{iT}(\gamma)\}_{i=1}^n$ 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 $\{Z^*_{i1}(\gamma)\} := \{Z^*_{i1}(\gamma), ..., Z^*_{iT}(\gamma)\}_{i=1}^n$ generated under the bootstrap DGP with parameter value $\gamma = (\theta, \alpha)$. To apply Lemma A.3 with $a_{nT} = 1$, 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 $\delta>0$. By Theorem 1, $\sup_{\gamma_0\in\Gamma_0^{\otimes}}P_{\gamma_0}(\|\hat{\gamma}-\gamma_0\|\geq\nu)=o(1)$. Invoking Lemma A.3 yields:

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

for any $\varepsilon>0$.

Part (ii). We apply Lemma A.4 by defining

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

By Theorem 2, $\hat{J}_{nT}(\gamma_0)-\tilde{J}_{nT}(\gamma_0)=o_p^U(1)$. It is also easy to show that $\sup_{n,T}\sup_{\gamma_0\in\Gamma_1^{\otimes}}(\mathbb{E}_{\gamma_0}\left\|\tilde{J}_{nT}(\gamma_0)\right\|^2)<\infty$. So

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}}\left\|\mathbb{E}_{\gamma_0}g_M[\hat{J}_{nT}(\gamma_0)]-\mathbb{E}_{\gamma_0}\tilde{J}_{nT}(\gamma_0)\right\|=o(1)$$

and

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

as desired.

Part (iii). 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 Part (i), $\lambda_{nT}(\gamma)=\lambda_{nT}^{*}(\gamma)$. It follows from Part (ii) that $\sup_{\gamma_0\in\Gamma_1^{\otimes}}\lambda_{nT}(\gamma_0)=o(1)$. Hence, by Lemma A.3, $\sup_{\gamma_0\in\Gamma_0^{\otimes}}P_{\gamma_0}(\lambda_{nT}^{*}(\hat{\gamma})\geq\varepsilon)=o(1)$.

Part (iv). HK have proved that $A_{nT}(\gamma_0)\xrightarrow{d}N[0,\Omega(\gamma_0)]$. Under Assumptions 2, 3 and 6, the result can be strengthened to

$$\sup_{\gamma_0\in\Gamma_1^{\otimes}}\left|P_{\gamma_0}(c'A_{nT}(\gamma_0)/\sqrt{c'\Omega(\gamma_0)c}<\tau)-\Phi(\tau)\right|=o(1)\tag{A.33}$$

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

$$
\begin{aligned}
& P\left\{\sqrt{nT}(\tilde{\theta}_{nT}-\theta_0)<\vartheta\right\} \\
= \; & P\left\{\sqrt{nT}(\tilde{\theta}_{nT}-\theta_0)<\vartheta, \hat{\gamma}\in\Gamma_1^{\otimes}\right\}+P\left\{\sqrt{nT}(\tilde{\theta}_{nT}-\theta_0)<\vartheta, \hat{\gamma}\notin\Gamma_1^{\otimes}\right\} \\
= \; & P\left\{\sqrt{nT}\left[\hat{\theta}_{nT}-\theta_0\right]-\mathbb{E}^*_{\hat{\gamma}}g_M\left[\sqrt{nT}\left(\hat{\theta}^*_{nT}-\hat{\theta}_{nT}\right)\right]<\vartheta,\ \hat{\gamma}\in\Gamma_1^{\otimes}\right\}+o\left(1\right) \\
= \; & P\left[A_{nT}(\gamma_0)-\sqrt{n/T}\left[B_{nT}\left(\hat{\gamma}\right)-B_{nT}(\gamma_0)\right]<\vartheta,\ \hat{\gamma}\in\Gamma_1^{\otimes}\right]+o\left(1\right) \\
= \; & P\left[A_{nT}(\gamma_0)<\vartheta,\hat{\gamma}\in\Gamma_1^{\otimes}\right]+o\left(1\right) \\
= \; & P\left[A_{nT}(\gamma_0)<\vartheta\right]-P\left[A_{nT}(\gamma_0)<\vartheta,\hat{\gamma}\notin\Gamma_1^{\otimes}\right]+o\left(1\right) \\
= \; & P\left[A_{nT}(\gamma_0)<\vartheta\right]+o(1)\to N\left[0,\Omega\left(\gamma_0\right)\right].
\end{aligned}
\tag{A.34}
$$

Because of (A.33) and that the $o(1)$ terms in the above equation hold uniformly over $\gamma_0\in\Gamma_0^{\otimes}$, we have

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

■

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

$$
\hat{\gamma}_k=\hat{\gamma}_{k-1}-\left[\mathcal{H}\left(\hat{\gamma}_{k-1};Z\right)\right]^{-1}\mathcal{S}\left(\hat{\gamma}_{k-1};Z\right) \tag{A.35}
$$

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

$$
\mathcal{H}\left(\hat{\gamma}_{k-1};Z\right) = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left.\frac{\partial^2\log l\left(\theta,\alpha_i;Z_{it}\right)}{\partial\left(\theta',\alpha'\right)'\partial\left(\theta',\alpha'\right)}\right|_{\theta=\hat{\theta}_{k-1},\alpha_i=\hat{\alpha}_{i,k-1}} \tag{A.36}
$$

$$
\mathcal{S}\left(\hat{\gamma}_{k-1};Z\right) = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left.\frac{\partial\log l\left(\theta,\alpha_i;Z_{it}\right)}{\partial\left(\theta',\alpha'\right)'}\right|_{\theta=\hat{\theta}_{k-1},\alpha_i=\hat{\alpha}_{i,k-1}} \tag{A.37}
$$

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

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

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

$$
\begin{aligned}
\hat{\gamma}_k-\hat{\gamma} &= \hat{\gamma}_{k-1}-\left[\mathcal{H}(\hat{\gamma}_{k-1};Z)\right]^{-1}\mathcal{S}\left(\hat{\gamma}_{k-1};Z\right)-\hat{\gamma} \\
&= \left[\mathcal{H}(\hat{\gamma}_{k-1};Z)\right]^{-1}\left[\mathcal{S}\left(\hat{\gamma};Z\right)-\mathcal{S}\left(\hat{\gamma}_{k-1};Z\right)-\mathcal{H}(\hat{\gamma}_{k-1};Z)\left(\hat{\gamma}-\hat{\gamma}_{k-1}\right)\right] \\
&= \frac{1}{2}\left[\mathcal{H}(\hat{\gamma}_{k-1};Z)\right]^{-1}\xi\left(\hat{\gamma}^{\dagger}_{k-1};Z\right)
\end{aligned}
\tag{A.38}
$$

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

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

and

$$\begin{aligned}\mathcal{H}_{\gamma_u}(\gamma;Z)&=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial}{\partial\gamma_u}\frac{\partial^2\log(\gamma;Z_{it})}{\partial\gamma\partial\gamma'}\\&:=\begin{pmatrix}\mathcal{H}_{\theta\theta\gamma_u}(\gamma;Z)&\mathcal{H}_{\theta\alpha\gamma_u}(\gamma;Z)\\\mathcal{H}_{\alpha\theta\gamma_u}(\gamma;Z)&\mathcal{H}_{\alpha\alpha\gamma_u}(\gamma;Z)\end{pmatrix}.\end{aligned}\tag{A.39}$$

A more explicit expression for $\xi\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)$ is

$$\begin{aligned}\xi_u\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)&=\left(\hat{\theta}_{k-1}-\hat{\theta}\right)'\mathcal{H}_{\theta\theta\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\theta}_{k-1}-\hat{\theta}\right)\\&+2\left(\hat{\theta}_{k-1}-\hat{\theta}\right)'\mathcal{H}_{\theta\alpha\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right)\\&+\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right)'\mathcal{H}_{\alpha\alpha\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right).\end{aligned}\tag{A.40}$$

Hence with probability approaching one uniformly over $\gamma_0\in\Gamma_1^{\otimes}$,

$$\begin{aligned}\left\|\xi_u\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\right\|&\le\left\|2\left(\hat{\theta}_{k-1}-\hat{\theta}\right)'\mathcal{H}_{\theta\theta\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\theta}_{k-1}-\hat{\theta}\right)\right\|\\&+\left\|2\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right)'\mathcal{H}_{\alpha\alpha\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\left(\hat{\alpha}_{k-1}-\hat{\alpha}\right)\right\|\\&\le2\left\|\mathcal{H}_{\theta\theta\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z\right)\right\|\left\|\hat{\theta}_{k-1}-\hat{\theta}\right\|^2\\&+\left\|\frac{2}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}\mathcal{H}_{\alpha_i\alpha_i\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z_{it}\right)\right]\left(\hat{\alpha}_{i,k-1}-\hat{\alpha}_i\right)^2\right\|\end{aligned}$$

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

$$\|\hat{\gamma}_k-\hat{\gamma}\|\le\zeta_{\theta,k-1}\left\|\hat{\theta}_{k-1}-\hat{\theta}\right\|^2+\zeta_{\alpha,k-1},\tag{A.41}$$

where

$$\zeta_{\theta,k-1}=\left\|\left[\mathcal{H}\left(\hat{\gamma}_{k-1};Z\right)\right]^{-1}\right\|\sum_{u=1}^{d_\gamma}\left\|\mathcal{H}_{\theta\theta\gamma_u}(\hat{\gamma}_{k-1}^{\dagger};Z)\right\|,\tag{A.42}$$

$$\zeta_{\alpha,k-1}=\left\|\left[\mathcal{H}\left(\hat{\gamma}_{k-1};Z\right)\right]^{-1}\right\|\sum_{u=1}^{d_\gamma}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\mathcal{H}_{\alpha_i\alpha_i\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z_{it}\right)\right|\left(\hat{\alpha}_{i,k-1}-\hat{\alpha}_i\right)^2.\tag{A.43}$$

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

$$
\begin{aligned}
& P_{\gamma_0}\left(T\left\|\hat{\gamma}_k-\hat{\gamma}\right\| \geq K\right) \\
\leq\; & P_{\gamma_0}\left(T\zeta_{\theta,k-1}\left\|\hat{\theta}_{k-1}-\hat{\theta}\right\|^2+T\zeta_{\alpha,k-1} \geq K\right) \\
\leq\; & P_{\gamma_0}\left(T\zeta_{\theta,k-1}\left\|\hat{\theta}_{k-1}-\hat{\theta}\right\|^2 \geq 0.5K\right)+P_{\gamma_0}\left(T\zeta_{\alpha,k-1} \geq 0.5K\right) \\
\leq\; & P_{\gamma_0}\left(T\left\|\hat{\theta}_{k-1}-\hat{\theta}\right\|^2 \geq \sqrt{0.5K}\right)+P_{\gamma_0}\left(\zeta_{\theta,k-1} \geq \sqrt{0.5K}\right)+P_{\gamma_0}\left(T\zeta_{\alpha,k-1} \geq 0.5K\right)
\end{aligned} \tag{A.44}
$$

Observing that

$$
\begin{aligned}
& P_{\gamma_0}\left(\zeta_{\theta,k-1} \geq \sqrt{0.5K}\right) \\
& = P_{\gamma_0}\left(\left\|\left(\mathcal{H}(\hat{\gamma}_{k-1};Z)\right)^{-1}\right\| \sum_{u=1}^{d_\gamma}\left\|\mathcal{H}_{\theta\theta\gamma_u}(\hat{\gamma}_{k-1}^{\dagger};Z)\right\| \geq \sqrt{0.5K}\right) \leq I_1+I_2,
\end{aligned} \tag{A.45}
$$

where

$$
I_1 = P_{\gamma_0}\left(\sum_{u=1}^{d_\gamma}\left\|n^{-1}\mathcal{H}_{\theta\theta\gamma_u}(\hat{\gamma}_{k-1}^{\dagger};Z)\right\| \geq (0.5K)^{1/4}\right), \tag{A.46}
$$

and

$$
I_2 = P_{\gamma_0}\left(\left\|\left(n^{-1}\mathcal{H}(\hat{\gamma}_{k-1};Z)\right)^{-1}\right\| \geq (0.5K)^{1/4}\right).
$$

Note that $\hat{\gamma}_{k-1}^{\dagger}$ is between $\hat{\gamma}_{k-1}$ and $\hat{\gamma}$ and $\hat{\gamma}=\gamma_0+o_p(1)$ uniformly. Using Lemma A.1, we have $\sup_{\gamma_0\in\Gamma_1^{\otimes}} I_1 = o(1)$, provided that K is large enough. Using the same argument, we can show that, when K is large enough, $\sup_{\gamma_0\in\Gamma_1^{\otimes}} I_2 = o(1)$. We have thus proved that

$$
\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\zeta_{\theta,k-1} \geq \sqrt{0.5K}\right) = o(1), \tag{A.47}
$$

when K is large enough. Therefore

$$
\begin{aligned}
& \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(T\left\|\hat{\gamma}_k-\hat{\gamma}\right\| \geq K\right) \\
\leq\; & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(T\left\|\hat{\theta}_{k-1}-\hat{\theta}\right\|^2 \geq \sqrt{0.5K}\right)+\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(T\zeta_{\alpha,k-1} \geq 0.5K\right)+o(1). \\
=\; & \sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(T\zeta_{\alpha,k-1} \geq 0.5K\right)+o(1)
\end{aligned} \tag{A.48}
$$

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

In view of $\sup_{\gamma_0\in\Gamma_1^{\otimes}} I_2 = o(1)$, to show that $\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(T\zeta_{\alpha,k-1} \geq 0.5K\right)=o(1)$, it suffices to prove

$$
\sup_{\gamma_0\in\Gamma_1^{\otimes}} P_{\gamma_0}\left(\frac{T}{n}\sum_{u=1}^{d_\gamma}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\mathcal{H}_{\alpha_i\alpha_i\gamma_u}\left(\hat{\gamma}_{k-1}^{\dagger};Z_{it}\right)\right|\left(\hat{\alpha}_{i,k-1}-\hat{\alpha}_i\right)^2 \geq (0.5K)^{3/4}\right)=o(1).
$$

Using Theorem 2, it is easy to show that there exists a constant $C_\alpha > 0$ such that

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left[\max_i T \left(\hat{\alpha}_{i,k-1} - \hat{\alpha}_i \right)^2 > C_\alpha \right] = o\left(1\right)$$

and so when K is large enough,

$$\begin{aligned}
& \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\frac{T}{n} \sum_{u=1}^{d_\gamma} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left| \mathcal{H}_{\alpha_i \alpha_i \gamma_u} \left(\hat{\gamma}_{k-1}^{\dagger}; Z_{it} \right) \right| \left(\hat{\alpha}_{i,k-1} - \hat{\alpha}_i \right)^2 \geq (0.5K)^{3/4} \right) \\
\leq \; & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\frac{1}{n} \sum_{u=1}^{d_\gamma} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left| \mathcal{H}_{\alpha_i \alpha_i \gamma_u} \left(\hat{\gamma}_{k-1}^{\dagger}; Z_{it} \right) \right| \max_i T \left(\hat{\alpha}_{i,k-1} - \hat{\alpha}_i \right)^2 \geq (0.5K)^{3/4} \right) \\
\leq \; & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(\max_u \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left| \mathcal{H}_{\alpha_i \alpha_i \gamma_u} \left(\hat{\gamma}_{k-1}^{\dagger}; Z_{it} \right) \right| \geq (0.5K)^{3/4} / C_\alpha \right) + o(1) \\
\leq \; & \sup_{\gamma_0 \in \Gamma_1^{\otimes}} \sum_{u=1}^{d_\gamma} P_{\gamma_0} \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left| \mathcal{H}_{\alpha_i \alpha_i \gamma_u} \left(\hat{\gamma}_{k-1}^{\dagger}; Z_{it} \right) \right| \geq (0.5K)^{3/4} / C_\alpha \right) + o(1) \\
= \; & o(1) \qquad \text{(A.49)}
\end{aligned}$$

where the last equality holds by the same arguments leading to Lemma A.1 but with a better bound as in the proof of Theorem 1.

To sum up the above steps, we have proved that for $k = 1$,

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

when K is large enough.

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

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} P_{\gamma_0} \left(T^{2^{k-1}} \left\| \hat{\gamma}_k \left(\gamma_0\right) - \hat{\gamma}\left(\gamma_0\right) \right\| > K \right) = o\left(1\right) \text{ for all } k \geq 1 \qquad \text{(A.50)}$$

where we have written $\hat{\gamma}_k := \hat{\gamma}_k\left(\gamma_0\right)$ and $\hat{\gamma} := \hat{\gamma}\left(\gamma_0\right)$ to emphasize their dependence on the true parameter γ_0. In particular, when $k \geq 2$.

$$\begin{aligned}
\sqrt{nT}\left[\hat{\theta}_{nT,k}\left(\gamma_0\right) - \theta_0\right] &= \sqrt{nT}\left[\hat{\theta}_{nT}\left(\gamma_0\right) - \theta_0\right] + O_p^U\left(\sqrt{nT} T^{-2^{k-1}}\right) \\
&= \sqrt{nT}\left[\hat{\theta}_{nT}\left(\gamma_0\right) - \theta_0\right] + o_p^U\left(1\right) \\
&= A_{nT}\left(\gamma_0\right) + \sqrt{n/T} B_{nT}\left(\gamma_0\right) + o_p^U\left(1\right). \qquad \text{(A.51)}
\end{aligned}$$

Letting

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

and invoking Lemma A.4, we have, for $k \geq 2$:

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

Combining this with Theorem 3(ii) yields

$$\sup_{\gamma_0 \in \Gamma_1^{\otimes}} \left\| \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT,k}(\gamma_0) - \theta_0 \right) \right] - \mathbb{E}_{\gamma_0} g_M \left[\sqrt{nT} \left(\hat{\theta}_{nT} - \theta_0 \right) \right] \right\| = o(1) \tag{A.52}$$

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.3 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 equation (24), the theorem follows from Theorems 3 and 4. ■

Proof of Theorem 6. Part (i). Under the smooth assumption on $\Delta(w, \tilde{Z}_{it}, \gamma_i)$, we have

$$\begin{aligned}
& \sqrt{nT}\left[\hat{\mu}(w) - \mu(w)\right] \\
= \; & \frac{\sqrt{nT}}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[\Delta(w, \tilde{Z}_{it}, \hat{\theta}_{nT}, \hat{\alpha}_i) - \Delta(w, \tilde{Z}_{it}, \theta_0, \alpha_{i0}) \right] \\
= \; & \left[\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta_\theta(w, \tilde{Z}_{it}, \gamma_{i0}) \right]' \sqrt{nT} \left(\hat{\theta}_{nT} - \theta_0 \right) + \frac{\sqrt{nT}}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta_{\alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0}) \left(\hat{\alpha}_i - \alpha_{i0} \right) \\
& + \frac{\sqrt{nT}}{2nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \Delta_{\alpha_i \alpha_i}(w, \tilde{Z}_{it}, \gamma_{i0}) \left(\hat{\alpha}_i - \alpha_{i0} \right)^2 + o_p^U(1).
\end{aligned} \tag{A.53}$$

Using Theorem 2, we have

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

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

and

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

So

$$
\sqrt{nT}\left[\hat{\mu}(w)-\mu(w)\right]=A_{nT}^{\mu}\left(\gamma_0\right)+\sqrt{\frac{n}{T}}B_{nT}^{\mu}\left(\gamma_0\right)+o_p^U\left(1\right).
$$

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.33) 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 $\gamma_0\in\Gamma_0^{\otimes}$.

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

$$
\sup_{\gamma_0\in\Gamma_1^{\otimes}}\left|\mathbb{E}_{\gamma_0}g_M\left[\sqrt{nT}\left(\hat{\mu}(w)-\mu(w)\right)\right]-\sqrt{\frac{n}{T}}B_{nT}^{\mu}\left(\gamma_0\right)\right|=o\left(1\right). \tag{A.54}
$$

This, combined with Lemma A.3, (A.54), 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

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

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

$$
\begin{aligned}
\sqrt{nT}\left[\tilde{\mu}(w) - \mu(w)\right] &= \sqrt{nT}\left[\hat{\mu}(w) - \mu(w)\right] - \mathbb{E}^*_{\hat{\gamma}} g_M \left[\sqrt{nT}\left(\hat{\mu}^*(w) - \hat{\mu}(w)\right)\right] \\
&= A^{\mu}_{nT}(\gamma_0) + \sqrt{\frac{n}{T}} B^{\mu}_{nT}(\gamma_0) - \sqrt{\frac{n}{T}} B^{\mu}_{nT}(\hat{\gamma}) + o_p(1) \\
&= A^{\mu}_{nT}(\gamma_0) + o_p(1) \xrightarrow{d} N\left[0, \Omega^{\mu}(\gamma_0)\right], \qquad \text{(A.55)}
\end{aligned}
$$

uniformly in that the underlying probability errors converge to zero uniformly over $\gamma_0 \in \Gamma^{\otimes}_0$.

Part (iii). It suffices to show that

$$
\sup_{\gamma_0 \in \Gamma^{\otimes}_0} P_{\gamma_0} \left\{ \left| \mathbb{E}^*_{\hat{\gamma}} g_M \left[\sqrt{nT}\left(\hat{\mu}^*(w) - \hat{\mu}(w)\right)\right] - \mathbb{E}^*_{\hat{\gamma}} g_M \left[\sqrt{nT}\left(\hat{\mu}^*_k(w) - \hat{\mu}(w)\right)\right] \right| \geq \delta \right\} = o(1),
$$

for $k \geq 2$. In view of Lemma A.3, this holds if

$$
\sup_{\gamma_0 \in \Gamma^{\otimes}_1} \left\| \mathbb{E} g_M \left[\sqrt{nT}\left(\hat{\mu}(w) - \mu(w)\right)\right] - \mathbb{E} g_M \left[\sqrt{nT}\left(\hat{\mu}_k(w) - \hat{\mu}(w)\right)\right] \right\| = o(1) \qquad \text{(A.56)}
$$

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.35). Noting for some $\check{\gamma}_i$ between $\hat{\gamma}_i$ and γ_{i0},

$$
\begin{aligned}
&\sqrt{nT}\left(\hat{\mu}(w) - \mu(w)\right) - \sqrt{nT}\left(\hat{\mu}_k(w) - \mu(w)\right) \\
&= \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[\Delta_\gamma(w, \tilde{Z}_{it}, \check{\gamma}_i) \sqrt{nT}\left(\hat{\gamma}_i - \hat{\gamma}_{i,k}\right)\right] = o^U_p(1) \qquad \text{(A.57)}
\end{aligned}
$$

for $k \geq 2$, (A.56) follows from the same argument for proving (A.52). ■

Proof of Theorem 7. Part (i). We first introduce the notations $\Gamma_{1/2} = \{\gamma_i \in \Gamma : \|\gamma_i - \Gamma_0\| < \nu/2\}$ and $\Gamma^{\otimes}_{1/2} = \left\{\gamma = (\theta, \alpha_1, \ldots, \alpha_n) : (\theta, \alpha_i) \in \Gamma_{1/2}\right\}$. Let $\Phi[\cdot;\Omega]$ be the CDF of the multivariate normal distribution with variance matrix Ω. Define $\lambda_{nT}(\gamma_0) = \sup_{\vartheta} \left| P_{\gamma_0}\left[\sqrt{nT}(\tilde{\theta}_{nT,k} - \theta_0) < \vartheta\right] - \Phi\left[\vartheta; \Omega(\gamma_0)\right] \right|$. It is not hard to show that $\sup_{\gamma_0 \in \Gamma^{\otimes}_{1/2}} |\lambda_{nT}(\gamma_0)| = o(1)$ and hence $\sup_{\gamma_0 \in \Gamma^{\otimes}_0} P_{\gamma_0}\left(|\lambda^*_{nT}(\hat{\gamma})| > \delta/2\right) =$

$o(1)$ by an argument similar to Lemma A.3. But

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

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

$$\begin{aligned}
&\sup_{\gamma_0\in\Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\hat{\gamma}}^{*}\left[\sqrt{nT}(\tilde{\theta}_{nT,k}^{*}-\hat{\theta}_{nT})<\vartheta\right]-P_{\gamma_0}\left[\sqrt{nT}(\tilde{\theta}_{nT,k}-\theta_0)<\vartheta\right]\right|\geq\delta\right\} \\
\leq&\sup_{\gamma_0\in\Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\hat{\gamma}}^{*}\left[\sqrt{nT}(\tilde{\theta}_{nT,k}^{*}-\hat{\theta}_{nT})<\vartheta\right]-\Phi\left[\vartheta,\Omega(\gamma_0)\right]\right|\geq\frac{\delta}{2}\right\} \\
&+\sup_{\gamma_0\in\Gamma_0^{\otimes}} P_{\gamma_0}\left\{\sup_{\vartheta}\left|P_{\gamma_0}\left[\sqrt{nT}(\tilde{\theta}_{nT,k}-\theta_0)<\vartheta\right]-\Phi\left[\vartheta,\Omega(\gamma_0)\right]\right|\geq\frac{\delta}{2}\right\} \\
=&\ o(1).
\end{aligned}$$

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 $\sqrt{nT}(\tilde{\theta}_{nT,kk}-\theta_0)=\left(\hat{\theta}_{nT,k}-\theta_0\right)-\mathbb{E}_{\hat{\gamma}_k}^{*}g_M\left[\sqrt{nT}\left(\hat{\theta}_{nT,kk}^{*}-\hat{\theta}_{nT,k}\right)\right]$. Using exactly the same argument that leads to Theorem 3(iii), 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 $\gamma_0\in\Gamma_{1/2}^{\otimes}$. Using this and (A.50), we have for $k\geq 2$

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

Combining this with Poyla's lemma, we have $\sup_{\gamma_0 \in \Gamma_{1/2}^{\otimes}} |\lambda_{nT}(\gamma_0)| = o(1)$. It then follows from an argument similar to Lemma A.3 that $\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0}(|\lambda_{nT}^*(\hat{\gamma})| > \delta/2) = o(1)$. The rest of the proof is the same as that in the proof of Part (i) and is omitted here. ■

Proof of Equation (52). It follows from Theorem 7 (i) that

$$\sup_{\gamma_0 \in \Gamma_0^{\otimes}} P_{\gamma_0} \left\{ \sup_{\tau} \left| P_{\hat{\gamma}}^* \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k}^* - \hat{\theta}_{nT}) \geq \tau \right] - P_{\gamma_0} \left[\sqrt{nT} c'(\tilde{\theta}_{nT,k} - \theta_0) \geq \tau \right] \right| \geq \delta \right\} = o(1).$$

So

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

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,

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

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

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

■

References

[1] Andersen, E. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society*, Series B, 32(2):283–301.

[2] Andrews, D. (2002). Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators. *Econometrica*, 70(1):119–162.

[3] 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.

[4] Arellano, M. and Hahn, J. (2006). Understanding Bias in Nonlinear Panel Models: Some Recent Developments. *Advances in Economics and Econometrics*, Ninth World Congress, Cambridge University Press.

[5] Amemiya, T. (1985). Advanced Econometrics, Harvard University Press.

[6] Bester, A. and Hansen, C. (2009). A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects. *Journal of Business & Economic Statistics*, 27(2): 131-148.

[7] Davidson, R. and MacKinnon, J. (1999). Bootstrap Testing in Nonlinear Models. *International Economic Review*, 40(2):487–508.

[8] Davidson, R. and MacKinnon, J. (2002). Fast Double Bootstrap Tests of Nonnested Linear Regression Models. *Econometric Review*, 21(4): 419-429.

[9] Davidson, R. and MacKinnon, J. (2007). Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap. *Computational Statistics and Data Analysis*, 51(7): 3259-3281.

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

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

[12] Fernandez-Val, I. (2009). Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models. *Journal of Econometrics*, 150(1): 71-85.

[13] Giacomini, R., Politis, D. and White, H. (2013): A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators, *Econometric Theory*, forthcoming.

[14] Greene, W. (2004). The Behavior of the Fixed Effects Estimator in Nonlinear Models. *The Econometrics Journal*, 7(1):98–119.

[15] Hahn, J. and Kuersteiner, G. (2011). Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. *Econometric Theory,* 27(6):1152-1191.

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

[17] Hahn, J. and Newey, W. (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica*, 72(4):1295–1319.

[18] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer Verlag.

[19] 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.*

[20] Honore, B. and Kyriazidou, E. (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica*, 68(4):839–874.

[21] Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2):391–413.

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

[23] Pace, L. and Salvan, A. (2006). Adjustments of the Profile Likelihood from a New Perspective. *Journal of Statistical Planning and Inference*, 136(10):3554–3564.

[24] Yokoyama, R. (1980): Moment Bounds for Stationary and Mixing Sequences. *Z. Wahrsch. Verw. Gebiete*, 52: 45–57.