Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects

Jinyong Hahn
UCLA

Guido Kuersteiner
MIT

April, 2003

ACKNOWLEDGMENT: We are grateful for helpful comments by Gary Chamberlain, Shakeeb Khan, Whitney Newey, and workshop participants of Federal Reserve Board, and Harvard/MIT. The second author gratefully acknowledges Financial support from NSF Grant SES 0095132.

Abstract

The fixed effects estimator of panel models can be severely biased because of the well-known incidental parameter problems. It is shown that such bias can be reduced as T grows with n. We consider asymptotics where n and T grow at the same rate as an approximation that allows us to compare bias properties. Under these asymptotics the bias corrected estimators are centered at the truth, whereas the fixed effects estimator is not. We also show how our alternative asymptotics is related to the higher order "large T" asymptotics.

1 Introduction

Panel data, consisting of observations across time for different individual economic agents, allows the possibility of controlling for unobserved individual heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for such heterogeneity can result in misleading inferences. One way of attempting to deal with the presence of unobserved individual effects in nonlinear models is to treat each such effect as a separate parameter to be estimated. Unfortunately, such estimators are typically subject to the incidental parameters problem noted by Neyman and Scott (1948). The estimators of the parameters of interest will be inconsistent if the number of individuals goes to infinity while the number of time periods is held fixed. In practical terms, the incidental parameters problem suggests that fixed effects estimators may be severely biased.

Hahn and Kuersteiner (2002) recently showed that the bias in a panel AR(1) model can be alleviated substantially by considering an alternative approximation where the number of individuals (n) and the number of time series observations (T) grow to infinity at the same rate. Hahn and Newey (2002) showed how the bias correction can be implemented in a static nonlinear panel data model with fixed effects under the same asymptotics. Alternative asymptotics are implicitly based on the intuition that the bias of fixed effects estimators should change with the number of time series observations.

To be more precise, let θ denote an R-dimensional parameter of interest, and let $\theta_{[T]}$ denote the limiting value, as the number of individuals goes to infinity, of a fixed effects estimator $\widehat{\theta}$ of θ. A consequence of the incidental parameters problem is that typically $\theta_{[T]} \neq \theta$. Note that the bias should be small for large enough T, i.e., $\lim_{T\to\infty} \theta_{[T]} = \theta$. Suppose that $\theta_{(T)}$ also obeys the relation $\theta_{[T]} = \theta + T^{-1}\beta + O\left(T^{-2}\right)$. If a reasonably precise estimator $\widehat{\beta}$ of β is available, we would expect that $\widehat{\widehat{\theta}} \equiv \widehat{\theta} - T^{-1}\widehat{\beta}$ would be less biased than the fixed effects estimator $\widehat{\theta}$. In order to implement this strategy, we need to have a convenient theoretical characterization of β.

In this paper, we examine asymptotic biases of general dynamic nonlinear panel models with fixed effects, and develop methods to remove them. The common intuition in Hahn and Kuersteiner (2002) and Hahn and Newey (2002) was that, under the alternative asymptotics where both n and T grow to infinity at the same rate, $\sqrt{nT}\left(\widehat{\theta} - \theta\right)$ will be centered at $\sqrt{\kappa}\beta$, where κ is the limit of $\frac{n}{T}$. It was shown that β can be consistently estimated, which led to a bias corrected estimator. We adopt the same strategy in this paper, and consider general nonlinear dynamic panel models with fixed effects. Our analysis could provide a useful alternative to Honoré and Kyriazidou (2000), who examined some dynamic binary response model. Their identification and estimation require conditioning on (possibly continuous) covariates taking identical values over time, which is not required in our approach. On the other hand, it is expected that our approach requires reasonably large T to be a good approximation.

In this paper, we also formalize the relationship between the "large n, large T" asymptotics and the higher order "large T" asymptotics. It is argued that the former is a simplified way of doing the latter, and the bias correction done under the latter would imply the bias correction under the former. This relationship provides a useful re-interpretation of the incidental parameters problem as a special case of higher order asymptotics.

We do not develop any "automatic" method of bias reduction, such as jackknife as in Hahn and Newey (2002). It is not clear how jackknife could be adapted to the situation where observations are dependent. Block bootstrap may provide an alternative solution, although we conjecture some modification is necessary.[1] Lancaster (2002), Woutersen (2002) and Arellano (2000) adopted the same asymptotics, and developed bias corrected estimator for various panel models with fixed effects. So far, all of their methods are confined to models with full parametric specification. We do not yet know the exact relationship between our approach and theirs, and are yet unable to generalize their methods to dynamic models.

2 Alternative Asymptotics

In this section, we characterize the approximate bias of the fixed effects estimator for a dynamic nonlinear panel model. We consider the asymptotic approximation where n and T grow to infinity at the same rate. We show that the fixed effects estimator is consistent and asymptotically normal, but has an asymptotic bias. We provide a formula for asymptotic bias under such asymptotics.

Suppose that we are given an estimator such that

$$\left(\widehat{\theta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_n\right) = \underset{\theta, \alpha_1, \ldots, \alpha_n}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(x_{it}; \theta, \alpha_i\right) \tag{1}$$

for some $\psi(\cdot)$. We assume that ψ is a sensible function that a time-series econometrician would use: If n is fixed, and $T \to \infty$, we assume that $\left(\widehat{\theta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_n\right)$ is consistent for $(\theta_0, \alpha_{10}, \ldots, \alpha_{n0})$, where θ_0 is a common parameter of interest, and α_{i0} is individual specific fixed effect. This can be justified by assuming a primitive condition such that, for each $\eta > 0$, $\inf_i \left[G_{(i)}\left(\theta_0, \alpha_{i0}\right) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} G_{(i)}\left(\theta, \alpha\right)\right] > 0$, where

$$\widehat{G}_{(i)}\left(\theta, \alpha_i\right) \equiv T^{-1} \sum_{t=1}^{T} \psi\left(x_{it}; \theta, \alpha_i\right), \quad G_{(i)}\left(\theta, \alpha_i\right) \equiv E\left[\psi\left(x_{it}; \theta, \alpha_i\right)\right]$$

It can be shown that this condition also implies that $\left(\widehat{\theta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_n\right)$ are consistent even when n grows to infinity. See Theorem 1 below.

We use the following notation in the paper:

$$\begin{aligned}
U_i\left(x_{it}; \theta, \alpha_i\right) &\equiv s_\theta\left(x_{it}; \theta, \alpha_i\right) - \rho_{i0} \cdot s_\alpha\left(x_{it}; \theta, \alpha_i\right), \\
V_i\left(x_{it}; \theta, \alpha_i\right) &\equiv s_\alpha\left(x_{it}; \theta, \alpha_i\right), \\
\rho_{i0} &\equiv \frac{E\left[s_{\theta\alpha}\left(x_{it}; \theta_0, \alpha_{i0}\right)\right]}{E\left[s_{\alpha\alpha}\left(x_{it}; \theta_0, \alpha_{i0}\right)\right]}, \\
s_\theta\left(x_{it}; \theta, \alpha_i\right) &\equiv \frac{\partial}{\partial\theta} \psi\left(x_{it}; \theta, \alpha_i\right), \\
s_\alpha\left(x_{it}; \theta, \alpha_i\right) &\equiv \frac{\partial}{\partial\alpha_i} \psi\left(x_{it}; \theta, \alpha_i\right), \\
\mathcal{I}_i &\equiv -E\left[\frac{\partial U_i\left(x_{it}; \theta, \alpha_i\right)}{\partial\theta'}\right].
\end{aligned}$$

For simplicity of notation, we will assume $\dim(\theta) = R$ and $\dim(\alpha) = 1$.

We now assume the following regularity conditions:

[1] Hahn, Kuersteiner, and Newey (2002) considered the bootstrap in the cross sectional context, and concluded that it would not work without some modification such as "truncation".

Condition 1 $n, T \to \infty$ *such that* $\frac{n}{T} \to \kappa$, *where* $0 < \kappa < \infty$.

Condition 2 *Suppose that, for each* i, $\{x_{it}, t = 1, 2, \ldots\}$ *is a stationary mixing sequence. Let* $\mathcal{A}_t^i = \sigma\left(x_{it}, x_{it-1}, x_{it-2}, ...\right)$, $\mathcal{B}_t^i = \sigma\left(x_{it}, x_{it+1}, x_{it+2}, ...\right)$ *and* $\alpha_i\left(m\right) = \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+m}^i} \left|P\left(A \cap B\right) - P\left(A\right) P\left(B\right)\right|$. *Assume that* $\sup_i \left|\alpha_i\left(m\right)\right| \leq C a^m$ *for some* a *such that* $0 < a < 1$ *and some* $C > 0$. *We assume that* $\{x_{it}, t = 1, 2, 3, ...\}$ *are independent across* i.

Condition 3 *Let* $\psi\left(x_{it}, \phi\right)$ *be a function indexed by the parameter* $\phi = \left(\theta, \alpha\right) \in \operatorname{int} \Phi$, *where* Φ *is a compact, convex subset of* $\mathbb{R}^p$ *and* $p \equiv \dim \phi = R + 1$. *Let* $\nu = \left(\nu_1, ..., \nu_k\right)$ *be a vector of non-negative integers* $v_i, |v| = \sum_{j=1}^k v_j$ *and* $D^v \psi\left(x_{it}, \phi\right) = \partial^{|\nu|} \psi\left(x_{it}, \phi\right) / \left(\partial \phi_1^{v_1} ... \partial \phi_k^{\nu_k}\right)$. *Assume that there exists a function* $M\left(x_{it}\right)$ *such that* $\left|D^v \psi\left(x_{it}, \phi_1\right) - D^v \psi\left(x_{it}, \phi_2\right)\right| \leq M\left(x_{it}\right) \left\|\phi_1 - \phi_2\right\|$ *for all* $\phi_1, \phi_2 \in \Phi$ *and* $|v| \leq 5$. *We also assume that* $M(x_{it})$ *satisfies* $\sup_{\phi \in \Phi} \left\|D^v \psi\left(x_{it}, \phi\right)\right\| \leq M(x_{it})$ *and* $\sup_i E\left[\left|M(x_{it})\right|^{10q+12+\delta}\right] < \infty$ *for some integer* $q \geq p/2 + 2$ *and for some* $\delta > 0$.

Condition 4 *Let* $\Sigma_{iT} = \operatorname{Var}\left(T^{-1/2} \sum_{t=1}^T U_i\left(x_{it}; \theta, \alpha_i\right)\right)$ *and assume that* $\inf_i \inf_T \lambda_{iT} > 0$ *where* λ_{iT} *is the smallest eigenvalue of* Σ_{iT}.

Condition 5 $\inf_i \left|E\left[\partial V_i\left(x_{it}; \theta_0, \alpha_{i0}\right) / \partial \alpha_i\right]\right| > 0$.

Condition 6 $\sup \left\|\mathcal{I}_i\right\| < \infty$ *such that* $\lim_{n \to \infty} n^{-1} \sum_{i=1}^n \mathcal{I}_i = \mathcal{I}$. *We also assume* $\left\|\mathcal{I}\right\| > 0$.

It can be shown that, under the alternative asymptotics, the parameters are consistent:

Theorem 1 *Assume that Conditions (1),(2) and (3) are satisfied. Then,* $\widehat{\theta}, \widehat{\alpha}_1, ..., \widehat{\alpha}_n$ *are consistent.*

Proof. See Appendix A.2. ■

It is shown in Appendix B that a Taylor series expansion leads to

$$\sqrt{nT}\left(\widehat{\theta} - \theta_0\right) = \Xi_1 + \frac{1}{2}\sqrt{\frac{n}{T}}\Xi_2 + o_p\left(1\right) \tag{2}$$

where

$$\begin{aligned}
\Xi_1 &\equiv \left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^n \sum_{t=1}^T U_{it}\right) \\
\Xi_2 &\equiv -2\left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1} \frac{1}{n}\sum_{i=1}^n \left[\frac{\sum_{t=1}^T V_{it}}{\sqrt{T} E\left[\frac{\partial V_i}{\partial \alpha_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T \left(U_i^{\alpha_i} - \frac{E\left[U_i^{\alpha_i \alpha_i}\right]}{2E\left[\frac{\partial V_i}{\partial \alpha_i}\right]} V_{it}\right)\right] \\
&\quad + \left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1} \frac{1}{n}\mathcal{T}
\end{aligned}$$

for $\mathcal{T}$ defined later in equation (30) in Appendix C, and in equation (31) as well. Here,

$$\begin{aligned}
U_{it} &\equiv U_i\left(x_{it}; \theta_0, \alpha_{i0}\right) \\
V_{it} &\equiv V_i\left(x_{it}; \theta_0, \alpha_{i0}\right) \\
U_i^{\alpha_i} &\equiv \frac{\partial}{\partial \alpha_i} U_i\left(x_{it}; \theta_0, \alpha_{i0}\right) \\
U_i^{\alpha_i \alpha_i} &\equiv \frac{\partial^2}{\partial \alpha_i^2} U_i\left(x_{it}; \theta_0, \alpha_{i0}\right).
\end{aligned}$$

It turns out that under our asymptotics where n and T tend to infinity jointly, the term Ξ_1 determines the asymptotic distribution of the estimator, while the term Ξ_2 turns out to be a pure bias term. These facts are established in the following two Lemmas.

Lemma 1 *Let Conditions (1),(2),(3) and (4) be satisfied. Then*

$$\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\to N\left(0,\Omega\right)$$

where $\Omega=\lim_n n^{-1}\sum_{i=1}^{n}\Sigma_{iT}$.

Proof. Follows from Lemma (6) in the Appendix. ■

Lemma 2 *Assume that Conditions (1),(2),(3), (4) and (6) hold. Let* $f_i^{VU^{\alpha}}=\sum_{l=-\infty}^{\infty}\operatorname{Cov}(V_{it},U_{it-l}^{\alpha_i})$ *and* $f_i^{VV}=\sum_{l=-\infty}^{\infty}\operatorname{Cov}(V_{it},V_{it-l})$. *Let*

$$f^{VU^{\alpha}}=\lim n^{-1}\sum_{i=1}^{n}\frac{f_i^{VU^{\alpha}}}{E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}$$

and

$$f^{VV}=\lim n^{-1}\sum_{i=1}^{n}\frac{E\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]}{2\left(E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]\right)^2}f_i^{VV}.$$

Define $\Psi\equiv f^{VU^{\alpha}}-f^{VV}$. *Then*

$$\begin{aligned}\Psi &= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}\left(x_{it};\theta,\alpha_i\right)}{E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_{it}^{\alpha_i}\left(x_{it};\theta,\alpha_i\right)-\frac{E\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]}{2E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}V_{it}\left(x_{it};\theta,\alpha_i\right)\right)\right],\\ \frac{1}{n}\mathcal{T} &= o_p\left(1\right),\end{aligned}$$

and

$$\frac{1}{2}\sqrt{\frac{n}{T}}\Xi_2=-\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\Psi+o_p\left(1\right).$$

Proof. See Appendix (C). ■

Using Lemmas (1) and (2) as well as (18) it follows that

$$\begin{aligned}\Xi_1 &\to N\left(0,\mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right),\\ \frac{1}{2}\sqrt{\frac{n}{T}}\Xi_2 &= -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\Psi+o_p\left(1\right).\end{aligned}$$

To summarize, we have the following result regarding the asymptotic distribution of the the estimator $\hat{\theta}$.

Theorem 2 *Assume that Conditions (1),(2),(3), (4) and (6) hold.*

$$\sqrt{nT}\left(\widehat{\theta}-\theta_0\right)\to N\left(\beta\sqrt{\kappa},\mathcal{I}^{-1}\Omega\left(\mathcal{I}'\right)^{-1}\right)$$

where

$$\beta\equiv-\mathcal{I}^{-1}\Psi.$$

3 Bias Correction

In this section, we consider various methods of estimating β in Theorem 2.

3.1 Case 1

In the fully parametric model, it may be possible to derive an analytic expression for β as a function $b(\theta_0, \alpha_{10}, \alpha_{20}, \ldots)$ of $(\theta_0, \alpha_{10}, \alpha_{20}, \ldots)$. A natural estimator would then be given by $b\left(\widehat{\theta}, \widehat{\alpha}_1, \widehat{\alpha}_2, \ldots\right)$, which would be consistent for β under some regularity conditions. For example, consider a linear dynamic panel model with fixed effects such that

$$y_{it} = \alpha_i + \theta y_{it-1} + \varepsilon_{it},$$

where $\varepsilon_{it} \sim N\left(0, \sigma^2\right)$ are independent over i and t. Note that we can take $x_{it} = (y_{it}, y_{it-1})$ in this example. Fixed effects OLS solves

$$\left(\widehat{\theta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_n\right) = \underset{\theta, \alpha_1, \ldots, \alpha_n}{\operatorname{argmax}} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(x_{it}; \theta, \alpha_i\right)$$

where $\psi\left(x_{it}; \theta, \alpha_i\right) = -\left(y_{it} - \alpha_i - \theta y_{it-1}\right)^2$. For this case, it can be shown that

$$\frac{1}{\sqrt{n}\sqrt{T}} \sum_{i=1}^{n} \sum_{t=1}^{T} U_i\left(x_{it}; \theta, \alpha_i\right) \rightarrow N\left(0, \frac{4\left(\sigma^2\right)^2}{1-\theta^2}\right),$$

and

$$\Psi = 2\frac{\sigma^2}{1-\theta}, \qquad \mathcal{I} = \frac{2\sigma^2}{1-\theta^2}.$$

(See Appendix E.) It follows that

$$\beta \equiv -\mathcal{I}^{-1}\Psi = -\left(1+\theta\right), \qquad \mathcal{I}^{-1}\Omega\mathcal{I}^{-1} = \frac{\frac{4\left(\sigma^2\right)^2}{1-\theta^2}}{\left(\frac{2\sigma^2}{1-\theta^2}\right)^2} = 1-\theta^2$$

and hence

$$\sqrt{nT}\left(\widehat{\theta} - \theta_0\right) \rightarrow N\left(-\sqrt{\kappa}\left(1+\theta\right), 1-\theta^2\right),$$

a result discussed by Hahn and Kuersteiner (2002). As a consequence, a natural estimator for β is given by $-\left(1+\widehat{\theta}\right)$.

3.2 Case 2

The estimation strategy discussed in the previous section is infeasible without a tightly specified model regarding x_{it}. In many cases, we may have to settle with the mixing condition as in Condition 2. We now develop a feasible estimator that is asymptotically unbiased under these general conditions. Note that natural estimators for $E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]$, $\widehat{E}\left[U_i^{\alpha_i\alpha_i}\right]$, and $\mathcal{I}_i$ are given by

$$\begin{aligned}
\widehat{E}\left[V_i^{\alpha_i}\right] &\equiv \frac{1}{T}\sum_{t=1}^{T} V_{it}^{\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) \\
\widehat{E}\left[V_i^{\theta}\right] &\equiv \frac{1}{T}\sum_{t=1}^{T} V_{it}^{\theta}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) \\
\widehat{E}\left[U_i^{\alpha_i\alpha_i}\right] &\equiv \frac{1}{T}\sum_{t=1}^{T} \widehat{U}_i^{\alpha_i\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) \\
\widehat{\mathcal{I}}_i &\equiv -\frac{1}{T}\sum_{t=1}^{T} \widehat{U}_i^{\theta}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right)
\end{aligned}$$

where

$$\begin{aligned}
\widehat{U}_i\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) &\equiv s_\theta\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) - \frac{\widehat{E}\left[V_i^{\theta}\right]}{\widehat{E}\left[V_i^{\alpha_i}\right]} V_{it}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) \\
\widehat{U}_i^{\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) &\equiv s_{\theta\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) - \frac{\widehat{E}\left[V_i^{\theta}\right]}{\widehat{E}\left[V_i^{\alpha_i}\right]} V_{it}^{\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) \\
\widehat{U}_i^{\alpha_i\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) &\equiv s_{\theta\alpha_i\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) - \frac{\widehat{E}\left[V_i^{\theta}\right]}{\widehat{E}\left[V_i^{\alpha_i}\right]} V_{it}^{\alpha_i\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) \\
\widehat{U}_i^{\theta}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) &\equiv s_{\theta\theta}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) - \frac{\widehat{E}\left[V_i^{\theta}\right]}{\widehat{E}\left[V_i^{\alpha_i}\right]} V_{it}^{\theta}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right)
\end{aligned}$$

In order to estimate the bias parameter β we form the spectral estimates

$$\widehat{f}_i^{VU^\alpha} = \sum_{l=-m}^{m} \widehat{\Gamma}_{il}^{VU^\alpha} \tag{3}$$

where

$$\widehat{\Gamma}_{il}^{VU^\alpha} = T^{-1}\sum_{t=\max(1,l)}^{\max(T,T+l)} V_i\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) U_i^{\alpha_i}\left(x_{it-l};\widehat{\theta},\widehat{\alpha}_i\right)'$$

and

$$\widehat{f}_i^{VV} = \sum_{l=-m}^{m} \widehat{\Gamma}_{il}^{VV} \tag{4}$$

where

$$\widehat{\Gamma}_{il}^{VV} = T^{-1}\sum_{t=\max(1,l)}^{\max(T,T+l)} V_i\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right) U_i^{\alpha_i}\left(x_{it-l};\widehat{\theta},\widehat{\alpha}_i\right)'.$$

The parameter m is a bandwidth parameter that needs to be chosen such that $m/T^{1/2} \to 0$ as $T \to \infty$. Note that $f_i^{VU^\alpha}$and f_i^{VV} are cross-spectra and spectra at zero frequency of the processes $U_i^{\alpha}\left(x_{it};\theta,\alpha\right)$ and $V_i\left(x_{it};\theta,\alpha_i\right)$. When such spectral quantities are estimated in the context of constructing confidence regions or test statistics, one needs to guarantee that the estimators are positive definite matrices. This is typically achieved by choosing appropriate kernel functions as pointed out by Newey and West (1987) and Andrews (1991). In the current context of bias correction, positivity of the estimates is of no concern and the main motivation for using any kernel other than a truncated kernel disappears.

We thus estimate β by

$$\widehat{\beta} \equiv -\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{f}_i^{VU^{\alpha}}}{\widehat{E}\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}-\frac{\widehat{E}\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]\widehat{f}_i^{VV}}{2\left(\widehat{E}\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]\right)^2}\right]$$

which is shown to be consistent in the next Theorem.

Theorem 3 *Assume that Conditions (1),(2),(3), (4) and (6) hold. Let* $m, T \to \infty$ *such that* $m/T^{1/2} \to 0$. *Then,*

$$\sqrt{nT}\left(\widehat{\widehat{\theta}}-\theta\right)\to N\left(0,\mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right).$$

where

$$\widehat{\widehat{\theta}} \equiv \widehat{\theta}-\frac{1}{T}\widehat{\beta} \tag{5}$$

4 Higher Order Time Series Asymptotics

We now consider the second order time series asymptotic approximation, where T is assumed to increase to infinity while n is being fixed. It is argued that the "large n, large T" asymptotics can be understood as a simplified way of doing the higher order "large T" asymptotics. This provides a new interpretation of the incidental parameter problem as a higher order bias problem.

It can be shown that an expansion similar to (2) is valid here as well[2]:

$$\sqrt{nT}\left(\widehat{\theta}-\theta_0\right)=\Xi_1+\frac{1}{2}\sqrt{\frac{n}{T}}\Xi_2+o_p\left(\frac{1}{\sqrt{T}}\right) \tag{6}$$

By an appropriate central limit theorem, it can also be shown that

$$\Xi_1 \to N\left(0,\mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right)$$

Writing

$$\widehat{\theta}=\theta_0+\frac{1}{\sqrt{T}}\frac{\Xi_1}{\sqrt{n}}+\frac{1}{T}\frac{\Xi_2}{2}+o_p\left(\frac{1}{T}\right)$$

we can see that

$$\overline{\Xi}_2 \equiv \lim_{T\to\infty} E\left[\Xi_2\right]$$

is the long run bias of order $\frac{1}{T}$ under "large T" asymptotics. We may therefore want to consider a bias reduced estimator

$$\widehat{\theta}-\frac{1}{T}\frac{\widehat{\Xi}_2}{2}$$

[2] This is because practically all the results in Appendices A and B go through even when n is fixed. Most notably, instead of the Central Limit theorem we use for the large n, large T case, a Central Limit Theorem for mixing sequences needs to be used here. The additional arguments needed are routine and therefore not reported.

where $\widehat{\Xi}_2$ is a consistent estimator for $\overline{\Xi}_2$ under "large T" asymptotics. In order to develop and analyze a consistent estimator for $\overline{\Xi}_2$, it is useful to recall that

$$\begin{aligned}\Xi_2 &\equiv -2\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\alpha_i}-\frac{E\left[U_i^{\alpha_i\alpha_i}\right]}{2E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}V_{it}\right)\right]\\ &\quad+\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\mathcal{T}\end{aligned}$$

Therefore, we have

$$\begin{aligned}\overline{\Xi}_2 &= -2\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\lim_{T\to\infty}E\left\{\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\alpha_i}-\frac{E\left[U_i^{\alpha_i\alpha_i}\right]}{2E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}V_{it}\right)\right]\right\}\\ &\quad+\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\lim_{T\to\infty}E\left[\mathcal{T}\right] \qquad (7)\end{aligned}$$

Note that $\frac{1}{n}\lim_{T\to\infty}E\left[\mathcal{T}\right]\neq 0$ in general under the "large T" asymptotics, although $\frac{1}{n}\operatorname{plim}_{T\to\infty}\mathcal{T}=0$ under the "large n, large T" asymptotics. Therefore, we can see that the asymptotic bias under the "large n, large T" asymptotics is a simplified version of the higher order bias under the "large T" asymptotics. This observation suggests that bias correction under "large T" asymptotics would imply bias correction under the "large n, large T" asymptotics.

In order to understand this implication, consider a bias corrected estimator under the "large T" asymptotics. It is reasonable to develop separate estimators for the two terms on the RHS of (7). Because the first term is the term that shows up in the "large n, large T" asymptotics, it is reasonable to use the bias correction device developed there. For this purpose, we can consider the probability limit of $\widehat{\beta}$ under "large T" asymptotics. It can be shown that

$$\widehat{\beta}=-\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{f}_i^{VU^\alpha}}{\widehat{E}\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}-\frac{\widehat{E}\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]\widehat{f}_i^{VV}}{2\widehat{E}\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}\right]$$

with $\widehat{f}_i^{VU^\alpha}$ defined in (3) and $\widehat{f}_i^{VV}$ defined in (4) consistently estimates $\frac{1}{2}$ times the first term on the RHS of (7) under the "large T" asymptotics as well.[3] We now consider the second term on the RHS of (7). It is also useful to note from (31) that $\mathcal{T}$ is the sum of several terms of the form

$$\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}k_{it}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-J^*}\prod_{j=1}^{J}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{E\left[m_{i,j}\right]}{E\left[m_{i,j}^*\right]}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}k_{it}^*\right) \qquad (8)$$

for some $k_{it}\equiv k\left(x_{it},\theta_0,\alpha_{i0}\right)$, $k_{it}\equiv k^*\left(x_{it},\theta_0,\alpha_{i0}\right)$, $m_{i,j}\equiv m_j\left(x_{it},\theta_0,\alpha_{i0}\right)$, $m_{i,j}^*\equiv m_j^*\left(x_{it},\theta_0,\alpha_{i0}\right)$ such that $E\left[k_{it}\right]=0$, $E\left[l_{it}\right]=0$, $\left|E\left[m_{i,j}^*\right]\right|>0$ and for J equal to 1, 2, or 3, and for J^* equal to 0 or 1. Writing (8) as

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=1}^{T}\sum_{t'=1}^{T}k_{it}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-J^*}\prod_{j=1}^{J}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{E\left[m_{i,j}\right]}{E\left[m_{i,j}^*\right]}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\right]k_{i't'}^*$$

[3] This holds because the arguments in Appendix D.3 reduce to the results in Andrews (1991) when $T\to\infty$ and n is fixed.

and noting that observations with $i \neq i'$ are independent of each other, we can see that the natural estimator of the limit (as $T \to \infty$) of its expectation would take the form

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{l=-m}^{m}\sum_{t=\max(1,l)}^{\max(T,T+l)}\widehat{k}_{it}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-J^*}\prod_{j=1}^{J}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{E}\left[m_{i,j}\right]}{\widehat{E}\left[m_{i,j}^*\right]}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1}\right]\widehat{k}_{it-l}^*, \tag{9}$$

with $m/T^{1/2} \to 0$, where

$$\begin{aligned}\widehat{k}_{it} &\equiv k\left(x_{it},\widehat{\theta},\widehat{\alpha}_i\right)\\ \widehat{k}_{it}^* &\equiv k^*\left(x_{it},\widehat{\theta},\widehat{\alpha}_i\right)\\ \widehat{E}\left[m_{i,j}\right] &\equiv T^{-1}\sum_{t=1}^{T}m_j\left(x_{it},\widehat{\theta},\widehat{\alpha}_i\right)\\ \widehat{E}\left[m_{i,j}^*\right] &\equiv T^{-1}\sum_{t=1}^{T}m_j^*\left(x_{it},\widehat{\theta},\widehat{\alpha}_i\right)\end{aligned}$$

Let $\widehat{\mathcal{T}}$ denote the sum of such terms. Using again standard arguments from the covariance estimation literature it follows as before that $\widehat{\mathcal{T}} = \mathcal{T} + o_p(1)$. We then have

$$\widehat{\Xi}_2 = 2\widehat{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1}\frac{1}{n}\widehat{\mathcal{T}}$$

and our bias reduced estimator will take the form

$$\widehat{\widehat{\widehat{\theta}}} \equiv \widehat{\theta} - \frac{1}{T}\widehat{\beta} - \frac{1}{nT}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1}\widehat{\mathcal{T}} \tag{10}$$

Note that the only difference between $\widehat{\widehat{\widehat{\theta}}}$ and $\widehat{\widehat{\theta}}$ is the last term $\frac{1}{nT}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i\right)^{-1}\widehat{\mathcal{T}}$. This is not surprising because this is related to the term that is ignored under "large n, large T" asymptotics.

5 Summary

In this paper, we provided a simple characterization of the asymptotic bias of a fixed effects estimator for dynamic nonlinear panel model with fixed effects. The asymptotic bias was based on the "large n, large T" asymptotics adopted by, e.g., Hahn and Kuersteiner (2002) and Hahn and Newey (2002). It was shown that this alternative asymptotic expansions are an alternative way of obtaining a higher order "large T" asymptotic approximation. A method of reducing bias based on these expansions was developed.

Appendix

A Consistency

A.1 General Lemmas

We provide a different version of Lahiri's (1992) Lemma 5.1 which is stated for bounded zero mean random variables.

Lemma 3 *Assume that* $\{W_t, t = 1, 2, \ldots\}$ *is a stationary, mixing sequence with* $E[W_t] = 0$ *and* $E\left[|W_t|^{2r+\delta}\right] < \infty$ *for any positive integer* r*, some* $\delta > 0$ *and all* t*. Let* $\mathcal{A}_t = \sigma(W_t, W_{t-1}, W_{t-2}, ...)$, $\mathcal{B}_t = \sigma(W_t, W_{t+1}, W_{t+2}, ...)$ *and* $\alpha(m) = \sup_t \sup_{A\in\mathcal{A}_t, B\in\mathcal{B}_{t+m}} |P(A\cap B) - P(A)P(B)|$. *Then, for any* m *such that* $1 \leq m < C(r)n$,

$$E\left[\left(\sum_{i=1}^{n} W_i\right)^{2r}\right] \leq C(r)\, E\left[|W_i|^{2r+\delta}\right]\left[n^r m^{2r} + n^{2r}\alpha(m)^{\frac{\delta}{2r+\delta}}\right]$$

where $C(r)$ *is a constant that depends on* r.

Proof. The proof is exactly identical to the proof of Lahiri (1992, p.198-200) except that, instead of using the mixing inequality in Lemma 27.2 of Billingsley (1986), we are using the mixing inequality in Corollary A.2 of Hall and Heyde (1980). For notation used here, we refer the reader to Lahiri (1992). All statements in Lahiri (1990) on p. 198 are unchanged and we pick up the proof starting on p. 199. We need to consider the term $\left|\sum_3 E\left[\prod_{t=1}^{j} W_{i_t}^{\alpha_t}\right]\right|$, where $(\alpha_1, .., \alpha_j)$ is a j-tuple such that $2r = \sum_{i=1}^{j}\alpha_i$, $j > r$. Let $A = \{t : \alpha_t = 1\}$ and let β_0 be the number of elements in A. Lahiri (1992) shows that $2(j-r) \leq \beta_0 \leq 2r$ which shows that A is non-empty when $j > r$. The sum $\sum_3$ is as defined in Lahiri and extends over all indices in the set $B_m = \{(i_1, ..., i_j) : 1 \leq i_1 \leq ... \leq i_j \leq n, |i_{t-1} - i_t| > m, |i_t - i_{t+1}| > m$ for some $t \in A\}$. Now consider $E\left[\prod_{t=1}^{j} W_{i_t}^{\alpha_t}\right]$ when $j > r$. Fix $\tau \in A$. Then, if $1 < \tau < j$ define $W_a = \prod_{t=1}^{\tau-1} W_{i_t}^{\alpha_t}$, $W_b = W_{i_\tau}$ and $W_c = \prod_{t=\tau+1}^{j} W_{i_t}^{\alpha_t}$ as well as $b_1 = \sum_{t=1}^{\tau-1}\alpha_t$ and $b_2 = \sum_{t=\tau}^{j}\alpha_t$ such that

$$\begin{aligned}
\left|E\left[\prod_{t=1}^{j} W_{i_t}^{\alpha_t}\right]\right| &\leq |E[W_aW_bW_c] - E[W_aW_b]E[W_c]| + |E[W_aW_b]|\,|E[W_c]| \\
&\leq 8\left(E\left[|W_aW_b|^{\frac{2r+\delta}{b_1+1}}\right]\right)^{\frac{b_1+1}{2r+\delta}}\left(E\left[|W_c|^{\frac{2r+\delta}{b_2}}\right]\right)^{\frac{b_2+1}{2r+\delta}}\alpha(m)^{\frac{\delta}{2r+\delta}} \\
&\quad +8\,|E[W_c]|\left(E\left[|W_a|^{\frac{2r+\delta}{b_1}}\right]E\left[|W_b|^{2r+\delta}\right]\right)^{\frac{b_1+1}{2r+\delta}}\alpha(m)^{\frac{\delta+b_2}{4r+\delta}} \\
&\leq 8E\left[|W_i|^{2r+\delta}\right]\alpha(m)^{\frac{\delta}{2r+\delta}} + 8E\left[|W_i|^{2r+\delta}\right]\alpha(m)^{\frac{\delta+b_2}{4r+\delta}}\ , \qquad (11)
\end{aligned}$$

where the second line follows from Corollary A.2 of Hall and Heyde (1980), and the last line is based on a repeated application of Hölder's inequality and makes use of stationarity. If $\tau = 1$, then define $W_a = W_{i_1}$ and $W_b = \prod_{t=\tau+1}^{j} W_{i_t}^{\alpha_t}$ such that

$$\begin{aligned}
\left|E\left[\prod_{t=1}^{j} W_{i_t}^{\alpha_t}\right]\right| &= |E[W_aW_b]| \\
&\leq 8\left(E\left[|W_a|^{2r+\delta}\right]\right)^{\frac{1}{2r+\delta}}\left(E\left[|W_b|^{\frac{2r+\delta}{2r-1}}\right]\right)^{\frac{2r-1}{2r+\delta}}\alpha(m)^{\frac{\delta}{2r+\delta}} \\
&\leq 8E\left[|W_i|^{2r+\delta}\right]\alpha(m)^{\frac{\delta}{2r+\delta}}
\end{aligned}$$

by the mixing inequality from Corollary A.2 of Hall and Heyde (1980). A similar argument holds for the case when $\tau = j$. Since $\alpha(m)^{\frac{\delta+b_2}{4r+\delta}} = o\left(\alpha(m)^{\frac{\delta}{4r+\delta}}\right)$ the second term in (11) can be subsumed into the constant $C(r)$. The remaining part of Lahiris proof is not affected by the changes made here because it does not involve mixing arguments. ■

Lemma 4 *Suppose that, for each i, $\{\xi_{it}, t = 1, 2, \ldots\}$ is a mixing sequence with $E[\xi_{it}] = 0$ for all i, t. Let $\mathcal{A}_t^i = \sigma(\xi_{it}, \xi_{it-1}, \xi_{it-2}, ...)$, $\mathcal{B}_t^i = \sigma(\xi_{it}, \xi_{it+1}, \xi_{it+2}, ...)$ and $\alpha_i(m) = \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+k}^i} |P(A \cap B) - P(A) P(B)|$. Assume that $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $0 < C < \infty$. We assume that $\{\xi_{it}, t = 1, 2, 3, ...\}$ are independent across i. We also assume that $n = O(T)$. Finally, assume that $E\left[|\xi_{it}|^{6+\delta}\right] < \infty$ for some $\delta > 0$. We then have*

$$\Pr\left[\max_{1\leq i\leq n}\left|\frac{1}{T}\sum_{t=1}^T \xi_{it}\right| > \eta\right] = o\left(T^{-1}\right)$$

for every $\eta > 0$. Now assume that $E\left[|\xi_{it}|^{10q+12+\delta}\right] < \infty$ for some $\delta > 0$ and some integer $q \geq 1$. Then,

$$\Pr\left[\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi_{it}\right| > \eta T^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-q}\right)$$

for every $\eta > 0$ and $0 < \upsilon < (100q + 120)^{-1}$.

Proof. By the Markov inequality

$$\Pr\left[\max_{1\leq i\leq n}\left|\frac{1}{T}\sum_{t=1}^T \xi_{it}\right| > \eta\right] = \Pr\left[\max_{1\leq i\leq n}\left|\sum_{t=1}^T \xi_{it}\right|^6 > \eta^6 T^6\right] \leq T^{-6}\eta^{-6}E\left[\max_{1\leq i\leq n}\left|\sum_{t=1}^T \xi_{it}\right|^6\right]$$

and by a an inequality for the Orlicz norm of a maximum of random variables (van der Vaart and Wellner, 1996, p.96) one obtains

$$E\left[\max_{1\leq i\leq n}\left|\sum_{t=1}^T \xi_{it}\right|^6\right] \leq n \max_i E\left[\left|\sum_{t=1}^T \xi_{it}\right|^6\right].$$

>From Lemma (3) it follows that

$$E\left[\left|\sum_{t=1}^T \xi_{it}\right|^6\right] \leq CE\left[|\xi_{it}|^{6+\delta}\right]\left(T^3 m^6 + T^6 \alpha_i(m)^{\frac{\delta}{6+\delta}}\right)$$

for any m such that $1 \leq m \leq CT$. Choose $m = T^\gamma$ and some γ such that $0 < \gamma \leq 1$. Then, for $\gamma < \frac{1}{4}$,

$$\begin{aligned}\Pr\left[\max_{1\leq i\leq n}\left|\frac{1}{T}\sum_{t=1}^T \xi_{it}\right| > \eta\right] &\leq nT^{-6}\eta^{-6}C\left(T^{3+6\gamma} + T^6 a^{\frac{\delta}{6+\delta}T^\gamma}\right) \\ &= O(T^{-2+6\gamma} + Ta^{T^\gamma}) = o\left(T^{-1}\right).\end{aligned}$$

For the second part of the Lemma, note that by previous arguments

$$\begin{aligned}
& T^q \Pr\left[\max_{1\le i\le n}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi_{it}\right| > \eta T^{\frac{1}{10}-\upsilon}\right] \\
= \; & T^q \Pr\left[\max_{1\le i\le n}\left|\sum_{t=1}^T \xi_{it}\right| > \eta T^{\frac{3}{5}-\upsilon}\right] \\
= \; & T^q \Pr\left[\max_{1\le i\le n}\left|\sum_{t=1}^T \xi_{it}\right|^{10q+12} > \eta^{(10q+12)} T^{\left(\frac{3}{5}-\upsilon\right)(10q+12)}\right] \\
\le \; & T^q T^{-\left(\frac{3}{5}-\upsilon\right)(10q+12)} \eta^{-(10p+12)} E\left[\max_{1\le i\le n}\left|\sum_{t=1}^T \xi_{it}\right|^{10q+12}\right] \\
= \; & O\left[T^{-5q-\frac{36}{5}+10\upsilon q+12\upsilon} n\cdot C\left(T^{5q+6+\gamma(10q+12)} + T^{(10q+12)} a^{\frac{\delta}{10q+12+\delta}T^\gamma}\right)\right] \\
= \; & O\left(T^{-\frac{1}{5}+10\upsilon q+12\upsilon+10\gamma q+12\gamma}\right) = o(1)
\end{aligned}$$

for $\gamma > 0$ sufficiently small. ■

Lemma 5 *Let $\xi(x_{it},\phi)$ a function indexed by the parameter $\phi \in \Phi$ where Φ is a convex subset of $\mathbb{R}^p$ with $E[\xi(x_{it},\phi)] = 0$ for all i,t and $\phi \in \Phi$. Assume that there exists a function $M(x_{it})$ such that $|\xi(x_{it},\phi_1) - \xi(x_{it},\phi_2)| \le M(x_{it})\|\phi_1 - \phi_2\|$ for all $\phi_1,\phi_2 \in \Phi$ and $\sup_\phi |\xi(x_{it},\phi)| \le M(x_{it})$. For each i, let x_{it} be a α-mixing process with exponentially decaying mixing coefficients satisfying $\sup_i |\alpha_i(m)| \le Ca^m$ for some a such that $0 < a < 1$ and some $0 < C < \infty$. Let q denote a positive integer such that $q \ge \frac{p+4}{2}$, where $p = \dim \phi$. We also assume that $E\left[|M(x_{it})|^{10q+12+\delta}\right] < \infty$ for some $\delta > 0$. Finally, assume that $n = O(T)$. We then have*

$$\Pr\left[\max_i \left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right| > T^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right)$$

for $0 < \upsilon < (100q+120)^{-1}$. Here, $\{\phi_i\}$ is an arbitrary nonstochastic sequence in Φ.

Proof. Note that we have

$$T\Pr\left[\max_i \left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right| > T^{\frac{1}{10}-\upsilon}\right] \le T\sum_{i=1}^n \Pr\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi_i)\right| > T^{\frac{1}{10}-\upsilon}\right]$$

Adapting an argument in Hall and Horowitz (1996) we chose $\varepsilon > 0$ and divide Φ into subsets $\Phi_1, \Phi_2, \ldots \Phi_{M(\varepsilon)}$ such that $\|\phi_1 - \phi_2\| < \frac{\varepsilon}{\sqrt{T}}$ whenever ϕ_1, ϕ_2 are in the same subset Φ_i. Then

$$\Pr\left[\sup_{\phi\in\Phi}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right| > T^{\frac{1}{10}-\upsilon}\right] \le \sum_{j=1}^{M(\varepsilon)} \Pr\left[\sup_{\phi\in\Phi_j}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right| > T^{\frac{1}{10}-\upsilon}\right]$$

Then, for some $\phi_j \in \Phi_j$ and any $\phi \in \Phi_j$

$$\begin{aligned}
\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi(x_{it},\phi)\right| &\le \left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi\left(x_{it},\phi_j\right)\right| + \frac{1}{\sqrt{T}}\sum_{t=1}^T \left|\xi(x_{it},\phi) - \xi\left(x_{it},\phi_j\right)\right| \\
&\le \left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \xi\left(x_{it},\phi_j\right)\right| + \left|\frac{\varepsilon}{T}\sum_{t=1}^T (M(x_{it}) - E[M(x_{it})])\right| + 2\varepsilon E[M(x_{it})]
\end{aligned}$$

such that

$$\Pr\left[\sup_{\phi\in\Phi_j}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi\right)\right|>T^{\frac{1}{10}-\upsilon}\right] \leq \Pr\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_j\right)\right|>\frac{T^{\frac{1}{10}-\upsilon}}{3}\right] + \Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}\left(M\left(x_{it}\right)-E\left[M\left(x_{it}\right)\right]\right)\right|>\frac{T^{\frac{1}{10}-\upsilon}}{3}\right].$$

By Lemma (4), it follows that both terms on the right are of order $o\left(T^{-q}\right)$, where the orders are uniform in i. Since $M\left(\varepsilon\right)=O\left(T^{p/2}\right)$ it follows that

$$T\Pr\left[\max_{i}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)\right|>T^{\frac{1}{10}-\upsilon}\right]=nT\cdot o\left(T^{-q}\right)\cdot O\left(T^{p/2}\right)=o\left(T^{2-q+p/2}\right)=o\left(1\right).$$

Lemma 6 *Assume that x_{it} satisfies Condition (2) and let $\xi\left(x_{it},\phi\right)$ be a function indexed by the parameter $\phi\in\operatorname{int}\Phi$ where Φ is a convex subset of $\mathbb{R}^p$. For any sequence $\phi_i\in\operatorname{int}\Phi$ assume $E\left[\xi\left(x_{it},\phi_i\right)\right]=0$. Moreover $\sup_\phi\left\|\xi\left(x_{it},\phi\right)\right\|\leq M(x_{it})$ such that $E\left[M(x_{it})^4\right]<\infty$. Then*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)\xrightarrow{d}N\left(0,f^{\xi\xi}\right)$$

where $f^{\xi\xi}=\lim n^{-1}\sum_{i=1}^{n}f_i^{\xi\xi}$ with $f_i^{\xi\xi}=\sum_{j=-\infty}^{\infty}E\left[\xi\left(x_{it},\phi_i\right)\xi\left(x_{it-j},\phi_i\right)'\right]$.

■

Proof. First note that by Condition (2) and the fact that the mixing property is preserved by measurable transformations of finitely many elements x_{it}, $\xi\left(x_{it},\phi_i\right)$ is mixing with exponentially decaying mixing coefficients. Let $p'\equiv\dim\left(\xi\right)$. By the Cramer-Wold theorem it is enough to consider $v_{i,n,T}=\ell'\Sigma_{nT}^{-1/2}Z_{iT}=\ell'\Sigma_{nT}^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)$ for all $\ell\in\mathbb{R}^{p'}$, $\|\ell\|=1$ where $\Sigma_{nT}=\sum_{i=1}^{n}\Sigma_{iT}^{\xi\xi}$ with $\Sigma_{iT}^{\xi\xi}=\operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)\right)$. The Lindeberg-Feller condition requires that for any $\varepsilon>0$

$$\sum_{i=1}^{n}E\left[v_{i,n,T}^2\mathbf{1}\left\{\left|v_{i,n,T}\right|>\varepsilon\right\}\right]\rightarrow 0.$$

Let $\xi_{it}\equiv\xi\left(x_{it},\phi_i\right)$, and note that

$$\begin{aligned} E\left[v_{i,n,T}^2\mathbf{1}\left\{\left|v_{i,n,T}\right|>\varepsilon\right\}\right] &= E\left[\left\|v_{i,n,T}\right\|^2\mathbf{1}\left\{\left|v_{i,n,T}\right|>\varepsilon\right\}\right]\leq\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^2E\left[\left\|Z_{iT}\right\|^2\mathbf{1}\left\{\left|v_{i,n,T}\right|>\varepsilon\right\}\right]\\ &\leq \varepsilon^{-2}\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^4E\left[\left\|Z_{iT}\right\|^4\right]=\varepsilon^{-2}\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^4T^{-2}\sum_{t_1,..,t_4}E\left[\xi_{it_1}'\xi_{it_2}\xi_{it_3}'\xi_{it_4}\right],\end{aligned}$$

where $\mathbf{1}\left\{\left|v_{i,n,T}\right|>\varepsilon\right\}\leq\mathbf{1}\left\{\left\|Z_{iT}\right\|^2>\varepsilon^2/\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^2\right\}$. Let the j-th element of ξ_{it} be denoted by $\xi_{j,it}$ such that

$$\begin{aligned} E\left[\xi_{it_1}'\xi_{it_2}\xi_{it_3}'\xi_{it_4}\right] &= \sum_{j_1,j_2}^{p}E\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\xi_{j_2,it_3}\xi_{j_2,it_4}\right]\\ &= \sum_{j_1,j_2}^{p}\left\{E\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right]E\left[\xi_{j_2,it_3}\xi_{j_2,it_4}\right]+E\left[\xi_{j_2,it_3}\xi_{j_1,it_2}\right]E\left[\xi_{j_1,it_1}\xi_{j_2,it_4}\right]\right.\\ &\quad\left.+E\left[\xi_{j_2,it_3}\xi_{j_1,it_1}\right]E\left[\xi_{j_1,it_2}\xi_{j_2,it_4}\right]+\operatorname{Cum}\left(\xi_{j_1,it_1},\xi_{j_1,it_2},\xi_{j_2,it_3},\xi_{j_2,it_4}\right)\right\}.\end{aligned}$$

>From Andrews (1991, Lemma 1) it follows that $\sum_{t_2,..,t_4} \sup_{t_1} \operatorname{Cum}\left(\xi_{j_1,it_1}, \xi_{j_1,it_2}, \xi_{j_2,it_3}, \xi_{j_2,it_4}\right) < \infty$.
>From Hall and Heyde (1980, Corollary A.2) it follows that $E(\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right] \leq 8E\left[\left|\xi_{j_1,it_1}\right|^4\right] E\left[\left|\xi_{j_1,it_2}\right|^4\right] \alpha\left(t_1-t_2\right)^{1/2}$ such that $T^{-1}\sum_{t_1,t_2} E(\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right] \leq c_1 \sum_{l=0}^{T}(1-l/T)(\sqrt{a})^l < \infty$ for all T and some constant $c_1 < \infty$. >>From $\ell' \Sigma_{nT}^{-1}\ell \leq \max_k \lambda_k^{-1} = 1/\min_k \lambda_k$ with λ_k eigenvalues of Σ_{nT} and the fact that $\min_k \lambda_k \geq n \inf_i \inf_T \lambda_{iT}$ implies that $\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^4 \leq c_2 n^{-2}$ for some constant $c_2 < \infty$. These arguments show that

$$E\left[v_{i,n,T}^2 \mathbf{1}\left\{\left|v_{i,n,T}\right| > \varepsilon\right\}\right] \leq cn^{-2}$$

uniformly in i, T for some constant $c < \infty$ such that the Lindeberg-Feller condition is satisfied. The result then follows from the fact that

$$\begin{aligned}\sup_i \left\|\Sigma_{iT}^{\xi\xi} - f_i^{\xi\xi}\right\| &\leq \frac{|l|}{T}\sum_{l=-T}^{T}\left\|\operatorname{Cov}\left(\xi_{it}, \xi_{it-l}'\right)\right\| + \sum_{|l|\geq T+1}\left\|\operatorname{Cov}\left(\xi_{it}, \xi_{it-l}'\right)\right\| \\ &\leq \frac{1}{T}\sum_{l=-\infty}^{\infty}|l|\left\|\operatorname{Cov}\left(\xi_{it}, \xi_{it-l}'\right)\right\| + c_1 \sum_{|l|\geq T+1}^{\infty}\left(1-|l|/T\right)\left(\sqrt{a}\right)^l \to 0 \text{ as } T\to\infty \quad (12)\end{aligned}$$

such that iterated and joint limits are the same such that $\Sigma_{nT} \to f^{\xi\xi}$. ■

A.2 Consistency

Definition 1

$$\widehat{G}_{(i)}(\theta,\alpha) \equiv \frac{1}{T}\sum_{t=1}^{T}\psi\left(x_{it};\theta,\alpha\right), \qquad G_{(i)}(\theta,\alpha) \equiv E\left[\psi\left(x_{it};\theta,\alpha\right)\right]$$

Lemma 7 *For all $\eta > 0$ that*

$$\Pr\left[\max_{1\leq i\leq n}\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| \geq \eta\right] = o\left(T^{-1}\right)$$

Proof. Let $\eta > 0$ be given. We note that

$$\Pr\left[\max_{1\leq i\leq n}\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| \geq \eta\right] \leq \sum_{i=1}^{n}\Pr\left[\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| \geq \eta\right]. \quad (13)$$

Let $\varepsilon > 0$ be chosen such that $2\varepsilon \max_i E\left[M\left(x_{it}\right)\right] < \frac{\eta}{3}$. Divide Υ into subsets $\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_{M(\varepsilon)}$ such that $\left|(\theta,\alpha) - \left(\theta',\alpha'\right)\right| < \varepsilon$ whenever (θ,α) and $\left(\theta',\alpha'\right)$ are in the same subset. Let (θ_j, α_j) denote *some* point in Υ_j for each j. Then,

$$\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| = \max_j \sup_{\Upsilon_j}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right|,$$

and therefore

$$\Pr\left[\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| > \eta\right] \leq \sum_{j=1}^{M(\varepsilon)}\Pr\left[\sup_{\Upsilon_j}\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| > \eta\right] \quad (14)$$

For $(\theta, \alpha) \in \Upsilon_j$, we have

$$\left|\widehat{G}_{(i)}(\theta, \alpha) - G_{(i)}(\theta, \alpha)\right| \leq \left|\widehat{G}_{(i)}(\theta_j, \alpha_j) - G_{(i)}(\theta_j, \alpha_j)\right| + \frac{\varepsilon}{T}\left|\sum_{t=1}^{T}(M(x_{it}) - E[M(x_{it})])\right| + 2\varepsilon E[M(x_{it})]$$

Then,

$$\begin{aligned}
\Pr\left[\sup_{\Upsilon_j}\left|\widehat{G}_{(i)}(\theta, \alpha) - G_{(i)}(\theta, \alpha)\right| > \eta\right] &\leq \Pr\left[\left|\widehat{G}_{(i)}(\theta_j, \alpha_j) - G_{(i)}(\theta_j, \alpha_j)\right| > \frac{\eta}{3}\right] \\
&\quad + \Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}(M(x_{it}) - E[M(x_{it})])\right| > \frac{\eta}{3\varepsilon}\right] \\
&= o\left(T^{-2}\right) \qquad (15)
\end{aligned}$$

by Lemma 4. Combining (13), (14), (15), and $n = O(T)$, we obtain the desired conclusion. ■

Theorem 4 $\Pr\left[\left|\widehat{\theta} - \theta_0\right| \geq \eta\right] = o\left(T^{-1}\right)$ *for every* $\eta > 0$*.*

Proof. Let η be given, and let $\varepsilon \equiv \inf_i \left[G_{(i)}(\theta_0, \alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} G_{(i)}(\theta, \alpha)\right] > 0$. With probability equal to $1 - o\left(\frac{1}{T}\right)$, we have

$$\begin{aligned}
\max_{|\theta-\theta_0|>\eta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta, \alpha_i) &\leq \max_{|(\theta,\alpha_i)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta, \alpha_i) \\
&< \max_{|(\theta,\alpha_i)-(\theta_0,\alpha_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}G_{(i)}(\theta, \alpha_i) + \frac{1}{3}\varepsilon \\
&< n^{-1}\sum_{i=1}^{n}G_{(i)}(\theta_0, \alpha_{i0}) - \frac{2}{3}\varepsilon \\
&< n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta_0, \alpha_{i0}) - \frac{1}{3}\varepsilon,
\end{aligned}$$

where the second and fourth inequalities are based on Lemma 7. Because

$$\max_{\theta,\alpha_1,\ldots,\alpha_n} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta, \alpha_i) \geq n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta_0, \alpha_{i0})$$

by definition, we can conclude that $\Pr\left[\left|\widehat{\theta} - \theta_0\right| \geq \eta\right] = o\left(T^{-1}\right)$. ■

Theorem 5 $\Pr\left[\max_{1\leq i\leq n}|\widehat{\alpha}_i - \alpha_{i0}| \geq \eta\right] = o\left(T^{-1}\right)$

Proof. We first prove that

$$T\Pr\left[\max_{1\leq i\leq n}\sup_{\alpha}\left|\widehat{G}_{(i)}\left(\widehat{\theta}, \alpha\right) - G_{(i)}(\theta_0, \alpha)\right| \geq \eta\right] = o(1) \qquad (16)$$

for every $\eta > 0$. Note that

$$\begin{aligned}
&\max_{1\leq i\leq n}\sup_{\alpha}\left|\widehat{G}_{(i)}\left(\widehat{\theta}, \alpha\right) - G_{(i)}(\theta_0, \alpha)\right| \\
\leq\ &\max_{1\leq i\leq n}\sup_{\alpha}\left|\widehat{G}_{(i)}\left(\widehat{\theta}, \alpha\right) - G_{(i)}\left(\widehat{\theta}, \alpha\right)\right| + \max_{1\leq i\leq n}\sup_{\alpha}\left|G_{(i)}\left(\widehat{\theta}, \alpha\right) - G_{(i)}(\theta_0, \alpha)\right| \\
\leq\ &\max_{1\leq i\leq n}\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}(\theta, \alpha) - G_{(i)}(\theta, \alpha)\right| + \max_{1\leq i\leq n}E[M(x_{it})]\cdot\left|\widehat{\theta} - \theta_0\right|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
T\Pr\left[\max_{1\le i\le n}\sup_{\alpha}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\alpha\right)-G_{(i)}\left(\theta_0,\alpha\right)\right|\ge\eta\right] &\le T\Pr\left[\max_{1\le i\le n}\sup_{(\theta,\alpha)}\left|\widehat{G}_{(i)}\left(\theta,\alpha\right)-G_{(i)}\left(\theta,\alpha\right)\right|\ge\frac{\eta}{2}\right]\\
&\quad +T\Pr\left[\left|\widehat{\theta}-\theta_0\right|\ge\frac{\eta}{2\left(1+\max_{1\le i\le n}E\left[M\left(x_{it}\right)\right]\right)}\right]\\
&= o\left(1\right)
\end{aligned}$$

by Lemma 7 and Theorem 4.

We now get back to the proof of Theorem 5. It suffices to prove that

$$T\Pr\left[\max_{1\le i\le n}\left|\widehat{\alpha}_i-\alpha_{i0}\right|\ge\eta\right]=o\left(1\right)$$

for every $\eta>0$. Let η be given, and let $\varepsilon\equiv\inf_i\left[G_{(i)}\left(\theta_0,\alpha_{i0}\right)-\sup_{\{\alpha_i:|\alpha_i-\alpha_{i0}|>\eta\}}G_{(i)}\left(\theta_0,\alpha_i\right)\right]>0$. Condition on the event $\left\{\max_{1\le i\le n}\sup_{\alpha}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\alpha\right)-G_{(i)}\left(\theta_0,\alpha\right)\right|\le\frac{1}{3}\varepsilon\right\}$, which has a probability equal to $1-o\left(\frac{1}{T}\right)$ by (16). We then have

$$\max_{|\alpha_i-\alpha_{i0}|>\eta}\widehat{G}_{(i)}\left(\widehat{\theta},\alpha_i\right)<\max_{|\alpha_i-\alpha_{i0}|>\eta}G_{(i)}\left(\theta_0,\alpha_i\right)+\frac{1}{3}\varepsilon<G_{(i)}\left(\theta_0,\alpha_{i0}\right)-\frac{2}{3}\varepsilon<\widehat{G}_{(i)}\left(\widehat{\theta},\alpha_{i0}\right)-\frac{1}{3}\varepsilon$$

This is inconsistent with $\widehat{G}_{(i)}\left(\widehat{\theta},\widehat{\alpha}_i\right)\ge\widehat{G}_{(i)}\left(\widehat{\theta},\alpha_{i0}\right)$, and therefore, $\left|\widehat{\alpha}_i-\alpha_{i0}\right|\le\eta$ for every i. ∎

B Expansion

Let

$$\widehat{\alpha}_i\left(\theta\right)\equiv\underset{a}{\operatorname{argmax}}\sum_{t=1}^{T}\psi\left(x_{it};\theta,a\right)$$

Notice that $\widehat{\theta}$ can be given an alternative characterization:

$$\widehat{\theta}\equiv\underset{c}{\operatorname{argmax}}\sum_{i=1}^{n}\sum_{t=1}^{T}\psi\left(x_{it};c,\widehat{\alpha}_i\left(c\right)\right)$$

Therefore, we can see that $\widehat{\theta}$ solves

$$0=\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\left(\widehat{\theta}\right)\right),$$

Let $F\equiv\left(F_1,\ldots,F_n\right)$ denote the collection of (marginal) distribution functions of x_{it}. Let $\widehat{F}\equiv\left(\widehat{F}_1,\ldots,\widehat{F}_n\right)$, where $\widehat{F}_i$ denotes the empirical distribution function for the stratum i. Define $F\left(\epsilon\right)\equiv F+\epsilon\sqrt{T}\left(\widehat{F}-F\right)$ for $\epsilon\in\left[0,T^{-1/2}\right]$. For each fixed θ and ϵ, let $\alpha_i\left(\theta,F_i\left(\epsilon\right)\right)$ be the solution to the estimating equation

$$0=\int V_i\left[\theta,\alpha_i\left(\theta,F_i\left(\epsilon\right)\right)\right]dF_i\left(\epsilon\right),$$

and let $\theta(F(\epsilon))$ be the solution to the estimating equation

$$0 = \sum_{i=1}^{n} \int U_i\left(x_{it}; \theta\left(F\left(\epsilon\right)\right), \alpha_i\left(\theta\left(F_i\left(\epsilon\right)\right), F_i\left(\epsilon\right)\right)\right) dF_i\left(\epsilon\right),$$

By a Taylor series expansion, we have

$$\theta\left(\widehat{F}\right) - \theta\left(F\right) = \frac{1}{\sqrt{T}}\theta^{\epsilon}\left(0\right) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^{2}\theta^{\epsilon\epsilon}\left(0\right) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^{3}\theta^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right), \tag{17}$$

where $\theta^{\epsilon}(\epsilon) \equiv d\theta(F(\epsilon))/d\epsilon$, $\theta^{\epsilon\epsilon}(\epsilon) \equiv d^2\theta(F(\epsilon))/d\epsilon^2$, ..., and $\widetilde{\epsilon}$ is somewhere in between 0 and $T^{-1/2}$. We therefore have

$$\sqrt{nT}\left(\theta\left(\widehat{F}\right) - \theta\left(F\right)\right) = \sqrt{nT}\frac{1}{\sqrt{T}}\theta^{\epsilon}\left(0\right) + \sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^{2}\theta^{\epsilon\epsilon}\left(0\right) + \frac{1}{6}\sqrt{\frac{n}{T}}\frac{1}{\sqrt{T}}\theta^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right). \tag{18}$$

Let

$$h_i\left(\cdot, \epsilon\right) \equiv U_i\left(\cdot; \theta\left(F\left(\epsilon\right)\right), \alpha_i\left(\theta\left(F\left(\epsilon\right)\right), F_i\left(\epsilon\right)\right)\right) \tag{19}$$

The first order condition may be written as

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int h_i\left(\cdot, \epsilon\right) dF_i\left(\epsilon\right) \tag{20}$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i\left(\cdot, \epsilon\right)}{d\epsilon} dF_i\left(\epsilon\right) + \frac{1}{n}\sum_{i=1}^{n}\int h_i\left(\cdot, \epsilon\right) d\Delta_{iT} \tag{21}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2 h_i\left(\cdot, \epsilon\right)}{d\epsilon^2} dF_i\left(\epsilon\right) + 2\frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i\left(\cdot, \epsilon\right)}{d\epsilon} d\Delta_{iT} \tag{22}$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3 h_i\left(\cdot, \epsilon\right)}{d\epsilon^3} dF_i\left(\epsilon\right) + 3\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2 h_i\left(\cdot, \epsilon\right)}{d\epsilon^2} d\Delta_{iT} \tag{23}$$

where $\Delta_{iT} \equiv \sqrt{T}\left(\widehat{F}_i - F_i\right)$.

B.1 $\theta^{\epsilon}(0)$

Because

$$\frac{dh_i\left(\cdot, \epsilon\right)}{d\epsilon} = \frac{\partial h_i\left(\cdot, \epsilon\right)}{\partial \theta'}\frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i\left(\cdot, \epsilon\right)}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial \theta'}\frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i\left(\cdot, \epsilon\right)}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial \epsilon}$$

we may rewrite (21) as

$$\begin{aligned} 0 = {} & \frac{1}{n}\sum_{i=1}^{n}\int \left(\frac{\partial h_i\left(\cdot, \epsilon\right)}{\partial \theta'}\frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i\left(\cdot, \epsilon\right)}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial \theta'}\frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i\left(\cdot, \epsilon\right)}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial \epsilon}\right) dF_i\left(\epsilon\right) \\ & + \frac{1}{n}\sum_{i=1}^{n}\int h_i\left(\cdot, \epsilon\right) d\Delta_{iT} \end{aligned} \tag{24}$$

Evaluating at $\epsilon = 0$, and noting that $E\left[U_i^{\alpha_i}\right] = 0$, we obtain

$$\theta^\epsilon(0) = \left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \int U_i d\Delta_{iT}\right) \tag{25}$$

We therefore have

$$\begin{aligned} \sqrt{nT}\frac{1}{\sqrt{T}}\theta^\epsilon(0) &= \left(\frac{1}{n}\sum_{i=1}^n \mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T U_i\right) \\ &\to N\left(0, \mathcal{I}^{-1}\Omega\left(\mathcal{I}'\right)^{-1}\right) \end{aligned}$$

B.2 α_i^θ and α_i^ϵ

In the ith stratum, $\alpha_i(\theta, F_i(\epsilon))$ solves the estimating equation

$$\int V_i(\cdot;\theta,\alpha_i(\theta,F_i(\epsilon)))\,dF_i(\epsilon) = 0 \tag{26}$$

Differentiating the LHS with respect to θ and ϵ, we obtain

$$\begin{aligned} 0 &= \int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta}dF_i(\epsilon) + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}, \\ 0 &= \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} + \int V_i(\cdot,\theta,\epsilon)\,d\Delta_{iT}. \end{aligned}$$

Observe that

$$\begin{aligned} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta} &= -\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)^{-1}\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta}dF_i(\epsilon)\right), \\ \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} &= -\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)^{-1}\left(\int V_i(\cdot,\theta,\epsilon)\,d\Delta_{iT}\right). \end{aligned}$$

Equating these equations to zero and solving for derivatives of α_i evaluated at $\epsilon = 0$ gives

$$\alpha_i^\theta = -\frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}, \tag{27}$$

$$\alpha_i^\epsilon = -\frac{\sum_{t=1}^T V_{it}}{\sqrt{T}E\left[\frac{\partial V_i}{\partial\alpha_i}\right]} = -\frac{T^{-1/2}\sum_{t=1}^T V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}, \tag{28}$$

where $\alpha_i^\theta \equiv \frac{\partial\alpha_i(\theta,F_i(0))}{\partial\theta}$, and $\alpha_i^\epsilon \equiv \frac{\partial\alpha_i(\theta,F_i(0))}{\partial\epsilon}$.

B.3 $\alpha_i^{\theta\theta}$, $\alpha_i^{\theta\epsilon}$, and $\alpha_i^{\epsilon\epsilon}$

Second order differentiation $\left(\frac{\partial^2}{\partial\theta^2}, \frac{\partial^2}{\partial\theta\partial\epsilon}, \frac{\partial^2}{\partial\epsilon^2}\right)$ of (26) yields

$$\begin{aligned} 0 = &\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\theta'}dF_i(\epsilon) + \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i\partial\theta'}dF_i(\epsilon)\right) \\ &+ \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i\partial\theta}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta'} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\theta'} \\ &+ \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta'}, \end{aligned}$$

$$
\begin{aligned}
0 &= \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\epsilon} \\
&\quad + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
&\quad + \int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta} d\Delta_{iT} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta},
\end{aligned}
$$

and

$$
\begin{aligned}
0 &= \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2 \\
&\quad + 2\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}.
\end{aligned}
$$

These three equalities characterizes $\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\theta'}$, $\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\epsilon}$, and $\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2}$.

Lemma 8

$$
T\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\widehat{\theta}(\epsilon)-\theta_0\right| \ge \eta\right] = o(1)
$$

and

$$
T\Pr\left[\max_{1\le i\le n}\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\widehat{\alpha}_i(\epsilon)-\alpha_{i0}\right| \ge \eta\right] = o(1)
$$

for every $\eta > 0$.

Proof. Only the first assertion is proved. The second assertion can be proved similarly. Let η be given, and let $\varepsilon \equiv \inf_i \left[G_{(i)}(\theta_0,\alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} G_{(i)}(\theta,\alpha)\right] > 0$. Recall that

$$
F(\epsilon) \equiv F + \epsilon\sqrt{T}\left(\widehat{F}-F\right), \qquad \epsilon \in \left[0, \frac{1}{\sqrt{T}}\right]
$$

We have

$$
\int g(\cdot;\theta,\alpha_i(\theta))\, dF_i(\epsilon) = \left(1-\epsilon\sqrt{T}\right) G_{(i)}(\theta,\alpha_i) + \epsilon\sqrt{T}\widehat{G}_{(i)}(\theta,\alpha_i)
$$

and

$$
\left|\int g(\cdot;\theta,\alpha_i(\theta))\, dF_i(\epsilon) - G_{(i)}(\theta,\alpha_i)\right| \le \left(1-\epsilon\sqrt{T}\right)\left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right| \le \left|\widehat{G}_{(i)}(\theta,\alpha) - G_{(i)}(\theta,\alpha)\right|.
$$

By Lemma 7, we have

$$
\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\max_{1\le i\le n}\sup_{(\theta,\alpha)}\left|\int g(\cdot;\theta,\alpha_i(\theta))\, dF_i(\epsilon) - G_{(i)}(\theta,\alpha)\right| \ge \eta\right] = o\left(T^{-1}\right)
$$

Therefore, for every $0 \leq \epsilon \leq \frac{1}{\sqrt{T}}$ with probability equal to $1 - o\left(\frac{1}{T}\right)$, we have

$$\begin{aligned}
\max_{|\theta-\theta_0|>\eta,\alpha_1,\ldots,\alpha_n} n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta,\alpha_i\left(\theta\right)\right) dF_i\left(\epsilon\right) &\leq \max_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta,\alpha_i\left(\theta\right)\right) dF_i\left(\epsilon\right) \\
&< \max_{|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta} n^{-1} \sum_{i=1}^{n} G_{(i)}\left(\theta,\alpha_i\right) + \frac{1}{3}\varepsilon \\
&< n^{-1} \sum_{i=1}^{n} G_{(i)}\left(\theta_0,\alpha_{i0}\right) - \frac{2}{3}\varepsilon \\
&< n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta_0,\alpha_{i0}\right) dF_i\left(\epsilon\right) - \frac{1}{3}\varepsilon.
\end{aligned}$$

We also have

$$\max_{\theta,\alpha_1,\ldots,\alpha_n} n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta,\alpha_i\right) dF_i\left(\epsilon\right) \geq n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta_0,\alpha_{i0}\right) dF_i\left(\epsilon\right)$$

by definition. It follows that

$$\max_{|\theta-\theta_0|>\eta,\alpha_1,\ldots,\alpha_n} n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta,\alpha_i\left(\theta\right)\right) dF_i\left(\epsilon\right) < \max_{\theta,\alpha_1,\ldots,\alpha_n} n^{-1} \sum_{i=1}^{n} \int g\left(\cdot;\theta,\alpha_i\right) dF_i\left(\epsilon\right) - \frac{1}{3}\varepsilon$$

for every $0 \leq \epsilon \leq \frac{1}{\sqrt{T}}$. We therefore obtain that $\Pr\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\widehat{\theta}\left(\epsilon\right) - \theta_0\right| \geq \eta\right] = o\left(\frac{1}{T}\right)$. ■

Lemma 9 *Suppose that $K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right)$ is equal to*

$$\frac{\partial^{m_1+m_2}\psi\left(x_{it};\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right)}{\partial\theta^{m_1}\partial\alpha_i^{m_2}}$$

for some $m_1 + m_2 \leq 1, \ldots, 5$. Then, for any $\eta > 0$, we have

$$\Pr\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\frac{1}{n}\sum_{i=1}^{n} \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - \frac{1}{n}\sum_{i=1}^{n} E\left[K_i\left(x_{it};\theta_0,\alpha_{i0}\right)\right]\right| > \eta\right] = o\left(T^{-1}\right)$$

and

$$\Pr\left[\max_{i} \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - E\left[K_i\left(x_{it};\theta_0,\alpha_{i0}\right)\right]\right| > \eta\right] = o\left(T^{-1}\right).$$

Also,

$$\Pr\left[\max_{i} \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) d\Delta_{iT}\right| > CT^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right)$$

for some constant $C > 0$ and $0 < \upsilon < \left(100q + 120\right)^{-1}$.

Proof. Note that we may write

$$\begin{aligned}
&\left\| \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i \right\| \\
\leq\; &\left\| \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - \int K_i\left(\cdot;\theta\left(0\right),\alpha_i\left(\theta\left(0\right),0\right)\right) dF_i\left(\epsilon\right) \right\| \\
&+ \left\| \int K_i\left(\cdot;\theta\left(0\right),\alpha_i\left(\theta\left(0\right),0\right)\right) dF_i\left(\epsilon\right) - \int K_i\left(\cdot;\theta\left(0\right),\alpha_i\left(\theta\left(0\right),0\right)\right) dF_i \right\| \\
\leq\; &\int M(x_{it})\left(\left\|\theta\left(\epsilon\right)-\theta\right\| + \left|\alpha_i\left(\theta(\epsilon),\epsilon\right)-\alpha_i\right|\right) d\left|F_i\left(\epsilon\right)\right| \\
&+ \epsilon\sqrt{T}\left\| \int K_i\left(\cdot;\theta\left(0\right),\alpha_i\left(\theta\left(0\right),0\right)\right) d\left(\widehat{F}_i - F_i\right) \right\|.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{i=1}^{n} \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - \frac{1}{n}\sum_{i=1}^{n} E\left[K_i\left(x_{it};\theta_0,\alpha_{i0}\right)\right] \right| \\
\leq\; &\left\|\theta\left(\epsilon\right)-\theta\right\| \cdot \frac{1}{n}\sum_{i=1}^{n}\left(E\left[M\left(x_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T} M\left(x_{it}\right)\right) \\
&+ \left(\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)-\alpha_i\right)^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n}\left(E\left[M\left(x_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T} M\left(x_{it}\right)\right)^2\right)^{1/2} \\
&+ \left\| \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T} K_i\left(x_{it};\theta\left(0\right),\alpha_i\left(\theta\left(0\right),0\right)\right) - E\left[K_i\left(x_{it};\theta\left(0\right),\alpha_i\left(\theta\left(0\right),0\right)\right)\right]\right)\right\|,
\end{aligned}$$

the RHS of which can be bounded by using Lemmas 4 and 8 in absolute value by some $\eta > 0$ with probability $1 - o\left(T^{-1}\right)$.

Because

$$\begin{aligned}
&\left| \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - E\left[K_i\left(x_{it};\theta_0,\alpha_{i0}\right)\right] \right| \\
\leq\; &\left|\theta\left(\epsilon\right)-\theta\right| \cdot \left(E\left[M\left(x_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T} M\left(x_{it}\right)\right) \\
&+ \left|\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)-\alpha_i\right| \cdot \left(E\left[M\left(x_{it}\right)\right] + \frac{1}{T}\sum_{t=1}^{T} M\left(x_{it}\right)\right) \\
&+ \left| \frac{1}{T}\sum_{t=1}^{T} M\left(x_{it}\right) - E\left[M\left(x_{it}\right)\right] \right|,
\end{aligned}$$

we can bound

$$\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left| \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) dF_i\left(\epsilon\right) - E\left[K_i\left(x_{it};\theta_0,\alpha_{i0}\right)\right] \right|$$

in absolute value by some $\eta > 0$ with probability $1 - o\left(T^{-1}\right)$.

Using Lemmas 5, we can also show that

$$\max_i \left| \int K_i\left(\cdot;\theta\left(\epsilon\right),\alpha_i\left(\theta\left(\epsilon\right),\epsilon\right)\right) d\Delta_{iT} \right|$$

can be bounded by in absolute value by $CT^{\frac{1}{10}-\upsilon}$ for some constant $C>0$ and υ such that $0\leq\upsilon<\frac{1}{160}$ with probability $1-o\left(T^{-1}\right)$. ■

Lemma 10

$$\begin{aligned}\Pr\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\alpha_i^{\theta}\left(\epsilon\right)\right|>C\right] &= o\left(T^{-1}\right)\\ \Pr\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\alpha_i^{\epsilon}\left(\epsilon\right)\right|>CT^{\frac{1}{10}-\upsilon}\right] &= o\left(T^{-1}\right)\end{aligned}$$

for some constant $C>0$ *and* $0<\upsilon<(100q+120)^{-1}$.

Proof. >From Appendix B.2, we obtain

$$\begin{aligned}\frac{\partial\alpha_i\left(\theta,F_i\left(\epsilon\right)\right)}{\partial\theta} &= -\left(\int\frac{\partial V_i\left(\cdot,\theta,\epsilon\right)}{\partial\alpha_i}dF_i\left(\epsilon\right)\right)^{-1}\left(\int\frac{\partial V_i\left(\cdot,\theta,\epsilon\right)}{\partial\theta}dF_i\left(\epsilon\right)\right),\\ \frac{\partial\alpha_i\left(\theta,F_i\left(\epsilon\right)\right)}{\partial\epsilon} &= -\left(\int\frac{\partial V_i\left(\cdot,\theta,\epsilon\right)}{\partial\alpha_i}dF_i\left(\epsilon\right)\right)^{-1}\left(\int V_i\left(\cdot,\theta,\epsilon\right)d\Delta_{iT}\right).\end{aligned}$$

Using Lemma 9, we can see that

$$\left(\int\frac{\partial V_i\left(\cdot,\theta,\epsilon\right)}{\partial\alpha_i}dF_i\left(\epsilon\right)\right)^{-1}$$

is uniformly bounded away from zero with probability $1-o\left(T^{-1}\right)$. We can also see that, with probability $1-o\left(T^{-1}\right)$,

$$\int\frac{\partial V_i\left(\cdot,\theta,\epsilon\right)}{\partial\theta}dF_i\left(\epsilon\right)$$

is uniformly bounded by some constant C, and

$$\int V_i\left(\cdot,\theta,\epsilon\right)d\Delta_{iT}$$

is uniformly bounded by $CT^{\frac{1}{10}-\upsilon}$. ■

Lemma 11

$$\Pr\left[\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\theta^{\epsilon}\left(\epsilon\right)\right|>CT^{\frac{1}{10}-\upsilon}\right]=o\left(T^{-1}\right)$$

for some constant $C>0$ *and* $0<\upsilon<(100q+120)^{-1}$.

Proof. >From (24), we have

$$\begin{aligned}\theta^{\epsilon}\left(\epsilon\right) &= -\left[\frac{1}{n}\sum_{i=1}^{n}\int\left(\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\theta'}+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\theta'}\right)dF_i\left(\epsilon\right)\right]^{-1}\\ &\quad\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\alpha_i}{\partial\epsilon}\left(\int\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}dF_i\left(\epsilon\right)\right)+\frac{1}{n}\sum_{i=1}^{n}\int h_i\left(\cdot,\epsilon\right)d\Delta_{iT}\right]\end{aligned}$$

Using Lemmas 9, and 10, we can bound the denominator of $\theta^{\epsilon}\left(\epsilon\right)$ by some $C>0$, and the numerator by some $CT^{\frac{1}{10}-\upsilon}$ with probability $1-o\left(T^{-1}\right)$. ■

Lemma 12

$$\begin{aligned}
\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\alpha_i^{\theta_r\theta_{r'}}(\epsilon)\right|>C\right] &= o\left(T^{-1}\right)\\
\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\alpha_i^{\theta_r\epsilon}(\epsilon)\right|>CT^{\frac{1}{10}-\upsilon}\right] &= o\left(T^{-1}\right)\\
\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\alpha_i^{\epsilon\epsilon}(\epsilon)\right|>C\left(T^{\frac{1}{10}-\upsilon}\right)^2\right] &= o\left(T^{-1}\right)
\end{aligned}$$

for some constant $C>0$ *and* $0<\upsilon<(100q+120)^{-1}$*. Here,* $\alpha_i^{\theta_r\theta_{r'}}\equiv\frac{\partial^2\alpha_i}{\partial\theta_r\partial\theta_{r'}}$*. We similarly define* $\alpha_i^{\theta_r\epsilon}$*.*

Proof. >From Appendix B.3, we have

$$\begin{aligned}
0 &= \int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\theta'}dF_i(\epsilon)+\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\left(\int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i\partial\theta'}dF_i(\epsilon)\right)\\
&\quad+\left(\int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i\partial\theta}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta'}+\left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\theta'}\\
&\quad+\left(\int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta'},
\end{aligned}$$

$$\begin{aligned}
0 &= \left(\int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\\
&\quad+\left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta\partial\epsilon}+\left(\int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}dF_i(\epsilon)\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta}\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\\
&\quad+\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta}d\Delta_{iT}+\left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}d\Delta_{iT}\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta},
\end{aligned}$$

and

$$\begin{aligned}
0 &= \left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)\frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2}+\left(\int\frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2}dF_i(\epsilon)\right)\left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2\\
&\quad+2\left(\int\frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i}d\Delta_{iT}\right)\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}.
\end{aligned}$$

The result then follows by applying the same argument as in the proof of Lemma 10. ■

Lemma 13

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}}\left|\theta^{\epsilon\epsilon}(\epsilon)\right|>C\left(T^{\frac{1}{10}-\upsilon}\right)^2\right]=o\left(T^{-1}\right)$$

for some constant $C>0$ *and* $0<\upsilon<(100q+120)^{-1}$*.*

Proof. The conclusion follows by using the characterization of $\theta^{\epsilon\epsilon}(\epsilon)$ in Appendix C, and Lemmas 9, 10, 11, and 12. ■

Lemma 14

$$\begin{aligned}
\Pr\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\theta_r\theta_{r'}\theta_{r''}}(\epsilon)\right| > C\right] &= o\left(T^{-1}\right) \\
\Pr\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\theta_r\theta_{r'}\epsilon}(\epsilon)\right| > CT^{\frac{1}{10}-\upsilon}\right] &= o\left(T^{-1}\right) \\
\Pr\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\theta_r\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^2\right] &= o\left(T^{-1}\right) \\
\Pr\left[\max_i \max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}} \left|\alpha_i^{\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^3\right] &= o\left(T^{-1}\right)
\end{aligned}$$

for some constant $C > 0$ *and* $0 < \upsilon < (100q + 120)^{-1}$.

Proof. It was seen in Appendix B.3 that

$$\begin{aligned}
0 = & \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\epsilon} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r\partial\epsilon} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\epsilon} \\
& + \int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\theta_r} d\Delta_{iT} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r},
\end{aligned}$$

and

$$\begin{aligned}
0 = & \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta, F_i(\epsilon))}{\partial\epsilon^2} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \left(\frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\epsilon}\right)^2 \\
& + 2\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\epsilon}.
\end{aligned}$$

We therefore obtain

$$
\begin{aligned}
0 \quad = \quad & \int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r \partial\theta_{r'} \partial\theta_{r''}} dF_i(\epsilon) + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i \partial\theta_r \partial\theta_{r'}} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r''}} \\
& + \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r''}} \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i \partial\theta_{r'}} dF_i(\epsilon) \right) + \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r} \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i \partial\theta_{r'} \partial\theta_{r''}} dF_i(\epsilon) \right) \\
& + \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r''}} \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2 \partial\theta_{r'}} dF_i(\epsilon) \right) \\
& + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i \partial\theta_r \partial\theta_{r''}} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'}} + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2 \partial\theta_r} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'}} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r''}} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i \partial\theta_r} dF_i(\epsilon) \right) \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'} \partial\theta_{r''}} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i \partial\theta_{r''}} dF_i(\epsilon) \right) \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'}} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r''}} \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'}} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon) \right) \frac{\partial^3 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'} \partial\theta_{r''}} \\
& + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2 \partial\theta_{r''}} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'}} \\
& + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^3} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'}} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r''}} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon) \right) \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r''}} \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'}} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon) \right) \frac{\partial\alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r} \frac{\partial^2 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_{r'} \partial\theta_{r''}},
\end{aligned}
$$

which characterizes $\frac{\partial^3 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'} \partial\theta_{r''}}$,

$$\begin{aligned}
0 \quad = \quad & \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r \partial\theta_{r'} \partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r \partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r \partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}\partial\epsilon} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_{r'} \partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\epsilon} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}} \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\epsilon} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^3\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\theta_{r'}\partial\epsilon} \\
& + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\theta_{r'} \partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^3} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\theta_{r'}} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}\partial\epsilon} \\
& + \int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r\partial\theta_{r'}} d\Delta_{iT} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}} \\
& + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_{r'}\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_{r'}} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\theta_{r'}},
\end{aligned}$$

which characterizes $\frac{\partial^3 \alpha_i(\theta, F_i(\epsilon))}{\partial\theta_r \partial\theta_{r'} \partial\epsilon}$,

$$\begin{aligned}
0 \quad = \quad & \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r \partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} \\
& + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^3\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\epsilon^2} \\
& + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r\partial\alpha_i^2} dF_i(\epsilon)\right) \left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2 + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^3} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2 \\
& +2\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\epsilon} \\
& +2\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\theta_r\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} + 2\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} d\Delta_{iT}\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r} \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \\
& +2\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\epsilon},
\end{aligned}$$

which characterizes $\frac{\partial^3\alpha_i(\theta,F_i(\epsilon))}{\partial\theta_r\partial\epsilon^2}$, and

$$\begin{aligned}
0 &= \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} \\
&\quad + \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} dF_i(\epsilon)\right) \frac{\partial^3\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^3} \\
&\quad + \left(\int \frac{\partial^3 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^3} dF_i(\epsilon)\right) \left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^3 + \left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} d\Delta_{iT}\right) \left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2 \\
&\quad +2\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} dF_i(\epsilon)\right) \frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon} \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2} \\
&\quad +2\left(\int \frac{\partial^2 V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i^2} d\Delta_{iT}\right) \left(\frac{\partial\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon}\right)^2 + 2\left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial\alpha_i} d\Delta_{iT}\right) \frac{\partial^2\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^2},
\end{aligned}$$

which characterizes $\frac{\partial^3\alpha_i(\theta,F_i(\epsilon))}{\partial\epsilon^3}$. Inspecting these derivatives and applying Lemmas 9, 10, and 12, we obtain the desired result. ■

Theorem 6

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\theta^{\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^3\right] = o\left(T^{-1}\right)$$

for some constant $C>0$ and $0<\upsilon<(100q+120)^{-1}$.

Proof. >From (23), we have

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3 h_i(\cdot,\epsilon)}{d\epsilon^3} dF_i(\epsilon) + 3\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2 h_i(\cdot,\epsilon)}{d\epsilon^2} d\Delta_{iT}$$

Combining Lemmas 9, 10, 11, 12, and 13, we can bound $\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2 h_i(\cdot,\epsilon)}{d\epsilon^2} d\Delta_{iT}$ by $C\left(T^{\frac{1}{10}-\upsilon}\right)^3$ with probability $1-o\left(T^{-1}\right)$. It was seen in Appendix C that the r-th component $\frac{d^2 h_i^{(r)}(\cdot,\epsilon)}{d\epsilon^2}$of $\frac{d^2 h_i(\cdot,\epsilon)}{d\epsilon^2}$ is equal to

$$\begin{aligned}
\frac{d^2 h_i^{(r)}(\cdot,\epsilon)}{d\epsilon^2} &= \frac{\partial\theta(\cdot,\epsilon)'}{\partial\epsilon} \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\theta\partial\theta'} \frac{\partial\theta(\cdot,\epsilon)}{\partial\epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\theta}{\partial\epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}\frac{\partial\theta}{\partial\epsilon} \\
&\quad + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2} \\
&\quad + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)^2 + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\alpha_i}{\partial\epsilon} \\
&\quad + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i}\left(\frac{\partial\theta'}{\partial\epsilon}\frac{\partial^2\alpha_i}{\partial\theta\partial\theta'}\right)\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\theta'\partial\epsilon}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2} \\
&\quad + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\alpha_i}{\partial\epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\epsilon}\right)^2 \\
&\quad + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\epsilon\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\epsilon^2}.
\end{aligned}$$

Using Lemmas 9, 10, 11, 12, and 13 again, we can conclude that $\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3 h_i(\cdot,\epsilon)}{d\epsilon^3} dF_i(\epsilon)$ is equal to $\left(\frac{1}{n}\sum_{i=1}^{n}\int \frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'} dF_i(\epsilon)\right)\frac{\partial^3\theta}{\partial\epsilon^3}$ plus terms that can all be bounded by $\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2 h_i(\cdot,\epsilon)}{d\epsilon^2} d\Delta_{iT}$ by $C\left(T^{\frac{1}{10}-\upsilon}\right)^3$

with probability $1-o\left(T^{-1}\right)$. Because $\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial h_i(\cdot,\epsilon)}{\partial\theta'}dF_i\left(\epsilon\right)\right)^{-1}$ is bounded away from 0 by Lemma 9, we obtain the desired conclusion. ■

C Proof of Lemma (2)

Note that

$$
\begin{aligned}
\frac{d^2 h_i\left(\cdot,\epsilon\right)}{d\epsilon^2} = {} & \mathcal{G}_i\left(\cdot,\epsilon\right)+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\theta'\partial\alpha_i}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\theta}{\partial\epsilon}+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\theta'\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}\frac{\partial\theta}{\partial\epsilon} \\
&+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2} \\
&+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)^2+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\alpha_i}{\partial\epsilon} \\
&+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}\left(\frac{\partial\theta'}{\partial\epsilon}\frac{\partial^2\alpha_i}{\partial\theta\partial\theta'}\right)\frac{\partial\theta}{\partial\epsilon}+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\theta'\partial\epsilon}\frac{\partial\theta}{\partial\epsilon}+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial^2\theta}{\partial\epsilon^2} \\
&+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\epsilon}+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon}\right)\frac{\partial\alpha_i}{\partial\epsilon}+\frac{\partial^2 h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i^2}\left(\frac{\partial\alpha_i}{\partial\epsilon}\right)^2 \\
&+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\epsilon\partial\theta'}\frac{\partial\theta}{\partial\epsilon}+\frac{\partial h_i\left(\cdot,\epsilon\right)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{\partial\epsilon^2},
\end{aligned}
$$

where $\mathcal{G}_i\left(\cdot,\epsilon\right)$ is a R-dimensional column vector such that its r-th element $\mathcal{G}_i^{(r)}\left(\cdot,\epsilon\right)$ is equal to

$$
\mathcal{G}_i^{(r)}\left(\cdot,\epsilon\right)=\frac{\partial\theta\left(\cdot,\epsilon\right)'}{\partial\epsilon}\frac{\partial^2 h_i^{(r)}\left(\cdot,\epsilon\right)}{\partial\theta\partial\theta'}\frac{\partial\theta\left(\cdot,\epsilon\right)}{\partial\epsilon},
$$

and $h_i^{(r)}\left(\cdot,\epsilon\right)$ denotes the r-th element of h_i.

Evaluating each term of (22) at $\epsilon=0$, and noting that $E\left[U_i^{\alpha_i}\right]=0$, we obtain

$$
\begin{aligned}
0 = {} & \frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial U_i}{\partial\theta'}\right]\theta^{\epsilon\epsilon}\left(0\right) \\
&+\frac{2}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\theta^{\epsilon}\left(0\right)+\frac{2}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\left(\theta^{\epsilon}\left(0\right)'\alpha_i^{\theta}\right)\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] \\
&+\mathcal{G}+\frac{2}{n}\sum_{i=1}^{n}\theta^{\epsilon}\left(0\right)'\alpha_i^{\theta}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\theta^{\epsilon}\left(0\right)+\frac{1}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}\left(0\right)'\alpha_i^{\theta}\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] \\
&+\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i^{\epsilon}\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] \\
&+\frac{2}{n}\sum_{i=1}^{n}\left(\int\frac{\partial U_i}{\partial\theta'}d\Delta_{iT}\right)\theta^{\epsilon}\left(0\right)+\frac{2}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}\left(0\right)'\alpha_i^{\theta}\right)\cdot\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT}+\frac{2}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT}
\end{aligned}
$$

where

$$
\mathcal{G}\equiv\begin{bmatrix}\theta^{\epsilon}\left(0\right)'\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^2 U_i^{(1)}}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}\left(0\right)\\ \vdots \\ \theta^{\epsilon}\left(0\right)'\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^2 U_i^{(R)}}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}\left(0\right)\end{bmatrix}
$$

from which we obtain

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)\theta^{\epsilon\epsilon}(0)=\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i^{\epsilon}\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]+\frac{2}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT}+\mathcal{T} \tag{29}$$

for

$$\begin{aligned}\frac{1}{n}\mathcal{T} \equiv\ & 2\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\right)\theta^{\epsilon}(0)+\left(\frac{2}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]\left(\alpha_i^{\theta}\right)'\right)\theta^{\epsilon}(0)\\ &+2\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\theta'}d\Delta_{iT}\right)\theta^{\epsilon}(0)+\left(\frac{2}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT}\cdot\left(\alpha_i^{\theta}\right)'\right)\theta^{\epsilon}(0)\\ &+\mathcal{G}+\frac{2}{n}\sum_{i=1}^{n}\theta^{\epsilon}(0)'\,\alpha_i^{\theta}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\theta^{\epsilon}(0)\\ &+\frac{1}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}(0)'\,\alpha_i^{\theta}\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]\end{aligned} \tag{30}$$

Note that we may rewrite

$$\begin{aligned}\mathcal{T} \equiv\ & 2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(-\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &+\left(\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(-\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]\left(-\frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &+2\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{\partial U_i}{\partial\theta'}-E\left[\frac{\partial U_i}{\partial\theta'}\right]\right)\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &+\left(\frac{2}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{\partial U_i}{\partial\theta'}-E\left[\frac{\partial U_i}{\partial\theta'}\right]\right)\cdot\left(-\frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &+\mathcal{G}^*\\ &+\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i'\right)^{-1}\left(\frac{2}{n}\sum_{i=1}^{n}\left(-\frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &+\frac{1}{n}\sum_{i=1}^{n}\left(\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i'\right)^{-1}\left(-\frac{E\left[\frac{\partial V_i}{\partial\theta}\right]}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right],\end{aligned} \tag{31}$$

where

$$\mathcal{G}^*\equiv\begin{bmatrix}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^2 U_i^{(1)}}{\partial\theta\partial\theta'}\right]\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ \vdots\\ \left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^2 U_i^{(R)}}{\partial\theta\partial\theta'}\right]\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\end{bmatrix}$$

Next note that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i^{\epsilon}\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] &= \frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\alpha_i\alpha_i}\right]\left(\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)^2 \\
\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT} &= -\frac{1}{n}\sum_{i=1}^{n}\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\alpha_i}-E\left[U_i^{\alpha_i}\right]\right)\right)
\end{aligned}
$$

and that by Lemma (6) it follows that

$$
\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right) \tag{32}
$$

$$
\left(\frac{1}{n}\sum_{i=1}^{n}\alpha_i^{\epsilon}E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right]\left(\alpha_i^{\theta}\right)'\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right) \tag{33}
$$

$$
\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\theta'}d\Delta_{iT}\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right) \tag{34}
$$

$$
\left(\frac{1}{n}\sum_{i=1}^{n}\int\frac{\partial U_i}{\partial\alpha_i}d\Delta_{iT}\cdot\left(\alpha_i^{\theta}\right)'\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right) \tag{35}
$$

and

$$
\theta^{\epsilon}(0)'\left(\frac{1}{n}\sum_{i=1}^{n}E\left[\frac{\partial^2 U_i^{(r)}}{\partial\theta\partial\theta'}\right]\right)\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O(1)\,O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right) \tag{36}
$$

$$
\frac{1}{n}\sum_{i=1}^{n}\theta^{\epsilon}(0)'\alpha_i^{\theta}\cdot E\left[\frac{\partial^2 U_i}{\partial\theta'\partial\alpha_i}\right]\theta^{\epsilon}(0) = O_p\left(\frac{1}{\sqrt{n}}\right)O(1)\,O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{n}\right) \tag{37}
$$

$$
\frac{1}{n}\sum_{i=1}^{n}\left(\theta^{\epsilon}(0)'\alpha_i^{\theta}\right)^2\cdot E\left[\frac{\partial^2 U_i}{\partial\alpha_i^2}\right] = O_p\left(\frac{1}{\sqrt{n}}\right)^2 O(1)=O_p\left(\frac{1}{n}\right). \tag{38}
$$

These arguments establish that

$$
\begin{aligned}
\theta^{\epsilon\epsilon}(0) &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\alpha_i\alpha_i}\right]\left(\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right)^2 \\
&\quad -2\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{T^{-1/2}\sum_{t=1}^{T}V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\alpha_i}-E\left[U_i^{\alpha_i}\right]\right)\right) \\
&\quad +O_p\left(\frac{1}{n}\right) \\
&= -2\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{t=1}^{T}V_{it}}{\sqrt{T}E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\alpha_i}-\frac{E\left[U_i^{\alpha_i\alpha_i}\right]}{2E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}V_{it}\right)\right]+o_p(1).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\theta^{\epsilon\epsilon}(0) \\
&= -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{V_{it}}{E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_i^{\alpha_i}-\frac{E\left[U_i^{\alpha_i\alpha_i}\right]}{2E\left[\frac{\partial V_i}{\partial\alpha_i}\right]}V_{it}\right)\right]\right\}+o_p(1)
\end{aligned}
$$

Let

$$Z_{iT} \equiv \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{V_{it}}{E\left[\frac{\partial V_i}{\partial \alpha_i}\right]} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(U_i^{\alpha_i} - \frac{E\left[U_i^{\alpha_i \alpha_i}\right]}{2E\left[\frac{\partial V_i}{\partial \alpha_i}\right]} V_{it} \right) \right]$$

Then, Z_{iT} are independent across i and

$$E\left[Z_{it}\right] = \frac{\Sigma_{iT}^{VU}}{E\left[\frac{\partial V_i}{\partial \alpha_i}\right]} - \frac{E\left[U_i^{\alpha_i \alpha_i}\right]}{2\left(E\left[\frac{\partial V_i}{\partial \alpha_i}\right]\right)^2} \Sigma_{iT}^{VV}$$

where $\Sigma_{iT}^{VU} \equiv T^{-1} \sum_{t,s=1}^{T} E\left[V_{it} U_i^{\alpha_i \prime}\right]$ and $\Sigma_{iT}^{VV} \equiv T^{-1} \sum_{t,s=1}^{T} E\left[V_{it} V_{is}'\right]$. Next note that

$$\begin{aligned} \operatorname{Var}(Z_{iT}) = T^{-2} \sum_{t_1,\ldots,t_4=1}^{T} \Bigg\{ & \frac{E\left[V_{it_1} V_{it_4} U_{it_2}^{\alpha_i} U_{it_3}^{\alpha_i'}\right] - \Sigma_{iT}^{VU} \Sigma_{iT}^{VU\prime}}{\left(E\left[\frac{\partial V_i}{\partial \alpha_i}\right]\right)^2} \\ & - \frac{E\left[U_{it_3}^{\alpha_i \alpha_i}\right] \left(E\left[V_{it_1} V_{it_4} V_{it_3} U_{it_2}^{\alpha_i'}\right] - \Sigma_{iT}^{VV} \Sigma_{iT}^{VU\prime}\right)}{2\left(E\left[\frac{\partial V_i}{\partial \alpha_i}\right]\right)^3} \\ & - \frac{\left(E\left[V_{it_1} V_{it_4} V_{it_2} U_{it_3}^{\alpha_i}\right] - \Sigma_{iT}^{VV} \Sigma_{iT}^{VU}\right) E\left[U_{it_2}^{\alpha_i \alpha_i}\right]}{2\left(E\left[\frac{\partial V_i}{\partial \alpha_i}\right]\right)^3} \\ & + \frac{E\left[U_{it_2}^{\alpha_i \alpha_i}\right] E\left[U_{it_3}^{\alpha_i \alpha_i}\right]' \left(E\left[V_{it_1} V_{it_2} V_{it_3} V_{it_4}\right] - \Sigma_{iT}^{VV} \Sigma_{iT}^{VV}\right)}{4\left(E\left[\frac{\partial V_i}{\partial \alpha_i}\right]\right)^4} \Bigg\} \end{aligned}$$

Note that $V_{it_k}, U_{it_{k'}}^{\alpha_i}$ are random variables measurable with respect to the filtration generated by x_{it}. By Condition (3) sufficient moments exist to apply Corollary A.2 of Hall and Heyde (1980, p.278) as well as Lemma 1 of Andrews (1991). First note that for any element (j_1, j_2) we have

$$\begin{aligned} E\left[V_{it_1} V_{it_4} U_{it_2}^{\alpha_i, j_1} U_{it_3}^{\alpha_i' j_2}\right] - \left[\Sigma_{iT}^{VU} \Sigma_{iT}^{VU\prime}\right]_{j_1, j_2} &= E\left[V_{it_1} V_{it_4}\right] E\left[U_{it_2}^{\alpha_i, j_1} U_{it_3}^{\alpha_i' j_2}\right] \\ &+ E\left[V_{it_1} U_{it_3}^{\alpha_i' j_2}\right] E\left[V_{it_4} U_{it_2}^{\alpha_i, j_1}\right] + \operatorname{Cum}\left(V_{it_1}, V_{it_4}, U_{it_2}^{\alpha_i, j_1}, U_{it_3}^{\alpha_i' j_2}\right) \end{aligned}$$

where $\operatorname{Cum}\left(V_{it_1}, V_{it_4}, U_{it_2}^{\alpha_i, j_1}, U_{it_3}^{\alpha_i' j_2}\right)$ is uniformly summable. For $\delta > 0$ and some constant $0 < c < \infty$ it follows from Corollary A.2 of Hall and Heyde (1980, p.278) and Condition (2) that

$$\sup_i \left| E\left[V_{it_1} U_{it_2}^{\alpha_i, j_1}\right] \right| \leq 8c \left(E\left[\left|V_{it_1}\right|^{2+\delta}\right]\right)^{\frac{1}{2+\delta}} \left(E\left[\left|U_{it_2}^{\alpha_i, j_1}\right|^{2+\delta}\right]\right)^{\frac{1}{2+\delta}} \left(a^{\frac{\delta}{2+\delta}}\right)^{|t_1 - t_2|}$$

with similar inequalities holding for the remaining second moments. These arguments establish that

$$\sup_i \left| T^{-2} \sum_{t_1,\ldots,t_4=1}^{T} E\left[V_{it_1} V_{it_4} U_{it_2}^{\alpha_i} U_{it_3}^{\alpha_i'}\right] - \Sigma_{iT}^{VU} \Sigma_{iT}^{VU\prime} \right| = O(1)$$

and the same can be established for the remaining terms in $\operatorname{Var}(Z_{iT})$. By the Markov inequality it follows that

$$\Pr\left[\left| \frac{1}{n} \sum_{i=1}^{n} \left(Z_{iT} - E Z_{iT}\right) \right| > \eta \right] < \frac{\sup_i \operatorname{Var}(Z_{iT})}{n\eta} \to 0$$

such that $\frac{1}{n}\sum_{i=1}^{n} Z_{iT} = \frac{1}{n}\sum_{i=1}^{n} E\left[Z_{iT}\right] + O_p(n^{-1/2})$. By the same argument as in (12) it then follows that $\sup_i \left|\Sigma_{iT}^{VU} - f_i^{VU^{\alpha}}\right| \to 0$ and $\sup_i \left|\Sigma_{iT}^{VV} - f_i^{VV}\right| \to 0$ as $T \to \infty$. Uniformity of convergence then implies that joint and iterated limits exist and agree such that $\frac{1}{n}\sum_{i=1}^{n} E\left[Z_{iT}\right] \to \Psi$ Therefore, we have

$$\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^{2}\theta^{\epsilon\epsilon}(0) = -\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\Psi + o_p(1).$$

D Proof of Theorem (3)

D.1 Expansion for $\widehat{\alpha}_i$

Let

$$\widehat{\alpha}_i(\epsilon) \equiv \operatorname*{argmax}_a \int \psi\left(x_{it};\widehat{\theta}(\epsilon), a\right) dF_i(\epsilon)$$

>From the first order condition

$$0 = \int V_i\left(\widehat{\theta}(\epsilon), \widehat{\alpha}_i(\epsilon)\right) dF_i(\epsilon).$$

Using the same arguments as earlier, we are looking for the expansion

$$\widehat{\alpha}_i(\epsilon) - \alpha_{i0} = \frac{1}{\sqrt{T}}\widehat{\alpha}_i^{\epsilon}(0) + \frac{1}{2T}\widehat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon})$$

for some $\tilde{\epsilon} \in \left[0, T^{-1/2}\right]$. Let

$$v_i(\cdot,\epsilon) \equiv V_i(\theta(F(\epsilon)), \alpha_i(F_i(\epsilon))).$$

The first order condition may be written as

$$0 = \int v_i(\cdot,\epsilon)\, dF_i(\epsilon)$$

Differentiating repeatedly with respect to ϵ, we obtain

$$0 = \int \frac{dv_i(\cdot,\epsilon)}{d\epsilon} dF_i(\epsilon) + \int v_i(\cdot,\epsilon)\, d\Delta_{iT} \tag{39}$$

$$0 = \int \frac{d^2 v_i(\cdot,\epsilon)}{d\epsilon^2} dF_i(\epsilon) + 2\int \frac{dv_i(\cdot,\epsilon)}{d\epsilon} d\Delta_{iT} \tag{40}$$

where $\Delta_{iT} \equiv \sqrt{T}\left(\widehat{F}_i - F_i\right)$.

D.2 $\widehat{\alpha}^{\epsilon}(0)$

Because

$$\frac{dv_i(\cdot,\epsilon)}{d\epsilon} = \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}$$

we may rewrite (39) as

$$0 = \int \left(\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial\alpha_i}{\partial\epsilon}\right) dF_i(\epsilon) + \int v_i(\cdot,\epsilon)\, d\Delta_{iT}$$

Evaluating at $\epsilon = 0$ we obtain

$$\widehat{\alpha}_i^{\epsilon}(0) = -\left(E\left[V_i^{\alpha}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\psi\left(x_{it};\theta_0,\alpha_{i0}\right)}{\partial\alpha} + E\left[V_i^{\theta}\right]\theta^{\epsilon}(0)\right) \tag{41}$$

where $\theta^{\epsilon}(0)$ is defined in (25). It also follows that

$$\widehat{\alpha}_i^{\epsilon}(\epsilon) = -\left(\int\frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\right)^{-1}\left[\int\left(\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\right)dF_i(\epsilon)\,\theta^{\epsilon}(\epsilon) + \int v_i(\cdot,\epsilon)\,d\Delta_{iT}\right]. \tag{42}$$

Next, consider

$$\begin{aligned}\frac{d^2 v_i(\cdot,\epsilon)}{d\epsilon^2} &= \frac{\partial\theta}{\partial\epsilon}'\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta\partial\theta'}\frac{\partial\theta}{\partial\epsilon} + \frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\frac{\partial^2\theta}{(\partial\epsilon)^2} + 2\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}\frac{\partial\theta}{\partial\epsilon}\frac{\partial\alpha_i}{\partial\epsilon} \\ &\quad + \frac{\partial^2 v_i(\cdot,\epsilon)}{(\partial\alpha_i)^2}\left(\frac{\partial\alpha_i}{\partial\epsilon}\right)^2 + \frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}\frac{\partial^2\alpha_i}{(\partial\epsilon)^2}\end{aligned}$$

such that $\widehat{\alpha}_i^{\epsilon\epsilon}(\epsilon)$ is characterized by

$$\begin{aligned}0 &= \theta^{\epsilon}(\epsilon)'\int\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta\partial\theta'}dF_i(\epsilon)\,\theta^{\epsilon}(\epsilon) + \int\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}dF_i(\epsilon)\,\theta^{\epsilon\epsilon}(\epsilon) \\ &\quad + 2\int\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'\partial\alpha_i}dF_i(\epsilon)\,\theta^{\epsilon}(\epsilon)\,\widehat{\alpha}_i^{\epsilon}(\epsilon) + \int\frac{\partial^2 v_i(\cdot,\epsilon)}{(\partial\alpha_i)^2}dF_i(\epsilon)\left(\widehat{\alpha}_i^{\epsilon}(\epsilon)\right)^2 \\ &\quad + \int\frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}dF_i(\epsilon)\,\widehat{\alpha}_i^{\epsilon\epsilon}(\epsilon) + \int\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}d\Delta_{iT}\theta^{\epsilon}(\epsilon) + \int\frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}d\Delta_{iT}^{\epsilon}\widehat{\alpha}i(\epsilon)\end{aligned} \tag{43}$$

Lemma 15 *Let Conditions (1),(2),(3) and (4) be satisfied. Then*

$$\begin{aligned}\Pr\left[\max_i\sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left|\widehat{\alpha}_i^{\epsilon}(\epsilon)\right| > T^{\frac{1}{10}-\upsilon}\right] &= o\left(T^{-1}\right), \\ \Pr\left[\max_i\left|\widehat{\alpha}_i^{\epsilon}(0)\right| > T^{\frac{1}{10}-\upsilon}\right] &= o\left(T^{-1}\right), \\ \Pr\left[\max_i\sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left|\widehat{\alpha}_i^{\epsilon\epsilon}(\epsilon)\right| > \left(T^{\frac{1}{10}-\upsilon}\right)^2\right] &= o\left(T^{-1}\right), \\ \Pr\left[\max_i\left|\sqrt{T}\left(\widehat{\alpha}_i-\alpha_{i0}\right)\right| > T^{1/10-\upsilon}\right] &= o\left(T^{-1}\right),\end{aligned}$$

for $0 < \upsilon < (100q+120)^{-1}$.

Proof. Note that the last claim follows as a consequence of the first three claims. In order to prove the first claim, we note that

$$\Pr\left[\sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left\|\theta^{\epsilon}(\epsilon)\right\| \geq T^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right)$$

from Lemma (11). By Lemma (9), we also have

$$\begin{aligned}\Pr\left[\max_i\sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left\|\int\left(\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\right)dF_i(\epsilon) - E\left[\frac{\partial v_i(\cdot,\epsilon)}{\partial\theta'}\right]\right\| > \eta\right] &= o\left(T^{-1}\right), \\ \Pr\left[\max_i\sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left|\int\frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}dF_i(\epsilon) - E\left[\frac{\partial v_i(\cdot,\epsilon)}{\partial\alpha_i}\right]\right| > \eta\right] &= o\left(T^{-1}\right).\end{aligned}$$

By Lemma (9) again, it follows that

$$\Pr\left[\max_i \sup_{\epsilon\in\left[0,1/\sqrt{T}\right]}\left|\int v_i(\cdot,\epsilon)\,d\Delta_{iT}\right| > T^{1/10-v}\right] = o\left(T^{-1}\right).$$

This proves the result for $\widehat{\alpha}_i^{\epsilon}(\epsilon)$. For $\widehat{\alpha}_i(0)$ the result follows directly from (41) and Lemmas (9) and (11). For $\widehat{\alpha}_i^{\epsilon\epsilon}(\epsilon)$, the result follows from representation (43) as well as Lemmas (9), (11), and (13). ■

D.3 Consistent Covariance Matrix Estimation for Dynamic Panel Models

Lemma 16 *Let $k_{it} = k(x_{it};\theta_0,\alpha_{i0})$ and $\widehat{k}_{it} = k\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right)$ where x_{it} satisfies Condition (2), k_{it} satisfies Condition (3) and $\widehat{\theta}$, $\widehat{\alpha}_i$ are defined in (1). Assume that $E[k_{it}] = 0$ for i,t. Let A_i be conformable matrix of constants such that $\max_i\|A_i\| < \infty$. Let $f_i^{kk} = \sum_{l=-\infty}^{\infty} E\left[k_{it}k_{it-l}'\right]$ and $f^{kk} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n A_i f_i^{kk}$. Then,*

$$\frac{1}{n}\sum_{i=1}^n A_i\left(\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l)}^{\min(T,T+l)}\widehat{k}_{it}\widehat{k}_{it-l}'\right) - f^{kk} = o_p(1)$$

where $m,T\to\infty$ such that $m = o\left(T^{1/2}\right)$.

Proof. Let $r_1 = \max(1,l)$ and $r_2 = \min(T,T+l)$ and define $K_{i,m} = \frac{1}{T}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2} k_{it}k_{it-l}'$. We first show that $\frac{1}{n}\sum_{i=1}^n A_iK_{i,m} - f^{kk} = o_p(1)$. This follows if $\frac{1}{n}\sum_{i=1}^n A_iE(K_{i,m}) - f^{kk} = o(1)$ and $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n A_iK_{i,m}\right) = o(1)$. Since $f^{kk} - n^{-1}\sum_{i=1}^n A_if_i^{kk} = o(1)$ by definition, we first consider

$$\begin{aligned}
&\left\|E[K_{i,m}] - f_i^{kk}\right\| \\
\leq\ & \textstyle\sum_{l=-m}^{m}\left|\frac{r_2-r_1}{T}-1\right|\left\|E\left[k_{it}k_{it-l}'\right]\right\| + \sum_{|l|>m}\left\|E\left[k_{it}k_{it-l}'\right]\right\| \\
\leq\ & \textstyle\sum_{l=-m}^{m}\frac{2|l|}{T}\left\|E\left[k_{it}k_{it-l}'\right]\right\| + \sum_{|l|>m}\left\|E\left[k_{it}k_{it-l}'\right]\right\| \\
\leq\ & \textstyle\sum_{l=-m}^{m}\frac{c_1|l|}{T}\left(a^{\frac{\delta}{2+\delta}}\right)^{|l|} + \left(a^{\frac{\delta}{2+\delta}}\right)^{m} c_2\sum_{l=1}\left(a^{\frac{\delta}{2+\delta}}\right)^{l} \to 0 \text{ as } m,T\to\infty
\end{aligned}$$

where the last inequality follows from Condition (2) and the fact that for any two elements k_{it,j_1} and k_{it-l,j_2} of k_{it} and k_{it-l} it follows from Corollary A.2 of Hall and Heyde (1980) that $|E[k_{it,j_1}k_{it-l,j_2}]| \leq 8\left(E\left[|k_{it,j_1}|^{2+\delta}\right]\right)^{\frac{1}{2+\delta}}\left(E\left[|k_{it-l,j_2}|^{2+\delta}\right]\right)^{\frac{1}{2+\delta}}\left(a^{\frac{\delta}{2+\delta}}\right)^{|l|}$ for some $\delta > 0$. Since the bound on $\|A_i\|\left\|E[K_{i,m}] - f_i^{kk}\right\|$ is uniform it therefore follows that $\frac{1}{n}\sum_{i=1}^n A_iE[K_{i,m}] - f^{kk} = o(1)$. Next we show that

$$\begin{aligned}
\left\|\mathrm{Var}\left(\frac{1}{n}\textstyle\sum_{i=1}^n A_iK_{i,m}\right)\right\| &\leq \frac{1}{n^2}\textstyle\sum_{i=1}^n\|A_i\|^2\left\|\mathrm{Var}(K_{i,m})\right\| \\
&\leq \sup\|A_i\|\frac{1}{n^2}\textstyle\sum_{i=1}^n\left\|\mathrm{Var}(K_{i,m})\right\| = o(1).
\end{aligned}$$

To show this we may assume without loss of generality that k_{it} is scalar. The variance can then be evaluated as

$$\begin{aligned}
& \operatorname{Var}(K_{i,m}) \\
= \; & \frac{1}{T^2}\sum\nolimits_{l_1,l_2=-m}^{m}\sum\nolimits_{t_1,t_2=r_1}^{r_2}\left(E\left[k_{it_1}k_{it-l_1}k_{it_2}k_{it_2-l_2}\right]-E\left[k_{it_1}k_{it-l_1}Ek_{it_2}k_{it_2-l_2}\right]\right) \\
= \; & \frac{1}{T^2}\sum\nolimits_{l_1,l_2=-m}^{m}\sum\nolimits_{t_1,t_2=r_1}^{r_2}\left(E\left[k_{it_1}k_{it_2}Ek_{it-l_1}k_{it_2-l_2}\right]+E\left[k_{it_1}k_{it_2-l_2}Ek_{it_2}k_{it-l_1}\right]\right) \\
& +\frac{1}{T^2}\sum\nolimits_{l_1,l_2=-m}^{m}\sum\nolimits_{t_1,t_2=r_1}^{r_2}\operatorname{Cum}\left(k_{it_1}k_{it-l_1}k_{it_2}k_{it_2-l_2}\right) \\
= \; & O(1)
\end{aligned}$$

where the last equality follows from the same arguments as in the proof of Lemma (6) such that $\operatorname{Var}(K_{i,m})$ is uniformly bounded in i. It now follows that $\frac{1}{n}\sum_{i=1}^{n}K_{i,m}-f^{kk}=o_p(1)$ by Markov's inequality.

Next we turn to showing that

$$\frac{1}{n}\sum\nolimits_{i=1}^{n}A_i\frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\left(\widehat{k}_{it}\widehat{k}_{it-l}'-k_{it}k_{it-l}'\right)=o_p(1).$$

We use the decomposition

$$\begin{aligned}
& \frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\left(\widehat{k}_{it}\widehat{k}_{it-l}'-k_{it}k_{it-l}'\right) \\
= \; & \frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\left(\widehat{k}_{it}-k_{it}\right)\left(\widehat{k}_{it-l}-k_{it-l}\right)' \\
& +\frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}k_{it}\left(\widehat{k}_{it-l}-k_{it-l}\right)'+\frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\left(\widehat{k}_{it}-k_{it}\right)k_{it-l}'
\end{aligned}$$

and consider the term $\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left(\widehat{k}_{it}-k_{it}\right)k_{is}'$. Use a first order Taylor approximation to

$$\widehat{k}_{it}-k_{it}=k_{it}^{\theta}\left(\widehat{\theta}-\theta\right)+k_{it}^{\alpha}\left(\widehat{\alpha}_i-\alpha_{i0}\right)$$

where $k_{it}^{\theta}=\partial k\left(x_{it};\tilde{\theta},\tilde{\alpha}_i\right)/\partial\theta'$ and $k_{it}^{\alpha}=\partial k\left(x_{it};\tilde{\theta}',\tilde{\alpha}_i'\right)/\partial\alpha$ with $\tilde{\theta},\tilde{\alpha}_i,\tilde{\theta}',\tilde{\alpha}_i'$ such that $\left\|\tilde{\theta}-\theta_0\right\|\leq\left\|\widehat{\theta}-\theta_0\right\|$, $\left\|\tilde{\theta}'-\theta_0\right\|\leq\left\|\widehat{\theta}-\theta_0\right\|$, etc. by the multivariate version of the mean value theorem. Note that each row of $\partial k\left(x_{it};\tilde{\theta},\tilde{\alpha}_i\right)/\partial\theta'$ needs to be evaluated at a different $\tilde{\theta}$ but in slight abuse of notation we do not make this explicit. Then

$$\begin{aligned}
& \frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\operatorname{vec}\left[A_i\left(\widehat{k}_{it}-k_{it}\right)k_{it-1}'\right] \\
= \; & \frac{1}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\left(I\otimes A_i\right)\left(k_{it-l}\otimes k_{it}^{\theta}\right)\left(\widehat{\theta}-\theta\right) \\
& +\frac{\left(\widehat{\alpha}_i-\alpha_{i0}\right)}{T}\sum\nolimits_{l=-m}^{m}\sum\nolimits_{t=r_1}^{r_2}\left(I\otimes A_i\right)\operatorname{vec}\left[k_{it}^{\alpha}k_{it-l}'\right] \qquad (44)
\end{aligned}$$

and consider $\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left(k_{it-l}\otimes k_{it}^{\theta}\right)$. Without loss of generality assume that $\left(k_{it-l}\otimes k_{it}^{\theta}\right)$ is a scalar. Then by the Cauchy-Schwartz inequality

$$\begin{aligned}
\left|\frac{1}{T}\sum\nolimits_{t=r_1}^{r_2}k_{it-l}k_{it}^{\theta}\right| \; & \leq \; \left(\frac{1}{T}\sum\nolimits_{t=1}^{T}k_{it-l}^2\right)^{1/2}\left(\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta,\alpha}\left(\partial k\left(x_{it};\theta,\alpha\right)/\partial\theta'\right)^2\right)^{1/2} \\
& \leq \; \left(\frac{1}{T}\sum\nolimits_{t=1}^{T}M(x_{it-l})^2\right)^{1/2}\left(\frac{1}{T}\sum\nolimits_{t=1}^{T}M(x_{it})^2\right)^{1/2}
\end{aligned}$$

such that $E\left|\frac{1}{T}\sum_{t=r_1}^{r_2} k_{it-l}k_{it}^{\theta}\right| \leq \left(\frac{1}{T}\sum_{t=1}^{T} E\left[M\left(x_{it-l}\right)^2\right]\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T} E\left[M\left(x_{it}\right)^2\right]\right)^{1/2} = O(1)$ uniformly in i. It thus follows from the Markov inequality that

$$\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left(k_{it-l}\otimes k_{it}^{\theta}\right)\left(\widehat{\theta}-\theta\right) = O_p(m/T).$$

We now turn to the second term in (44) where

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\left(\widehat{\alpha}_i-\alpha_{i0}\right)}{T}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\operatorname{vec}\left[k_{it}^{\alpha}k_{it-l}'\right] \qquad (45)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\alpha}_i^{\epsilon}(0)}{T^{3/2}}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes k_{it}^{\alpha}\right]\\
&+\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon})}{T^{2}}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes k_{it}^{\alpha}\right].
\end{aligned}$$

Define $k_{it}^{\alpha}(\theta,\alpha) = \partial k\left(x_{it};\theta,\alpha\right)/\partial\alpha$ and $\bar{k}_{it}^{\alpha}(\theta,\alpha) = k_{it}^{\alpha}(\theta,\alpha) - E\left[k_{it}^{\alpha}(\theta,\alpha)\right]$ and consider

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\alpha}_i^{\epsilon}(0)}{T^{3/2}}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes k_{it}^{\alpha}\right]\right\|\\
\leq\;&\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-m}^{m}\left\|\frac{1}{T^{3/2}}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right]\right\|\left|\widehat{\alpha}_i^{\epsilon}(0)\right|\left\|I\otimes A_i\right\|\\
&+\left\|\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\alpha}_i^{\epsilon}(0)}{T^{3/2}}\sum_{t=1}^{T}\left(m+\min(\min(t,m),\min(T-t,m))\right)\left[k_{it-l}\otimes E\left[k_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right]\right]\right\|.
\end{aligned}$$

For the first term we assume without loss of generality that $k_{it-l}\otimes\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)$ is scalar such that

$$\left|\frac{1}{T}\sum_{t=r_1}^{r_2} k_{it-l}\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right| \leq \frac{1}{T}\sum_{t=r_1}^{r_2}\left|k_{it-l}\right|\left|\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right| \leq \left(\frac{1}{T}\sum_{t=r_1}^{r_2}\left|k_{it-l}\right|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=r_1}^{r_2}\left|\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right|^2\right)^{1/2}$$

by the Hölder inequality. Then

$$\begin{aligned}
E\left[\left|\frac{1}{T}\sum_{t=r_1}^{r_2} k_{it-l}\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right|^2\right] &\leq \left(E\left[\left(\frac{1}{T}\sum_{t=r_1}^{r_2} k_{it-l}^2\right)^2\right]\right)^{1/2}\left(E\left[\left(\frac{1}{T}\sum_{t=r_1}^{r_2}\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)^2\right)^2\right]\right)^{1/2}\\
&\leq \left(E\left[\left(\frac{1}{T}\sum_{t=r_1}^{r_2} M(x_{it-l})^2\right)^2\right]\right)^{1/2}\left(E\left[\left(\frac{1}{T}\sum_{t=r_1}^{r_2} M(x_{it})^2\right)^2\right]\right)^{1/2}
\end{aligned}$$

which is uniformly bounded in i.By (41) it follows that $E\left[\left|\widehat{\alpha}_i^{\epsilon}(0)\right|^2\right]$ is bounded uniformly in i. From the Markov inequality it then follows that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-m}^{m}\left\|\frac{1}{T^{3/2}}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes\bar{k}_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right]\right\|\left|\widehat{\alpha}_i^{\epsilon}(0)\right|\left\|I\otimes A_i\right\| = O_p\left(m/T^{1/2}\right).$$

For the second term let $m_t = (m+\min(\min(t,m),\min(T-t,m))$ and again assume that $k_{it-l}\otimes k_{it}^{\alpha}$ is scalar. Then

$$\begin{aligned}
&E\left|\frac{1}{T^{3/2}}\sum_{t=1}^{T} m_t\left[k_{it-l}E\left[k_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right]\right]\right|^2\\
&\qquad\leq \frac{1}{T^3}\sum_{s,t=1}^{T} m_t m_s\left|E\left[k_{it}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right]\right|\left|E\left[k_{is}^{\alpha}\left(\tilde{\theta}',\tilde{\alpha}'\right)\right]\right|\left|E\left[k_{it}k_{is}\right]\right| = O(m^2/T^2)
\end{aligned}$$

by the mixing inequality in Hall and Heyde (1980). This establishes that

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\alpha}_i^{\epsilon}(0)}{T^{3/2}}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes k_{it}^{\alpha}\right]\right\| = O_p(m/T^{1/2}).$$

For the second term in (45) use the bound

$$\left\| \frac{1}{n} \sum_{i=1}^{n} (I \otimes A_i) \frac{\widehat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon})}{T^2} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left[k_{it-l} \otimes k_{it}^{\alpha} \right] \right\|$$
$$\leq \sup_i \|I \otimes A_i\| \max_i \sup_{\epsilon \in [0,1/\sqrt{T}]} \left| \frac{\widehat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon})}{T^{1/5-2v}} \right| \frac{1}{nT^{4/5+2v}} \sum_{i=1}^{n} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \left[k_{it-l} \otimes k_{it}^{\alpha} \right] \right\|$$

where $E \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \left[k_{it-l} \otimes k_{it}^{\alpha} \right] \right\| = O(1)$ by previous arguments and $\max_i \sup_{\epsilon \in [0,1/\sqrt{T}]} \left| \frac{\widehat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon})}{T^{1/5-2v}} \right| = o_p(1)$ by Lemma (15). It follows that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} (I \otimes A_i) \frac{\widehat{\alpha}_i^{\epsilon\epsilon}(\tilde{\epsilon})}{T^2} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left[k_{it-l} \otimes k_{it}^{\alpha} \right] \right\| = o_p(m/T^{1/2}).$$

We now turn to

$$\frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(\widehat{k}_{it} - k_{it}\right)\left(\widehat{k}_{it-l} - k_{it-l}\right)'$$
$$= \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left(k_{it-l}^{\theta} \otimes k_{it}^{\theta}\right) \operatorname{vec}\left(\widehat{\theta} - \theta\right)\left(\widehat{\theta} - \theta\right)' + \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left(k_{it-l}^{\theta} \otimes k_{it}^{\alpha}\right) (\widehat{\alpha}_i - \alpha_{i0}) \operatorname{vec}\left(\widehat{\theta} - \theta\right)'$$
$$+ \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left(k_{it-l}^{\alpha} \otimes k_{it}^{\theta}\right)\left(\widehat{\theta} - \theta\right)(\widehat{\alpha}_i - \alpha_{i0}) + \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} (\widehat{\alpha}_i - \alpha_{i0})^2 \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right)$$

such that

$$\begin{aligned}
& \left\| \frac{1}{n} \sum_{i=1}^{n} (I \otimes A_i) \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} (\widehat{\alpha}_i - \alpha_{i0})^2 \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right) \right\| \\
\leq \; & \sup_i \|I \otimes A_i\| \max_i |\widehat{\alpha}_i - \alpha_{i0}| \frac{1}{n} \sum_{i=1}^{n} |\widehat{\alpha}_i - \alpha_{i0}| \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right) \right\| \\
\leq \; & \sup_i \|I \otimes A_i\| \max_i |\widehat{\alpha}_i - \alpha_{i0}| \frac{1}{n} \sum_{i=1}^{n} \frac{|\widehat{\alpha}_i^{\epsilon}(0)|}{\sqrt{T}} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right) \right\| \\
& + \sup_i \|I \otimes A_i\| \max_i |\widehat{\alpha}_i - \alpha_{i0}| \max_i \sup_{\epsilon \in [0,1/\sqrt{T}]} |\widehat{\alpha}_i^{\epsilon\epsilon}(\epsilon)| \frac{1}{2nT} \sum_{i=1}^{n} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right) \right\| \\
= \; & o_p(m/T^{1/2})
\end{aligned}$$

where by the same arguments as before $|\widehat{\alpha}_i^{\epsilon}(0)|$ is uniformly bounded and $\left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right) \right\| = O(1)$ uniformly in i such that the first term is $o_p(m/T^{1/2})$ by the Markov inequality and the fact that $\max_i |\widehat{\alpha}_i - \alpha_{i0}| = o_p(1)$. The second term is $o_p(m/T^{1/2})$ because $\max_i \sup_{\epsilon \in [0,1/\sqrt{T}]} |\widehat{\alpha}_i^{\epsilon\epsilon}(\epsilon)| = O_p(T^{\frac{1}{5}-2v})$ by Lemma (15) and $\frac{1}{nT} \sum_{i=1}^{n} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\alpha} k_{it-l}^{\alpha'}\right) \right\| = O_p(mT^{-1})$ by the Markov inequality. All the remaining terms in $\frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(\widehat{k}_{it} - k_{it}\right)\left(\widehat{k}_{it-l} - k_{it-l}\right)'$ are $o_p\left(m/T^{1/2}\right)$ by similar arguments. ■

Proof of Theorem (3). We only need to show that $\widehat{\beta}-\beta=o_p(1)$. First consider $\widehat{E}\left[V_i^{\alpha_i}\right]=\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right)$. We have

$$\begin{aligned}\left\|\widehat{E}\left[V_i^{\alpha_i}\right]-E\left[V_i^{\alpha_i}\right]\right\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\alpha_i}\left(x_{it};\widehat{\theta},\widehat{\alpha}_i\right)-\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)\right\| \\ &\quad+\left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)-E\left[V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)\right]\right\| \\ &\leq \left(\frac{1}{T}\sum_{t=1}^{T}\|M(x_{it})\|\right)\left(\left\|\widehat{\theta}-\theta\right\|+\max_i\left|\widehat{\alpha}_i-\alpha_{i0}\right|\right) \\ &\quad+\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)-E\left[V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)\right]\right\|\end{aligned}$$

so that

$$\begin{aligned}\max_i\left\|\widehat{E}\left[V_i^{\alpha_i}\right]-E\left[V_i^{\alpha_i}\right]\right\| &\leq \left(\max_i\frac{1}{T}\sum_{t=1}^{T}\|M(x_{it})\|\right)\left(\left\|\widehat{\theta}-\theta\right\|+\max_i\left|\widehat{\alpha}_i-\alpha_{i0}\right|\right) \\ &\quad+\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)-E\left[V_{it}^{\alpha_i}\left(x_{it};\theta_0,\alpha_{i0}\right)\right]\right\|\end{aligned}$$

By Lemma (5) the second term tends to zero with probability $1-o(T^{-1})$. Applying Lemmas (4) and (5) to the first term, we obtain

$$\max_i\left\|\widehat{E}\left[V_i^{\alpha_i}\right]-E\left[V_i^{\alpha_i}\right]\right\|=o_p\left(1\right)$$

In the same way, we obtain

$$\begin{aligned}\max_i\left\|\widehat{E}\left[V_i^{\theta}\right]-E\left[V_i^{\theta}\right]\right\| &= o_p\left(1\right) \\ \max_i\left\|\widehat{E}\left[U_i^{\alpha_i\alpha_i}\right]-E\left[U_i^{\alpha_i\alpha_i}\right]\right\| &= o_p\left(1\right)\end{aligned}$$

Let $\widehat{E}\left[V_i^{\alpha_i}\right]=E\left[V_i^{\alpha_i}\right]+o_p(1)$, which holds uniformly in i. Thus,

$$\max_i\left\|\widehat{\mathcal{I}}_i-\mathcal{I}_i\right\|\leq\sup_i E\left[\|M(x_{it})\|\right]\left(\left\|\widehat{\theta}-\theta\right\|+\max_i\left|\widehat{\alpha}_i-\alpha_{i0}\right|\right)+o_p(1).$$

Since $\left|\widehat{\alpha}_i-\alpha_{i0}\right|\leq\frac{1}{\sqrt{T}}\left|\widehat{\alpha}_i^{\epsilon}\left(0\right)\right|+\frac{1}{2T}\left|\widehat{\alpha}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)\right|$ with $\max_i T^{-\frac{1}{10}}\left|\widehat{\alpha}_i^{\epsilon}\left(0\right)\right|=o_p(1)$ and $\max_i T^{-\frac{2}{10}}\left|\widehat{\alpha}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)\right|=o_p(1)$ by Lemma (15), it follows that $\max_i\left\|\widehat{\mathcal{I}}_i-\mathcal{I}_i\right\|=o_p(1)$ such that $n^{-1}\sum_{i=1}^{n}\widehat{\mathcal{I}}_i-\mathcal{I}=o_p(1)$.

Using these results we now have

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{f}_i^{VU^{\alpha}}}{\widehat{E}\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}-\frac{\widehat{f}_i^{VU^{\alpha}}}{E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}\right]=o_p(1)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{E}\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]\widehat{f}_i^{VV}}{2\left(\widehat{E}\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]\right)^2}-\frac{E\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]\widehat{f}_i^{VV}}{2\left(E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]\right)^2}\right]=o_p(1).$$

In order to establish the result we thus need to apply Lemma (16) to show that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{E\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]\left(\widehat{f}_i^{VU^\alpha}-f_i^{VU^\alpha}\right)}{E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]}\right]=o_p(1)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\widehat{f}_i^{VV}-f_i^{VV}}{\left(E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]\right)^2}\right]=o_p(1).$$

The result follows by Lemma (16) since $\inf\left|E\left[\frac{\partial V_i(x_{it};\theta,\alpha_i)}{\partial\alpha_i}\right]\right|>0$ by Condition (5) and $\left\|E\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]\right\|\leq E\left[M\left(x_{it}\right)\right]<\infty$. ■

E Linear Dynamic Panel Model with Fixed Effects

It can be shown that

$$\begin{aligned}
s_\theta\left(x_{it};\theta,\alpha_i\right) &= 2y_{it-1}\left(y_{it}-\alpha_i-\theta y_{it-1}\right)\\
s_{\theta\alpha}\left(x_{it};\theta,\alpha_i\right) &= -2y_{it-1}\\
s_\alpha\left(x_{it};\theta,\alpha_i\right) &= 2\left(y_{it}-\alpha_i-\theta y_{it-1}\right)\\
s_{\alpha\alpha}\left(x_{it};\theta,\alpha_i\right) &= -2\\
\rho_{i0} &= \frac{E\left[-2y_{it-1}\right]}{-2}=E\left[y_{it-1}\right]\\
V_{it} &= 2\left(y_{it}-\alpha_i-\theta y_{it-1}\right)=2\varepsilon_{it}\\
V_{it}^\alpha &= -2\\
U_i\left(x_{it};\theta,\alpha_i\right) &= 2y_{it-1}\left(y_{it}-\alpha_i-\theta y_{it-1}\right)-\rho_{i0}\cdot 2\left(y_{it}-\alpha_i-\theta y_{it-1}\right)\\
&= 2\left(y_{it-1}-\rho_{i0}\right)\left(y_{it}-\alpha_i-\theta y_{it-1}\right)\\
&= 2\left(y_{it-1}-\rho_{i0}\right)\varepsilon_{it}\\
U_i^\theta\left(x_{it};\theta,\alpha_i\right) &= -2y_{it-1}^2+2\rho_{i0}y_{it-1}=-2y_{it-1}\left(y_{it-1}-\rho_{i0}\right)\\
U_i^\alpha\left(x_{it};\theta,\alpha_i\right) &= -2\left(y_{it-1}-\rho_{i0}\right)\\
U_i^{\alpha\alpha}\left(x_{it};\theta,\alpha_i\right) &= 0\\
V_i\left(x_{it};\theta,\alpha_i\right)U_i^{\alpha_i}\left(x_{it};\theta,\alpha_i\right) &= -4\varepsilon_{it}\left(y_{it-1}-\rho_{i0}\right)
\end{aligned}$$

We can see that

$$\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_i\left(x_{it};\theta,\alpha_i\right)=\frac{1}{\sqrt{n}\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}2\left(y_{it-1}-\rho_{i0}\right)\varepsilon_{it}\rightarrow N\left(0,\frac{4\left(\sigma^2\right)^2}{1-\theta^2}\right)\equiv N\left(0,\Omega\right)$$

$$
\begin{aligned}
\Psi &= \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{V_{it}\left(x_{it};\theta,\alpha_i\right)}{E\left[\frac{\partial V_i\left(x_{it};\theta,\alpha_i\right)}{\partial \alpha_i}\right]} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(U_{it}^{\alpha_i}\left(x_{it};\theta,\alpha_i\right) - \frac{E\left[U_i^{\alpha_i\alpha_i}\left(x_{it};\theta,\alpha_i\right)\right]}{2E\left[\frac{\partial V_i\left(x_{it};\theta,\alpha_i\right)}{\partial \alpha_i}\right]} V_{it}\left(x_{it};\theta,\alpha_i\right) \right) \right] \\
&= \operatorname{plim} 2 \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(y_{it-1} - E\left[y_{it-1}\right]\right) \right] \right\} \\
&= 2\frac{\sigma^2}{1-\theta}
\end{aligned}
$$

and

$$
\mathcal{I} = -E\left[U_i^{\theta}\left(x_{it};\theta,\alpha_i\right)\right] = 2E\left[y_{it-1}\left(y_{it-1}-\rho_{i0}\right)\right] = 2E\left[\left(y_{it-1}-\rho_{i0}\right)^2\right] = \frac{2\sigma^2}{1-\theta^2}
$$

It follows that

$$
\begin{aligned}
\beta &\equiv -\mathcal{I}^{-1}\Psi = -\frac{2\sqrt{\kappa}\left(1-\theta\right)\sigma^2}{\frac{2\sigma^2}{1-\theta^2}} = -\sqrt{\kappa}\left(1+\theta\right) \\
\mathcal{I}^{-1}\Omega\mathcal{I}^{-1} &= \frac{\frac{4\left(\sigma^2\right)^2}{1-\theta^2}}{\left(\frac{2\sigma^2}{1-\theta^2}\right)^2} = 1-\theta^2
\end{aligned}
$$

and hence

$$
\sqrt{nT}\left(\widehat{\theta}-\theta_0\right) \to N\left(-\sqrt{\kappa}\left(1+\theta\right), 1-\theta^2\right)
$$

References

[1] Andrews, D. W. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

[2] Arellano, M. (2000): "Discrete Choices with Panel Data", Investigaciones Económicas Lecture, XXV Simposio de Análisis Económico, Bellaterra.

[3] Billingsley, P. (1986): *Probability and Measure*. John Wiley and Sons, New York.

[4] Hahn, J. and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", *Econometrica,* 70, 1639-1657.

[5] Hahn, J., G. Kuersteiner, and W.K. Newey (2002): "Higher Order Properties of Bootstrap and Jackknife Bias Corrected Maximum Likelihood Estimators", *unpublished manuscript.*

[6] Hahn, J. and W.K. Newey (2002): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *unpublished manuscript.*

[7] Hall, P., and C. Heyde (1980): *Martingale Limit Theory and its Application*. Academic Press.

[8] Hall, P. and J. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators", *Econometrica* 64, 891-916.

[9] Honore, B., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica* 68, pp. 839 - 874.

[10] Lahiri, S. (1992): "Edgeworth Correction by Moving Block Bootstrap for Stationary and Nonstationary Data," in *Exploring the Limits of Bootstrap*, ed. by R. LePage, and L. Billard, pp. 183–214. Wiley.

[11] Lancaster, T. (1997): "Orthogonal Parameters and Panel Data", *Review of Economic Studies* 69, 647–666.

[12] Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

[13] Neyman, J., and E. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations", *Econometrica* 16, pp. 1 -31.

[14] Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer Verlag.

[15] Woutersen, T. (2002): "Robustness against Incidental Parameters", *unpublished manuscript.*