Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Bootstrap Inference For Misspecified Moment Condition Models: Supplementary Material

Mihai Giurcanu · **Brett Presnell**

Received: date / Revised: date

A Proofs of Section 2

Lemma A.1 shows the consistency of GMM estimators using similar arguments as in Theorem 5.7 and Example 19.8 of van der Vaart (1998).

Lemma A.1 *Suppose that conditions A–E hold. Then (i)* $\sup_{\theta\in\Theta}|\hat{Q}(\theta,\hat{W})-Q(\theta,W_0)|\xrightarrow{P}0$*; (ii)* $\hat{\eta}\xrightarrow{P}\eta_0$*; and (iii)* $\Pr(\hat{\Psi}(\hat{\eta})=0)\to 1$.

Proof (Theorem 1) Taylor expansion of $\hat{\Psi}(\hat{\eta})$ about $\eta\in\mathbb{R}^{2p}$ yields:

$$0=\hat{\Psi}(\eta_0)+\begin{pmatrix}\hat{A}(\theta_1,\hat{W}) & O\\ \hat{B}(\theta_1,\theta_2) & \hat{A}(\theta_2,\hat{V}(\theta_1)^{-1})\end{pmatrix}(\hat{\eta}-\eta_0)+o_P(\|\hat{\eta}-\eta_0\|)\,, \tag{A.1}$$

$\hat{A}(\zeta_1,W)=\big(\partial_j\hat{C}(\zeta_1,W):1\le j\le p\big)\in\mathbb{R}^{p\times p}$, and

$$\hat{B}(\zeta_1,\zeta_2)=\big(-\hat{G}(\zeta_2)^T\hat{V}(\zeta_1)^{-1}\big(\partial_j\hat{V}(\zeta_1)\big)\hat{V}(\zeta_1)^{-1}\bar{g}(\zeta_2):1\le j\le p\big)\in\mathbb{R}^{p\times p}\,.$$

By (A.1),

$$n^{1/2}(\hat{\eta}-\eta_0)=-n^{1/2}\begin{pmatrix}A_1^{-1} & O\\ -A_2^{-1}BA_1^{-1} & A_2^{-1}\end{pmatrix}\hat{\Psi}(\eta_0)+o_P(1)\,. \tag{A.2}$$

Simple algebra shows that both A_1 and A_2 are symmetric matrices. Since $\operatorname{vec}(AXB)=(B^T\otimes A)\operatorname{vec}(X)$, where $\operatorname{vec}(\cdot)$ vectorizes the columns of matrix, by (5),

$$\begin{aligned}\hat{\Psi}(\eta_0)&=\begin{pmatrix}G_1^TW_0(\bar{g}(\theta_1)-\mu_1)\\ G_2^TV_1^{-1}(\bar{g}(\theta_2)-\mu_2)\end{pmatrix}+\begin{pmatrix}\big((W_0\mu_1)^T\otimes I_p\big)\operatorname{vec}\big(\hat{G}(\theta_1)^T-G_1^T\big)\\ \big((V_1^{-1}\mu_2)^T\otimes I_p\big)\operatorname{vec}\big(\hat{G}(\theta_2)^T-G_2^T\big)\end{pmatrix}\\ &\quad+\begin{pmatrix}\big(\mu_1^T\otimes G_1^T\big)\operatorname{vec}(\hat{W}-W_0)\\ -\big((V_1^{-1}\mu_2)^T\otimes(G_2^TV_1^{-1})\big)\operatorname{vec}(\hat{V}(\theta_1)-V_1)\end{pmatrix}+o_P(n^{-1/2})\\ &=\Pi\big(\hat{\psi}(\eta_0)-\psi(\eta_0)\big)+o_P(n^{-1/2})\,.\end{aligned} \tag{A.3}$$

Mihai Giurcanu E-mail: giurcanu@uchicago.edu (corresponding author)
Department of Public Health Sciences, University of Chicago, 5841 S Maryland Ave, Room R325, Chicago, IL, 60637, USA

Brett Presnell E-mail: presnell@ufl.edu
Department of Statistics, University of Florida, 225 Griffin-Floyd Hall, Gainesville, FL, 32611, USA

Hence $n^{1/2}\hat{\Psi}(\eta_0) \xrightarrow{\mathrm{d}} N(0,\Xi)$, and by (A.2), then (8a) holds, as stated. By the delta method, $n^{1/2}(h(\hat{\theta}_{\mathrm{T}}) - h(\theta_2)) \xrightarrow{\mathrm{d}} N(0, H_2\Gamma_{22}H_2^T)$, and thus, (7a) holds.

The constrained estimator $\tilde{\theta}_{\mathrm{T}}$ minimizes $\hat{Q}(\theta,\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$ subject to $h(\theta)=0$. Let $(1/2)\hat{Q}(\theta,\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) + \lambda^T h(\theta)$ be the Lagrangian function. Taking the partial derivatives of the Lagrangian and solving for θ and λ yields:

$$\begin{cases} \hat{C}(\tilde{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) + H(\tilde{\theta}_{\mathrm{T}})^T\hat{\lambda} = 0\,, \\ h(\tilde{\theta}_{\mathrm{T}}) = 0\,. \end{cases} \tag{A.4}$$

Since $\mathrm{H}_0 : \theta_0 \in \Theta_0$ is true, then $\tilde{\theta}_{\mathrm{T}} = \theta_2 + O_P(n^{-1/2})$, and by (A.4), we have $\hat{\lambda} = O_P(n^{-1/2})$. Taylor expansions of (A.4) about $(\theta_2, 0) \in \mathbb{R}^p \times \mathbb{R}^r$ yields

$$0 = \begin{pmatrix} \hat{C}(\theta_2,\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \\ 0 \end{pmatrix} + \begin{pmatrix} A_2 & H_2^T \\ H_2 & O \end{pmatrix} \begin{pmatrix} \tilde{\theta}_{\mathrm{T}} - \theta_2 \\ \hat{\lambda} \end{pmatrix} + O_P(n^{-1})\,.$$

Hence

$$n^{1/2}(\tilde{\theta}_{\mathrm{T}} - \theta_2) = -n^{1/2}(A_2^{-1} - \Delta_2)\hat{C}(\theta_2,\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) + o_P(1)\,. \tag{A.5}$$

Similarly to (A.2),

$$n^{1/2}\hat{C}(\theta_2,\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) = -n^{1/2}A_2(\hat{\theta}_{\mathrm{T}} - \theta_2) + o_P(1)\,, \tag{A.6}$$

and by (A.5), we obtain

$$n^{1/2}(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) = n^{1/2}\Delta_2\hat{C}(\theta_2,\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) + o_P(1)\,. \tag{A.7}$$

By (A.6), (A.7), and (8a), $n^{1/2}(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) \xrightarrow{\mathrm{d}} N(0,\Omega_2)$. One-term Taylor expansion of $\hat{Q}(\tilde{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$ about $\hat{\theta}_{\mathrm{T}}$ yields

$$\begin{aligned} \hat{L}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) &= 2n\hat{C}(\hat{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})^T(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) \\ &\quad + n(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}})^T\hat{A}(\hat{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) + o_P(1) = n(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}})^T A_2(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) + o_P(1)\,, \end{aligned}$$

and thus, (7b) holds. Taylor expansion of $\hat{C}(\tilde{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$ about $\hat{\theta}_{\mathrm{T}}$ yields

$$\begin{aligned} n^{1/2}\hat{C}(\tilde{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) &= n^{1/2}\hat{C}(\hat{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \\ &\quad + n^{1/2}\hat{A}(\hat{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) + o_P(1) = n^{1/2}A_2(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) + o_P(1)\,, \end{aligned}$$

and thus, (7c) holds. Similarly, $n^{-1/2}\hat{L}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) = 2n^{1/2}\hat{C}(\hat{\theta}_{\mathrm{T}},\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1})^T(\tilde{\theta}_{\mathrm{T}} - \hat{\theta}_{\mathrm{T}}) + o_P(1)$. Hence (8b) holds, where $\sigma_L^2 = 4\mu_2^T V_2^{-1}G_2\Omega_2 G_2^T V_2^{-1}\mu_2$. Note further that

$$n^{-1}\hat{S}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) = \psi_S\big(\bar{g}(\tilde{\theta}_{\mathrm{T}}), \mathrm{vec}(\hat{V}(\hat{\theta}_{\mathrm{T}})), \mathrm{vec}(\hat{G}(\tilde{\theta}_{\mathrm{T}})), \mathrm{vec}(\hat{G}(\hat{\theta}_{\mathrm{T}}))\big)\,, \tag{A.8}$$

where $\psi_S(g,v,g_1,g_2) : \mathbb{R}^q \times \mathbb{R}^{q^2} \times \mathbb{R}^{pq} \times \mathbb{R}^{pq} \mapsto \mathbb{R}$ is given by:

$$\psi_S(g,v,g_1,g_2) = g^T V^{-1}G_1(G_2^T V^{-1}G_2)^{-1}G_1^T V^{-1}g\,, \tag{A.9}$$

where $V = \mathrm{uvec}_{q,q}(v) \in \mathbb{R}^{q\times q}$, $G_1 = \mathrm{uvec}_{q,p}(g_1) \in \mathbb{R}^{q\times p}$, $G_2 = \mathrm{uvec}_{q,p}(g_2) \in \mathbb{R}^{q\times p}$, and $\mathrm{uvec}_{p,q} : \mathbb{R}^{pq} \to \mathbb{R}^{p\times q}$ is the inverse of the vec operator. Since

$$n^{1/2}\begin{pmatrix} \bar{g}(\tilde{\theta}_{\mathrm{T}}) - \mu_2 \\ \mathrm{vec}(\hat{V}(\hat{\theta}_{\mathrm{T}}) - V_2) \\ \mathrm{vec}(\hat{G}(\tilde{\theta}_{\mathrm{T}}) - G_2) \\ \mathrm{vec}(\hat{G}(\hat{\theta}_{\mathrm{T}}) - G_2) \end{pmatrix} = n^{1/2}\begin{pmatrix} \bar{g}(\tilde{\theta}_{\mathrm{T}}) - \bar{g}(\theta_2) \\ \mathrm{vec}(\hat{V}(\hat{\theta}_{\mathrm{T}}) - \hat{V}(\theta_2)) \\ \mathrm{vec}(\hat{G}(\tilde{\theta}_{\mathrm{T}}) - \hat{G}(\theta_2)) \\ \mathrm{vec}(\hat{G}(\hat{\theta}_{\mathrm{T}}) - \hat{G}(\theta_2)) \end{pmatrix} + n^{1/2}\begin{pmatrix} \bar{g}(\theta_2) - \mu_2 \\ \mathrm{vec}(\hat{V}(\theta_2) - V_2) \\ \mathrm{vec}(\hat{G}(\theta_2) - D_2) \\ \mathrm{vec}(\hat{G}(\theta_2) - D_2) \end{pmatrix} + o_P(1)\,,$$

and $n^{1/2}(\bar{g}(\tilde{\theta}_{\mathrm{T}}) - \bar{g}(\theta_2)) = n^{1/2}G_0(\tilde{\theta}_{\mathrm{T}} - \theta_2) + o_P(1)$, and

$$\begin{aligned} n^{1/2}\,\mathrm{vec}\big(\hat{V}(\hat{\theta}_{\mathrm{T}}) - \hat{V}(\theta_2)\big) &= n^{1/2}\,\mathrm{vec}\big(\nabla V(\theta_2)((\hat{\theta}_{\mathrm{T}} - \theta_2) \otimes I_q)\big) + o_P(1)\,, \\ n^{1/2}\,\mathrm{vec}\big(\hat{G}(\tilde{\theta}_{\mathrm{T}}) - \hat{G}(\theta_2)\big) &= n^{1/2}\,\mathrm{vec}\big(\nabla G(\theta_2)((\tilde{\theta}_{\mathrm{T}} - \theta_2) \otimes I_q)\big) + o_P(1)\,, \\ n^{1/2}\,\mathrm{vec}\big(\hat{G}(\hat{\theta}_{\mathrm{T}}) - \hat{G}(\theta_2)\big) &= n^{1/2}\,\mathrm{vec}\big(\nabla G(\theta_2)((\hat{\theta}_{\mathrm{T}} - \theta_2) \otimes I_q)\big) + o_P(1)\,, \end{aligned}$$

by (A.2), (A.3), (A.5), and condition E, there exist a matrix D_S such that

$$n^{1/2}\begin{pmatrix}\bar g(\tilde\theta_{\mathrm T})-\mu_2\\ \operatorname{vec}(\hat V(\tilde\theta_{\mathrm T})-V_2)\\ \operatorname{vec}(\hat G(\tilde\theta_{\mathrm T})-G_2)\\ \operatorname{vec}(\hat G(\hat\theta_{\mathrm T})-G_2)\end{pmatrix}\xrightarrow{\mathrm d} N(0,D_S\Sigma D_S^T)\,.$$

By the delta method, then (8c) holds, where $\psi_2=\mu_2^TV_2^{-1}G_2(G_2^TV_2^{-1}G_2)^{-1}G_2^TV_2^{-1}\mu_2$, $\sigma_S^2=\nabla\psi_SD_S\Sigma D_S^T\nabla\psi_S^T$, and $\nabla\psi_S=\nabla\psi_S(\mu_2,\operatorname{vec}(V_2),\operatorname{vec}(G_2),\operatorname{vec}(G_2))$. Since

$$n^{1/2}(\bar g(\hat\theta_{\mathrm T})-\mu_2)=n^{1/2}(\bar g(\hat\theta_{\mathrm T})-\bar g(\theta_2))+n^{1/2}(\bar g(\theta_2)-\mu_2)+o_P(1)\,,$$

by (A.2) and condition E, there exists a matrix D_J such that

$$n^{1/2}(\bar g(\hat\theta_{\mathrm T})-\mu_2)\xrightarrow{\mathrm d} N(0,D_J\Sigma D_J^T)\,. \tag{A.10}$$

Hence (8d) holds, where $\lambda_1^J,\dots,\lambda_q^J$ are the eigenvalues of $(D_J\Sigma D_J^T)^{1/2}V_2^{-1}(D_J\Sigma D_J^T)^{1/2}$.

Finally, write $n^{-1}\hat J(\hat V(\hat\theta_{\mathrm O})^{-1})=\psi_J(\bar g(\hat\theta_{\mathrm T}),\operatorname{vec}(\hat V(\hat\theta_{\mathrm O}))$, where $\psi_J:\mathbb R^q\times\mathbb R^{q^2}\mapsto\mathbb R$ is given by $\psi_J(g,v)=g^TV^{-1}g$ and $V=\operatorname{uvec}_{q,q}(v)\in\mathbb R^{q\times q}$. Since

$$n^{1/2}\begin{pmatrix}\bar g(\hat\theta_{\mathrm T})-\mu_2\\ \operatorname{vec}(\hat V(\hat\theta_{\mathrm O})-V_1)\end{pmatrix}=n^{1/2}\begin{pmatrix}\bar g(\hat\theta_{\mathrm T})-\bar g(\theta_2)\\ \operatorname{vec}(\hat V(\hat\theta_{\mathrm O})-\hat V(\theta_1))\end{pmatrix}+n^{1/2}\begin{pmatrix}\bar g(\theta_2)-\mu_2\\ \operatorname{vec}(\hat V(\theta_1)-V_1)\end{pmatrix}+o_P(1)\,,$$

and $n^{1/2}(\bar g(\hat\theta_{\mathrm T})-\bar g(\theta_2))=n^{1/2}G_2(\hat\theta_{\mathrm T}-\theta_2)+o_P(1)$, and

$$n^{1/2}\operatorname{vec}\big(\hat V(\hat\theta_{\mathrm O})-\hat V(\theta_1)\big)=n^{1/2}\operatorname{vec}\big(\nabla V(\theta_1)((\hat\theta_{\mathrm O}-\theta_1)\otimes I_q)\big)+o_P(1)\,,$$

by (A.2) and condition E, there exists a matrix E_J such that

$$n^{1/2}\begin{pmatrix}\bar g(\hat\theta_{\mathrm T})-\mu_2\\ \operatorname{vec}(\hat V(\hat\theta_{\mathrm O})-V_1)\end{pmatrix}\xrightarrow{\mathrm d} N\big(0,E_J\Sigma E_J^T\big)\,.$$

Thus, (8e) holds, where $\sigma_J^2=\nabla\psi_JE_J\Sigma E_J^T\nabla\psi_J^T$ and $\nabla\psi_J=\nabla\psi_J(\mu_2,\operatorname{vec}(V_1))$.

Remark A.1 If the model (1) is correctly specified, then $\theta_1=\theta_2=\theta_0$, $\mu_1=\mu_2=0$, $A_1=D(\theta_0,W_0)$, $B=O$, $A_2=D_0$, and $G_1=G_2=G_0$. By (6), $\Pi_{11}=G_0^TW_0$, $\Pi_{12}=O$, $\Pi_{13}=O$, $\Pi_{24}=O$, $\Pi_{25}=G_0^TV_0^{-1}$, and $\Pi_{26}=O$. Hence $\Xi_{11}=G_0^TW_0V_0W_0G_0$, $\Xi_{12}=\Xi_{21}=D(\theta_0,W_0)$, and $\Xi_{22}=D_0$. Furthermore, $\Gamma_{11}=D(\theta_0,W_0)^{-1}G_0^TW_0V_0W_0G_0D(\theta_0,W_0)^{-1}$, $\Gamma_{12}=\Gamma_{21}=\Gamma_{22}=D_0^{-1}$, and thus, (2a) holds. Note further that $\Omega_2=D_0^{-1}H_0^T\big(H_0D_0^{-1}H_0^T\big)^{-1}H_0D_0^{-1}$. Since $\Omega_2^{1/2}D_0\Omega_2^{1/2}$ is idempotent of rank r, by (7a), (7b), and (7c), then (2b) holds. By (A.10), then $V_0^{-1/2}D_J\Sigma D_J^TV_0^{-1/2}=\Lambda_2$, where $\Lambda_2=I_q-V_0^{-1/2}G_0D_0^{-1}G_0^TV_0^{-1/2}$ is idempotent of rank $q-p$. Hence (4c) holds.

Proof (Lemma 1) Assume that $\mathrm H_0:\theta_0\in\Theta_0$ is true. Note first that

$$\begin{aligned}\Pr\big(\hat T(\hat V(\hat\theta_{\mathrm T})^{-1})\le x\big)&=\Pr\big(\hat Z^T\Upsilon_0\hat Z\le x\big)-n^{-1/2}\frac{d}{dx}\mathrm E\big(\hat Z^T\Upsilon_1(\hat Z\otimes I_k)\hat ZI(\hat Z^T\Upsilon_0\hat Z\le x)\big)\\&\quad-n^{-1}\frac{d}{dx}\mathrm E\big(\hat Z^T\Upsilon_2(\hat Z\otimes I_{k^2})(\hat Z\otimes I_k)\hat ZI(\hat Z^T\Upsilon_0\hat Z\le x)\big)\\&\quad-n^{-3/2}\frac{d}{dx}\mathrm E\big(\hat Z^T\Upsilon_3(\hat Z\otimes I_{k^3})(\hat Z\otimes I_{k^2})(\hat Z\otimes I_k)\hat ZI(\hat Z^T\Upsilon_0\hat Z\le x)\big)\\&\quad+(1/2)n^{-1}\frac{d^2}{dx^2}\mathrm E\big((\hat Z^T\Upsilon_1(\hat Z\otimes I_k)\hat Z)^2I(\hat Z^T\Upsilon_0\hat Z\le x)\big)\\&\quad+n^{-3/2}\frac{d^2}{dx^2}\mathrm E\big((\hat Z^T\Upsilon_1(\hat Z\otimes I_k)\hat Z)(\hat Z^T\Upsilon_2(\hat Z\otimes I_{k^2})(\hat Z\otimes I_k)\hat Z)I(\hat Z^T\Upsilon_0\hat Z\le x)\big)\\&\quad-(1/6)n^{-3/2}\frac{d^3}{dx^3}\mathrm E\big((\hat Z^T\Upsilon_1(\hat Z\otimes I_k)\hat Z)^3I(\hat Z^T\Upsilon_0\hat Z\le x)\big)+O(n^{-2})\,,\end{aligned}$$

where $I(E)$ is the indicator function of an event E. Here and in other places, we use

$$\begin{aligned}
\Pr\big(R+n^{-1/2}S+n^{-1}T+n^{-3/2}U \le x\big) &= \Pr(R\le x)+\int\int_x^{x-n^{-1/2}s-n^{-1}t-n^{-3/2}u} dF(r,s,t,u)\\
&= \Pr(R\le x)+\int\big(-n^{-1/2}s-n^{-1}t-n^{-3/2}u\big)\,d\partial_1 F(x,s,t,u)\\
&\quad+(1/2)\int\big(n^{-1/2}s+n^{-1}t+n^{-3/2}u\big)^2\,d\partial_1^2 F(x,s,t,u)\\
&\quad+(1/6)\int\big(-n^{-1/2}s-n^{-1}t-n^{-3/2}u\big)^3\,d\partial_1^3 F(x,s,t,u)+O(n^{-2})\\
&= \Pr(R\le x)-n^{-1/2}\frac{d}{dx}\,\mathrm{E}\big(SI(R\le x)\big)-n^{-1}\frac{d}{dx}\,\mathrm{E}\big(TI(R\le x)\big)\\
&\quad-n^{-3/2}\frac{d}{dx}\,\mathrm{E}\big(UI(R\le x)\big)+(1/2)n^{-1}\frac{d^2}{dx^2}\,\mathrm{E}\big(S^2I(R\le x)\big)\\
&\quad+n^{-3/2}\frac{d^2}{dx^2}\,\mathrm{E}\big(STI(R\le x)\big)-(1/6)n^{-3/2}\frac{d^3}{dx^3}\,\mathrm{E}\big(S^3I(R\le x)\big)+O(n^{-2})\,,
\end{aligned}$$

where (R,S,T,V) is a random vector with distribution function $F(r,s,t,v)$. By a change of variable $y=-z$,

$$\int I(z^TAz\le x)p(z)\phi(z)\,dz=0\,,$$

for every odd polynomial $p(z)$ and $A\in\mathbb{R}^{k\times k}$. Hence

$$\begin{aligned}
\Pr\big(\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1})\le x\big) &= \int I(z^T\Upsilon_0 z\le x)\big(1+n^{-1}p_2(z)\big)\phi(z)\,dz\\
&\quad-n^{-1}\frac{d}{dx}\int z^T\Upsilon_1(z\otimes I_k)zI(z^T\Upsilon_0 z\le x)p_1(z)\phi(z)\,dz\\
&\quad-n^{-1}\frac{d}{dx}\int z^T\Upsilon_2(z\otimes I_{k^2})(z\otimes I_k)zI(z^T\Upsilon_0 z\le x)\phi(z)\,dz\\
&\quad+(1/2)n^{-1}\frac{d^2}{dx^2}\int\big(z^T\Upsilon_1(z\otimes I_k)z\big)^2 I(z^T\Upsilon_0 z\le x)\phi(z)\,dz+O(n^{-2})=F(x;r)+n^{-1}R(x)+O(n^{-2})\,,
\end{aligned}$$

where $F(x;r)$ denotes the distribution function of χ^2_r. Hence,

$$\begin{aligned}
\Pr\big(\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1})>\chi^2_{r;\alpha}\big) &= 1-F(\chi^2_{r;\alpha};r)-n^{-1}R(\chi^2_{r;\alpha})+O(n^{-2})\\
&=\alpha+n^{-1}a(\alpha)+O(n^{-2})\,,
\end{aligned}$$

where $a(\alpha)=-R(\chi^2_{r;\alpha})$. Similar results hold for the LRT, ST, and J-test statistics.

B Proofs of Section 4

Let $\hat{\Psi}^*(\zeta)$ be the bootstrap version of $\hat{\Psi}(\zeta)$ given by $\hat{\Psi}^*(\zeta)=\big(\hat{C}^*(\zeta_1,\hat{W}^*)^T,\hat{C}^*(\zeta_2,\hat{V}^*(\zeta_1)^{-1})^T\big)^T$. In the proofs of this section, we use the following notation. Let $\hat{T}^*\in\mathbb{R}^k$ denote a bootstrap quantity. We say that $\hat{T}^*$ converges to 0 conditionally in probability, and write $\hat{T}^*=o_P^*(1)$, if $\lim_{n\to\infty}\Pr(\|\hat{T}^*\|\ge\varepsilon|X_{1:n})\xrightarrow{P}0$ for all $\varepsilon>0$. By definition, $\hat{T}^*$ converges (unconditionally) in probability to 0 if $\Pr(\|\hat{T}^*\|\ge\varepsilon)\to 0$ for all $\varepsilon>0$. By the bounded convergence theorem (Billingsley 1995, Theorem 16.5), conditional convergence in probability implies convergence in probability. Since convergence in L^1 implies convergence in probability, the converse holds as well. Similarly, we say that $\hat{T}^*$ is conditionally bounded in probability, and write $\hat{T}^*=O_P^*(1)$, if for all $\varepsilon>0$ there exists $M=M(\varepsilon)$ such that $\Pr(\|\hat{T}^*\|\ge M|X_{1:n})\le\varepsilon+o_P(1)\le 1$. By definition, $\hat{T}^*$ is bounded in probability if for all $\varepsilon>0$ there exists $M=M(\varepsilon)$ such that $\Pr(\|\hat{T}^*\|\ge M)\le\varepsilon$. It readily follows that if a random variable is conditionally bounded in probability then it is bounded in probability. The proof of Lemma B.1 is similar to the proof of Lemma A.1.

Lemma B.1 *Suppose that conditions A through E hold. Then (i)* $\sup_{\theta\in\Theta}|\hat{Q}^*(\theta,\hat{W}^*) - Q(\theta,W_0)| \xrightarrow{P} 0$, *(ii)* $\hat{\eta}^* \xrightarrow{P} \eta_0$, *and (iii)* $\Pr(\hat{\Psi}^*(\hat{\eta}^*) = 0|X_{1:n}) \xrightarrow{P} 1$.

Proof (Theorem 8) Similarly to (A.2),

$$n^{1/2}(\hat{\eta}^* - \hat{\eta}) = -n^{1/2}\begin{pmatrix} A_1^{-1} & O \\ -A_2^{-1}BA_1^{-1} & A_2^{-1} \end{pmatrix}\hat{\Psi}^*(\hat{\eta}) + o_P(1)\,. \tag{B.1}$$

Similarly to (A.3),

$$\begin{aligned}\hat{\Psi}^*(\hat{\eta}) &= \begin{pmatrix} G_1^T W_0(\bar{g}^*(\hat{\theta}_\mathrm{O}) - \bar{g}(\hat{\theta}_\mathrm{O})) \\ G_2^T V_1^{-1}(\bar{g}^*(\hat{\theta}_\mathrm{T}) - \bar{g}(\hat{\theta}_\mathrm{T})) \end{pmatrix} + \begin{pmatrix} \big((W_0\mu_1)^T \otimes I_p\big)\operatorname{vec}\big(\hat{G}^*(\hat{\theta}_\mathrm{O})^T - \hat{G}(\hat{\theta}_\mathrm{O})^T\big) \\ \big((V_1^{-1}\mu_2)^T \otimes I_p\big)\operatorname{vec}\big(\hat{G}^*(\hat{\theta}_\mathrm{T})^T - \hat{G}(\hat{\theta}_\mathrm{T})^T\big) \end{pmatrix} \\ &\quad + \begin{pmatrix} \big(\mu_1^T \otimes G_1^T\big)\operatorname{vec}(\hat{W}^* - \hat{W}) \\ -\big((V_1^{-1}\mu_2)^T \otimes (G_2^T V_1^{-1})\big)\operatorname{vec}(\hat{V}^*(\hat{\theta}_\mathrm{O}) - \hat{V}(\hat{\theta}_\mathrm{O})) \end{pmatrix} + o_P(n^{-1/2}) \\ &= \Pi\big(\hat{\psi}^*(\hat{\eta}) - \hat{\psi}(\hat{\eta})\big) + o_P(n^{-1/2})\,. \end{aligned}$$

By (31b) and (B.1), then (22a) holds. By the delta method for the bootstrap and (22a),

$$\mathscr{L}_\mathrm{S}(n^{1/2}h^*(\hat{\theta}_\mathrm{T}^*)|X_{1:n}) \xrightarrow{P} N(0, H_2\Gamma_{22}H_2^T)\,,$$

and thus, (22b) holds, as stated. The SB estimator $\tilde{\theta}_\mathrm{T}^*$ minimizes $\hat{Q}^*(\theta,\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1})$ subject to $h^*(\theta) = 0$. Let $(1/2)\hat{Q}^*(\theta,\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) + \lambda^T h^*(\theta)$ be the SB version of the Lagrangian and let $\hat{\lambda}^*$ be the SB version of $\hat{\lambda}$. Taking the partial derivatives of the SB Lagrangian and solving for θ and λ yields

$$\begin{cases} \hat{C}^*(\tilde{\theta}_\mathrm{T}^*,\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) + H(\tilde{\theta}_\mathrm{T}^*)^T\hat{\lambda}^* = 0\,, \\ h^*(\tilde{\theta}_\mathrm{T}^*) = 0\,. \end{cases} \tag{B.2}$$

Since $\tilde{\theta}_\mathrm{T}^* = \hat{\theta}_\mathrm{T} + O_P(n^{-1/2})$, then $\hat{\lambda}^* = O_P(n^{-1/2})$. Taylor expansion of (B.2) about $(\hat{\theta}_\mathrm{T}, 0) \in \mathbb{R}^p \times \mathbb{R}^r$ yields

$$0 = \begin{pmatrix} \hat{C}^*(\hat{\theta}_\mathrm{T},\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) \\ 0 \end{pmatrix} + \begin{pmatrix} A_2 & H_2^T \\ H_2 & O \end{pmatrix}\begin{pmatrix} \tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T} \\ \hat{\lambda}^* \end{pmatrix} + O_P(n^{-1})\,.$$

Hence

$$n^{1/2}(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}) = -n^{1/2}(A_2^{-1} - \Delta_2)\hat{C}^*(\hat{\theta}_\mathrm{T},\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) + o_P(1)\,. \tag{B.3}$$

Similarly to (B.1),

$$n^{1/2}\hat{C}^*(\hat{\theta}_\mathrm{T},\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) = -n^{1/2}A_2(\hat{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}) + o_P(1)\,, \tag{B.4}$$

and by (B.3),

$$n^{1/2}(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}^*) = n^{1/2}\Delta_2\hat{C}^*(\hat{\theta}_\mathrm{T},\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) + o_P(1)\,. \tag{B.5}$$

By (B.4), (B.5), and (22a), $\mathscr{L}_\mathrm{S}(n^{1/2}(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}^*)|X_{1:n}) \xrightarrow{P} N(0,\Omega_2)$. Taylor expansion of $\hat{Q}^*(\tilde{\theta}_\mathrm{T}^*,\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1})$ about $\hat{\theta}_\mathrm{T}^*$ yields:

$$\hat{L}^*\big(\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}\big) = n(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}^*)^T A_2(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}^*) + o_P(1)\,, \tag{B.6}$$

and thus, (22c) holds. Similarly,

$$n^{-1/2}\hat{L}^*(\hat{V}^*(\hat{\theta}_\mathrm{T}^*)^{-1}) = 2n^{1/2}\hat{C}^*(\hat{\theta}_\mathrm{T}^*,\hat{V}^*(\hat{\theta}_\mathrm{T}^*)^{-1})^T(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}^*) + o_P(1)\,,$$

and thus, (22d) holds. Taylor expansion of $\hat{C}^*(\tilde{\theta}_\mathrm{T}^*,\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1})$ about $\hat{\theta}_\mathrm{T}^*$ yields

$$n^{1/2}\hat{C}^*(\tilde{\theta}_\mathrm{T}^*,\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) = n^{1/2}A_2(\tilde{\theta}_\mathrm{T}^* - \hat{\theta}_\mathrm{T}^*) + o_P(1)\,, \tag{B.7}$$

and thus, (22e) holds. By the delta method for the bootstrap, (22f) holds. Taylor expansion of $\bar{g}^*(\hat{\theta}_\mathrm{T}^*)$ about $\hat{\theta}_\mathrm{T}$, similarly to (A.10),

$$\mathscr{L}_\mathrm{S}(n^{1/2}(\bar{g}^*(\hat{\theta}_\mathrm{T}^*) - \bar{g}(\hat{\theta}_\mathrm{T}))|X_{1:n}) \xrightarrow{P} N\big(0, D_J\Sigma D_J^T\big)\,,$$

and thus, (22g) holds. Finally, write $n^{-1}\hat{J}^*(\hat{V}^*(\hat{\theta}_\mathrm{O}^*)^{-1}) = \psi_J(\bar{g}^*(\hat{\theta}_\mathrm{T}^*), \operatorname{vec}(\hat{V}^*(\hat{\theta}_\mathrm{O}^*)))$. By the delta theorem for the bootstrap, then (22h) holds, as stated.

Proof (Theorem 9) By Lemma B.1, $\hat{\theta}^*_{\mathrm{O}} = \theta_2 + o_P(1)$ and $\hat{\theta}^*_{\mathrm{T}} = \theta_2 + o_P(1)$. Hence $\hat{A}^*(\hat{\theta}_{\mathrm{T}}, \hat{W}^*) \xrightarrow{P} A_3$, $\hat{B}^*(\hat{\theta}_{\mathrm{T}}, \hat{\theta}_{\mathrm{T}}) = o_P(1)$, and $\hat{A}^*(\hat{\theta}_{\mathrm{T}}, \hat{V}^*(\hat{\theta}_{\mathrm{T}})^{-1}) \xrightarrow{P} A_4$. Similarly to (B.1),

$$n^{1/2}(\hat{\eta}^* - \hat{\eta}_{\mathrm{c}}) = -n^{1/2} \begin{pmatrix} A_3^{-1} & O \\ O & A_4^{-1} \end{pmatrix} \hat{\Psi}^*(\hat{\eta}_{\mathrm{c}}) + o_P(1)\,. \tag{B.8}$$

Since

$$\mathscr{L}_{\mathrm{C}}(n^{1/2}\hat{\Psi}^*(\hat{\eta}_{\mathrm{c}})|X_{1:n}) \xrightarrow{P} N\left(0, \begin{pmatrix} G_2^T W_0 V_2 W_0 G_2 & A_3 \\ A_3 & A_4 \end{pmatrix}\right),$$

then (24a) holds. By the delta method for the bootstrap, then

$$\mathscr{L}_{\mathrm{C}}(n^{1/2}h^*(\hat{\theta}^*_{\mathrm{T}})|X_{1:n}) \xrightarrow{P} N(0, H_2 A_4^{-1} H_2^T)\,,$$

and since $(H_2A_4^{-1}H_2^T)^{1/2}H_2A_4^{-1}H_2^T(H_2A_4^{-1}H_2^T)^{1/2}$ is idempotent of rank r, then (24b) holds. Similarly to (B.5),

$$n^{1/2}(\tilde{\theta}^*_{\mathrm{T}} - \hat{\theta}^*_{\mathrm{T}}) = n^{1/2}\Delta_4 \hat{C}^*(\hat{\theta}_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}) + o_P(1)\,,$$

where $\Delta_4 = A_4^{-1}H_2^T(H_2A_4^{-1}H_2^T)^{-1}H_2A_4^{-1}$. Hence $\mathscr{L}_{\mathrm{C}}(n^{1/2}(\tilde{\theta}^*_{\mathrm{T}} - \hat{\theta}^*_{\mathrm{T}})|X_{1:n}) \xrightarrow{P} N(0, \Delta_4)$. Since the matrix $A_4^{-1/2}\Delta_4 A_4^{-1/2}$ is idempotent of rank r, (24c) holds. Since $\bar{g}^*(\hat{\theta}^*_{\mathrm{T}}) = o_P(1)$, then

$$\hat{C}^*(\hat{\theta}^*_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) = \hat{G}^*(\hat{\theta}^*_{\mathrm{T}})^T(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1} - \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1})\bar{g}^*(\hat{\theta}^*_{\mathrm{T}}) = o_P(n^{-1/2})\,. \tag{B.9}$$

Hence $\hat{L}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) = \hat{L}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}) + o_P(1)$, and thus, (24d) holds. Similarly to (B.7) and by (B.9), then (24e) and (24f) hold. By (B.8), Taylor expansion of $\bar{g}^*(\hat{\theta}^*_{\mathrm{T}})$ about $\hat{\theta}_{\mathrm{T}}$ yields:

$$n^{1/2}\bar{g}^*(\hat{\theta}^*_{\mathrm{T}}) = n^{1/2}(I_q - G_2A_4^{-1}G_2^TV_2^{-1})\bar{g}^*(\hat{\theta}_{\mathrm{T}}) + o_P(1)\,. \tag{B.10}$$

Since $\mathscr{L}_{\mathrm{C}}(\bar{g}^*(\hat{\theta}_{\mathrm{T}})|X_{1:n}) \xrightarrow{P} N(0, V_2)$ and $\Lambda_4 = I_q - V_2^{-1/2}G_2A_4^{-1}G_2^TV_2^{-1/2}$ is idempotent of rank $q-p$, then (24g) holds, as stated.

C Proofs of Section 3

Proof (Theorem 2) By Remark A.1 and Theorem 8, then (9a), (9b), (9c), and (9d) hold. By a Taylor expansion of $\bar{g}^*(\hat{\theta}^*_{\mathrm{T}})$ about $\hat{\theta}_{\mathrm{T}}$ and (9a), then

$$n^{1/2}\bar{g}^*(\hat{\theta}^*_{\mathrm{T}}) = n^{1/2}(I_q - G_0D_0^{-1}G_0^TV_0^{-1})\bar{g}^*(\hat{\theta}_{\mathrm{T}}) + o_P(1)\,. \tag{C.1}$$

Hence

$$\hat{J}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}) = n\bar{g}^*(\hat{\theta}_{\mathrm{T}})^TV_0^{-1/2}\Lambda_2V_0^{-1/2}\bar{g}^*(\hat{\theta}_{\mathrm{T}}) + o_P(1)\,, \tag{C.2}$$

where recall that $\Lambda_2 = I_q - V_0^{-1/2}G_0D_0^{-1}G_0^TV_0^{-1/2}$. Since

$$\mathscr{L}_{\mathrm{S}}(n^{1/2}V_0^{-1/2}(\bar{g}^*(\hat{\theta}_{\mathrm{T}}) - \bar{g}(\hat{\theta}_{\mathrm{T}}))|X_{1:n}) \xrightarrow{P} N(0, I_q)$$

and $n^{1/2}V_0^{-1/2}\bar{g}(\hat{\theta}_{\mathrm{T}}) \xrightarrow{\mathrm{d}} Z$, where $Z \sim N(0, \Lambda_2)$, by Slutsky's theorem in the product metric space $\mathscr{P}^q \times \mathbb{R}^q$ (van der Vaart 1998, Theorem 18.10(v)),

$$\big(\mathscr{L}_{\mathrm{S}}(n^{1/2}V_0^{-1/2}(\bar{g}^*(\hat{\theta}_{\mathrm{T}}) - \bar{g}(\hat{\theta}_{\mathrm{T}}))|X_{1:n}), n^{1/2}V_0^{-1/2}\bar{g}(\hat{\theta}_{\mathrm{T}})\big) \xrightarrow{\mathrm{d}} (N(0, I_q), Z)\,.$$

Let the mapping $\varphi : \mathscr{P}^q \times \mathbb{R}^q \mapsto \mathscr{P}^1$ be given by $\varphi(P, x) = \mathscr{L}\big((X+x)^T\Lambda_2(X+x)\big)$, where $X \sim P$, $P \in \mathscr{P}^q$, and $x \in \mathbb{R}^q$. By Corollary 13 of Fristedt and Gray (1997, p. 414) and (C.2),

$$\mathscr{L}_{\mathrm{S}}(\hat{J}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1})|X_{1:n}) = \varphi\big(\mathscr{L}_{\mathrm{S}}(n^{1/2}V_0^{-1/2}(\bar{g}^*(\hat{\theta}_{\mathrm{T}}) - \bar{g}(\hat{\theta}_{\mathrm{T}}))|X_{1:n}), n^{1/2}V_0^{-1/2}\bar{g}(\hat{\theta}_{\mathrm{T}})\big) + o_P(1)\,.$$

Since φ is continuous, by the continuous mapping theorem,

$$\varphi\big(\mathscr{L}_{\mathrm{S}}(n^{1/2}V_0^{-1/2}(\bar{g}^*(\hat{\theta}_{\mathrm{T}}) - \bar{g}(\hat{\theta}_{\mathrm{T}}))|X_{1:n}), V_0^{-1/2}n^{1/2}\bar{g}(\hat{\theta}_{\mathrm{T}})\big) \xrightarrow{\mathrm{d}} \varphi(N(0, I_q), Z)\,.$$

Since $\varphi(N(0, I_q), x) = \chi^2_{q-p}(x^T\Lambda_2x)$ and $Z^T\Lambda_2Z \sim \chi^2_{q-p}$, then (9e) holds, as stated.

Proof (Corollary 1) By Theorem 2 and continuous mapping theorem, then

$$\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) - \hat{\xi}^{\mathrm{J}}_{\mathrm{S},\alpha} \xrightarrow{\mathrm{d}} U^2 - \chi^2_{q-p;\alpha}(U^2)\,,$$

where $U^2 \sim \chi^2_{q-p}$ and $\chi^2_{q-p;\alpha}(u^2)$ is the upper α-quantile of the non-central chi-squared distribution $\chi^2_{q-p}(u^2)$ with non-centrality parameter u^2. Hence,

$$\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) > \hat{\xi}^{\mathrm{J}}_{\mathrm{S},\alpha}\big) \to \Pr(U - \chi^2_{q-p;\alpha}(U) > 0)\,. \tag{C.3}$$

By a median-mode inequality for non-central chi-squared distributions (see, e.g., Sen (1989, Theorem 2) or Robert (1990, Proposition 4.1)), $U^2 - \chi^2_{q-p;\alpha}(U^2) < 0$ with probability one for $\alpha \le 1/2$. This completes the proof of the corollary.

Proof (Theorem 3) Note first that

$$\begin{aligned}
\Pr{}_{\mathrm{S}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \le x|X_{1:n}\big) &= \Pr{}_{\mathrm{S}}\big(\hat{Z}^{*T}\hat{\Upsilon}_0\hat{Z}^* \le x|X_{1:n}\big) \\
&\quad - n^{-1/2}\frac{d}{dx}\mathrm{E}_{\mathrm{S}}\big(\hat{Z}^{*T}\hat{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^* I\big(\hat{Z}^{*T}\hat{\Upsilon}_0\hat{Z}^* \le x\big)|X_{1:n}\big) \\
&\quad - n^{-1}\frac{d}{dx}\mathrm{E}_{\mathrm{S}}\big(\hat{Z}^{*T}\hat{\Upsilon}_2(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\hat{Z}^* I\big(\hat{Z}^{*T}\hat{\Upsilon}_0\hat{Z}^* \le x\big)|X_{1:n}\big) \\
&\quad - n^{-3/2}\frac{d}{dx}\mathrm{E}_{\mathrm{S}}\big(\hat{Z}^{*T}\hat{\Upsilon}_3(\hat{Z}^*\otimes I_{k^3})(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\hat{Z}^* I\big(\hat{Z}^{*}\hat{\Upsilon}_0\hat{Z}^* \le x\big)|X_{1:n}\big) \\
&\quad + (1/2)n^{-1}\frac{d^2}{dx^2}\mathrm{E}_{\mathrm{S}}\big((\hat{Z}^{*T}\hat{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^*)^2 I\big(\hat{Z}^{*T}\hat{\Upsilon}_0\hat{Z}^* \le x\big)|X_{1:n}\big) \\
&\quad + n^{-3/2}\frac{d^2}{dx^2}\mathrm{E}_{\mathrm{S}}\big((\hat{Z}^{*T}\hat{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^*)(\hat{Z}^{*T}\hat{\Upsilon}_2(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\hat{Z}^*)I\big(\hat{Z}^{*T}\hat{\Upsilon}_0\hat{Z}^* \le x\big)|X_{1:n}\big) \\
&\quad - (1/6)n^{-3/2}\frac{d^3}{dx^3}\mathrm{E}_{\mathrm{S}}\big((\hat{Z}^{*T}\hat{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^*)^3 I\big(\hat{Z}^{*T}\hat{\Upsilon}_0\hat{Z}^* \le x\big)|X_{1:n}\big) + O_P(n^{-2})\,.
\end{aligned}$$

We further obtain

$$\begin{aligned}
\Pr{}_{\mathrm{S}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \le x|X_{1:n}\big) &= \int I\big(z^T\hat{\Upsilon}_0 z \le x\big)\big(1 + n^{-1/2}\hat{p}_1(z) + n^{-1}\hat{p}_2(z) + n^{-3/2}\hat{p}_3(z)\big)\phi(z)\,dz \\
&\quad - n^{-1/2}\frac{d}{dx}\int z^T\hat{\Upsilon}_1(z\otimes I_k)z\big(1 + n^{-1/2}\hat{p}_1(z) + n^{-1}\hat{p}_2(z)\big)I\big(z^T\hat{\Upsilon}_0 z \le x\big)\phi(z)\,dz \\
&\quad - n^{-1}\frac{d}{dx}\int z^T\hat{\Upsilon}_2(z\otimes I_{k^2})(z\otimes I_k)z\big(1 + n^{-1/2}\hat{p}_1(z)\big)I\big(z^T\hat{\Upsilon}_0 z \le x\big)\phi(z)\,dz \\
&\quad - n^{-3/2}\frac{d}{dx}\int z^T\hat{\Upsilon}_3(z\otimes I_{k^3})(z\otimes I_{k^2})(z\otimes I_k)zI\big(z^T\hat{\Upsilon}_0 z \le x\big)\phi(z)\,dz \\
&\quad + (1/2)n^{-1}\frac{d^2}{dx^2}\int\big(z^T\hat{\Upsilon}_1(z\otimes I_k)z\big)^2\big(1 + n^{-1/2}\hat{p}_1(z)\big)I\big(z^T\hat{\Upsilon}_0 z \le x\big)\phi(z)\,dz \\
&\quad + n^{-3/2}\frac{d^2}{dx^2}\int\big(z^T\hat{\Upsilon}_1(z\otimes I_k)z\big)\big(z^T\hat{\Upsilon}_2(z\otimes I_{k^2})(z\otimes I_k)z\big)I\big(z^T\hat{\Upsilon}_0 z \le x\big)\phi(z)\,dz \\
&\quad - (1/6)n^{-3/2}\frac{d^3}{dx^3}\int\big(z^T(z\otimes I_k)z\big)^3 I\big(z^T\hat{\Upsilon}_0 z \le x\big)\phi(z)\,dz + O_P(n^{-2})\,.
\end{aligned}$$

Hence

$$\Pr{}_{\mathrm{S}}(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \le x|X_{1:n}) = F(x;r) + n^{-1}\hat{R}(x) + O_P(n^{-2})\,,$$

where $\hat{R}(x) = R(x) + O_P(n^{-1/2})$. Let $\hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha} = \chi^2_{r;\alpha} + n^{-1}\hat{q}_1(\alpha) + O_P(n^{-2})$ be an empirical Cornish-Fisher expansion of the upper α-quantile of $\mathscr{L}_{\mathrm{S}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1})|X_{1:n}\big)$. Taylor expansion of $\Pr_{\mathrm{S}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \le \hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha}|X_{1:n}\big)$ about $\chi^2_{r;\alpha}$ yields

$$1-\alpha = \Pr{}_{\mathrm{S}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \le \hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha}|X_{1:n}\big) = 1 - \alpha + n^{-1}F'(\chi^2_{r;\alpha};r)\hat{q}_1(\alpha) + n^{-1}\hat{R}(\chi^2_{r;\alpha}) + O_P(n^{-2})\,.$$

Hence, $\hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha} = \chi^2_{r;\alpha} - n^{-1}\big(F'(\chi^2_{r;\alpha})\big)^{-1}\hat{R}(\chi^2_{r;\alpha})$. Thus, $\Pr\big(\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) > \hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha}\big) = \alpha + O(n^{-2})$, as stated. Similar proofs hold for the LRT, ST, and J test statistics.

Proof (Theorem 4) By Remark A.1 and (24a)–(24g), then (14a)–(14e) hold. This completes the proof of the theorem.

Proof (Theorem 5) Proofs of (15a)–(15c) are similar to the proofs of (11a)–(11c). The proof of (15d) follows similarly using the following expansion of the CB version of the J-test statistic:

$$\hat{J}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) = \hat{Z}^{*T}\big(\hat{\Phi}_0 + n^{-1/2}\hat{\Phi}_1(\hat{Z}^*\otimes I_k) + n^{-1}\hat{\Phi}_2(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k) + n^{-3/2}\hat{\Phi}_3(\hat{Z}^*\otimes I_{k^3})(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\big)\hat{Z}^* + O_P(n^{-2})\,.$$

Lemmas C.1–C.3 describe the asymptotic properties of the ELB weights using similar methods of proof as Owen (1990).

Lemma C.1 *Suppose the model* (1) *and conditions A–E hold. Then there exists* $\varepsilon > 0$ *such that*

$$\inf_{\|\theta-\theta_0\|\le\varepsilon}\inf_{\|\rho\|=1}\Pr(\rho^T g(X,\theta) < 0) > 0\,.$$

Proof (Lemma C.1) Let $\mathbb{S}^{q-1} = \{\rho\in\mathbb{R}^q : \|\rho\| = 1\}$ denote the unit sphere in $\mathbb{R}^q$. Suppose, to the contrary, that the conclusion of the lemma is false. Then, there exist $\theta_n\in\mathbb{R}^p$ and $\rho_n\in\mathbb{S}^{q-1}$ such that $\theta_n\to\theta_0$ and $\Pr(\rho_n^T g(X,\theta_n) < 0) < n^{-1}$. Since $\mathbb{S}^{q-1}$ is compact, there exists $\rho_0\in\mathbb{S}^{p-1}$ and a subsequence m_n such that $\rho_{m_n}\to\rho_0$. Since $g(x,\theta)$ is continuous at θ_0, then $\rho_{m_n}^T g(x,\theta_{m_n})\to\rho_0^T g(x,\theta_0)$ for P_0-almost all $x\in\mathscr{X}$. Hence, if $\rho_0^T g(x,\theta_0) < 0$ then $\rho_{m_n}^T g(x,\theta_{m_n}) < 0$ for n sufficiently large. Thus, for P_0-almost all $x\in\mathscr{X}$,

$$I(\rho_0^T g(x,\theta_0) < 0) \le \liminf_n I(\rho_{m_n}^T g(x,\theta_{m_n}) < 0)\,.$$

By Fatou's lemma,

$$\Pr(\rho_0^T g(X,\theta_0) < 0) = \mathrm{E}\big(I(\rho_0^T g(X,\theta_0) < 0)\big) \le \mathrm{E}\big(\liminf_n I(\rho_{m_n}^T g(X,\theta_{m_n}) < 0)\big) \le \liminf_n \mathrm{E}\big(I(\rho_{m_n}^T g(X,\theta_{m_n}) < 0)\big) \le \liminf_n m_n^{-1} = 0\,.$$

Since $\mathrm{E}(\rho_0^T g(X,\theta_0)) = 0$, then $\rho_0^T g(X,\theta_0) = 0$ with probability one. Hence $\mathrm{Var}(\rho_0^T g(X,\theta_0)) = \rho_0^T V_0\rho_0 = 0$, contradicting the assumption that V_0 is nonsingular.

Lemma C.2 *Suppose the model* (1) *and conditions A–E hold. Then*

$$\Pr\big\{0\in\mathrm{int}\big(\mathrm{conv}\{g(X_i,\hat{\theta}_T) : 1\le i\le n\}\big)\big\}\to 1\,,$$

where $\mathrm{int}(A)$ *denotes the interior of a set A and* $\mathrm{conv}\{g(X_i,\hat{\theta}_T) : 1\le i\le n\}$ *denotes the convex hull of* $\{g(X_i,\hat{\theta}_T) : 1\le i\le n\}$.

Proof (Lemma C.2) Let ε be given as in Lemma C.1. It suffices to show that $\Pr(A_n)\to 0$, where

$$A_n = \big\{0\notin\cap_{\theta\in\bar{B}(\theta_0,\varepsilon)}\mathrm{conv}\{g(X_i,\theta) : 1\le i\le n\}\big\}\,,$$

and $\bar{B}(\theta_0,\varepsilon)$ denotes the closed ball of radius ε about θ_0. By the separating hyperplane theorem (see, e.g., Rockafellar 1970, Theorem 11.4),

$$A_n = \Big\{\sup_{\theta\in\bar{B}(\theta_0,\varepsilon)}\sup_{\|\rho\|=1} n^{-1}\sum_{i=1}^n I\big(\rho^T g(X_i,\theta) > 0\big) = 1\Big\}\,.$$

Since

$$\sup_{\theta\in\bar{B}(\theta_0,\varepsilon)}\sup_{\|\rho\|=1}\Pr(\rho^T g(X,\theta)\ge 0) = 1 - \inf_{\theta\in\bar{B}(\theta_0,\varepsilon)}\inf_{\|\rho\|=1}\Pr(\rho^T g(X,\theta) < 0)\,,$$

by Lemma C.1, then

$$\sup_{\theta\in\bar{B}(\theta_0,\varepsilon)}\sup_{\|\rho\|=1}\Pr\big(\rho^T g(X,\theta)\ge 0\big) < 1\,.$$

Thus $\Pr(\rho^T g(X,\theta) > 0) < 1$ for all $\theta \in \bar{B}(\theta_0,\varepsilon)$ and $\rho \in \mathbb{S}^{q-1}$. For $(\rho,\theta) \in \mathbb{S}^{q-1} \times \bar{B}(\theta_0,\varepsilon)$, let

$$g_m(x;\rho,\theta) = \max_{(\rho',\theta') \in \bar{B}((\rho,\theta),m^{-1})} \rho'^T g(x,\theta') .$$

Since $g(x,\theta)$ is continuous in θ for almost all $x \in \mathscr{X}$, $g_m(X;\rho,\theta) \to \rho^T g(X,\theta)$ with probability one. Thus, for all $\theta \in \bar{B}(\theta_0,\varepsilon)$ and $\rho \in \mathbb{S}^{q-1}$, there exists $m_{\rho,\theta} = m(\rho,\theta)$ such that $\Pr\big(g_{m_{\rho,\theta}}(X;\rho,\theta) > 0\big) < 1$. Since

$$\mathbb{S}^{p-1} \times \bar{B}(\theta_0,\varepsilon) \subseteq \bigcup_{(\rho,\theta)\in\mathbb{S}^{p-1}\times\bar{B}(\theta_0,\varepsilon)} B\big((\rho,\theta),m_{\rho,\theta}^{-1}\big)$$

and $\mathbb{S}^{p-1} \times \bar{B}(\theta_0,\varepsilon)$ is compact, there exists a finite collection of open balls such that

$$\mathbb{S}^{q-1} \times \bar{B}(\theta_0,\varepsilon) \subseteq \bigcup_{s=1}^{k} B\big((\rho_s,\theta_s),m_{\rho_s,\theta_s}^{-1}\big) .$$

Therefore, with probability one,

$$\sup_{\theta\in\bar{B}(\theta_0,\varepsilon)} \sup_{\|\rho\|=1} n^{-1}\sum_{i=1}^n I\big(\rho^T g(X_i,\theta) > 0\big) \le \max_{1\le s\le k} n^{-1}\sum_{i=1}^n I\big(g_{m_{\rho_s,\theta_s}}(X_i;\rho_s,\theta_s) > 0\big) .$$

Since

$$n^{-1}\sum_{i=1}^n I\big(g_{m_{\rho_s,\theta_s}}(X_i;\rho_s,\theta_s) > 0\big) \to \Pr\big(g_{m_{\rho_s,\theta_s}}(X;\rho_s,\theta_s) > 0\big) < 1$$

with probability one for all $s = 1,\dots,k$, we conclude that $\Pr(A_n) \to 0$, as stated.

It is known that when $0 \in \operatorname{int}\big(\operatorname{conv}\{g(X_i,\hat{\theta}_{\mathrm{T}}) : 1 \le i \le n\}\big)$, the ELB weights $\hat{w}_i$ have the following expressions (Owen 1990):

$$\hat{w}_i = \frac{1}{n\big(1+\hat{\lambda}^T g(X_i,\hat{\theta}_{\mathrm{T}})\big)}, \text{ where } \hat{\lambda} \text{ satisfies } \sum_{i=1}^n \hat{w}_i g(X_i,\hat{\theta}_{\mathrm{T}}) = 0 . \tag{C.4}$$

Lemma C.3 *Suppose the model* (1) *and conditions A–E hold. Then (i)* $\hat{\lambda} = O_P(n^{-1/2})$*; (ii)* $\max_{1\le i\le n}\hat{w}_i = O_P(n^{-1})$*; and (iii)* $\max_{1\le i\le n}|\hat{w}_i - n^{-1}| = o_P(n^{-1})$.

Proof (Lemma C.3) By (C.4), we write

$$n^{-1} - \hat{w}_i = \hat{w}_i g(X_i,\hat{\theta}_{\mathrm{T}})^T\hat{\lambda} , \quad 1 \le i \le n . \tag{C.5}$$

Left multiplication of (C.5) by $g(X_i,\hat{\theta}_{\mathrm{T}})$ and summing over i yields $\tilde{V}(\hat{\theta}_{\mathrm{T}})\hat{\lambda} = \bar{g}(\hat{\theta}_{\mathrm{T}})$, where $\tilde{V}(\theta) = \sum_{i=1}^n \hat{w}_i g(X_i,\theta)g(X_i,\theta)^T$. Write $\hat{\lambda} = \|\hat{\lambda}\|\hat{\rho}$, with $\|\hat{\rho}\| = 1$. Since $\hat{w}_i \ge n^{-1}(1+\|\hat{\lambda}\|\hat{M})^{-1}$, where $\hat{M} = \max_{1\le i\le n}\|g(X_i,\hat{\theta}_{\mathrm{T}})\|$, we have

$$\begin{aligned}\|\hat{\lambda}\|\hat{\rho}^T\hat{V}_{\mathrm{U}}(\hat{\theta}_{\mathrm{T}})\hat{\rho} &\le (1+\|\hat{\lambda}\|\hat{M})\|\hat{\lambda}\|\sum_{i=1}^n \hat{w}_i\hat{\rho}^T g(X_i,\hat{\theta}_{\mathrm{T}})g(X_i,\hat{\theta}_{\mathrm{T}})^T\hat{\rho} \\ &= (1+\|\hat{\lambda}\|\hat{M})\hat{\rho}^T\tilde{V}(\hat{\theta}_{\mathrm{T}})\hat{\lambda} = (1+\|\hat{\lambda}\|\hat{M})\hat{\rho}^T\bar{g}(\hat{\theta}_{\mathrm{T}}) ,\end{aligned}$$

where recall that $\hat{V}_{\mathrm{U}}(\theta) = n^{-1}\sum_{i=1}^n g(X_i,\theta)g(X_i,\theta)^T$. Thus, with probability one,

$$\|\hat{\lambda}\|\big(\hat{\rho}^T\hat{V}_{\mathrm{U}}(\hat{\theta}_{\mathrm{T}})\hat{\rho} - \hat{M}\hat{\rho}^T\bar{g}(\hat{\theta}_{\mathrm{T}})\big) \le \hat{\rho}^T\bar{g}(\hat{\theta}_{\mathrm{T}}) .$$

Let $\sigma_q > 0$ be the smallest eigenvalue of V_0. Since $\hat{M} = o_P(n^{1/2})$, $\hat{V}_{\mathrm{U}}(\hat{\theta}_{\mathrm{T}}) = V_0 + o_P(1)$, and $\bar{g}(\hat{\theta}_{\mathrm{T}}) = O_P(n^{-1/2})$, then $\|\hat{\lambda}\|(\sigma_q + o_P(1)) \le O_P(n^{-1/2})$. Hence $\hat{\lambda} = O_P(n^{-1/2})$, as stated at (i). Since $\|\hat{\lambda}\|\hat{M} = o_P(1)$, then $\Pr(\|\hat{\lambda}\|\hat{M} < 1) \to 1$. Note that on the event $\{\|\hat{\lambda}\|\hat{M} < 1\}$, we have

$$(1+\hat{\lambda}^T g(X_i,\hat{\theta}_{\mathrm{T}}))^{-1} \le (1-\|\hat{\lambda}\|\hat{M})^{-1} .$$

Thus, $\max_{1\le i\le n}\hat{w}_i = O_P(n^{-1})$, as stated at (ii). Finally,

$$\max_{1\le i\le n}|\hat{w}_i - n^{-1}| \le \|\hat{\lambda}\|\max_{1\le i\le n}\hat{w}_i\|g(X_i,\hat{\theta}_{\mathrm{T}})\| \le O_P(n^{-1/2})O_P(n^{-1})o_P(n^{1/2}) = o_P(n^{-1}) ,$$

as stated at (iii). This completes the proof of the lemma.

Lemma C.4 *Suppose that the model* (1) *and conditions A–E hold. Then,*

$$\mathscr{L}_{\mathrm{E}}(n^{1/2}\bar{g}^*(\hat{\theta}_T)|X_{1:n}) \xrightarrow{P} N(0,V_0)\,.$$

Proof (Lemma C.4) Note that $\bar{g}^*(\hat{\theta}_{\mathrm{T}})$ is the average of n conditionally i.i.d. random variables, with conditional mean

$$\mathrm{E}_{\mathrm{E}}(g^*(X^*,\hat{\theta}_{\mathrm{T}})|X_{1:n}) = \sum_{i=1}^n \hat{w}_i g(X_i,\hat{\theta}_{\mathrm{T}}) = 0$$

and conditional covariance matrix

$$\tilde{V}(\hat{\theta}_{\mathrm{T}}) = \mathrm{Var}_{\mathrm{E}}(g^*(X^*,\hat{\theta}_{\mathrm{T}})|X_{1:n}) = \sum_{i=1}^n \hat{w}_i g(X_i,\hat{\theta}_{\mathrm{T}}) g(X_i,\hat{\theta}_{\mathrm{T}})^T\,.$$

By Lemma C.3,

$$\|\tilde{V}(\hat{\theta}_{\mathrm{T}}) - \hat{V}_{\mathrm{U}}(\hat{\theta}_{\mathrm{T}})\| \le \max_{1\le i\le n}|\hat{w}_i - n^{-1}| \sum_{i=1}^n \|g(X_i,\hat{\theta}_{\mathrm{T}})\|^2 = o_P(n^{-1})O_P(n) = o_P(1)\,,$$

and since $\hat{V}_{\mathrm{U}}(\hat{\theta}_{\mathrm{T}}) = V_0 + o_P(1)$, then $\tilde{V}(\hat{\theta}_{\mathrm{T}}) = V_0 + o_P(1)$. To complete the proof, we prove that a (conditional) Lindeberg condition holds; i.e., for every $\varepsilon > 0$,

$$\mathrm{E}_{\mathrm{E}}\big(\|g^*(X^*,\hat{\theta}_{\mathrm{T}})\|^2 I(\|g^*(X^*,\hat{\theta}_{\mathrm{T}})\| \ge \varepsilon n^{1/2})|X_{1:n}\big) = o_P(1)\,.$$

To this end, note that

$$\begin{aligned}\mathrm{E}_{\mathrm{E}}\big(\|g^*(X^*,\hat{\theta}_{\mathrm{T}})\|^2 I(\|g^*(X^*,\hat{\theta}_{\mathrm{T}})\| \ge \varepsilon n^{1/2})|X_{1:n}\big) &= \sum_{i=1}^n \hat{w}_i \|g(X_i,\hat{\theta}_{\mathrm{T}})\|^2 I(\|g(X_i,\hat{\theta}_{\mathrm{T}})\| \ge \varepsilon n^{1/2}) \\ &\le \big\{\max_{1\le i\le n} \hat{w}_i\big\} I(\hat{M} \ge \varepsilon n^{1/2}) \sum_{i=1}^n \|g(X_i,\hat{\theta}_{\mathrm{T}})\|^2 = O_P(n^{-1})o_P(1)O_P(n) = o_P(1)\,.\end{aligned}$$

This completes the proof of the lemma.

Proof (Theorem 6) Similarly to (B.1),

$$n^{1/2}(\hat{\eta}^* - \hat{\eta}) = -n^{1/2}\begin{pmatrix} A_1^{-1} & O \\ O & A_2^{-1}\end{pmatrix}\hat{\Psi}^*(\hat{\eta}) + o_P(1)\,. \tag{C.6}$$

Similarly to the proof of Lemma C.4,

$$\mathscr{L}_{\mathrm{E}}\left(n^{1/2}\begin{pmatrix}\bar{g}^*(\hat{\theta}_{\mathrm{O}}) - \tilde{g}(\hat{\theta}_{\mathrm{O}}) \\ \bar{g}^*(\hat{\theta}_{\mathrm{T}})\end{pmatrix}\Big|X_{1:n}\right) \xrightarrow{P} N\left(0, \begin{pmatrix} V_0 & V_0 \\ V_0 & V_0\end{pmatrix}\right)\,, \tag{C.7}$$

where $\tilde{g}(\theta) = \sum_{i=1}^n \hat{w}_i g(X_i,\theta)$. Since $n^{1/2}\hat{G}^*(\hat{\theta}_{\mathrm{O}})\hat{W}^*\tilde{g}(\hat{\theta}_{\mathrm{O}}) = o_P(1)$, by (C.6) and (C.7), then (20a) holds. By the delta method for the bootstrap, then (20b) holds. Similarly to (B.5),

$$\mathscr{L}_{\mathrm{E}}(n^{1/2}(\tilde{\theta}_{\mathrm{T}}^* - \hat{\theta}_{\mathrm{T}})|X_{1:n}) \xrightarrow{P} N(0,\Delta_2)\,.$$

Similarly to (B.6) and (B.7), then (20c) and (20d) hold. Similarly to (C.1),

$$n^{1/2}V_0^{-1/2}\bar{g}^*(\hat{\theta}_{\mathrm{T}}^*) = n^{1/2}\Lambda_2 V_0^{-1/2}\bar{g}^*(\hat{\theta}_{\mathrm{T}}) + o_P(1)\,,$$

and by Lemma C.4, then (20e) holds, as stated.

Proof (Theorem 7) Note first that by (19), $\lim_{n\to\infty} n^{-\beta}\chi^2_{q-p;\alpha_n} = 1$. Next, by a large deviation theorem (see, e.g., Dembo and Zeitouni 2009, Theorem 3.7.1), for every $C > 0$, there exists $c > 0$ such that

$$\limsup_{n\to\infty} n^{-\beta} \log\big(\Pr(\|\hat{Z}\| \geq Cn^{\beta/2})\big) \leq -c\,.$$

Hence, for n large enough, $\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \leq \chi^2_{q-p;\alpha_n}\big) \geq 1-\exp(-cn^{\beta})$. Thus, on an event of probability exponentially close to 1, the ELB weights are not modified to the uniform weights. Note further that

$$\begin{aligned}
\Pr{}_{\mathrm{E}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \leq x|X_{1:n}\big) &= \Pr{}_{\mathrm{E}}\big(\hat{Z}^{*T}\tilde{\Upsilon}_0\hat{Z}^* \leq x|X_{1:n}\big)\\
&\quad - n^{-1/2}\frac{d}{dx}\,\mathrm{E}_{\mathrm{E}}\big(\hat{Z}^{*T}\tilde{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^* I\big(\hat{Z}^{*T}\tilde{\Upsilon}_0\hat{Z}^* \leq x\big)|X_{1:n}\big)\\
&\quad - n^{-1}\frac{d}{dx}\,\mathrm{E}_{\mathrm{E}}\big(\hat{Z}^{*T}\tilde{\Upsilon}_2(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\hat{Z}^* I\big(\hat{Z}^{*T}\tilde{\Upsilon}_0\hat{Z}^* \leq x\big)|X_{1:n}\big)\\
&\quad - n^{-3/2}\frac{d}{dx}\,\mathrm{E}_{\mathrm{E}}\big(\hat{Z}^{*T}\tilde{\Upsilon}_3(\hat{Z}^*\otimes I_{k^3})(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\hat{Z}^* I\big(\hat{Z}^*\tilde{\Upsilon}_0\hat{Z}^* \leq x\big)|X_{1:n}\big)\\
&\quad + (1/2)n^{-1}\frac{d^2}{dx^2}\,\mathrm{E}_{\mathrm{E}}\big((\hat{Z}^{*T}\tilde{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^*)^2 I\big(\hat{Z}^{*T}\tilde{\Upsilon}_0\hat{Z}^* \leq x\big)|X_{1:n}\big)\\
&\quad + n^{-3/2}\frac{d^2}{dx^2}\,\mathrm{E}_{\mathrm{E}}\big((\hat{Z}^{*T}\tilde{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^*)(\hat{Z}^{*T}\tilde{\Upsilon}_2(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\hat{Z}^*) I\big(\hat{Z}^{*T}\tilde{\Upsilon}_0\hat{Z}^* \leq x\big)|X_{1:n}\big)\\
&\quad - (1/6)n^{-3/2}\frac{d^3}{dx^3}\,\mathrm{E}_{\mathrm{E}}\big((\hat{Z}^{*T}\tilde{\Upsilon}_1(\hat{Z}^*\otimes I_k)\hat{Z}^*)^3 I\big(\hat{Z}^{*T}\tilde{\Upsilon}_0\hat{Z}^* \leq x\big)|X_{1:n}\big) + O_P(n^{-2})\,,
\end{aligned}$$

where $\tilde{\Upsilon}_j$ are the ELB versions of Υ_j. We further obtain

$$\begin{aligned}
\Pr{}_{\mathrm{E}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \leq x|X_{1:n}\big) &= \int I\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\big(1 + n^{-1/2}\tilde{p}_1(z) + n^{-1}\tilde{p}_2(z) + n^{-3/2}\tilde{p}_3(z)\big)\phi(z)\,dz\\
&\quad - n^{-1/2}\frac{d}{dx}\int z^T\tilde{\Upsilon}_1(z\otimes I_k)z\big(1 + n^{-1/2}\tilde{p}_1(z) + n^{-1}\tilde{p}_2(z)\big)I\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\phi(z)\,dz\\
&\quad - n^{-1}\frac{d}{dx}\int z^T\tilde{\Upsilon}_2(z\otimes I_{k^2})(z\otimes I_k)z\big(1 + n^{-1/2}\tilde{p}_1(z)\big)I\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\phi(z)\,dz\\
&\quad - n^{-3/2}\frac{d}{dx}\int z^T\tilde{\Upsilon}_3(z\otimes I_{k^3})(z\otimes I_{k^2})(z\otimes I_k)zI\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\phi(z)\,dz\\
&\quad + (1/2)n^{-1}\frac{d^2}{dx^2}\int \big(z^T\tilde{\Upsilon}_1(z\otimes I_k)z\big)^2\big(1 + n^{-1/2}\tilde{p}_1(z)\big)I\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\phi(z)\,dz\\
&\quad + n^{-3/2}\frac{d^2}{dx^2}\int \big(z^T\tilde{\Upsilon}_1(z\otimes I_k)z\big)\big(z^T\tilde{\Upsilon}_2(z\otimes I_{k^2})(z\otimes I_k)z\big)I\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\phi(z)\,dz\\
&\quad - (1/6)n^{-3/2}\frac{d^3}{dx^3}\int \big(z^T(z\otimes I_k)z\big)^3 I\big(z^T\tilde{\Upsilon}_0 z \leq x\big)\phi(z)\,dz + O_P(n^{-2})\,,
\end{aligned}$$

where $\tilde{p}_j(z)$ is the ELB version of $p_j(z)$. Hence

$$\Pr{}_{\mathrm{E}}(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \leq x|X_{1:n}) = F(x;r) + n^{-1}\tilde{R}(x) + O_P(n^{-2})\,,$$

where $\tilde{R}(x) = R(x) + O_P(n^{-1/2})$. Let $\hat{\xi}^{\mathrm{T}}_{\mathrm{E},\alpha} = \chi^2_{r;\alpha} + n^{-1}\tilde{q}_1(\alpha) + O_P(n^{-2})$ be an empirical Cornish-Fisher expansion of the upper α-quantile of $\mathscr{L}_{\mathrm{E}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1})|X_{1:n}\big)$. Taylor expansion of $\Pr_{\mathrm{E}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \leq \hat{\xi}^{\mathrm{T}}_{\mathrm{E},\alpha}|X_{1:n}\big)$ about $\chi^2_{r;\alpha}$ yields:

$$1-\alpha = \Pr{}_{\mathrm{E}}\big(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) \leq \hat{\xi}^{\mathrm{T}}_{\mathrm{E},\alpha}|X_{1:n}\big) = 1-\alpha + n^{-1}F'(\chi^2_{r;\alpha};r)\tilde{q}_1(\alpha) + n^{-1}\tilde{R}(\chi^2_{r;\alpha}) + O_P(n^{-2})\,.$$

Hence, $\hat{\xi}^{\mathrm{T}}_{\mathrm{E},\alpha} = \chi^2_{r;\alpha} - n^{-1}\big(F'(\chi^2_{r;\alpha})\big)^{-1}\tilde{R}(\chi^2_{r;\alpha})$. Hence, $\Pr\big(\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) > \hat{\xi}^{\mathrm{T}}_{\mathrm{E},\alpha}\big) = \alpha + O(n^{-2})$, as stated. Similar results hold for the other test statistics.

References

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, New York.

Dembo, A. and Zeitouni, O. (2009). *Large Deviations Techniques and Applications*. Springer, Berlin.

Fristedt, B. and Gray, L. (1997). *A Modern Approach to Probability Theory*. Birkhäuser, Boston.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18:90–120.

Robert, C. (1990). On some accurate bounds for the quantiles of a non-central chi squared distribution. *Statistics & Probability Letters*, 10:101–106.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

Sen, P. K. (1989). The mean-median-mode inequality and noncentral chi square distributions. *Sankhya*, 51:106–114.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Springer-Verlag, New York.