Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects: Supplementary Appendix

Jinyong Hahn
UCLA

Guido Kuersteiner
MIT

D Some Lemmas

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

Lemma 7 *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 $\left|\sum_3 E\left[\prod_{t=1}^j W_{i_t}^{\gamma_t}\right]\right|$, where $(\gamma_1, .., \gamma_j)$ is a j-tuple such that $2r = \sum_{i=1}^j \gamma_i$, $j > r$. Let $A = \{t : \gamma_t = 1\}$ 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 $B_m = \{(i_1, ..., i_j) : 1 \leq i_1 \leq ... \leq i_j \leq n, |i_{t-1} - i_t| > m, |i_t - i_{t+1}| > m \text{ for some } t \in A\}$. Now consider $E\left[\prod_{t=1}^j W_{i_t}^{\gamma_t}\right]$ when $j > r$. Fix $\tau \in A$. Then, if $1 < \tau < j$ define $W_a = \prod_{t=1}^{\tau-1} W_{i_t}^{\gamma_t}$, $W_b = W_{i_\tau}$ and $W_c = \prod_{t=\tau+1}^j W_{i_t}^{\gamma_t}$ as well as $b_1 = \sum_{t=1}^{\tau-1}\gamma_t$ and $b_2 = \sum_{t=\tau}^j \gamma_t$ such that

$$\begin{aligned}
\left|E\left[\prod_{t=1}^j W_{i_t}^{\gamma_t}\right]\right| &\leq |E[W_a W_b W_c] - E[W_a W_b] E[W_c]| + |E[W_a W_b]| \, |E[W_c]| \\
&\leq 8\left(E\left[|W_a W_b|^{\frac{2r+\delta}{b_1+1}}\right]\right)^{\frac{b_1+1}{2r+\delta}} \left(E\left[|W_c|^{\frac{2r+\delta}{b_2}}\right]\right)^{\frac{b_2+1}{2r+\delta}} \alpha(m)^{\frac{\delta}{2r+\delta}} \\
&\quad + 8|E[W_c]| \left(E\left[|W_a|^{\frac{2r+\delta}{b_1}}\right] E\left[|W_b|^{2r+\delta}\right]\right)^{\frac{b_1+1}{2r+\delta}} \alpha(m)^{\frac{\delta+b_2}{4r+\delta}} \\
&\leq 8E\left[|W_i|^{2r+\delta}\right]\alpha(m)^{\frac{\delta}{2r+\delta}} + 8E\left[|W_i|^{2r+\delta}\right]\alpha(m)^{\frac{\delta+b_2}{4r+\delta}}, \qquad (25)
\end{aligned}$$

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 $W_b = \prod_{t=\tau+1}^j W_{i_t}^{\gamma_t}$ such that

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

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 (25) 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 8 *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 $\sup_i |\alpha_i(m)| \le Ca^m$ 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*

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

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

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

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

Proof. By the Markov inequality

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

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

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

From Lemma (7) it follows that

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

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

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

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. ■

E Consistency

Recall that

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

We restate Lemma 1 below for convenience:

Lemma 9 *For all $\eta > 0$ that*

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

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

$$\Pr\left[\max_{1\le i\le n}\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \ge \eta\right] \le \sum_{i=1}^{n}\Pr\left[\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| \ge \eta\right]. \tag{26}$$

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

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

and therefore

$$\Pr\left[\sup_{(\theta,\gamma)}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| > \eta\right] \le \sum_{j=1}^{M(\varepsilon)}\Pr\left[\sup_{\Upsilon_j}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| > \eta\right] \tag{27}$$

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

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

Then,

$$\begin{aligned}
\Pr\left[\sup_{\Upsilon_j}\left|\widehat{G}_{(i)}(\theta,\gamma) - G_{(i)}(\theta,\gamma)\right| > \eta\right] &\le \Pr\left[\left|\widehat{G}_{(i)}(\theta_j,\gamma_j) - G_{(i)}(\theta_j,\gamma_j)\right| > \frac{\eta}{3}\right] \\
&\quad + \Pr\left[\frac{1}{T}\left|\sum_{t=1}^{T}(M(x_{it}) - E[M(x_{it})])\right| > \frac{\eta}{3\varepsilon}\right] \\
&= o\left(T^{-2}\right) \qquad (28)
\end{aligned}$$

by Lemma 8. Combining (26), (27), (28), and $n = O(T)$, we obtain the desired conclusion. ■

Equations (4) and (5) are formally established by the two Theorems below:

Theorem 3 $\Pr\left[\left|\widehat{\theta} - \theta_0\right| \ge \eta\right] = o\left(T^{-1}\right)$ *for every $\eta > 0$.*

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

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

where the second and fourth inequalities are based on Lemma 1. Because

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

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

Theorem 4 $\Pr\left[\max_{1\le i\le n}|\widehat{\gamma}_i - \gamma_{i0}| \ge \eta\right] = o\left(T^{-1}\right)$

Proof. We first prove that

$$T\Pr\left[\max_{1\le i\le n}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\gamma\right) - G_{(i)}(\theta_0,\gamma)\right| \ge \eta\right] = o(1) \tag{29}$$

for every $\eta > 0$. Note that

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

Therefore,

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

by Lemma 1 and Theorem 4.

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

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

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 $\left\{\max_{1\le i\le n}\sup_{\gamma}\left|\widehat{G}_{(i)}\left(\widehat{\theta},\gamma\right) - G_{(i)}(\theta_0,\gamma)\right| \le \frac{1}{3}\varepsilon\right\}$, which has a probability equal to $1-o\left(\frac{1}{T}\right)$

by (29). We then have

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

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

F Proof of Lemma 3

First note that by Condition 3 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 $p' \equiv \dim(\xi)$. 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 $\varepsilon>0$

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

In order to verify this condition, let $\xi_{it}\equiv\xi\left(x_{it},\phi_i\right)$, and note that

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

where $\mathbf{1}\left\{|v_{i,n,T}|>\varepsilon\right\}\leq\mathbf{1}\left\{\|Z_{iT}\|^2>\varepsilon^2/\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^2\right\}$. Observe that (i) $\ell'\Sigma_{nT}^{-1}\ell\leq\max_k\lambda_k^{-1}=1/\min_k\lambda_k$ with λ_k eigenvalues of Σ_{nT}; and (ii) $\min_k\lambda_k\geq n\inf_i\inf_T\lambda_{iT}^{\xi}$. It follows that $\left\|\ell'\Sigma_{nT}^{-1}\ell\right\|^4\leq c_2n^{-2}$ for some constant $c_2<\infty$.

We now take care of $E\left[\xi_{it_1}'\xi_{it_2}\xi_{it_3}'\xi_{it_4}\right]$. For this purpose, let the j-th element of ξ_{it} be denoted by $\xi_{j,it}$. We then have

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

where $\operatorname{Cum}\left(\xi_{j_1,it_1},\xi_{j_1,it_2},\xi_{j_2,it_3},\xi_{j_2,it_4}\right)$ is a fourth order cumulant term. The terms on the right side can be bounded by making following observations. First, from Andrews (1991, Lemma 1), it follows that $\sum_{t_2,..,t_4}\sup_{t_1}\operatorname{Cum}\left(\xi_{j_1,it_1},\xi_{j_1,it_2},\xi_{j_2,it_3},\xi_{j_2,it_4}\right)<\infty$. Second, from Hall and Heyde (1980, Corollary A.2), we have $E\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right]\leq 8E\left[\left|\xi_{j_1,it_1}\right|^4\right]E\left[\left|\xi_{j_1,it_2}\right|^4\right]\alpha\left(t_1-t_2\right)^{1/2}$ such that $T^{-1}\sum_{t_1,t_2}E(\left[\xi_{j_1,it_1}\xi_{j_1,it_2}\right]\leq c_1\sum_{l=0}^{T}\left(1-l/T\right)\left(\sqrt{a}\right)^l<\infty$ for all T and some constant $c_1<\infty$.

These arguments show that

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

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

$$\begin{aligned}
\sup_i \left\|\Sigma_{iT}^{\xi\xi}-f_i^{\xi\xi}\right\| &\leq \frac{|l|}{T}\sum_{l=-T}^{T}\sup_i\left\|\operatorname{Cov}\left(\xi_{it},\xi_{it-l}'\right)\right\|+\sum_{|l|\geq T+1}\sup_i\left\|\operatorname{Cov}\left(\xi_{it},\xi_{it-l}'\right)\right\| \\
&\leq \frac{1}{T}\sum_{l=-\infty}^{\infty}|l|\sup_i\left\|\operatorname{Cov}\left(\xi_{it},\xi_{it-l}'\right)\right\|+c_1\sum_{|l|\geq T+1}\left(1-|l|/T\right)\left(\sqrt{a}\right)^l \\
&\to 0 \text{ as } T\to\infty. \qquad (30)
\end{aligned}$$

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

G Proof of Lemma 4

Note that we have

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

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

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

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

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

such that

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

By Lemma (8), 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

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

H Proof of Lemma 5

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

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

from Lemma (13). By Lemma (11), we also have

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

By Lemma (11) again, it follows that

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

This proves the result for $\widehat{\gamma}_{i}^{\epsilon}\left(\epsilon\right)$. For $\widehat{\gamma}_{i}(0)$ the result follows directly from (22) and Lemmas (11) and (13). For $\widehat{\gamma}_{i}^{\epsilon\epsilon}\left(\epsilon\right)$, the result follows from representation (24) as well as Lemmas (11), (13), and (15).

I Proof of Lemma 6

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 $\frac{1}{n}\sum_{i=1}^{n}A_iE\left(K_{i,m}\right)-f^{kk}=o(1)$ and $\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}A_iK_{i,m}\right)=o(1)$. Since $f^{kk}-n^{-1}\sum_{i=1}^{n}A_if_i^{kk}=o(1)$ by definition, we first consider

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

where the last inequality follows from Condition 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 $\delta>0$. Since the bound on $\left\|A_i\right\|\left\|E\left[K_{i,m}\right]-f_i^{kk}\right\|$ is uniform it therefore follows that $\frac{1}{n}\sum_{i=1}^{n}A_iE\left[K_{i,m}\right]-f^{kk}=o(1)$. Next we show that

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

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

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

where the last equality follows from the same arguments as in the proof of Lemma (3) such that $\operatorname{Var}\left(K_{i,m}\right)$ 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

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

We use the decomposition

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

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

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

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

$$\begin{aligned}
& \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \operatorname{vec}\left[A_i\left(\widehat{k}_{it} - k_{it}\right)k_{it-1}'\right] \\
= & \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left(I \otimes A_i\right)\left(k_{it-l} \otimes k_{it}^{\theta}\right)\left(\widehat{\theta} - \theta\right) \\
& + \frac{\left(\widehat{\gamma}_i - \gamma_{i0}\right)}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} \left(I \otimes A_i\right) \operatorname{vec}\left[k_{it}^{\gamma}k_{it-l}'\right] \qquad (31)
\end{aligned}$$

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

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

such that $E\left|\frac{1}{T}\sum_{t=r_1}^{r_2}k_{it-l}k_{it}^{\theta}\right| \leq \left(\frac{1}{T}\sum_{t=1}^{T}E\left[M\left(x_{it-l}\right)^2\right]\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}E\left[M\left(x_{it}\right)^2\right]\right)^{1/2} = O(1)$ 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 (31) where

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\left(\widehat{\gamma}_i-\gamma_{i0}\right)}{T}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\operatorname{vec}\left[k_{it}^{\gamma}k_{it-l}'\right] \qquad (32)\\
= & \frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\gamma}_i^{\epsilon}\left(0\right)}{T^{3/2}}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes k_{it}^{\gamma}\right]\\
& +\frac{1}{n}\sum_{i=1}^{n}\left(I\otimes A_i\right)\frac{\widehat{\gamma}_i^{\epsilon\epsilon}\left(\tilde{\epsilon}\right)}{T^{2}}\sum_{l=-m}^{m}\sum_{t=r_1}^{r_2}\left[k_{it-l}\otimes k_{it}^{\gamma}\right].
\end{aligned}$$

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

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

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

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

by the Hölder inequality. Then

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

which is uniformly bounded in i.By (22) it follows that $E\left[\left|\widehat{\gamma}_i^{\epsilon}\left(0\right)\right|^2\right]$ is bounded uniformly in i. From the Markov inequality it then follows that

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

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

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

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

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

For the second term in (32) use the bound

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

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

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

We now turn to

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

such that

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^{n} (I \otimes A_i) \frac{1}{T} \sum_{l=-m}^{m} \sum_{t=r_1}^{r_2} (\widehat{\gamma}_i - \gamma_{i0})^2 \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma\prime}\right) \right\| \\
\leq\ & \sup_i \|I \otimes A_i\| \max_i |\widehat{\gamma}_i - \gamma_{i0}| \frac{1}{n} \sum_{i=1}^{n} |\widehat{\gamma}_i - \gamma_{i0}| \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma\prime}\right) \right\| \\
\leq\ & \sup_i \|I \otimes A_i\| \max_i |\widehat{\gamma}_i - \gamma_{i0}| \frac{1}{n} \sum_{i=1}^{n} \frac{|\widehat{\gamma}_i^{\epsilon}(0)|}{\sqrt{T}} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma\prime}\right) \right\| \\
& + \sup_i \|I \otimes A_i\| \max_i |\widehat{\gamma}_i - \gamma_{i0}| \max_i \sup_{\epsilon \in [0,1/\sqrt{T}]} |\widehat{\gamma}_i^{\epsilon\epsilon}(\epsilon)| \frac{1}{2nT} \sum_{i=1}^{n} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma\prime}\right) \right\| \\
=\ & o_p(m/T^{1/2})
\end{aligned}$$

where by the same arguments as before $|\widehat{\gamma}_i^{\epsilon}(0)|$ is uniformly bounded and $\left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma\prime}\right) \right\| = O(1)$ uniformly in i such that the first term is $o_p(m/T^{1/2})$ by the Markov inequality and the fact that $\max_i |\widehat{\gamma}_i - \gamma_{i0}| = o_p(1)$. The second term is $o_p(m/T^{1/2})$ because $\max_i \sup_{\epsilon \in [0,1/\sqrt{T}]} |\widehat{\gamma}_i^{\epsilon\epsilon}(\epsilon)| = O_p(T^{\frac{1}{5}-2v})$ by Lemma (5) and $\frac{1}{nT} \sum_{i=1}^{n} \sum_{l=-m}^{m} \left\| \frac{1}{T} \sum_{t=r_1}^{r_2} \operatorname{vec}\left(k_{it}^{\gamma} k_{it-l}^{\gamma\prime}\right) \right\| = O_p(mT^{-1})$ 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.

J Proof of (7)

Recall that

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

and $\theta^{\epsilon}(0)$, $\theta^{\epsilon\epsilon}(0)$, and $\theta^{\epsilon\epsilon\epsilon}(\widetilde{\epsilon})$ are characterized by

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

Theorem 5 below shows that

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

which proves (7). For example, θ^{ϵ}, γ_i^{θ}, and γ_i^{ϵ} are characterized by

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

and

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

J.1 $\gamma_i^{\theta\theta}$, $\gamma_i^{\theta\epsilon}$, and $\gamma_i^{\epsilon\epsilon}$

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 (16) yields

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

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

and

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

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

Lemma 10

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

and

$$T\Pr\left[\max_{1\leq i\leq n}\max_{0\leq\epsilon\leq\frac{1}{\sqrt{T}}}\left|\widehat{\gamma}_i\left(\epsilon\right)-\gamma_{i0}\right|\geq\eta\right]=o\left(1\right)$$

for every $\eta>0$.

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

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

We have

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

and

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

By Lemma 1, we have

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

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

$$\begin{aligned}\max_{|\theta-\theta_0|>\eta,\gamma_1,\ldots,\gamma_n}n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta,\gamma_i\left(\theta\right)\right)dF_i\left(\epsilon\right) &\leq \max_{|(\theta,\gamma)-(\theta_0,\gamma_{i0})|>\eta}n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta,\gamma_i\left(\theta\right)\right)dF_i\left(\epsilon\right)\\ &< \max_{|(\theta,\gamma)-(\theta_0,\gamma_{i0})|>\eta}n^{-1}\sum_{i=1}^n G_{(i)}\left(\theta,\gamma_i\right)+\frac{1}{3}\varepsilon\\ &< n^{-1}\sum_{i=1}^n G_{(i)}\left(\theta_0,\gamma_{i0}\right)-\frac{2}{3}\varepsilon\\ &< n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta_0,\gamma_{i0}\right)dF_i\left(\epsilon\right)-\frac{1}{3}\varepsilon.\end{aligned}$$

We also have

$$\max_{\theta,\gamma_1,\ldots,\gamma_n}n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta,\gamma_i\right)dF_i\left(\epsilon\right)\geq n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta_0,\gamma_{i0}\right)dF_i\left(\epsilon\right)$$

by definition. It follows that

$$\max_{|\theta-\theta_0|>\eta,\gamma_1,\ldots,\gamma_n}n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta,\gamma_i\left(\theta\right)\right)dF_i\left(\epsilon\right)<\max_{\theta,\gamma_1,\ldots,\gamma_n}n^{-1}\sum_{i=1}^n\int g\left(\cdot;\theta,\gamma_i\right)dF_i\left(\epsilon\right)-\frac{1}{3}\varepsilon$$

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 11 *Suppose that* $K_i(\cdot;\theta(\epsilon),\gamma_i(\theta(\epsilon),\epsilon))$ *is equal to*

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

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

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

and

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

Also,

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

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

Proof. Note that we may write

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

Therefore, we have

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

the RHS of which can be bounded by using Lemmas 8 and 10 in absolute value by some $\eta>0$ with probability $1-o\left(T^{-1}\right)$.

Because

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

we can bound

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

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

Using Lemmas 4, we can also show that

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

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 12

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

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

Proof. We have

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

Using Lemma 11, we can see that

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

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

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

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

Lemma 13

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

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

Proof. We have

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

Using Lemmas 11, and 12, we can bound the denominator of $\theta^{\epsilon}(\epsilon)$ by some $C>0$, and the numerator by some $CT^{\frac{1}{10}-\upsilon}$ with probability $1-o\left(T^{-1}\right)$. ■

Lemma 14

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

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

Proof. From Section J.1, we have

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

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

and

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

The result then follows by applying the same argument as in the proof of Lemma 12. ■

Lemma 15

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

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

Proof. The conclusion follows by using the characterization of $\theta^{\epsilon\epsilon}(\epsilon)$, and Lemmas 11, 12, 13, and 14. ■

Lemma 16

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

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

Proof. It was seen in Section J.1 that

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

and

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

We therefore obtain

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

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

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

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

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

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

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

which characterizes $\frac{\partial^3 \gamma_i(\theta, F_i(\epsilon))}{\partial \epsilon^3}$. Inspecting these derivatives and applying Lemmas 11, 12, and 14, we obtain the desired result. ■

Theorem 5

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

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

Proof. From (14), 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 11, 12, 13, 14, and 15, 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 $1 - o\left(T^{-1}\right)$. Recall 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 \gamma_i} \left(\frac{\partial \gamma_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 \gamma_i} \frac{\partial \gamma_i}{\partial \epsilon} \frac{\partial \theta}{\partial \epsilon} \\
&+ \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial \theta'} \frac{\partial^2 \theta}{\partial \epsilon^2} \\
&+ \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial \theta' \partial \gamma_i} \frac{\partial \theta}{\partial \epsilon} \frac{\partial \gamma_i}{\partial \theta'} \frac{\partial \theta}{\partial \epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i^2} \left(\frac{\partial \gamma_i}{\partial \theta'} \frac{\partial \theta}{\partial \epsilon}\right)^2 + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i^2} \left(\frac{\partial \gamma_i}{\partial \theta'} \frac{\partial \theta}{\partial \epsilon}\right) \frac{\partial \gamma_i}{\partial \epsilon} \\
&+ \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i} \left(\frac{\partial \theta'}{\partial \epsilon} \frac{\partial^2 \gamma_i}{\partial \theta \partial \theta'}\right) \frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i} \frac{\partial^2 \gamma_i}{\partial \theta' \partial \epsilon} \frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i} \frac{\partial \gamma_i}{\partial \theta'} \frac{\partial^2 \theta}{\partial \epsilon^2} \\
&+ \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial \theta' \partial \gamma_i} \frac{\partial \theta}{\partial \epsilon} \frac{\partial \gamma_i}{\partial \epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i^2} \left(\frac{\partial \gamma_i}{\partial \theta'} \frac{\partial \theta}{\partial \epsilon}\right) \frac{\partial \gamma_i}{\partial \epsilon} + \frac{\partial^2 h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i^2} \left(\frac{\partial \gamma_i}{\partial \epsilon}\right)^2 \\
&+ \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i} \frac{\partial^2 \gamma_i}{\partial \epsilon \partial \theta'} \frac{\partial \theta}{\partial \epsilon} + \frac{\partial h_i^{(r)}(\cdot,\epsilon)}{\partial \gamma_i} \frac{\partial^2 \gamma_i}{\partial \epsilon^2}.
\end{aligned}$$

Using Lemmas 11, 12, 13, 14, and 15 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 $\left(\frac{1}{n}\sum_{i=1}^{n} \int \frac{\partial h_i(\cdot,\epsilon)}{\partial \theta'} dF_i(\epsilon)\right)^{-1}$ is bounded away from 0 by Lemma 11, we obtain the desired conclusion. ■

Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects: Supplementary Appendix For the Dynamic Tobit Model

Jinyong Hahn	Guido Kuersteiner
UCLA	MIT

Definition 1 *Let x_{it} and ε_{it} be processes defined on a probability space $(S, \mathfrak{S}, P)$ taking values in $\mathbb{R}^p$ and $\mathbb{R}^d$ respectively. Then the process x_{it} is called uniformly near epoch dependent (NED) of size $-q$ on ε_{it} if*

$$v(m) = \sup_i \left(E \left\| x_{it} - E\left(x_{it} | \varepsilon_{it+m}, ..., \varepsilon_{it-m}\right) \right\|^2 \right)^{1/2}$$

satisfies $v(m) = O(m^{-\lambda})$ for some $\lambda > q$.

Lemma 17 *Let $y_{it} = \max(0, \phi_0 y_{it-1} + \varepsilon_{it})$ for $t \geq 1$ with $|\phi_0| < 1$ and y_{i0} such that $E|y_{i0}|^r < \infty$ for some $r > 2$. Then y_{it} is uniformly NED on ε_{it} with $v(m) \leq Ca^m$ for some constants $0 < C < \infty$ and $0 < a < 1$. If $\sup_i \sup_t E|\varepsilon_{it}|^r < \infty$ and $\sup_i E|y_{i0}|^r < \infty$ for some $r \geq 2$ then $\sup_t E|y_{it}|^r < \infty$.*

Proof. Let $g(y_{it-1}, \varepsilon_{it}) = \max(0, \phi_0 y_{it-1} + \varepsilon_{it})$. Then $\left|g(y_{it-1}, \varepsilon_{it}) - g(y^*_{it-1}, \varepsilon_{it})\right| \leq |\phi_0| \left|y_{it-1} - y^*_{it-1}\right|$ and $|g(\bar{y}, \bar{\varepsilon})| < \infty$ for all $|\bar{y}| < \infty, |\bar{\varepsilon}| < \infty$ such that the result follows by Pötscher and Prucha (1991, Theorem 6.10 and Theorem 6.11). ■

Lemma 18 *Let ε_{it} be α-mixing with uniform mixing coefficients $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$. Let y_{it} take values in $\mathbb{R}$, $\sup_i E|y_{it}|^r < \infty$ for some $r \geq 2$ and be uniformly NED on ε_{it} such that $v(m) \leq Ca^m$. Then, for some bounded constant $C_0 > 0$, any q with $2 < q \leq r$ and some $0 \leq \bar{a} < 1$*

$$|Ey_{it}y_{is} - Ey_{it}Ey_{is}| \leq C_0 \left(\bar{a}^{|t-s|}\right)^{1-r^{-1}-q^{-1}}.$$

Proof. Let $\eta_{it} = y_{it} - Ey_{it}$ such that

$$\begin{aligned} Ey_{it}y_{is} - Ey_{it}Ey_{is} &= E\eta_{it}\eta_{is} \\ &= E\left(\eta_{it} - E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right)\left(\eta_{is} - E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right) \\ &\quad + E\left(\left(\eta_{it} - E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right) E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right) \\ &\quad + E\left(E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\left(\eta_{is} - E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right)\right) \\ &\quad + E\left(E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right] E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right). \end{aligned}$$

By the Cauchy-Schwartz inequality and the fact that y_{it} is NED on ε_{it} it follows that

$$\begin{aligned} & \left|E\left(\eta_{it} - E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right)\left(\eta_{is} - E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right)\right| \\ \leq & \left(E\left|\eta_{it} - E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right|^2\right)^{1/2}\left(E\left|\eta_{is} - E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right|^2\right)^{1/2} \\ \leq & \left(v(|t-s|/2)\right)^2 \leq Ca^{2|t-s|}. \end{aligned}$$

Again by the Cauchy-Schwartz inequality we find that for some constant C_1

$$\begin{aligned} & \left|E\left(\left(\eta_{it} - E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right) E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right)\right| \\ \leq & \left(E\left|\eta_{it} - E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right|^2\right)^{1/2}\left(E|\eta_{is}|^2\right)^{1/2} \\ \leq & \left(E|\eta_{is}|^2\right)^{1/2}\left(v(|t-s|/2)\right)^2 \leq C_1 a^{2|t-s|}. \end{aligned}$$

by the fact that y_{it} is NED on ε_{it}. Finally, we note that $E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]$ is a function of a finite number of ε_{it} and thus is mixing. Moreover, note that $E\left(E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]\right) = E\eta_{it} = 0$ such that for any q with $2 < q \leq r$

$$\left|E\left(E\left[\eta_{it}|\varepsilon_{it+|t-s|/2}...\varepsilon_{it-|t-s|/2}\right]E\left[\eta_{is}|\varepsilon_{is+|t-s|/2}...\varepsilon_{is-|t-s|/2}\right]\right)\right| \leq 8\left(E\left|\eta_{it}\right|^r\right)^{1/r}\left(E\left|\eta_{is}\right|^q\right)^{1/q}\left(a^{|t-s|}\right)^{1-r^{-1}-q^{-1}}$$
$$\leq 32\left(E\left|y_{it}\right|^r\right)^{1/r}\left(E\left|y_{is}\right|^q\right)^{1/q}\left(a^{|t-s|}\right)^{1-r^{-1}-q^{-1}}$$

where the first inequality follows by Corollary A.2 of Hall and Heyde (1980) and the second inequality uses the Minkowski and Jensen inequalities. Combining the inequalities then gives the result. ■

Proposition 2 *Let ε_{it} be iid across i and t with $E\left|\varepsilon_{it}\right|^r < \infty$ for some $r > 2$ and z_{it} is strictly stationary and mixing with $\sup_i \left|\alpha_i(m)\right| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$ and $\sup_i E\left|z_{it}\right|^r < \infty$. Then $y_{it} = \max(0, \phi_0 y_{it-1} + z_{it}'\beta_0 + \gamma_{i0} + \varepsilon_{it})$ with $|\phi_0| < 1$ has a stationary solution y_{it}, $\sup_i E\left|y_{it}\right|^r < \infty$, $x_{it} = (y_{it}, z_{it})$ is NED with exponentially decaying coefficients $v(m)$ and for any function $\xi(x_{it})$ such that $|\xi(x_{it}) - \xi(x_{it}^m)| \leq C(x_{it}, x_{it}^m)\left\|x_{it} - x_{it}^m\right\|$ where $x_{it}^m = E\left(x_{it}|z_{it+m}^*, ...z_{it-m}^*\right)$ with $\sup_i \sup_m \sup_t E\left|C(x_{it}, x_{it}^m)\right|^2 < \infty$,*

$$\sup_i \sup_m \sup_t E\left(C(x_{it}, x_{it}^m)\left\|x_{it} - x_{it}^m\right\|\right)^r < \infty$$

and $\sup_i E\left\|\xi(x_{it})\right\|^r < \infty$ for $r > 2$ it follows that $\sup_i \left|\operatorname{Cov}\left(\xi(x_{it}), \xi(x_{is})\right)\right| \leq C_0 b^{|t-s|}$ for some constants $C_0 > 0$ and $0 \leq b < 1$.

Proof. We note that $z_{it}^* = z_{it}'\beta_0 + \gamma_{i0} + \varepsilon_{it}$ is strictly stationary and mixing with mixing coefficients $\alpha_i(m)$ such that $\sup_i \left|\alpha_i(m)\right| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$. To show stationarity we adapt the argument of de Jong and Woutersen (2003). Let y_{it}^m be a process that satisfies $y_{is}^m = \max(0, \phi_0 y_{is-1}^m + z_{is}^*)$ for $s = t, t-1, ..., t-m$ and $y_{is}^m = 0$ for $s < t-m$. Note that $y_{it}^m = f(z_{it}^*, ..., z_{it-m}^*)$ for some measurable map f such that y_{it}^m is stationary by construction. Assume that y_{it}^* is a solution to $y_{it} = \max(0, \phi_0 y_{it-1} + z_{it}^*)$. Consider

$$\begin{aligned} \left|y_{it}^* - y_{it}^m\right| &= \left|\max(0, \phi_0 y_{it-1}^* + z_{it}^*) - \max(0, \phi_0 y_{it-1}^m + z_{it}^*)\right| \\ &\leq \left|\phi_0\right|\left|y_{it-1}^* - y_{it-}^m\right| \leq .... \leq \left|\phi_0\right|^m\left|y_{it-m}^*\right| \end{aligned} \tag{33}$$

as well as

$$\left|y_{it}^*\right| \leq \left|y_{it}^m\right| + \left|\phi_0\right|^m\left|y_{it-m}^*\right|$$

such that

$$\left|y_{it}^*\right| \leq \left(1 - \left|\phi_0\right|^m\right)^{-1}\left|y_{it}^m\right| \tag{34}$$

By Theorem 6.10 of Pötscher and Prucha (1991) it follows that $\sup_i \sup_{t\geq 0} E\left|y_{it}^m\right|^r < \infty$ uniformly in m. Then, from the previous Inequality 34 we have $\sup_i \sup_{t\geq 0} E\left|y_{it}^*\right| < \infty$. Form Inequality (33) it follows that $\sup_i \sup_{t\geq 0} E\left|y_{it}^* - y_{it}^m\right|^r \leq \left|\phi_0\right|^{rm} \sup_i \sup_{t\geq 0} E\left|y_{it}^*\right| \to 0$ as $m \to \infty$. We have shown convergence in r-th mean of y_{it}^* to y_{it}^m as $m \to \infty$ uniformly in i and $t \geq 0$. This means that the solution y_{it}^* converges in distribution to y_{it}^m as $m \to \infty$ where y_{it}^m is stationary by construction. Let y_{it} be the limit of y_{it}^m as $m \to \infty$. Again, by Theorem 6.10 of Pötscher and Prucha (1991) this limit exists in r-th mean with $\sup_i \sup_{t\geq 0} E\left|y_{it}\right|^r < \infty$. Then, by Lemma (17), the sequence y_{it} is NED on z_{it}^* with $v(m) = O(b^{-m})$ for some $0 \leq b < 1$. It follows immediately form Definition 1 that x_{it} is NED on z_{it}^* of the same order. By stationarity and the moment bound x_{it} has a tight distribution. Then, Theorem 6.7 of Pötscher and Prucha (1991) implies that $\xi(x_{it})$ is NED on z_{it}^* and the last assertion follows by Lemma (18). ■

Proposition 3 *Let ε_{it} be iid across i and t with $\varepsilon_{it}\tilde{}N(0,\sigma_\varepsilon^2)$ and z_{it} is strictly stationary and mixing with $\sup_i |\alpha_i(m)| \leq Ca^m$ for some a such that $0 < a < 1$ and some $C > 0$ and $\sup_i E|z_{it}|^r < \infty$ for some $r > 7 + 10q + 12 + \delta$ with $q \geq p/2 + 2$ and for some $\delta > 0$. Let $y_{it} = \max(0, \phi_0 y_{it-1} + z_{it}'\beta_0 + \gamma_{i0} + \varepsilon_{it})$ and define $x_{it} = (y_{it}, y_{it-1}, z_{it})$ and*

$$\psi(x_{it};\theta,\gamma_i) = 1\{y_{it}=0\}\log\Lambda\left(\left(\phi y_{it-1} + z_{it}'\beta + \gamma_i\right)/\sigma_\varepsilon^2\right) + 1\{y_{it}>0\}\log\lambda\left(\left(y_{it} - \phi y_{it-1} - z_{it}'\beta - \gamma_i\right)/\sigma_\varepsilon^2\right)$$

where Λ is the cumulative distribution function of the standard normal distribution and λ is the corresponding density. Then there exists a function $M(x_{it})$ such that $|D^v\psi(x_{it},\phi_1) - D^v\psi(x_{it},\phi_2)| \leq M(x_{it})\|\phi_1 - \phi_2\|$ for all $\phi_1, \phi_2 \in \Phi$ and $|v| \leq 5$, $\sup_{\phi\in\Phi}\|D^v\psi(x_{it},\phi)\| \leq M(x_{it})$ and $\sup_i E\left[|M(x_{it})|^{10q+12+\delta}\right] < \infty$ for some integer $q \geq p/2+2$ and for some $\delta > 0$. Furthermore, $\sup_{\phi\in\Phi}|D^v\psi(x_{it},\phi) - D^v\psi(x_{it}^m,\phi)| \leq C(x_{it},x_{it}^m)\|x_{it} - x_{it}^m\|$ where $x_{it}^m = E\left(x_{it}|z_{it+m}^*, ...z_{it-m}^*\right)$ with $z_{it}^* = z_{it}'\beta + \gamma_i + \varepsilon_{it}$ and $\sup_i\sup_m\sup_t E|C(x_{it},x_{it}^m)|^2 < \infty$,

$$\sup_i \sup_m \sup_t E\left(C(x_{it},x_{it}^m)\|x_{it} - x_{it}^m\|\right)^r < \infty$$

for some $r > 2$.

Proof. By Proposition 2 it follows that x_{it} can be chosen to be strictly stationary and NED with exponentially decaying coefficients and $\sup_i E\|x_{it}\|^r < \infty$ for $r > 7 + 10q + 12 + \delta$ with $q \geq p/2+2$ and for some $\delta > 0$. Consider $\xi(x) = \log\left(\int_{-\infty}^x (2\pi)^{-1/2}\exp\left(-1/2u^2\right)du\right)$, with the first five derivatives

$$\frac{d\xi(x)}{dx} = \frac{1}{2\sqrt{\pi}}\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}},$$

$$\frac{d^2\xi(x)}{(dx)^2} = -\frac{1}{2\pi}\frac{e^{2\left(-\frac{1}{2}x^2\right)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} - \frac{1}{2\sqrt{\pi}}x\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}}$$

$$\begin{aligned}\frac{d^3\xi(x)}{(dx)^3} &= \frac{1}{\pi}x\frac{e^{-x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} + \frac{1}{2\pi}x\frac{e^{2\left(-\frac{1}{2}x^2\right)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} - \frac{1}{2\sqrt{\pi}}\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}} \\ &+\frac{1}{2\sqrt{\pi}}x^2\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}} + \frac{1}{2\pi^{\frac{3}{2}}}\sqrt{2}e^{-x^2}\frac{e^{-\frac{1}{2}x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^3}\end{aligned}$$

$$\begin{aligned}\frac{d^4\xi(x)}{(dx)^4} &= \frac{3}{2\pi}\frac{e^{-x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} - \frac{3}{\pi}x^2\frac{e^{-x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} + \frac{1}{2\pi}\frac{e^{2\left(-\frac{1}{2}x^2\right)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} \\ &-\frac{1}{2\pi}x^2\frac{e^{2\left(-\frac{1}{2}x^2\right)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2} - \frac{3}{2\pi^2}e^{-x^2}\frac{e^{2\left(-\frac{1}{2}x^2\right)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^4} \\ &+\frac{3}{2\sqrt{\pi}}x\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}} - \frac{1}{2\sqrt{\pi}}x^3\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}} - \frac{3}{\pi^{\frac{3}{2}}}x\sqrt{2}e^{-x^2}\frac{e^{-\frac{1}{2}x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^3}\end{aligned}$$

$$
\begin{aligned}
\frac{d^5\xi(x)}{(dx)^5} &= -\frac{11}{\pi}x\frac{e^{-x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2}+\frac{7}{\pi}x^3\frac{e^{-x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2}+\frac{6}{\pi^2}x\frac{e^{2(-x^2)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^4} \\
&\quad -\frac{3}{2\pi}x\frac{e^{2(-\frac{1}{2}x^2)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2}+\frac{1}{2\pi}x^3\frac{e^{2(-\frac{1}{2}x^2)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^2}+\frac{9}{\pi^2}xe^{-x^2}\frac{e^{2(-\frac{1}{2}x^2)}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^4} \\
&\quad +\frac{3}{2\sqrt{\pi}}\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}}-\frac{3}{\sqrt{\pi}}x^2\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}}+\frac{1}{2\sqrt{\pi}}x^4\sqrt{2}\frac{e^{-\frac{1}{2}x^2}}{\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}} \\
&\quad -\frac{5}{\pi^{\frac{3}{2}}}\sqrt{2}e^{-x^2}\frac{e^{-\frac{1}{2}x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^3}+\frac{3}{\pi^{\frac{5}{2}}}\sqrt{2}e^{2(-x^2)}\frac{e^{-\frac{1}{2}x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^5}+\frac{25}{2\pi^{\frac{3}{2}}}x^2\sqrt{2}e^{-x^2}\frac{e^{-\frac{1}{2}x^2}}{\left(\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}\right)^3}
\end{aligned}
$$

where

$$
\frac{1}{2}\operatorname{erf}\left(\frac{1}{2}x\sqrt{2}\right)+\frac{1}{2}=\Lambda(x).
$$

Therefore, $\frac{d^{|v|}\xi(x)}{(dx)^{|v|}}$ is bounded except for $x\to-\infty$. We note that $\frac{e^{-x^2/2}}{\Lambda(x)}=O(-x)$ as $x\to-\infty$ which implies that

$$
\frac{d^{|v|}\xi(x)}{(dx)^{|v|}}=O(|x|^{|v|}) \text{ as } x\to-\infty
$$

and bounded elsewhere. We thus have the bound $\left|\frac{d^{|v|}\xi(x)}{(dx)^{|v|}}\right|\le C\,|x|^{|v|}$ for some constant C. This now implies that

$$
\sup_{\phi\in\Phi}\|D^v\psi(x_{it},\phi)\|\le C\sup_{\phi\in\Phi}\left|x_{it}'\tilde{\phi}\right|^{|v|}\le C\,\|x_{it}\|^{|v|}\sup_{\phi\in\Phi}\left|\tilde{\phi}\right|^{|v|}
$$

where $\tilde{\phi}=\left(1,\phi'\right)'$. The assertion the follows from $\sup_i E\,\|x_{it}\|^{|v|+10q+12+\delta}<\infty$ since $\sup_{\phi\in\Phi}\left|\tilde{\phi}\right|^{|v|}$ is bounded when Φ is compact. By the chain rule of differentiation and the mean value theorem we have

$$
|D^v\psi(x_{it},\phi_1)-D^v\psi(x_{it},\phi_2)|\le C\left|x_{it}'\tilde{\phi}^*\right|^{|v|+1}\|x_{it}\|\,\|\phi_1-\phi_2\|
$$

where $\left\|\tilde{\phi}^*-\tilde{\phi}_1\right\|\le\left\|\tilde{\phi}_1-\tilde{\phi}_2\right\|$ and $\left\|\tilde{\phi}^*-\tilde{\phi}_2\right\|\le\left\|\tilde{\phi}_1-\tilde{\phi}_2\right\|$ such that we again need $\sup_i E\,\|x_{it}\|^{|v|+2+10q+12+\delta}<\infty$.

For the last part we note again that by the mean value theorem

$$
|D^v\psi(x_{it},\phi)-D^v\psi(x_{it}^m,\phi)|\le\sup_{\phi\in\Phi}\|\phi\|\,C\left|x_{it}^{*\prime}\tilde{\phi}\right|^{|v|+1}\|x_{it}-x_{it}^m\|
$$

where $x_{it}^*=x_{it}+\delta\left(x_{it}^m-x_{it}\right)$ for some $\delta\in[0,1]$. The result then follows if $\sup_i E\,\|x_{it}\|^{2|v|+6+\delta}<\infty$. Since $2\,|v|+6+\delta\le16+\delta<7+10q+12+\delta$ the assumptions imply the result. ■