Supplementary Appendix to "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects"

Manuel Arellano and Jinyong Hahn

A Some Auxiliary Lemmas

Throughout this appendix, we will let $F \equiv (F_1, \ldots, F_n)$ denote the collection of (marginal) distribution functions of x_{it} and $\widehat{F} \equiv \left(\widehat{F}_1, \ldots, \widehat{F}_n\right)$, where $\widehat{F}_i$ denotes the empirical distribution function for the i-th observation. Define $F(\epsilon) \equiv F + \epsilon\sqrt{T}\left(\widehat{F} - F\right)$ for $\epsilon \in \left[0, T^{-1/2}\right]$, and $\Delta_{iT} \equiv \sqrt{T}\left(\widehat{F}_i - F_i\right)$. We first provide a different version of Lahiri's (1992) Lemma 5.1, which is stated for bounded zero mean random variables.

Lemma 1 (Hahn and Kuersteiner, 2004) *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* $\mathcal{A}_t = \sigma(W_t, W_{t-1}, W_{t-2}, \ldots)$, $\mathcal{B}_t = \sigma(W_t, W_{t+1}, W_{t+2}, \ldots)$, *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.

Lemma 2 (Hahn and Kuersteiner, 2004) *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* $\mathcal{A}_t^i = \sigma\left(\xi_{it}, \xi_{it-1}, \xi_{it-2}, \ldots\right)$, $\mathcal{B}_t^i = \sigma\left(\xi_{it}, \xi_{it+1}, \xi_{it+2}, \ldots\right)$, *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* $\{\xi_{it}, t = 1, 2, 3, \ldots\}$ *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* $0 < \upsilon < (100q + 120)^{-1}$.

Lemma 3 (Hahn and Kuersteiner, 2004) *Let* $\xi(x_{it}, \phi)$ *be 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* $\mathbf{M}(x_{it})$ *such that* $|\xi(x_{it}, \phi_1) - \xi(x_{it}, \phi_2)| \leq \mathbf{M}(x_{it})\|\phi_1 - \phi_2\|$ *for all* $\phi_1, \phi_2 \in \Phi$ *and* $\sup_\phi |\xi(x_{it}, \phi)| \leq \mathbf{M}(x_{it})$. *For each* i, *let* x_{it} *be a* α*-mixing process with exponentially decaying mixing coefficients* $\underline{\alpha}_i(m)$ *satisfying* $\sup_i |\underline{\alpha}_i(m)| \leq Ca^m$ *for some* a *such that* $0 < a < 1$ *and some* $0 < C < \infty$. *Let* q *denote a positive integer such that* $q \geq \frac{p+4}{2}$, *where* $p = \dim\phi$. *We also assume that* $E\left[|\mathbf{M}(x_{it})|^{10q+12+\delta}\right] < \infty$ *for some* $\delta > 0$. *Finally, assume that* $n = O(T)$. *We then have*

$\Pr\left[\max_i \left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)\right| > T^{\frac{1}{10}-v}\right] = o\left(T^{-1}\right)$ *for* $0 < v < (100q+120)^{-1}$. *Here,* $\{\phi_i\}$ *is an arbitrary nonstochastic sequence in* Φ.

Lemma 4 (Hahn and Kuersteiner, 2004) *Assume that* x_{it} *satisfies Assumption 3, 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* $E\left[\xi\left(x_{it},\phi_i\right)\right]=0$. *Further assume that* $\sup_\phi \|\xi\left(x_{it},\phi\right)\| \leq \mathbf{M}\left(x_{it}\right)$ *for some* $\mathbf{M}\left(x_{it}\right)$ *such that* $E\left[\mathbf{M}\left(x_{it}\right)^4\right] < \infty$. *Let* $\Sigma_{nT} = \sum_{i=1}^{n}\Sigma_{iT}^{\xi\xi}$ *with* $\Sigma_{iT}^{\xi\xi} = \operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right)\right)$. *Denote the smallest eigenvalue of* $\Sigma_{iT}^{\xi\xi}$ *by* λ_{iT}^{ξ}, *and assume that* $\inf_i \inf_T \lambda_{iT}^{\xi} > 0$. *Then,*

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\xi\left(x_{it},\phi_i\right) \Rightarrow N\left(0,f^{\xi\xi}\right), \quad \text{and} \quad \sup_i \left\|\Sigma_{iT}^{\xi\xi} - f_i^{\xi\xi}\right\| \to 0,$$

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

Lemma 5 *Let* $k_{it} = k\left(x_{it};\theta,\gamma_i(\theta)\right)$ *and* $\widehat{k}_{it} = k\left(x_{it};\theta,\widehat{\gamma}_i(\theta)\right)$ *where* x_{it} *satisfies Assumption 3,* k *satisfies Assumption 4 and* $\widehat{\theta}$, $\widehat{\gamma}_i$ *are defined in (1). Assume that* $E\left[k_{it}\right]=0$ *for* i,t. *Let* $f_i^{kk} \equiv \sum_{l=-\infty}^{\infty} E\left[k_{it}k_{it-l}'\right]$ *and* $f^{kk} \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} f_i^{kk}$. *Then,*

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

where $m,T \to \infty$ *such that* $m = o\left(T^{2/5}\right)$.

Proof. The proof is almost identical to a similar result found in Hahn and Kuersteiner (2004). Let $r_1 = \max(1,l)$ and $r_2 = \min(T,T+l)$ and define $K_{i,m} = \frac{1}{T}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2} k_{it}k_{it-l}'$.

We first show that $\frac{1}{n}\sum_{i=1}^{n} K_{i,m} - f^{kk} = o_p(1)$. This follows if $\frac{1}{n}\sum_{i=1}^{n} E\left[K_{i,m}\right] - f^{kk} = o(1)$ and $\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} K_{i,m}\right) = o(1)$. Since $f^{kk} - n^{-1}\sum_{i=1}^{n} f_i^{kk} = o(1)$ by definition, we first consider

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

where the last inequality follows from Condition 3 and the fact that

$$|E[k_{it,j_1}k_{it-l,j_2}]| \le 8\left(E\left[|k_{it,j_1}|^{2+\delta}\right]\right)^{\frac{1}{2+\delta}}\left(E\left[|k_{it-l,j_2}|^{2+\delta}\right]\right)^{\frac{1}{2+\delta}}\left(a^{\frac{\delta}{2+\delta}}\right)^{|l|}$$

for any two elements $k_{it,j_1}$and k_{it-l,j_2} of k_{it} and k_{it-l} for some $\delta > 0$, which can be proved by Corollary A.2 of Hall and Heyde (1980). Since the bound on $\left\|E[K_{i,m}] - f_i^{kk}\right\|$ is uniform it therefore follows that $\frac{1}{n}\sum_{i=1}^{n} E[K_{i,m}] - f^{kk} = o(1)$.

Next we show that

$$\left\|\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} K_{i,m}\right)\right\| \le \frac{1}{n^2}\sum_{i=1}^{n}\|\text{Var}(K_{i,m})\| = o(1).$$

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

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

such that $\text{Var}(K_{i,m})$ 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

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

We use the decomposition

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

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

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

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

$$\begin{aligned}
& \frac{1}{T}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2}\operatorname{vec}\left[\left(\widehat{k}_{it}-k_{it}\right)k_{it-1}'\right] \qquad (34)\\
= & \frac{1}{T}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2}\left(k_{it-l}\otimes k_{it}^{\theta}\right)\left(\widehat{\theta}-\theta\right)\\
& +\frac{\left(\widehat{\gamma}_i-\gamma_{i0}\right)}{T}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2}\operatorname{vec}\left[k_{it}^{\gamma}k_{it-l}'\right]
\end{aligned}$$

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

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

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

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

We now turn to the second term in (34). Noting that

$$T^{2/5}\max_i\left|\widehat{\gamma}_i-\gamma_{i0}\right|=o_p\left(1\right)$$

by Lemma (7), we obtain

$$\begin{aligned}
& \left|\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\widehat{\gamma}_i-\gamma_{i0}\right)}{T}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2}\operatorname{vec}\left[k_{it}^{\gamma}k_{it-l}'\right]\right|\\
\leq & \frac{1}{T^{7/5}}T^{2/5}\max_i\left|\widehat{\gamma}_i-\gamma_{i0}\right|\cdot\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2}\left\|\operatorname{vec}\left[k_{it}^{\gamma}k_{it-l}'\right]\right\|\\
\leq & o_p\left(T^{-7/5}\right)\cdot\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-m}^{m} w_{T,l}\sum_{t=r_1}^{r_2}\operatorname{vec}\left[M_{it}M_{it-l}'\right]\\
= & o_p\left(T^{-7/5}\right)\sum_{l=-m}^{m}\left(1-\frac{|l|}{m+1}\right)\left(T-|l|\right)\\
= & o_p\left(T^{-7/5}\right)O\left(Tm\right)\\
= & o_p\left(\frac{m}{T^{2/5}}\right)
\end{aligned}$$

We now turn to

$$
\begin{aligned}
& \frac{1}{T} \sum_{l=-m}^{m} w_{T,l} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(\widehat{k}_{it}-k_{it}\right)\left(\widehat{k}_{it-l}-k_{it-l}\right)' \\
= & \frac{1}{T} \sum_{l=-m}^{m} w_{T,l} \sum_{t=r_1}^{r_2}\left(k_{it-l}^{\theta} \otimes k_{it}^{\theta}\right) \operatorname{vec}\left(\widehat{\theta}-\theta\right)\left(\widehat{\theta}-\theta\right)' \\
& +\frac{1}{T} \sum_{l=-m}^{m} w_{T,l} \sum_{t=r_1}^{r_2}\left(k_{it-l}^{\theta} \otimes k_{it}^{\gamma}\right)\left(\widehat{\gamma}_i-\gamma_{i0}\right) \operatorname{vec}\left(\widehat{\theta}-\theta\right)' \\
& +\frac{1}{T} \sum_{l=-m}^{m} w_{T,l} \sum_{t=r_1}^{r_2}\left(k_{it-l}^{\gamma} \otimes k_{it}^{\theta}\right)\left(\widehat{\theta}-\theta\right)\left(\widehat{\gamma}_i-\gamma_{i0}\right) \\
& +\frac{1}{T} \sum_{l=-m}^{m} w_{T,l} \sum_{t=r_1}^{r_2}\left(\widehat{\gamma}_i-\gamma_{i0}\right)^2 \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma'}\right)
\end{aligned}
$$

All the terms on the RHS are $o_p\left(m/T^{2/5}\right)$ by similar arguments. ■

Lemma 6 *Under Assumptions 1, 2, 3, 4, 5, 6, and 7, we have*
(i) $n^{-1}\sum_{i=1}^{n}\overline{\mathcal{I}}_i-\mathcal{I}=o_p(1)$;
(ii) $\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\gamma_i}\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-E\left[V_i^{\gamma_i}\right]\right\|=o_p\left(1\right)$;
(iii) $\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\theta}\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-E\left[V_i^{\theta}\right]\right\|=o_p\left(1\right)$;
(iv) $\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}U_{it}^{\gamma_i\gamma_i}\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-E\left[U_i^{\gamma_i\gamma_i}\right]\right\|=o_p\left(1\right)$;
(v) $\max_i\left\|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\gamma_i\gamma_i}\left(\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)-E\left[V_i^{\gamma_i\gamma_i}\right]\right\|=o_p\left(1\right)$.

Proof. We only prove the first result. The rest can be proved using the same argument as in Hahn and Kuersteiner (2004). Note that

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

Since

$$
\left|\widehat{\gamma}_i-\gamma_{i0}\right| \leq \frac{1}{\sqrt{T}}\left|\widehat{\gamma}_i^{\epsilon}\left(0\right)\right|+\frac{1}{2T}\left|\widehat{\gamma}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)\right|
$$

with $\max_i T^{-\frac{1}{10}}\left|\widehat{\gamma}_i^{\epsilon}\left(0\right)\right|=o_p(1)$ and $\max_i T^{-\frac{2}{10}}\left|\widehat{\gamma}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)\right|=o_p(1)$ by Lemma 14, it follows that $\max_i\left\|\overline{\mathcal{I}}_i-\mathcal{I}_i\right\|=o_p(1)$ such that

$$
n^{-1}\sum_{i=1}^{n}\overline{\mathcal{I}}_i-\mathcal{I}=o_p(1).
$$

■

Lemma 7 (Hahn and Kuersteiner, 2004) *Let Assumptions 1, 2, 3, 4 and 5 be satisfied. Then* $\Pr\left[\max_i\left|\sqrt{T}\left(\widehat{\gamma}_i-\gamma_{i0}\right)\right|>T^{1/10-\upsilon}\right]=o\left(T^{-1}\right)$ *for* $0<\upsilon<\left(100q+120\right)^{-1}$.

Lemma 8 *Let Assumptions 1, 2, 3, 4 and 5 be satisfied. Then* $\Pr\left[\max_i\left|\sqrt{T}\left(\widehat{\gamma}_i\left(\theta_0\right)-\gamma_{i0}\right)\right|>T^{1/10-\upsilon}\right]=o\left(T^{-1}\right)$ *for* $0<\upsilon<\left(100q+120\right)^{-1}$.

Proof. It can be proved in the same way as in Hahn and Kuersteiner (2004), and is omitted. ■

Lemma 9 *Let* $k_{it} = k(x_{it}; \theta_0, \gamma_{i0})$ *and* $\widehat{k}_{it} = k(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0))$ *where* x_{it} *satisfies Assumption 3,* k_{it} *satisfies Assumption 4 and* $\widehat{\theta}$*,* $\widehat{\gamma}_i$ *are defined in (1). Assume that* $E[k_{it}] = 0$ *for* i, t*. Let* $f_i^{kk} = \sum_{l=-\infty}^{\infty} E\left[k_{it}k_{it-l}'\right]$*. Then,* $\sup_i \left\| \sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[\widehat{k}_{it}\widehat{k}_{it-l}'\right] - f^{kk} \right\| = o_p(1)$*, where* $m, T \to \infty$ *such that* $m = o\left(T^{2/5}\right)$*.*

Proof. For notational simplicity, we may assume without loss of generality that k_{it} is scalar. Let $K_{i,m} = \sum_{l=-m}^{m} w_{T,l} E[k_{it}k_{it-l}]$. We first consider

$$
\begin{aligned}
& \left\| K_{i,m} - f_i^{kk} \right\| \\
\leq\ & \textstyle\sum_{l=-m}^{m} \left| \frac{r_2 - r_1 + 1}{T} w_{T,l} - 1 \right| \left\| E[k_{it}k_{it-l}] \right\| + \sum_{|l|>m} \left\| E[k_{it}k_{it-l}] \right\| \\
\leq\ & \textstyle\sum_{l=-m}^{m} \left(\frac{1}{T} + \frac{1}{m} \right) |l| \left\| E[k_{it}k_{it-l}] \right\| + \sum_{|l|>m} \left\| E[k_{it}k_{it-l}] \right\| \\
\leq\ & \textstyle\sum_{l=-m}^{m} c_1 \left(\frac{1}{T} + \frac{1}{m} \right) |l| \left(a^{\frac{\delta}{2+\delta}} \right)^{|l|} + \left(a^{\frac{\delta}{2+\delta}} \right)^{m} c_2 \sum_{l=1} \left(a^{\frac{\delta}{2+\delta}} \right)^{l} \to 0 \text{ as } m, T \to \infty
\end{aligned}
$$

where the last inequality follows from Assumption 3 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

$$
\left| E\left[k_{it,j_1} k_{it-l,j_2}\right] \right| \leq 8 \left(E\left[\left| k_{it,j_1} \right|^{2+\delta} \right] \right)^{\frac{1}{2+\delta}} \left(E\left[\left| k_{it-l,j_2} \right|^{2+\delta} \right] \right)^{\frac{1}{2+\delta}} \left(a^{\frac{\delta}{2+\delta}} \right)^{|l|}
$$

for some $\delta > 0$. It follows that

$$
\sup_i \left\| K_{i,m} - f_i^{kk} \right\| = o(1).
$$

Now, let

$$
\widehat{K}_{i,m} = \sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[\widehat{k}_{it}\widehat{k}_{it-l}\right] = \sum_{l=-m}^{m} w_{T,l} \int \widehat{k}_{it}\widehat{k}_{it-l}\widehat{p}_{i,t,l}\, d(x_{it}, x_{it-l}).
$$

where

$$
\begin{aligned}
\widehat{k}_{it} &\equiv k_{it}(x_{it}; \theta_0, \widehat{\gamma}_i(\theta_0)) \\
\widehat{p}_{i,t,l} &\equiv p_{i,t,l}\left(x_{it}, x_{it-l}; \widehat{\theta}, \widehat{\gamma}_i\right)
\end{aligned}
$$

Here, $p_{i,t,l}(x_{it}, x_{it-l}; \theta, \gamma_i)$ denotes the joint density of (x_{it}, x_{it-l}). Consider $\widehat{K}_{i,m} - K_{i,m}$

$$
\widehat{K}_{i,m} - K_{i,m} = \sum_{l=-m}^{m} w_{T,l} \int \left(\widehat{k}_{it}\widehat{k}_{it-l}\widehat{p}_{it} - k_{it}k_{it-l}p_{it} \right) d(x_{it}, x_{it-l})
$$

We use the mean value theorem and write

$$\begin{aligned}\widehat{k}_{it}\widehat{k}_{it-l}\widehat{p}_{it}-k_{it}k_{it-l}p_{it} &= \widetilde{k}_{it}^{\gamma}\widetilde{k}_{it-l}\widetilde{p}_{it}\left(\widehat{\gamma}_i\left(\theta_0\right)-\gamma_{i0}\right)+\widetilde{k}_{it}\widetilde{k}_{it-l}^{\gamma}\widetilde{p}_{it}\left(\widehat{\gamma}_i\left(\theta_0\right)-\gamma_{i0}\right)\\ &\quad +\widetilde{k}_{it}\widetilde{k}_{it-l}\widetilde{p}_{it}^{\theta}\left(\widehat{\theta}-\theta\right)+\widetilde{k}_{it}\widetilde{k}_{it-l}\widetilde{p}_{it}^{\gamma}\left(\widehat{\gamma}_i-\gamma_{i0}\right)\end{aligned}$$

where $\widetilde{k}_{it}^{\theta}=\partial k\left(x_{it};\tilde{\theta},\tilde{\gamma}_i\right)/\partial\theta$, etc. Note that we may write $\widetilde{p}_{it}^{\theta}=\widetilde{u}_{it}^{\theta}\widetilde{p}_{it}$ and $\widetilde{f}_{it}^{\gamma}=\widetilde{v}_{it}^{\theta}\widetilde{p}_{it}$. By Assumptions 4 and 8, we obtain

$$\left\|\widehat{K}_{i,m}-K_{i,m}\right\|\leq m\mathbf{M}\left(\sup_i\left\|\widehat{\gamma}_i\left(\theta_0\right)-\gamma_{i0}\right\|+\left\|\widehat{\theta}-\theta\right\|+\sup_i\left\|\widehat{\gamma}_i-\gamma_{i0}\right\|\right)$$

for some finite constant $\mathbf{M}$, or

$$\max_i\left\|\widehat{K}_{i,m}-K_{i,m}\right\|=O_p\left(\frac{m}{T^{2/5}}\right)$$

by Lemmas 7 and 8. ■

Lemma 10 *Let $k_{it}=k\left(x_{it};\theta_0,\gamma_{i0}\right)$ and $\widehat{k}_{it}=k\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right)$ where x_{it} satisfies Assumption 3, k_{it} satisfies Assumption 4 and θ^*, γ_i^* are such that $\left\|\theta^*-\theta\right\|=O_p\left(T^{-2/5}\right)$ and $\sup_i\left\|\gamma_i^*-\gamma_{i0}\right\|=O_p\left(T^{-2/5}\right)$. Then, $\sup_i\left\|\sum_{l=-m}^{m}w_{T,l}E_{\theta^*,\gamma_i^*}\left[\widehat{k}_{it}\widehat{k}_{it-l}'\right]-f^{kk}\right\|=o_p(1)$, where $m,T\to\infty$ such that $m=o\left(T^{2/5}\right)$.*

Proof. Similar to the proof of Lemma 9, and omitted. ■

Lemma 11 (Hahn and Kuersteiner, 2004) $\Pr\left[\max_{1\leq i\leq n}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\widehat{\gamma}_i\left(\epsilon\right)-\gamma_{i0}\right|\geq\eta\right]=o\left(T^{-1}\right)$ *for every $\eta>0$.*

Lemma 12 *Suppose that $K_i\left(\cdot;\theta_0,\gamma_i\left(\theta_0,\epsilon\right)\right)$ is equal to*

$$\frac{\partial^{m_1+m_2}\psi\left(x_{it};\theta_0,\gamma_i\left(\theta_0,\epsilon\right)\right)}{\partial\gamma_i^m}$$

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

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

and

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

Also,

$$\Pr\left[\max_i\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\int K_i\left(\cdot;\theta_0,\gamma_i\left(\theta_0,\epsilon\right)\right)d\Delta_{iT}\right|>CT^{\frac{1}{10}-\upsilon}\right]=o\left(T^{-1}\right)$$

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

Proof. Note that we may write

$$\begin{aligned} & \left\| \int K_i\left(\cdot;\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)dF_i\left(\epsilon\right) - \int K_i\left(\cdot;\theta_0,\gamma_{i0}\right)dF_i \right\| \\ \leq & \left\| \int K_i\left(\cdot;\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)dF_i\left(\epsilon\right) - \int K_i\left(\cdot;\theta_0,\gamma_{i0}\right)dF_i\left(\epsilon\right) \right\| \\ & + \left\| \int K_i\left(\cdot;\theta_0,\gamma_{i0}\right)dF_i\left(\epsilon\right) - \int K_i\left(\cdot;\theta_0,\gamma_{i0}\right)dF_i \right\| \\ \leq & \int M(x_{it})\left(\left|\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)-\gamma_{i0}\right|\right)d\left|F_i\left(\epsilon\right)\right| \\ & + \epsilon\sqrt{T}\left\| \int K_i\left(\cdot;\theta_0,\gamma_{i0}\right)d\left(\widehat{F}_i - F_i\right) \right\|. \end{aligned}$$

Therefore, we have

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

the RHS of which can be bounded by using Lemmas 2 and 11 in absolute value by some $\eta > 0$ with probability $1 - o\left(T^{-1}\right)$.

Because

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

we can bound

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

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

Using Lemmas 3, we can also show that

$$\max_i \left| \int K_i\left(\cdot;\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)d\Delta_{iT} \right|$$

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 $1 - o\left(T^{-1}\right)$. ■

B Consistency

Let

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

where $\widehat{\gamma}_i(\theta) \equiv \operatorname{argmax}_a \sum_{t=1}^{T}\psi(x_{it};\theta,a)$.

Lemma 13 (Hahn and Kuersteiner, 2004) *For all $\eta > 0$, it follows that*

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

Recall now that $\widetilde{\theta}$ is a solution to (19).

Theorem 11 $\Pr\left[\left|\widetilde{\theta}-\theta_0\right| \ge \eta\right] = o\left(T^{-1}\right)$ *for every $\eta > 0$.*

Proof. Let η be given, and let $\varepsilon \equiv \inf_i\left[G_{(i)}(\theta_0,\gamma_{i0}) - \sup_{\{(\theta,\gamma):|(\theta,\gamma)-(\theta_0,\gamma_{i0})|>\eta\}} G_{(i)}(\theta,\gamma)\right] > 0$. Because of Condition 1, we have

$$\left|\frac{1}{T}B_n(\theta)\right| \le \frac{1}{6}\varepsilon$$

with probability equal to $1 - o\left(\frac{1}{T}\right)$. Also, because of Lemma 13, we have

$$\max_{1\le i\le n}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \le \frac{1}{6}\varepsilon$$

with probability equal to $1 - o\left(\frac{1}{T}\right)$. It follows that

$$\begin{aligned}
&\max_{|\theta-\theta_0|>\eta,\gamma_1,\ldots,\gamma_n} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta,\gamma_i) - \frac{1}{T}B_n(\theta) \\
\le\ &\max_{|(\theta,\gamma_i)-(\theta_0,\gamma_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta,\gamma_i) - \frac{1}{T}B_n(\theta) \\
\le\ &\max_{|(\theta,\gamma_i)-(\theta_0,\gamma_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta,\gamma_i) + \frac{1}{6}\varepsilon \\
\le\ &\max_{|(\theta,\gamma_i)-(\theta_0,\gamma_{i0})|>\eta} n^{-1}\sum_{i=1}^{n}G_{(i)}(\theta,\gamma_i) + \frac{1}{3}\varepsilon \\
\le\ &n^{-1}\sum_{i=1}^{n}G_{(i)}(\theta_0,\gamma_{i0}) - \frac{2}{3}\varepsilon \\
\le\ &n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta_0,\gamma_{i0}) - \frac{1}{T}B_n(\theta_0) - \frac{1}{3}\varepsilon
\end{aligned}$$

Because

$$\max_{\theta,\gamma_1,\ldots,\gamma_n} n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta,\gamma_i)-\frac{1}{T}B_n(\theta)\geq n^{-1}\sum_{i=1}^{n}\widehat{G}_{(i)}(\theta_0,\gamma_{i0})-\frac{1}{T}B_n(\theta_0)$$

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

Theorem 12 (Hahn and Kuersteiner, 2004) $\Pr\left[\max_{1\leq i\leq n}\left|\widehat{\gamma}_i-\gamma_{i0}\right|\geq\eta\right]=o\left(T^{-1}\right)$

Theorem 13 *Let* $\overline{\theta}$ *be such that* $\Pr\left[\left|\overline{\theta}-\theta_0\right|\geq\eta\right]=o\left(T^{-1}\right)$ *for every* $\eta>0$. *Then,*

$$\Pr\left[\max_{1\leq i\leq n}\left|\widehat{\gamma}_i\left(\overline{\theta}\right)-\gamma_{i0}\right|\geq\eta\right]=o\left(T^{-1}\right)$$

for every $\eta>0$.

Proof. We first prove that

$$T\Pr\left[\max_{1\leq i\leq n}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\overline{\theta},\gamma\right)-G_{(i)}\left(\theta_0,\gamma\right)\right|\geq\eta\right]=o(1) \tag{35}$$

for every $\eta>0$. Note that

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

Therefore,

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

by Lemma 13 and Theorem 11.

We now get back to the proof of Theorem 13. It suffices to prove that

$$T\Pr\left[\max_{1\leq i\leq n}\left|\widehat{\gamma}_i\left(\overline{\theta}\right)-\gamma_{i0}\right|\geq\eta\right]=o(1)$$

for every $\eta>0$. Let η be given, and let $\varepsilon\equiv\inf_i\left[G_{(i)}(\theta_0,\gamma_{i0})-\sup_{\{\gamma_i:|\gamma_i-\gamma_{i0}|>\eta\}}G_{(i)}(\theta_0,\gamma_i)\right]>0$. Condition on the event

$$\max_{1\leq i\leq n}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\overline{\theta},\gamma\right)-G_{(i)}\left(\theta_0,\gamma\right)\right|\leq\frac{1}{3}\varepsilon,$$

which has a probability equal to $1 - o\left(\frac{1}{T}\right)$ by (35). We then have

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

This is inconsistent with $\widehat{G}_{(i)}\left(\overline{\theta}, \widehat{\gamma}_i\left(\overline{\theta}\right)\right) \geq \widehat{G}_{(i)}\left(\overline{\theta}, \gamma_{i0}\right)$, and therefore, $\left|\widehat{\gamma}_i\left(\overline{\theta}\right) - \gamma_{i0}\right| \leq \eta$ for every i. ■

Corollary 1 $\Pr\left[\max_{1 \leq i \leq n}\left|\widehat{\gamma}_i\left(\widetilde{\theta}\right) - \gamma_{i0}\right| \geq \eta\right] = o\left(T^{-1}\right)$.

Proof. It follows from Theorem 13 above. ■

C Justification of (26)

We analyze the asymptotic distribution of

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right) \tag{36}$$

Let $F\equiv(F_1,\ldots,F_n)$ 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 observation i. Define $F(\epsilon)\equiv F+\epsilon\sqrt{T}\left(\widehat{F}-F\right)$ for $\epsilon\in\left[0,T^{-1/2}\right]$. For each fixed θ and ϵ, let $\gamma_i\left(\theta,F_i\left(\epsilon\right)\right)$ be the solution to the estimating equation

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

and let $\mu\left(F\left(\epsilon\right)\right)$ be the solution to the estimating equation

$$0=\sum_{i=1}^{n}\int\left(U_i\left(x_{it};\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)-\mu\left(F\left(\epsilon\right)\right)\right)dF_i\left(\epsilon\right).$$

Note that $\mu\left(F\left(0\right)\right)=0$, and

$$\begin{aligned}\mu\left(\widehat{F}\right) &\equiv \mu\left(F\left(\frac{1}{\sqrt{T}}\right)\right)=\frac{1}{n}\sum_{i=1}^{n}U_i\left(x_{it};\theta_0,\gamma_i\left(\theta_0,F_i\left(\frac{1}{\sqrt{T}}\right)\right)\right)\\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}U\left(x_{it};\theta_0,\widehat{\gamma}_i\left(\theta_0\right)\right).\end{aligned}$$

By a Taylor series expansion, we have

$$\mu\left(\widehat{F}\right)-\mu\left(F\right)=\frac{1}{\sqrt{T}}\mu^{\epsilon}\left(0\right)+\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\mu^{\epsilon\epsilon}\left(0\right)+\frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3\mu^{\epsilon\epsilon\epsilon}\left(\widetilde{\epsilon}\right), \tag{37}$$

where $\mu^{\epsilon}\left(\epsilon\right)\equiv d\mu\left(F\left(\epsilon\right)\right)/d\epsilon$, $\mu^{\epsilon\epsilon}\left(\epsilon\right)\equiv d^2\mu\left(F\left(\epsilon\right)\right)/d\epsilon^2$, ..., and $\widetilde{\epsilon}$ is somewhere in between 0 and $T^{-1/2}$. It is shown later in Appendix C.2 that the last term is of order $o_p(1)$. We will therefore work with the expansion

$$\sqrt{nT}\left(\mu\left(\widehat{F}\right)-\mu\left(F\right)\right)=\sqrt{nT}\frac{1}{\sqrt{T}}\mu^{\epsilon}\left(0\right)+\sqrt{nT}\frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2\mu^{\epsilon\epsilon}\left(0\right)+o_p\left(1\right). \tag{38}$$

The expansion (26) follows from combining (38) with (44) and (47) below.

C.1 Details of Expansion (37)

C.1.1 $\mu^{\epsilon}\left(0\right)$

In order to obtain (44) and (47), we let

$$h_i\left(\cdot,\epsilon\right)\equiv U_i\left(\cdot;\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)-\mu\left(F\left(\epsilon\right)\right) \tag{39}$$

The first order condition may be written as

$$0 = \frac{1}{n}\sum_{i=1}^{n}\int h_i(\cdot,\epsilon)\, dF_i(\epsilon) \tag{40}$$

Differentiating repeatedly with respect to ϵ, we obtain

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i(\cdot,\epsilon)}{d\epsilon} dF_i(\epsilon) + \frac{1}{n}\sum_{i=1}^{n}\int h_i(\cdot,\epsilon)\, d\Delta_{iT} &(41)\\
0 &= \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2} dF_i(\epsilon) + 2\frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i(\cdot,\epsilon)}{d\epsilon} d\Delta_{iT} &(42)\\
0 &= \frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3h_i(\cdot,\epsilon)}{d\epsilon^3} dF_i(\epsilon) + 3\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i(\cdot,\epsilon)}{d\epsilon^2} d\Delta_{iT} &(43)
\end{aligned}$$

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

Equation (41) can be rewritten as

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\,\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon)) - \mu^{\epsilon}(F(\epsilon))\right) dF_i(\epsilon)\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\int \left(U_i(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon))) - \mu(F(\epsilon))\right) d\Delta_{iT}
\end{aligned}$$

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

$$\mu^{\epsilon}(0) = \frac{1}{n}\sum_{i=1}^{n}\int U_i d\Delta_{iT} \tag{44}$$

C.1.2 γ_i^{ϵ}

In the ith observation, $\gamma_i(\theta_0, F_i(\epsilon))$ solves the estimating equation

$$\int V_i(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\, dF_i(\epsilon) = 0 \tag{45}$$

Differentiating the LHS with respect to ϵ, we obtain

$$0 = \left(\int \frac{\partial V_i(\cdot,\theta,\epsilon)}{\partial \gamma_i'} dF_i(\epsilon)\right) \frac{\partial \gamma_i(\theta, F_i(\epsilon))}{\partial \epsilon} + \int V_i(\cdot,\theta,\epsilon)\, d\Delta_{iT}.$$

Evaluating the expression at $\epsilon = 0$, we obtain gives

$$\gamma_i^{\epsilon} \equiv \frac{\partial \gamma_i(\theta_0, F_i(0))}{\partial \epsilon} = -\left(E\left[\frac{\partial V_i}{\partial \gamma_i'}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_{it}\right). \tag{46}$$

C.1.3 $\mu^{\epsilon\epsilon}(0)$

Equation (42) can be rewritten as

$$\begin{aligned}
0 &= -\frac{1}{n}\sum_{i=1}^{n}\int \mu^{\epsilon\epsilon}(F(\epsilon))\,dF_i(\epsilon) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\left(\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))\otimes\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))\right)\right)dF_i(\epsilon) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\,\gamma_i^{\epsilon\epsilon}(\theta_0,F_i(\epsilon))\right)dF_i(\epsilon) \\
&+\frac{2}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\,\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))-\mu^{\epsilon}(F(\epsilon))\right)d\Delta_{iT}
\end{aligned}$$

where $U_i^{\gamma_i\gamma_i}\equiv\partial^2 U_i/(\partial\gamma_i\otimes\partial\gamma_i)$. Evaluating at $\epsilon=0$, and noting that $E\left[U_i^{\gamma_i}\right]=0$, we obtain

$$\begin{aligned}
\mu^{\epsilon\epsilon}(0) &= \frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\left(\gamma_i^{\epsilon}\otimes\gamma_i^{\epsilon}\right)+\frac{2}{n}\sum_{i=1}^{n}\left(\int U_i^{\gamma_i}(\cdot;\theta_0,\gamma_{i0})\,d\Delta_{iT}\right)\gamma_i^{\epsilon}(\theta_0,F_i(0)) \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\left(\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)\otimes\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)\right) \\
&\quad -\frac{2}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)
\end{aligned}$$

or

$$\begin{aligned}
\mu^{\epsilon\epsilon}(0) &= \frac{1}{n}\sum_{i=1}^{n}E\left[U_i^{\gamma_i\gamma_i}\right]\left[\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\otimes\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\right]\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)\otimes\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)\right] \\
&\quad -\frac{2}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}U_{it}^{\gamma_i}\right)\left(E\left[\frac{\partial V_i}{\partial\gamma_i'}\right]\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right) \qquad (47)
\end{aligned}$$

C.1.4 $\gamma_i^{\epsilon\epsilon}$

Second order differentiation of (45) yields

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

which characterizes $\gamma_i^{\epsilon\epsilon}$.

C.2 Bounding Remainder Term in (37)

Lemma 14 below allows us to ignore the last term in equation (37).

Lemma 14

$$\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\epsilon}(\epsilon)\right| > CT^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right) \tag{48}$$

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\mu^{\epsilon}(\epsilon)\right| > CT^{\frac{1}{10}-\upsilon}\right] = o\left(T^{-1}\right) \tag{49}$$

$$\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^2\right] = o\left(T^{-1}\right) \tag{50}$$

$$\Pr\left[\max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\mu^{\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^2\right] = o\left(T^{-1}\right) \tag{51}$$

$$\Pr\left[\max_i \max_{0\le\epsilon\le\frac{1}{\sqrt{T}}} \left|\gamma_i^{\epsilon\epsilon\epsilon}(\epsilon)\right| > C\left(T^{\frac{1}{10}-\upsilon}\right)^3\right] = o\left(T^{-1}\right) \tag{52}$$

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

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

Proof. Proof is almost identical to the argument in Hahn ad Kuersteiner (2004), and so only the last equality is explicitly established here. From (43), we have

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

where

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3 h_i(\cdot,\epsilon)}{d\epsilon^3} dF_i(\epsilon) \\
= &-\frac{1}{n}\sum_{i=1}^{n}\int \mu^{\epsilon\epsilon\epsilon}(F(\epsilon))\, dF_i(\epsilon) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int U_i^{\gamma_i\gamma_i\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\left(\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))\otimes\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))\otimes\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))\right) dF_i(\epsilon) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\left(\gamma_i^{\epsilon\epsilon}(\theta_0,F_i(\epsilon))\otimes\gamma_i^{\epsilon}(\theta_0,F_i(\epsilon))\right)\right) dF_i(\epsilon) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\left(\gamma_i^{\epsilon\epsilon}(\theta_0,F_i(\epsilon))\otimes\gamma_i^{\epsilon\epsilon}(\theta_0,F_i(\epsilon))\right)\right) dF_i(\epsilon) \\
&+\frac{1}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i}(\cdot;\theta_0,\gamma_i(\theta_0,F_i(\epsilon)))\,\gamma_i^{\epsilon\epsilon\epsilon}(\theta_0,F_i(\epsilon))\right) dF_i(\epsilon)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{3}{n}\sum_{i=1}^{n}\int \frac{d^2 h_i(\cdot,\epsilon)}{d\epsilon^2} d\Delta_{iT} \\
= &-\frac{3}{n}\sum_{i=1}^{n}\int \mu^{\epsilon\epsilon}\left(F\left(\epsilon\right)\right) d\Delta_{iT} \\
&+\frac{3}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i\gamma_i}\left(\cdot;\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)\left(\gamma_i^{\epsilon}\left(\theta_0,F_i\left(\epsilon\right)\right)\otimes\gamma_i^{\epsilon}\left(\theta_0,F_i\left(\epsilon\right)\right)\right)\right) d\Delta_{iT} \\
&+\frac{3}{n}\sum_{i=1}^{n}\int \left(U_i^{\gamma_i\gamma_i}\left(\cdot;\theta_0,\gamma_i\left(\theta_0,F_i\left(\epsilon\right)\right)\right)\left(\gamma_i^{\epsilon}\left(\theta_0,F_i\left(\epsilon\right)\right)\otimes\gamma_i^{\epsilon}\left(\theta_0,F_i\left(\epsilon\right)\right)\right)\right) d\Delta_{iT}
\end{aligned}$$

Combining Lemma 12 in Appendix A and (48)-(52), 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 $1-o\left(T^{-1}\right)$. Likewise, using Lemmas 12, and (48)-(52) again, we can conclude that $\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^3 h_i(\cdot,\epsilon)}{d\epsilon^3} dF_i\left(\epsilon\right)$ is equal to $-\mu^{\epsilon\epsilon\epsilon}\left(F\left(\epsilon\right)\right)$ 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 $1-o\left(T^{-1}\right)$. ■

D Proof of Theorem 3

Without loss of generality, we may write

$$2B_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\ln\det\left(\frac{1}{T}H_i(\theta,\widehat{\gamma}_i(\theta))\right) + \frac{1}{n}\sum_{i=1}^{n}\ln\det\left(\frac{1}{T}\Upsilon_i(\theta,\widehat{\gamma}_i(\theta))\right) \tag{53}$$

We begin with the first component on the RHS of (53). By Assumption 4, each component of $H_i(\theta,\widehat{\gamma}_i(\theta))$ is bounded above by $\sum_{t=1}^{T} M(x_{it})$ such that $\sup_i E\left[|M(x_{it})|^{10q+12+\delta}\right] < \infty$ for some integer $q \geq (\dim(\theta) + \dim(\gamma))/2 + 2$ and for some $\delta > 0$.

Lemma 15 *Suppose that A is an $n \times n$ matrix. Then*

$$|\det(A)| \leq n! \cdot \max(|a_{ij}|)^n$$

Proof. By definition, we have

$$\det(A) = \sum (-1)^{\phi(j_1,\ldots,j_n)} \prod_{i=1}^{n} a_{ij_i}$$

where the summation is taken over all permutations $(j_1,\ldots,j_n)$ of the set of integers $(1,\ldots,n)$ and $\phi(j_1,\ldots,j_n)$ is the number of transpositions required change $(1,\ldots,n)$ into $(j_1,\ldots,j_n)$. Because the number of all permutations is equal to $n!$, we obtain the desired conclusion. ■

Using Lemma 15, we then obtain that

$$\ln\det\left(\frac{1}{T}H_i(\theta,\widehat{\gamma}_i(\theta))\right) \leq \ln r! + r\ln\left(\frac{1}{T}\sum_{t=1}^{T} M(x_{it})\right)$$

where $r = \dim(\gamma)$. It follows that

$$\left|-\frac{1}{n}\sum_{i=1}^{n}\ln\det\left(\frac{1}{T}H_i(\theta,\widehat{\gamma}_i(\theta))\right)\right| \leq \ln r! + r\frac{1}{n}\sum_{i=1}^{n}\left|\ln\left(\frac{1}{T}\sum_{t=1}^{T} M(x_{it})\right)\right|$$

By Lemma 2, we have

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

from which we obtain[12]

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

It follows that

$$\Pr\left[\left|-\frac{1}{n}\sum_{i=1}^{n}\ln\det\left(\frac{1}{T}H_i(\theta,\widehat{\gamma}_i(\theta))\right)\right| > \ln r! + r\frac{1}{n}\sum_{i=1}^{n}\ln(E[M(x_{it})]) + \eta\right] = o\left(T^{-1}\right)$$

[12] In addition to the Condition 4, we need to impose that the minimum of $E[M(x_{it})]$ is bounded away from zero to make this inequality valid.

from which we conclude that

$$\Pr\left[\frac{1}{T}\left|-\frac{1}{n}\sum_{i=1}^{n}\ln\det\left(\frac{1}{T}H_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)\right)\right|>\eta\right]=o\left(T^{-1}\right)$$

for all $\eta>0$.

We now take care of the second component on the RHS of (53). By Assumption 4, each component of $\Upsilon_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)$ is bounded above by $\sum_{l=-m}^{m}w_{T,l}\left(\sum_{t=\max(1,l+1)}^{\min(T,T+l)}M\left(x_{it}\right)M\left(x_{it-l}\right)\right)$. Using Lemma 15, we can then conclude that

$$\ln\det\left(\frac{1}{T}\Upsilon_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)\right)\le\ln r!+r\ln\left(\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}M\left(x_{it}\right)M\left(x_{it-l}\right)\right)$$

Using Lemma 2 again, we have

$$\Pr\left[\max_{1\le i\le n}\left|\frac{1}{T}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\left(M\left(x_{it}\right)M\left(x_{it-l}\right)-E\left[M\left(x_{it}\right)M\left(x_{it-l}\right)\right]\right)\right|>\eta\right]=o\left(T^{-1}\right)$$

and we obtain

$$\Pr\left[\max_{1\le i\le n}\left|\ln\left(\frac{1}{T}\sum_{l=-m}^{m}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}M\left(x_{it}\right)M\left(x_{it-l}\right)\right)-\ln\left(\sum_{l=-m}^{m}E\left[M\left(x_{it}\right)M\left(x_{it-l}\right)\right]\right)\right|>m\eta\right]=o\left(T^{-1}\right)$$

It follows that

$$\Pr\left[\ln\det\left(\frac{1}{T}\Upsilon_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)\right)>\ln r!+r\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-m}^{m}E\left[M\left(x_{it}\right)M\left(x_{it-l}\right)\right]+m\eta\right]=o\left(T^{-1}\right)$$

Because $E\left[M\left(x_{it}\right)M\left(x_{it-l}\right)\right]\le\sqrt{E\left[M\left(x_{it}\right)^2\right]E\left[M\left(x_{it-l}\right)^2\right]}=E\left[M\left(x_{it}\right)^2\right]$, we have

$$\Pr\left[\ln\det\left(\frac{1}{T}\Upsilon_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)\right)>\ln r!+2m\cdot r\frac{1}{n}\sum_{i=1}^{n}E\left[M\left(x_{it}\right)^2\right]+m\eta\right]=o\left(T^{-1}\right)$$

or

$$\Pr\left[\frac{1}{T}\ln\det\left(\frac{1}{T}\Upsilon_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)\right)>\frac{\ln r!}{T}+\frac{2m}{T}r\sup_i E\left[M\left(x_{it}\right)^2\right]+\frac{m}{T}\eta\right]=o\left(T^{-1}\right)$$

Therefore, we obtain

$$\Pr\left[\frac{1}{T}\ln\det\left(\frac{1}{T}\Upsilon_i\left(\theta,\widehat{\gamma}_i\left(\theta\right)\right)\right)>\eta\right]=o\left(T^{-1}\right)$$

for all $\eta>0$.

E Proof of Theorem 4

We can verify by inspection that $\frac{\partial S_n(\theta)}{\partial \theta}$ can be expressed as a sum of terms, all of which are cross section averages of some smooth functions of the form

$$\frac{1}{T}\sum_{t=1}^{T} D^{v}\psi\left(x_{it},\theta,\widehat{\gamma}_i\left(\theta\right)\right), \quad \frac{1}{T}\sum\nolimits_{l=-m}^{m} w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial\psi\left(x_{it},\theta,\widehat{\gamma}_i\left(\theta\right)\right)}{\partial\gamma'}\otimes D^{v}\psi\left(x_{it-l},\theta,\widehat{\gamma}_i\left(\theta\right)\right),$$

$$\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi\left(x_{it},\theta,\widehat{\gamma}_i\left(\theta\right)\right)}{\partial\gamma\partial\gamma'}\right)^{-1}, \quad \left(\frac{1}{T}\sum\nolimits_{l=-m}^{m} w_{T,l}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\frac{\partial\psi\left(x_{it},\theta,\widehat{\gamma}_i\left(\theta\right)\right)}{\partial\gamma}\frac{\partial\psi\left(x_{it-l},\theta,\widehat{\gamma}_i\left(\theta\right)\right)}{\partial\gamma'}\right)^{-1}$$

with $|v|\leq 4$. Here, $\phi\equiv(\theta,\gamma)$, and $D^{v}\psi\left(x_{it},\phi\right)\equiv\partial^{|\nu|}\psi\left(x_{it},\phi\right)/\left(\partial\phi_1^{v_1}...\partial\phi_k^{\nu_k}\right)$, where $\nu=(\nu_1,...,\nu_k)$ be a vector of non-negative integers v_i, and $|v|=\sum_{j=1}^{k}v_j$. By Assumptions 4, 6, and Lemma 5, we can see that all these terms are $O_p\left(1\right)$ uniformly over i and θ.

F Proof of Theorem 6

Because of the result in the previous section, we only need to consider $\Upsilon_i(\theta, \widehat{\gamma}_i(\theta))$. By Assumption 4, each component of $\Upsilon_i(\theta, \widehat{\gamma}_i(\theta))$ is bounded above by $\sum_{l=-m}^{m} E_{\widehat{\theta},\widehat{\gamma}_i}[M(x_{it}) M(x_{it-l})]$. By Assumption 8, we have

$$\sup \sum_{l=-m}^{m} E_{\widehat{\theta},\widehat{\gamma}_i}[M(x_{it}) M(x_{it-l})] \leq 2mK$$

where $K = \sup_{(\theta,\gamma)\in\Phi} \sup_l E_{\theta,\gamma}[M(x_{it}) M(x_{it-l})]$, and

$$\ln\det(\Upsilon_i(\theta, \widehat{\gamma}_i(\theta))) \leq \ln r! + 2rK\ln m$$

It follows that

$$\Pr[\ln\det(\Upsilon_i(\theta, \widehat{\gamma}_i(\theta))) > \ln r! + 2rK\ln m + \eta] = o\left(T^{-1}\right)$$

Therefore, we obtain

$$\Pr\left[\frac{1}{T}\ln\det\left(\frac{1}{T}\Upsilon_i(\theta, \widehat{\gamma}_i(\theta))\right) > \eta\right] = o\left(T^{-1}\right)$$

for all $\eta > 0$ as long as $\frac{\ln m}{T} = o(1)$.

We note that all the above results hold even when the preliminary estimates $\left(\widehat{\theta}, \widehat{\gamma}_i\right)$ are replaced by some (θ^*, γ_i^*).

G Proof of Theorem 8

By differentiating B_n, we obtain that

$$S_n(\theta_0) = [2] + [3] + [4]' + [5]'$$

where

$$[2] = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\psi_{it}}{\partial\theta\left(\partial\gamma'\otimes\partial\gamma'\right)}\right)\mathrm{vec}\left(\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_{it}}{\partial\gamma\partial\gamma'}\right)^{-1}\right)$$

$$[3] = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^3\psi_{it}}{\partial\gamma\left(\partial\gamma'\otimes\partial\gamma'\right)}\right)\mathrm{vec}\left(\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_{it}}{\partial\gamma\partial\gamma'}\right)^{-1}\right)$$

$$\begin{aligned}[4]' &= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\textstyle\sum_{l=-m}^{m} w_{T,l}E_{\widehat{\theta},\widehat{\gamma}_i}\left[\frac{\partial}{\partial\theta}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}}{\partial\gamma'}\right)\right)\right]\right] \\ &\quad\cdot\mathrm{vec}\left(\left(\textstyle\sum_{l=-m}^{m} w_{T,l}E_{\widehat{\theta},\widehat{\gamma}_i}\left[\frac{\partial\psi_{it}}{\partial\gamma}\frac{\partial\psi_{it-l}}{\partial\gamma'}\right]\right)^{-1}\right)\end{aligned}$$

and

$$\begin{aligned}[5]' &= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta}\left[\textstyle\sum_{l=-m}^{m} w_{T,l}E_{\widehat{\theta},\widehat{\gamma}_i}\left[\frac{\partial}{\partial\gamma}\left(\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it}}{\partial\gamma'}\right)\right)\right]\right] \\ &\quad\cdot\mathrm{vec}\left(\left(\textstyle\sum_{l=-m}^{m} w_{T,l}E_{\widehat{\theta},\widehat{\gamma}_i}\left[\frac{\partial\psi_{it}}{\partial\gamma}\frac{\partial\psi_{it}}{\partial\gamma'}\right]\right)^{-1}\right)\end{aligned}$$

We can see that [2] and [3] are identical to the ones in the previous section. Because we have already established

$$[2] + [3] = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}E\left[U_{it}^{\gamma\gamma}\right]\mathrm{vec}\left(\left(E\left[V_{it}^{\gamma}\right]\right)^{-1}\right) + o_p(1)$$

we will focus on $[4]'$ and $[5]'$ here.

Because

$$\begin{aligned}\frac{\partial}{\partial\theta}\left(\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it-l}(\theta,\gamma)}{\partial\gamma'}\right)\right) &= \left(U_{it}^{\gamma}+\rho_i V_{it}^{\gamma}\right)\otimes V_{it-l}' + V_{it}'\otimes\left(U_{it-l}^{\gamma}+\rho_i V_{it-l}^{\gamma}\right) \\ \frac{\partial}{\partial\gamma}\left(\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\otimes\left(\frac{\partial\psi_{it}(\theta,\gamma)}{\partial\gamma'}\right)\right) &= V_{it}^{\gamma}\otimes V_{it-l}' + V_{it}'\otimes V_{it-l}^{\gamma}\end{aligned}$$

and

$$\frac{\partial\widehat{\gamma}_i'(\theta)}{\partial\theta} = -\rho_i + o_p(1)$$

we can write

$$[4]'+[5]' = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[\begin{array}{c} U_{it}^{\gamma}(\theta_0,\widehat{\gamma}_i(\theta_0))\otimes V_{it-l}(\theta_0,\widehat{\gamma}_i(\theta_0))' \\ +V_{it}(\theta_0,\widehat{\gamma}_i(\theta_0))'\otimes U_{it-l}^{\gamma}(\theta_0,\widehat{\gamma}_i(\theta_0)) \end{array}\right]\right]$$
$$\cdot \operatorname{vec}\left(\left(\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[V_{it}(\theta_0,\widehat{\gamma}_i(\theta_0)) V_{it-l}(\theta_0,\widehat{\gamma}_i(\theta_0))'\right]\right)^{-1}\right)+o_p(1)$$

Using Lemma 9, we obtain

$$\max_i\left|\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[V_{it}(\theta_0,\widehat{\gamma}_i(\theta_0)) V_{it-l}(\theta_0,\widehat{\gamma}_i(\theta_0))'\right]-\sum_{l=-\infty}^{\infty} E\left[V_{it}V_{it-l}'\right]\right| = o_p(1)$$

Furthermore, if the conditional likelihood is properly defined, then we should have V_{it} serially uncorrelated, which implies that

$$\max_i\left|\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[V_{it}(\theta_0,\widehat{\gamma}_i(\theta_0)) V_{it-l}(\theta_0,\widehat{\gamma}_i(\theta_0))'\right]-E\left[V_{it}V_{it}'\right]\right|$$
$$=\max_i\left|\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left[V_{it}(\theta_0,\widehat{\gamma}_i(\theta_0)) V_{it-l}(\theta_0,\widehat{\gamma}_i(\theta_0))'\right]+E\left[V_{it}^{\gamma}\right]\right| = o_p(1)$$

where the first equality is based on the information equality. Therefore, we obtain

$$[4]'+[5]'$$
$$= -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l=-m}^{m} w_{T,l} E_{\widehat{\theta},\widehat{\gamma}_i}\left(\begin{array}{c} U_{it}^{\gamma}(\theta_0,\widehat{\gamma}_i(\theta_0))\otimes V_{it-l}(\theta_0,\widehat{\gamma}_i(\theta_0))' \\ +V_{it}(\theta_0,\widehat{\gamma}_i(\theta_0))'\otimes U_{it-l}^{\gamma}(\theta_0,\widehat{\gamma}_i(\theta_0)) \end{array}\right)\right]\cdot \operatorname{vec}\left(E\left[V_{it}^{\gamma}\right]^{-1}\right)+o_p(1)$$

Using Lemma 9 again, we obtain

$$[4]'+[5]' = -\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\otimes V_{it-l}'+V_{it}'\otimes U_{it-l}^{\gamma}\right]\operatorname{vec}\left(E\left[V_{it}^{\gamma}\right]^{-1}\right)+o_p(1)$$

Because we have[13]

$$\begin{aligned}
\left(U_{it}^{\gamma}\otimes V_{it-l}'\right)\operatorname{vec}\left(E\left[V_{it}^{\gamma}\right]^{-1}\right) &= U_{it}^{\gamma}E\left[V_{it}^{\gamma}\right]^{-1}V_{it-l} = -U_{it}^{\gamma}\widetilde{V}_{it-l} \\
\left(V_{it}'\otimes U_{it-l}^{\gamma}\right)\operatorname{vec}\left(E\left[V_{it}^{\gamma}\right]^{-1}\right) &= U_{it-l}^{\gamma}E\left[V_{it}^{\gamma}\right]^{-1}V_{it} = -U_{it-l}^{\gamma}\widetilde{V}_{it}
\end{aligned}$$

it follows that

$$\begin{aligned}
[4]'+[5]' &= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}+U_{it-l}^{\gamma}\widetilde{V}_{it}\right]+o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=-\infty}^{\infty} E\left[U_{it}^{\gamma}\widetilde{V}_{it-l}\right]+o_p(1)
\end{aligned}$$

We note that, because of Lemma 10, all the above results hold even when the preliminary estimates $\left(\widehat{\theta},\widehat{\gamma}_i\right)$ are replaced by some (θ^*,γ_i^*) as long as $\|\theta^*-\theta_0\| = O_p\left(T^{-2/5}\right)$ and $\sup_i\|\gamma_i^*-\gamma_{i0}\| = O_p\left(T^{-2/5}\right)$.

[13] See, e.g., Magnus & Neudecker (1988, p. 31, eq. (3)).

References

[1] Hahn, J., and G. Kuersteiner (2004): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects", *unpublished manuscript.*

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

[3] 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.

[4] Magnus, J. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley.