Bootstrap inference for fixed-effect models

Supplementary material

Ayden Higgins* — University of Oxford

Koen Jochmans† — Toulouse School of Economics

November 6, 2023

Abstract

This document contains auxiliary theorems and lemmata, with proofs, that are used in the proofs in the main text.

1 Auxiliary theorems

Theorem S.1 (Uniform asymptotic expansion)**.** *Let Assumptions 1–6 hold. Then*

$$\sqrt{nm}(\hat{\varphi} - \varphi) = \frac{1}{\sqrt{nm}} \sum_{i=1}^{n} \sum_{t=1}^{m} \Omega_{nm}^{-1} u_{it} + \sqrt{\frac{n}{m}} \beta_{nm} + e_{nm}$$

where the three right-hand side terms are, respectively, $O_P(1)$*,* $O(1)$*, and* $o_P(1)$ *uniformly in* $\theta \in \Theta_1$*.*

Proof. Let z_{it} be generated with parameters φ_0, η_{i0} and collect φ_0 and all η_{i0} in the vector θ_0. For notational simplicity we will presume throughout this proof that both φ_0 and η_{i0} are scalars. The proof follows the same strategy as those in Hahn and Newey (2004) and Hahn and Kuersteiner (2011), with the main difference being that we show the result to hold uniformly in a neighborhood around θ_0.

*Address: Department of Economics, University of Oxford, 10 Manor Road, Oxford OX1 3UQ, United Kingdom. E-mail: `ayden.higgins@economics.ox.ac.uk`.

†Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: `koen.jochmans@tse-fr.eu`.

Let

$$v(\varphi,\eta_i|z_{it}) \coloneqq \frac{\partial \ell(\varphi,\eta_i|z_{it})}{\partial \eta_i}, \qquad w(\varphi,\eta_i|z_{it}) \coloneqq \frac{\partial \ell(\varphi,\eta_i|z_{it})}{\partial \varphi},$$

and, with the projection coefficient $\rho_{i,m}$ defined in the main text,

$$u(\varphi,\eta_i|z_{it}) \coloneqq w(\varphi,\eta_i|z_{it}) - \rho_{i,m}\, v(\varphi,\eta_i|z_{it}).$$

We will let $v_{it} \coloneqq v(\varphi_0,\eta_{i0}|z_{it})$ and $u_{it} \coloneqq u(\varphi_0,\eta_{i0}|z_{it})$. We will use a similar shorthand for derivatives, for example, $v_{it}^{\eta_i} \coloneqq {}^{\partial v(\varphi_0,\eta_{i0}|z_{it})}/_{\partial \eta_i}$, $v_{it}^{\eta_i\eta_i} \coloneqq {}^{\partial^2 v(\varphi_0,\eta_{i0}|z_{it})}/_{\partial \eta_i^2}$, and so on. Let F_{it} be the distribution of z_{it}, write $\hat{F}_{it}$ for the corresponding empirical distribution, and consider linear combinations of the form

$$G_{it}(z|\epsilon) \coloneqq F_{it}(z) + \epsilon\sqrt{m}(\hat{F}_{it}(z) - F_{it}(z)),$$

where $0 \leq \epsilon \leq m^{-1/2}$. For fixed values φ and ϵ, let $\eta_i(\varphi,\epsilon)$ satisfy

$$\sum_{t=1}^{m} \int v(\varphi,\eta_i(\varphi,\epsilon)|z)\, dG_{it}(z|\epsilon) = 0. \tag{S.1}$$

Similarly, for fixed ϵ, let $\varphi(\epsilon)$ satisfy

$$\sum_{i=1}^{n}\sum_{t=1}^{m} \int u(\varphi(\epsilon),\eta_i(\varphi(\epsilon),\epsilon)|z)\, dG_{it}(z|\epsilon) = 0. \tag{S.2}$$

Note that setting $\epsilon = m^{-1/2}$ gives the maximum-likelihood estimator, $\hat{\theta}$, while setting $\epsilon = 0$ gives θ_0. By an expansion around $\epsilon = 0$,

$$\varphi(m^{-1/2}) - \varphi(0) = \frac{1}{\sqrt{m}}\frac{\partial \varphi(0)}{\partial \epsilon} + \frac{1}{2!}\left(\frac{1}{\sqrt{m}}\right)^2 \frac{\partial^2 \varphi(0)}{\partial \epsilon^2} + \frac{1}{3!}\left(\frac{1}{\sqrt{m}}\right)^3 \frac{\partial^3 \varphi(\tilde{\epsilon})}{\partial \epsilon^3} \tag{S.3}$$

for some $0 \leq \tilde{\epsilon} \leq m^{-1/2}$. We now investigate each of the three right-hand side terms, in turn.

For the first term, to calculate ${}^{\partial\varphi(0)}/_{\partial\epsilon}$, differentiate the expression in (S.2) with respect to ϵ to obtain

$$0 = \sum_{i=1}^{n}\sum_{t=1}^{m} \int \frac{\partial \bar{u}(\epsilon|z)}{\partial \epsilon}\, dG_{it}(z|\epsilon) + \sum_{i=1}^{n}\sum_{t=1}^{m} \int \bar{u}(\epsilon|z)\, d\,\frac{\partial G_{it}(z|\epsilon)}{\partial \epsilon} \tag{S.4}$$

where $\bar{u}(\epsilon|z) := u(\varphi(\epsilon), \eta_i(\varphi(\epsilon), \epsilon)|z)$. With a minor abuse of notation,

$$\frac{\partial \bar{u}(\epsilon|z)}{\partial \epsilon} = \frac{\partial u(\varphi(\epsilon), \eta_i(\varphi(\epsilon), \epsilon)|z)}{\partial \varphi} \frac{\partial \varphi(\epsilon)}{\partial \epsilon} + \frac{\partial u(\varphi(\epsilon), \eta_i(\varphi(\epsilon), \epsilon)|z)}{\partial \eta_i} \left(\frac{\partial \eta_i(\varphi(\epsilon), \epsilon)}{\partial \varphi} \frac{\partial \varphi(\epsilon)}{\partial \epsilon} + \frac{\partial \eta_i(\varphi(\epsilon), \epsilon)}{\partial \epsilon} \right),$$

and

$$\frac{\partial G_{it}(z|\epsilon)}{\partial \epsilon} = \sqrt{m}(\hat{F}_{it} - F_{it}).$$

Evaluating (S.4) at $\epsilon = 0$ and exploiting that

$$\sum_{t=1}^{m} \int \frac{\partial u(\varphi(0), \eta_i(\varphi(0), 0)|z)}{\partial \eta_i} \, dG_{it}(z|0) = \sum_{t=1}^{m} \mathbb{E}_{\theta_0}\left(w_{it}^{\eta_i}\right) - \rho_{i,m}\, \mathbb{E}_{\theta_0}\left(v_{it}^{\eta_i}\right)$$

is zero by definition of $\rho_{i,m}$ we may re-arrange (S.4) to obtain

$$\frac{\partial \varphi(0)}{\partial \epsilon} = \frac{1}{n\sqrt{m}} \sum_{i=1}^{n} \sum_{t=1}^{m} \mathit{\Omega}_{nm}^{-1} \, u_{it}, \tag{S.5}$$

where we have used the definition of $\mathit{\Omega}_{nm}$ given in the main text along with the fact that Assumption 6 guarantees its inverse is well-defined, and we have exploited the observation that $\mathbb{E}_{\theta_0}(u_{it}) = 0$. By Markov's inequality, the independence of the data over i, and a strong-mixing inequality (Doukhan, 1994, pp. 25–30), we have

$$\begin{aligned}
\sup_{\theta_0 \in \Theta_1} \mathbb{P}_{\theta_0}\left(\left| \sqrt{n}\, \frac{\partial \varphi(0)}{\partial \epsilon} \right|^2 > \varepsilon \right) &= \sup_{\theta_0 \in \Theta_1} \mathbb{P}_{\theta_0}\left(\left| \frac{1}{\sqrt{nm}} \sum_{i=1}^{n} \sum_{t=1}^{m} \mathit{\Omega}_{nm}^{-1} \, u_{it} \right|^2 > \varepsilon \right) \\
&\leq \frac{1}{nm\varepsilon} \sup_{\theta_0 \in \Theta_1} \mathbb{E}_{\theta_0}\left(\left| \sum_{i=1}^{n} \sum_{t=1}^{m} \mathit{\Omega}_{nm}^{-1} \, u_{it} \right|^2 \right) \\
&\leq \frac{1}{m\varepsilon} \max_{1 \leq i \leq n} \sup_{\theta_0 \in \Theta_1} \mathbb{E}_{\theta_0}\left(\left| \sum_{t=1}^{m} \mathit{\Omega}_{nm}^{-1} \, u_{it} \right|^2 \right) \\
&\lesssim \frac{1}{\varepsilon},
\end{aligned}$$

so that ${}^{\partial \varphi(0)}/_{\partial \epsilon} = O_P(n^{-1/2})$ uniformly over Θ_1. Here and later we use $A \lesssim B$ to indicate that there exists a finite constant c such that $A \leq c\, B$.

Before calculating ${}^{\partial^2\varphi(0)}/_{\partial\epsilon^2}$ we observe that differentiating (S.1) with respect to φ gives

$$\sum_{t=1}^{m}\int \frac{\partial v(\varphi,\eta_i(\varphi,\epsilon)|z)}{\partial\varphi}\,dG_{it}(z|\epsilon)+\sum_{t=1}^{m}\int \frac{\partial v(\varphi,\eta_i(\varphi,\epsilon)|z)}{\partial\eta_i}\,dG_{it}(z|\epsilon)\,\frac{\partial\eta_i(\varphi,\epsilon)}{\partial\varphi}=0.$$

Re-arranging and evaluating at $\epsilon=0$ yields

$$\frac{\partial\eta_i(\varphi,0)}{\partial\varphi}=-\,\mathbb{E}_{\theta_0}\left(\frac{1}{m}\sum_{t=1}^{m}v_{it}^{\varphi}\right)\Bigg/\,\mathbb{E}_{\theta_0}\left(\frac{1}{m}\sum_{t=1}^{m}v_{it}^{\eta_i}\right)=-\rho_{i,m}.$$

In the same way, differentiating (S.1) with respect to ϵ reveals that

$$\frac{\partial\eta_i(\varphi,0)}{\partial\epsilon}=-\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m}v_{it}\right)\Bigg/\,\mathbb{E}_{\theta_0}\left(\frac{1}{m}\sum_{t=1}^{m}v_{it}^{\eta_i}\right)=:\psi_{i,m},$$

which is the asymptotically-linear representation of the maximum-likelihood estimator of η_{i0}. With these expressions at hand we turn to ${}^{\partial^2\varphi(0)}/_{\partial\epsilon^2}$. Differentiating (S.4) again with respect to ϵ gives

$$0=\sum_{i=1}^{n}\sum_{t=1}^{m}\int\frac{\partial^2\bar{u}(\epsilon|z)}{\partial\epsilon^2}\,dG_{it}(z|\epsilon)+2\sum_{i=1}^{n}\sum_{t=1}^{m}\int\frac{\partial\bar{u}(\epsilon|z)}{\partial\epsilon}\,d\,\frac{\partial G_{it}(z|\epsilon)}{\partial\epsilon},\tag{S.6}$$

where the second derivative of $\bar{u}(\epsilon|z)$ follows from the chain rule and consists of many terms. Evaluating each of these terms at $\epsilon=0$, re-arranging, and recalling again the expression for ${}^{\partial G_{it}(z|0)}/_{\partial\epsilon}$ and the fact that $\mathbb{E}_{\theta_0}(u_{it}^{\eta_i})=0$ gives

$$\left(-\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi}\right)\right)\frac{\partial^2\varphi(0)}{\partial\epsilon^2}=2\sum_{i=1}^{n}\left(\sqrt{m}\sum_{t=1}^{m}u_{it}^{\eta_i}\right)\psi_{i,m}+\frac{1}{2}\left(\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\eta_i\eta_i}\right)\right)\psi_{i,m}^2+r_{nm}$$

with

$$\begin{aligned}r_{nm}:=&\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi\varphi}\right)\left(\frac{\partial\varphi(0)}{\partial\epsilon}\right)^2\\&+2\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi\eta_i}\right)\frac{\partial\varphi(0)}{\partial\epsilon}\left(\psi_{i,m}-\frac{\partial\varphi(0)}{\partial\epsilon}\,\rho_{i,m}\right)\\&-2\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\eta_i\eta_i}\right)\frac{\partial\varphi(0)}{\partial\epsilon}\left(\psi_{i,m}-\frac{1}{2}\frac{\partial\varphi(0)}{\partial\epsilon}\,\rho_{i,m}\right)\rho_{i,m}\\&+2\sum_{i=1}^{n}\sum_{t=1}^{m}\left(u_{it}^{\varphi}-\mathbb{E}_{\theta_0}(u_{it}^{\varphi})-\rho_{i,m}\,u_{it}^{\eta_i}\right)\frac{\partial\varphi(0)}{\partial\epsilon}\,\sqrt{m}.\end{aligned}$$

Each term on the right-hand side of this expression will be asymptotically negligible for our purposes. For example,

$$\left|\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi\varphi}\right)\left(\frac{\partial\varphi(0)}{\partial\epsilon}\right)^2\right| \leq \sum_{i=1}^{n}\sum_{t=1}^{m}\left|\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi\varphi}\right)\right|\left|\left(\frac{\partial\varphi(0)}{\partial\epsilon}\right)^2\right| = O(nm)\,O_P(n^{-1})$$

and, therefore, $O_P(m)$, uniformly over Θ_1 by the moment requirements in Assumption 3 and the convergence rate on ${}^{\partial\varphi(0)}/_{\partial\epsilon}$ obtained above. Similarly, using the definition of $\psi_{i,m}$ together with Assumptions 3 and 6 we obtain, by the same arguments as those employed below (S.5), that

$$\left|\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi\eta_i}\right)\,\psi_{i,m}\right| = \left|\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(\frac{1}{m}\sum_{t=1}^{m}u_{it}^{\varphi\eta_i}\right)\left(\mathbb{E}_{\theta_0}\left(\frac{1}{m}\sum_{t=1}^{m}v_{it}^{\eta_i}\right)\right)^{-1}\sqrt{m}\sum_{t=1}^{m}v_{it}\right|$$

is $O_P(\sqrt{n}m)$ uniformly over Θ_1. Hence,

$$\left|\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}\left(u_{it}^{\varphi\eta_i}\right)\,\psi_{i,m}\,\frac{\partial\varphi(0)}{\partial\epsilon}\right| = O_P(m)$$

uniformly over Θ_1. The remaining terms that make up r_{nm} can be dealt with in a similar way. Consequently, letting

$$\chi_{i,m} := \left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m}u_{it}^{\eta_i}\right)\,\psi_{i,m} + \frac{1}{2}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(u_{it}^{\eta_i\eta_i})\right)\,\psi_{i,m}^2,$$

we have shown that

$$\Omega_{nm}\,\frac{\partial^2\varphi(0)}{\partial\epsilon^2} = \frac{2}{n}\sum_{i=1}^{n}\chi_{i,m} + O_P(n^{-1}),$$

where the order of the remainder term is uniformly over Θ_1. We next establish that, uniformly over Θ_1,

$$\frac{1}{n}\sum_{i=1}^{n}\chi_{i,m} = b_{nm} + o_P(1), \qquad b_{nm} := \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}(\chi_{i,m}),$$

that is, we demonstrate

$$\sup_{\theta_0\in\Theta_1}\mathbb{P}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))\right|^2 > \varepsilon\right) = o(1) \tag{S.7}$$

for any $\varepsilon > 0$. By Markov's inequality and independence of the observations across i we have

$$\begin{aligned}
\sup_{\theta_0 \in \Theta_1} \mathbb{P}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))\right|^2 > \varepsilon\right) &\leq \frac{1}{\varepsilon}\sup_{\theta_0 \in \Theta_1} \mathbb{E}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))\right|^2\right) \\
&\leq \frac{1}{\varepsilon n^2}\sum_{i=1}^{n}\sup_{\theta_0 \in \Theta_1} \mathbb{E}_{\theta_0}\left((\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))^2\right) \\
&\leq \frac{1}{\varepsilon n}\max_{1 \leq i \leq n}\sup_{\theta_0 \in \Theta_1} \mathbb{E}_{\theta_0}\left((\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))^2\right),
\end{aligned}$$

and so it suffices to show that

$$\max_{1 \leq i \leq n}\sup_{\theta_0 \in \Theta_1} \mathbb{E}_{\theta_0}\left((\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))^2\right) = o(n).$$

To begin we use the expression for $\psi_{i,m}$ to re-write $\chi_{i,m}$ as

$$\begin{aligned}
\chi_{i,m} = &-\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(v_{it}^{\eta_i})\right)^{-1}\left(\frac{1}{m}\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}u_{it_1}^{\eta_i}v_{it_2}\right) \\
&+\frac{1}{2}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(v_{it}^{\eta_i})\right)^{-2}\left(\frac{1}{m}\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}v_{it_1}v_{it_2}\right)\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(u_{it}^{\eta_i\eta_i})\right),
\end{aligned}$$

and introduce the shorthand notation

$$\zeta_{i,m} \coloneqq \left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(v_{it}^{\eta_i})\right)^{-1}, \qquad \xi_{i,m} \coloneqq \frac{1}{2}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(v_{it}^{\eta_i})\right)^{-2}\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_{\theta_0}(u_{it}^{\eta_i\eta_i})\right),$$

both of which are well-behaved under our assumptions. Then

$$\chi_{i,m} = \xi_{i,m}\left(\frac{1}{m}\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}v_{it_1}v_{it_2}\right) - \zeta_{i,m}\left(\frac{1}{m}\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}u_{it_1}^{\eta_i}v_{it_2}\right),$$

and so

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left((\chi_{i,m} - \mathbb{E}_{\theta_0}(\chi_{i,m}))^2\right) = &\frac{\zeta_{i,m}^2}{m^2}\sum_{t_1,\ldots,t_4}\left(\mathbb{E}_{\theta_0}(u_{it_1}^{\eta_i}v_{it_2}u_{it_3}^{\eta_i}v_{it_4}) - \mathbb{E}_{\theta_0}(u_{it_1}^{\eta_i}v_{it_2})\mathbb{E}_{\theta_0}(u_{it_3}^{\eta_i}v_{it_4})\right) \\
&+\frac{\xi_{i,m}^2}{m^2}\sum_{t_1,\ldots,t_4}\left(\mathbb{E}_{\theta_0}(v_{it_1}v_{it_2}v_{it_3}v_{it_4}) - \mathbb{E}_{\theta_0}(v_{it_1}v_{it_2})\mathbb{E}_{\theta_0}(v_{it_3}v_{it_4})\right) \\
&-\frac{2\xi_{i,m}\zeta_{i,m}}{m^2}\sum_{t_1,\ldots,t_4}\left(\mathbb{E}_{\theta_0}(u_{it_1}^{\eta_i}v_{it_2}v_{it_3}v_{it_4}) - \mathbb{E}_{\theta_0}(u_{it_1}^{\eta_i}v_{it_2})\mathbb{E}_{\theta_0}(v_{it_3}v_{it_4})\right).
\end{aligned}$$

Take the first term on the right-hand side. We have

$$\begin{aligned}\mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2} u^{\eta_i}_{it_3} v_{it_4}) - \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2})\mathbb{E}_{\theta_0}(u^{\eta_i}_{it_3} v_{it_4}) &= \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} u^{\eta_i}_{it_3})\,\mathbb{E}_{\theta_0}(v_{it_2} v_{it_4}) \\ &\quad + \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2})\,\mathbb{E}_{\theta_0}(u^{\eta_i}_{it_3} v_{it_4}) \\ &\quad + \mathrm{cum}_4(u^{\eta_i}_{it_1}, v_{it_2}, u^{\eta_i}_{it_3}, v_{it_4}),\end{aligned}$$

where cum_4 refers to the fourth-order cumulant of the joint distribution of its arguments. As in Hahn and Kuersteiner (2011), Assumptions 2 and 3 allow us to apply Corollary A.2 of Hall and Heyde (1980) to obtain

$$\sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} |\mathbb{E}_{\theta_0}(v_{it_2} v_{it_4})| \lesssim \max_{1\le i\le n} \sup_{\theta_0\in\Theta_1} \alpha_i(\theta_0, |t_2 - t_4|) = O(r^{|t_2-t_4|}),$$

where $0 < r < 1$, and, therefore,

$$\sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \sum_{t_2=1}^{m} \sum_{t_4=1}^{m} |\mathbb{E}_{\theta_0}(v_{it_2} v_{it_4})| = O(m).$$

In the same way we obtain

$$\begin{aligned}\sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \sum_{t_1=1}^{m} \sum_{t_3=1}^{m} |\mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} u^{\eta_i}_{it_3})| &= O(m), \\ \sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \sum_{t_1=1}^{m} \sum_{t_2=1}^{m} |\mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2})| &= O(m),\end{aligned}$$

whereas, from Andrews (1991, Lemma 1),

$$\sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \left| \sum_{t_1,\ldots,t_4} \mathrm{cum}_4(u^{\eta_i}_{it_1}, v_{it_2}, u^{\eta_i}_{it_3}, v_{it_4}) \right| = O(m^2).$$

It follows that

$$\sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \left| \frac{1}{m^2} \sum_{t_1,\ldots,t_4} \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2} u^{\eta_i}_{it_3} v_{it_4}) - \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2})\mathbb{E}_{\theta_0}(u^{\eta_i}_{it_3} v_{it_4}) \right| = O(1).$$

In the same way,

$$\begin{aligned}\sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \left| \frac{1}{m^2} \sum_{t_1,\ldots,t_4} \mathbb{E}_{\theta_0}(v_{it_1} v_{it_2} v_{it_3} v_{it_4}) - \mathbb{E}_{\theta_0}(v_{it_1} v_{it_2})\mathbb{E}_{\theta_0}(v_{it_3} v_{it_4}) \right| &= O(1), \\ \sup_{\theta_0\in\Theta_1} \max_{1\le i\le n} \left| \frac{1}{m^2} \sum_{t_1,\ldots,t_4} \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2} v_{it_3} v_{it_4}) - \mathbb{E}_{\theta_0}(u^{\eta_i}_{it_1} v_{it_2})\mathbb{E}_{\theta_0}(v_{it_3} v_{it_4}) \right| &= O(1),\end{aligned}$$

which is more than enough to imply that

$$\max_{1\le i\le n}\sup_{\theta_0\in\Theta_1}\mathbb{E}_{\theta_0}\left((\chi_{i,m}-\mathbb{E}_{\theta_0}(\chi_{i,m}))^2\right)=o(n),$$

so that (S.7) holds. Thus,

$$\frac{\partial^2\varphi(0)}{\partial\epsilon^2}=\frac{2}{n}\sum_{i=1}^n\Omega_{nm}^{-1}\,\mathbb{E}_{\theta_0}(\chi_{i,m})+o_P(1)=2\Omega_{nm}^{-1}\,b_{nm}+o_P(1)=2\beta_{nm}+o_P(1)\tag{S.8}$$

uniformly over Θ_1.

Finally, following the same arguments as those in the supplementary appendices to Hahn and Newey (2004) (using suitably uniform versions of Lemmas 5 to 11 of Hahn and Kuersteiner 2011, which may be shown by relying on our Lemmas S.1 and S.2) we obtain

$$\sup_{\theta_0\in\Theta_1}\mathbb{P}_\theta\left(\max_{0\le\epsilon\le m^{-1/2}}\left|\frac{\partial^3\varphi(\epsilon)}{\partial\epsilon^3}\right|>c\,m^{3s}\right)=o(m^{-1})$$

for some finite $c>0$ and $0<s<1/10$. This implies that $\partial^3\varphi(\epsilon)/\partial\epsilon^3=O_P(m^{3s})$ uniformly in $0\le\epsilon\le m^{-1/2}$, and so

$$\left(\frac{1}{\sqrt{m}}\frac{\partial^3\varphi(\tilde{\epsilon})}{\partial\epsilon^3}\right)=o_P(1)$$

uniformly in $\theta_0\in\Theta_1$.

Then, combining the expansion in (S.3) with the expressions obtained in (S.5) and (S.8) we find that

$$\sqrt{nm}(\hat{\varphi}-\varphi_0)=\frac{1}{\sqrt{nm}}\sum_{i=1}^n\sum_{t=1}^m\Omega_{nm}^{-1}\,u_{it}+\sqrt{\frac{n}{m}}\beta_{nm}+e_{nm}$$

where (uniformly over Θ_1) the first term on the right-hand side has been shown to be $O_P(1)$, the second term satisfies

$$\begin{aligned}\sup_{\theta_0\in\Theta_1}|\beta_{nm}|\le&\left(\sup_{\theta_0\in\Theta_1}\max_{1\le i\le n}|\xi_{i,m}|\right)\left(\sup_{\theta_0\in\Theta_1}\max_{1\le i\le n}\frac{1}{m}\sum_{t_1=1}^m\sum_{t_2=1}^m|\mathbb{E}_{\theta_0}(v_{it_1}v_{it_2})|\right)\\&+\left(\sup_{\theta_0\in\Theta_1}\max_{1\le i\le n}|\zeta_{i,m}|\right)\left(\sup_{\theta_0\in\Theta_1}\max_{1\le i\le n}\frac{1}{m}\sum_{t_1=1}^m\sum_{t_2=1}^m|\mathbb{E}_{\theta_0}(u_{it_1}^{\eta_i}v_{it_2})|\right)=O(1)\end{aligned}$$

under our assumptions by another application of Hall and Heyde (1980, Corollary A.2), and the remainder term is $o_P(1)$. This completes the proof. □

Theorem S.2 (Uniform asymptotic normality)**.** *Let Assumptions 1–6 hold. Then*

$$\sup_{\theta\in\Theta_1} \left|\mathbb{P}_\theta(\sqrt{nm}(\hat{\varphi} \ -\varphi) \leq a) - \mathbb{P}_\theta(v_\theta \leq a)\right| = o(1)$$

for any a.

Proof. From Theorem S.1,

$$\sqrt{nm}(\hat{\varphi}-\varphi) = \frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m} \varOmega_{nm}^{-1}\, u_{it} + \sqrt{\frac{n}{m}}\beta_{nm} + e_{nm},$$

where

$$\sup_{\theta\in\Theta_1} \mathbb{P}_\theta(\|e_{nm}\|_2 > \varepsilon) = o(1).$$

We first show that

$$\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m} \varOmega_{nm}^{-1}\, u_{it} \xrightarrow{L} N(0,\varSigma_\theta) \tag{S.9}$$

uniformly in $\theta\in\Theta_1$. To do so, by the Cramer-Wold device, it suffices to show that, for any (non-random) vector c of conformable dimension, ${}^{1}/_{\sqrt{nm}} \sum_{i=1}^{n}\sum_{t=1}^{m} c'\varOmega_{nm}^{-1}\, u_{it} \xrightarrow{L} N(0, c'\varSigma_\theta\, c)$ holds uniformly in $\theta\in\Theta_1$. Let

$$w_i := \frac{\varOmega_{nm}^{-1}}{\sqrt{n}}\frac{\sum_{t=1}^{m} u_{it}}{\sqrt{m}}$$

so that

$$\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m} c'\varOmega_{nm}^{-1}\, u_{it} = \sum_{i=1}^{n} c'w_i.$$

By the mean-zero property of the efficient score and the information equality, respectively,

$$\mathbb{E}_\theta(c'w_i) = 0, \qquad \sigma_{nm}^2 := \sum_{i=1}^{n}\mathbb{E}_\theta(c'w_i w_i' c) = c'\varOmega_{nm}^{-1}\, c = O(1),$$

uniformly in $\theta\in\Theta_1$. The Berry-Esseen inequality gives

$$\sup_a \left|\mathbb{P}_\theta\left(\sum_{i=1}^{n}\frac{c'w_i}{\sigma_{nm}} \leq a\right) - \varPhi\left(a\right)\right| \lesssim \sum_{i=1}^{n}\mathbb{E}_\theta(|c'w_i|^3)\left(\sum_{i=1}^{n}\mathbb{E}_\theta(|c'w_i|^2)\right)^{-3/2} \lesssim \sum_{i=1}^{n}\mathbb{E}_\theta(|c'w_i|^3).$$

The mixing condition in Assumption 2 and the moment requirements in Assumption 3 guarantee that

$$\sum_{i=1}^{n}\mathbb{E}_\theta(|c'w_i|^3) = O(n^{-1/2}),$$

uniformly in $\theta \in \Theta_1$. Therefore,

$$\sup_{\theta\in\Theta_1}\sup_a \left|\mathbb{P}_\theta\left(\sum_{i=1}^n \frac{c'w_i}{\sigma_{nm}} \leq a\right) - \Phi(a)\right| \lesssim \sup_{\theta\in\Theta_1}\left(\sum_{i=1}^n \mathbb{E}_\theta(|c'w_i|^3)\right) = o(1). \tag{S.10}$$

Next, the continuous-mapping theorem, together with the fact that, under our assumptions,

$$\lim_{n,m\to\infty} \Omega_{nm}^{-1} = \Sigma_\theta$$

yields

$$\sup_{\theta\in\Theta_1}\sup_a \left|\mathbb{P}_\theta\left(\sum_{i=1}^n \frac{c'w_i}{\sqrt{c'\Sigma_\theta\, c}} \leq a\right) - \Phi(a)\right| = o(1),$$

from which (S.9) follows.

As this result is uniform in a, and $\sqrt{n/m}\beta_{nm} = \gamma\beta_\theta + o(1)$ uniformly in $\theta \in \Theta_1$, we find

$$\sup_{\theta\in\Theta_1}\sup_a \left|\mathbb{P}_\theta\left(\sum_{i=1}^n \frac{c'w_i}{\sqrt{c'\Sigma_\theta\, c}} \leq a - \sqrt{n/m}\, c'\beta_{nm}\right) - \Phi(a - \gamma c'\beta_\theta)\right| = o(1), \tag{S.11}$$

which accounts for the asymptotic bias in the limit distribution.

Finally, an application of Lemma S.4 with

$$x_{nm} = \frac{1}{\sqrt{nm}}\sum_{i=1}^n\sum_{t=1}^m \Omega_{nm}^{-1}\, u_{it} + \sqrt{\frac{n}{m}}\beta_{nm}, \qquad y_{nm} = x_{nm} + e_{nm},$$

and $z \sim N(\beta_\theta, \Sigma_\theta)$ yields the result of the theorem. □

Theorem S.3 (Uniform consistency of the plug-in estimator of the information matrix)**.** *Let Assumptions 1–6 hold. Then*

$$\sup_{\theta\in\Theta_1} \mathbb{P}_\theta\left(\left\|\hat{\Omega}_{nm,\theta} - \Omega_{nm,\theta}\right\|_2 > \varepsilon\right) = o(1)$$

for any $\varepsilon > 0$.

Proof. We introduce the notational shorthand

$$V_{it} := \begin{pmatrix} V_{it}^{11} & V_{it}^{12} \\ V_{it}^{21} & V_{it}^{22} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 \ell(\varphi,\eta_i|z_{it})}{\partial\varphi\partial\varphi'} & \frac{\partial^2 \ell(\varphi,\eta_i|z_{it})}{\partial\varphi\partial\eta_i'} \\ \frac{\partial^2 \ell(\varphi,\eta_i|z_{it})}{\partial\eta_i\partial\varphi'} & \frac{\partial^2 \ell(\varphi,\eta_i|z_{it})}{\partial\eta_i\partial\eta_i'} \end{pmatrix},$$

where the derivatives are evaluated at the parameter values that were used to generate the data. In the same manner, we write the plug-in estimator constructed using $\hat{\varphi}, \hat{\eta}_i$ as $\hat{V}_{it}$. Then

$$\mathit{\Omega}_{nm,\theta} = -\frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\mathbb{E}_\theta(V_{it}^{11}) - \left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_\theta(V_{it}^{12})\right)\left(\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}_\theta(V_{it}^{22})\right)^{-1}\mathbb{E}_\theta(V_{it}^{21})\right),$$

and its plug-in estimator is

$$\hat{\mathit{\Omega}}_{nm,\theta} \coloneqq -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{t=1}^{m}\hat{V}_{it}^{11} - \left(\frac{1}{m}\sum_{t=1}^{m}\hat{V}_{it}^{12}\right)\left(\frac{1}{m}\sum_{t=1}^{m}\hat{V}_{it}^{22}\right)^{-1}\frac{1}{m}\sum_{t=1}^{m}\hat{V}_{it}^{21}\right).$$

To show Theorem S.3 it suffices to establish that, for all $\varepsilon > 0$,

$$\sup_{\theta\in\mathit{\Theta}_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^{m}(\hat{V}_{it}^{11} - \mathbb{E}_\theta(V_{it}^{11}))\right\|_2 > \varepsilon\right) = o(1),$$
$$\sup_{\theta\in\mathit{\Theta}_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^{m}(\hat{V}_{it}^{12} - \mathbb{E}_\theta(V_{it}^{12}))\right\|_2 > \varepsilon\right) = o(1),$$
$$\sup_{\theta\in\mathit{\Theta}_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^{m}(\hat{V}_{it}^{22} - \mathbb{E}_\theta(V_{it}^{22}))\right\|_2 > \varepsilon\right) = o(1).$$

The proof for each of the four terms is similar and so we only provide details for the first of them.

To begin we note that

$$\sup_{\theta\in\mathit{\Theta}_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^{m}(\hat{V}_{it}^{11} - \mathbb{E}_\theta(V_{it}^{11}))\right\|_2 > \varepsilon\right)$$

is bounded from above by

$$\sup_{\theta\in\mathit{\Theta}_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^{m}(\hat{V}_{it}^{11} - V_{it}^{11})\right\|_2 > \frac{\varepsilon}{2}\right) + \sup_{\theta\in\mathit{\Theta}_1}\mathbb{P}_\theta\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^{m}(V_{it}^{11} - \mathbb{E}_\theta(V_{it}^{11}))\right\|_2 > \frac{\varepsilon}{2}\right).$$

To deal with the first of these terms let $\tilde{V}_{it}^{111}$ be the vector that collects all third-order derivatives with respect to φ and let $\tilde{V}_{it}^{112}$ denote derivatives with respect to φ (twice) and η_i. The tilde is used to indicate that these derivatives are evaluated at values $(\tilde{\varphi}, \tilde{\eta}_i)$ that (elementwise) lie between $(\hat{\varphi}, \hat{\eta}_i)$ and (φ, η_i). A mean-value expansion around (φ, η_i) yields

$$\begin{aligned}\left\| \frac{1}{m} \sum_{t=1}^{m} (\hat{V}_{it}^{11} - V_{it}^{11}) \right\|_2 &\leq \frac{1}{m} \sum_{t=1}^{m} \left\| \hat{V}_{it}^{11} - V_{it}^{11} \right\|_2 \\ &\leq \frac{1}{m} \sum_{t=1}^{m} \left\| \tilde{V}_{it}^{111} \right\|_2 \|\hat{\varphi} - \varphi\|_2 + \frac{1}{m} \sum_{t=1}^{m} \left\| \tilde{V}_{it}^{112} \right\|_2 \|\hat{\eta}_i - \eta_i\|_2 \\ &\leq \frac{1}{m} \sum_{t=1}^{m} \left\| \tilde{V}_{it}^{111} \right\|_1 \|\hat{\varphi} - \varphi\|_2 + \frac{1}{m} \sum_{t=1}^{m} \left\| \tilde{V}_{it}^{112} \right\|_1 \|\hat{\eta}_i - \eta_i\|_2.\end{aligned}$$

The uniform bound on the derivatives in Assumption 3(ii) implies that

$$\frac{1}{m} \sum_{t=1}^{m} \left\| \tilde{V}_{it}^{111} \right\|_1 \lesssim \frac{1}{m} \sum_{t=1}^{m} b(z_{it}),$$
$$\frac{1}{m} \sum_{t=1}^{m} \left\| \tilde{V}_{it}^{112} \right\|_1 \lesssim \frac{1}{m} \sum_{t=1}^{m} b(z_{it}).$$

Therefore,

$$\max_{1 \leq i \leq n} \left\| \frac{1}{m} \sum_{t=1}^{m} (\hat{V}_{it}^{11} - V_{it}^{11}) \right\|_2 \lesssim \left(\max_{1 \leq i \leq n} \frac{1}{m} \sum_{t=1}^{m} b(z_{it}) \right) \left(\|\hat{\varphi} - \varphi\|_2 + \max_{1 \leq i \leq n} \|\hat{\eta}_i - \eta_i\|_2 \right).$$

Now, the mixing conditions in Assumption 2 and the moment conditions on the bounding function b in Assumption 3(iii) imply that

$$\sup_{\theta \in \Theta_1} \mathbb{P}_\theta \left(\max_{1 \leq i \leq n} \left| \frac{1}{m} \sum_{t=1}^{m} (b(z_{it}) - \mathbb{E}_\theta(b(z_{it}))) \right| > \varepsilon \right) = o(1)$$

by an application of Lemma S.1. Also, $1/m \sum_{t=1}^{m} \mathbb{E}_\theta(b(z_{it}))$ converges to its limit uniformly over Θ_1 by Assumption 3(iv). At the same time, by Theorem 1 in Kim and Sun (2016) we have that

$$\sup_{\theta \in \Theta_1} \mathbb{P}_\theta \left(\|\hat{\varphi} - \varphi\|_2 > \varepsilon \right) = o(1), \qquad \sup_{\theta \in \Theta_1} \mathbb{P}_\theta \left(\max_{1 \leq i \leq n} \|\hat{\eta}_i - \eta_i\|_2 > \varepsilon \right) = o(1).$$

Taken together these results yield

$$\sup_{\theta \in \Theta_1} \mathbb{P}_\theta \left(\max_{1 \leq i \leq n} \left\| \frac{1}{m} \sum_{t=1}^{m} (\hat{V}_{it}^{11} - V_{it}^{11}) \right\|_2 > \frac{\varepsilon}{2} \right) = o(1)$$

follows. Next, again by Assumptions 2 and 3, an application of S.2 gives

$$\sup_{\theta\in\Theta_1} \mathbb{P}_\theta\left(\max_{1\leq i\leq n} \left\| \frac{1}{m}\sum_{t=1}^{m} (V_{it}^{11} - \mathbb{E}_\theta(V_{it}^{11})) \right\|_2 > \frac{\varepsilon}{2} \right) = o(1).$$

Hence,

$$\sup_{\theta\in\Theta_1} \mathbb{P}_\theta\left(\max_{1\leq i\leq n} \left\| \frac{1}{m}\sum_{t=1}^{m} (\hat{V}_{it}^{11} - \mathbb{E}_\theta(V_{it}^{11})) \right\|_2 > \varepsilon \right) = o(1),$$

and the proof is complete. □

Theorem S.4 (Distribution of the likelihood-ratio statistic)**.** *Let Assumptions 1–6 hold. Suppose that the true parameter value lies in the interior of the set $\Theta\cap\{\varphi\in V_\varphi : \phi(\varphi)=0\}$ and that ϕ is five times continuously-differentiable on V_φ with bounded derivatives and Jacobian matrix with maximal row rank. Then*

$$\sup_{\theta\in\Theta_1} |\mathbb{P}_\theta(\hat{w}\leq a) - \mathbb{P}_\theta(w_\theta \leq a)| = o(1)$$

for any a, where w_θ has a non-central χ^2-distribution with $\dim\phi$ degrees of freedom and non-centrality parameter $\gamma\, \beta_\theta' \Phi_\varphi' (\Phi_\varphi \Sigma_\theta \Phi_\varphi')^{-1} \Phi_\varphi \beta_\theta$.

Proof. Recall the profile likelihood

$$\sum_{i=1}^{n}\sum_{t=1}^{m} \ell(\varphi, \hat{\eta}_i(\varphi)|z_{it}), \qquad \hat{\eta}_i(\varphi) \coloneqq \arg\max_{\eta_i} \sum_{t=1}^{m} \ell(\varphi,\eta_i|z_{it}).$$

By a standard expansion,

$$\sum_{i=1}^{n}\sum_{t=1}^{m} (\ell(\check{\varphi}, \hat{\eta}_i(\check{\varphi})|z_{it}) - \ell(\hat{\varphi}, \hat{\eta}_i(\hat{\varphi})|z_{it})) = \frac{1}{2}(\check{\varphi}-\hat{\varphi})' \sum_{i=1}^{n}\sum_{t=1}^{m} \frac{\partial^2 \ell(\tilde{\varphi}, \hat{\eta}_i(\tilde{\varphi})|z_{it})}{\partial\varphi\partial\varphi'} (\check{\varphi}-\hat{\varphi}),$$

where $\tilde{\varphi}$ lies (elementwise) between $\check{\varphi}$ and $\hat{\varphi}$. It is straightforward to adapt the proof of Theorem 1 of Kim and Sun (2016) to yield a consistency result for the constrained estimator. Moreover,

$$\sup_{\theta\in\Theta_1} \mathbb{P}_\theta(\|\check{\varphi}-\varphi\|_2 > \varepsilon) = o(m^{-1}), \qquad \sup_{\theta\in\Theta_1} \mathbb{P}_\theta\left(\max_{1\leq i\leq n} \|\check{\eta}_i - \eta_i\|_2 > \varepsilon\right) = o(m^{-1}),$$

for any $\varepsilon>0$. Combined with (A.3) this then equally yields

$$\sup_{\theta\in\Theta_1} \mathbb{P}_\theta(\|\tilde{\varphi}-\varphi\|_2 > \varepsilon) = o(m^{-1}), \qquad \sup_{\theta\in\Theta_1} \mathbb{P}_\theta\left(\max_{1\leq i\leq n} \|\tilde{\eta}_i - \eta_i\|_2 > \varepsilon\right) = o(m^{-1}),$$

for any $\varepsilon > 0$. Proceeding in the same manner as in the proof of Theorem S.2 readily gives

$$\frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\partial^2 \ell(\tilde{\varphi},\hat{\eta}_i(\tilde{\varphi})|z_{it})}{\partial\varphi\partial\varphi'} = -\Omega_{nm} + o_P(1)$$

uniformly on Θ_1. Consequently, by rearranging terms we obtain the conventional quadratic approximation

$$\hat{w} = \sqrt{nm}(\check{\varphi}-\hat{\varphi})'\Omega_{nm}\sqrt{nm}(\check{\varphi}-\hat{\varphi}) + o_P(1)$$

uniformly on Θ_1. By Theorem S.6,

$$\sqrt{nm}(\hat{\varphi}-\check{\varphi}) = \Omega_{nm}^{-1}\Phi'_{\varphi}(\Phi_{\varphi}\,\Omega_{nm}^{-1}\Phi'_{\varphi})^{-1}\Phi_{\varphi}\sqrt{nm}(\hat{\varphi}-\varphi) + o_P(1)$$

uniformly on Θ_1. Further, by the uniform asymptotic-normality result of Theorem S.2, $\Phi_{\varphi}\sqrt{nm}(\hat{\varphi}-\varphi) = \Phi_{\varphi}\,v_{\theta} + o_P(1)$ uniformly in $\theta \in \Theta_1$. Also, $\Phi_{\varphi}\,v_{\theta} \sim N(\gamma\,\Phi_{\varphi}\beta_{\theta}, \Phi_{\varphi}\Sigma_{\theta}\,\Phi'_{\varphi})$ and $\lim_{n,m\to\infty}\Omega_{nm}^{-1} = \Sigma_{\theta}$. Hence, $\hat{w}$ converges in distribution to a non-central χ^2-distribution with non-centrality parameter $\gamma\,\beta'_{\theta}\Phi'_{\varphi}(\Phi_{\varphi}\Sigma_{\theta}\Phi'_{\varphi})^{-1}\Phi_{\varphi}\beta_{\theta}$ uniformly in $\theta \in \Theta_1$. This completes the proof of the theorem. □

Theorem S.5 (Distribution of the score statistic)**.** *Let Assumptions 1–6 hold. Suppose that the true parameter value lies in the interior of the set* $\Theta \cap \{\varphi \in V_{\varphi} : \phi(\varphi) = 0\}$ *and that* ϕ *is five times continuously-differentiable on* V_{φ} *with bounded derivatives and Jacobian matrix with maximal row rank. Let*

$$\hat{s} := \left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\partial \ell(\check{\varphi},\hat{\eta}_i(\check{\varphi})|z_{it})}{\partial\varphi}\right)'\check{\Sigma}\left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\partial \ell(\check{\varphi},\hat{\eta}_i(\check{\varphi})|z_{it})}{\partial\varphi}\right),$$

for $\check{\Sigma}$ *the plug-in estimator of* Σ *based on the constrained maximum-likelihood estimator. Then*

$$\sup_{\theta\in\Theta_1}|\mathbb{P}_{\theta}(\hat{s}\le a) - \mathbb{P}_{\theta}(w_{\theta}\le a)| = o(1)$$

for any a*, where* w_{θ} *has a non-central* χ^2*-distribution with* $\dim\phi$ *degrees of freedom and non-centrality parameter* $\gamma\,\beta'_{\theta}\Phi'_{\varphi}(\Phi_{\varphi}\Sigma_{\theta}\Phi'_{\varphi})^{-1}\Phi_{\varphi}\beta_{\theta}$.

Proof. The Lagrangian problem associated with the constraint $\phi(\varphi) = 0$ has first-order condition

$$\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\partial \ell(\check{\varphi},\hat{\eta}_i(\check{\varphi})|z_{it})}{\partial\varphi} - \Phi'_{\check{\varphi}}\check{\lambda} = 0,$$

where $\Phi_{\check{\varphi}}$ is the Jacobian of the constraint evaluated at $\check{\varphi}$. Combining this with (S.14) gives

$$\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}\frac{\partial \ell(\check{\varphi},\hat{\eta}_i(\check{\varphi})|z_{it})}{\partial \varphi} = \frac{1}{\sqrt{nm}}\Phi_{\varphi}'\check{\lambda} = \Phi_{\varphi}'(\Phi_{\varphi}\Omega_{nm}^{-1}\Phi_{\varphi}')^{-1}\Phi_{\varphi}\sqrt{nm}(\hat{\varphi}-\varphi)+o_P(1).$$

The uniform asymptotic normality of $\sqrt{nm}(\hat{\varphi}-\varphi)$ obtained in Theorem S.2 then readily yields that $(nm)^{-1}\check{\lambda}'\Phi\Omega_{nm}^{-1}\Phi'\check{\lambda}$ has a non-central χ^2 limit distribution with non-centrality parameter $\gamma\,\beta_\theta'\Phi_{\varphi}'(\Phi_{\varphi}\Sigma_\theta\Phi_{\varphi}')^{-1}\Phi_{\varphi}\beta_\theta$.. The uniform consistency of $\check{\Sigma}$ implied by Theorem S.3 and the consistency of the constrained estimator then give the result of the theorem. □

Theorem S.6 (Uniform asymptotic expansion)**.** *Let Assumptions 1–6 hold. Suppose that the true parameter value lies in the interior of the set* $\Theta\cap\{\varphi\in V_\varphi : \phi(\varphi)=0\}$ *and that* ϕ *is five times continuously-differentiable on* V_φ *with bounded derivatives and Jacobian matrix with maximal row rank. Then*

$$\sqrt{nm}(\hat{\varphi}-\check{\varphi}) = \Omega_{nm}^{-1}\Phi_{\varphi}'(\Phi_{\varphi}\,\Omega_{nm}^{-1}\Phi_{\varphi}')^{-1}\Phi_{\varphi}\sqrt{nm}(\hat{\varphi}-\varphi)+o_P(1)$$

uniformly in $\theta\in\Theta_1$*.*

Proof. The proof of the theorem proceeds in a similar manner as the proof of Theorem S.1 and we use the same notation wherever possible. In particular, for $0\leq\epsilon\leq m^{-1/2}$, let $\varphi(\epsilon)$ and $\lambda(\epsilon)$ satisfy

$$\sum_{i=1}^{n}\sum_{t=1}^{m}\int u(\varphi(\epsilon),\eta_i(\varphi(\epsilon),\epsilon)|z)\,dG_{it}(z|\epsilon) - \Phi_{\varphi(\epsilon)}'\lambda(\epsilon)=0 \tag{S.12}$$

and $\phi(\varphi(\epsilon))=0$. Here, $\eta_i(\varphi,\epsilon)$ satisfies (S.1) for fixed φ and ϵ and $\lambda(\epsilon)$ is the Lagrange multiplier. Setting $\epsilon=m^{-1/2}$ gives the constrained maximum-likelihood estimator while setting $\epsilon=0$ gives the true parameter value. By a third-order expansion around $\epsilon=0$ we have

$$\begin{pmatrix}\check{\varphi}-\varphi\\ \check{\lambda}\end{pmatrix} = \frac{1}{\sqrt{m}}\begin{pmatrix}\frac{\partial\varphi(0)}{\partial\epsilon}\\ \frac{\partial\lambda(0)}{\partial\epsilon}\end{pmatrix} + \frac{1}{2!}\left(\frac{1}{\sqrt{m}}\right)^2\begin{pmatrix}\frac{\partial^2\varphi(0)}{\partial\epsilon^2}\\ \frac{\partial^2\lambda(0)}{\partial\epsilon^2}\end{pmatrix} + \frac{1}{3!}\left(\frac{1}{\sqrt{m}}\right)^3\begin{pmatrix}\frac{\partial^3\varphi(\tilde{\epsilon})}{\partial\epsilon^3}\\ \frac{\partial^3\lambda(\tilde{\epsilon})}{\partial\epsilon^3}\end{pmatrix}$$

for some $0 \le \tilde{\epsilon} \le m^{-1/2}$. Like before, the first of the right-hand side terms will satisfy a central limit theorem, the second will introduce bias, and the third will be asymptotically negligible.

For the first term we again begin by differentiating (S.12) with respect to ϵ. This yields

$$\sum_{i=1}^{n}\sum_{t=1}^{m} \frac{\partial \bar{u}(\epsilon|z)}{\partial \epsilon}\, dG_{it}(z|\epsilon) + \sum_{i=1}^{n}\sum_{t=1}^{m} \bar{u}(\epsilon|z)\, d\frac{\partial G_{it}(z|\epsilon)}{\partial \epsilon} - \frac{\partial \varPhi'_{\varphi(\epsilon)}}{\partial \epsilon}\lambda(\epsilon) - \varPhi'_{\varphi(\epsilon)}\frac{\partial \lambda(\epsilon)}{\partial \epsilon} = 0.$$

The first two terms coincide with those in (S.4). Evaluating at $\epsilon = 0$ and using that $\lambda(0) = 0$ gives

$$-nm\, \varOmega_{nm}\frac{\partial \varphi(0)}{\partial \epsilon} + \sqrt{m}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it} - \varPhi'_{\varphi(0)}\frac{\partial \lambda(0)}{\partial \epsilon} = 0.$$

Next, differentiate the constraint $\phi(\varphi(\epsilon)) = 0$ with respect to ϵ and evaluate at zero to obtain

$$\varPhi_{\varphi(0)}\frac{\partial \varphi(0)}{\partial \epsilon} = 0.$$

Combining both equations and using the shorthand notation $\varPhi = \varPhi_{\varphi(0)}$ yields the system

$$\begin{pmatrix} nm\, \varOmega_{nm} & \varPhi' \\ \varPhi & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial \varphi(0)}{\partial \epsilon} \\ \frac{\partial \lambda(0)}{\partial \epsilon} \end{pmatrix} = \begin{pmatrix} \sqrt{m}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it} \\ 0 \end{pmatrix}.$$

By a block-inversion formula,

$$\begin{pmatrix} nm\, \varOmega_{nm} & \varPhi' \\ \varPhi & 0 \end{pmatrix}^{-1} = \begin{pmatrix} (nm)^{-1}\ (\varOmega_{nm}^{-1} - \varOmega_{nm}^{-1}\varPhi'(\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\varOmega_{nm}^{-1}) & \varOmega_{nm}^{-1}\varPhi'(\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1} \\ (\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\varOmega_{nm}^{-1} & nm\, (\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1} \end{pmatrix},$$

and so

$$\frac{\partial \varphi(0)}{\partial \epsilon} = \left(\varOmega_{nm}^{-1} - \varOmega_{nm}^{-1}\varPhi'(\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\varOmega_{nm}^{-1}\right)\frac{1}{n\sqrt{m}}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it},$$

and, similarly,

$$\frac{\partial \lambda(0)}{\partial \epsilon} = (\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\varOmega_{nm}^{-1}\sqrt{m}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it}.$$

The same argument as used in the proof of Theorem S.1 yields that ${\partial \varphi(0)}/{\partial \epsilon} = O_P(n^{-1/2})$ and ${\partial \lambda(0)}/{\partial \epsilon} = O_P(\sqrt{n}m)$ uniformly over $\varTheta_1$

Moving on to the second term in the expansion we differentiate (S.12) twice with respect to ϵ.

$$\sum_{i=1}^{n}\sum_{t=1}^{m}\int \frac{\partial^2 \bar{u}(\epsilon|z)}{\partial \epsilon^2}\, dG_{it}(z|\epsilon) + 2\sum_{i=1}^{n}\sum_{t=1}^{m}\int \frac{\partial \bar{u}(\epsilon|z)}{\partial \epsilon}\, d\,\frac{\partial G_{it}(z|\epsilon)}{\partial \epsilon} + \frac{\partial^2 \Phi'_{\varphi(\epsilon)}\lambda(\epsilon)}{\partial \epsilon^2} = 0. \quad \text{(S.13)}$$

All but the last term previously appeared in the proof of Theorem S.1. Using that $\lambda(0) = 0$, that $\partial\varphi(0)/\partial\epsilon = O_P(n^{-1/2})$ and $\partial\lambda(0)/\partial\epsilon = O_P(\sqrt{n}m)$ uniformly on Θ_1, and that the first two derivatives of ϕ are bounded gives

$$\frac{\partial^2 \Phi'_{\varphi(0)}\lambda(0)}{\partial \epsilon^2} = \Phi' \frac{\partial^2 \lambda(0)}{\partial \epsilon^2} + O_P(m)$$

uniformly on Θ_1. Combining this with the analysis of the first two terms at $\epsilon = 0$ in the proof of Theorem S.1 implies that (S.13) at $\epsilon = 0$ equals

$$-nm\, \Omega_{nm} \frac{\partial^2 \varphi(0)}{\partial \epsilon^2} + 2nm\, b_{nm} - \Phi'_{\varphi(0)} \frac{\partial^2 \lambda(0)}{\partial \epsilon^2} + O_P(m) = 0$$

uniformly on Θ_1. Next, differentiate the constraint $\phi(\varphi(\epsilon)) = 0$ with respect to ϵ twice and evaluate at zero to obtain

$$\Phi_{\varphi(0)} \frac{\partial^2 \varphi(0)}{\partial \epsilon^2} + O_P(n^{-1}) = 0$$

uniformly on Θ_1. Combining both equations gives the system

$$\begin{pmatrix} nm\, \Omega_{nm} & \Phi' \\ \Phi & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 \varphi(0)}{\partial \epsilon^2} \\ \frac{\partial^2 \lambda(0)}{\partial \epsilon^2} \end{pmatrix} = \begin{pmatrix} 2nm\, b_{nm} \\ 0 \end{pmatrix} + \begin{pmatrix} O_P(m) \\ O_P(n^{-1}) \end{pmatrix}.$$

So,

$$\frac{\partial^2 \varphi(0)}{\partial \epsilon^2} = \left(\Omega_{nm}^{-1} - \Omega_{nm}^{-1}\Phi'(\Phi\Omega_{nm}^{-1}\Phi')^{-1}\Phi\Omega_{nm}^{-1}\right) 2b_{nm} + O_P(n^{-1}),$$

and, similarly,

$$\frac{\partial^2 \lambda(0)}{\partial \epsilon^2} = (\Phi\Omega_{nm}^{-1}\Phi')^{-1}\Phi\Omega_{nm}^{-1} 2nm\, b_{nm} + O_P(m),$$

uniformly on Θ_1.

For the third term in the expansion, finally, the same arguments as in the supplementary materials to Hahn and Newey (2004) can be used to establish that it is asymptotically negligible.

Combining results yields that, up to $o_P(1)$,

$$\sqrt{nm}(\check{\varphi} - \varphi) = \left(\varOmega_{nm}^{-1} - \varOmega_{nm}^{-1}\varPhi'(\varPhi\,\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\varOmega_{nm}^{-1}\right)\left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it} + \sqrt{\frac{n}{m}}b_{nm}\right)$$

uniformly on $\varTheta_1$. From Theorem S.1,

$$\sqrt{nm}(\hat{\varphi} - \varphi) = \varOmega_{nm}^{-1}\left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it} + \sqrt{\frac{n}{m}}b_{nm}\right) + o_P(1)$$

uniformly on $\varTheta_1$. Taking differences between these equations and again using Theorem S.1 then yields

$$\sqrt{nm}(\check{\varphi} - \hat{\varphi}) = -\varOmega_{nm}^{-1}\varPhi'(\varPhi\,\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\,\sqrt{nm}(\hat{\varphi} - \varphi) + o_P(1),$$

uniformly on $\varTheta_1$, which is the result of the theorem.

The calculations above also imply that

$$\frac{1}{\sqrt{nm}}\check{\lambda} = (\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\varOmega_{nm}^{-1}\left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m} u_{it} + \sqrt{\frac{n}{m}}b_{nm}\right) + o_P(1),$$

uniformly on $\varTheta_1$ and so, again by Theorem S.1,

$$\frac{1}{\sqrt{nm}}\check{\lambda} = (\varPhi\varOmega_{nm}^{-1}\varPhi')^{-1}\varPhi\,\sqrt{nm}(\hat{\varphi} - \varphi) + o_P(1) \tag{S.14}$$

uniformly on $\varTheta_1$. From this it readily follows that the Lagrange-multiplier statistic and the likelihood-ratio statistic share the same limit distribution (under the null). □

2 Auxiliary lemmata

Lemma S.1 (A uniform version of Lemma 1 of Hahn and Kuersteiner (2011))**.** *For $i = 1, \dots, n$, let $\{\xi_{it}, t = 1, 2, \dots m\}$ be a vector-valued sequence generated through a data generating process indexed by parameter $\psi_i \in \mathcal{P}$. Let*

$$a_i(\psi_i, h) := \sup_{1\leq t\leq m}\sup_{A\in\mathcal{A}_{it}(\psi_i)}\sup_{B\in\mathcal{B}_{it+h}(\psi_i)}|\mathbb{P}_{\psi_i}(A\cap B) - \mathbb{P}_{\psi_i}(A)\,\mathbb{P}_{\psi_i}(B)|,$$

where $\mathcal{A}_{it}(\psi_i)$ *and* $\mathcal{B}_{it}(\psi_i)$ *are the sigma algebras generated by the sequences* $\xi_{it}, \xi_{it-1}, \ldots$ *and* $\xi_{it}, \xi_{it+1}, \ldots$. *Assume that*

(i) $\mathbb{E}_{\psi_i}(\xi_{it}) = 0$ *for all* (i,t) *and* $\psi_i \in \mathcal{P}$,

(ii) the mixing coefficients satisfy

$$\sup_{1\le i\le n}\sup_{\psi_i\in\mathcal{P}}|a_i(\psi_i,h)| = O(r^h)$$

for some constant $0 < r < 1$,

(iii) $\sup_{1\le i\le n}\sup_{1\le t\le m}\sup_{\psi_i\in\mathcal{P}}\mathbb{E}_{\psi_i}(\|\xi_{it}\|_2^q) \le c$ *for some* $q \ge 2$ *and constant* $0 < c < \infty$.

Then, as $n, m \to \infty$ *so that* ${}^n\!/\!_m \to \gamma^2$ *with* $0 < \gamma < \infty$,

if $q > 4$,

$$\sup_{\psi_1,\ldots,\psi_n\in\mathcal{P}^n}\mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\le i\le n}\left\|\frac{1}{m}\sum_{t=1}^m \xi_{it}\right\|_2 > \varepsilon\right) = o(m^{-1})$$

for all $\varepsilon > 0$; *while*

if $q \ge 2$ *and* $s > 0$,

$$\sup_{\psi_1,\ldots,\psi_n\in\mathcal{P}^n}\mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\le i\le n}\left\|\frac{1}{\sqrt{m}}\sum_{t=1}^m \xi_{it}\right\|_2 > m^s\varepsilon\right) = O(m^{1-qs})$$

for all $\varepsilon > 0$.

Proof. It suffices to prove the second part of the theorem. Consider a fixed value $\varepsilon > 0$. We have

$$\mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\le i\le n}\left\|\frac{1}{\sqrt{m}}\sum_{t=1}^m \xi_{it}\right\|_2 > m^s\varepsilon\right) \le \sum_{i=1}^n \mathbb{P}_{\psi_i}\left(\left\|\frac{1}{\sqrt{m}}\sum_{t=1}^m \xi_{it}\right\|_2 > m^s\varepsilon\right).$$

By and application of Markov's inequality and a strong-mixing moment inequality of Doukhan (1994, Theorem 2 and Remark 2, pp. 25–30), we have that

$$\begin{aligned}\mathbb{P}_{\psi_i}\left(\left\|\frac{1}{\sqrt{m}}\sum_{t=1}^m \xi_{it}\right\|_2 > m^s\varepsilon\right) &\le \frac{1}{\varepsilon^q}\frac{1}{m^{(2s+1)q/2}}\quad \mathbb{E}_{\psi_i}\left(\left\|\sum_{t=1}^m \xi_{it}\right\|_2^q\right)\\ &\lesssim \frac{1}{\varepsilon^q}\frac{1}{m^{(2s+1)q/2}}\; m^{q/2}\\ &\lesssim \frac{1}{\varepsilon^q}\,\frac{1}{m^{qs}}\end{aligned}$$

for all $1 \leq i \leq n$, with the upper bound being independent of both i and ψ_i. Consequently, we obtain that

$$\mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\leq i\leq n}\left\|\frac{1}{\sqrt{m}}\sum_{t=1}^{m}\xi_{it}\right\|_2 > m^s\varepsilon\right) \lesssim \frac{n}{m}\,\frac{1}{m^{qs-1}} = O(m^{(1-qs)}).$$

This completes the proof. □

Lemma S.2 (A uniform version of Lemma 2 of Hahn and Kuersteiner (2011))**.** *For $i = 1,\ldots,n$, let $\{\xi(z_{it},\phi_i), t = 1,2,\ldots m\}$ be a vector-valued sequence of functions of data z_{it} and a parameter $\phi_i \in \mathcal{Q}$, for $\mathcal{Q}$ compact. The z_{it} are generated through a data generating process indexed by parameter $\psi_i \in \mathcal{P}$. Let*

$$a_i(\psi_i,h) := \sup_{1\leq t\leq m}\ \sup_{A\in\mathcal{A}_{it}(\psi_i)}\ \sup_{B\in\mathcal{B}_{it+h}(\psi_i)} |\mathbb{P}_{\psi_i}(A\cap B) - \mathbb{P}_{\psi_i}(A)\,\mathbb{P}_{\psi_i}(B)|,$$

where $\mathcal{A}_{it}(\psi_i)$ and $\mathcal{B}_{it}(\psi_i)$ are the sigma algebras generated by the sequences $z_{it}, z_{it-1},\ldots$ and $z_{it}, z_{it+1},\ldots$. Assume that

(i) $\mathbb{E}_{\psi_i}(\xi(z_{it},\phi_i)) = 0$ for all (i,t), $\psi_i \in \mathcal{P}$ and $\phi_i \in \mathcal{Q}$,

(ii) the mixing coefficients satisfy

$$\sup_{1\leq i\leq n}\ \sup_{\psi_i\in\mathcal{P}} |a_i(\psi_i,h)| = O(r^h)$$

for some constant $0 < r < 1$,

(iii) there exists a function b such that $\sup_{\phi_i\in\mathcal{Q}}\|\xi(z_{it},\phi_i)\|_2 \leq b(z_{it})$, for all $\phi_1,\phi_2 \in \mathcal{Q}$,

$$\|\xi(z_{it},\phi_1) - \xi(z_{it},\phi_2)\|_2 \leq b(z_{it})\ \|\phi_1-\phi_2\|_2,$$

and $\sup_{1\leq i\leq n}\sup_{1\leq t\leq m}\sup_{\phi_i\in\mathcal{P}}\mathbb{E}_{\psi_i}(b(z_{it})^q) < c$ for some $q \geq 2$ and $0 < c < \infty$.

Then, as $n,m\to\infty$ so that $^n/_m \to \gamma^2$ with $0<\gamma<\infty$,

if $q > 6$,

$$\sup_{\psi_1,\ldots,\psi_n\in\mathcal{P}^n}\mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\leq i\leq n}\left\|\frac{1}{m}\sum_{t=1}^{m}\xi(z_{it},\phi_i)\right\|_2 > \varepsilon\right) = o(m^{-1})$$

for all $\varepsilon > 0$; while

If $q \geq 2$ and $s > 0$ are such that $qs > 3 + {}^{\dim(\phi)}/2$,

$$\sup_{\psi_1,\ldots,\psi_n \in \mathcal{P}^n} \mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\leq i\leq n} \left\| \frac{1}{\sqrt{m}} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > m^s \varepsilon \right) = o(m^{-1})$$

for all $\varepsilon > 0$.

Proof. Fix $\varepsilon > 0$. We begin by noting that

$$\begin{aligned}
&\sup_{\psi_1,\ldots,\psi_n \in \mathcal{P}^n} \mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\max_{1\leq i\leq n} \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > \varepsilon \right) \\
&\leq \sup_{\psi_1,\ldots,\psi_n \in \mathcal{P}^n} \mathbb{P}_{\psi_1,\ldots,\psi_n}\left(\bigcup_{i=1}^{n} \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > \varepsilon \right) \\
&\leq \sum_{i=1}^{n} \sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > \varepsilon \right).
\end{aligned}$$

Because $\mathcal{Q}$ is compact we can divide it into a finite number k_δ of subsets $\mathcal{Q}_1, \ldots \mathcal{Q}_{k_\delta}$ such that $\|\phi_1 - \phi_2\|_2 \leq \delta$ whenever ϕ_1 and ϕ_2 lie in the same subset. With this covering in hand,

$$\begin{aligned}
\sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > \varepsilon \right) &\leq \sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\sup_{\phi \in \mathcal{Q}} \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi) \right\|_2 > \varepsilon \right) \\
&\leq \sum_{k=1}^{k_\delta} \sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\sup_{\phi \in \mathcal{Q}_k} \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi) \right\|_2 > \varepsilon \right),
\end{aligned}$$

for each $i = 1, \ldots, n$. Further, for each subset $\mathcal{Q}_k$, letting $\phi_{(k)} \in \mathcal{Q}_k$ we can invoke Condition (iii) to obtain

$$\begin{aligned}
\sup_{\phi \in \mathcal{Q}_k} \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi) \right\|_2 &\leq \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_{(k)}) \right\|_2 + \frac{1}{m} \sum_{t=1}^{m} \sup_{\phi \in \mathcal{Q}_k} \left\| \xi(z_{it}, \phi) - \xi(z_{it}, \phi_{(k)}) \right\|_2 \\
&\leq \left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_{(k)}) \right\|_2 \\
&\quad + \frac{\delta}{m} \sum_{t=1}^{m} |b(z_{it}) - \mathbb{E}_{\psi_i}(b(z_{it}))| + 2\,\delta\,\mathbb{E}_{\psi_i}(b(z_{it})).
\end{aligned}$$

Set δ so that $2\,\delta\,\mathbb{E}_{\psi_i}(b(z_{it})) < {}^{\varepsilon}/3$. Then, combining the last two bounding inequalities yields

$$\begin{aligned}
\sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > \varepsilon \right) &\leq \sum_{k=1}^{k_\delta} \sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\left\| \frac{1}{m} \sum_{t=1}^{m} \xi(z_{it}, \phi_{(k)}) \right\|_2 > \frac{\varepsilon}{3} \right) \\
&\quad + \sum_{k=1}^{k_\delta} \sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\frac{\delta}{m} \sum_{t=1}^{m} |b(z_{it}) - \mathbb{E}_{\psi_i}(b(z_{it}))| > \frac{\varepsilon}{3} \right).
\end{aligned}$$

Here, each of the right-hand side terms satisfies the conditions of Lemma S.1 and are, therefore, both $o(m^{-1})$ by an application of the first result given there. The first statement in the theorem then follows from the fact that $k_\delta = O(1)$ and that ${}^{n}\!/_{m} = O(1)$.

To show the second part we proceed in the same manner, only now partitioning $\mathcal{Q}$ into subsets such that $\|\phi_1 - \phi\|_2 \leq {}^{\delta}\!/\!\sqrt{m}$ for some $\delta > 0$. The number of sets needed to do so is of the order $m^{\dim(\phi)/2}$, and each of them yields terms to which the second part of Lemma S.1 can be applied, showing them to be at most of order m^{1-qs} uniformly. This then yields

$$\sup_{\psi_i \in \mathcal{P}} \mathbb{P}_{\psi_i}\left(\max_{1\leq i \leq n} \left\| \frac{1}{\sqrt{m}} \sum_{t=1}^{m} \xi(z_{it}, \phi_i) \right\|_2 > m^s \varepsilon \right) = O(n)\, O(m^{1-qs+\dim(\phi)/2}) = o(m^{-1}),$$

using that $O({}^{n}\!/_{m}) = O(1)$ and that $qs > 3 + {}^{\dim(\phi)}\!/_{2}$. This completes the proof of the lemma. □

Lemma S.3. *Let x and y be two random vectors of length k. Write ι for the k-vector of ones. Then*

$$\mathbb{P}(y \leq a) \leq \mathbb{P}(x \leq a + \iota\varepsilon) + \mathbb{P}(\|y - x\|_2 > \sqrt{k}\varepsilon)$$

for all a and any $\varepsilon > 0$.

Proof. Fix $\varepsilon > 0$. We have

$$\mathbb{P}(y \leq a) = \mathbb{P}(y \leq a\,,\, x \leq a + \iota\varepsilon) + \mathbb{P}(y \leq a\,,\, x > a + \iota\varepsilon).$$

Now, $\mathbb{P}(y \leq a\,,\, x \leq a + \iota\varepsilon) \leq \mathbb{P}(x \leq a + \iota\varepsilon)$ while

$$\begin{aligned} \mathbb{P}(y \leq a\,,\, x > a + \iota\varepsilon) &= \mathbb{P}(y - x \leq a - x\,,\, a - x < -\iota\varepsilon) \\ &\leq \mathbb{P}(y - x < -\iota\varepsilon) + \mathbb{P}(y - x > \iota\varepsilon) \\ &\leq \mathbb{P}(\|y - x\|_1 > k\varepsilon) \\ &\leq \mathbb{P}(\|y - x\|_2 > \sqrt{k}\varepsilon). \end{aligned}$$

Combining results completes the proof of the lemma. □

Lemma S.4. *Let y_{nm}, x_{nm}, and z be random vectors of size k whose probability functions are indexed by the parameter $\theta \in \Theta$. Assume that* *(i) The function* $\mathbb{P}_\theta(z \le a)$ *is continuous in* a *for all* $\theta \in \Theta$*,*

(ii) as $n, m \to \infty$*,* $\sup_{\theta\in\Theta} \mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \varepsilon) = o(1)$ *for all* $\varepsilon > 0$*,*

(iii) as $n, m \to \infty$*,* $\sup_{\theta\in\Theta}|\mathbb{P}_\theta(x_{nm} \le a) - \mathbb{P}_\theta(z \le a)| = o(1)$ *for all* a*.*

Then,

$$\sup_{\theta\in\Theta}|\mathbb{P}_\theta(y_{nm} \le a) - \mathbb{P}_\theta(z \le a)| = o(1)$$

as $n, m \to \infty$*.*

Proof. For any $\theta \in \Theta$ and $\varepsilon > 0$, an application of Lemma S.3 gives

$$\mathbb{P}_\theta(y_{nm} \le a) \le \mathbb{P}_\theta(x_{nm} \le a + \iota\varepsilon) + \mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon)$$

and so, for any $a_+ > a + \iota\varepsilon$,

$$\mathbb{P}_\theta(y_{nm} \le a) \le \mathbb{P}_\theta(x_{nm} \le a_+) + \mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon). \tag{S.15}$$

By an application of the same lemma, for any $a_- < a - \iota\varepsilon$,

$$\mathbb{P}_\theta(x_{nm} \le a_-) \le \mathbb{P}_\theta(y_{nm} \le a) + \mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon). \tag{S.16}$$

Taken together, (S.15) and (S.16) imply that

$$\begin{aligned}
&\mathbb{P}_\theta(x_{nm} \le a_-) - \mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon) \\
\le\, &\mathbb{P}_\theta(y_{nm} \le a) \\
\le\, &\mathbb{P}_\theta(x_{nm} \le a_+) + \mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon).
\end{aligned}$$

Subtracting $\mathbb{P}_\theta(z \le a)$ from each of the terms in the above inequalities and re-arranging shows that

$$\begin{aligned}
|\mathbb{P}_\theta(y_{nm} \le a) - \mathbb{P}_\theta(z \le a)| \le\, & 2\,\mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon) \\
& + |\mathbb{P}_\theta(x_{nm} \le a_-) - \mathbb{P}_\theta(z \le a)| + |\mathbb{P}_\theta(x_{nm} \le a_+) - \mathbb{P}_\theta(z \le a)|.
\end{aligned}$$

Applying an adding and substracting stategy to the terms on the right-hand side now gives

$$
\begin{aligned}
\sup_{\theta\in\Theta}|\mathbb{P}_\theta(y_{nm} \le a) - \mathbb{P}_\theta(z \le a)| \le 2\sup_{\theta\in\Theta}\mathbb{P}_\theta(\|y_{nm} - x_{nm}\|_2 > \sqrt{k}\varepsilon) \\
+ \sup_{\theta\in\Theta}|\mathbb{P}_\theta(x_{nm} \le a_-) - \mathbb{P}_\theta(z \le a_-)| \\
+ \sup_{\theta\in\Theta}|\mathbb{P}_\theta(z \le a_-) - \mathbb{P}_\theta(z \le a)| \\
+ \sup_{\theta\in\Theta}|\mathbb{P}_\theta(x_{nm} \le a_+) - \mathbb{P}_\theta(z \le a_+)| \\
+ \sup_{\theta\in\Theta}|\mathbb{P}_\theta(z \le a_+) - \mathbb{P}_\theta(z \le a)|.
\end{aligned}
$$

Here, as $n, m \to \infty$, the first right-hand side term is $o(1)$ by Condition (ii); the second and fourth term are both $o(1)$ by Condition (iii); and, due to the fact that a_- and a_+ can be chosen to be arbitrarily close, the third and fifth term can be made arbitrarily small by Condition (i). The result has thus been shown and the proof of the lemma is, therefore, complete. □

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Doukhan, P. (1994). *Mixing: Properties and examples.* Springer Texts in Statistics. Springer-Verlag.

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

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

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications.* Academic Press.

Kim, M. S. and Y. Sun (2016). Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models. *Econometric Theory 32*, 1523–1568.