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} |Q(\theta, \hat{W}) - Q(\theta, W_0)| \xrightarrow{L^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:

\[ 0 = \hat{\Psi}(\eta_0) + \left( \frac{\partial \hat{A}(\theta_1)}{\partial \theta_1} \right) \tilde{W}(\theta_1, \theta_2) \frac{O}{B^T} (\hat{\eta} - \eta_0) + o_P(\|\hat{\eta} - \eta_0\|). \]  

(A.1)

\[ \hat{A}(\zeta_1, \zeta_2) = (\partial \hat{C}(\zeta_1, \zeta_2) : 1 \leq j \leq p) \in \mathbb{R}^{p \times p}, \]  

and

\[ \hat{B}(\zeta_1, \zeta_2) = (-\hat{G}(\zeta_1)^T \hat{V}(\zeta_1)^{-1} (\partial \hat{V}(\zeta_1)) \hat{V}(\zeta_1)^{-1} \hat{g}(\zeta) : 1 \leq j \leq p) \in \mathbb{R}^{p \times p}. \]

By (A.1),

\[ n^{1/2}(\hat{\eta} - \eta_0) = -n^{1/2} \left( \begin{array}{c} A_1^{-1} O \\ -A_2^{-1} B A_1^{-1} A_2^{-1} \end{array} \right) \hat{\Psi}(\eta_0) + o_P(1). \]  

(A.2)

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

\[ \hat{\Psi}(\eta_0) = \left( \begin{array}{c} G W_0 (\hat{\theta}_1) - \mu_1 \\ G V_1^{-1} (\hat{\theta}_2) - \mu_2 \end{array} \right) \]  

\[ + \left( \begin{array}{c} (W_0 \mu_1^T \otimes I_p) \text{vec}(\tilde{G}(\hat{\theta}_1)^T - G^T_1) \\ (V_1^{-1} \mu_2^T \otimes I_p) \text{vec}(G(\hat{\theta}_2)^T - G^T_2) \end{array} \right) \]  

\[ + \left( \begin{array}{c} \mu_1^T \otimes G^T_1 \text{vec}(\tilde{W} - W_0) \\ -((V_1^{-1} \mu_2^T \otimes G^T_2 V_1^{-1}) \text{vec}(\tilde{V}_1) - V_1) \end{array} \right) + o_P(n^{-1/2}). \]  

(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 $n^{1/2} \Phi(\theta_0) \xrightarrow{d} N(0, \Sigma)$, and by (A.2), then (8a) holds, as stated. By the delta method, $n^{1/2}(h(\hat{\theta}_1) - h(\theta_1)) \xrightarrow{d} N(0, H_0 F_{22} H_0^T)$, and thus, (7a) holds.

The constrained estimator $\hat{\theta}_1$ minimizes $\tilde{Q}(\theta, \hat{V}(\hat{\theta}_0)^{-1})$ subject to $h(\theta) = 0$. Let $(1/2) \tilde{Q}(\theta, \hat{V}(\hat{\theta}_0)^{-1}) + \lambda^2 h(\theta)$ be the Lagrangian function. Taking the partial derivatives of the Lagrangian and solving for $\theta$ and $\lambda$ yields:

\[
\begin{cases}
\hat{C}(\hat{\theta}, \hat{V}(\hat{\theta}_0)^{-1}) + H(\hat{\theta}_1)\lambda = 0, \\
h(\hat{\theta}_1) = 0.
\end{cases}
\]  

(A.4)

Since $H_0 : \theta_0 \in \Theta_0$ is true, then $\hat{\theta}_1 = \theta_2 + O_p(n^{-1/2})$, and by (A.4), we have $\lambda = O_p(n^{-1/2})$. Taylor expansions of (A.4) about $(\hat{\theta}_1, 0) \in \mathbb{R}^p \times \mathbb{R}^r$ yields

\[
0 = \begin{pmatrix}
\hat{C}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) \\
h(\hat{\theta}_1)
\end{pmatrix} = A_\theta \left( \begin{pmatrix} H_2^T \lambda \end{pmatrix} + O_p(n^{-1/2}) \right)
\]

Hence

\[
n^{1/2}(\hat{\theta}_1 - \theta_0) = -n^{1/2}(A_\theta^2 \lambda_2 - \lambda_2) C(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) + o_p(1).
\]  

(A.5)

Similarly to (A.2),

\[
n^{1/2} \hat{C}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) = -n^{1/2} A_\lambda (\hat{\theta}_1 - \theta_2) + o_p(1),
\]  

(A.6)

and by (A.5), we obtain

\[
n^{1/2}(\hat{\theta}_1 - \theta_1) = n^{1/2} A_\lambda C(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) + o_p(1).
\]  

(A.7)

By (A.6), (A.7), and (8a), $n^{1/2}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, \Omega_2)$. One-term Taylor expansion of $\tilde{Q}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1})$ about $\hat{\theta}_1$ yields

\[
\tilde{L}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) = 2n^{1/2} \lambda \hat{\theta}_1 (\hat{\theta}_1 - \hat{\theta}_1)^T (\hat{\theta}_1 - \hat{\theta}_1)
\]

\[
+ n(\hat{\theta}_1 - \hat{\theta}_1)^T \hat{\Lambda} (\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1})(\hat{\theta}_1 - \hat{\theta}_1) + o_p(1) = n(\hat{\theta}_1 - \hat{\theta}_1)^T A_\lambda (\hat{\theta}_1 - \hat{\theta}_1) + o_p(1),
\]

and thus, (7b) holds. Taylor expansion of $\hat{C}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1})$ about $\hat{\theta}_1$ yields

\[
n^{1/2} \hat{C}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) = n^{1/2} \hat{C}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1})
\]

\[
+ n^{1/2} A_\lambda (\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1})(\hat{\theta}_1 - \hat{\theta}_1) + o_p(1) = n^{1/2} A_\lambda (\hat{\theta}_1 - \hat{\theta}_1) + o_p(1),
\]

and thus, (7c) holds. Similarly, $n^{-1} \tilde{L}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) = 2n^{-1/2} \lambda_2 \hat{\theta}_1 (\hat{\theta}_1 - \hat{\theta}_1)^T (\hat{\theta}_1 - \hat{\theta}_1) + o_p(1)$. Hence (8b) holds, where $\sigma^2_{\Delta} = 4 \mu_2 V_1^{-1} G_2 G_2^T V_1^{-1} \mu_2$. Note further that

\[
n^{-1} \tilde{L}(\hat{\theta}_1, \hat{V}(\hat{\theta}_0)^{-1}) = \Psi_\theta \left( \hat{g}(\hat{\theta}_1), \text{vec}(\hat{V}(\hat{\theta}_1)), \text{vec}(\hat{G}(\hat{\theta}_1)), \text{vec}(\hat{G}(\hat{\theta}_1)) \right),
\]  

(A.8)

where $\Psi_\theta : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^{n \times p} \times \mathbb{R}^{n \times p} \mapsto \mathbb{R}$ is given by:

\[
\Psi_\theta (\hat{g}(\hat{\theta}_1), \text{vec}(\hat{V}(\hat{\theta}_1)), \text{vec}(\hat{G}(\hat{\theta}_1)), \text{vec}(\hat{G}(\hat{\theta}_1))) = \left( g^T V^{-1} G_1 (G_2^T V^{-1} G_2)^{-1} G_1^T V^{-1} g \right),
\]  

(A.9)

where $V = \text{vec}(\theta_0 \in \mathbb{R}^{n \times p}, G_1 = \text{vec}(\theta_1 \in \mathbb{R}^{n \times p}, G_2 = \text{vec}(\theta_2 \in \mathbb{R}^{n \times p},$ and $\text{vec}(\theta \in \mathbb{R}^{n \times p} \mapsto \mathbb{R}^{n \times p}$ is the inverse of the vec operator. Since

\[
n^{1/2} \left( \begin{array}{c}
\hat{g}(\hat{\theta}_1) - \mu_2 \\
\text{vec}(\hat{V}(\hat{\theta}_1) - V_2) \\
\text{vec}(\hat{G}(\hat{\theta}_1) - G_2)
\end{array} \right) = n^{1/2} \left( \begin{array}{c}
\hat{g}(\hat{\theta}_1) - \hat{g}(\theta_2) \\
\text{vec}(\hat{V}(\hat{\theta}_1) - \hat{V}(\theta_2)) \\
\text{vec}(\hat{G}(\hat{\theta}_1) - \hat{G}(\theta_2))
\end{array} \right) + o_p(1),
\]

and $n^{1/2} \hat{g}(\hat{\theta}_1) = n^{1/2} G_0 (\hat{\theta}_1 - \theta_1) + o_p(1), and

\[
n^{1/2} \text{vec}(\hat{V}(\hat{\theta}_1) - \hat{V}(\theta_2)) = n^{1/2} \text{vec} (VV(\theta_2) ((\hat{\theta}_1 - \theta_1) \otimes L_4)) + o_p(1),
\]

\[
n^{1/2} \text{vec}(\hat{G}(\hat{\theta}_1) - \hat{G}(\theta_2)) = n^{1/2} \text{vec} (VG(\theta_2) ((\hat{\theta}_1 - \theta_1) \otimes L_4)) + o_p(1),
\]

\[
n^{1/2} \text{vec}(\hat{G}(\hat{\theta}_1) - \hat{G}(\theta_2)) = n^{1/2} \text{vec} (VG(\theta_2) ((\hat{\theta}_1 - \theta_1) \otimes L_4)) + o_p(1),
\]

\[
and n^{1/2} \text{vec}(\hat{G}(\hat{\theta}_1) - \hat{G}(\theta_2)) = n^{1/2} G_0 (\hat{\theta}_1 - \theta_1) + o_p(1), and

\[
n^{1/2} \text{vec}(\hat{V}(\hat{\theta}_1) - \hat{V}(\theta_2)) = n^{1/2} \text{vec} (VV(\theta_2) ((\hat{\theta}_1 - \theta_1) \otimes L_4)) + o_p(1),
\]

\[
n^{1/2} \text{vec}(\hat{G}(\hat{\theta}_1) - \hat{G}(\theta_2)) = n^{1/2} \text{vec} (VG(\theta_2) ((\hat{\theta}_1 - \theta_1) \otimes L_4)) + o_p(1),
\]

\[
n^{1/2} \text{vec}(\hat{G}(\hat{\theta}_1) - \hat{G}(\theta_2)) = n^{1/2} \text{vec} (VG(\theta_2) ((\hat{\theta}_1 - \theta_1) \otimes L_4)) + o_p(1),
\]
by (A.2), (A.3), (A.5), and condition E, there exist a matrix $D_2$ such that

$$\begin{pmatrix} \hat{g}(\hat{\theta}_1) - \mu_2 \\ \text{vec}(\hat{V}(\hat{\theta}_1) - V_1) \\ \text{vec}(\hat{g}(\hat{\theta}_1) - G_2) \\ \text{vec}(\hat{g}(\hat{\theta}_1) - G_3) \end{pmatrix} \xrightarrow{n^{1/2}} N(0, D_2 \Sigma D_2^T).$$

By the delta method, then (8c) holds, where $\psi_2 = \mu_2 V_1^{-1} G_2 (G_2^T V_2^{-1} G_2)^{-1} G_2^T V_2^{-1} \mu_2$, $\sigma_2^2 = \nabla \psi_2 D_2 \Sigma D_2^T \nabla \psi_2^T$, and $\nabla \psi_2 = \nabla \psi_2(\mu_2, \text{vec}(V_2), \text{vec}(G_2))$. Since

$$n^{1/2}(\hat{g}(\hat{\theta}_1) - \mu_2) = n^{1/2}(\hat{g}(\hat{\theta}_1) - \bar{g}(\theta_2)) + n^{1/2}(\bar{g}(\theta_2) - \mu_2) + o_p(1),$$

by (A.2) and condition E, there exists a matrix $D_2$ such that

$$n^{1/2}(\hat{g}(\hat{\theta}_1) - \mu_2) \xrightarrow{\text{d}} N(0, D_2 \Sigma D_2^T).$$

Hence (8d) holds, where $\lambda_1^2, \ldots, \lambda_q^2$ are the eigenvalues of $(D_2 \Sigma D_2^T)^{1/2} V_0^{-1} (D_2 \Sigma D_2^T)^{1/2}$.

Finally, write $n^{-1/2}(\hat{V}(\hat{\theta}_1) - V_0) = \psi_2(\hat{g}(\hat{\theta}_1), \text{vec}(\hat{V}(\hat{\theta}_1)))$, where $\psi_2 : \mathbb{R}^q \times \mathbb{R}^q \rightarrow \mathbb{R}$ is given by $\psi_2(g, v) = g^T V^{-1} g$ and $V = \text{vec}_{q \times q}(\nu) \in \mathbb{R}^{q^2 \times q^2}$. Since

$$n^{1/2}(\hat{g}(\hat{\theta}_1) - \mu_2) = n^{1/2}(\hat{g}(\hat{\theta}_1) - \bar{g}(\theta_2)) + n^{1/2}(\bar{g}(\theta_2) - \mu_2) + o_p(1),$$

and $n^{1/2}(\hat{V}(\hat{\theta}_1) - \bar{V}(\theta_2)) = n^{1/2} G_2(\hat{\theta}_1 - \theta_2) + o_p(1)$, and

$$n^{1/2} \text{vec}(\hat{V}(\hat{\theta}_1) - \bar{V}(\theta_2)) = n^{1/2} \text{vec}(\text{vec}(\hat{V}(\hat{\theta}_1) - \bar{V}(\theta_2))) + o_p(1),$$

by (A.2) and condition E, there exists a matrix $E_2$ such that

$$n^{1/2} \left( \begin{array}{c} \hat{g}(\hat{\theta}_1) - \mu_2 \\ \text{vec}(\hat{V}(\hat{\theta}_1) - V_1) \end{array} \right) \xrightarrow{\text{d}} N(0, E_2 \Sigma E_2^T).$$

Thus, (8c) holds, where $\sigma_2^2 = \nabla \psi_2 E_2 \Sigma E_2^T \nabla \psi_2^T$ and $\nabla \psi_2 = \nabla \psi_2(\mu_2, \text{vec}(V_1))$.

Remark A.1 If the model (1) is correctly specified, then $\hat{\theta}_1 = \theta$, $\mu_1 = \mu_2 = 0$, $A_1 = D(\theta_1, W_1)$, $B = 0, A_2 = D_0$, and $G_1 = G_2 = G_0$. By (6), $\Pi_{12} = G_0^T W_0$, $\Pi_{21} = 0$, $\Pi_{13} = 0$, $\Pi_{24} = 0$, $\Pi_{25} = G_0^T V_0^{-1}$, and $\Pi_{26} = 0$. Hence $\Xi_{11} = G_0^T W_0 W_0 G_0$, $\Xi_{12} = G_0^T W_0$, and $\Xi_{22} = D_0$. Furthermore, $\Gamma_{11} = D(\theta_1, W_1)^{-1} G_0^T W_0 W_0 G_0 D(\theta_1, W_1)^{-1}$, $\Gamma_{12} = \Gamma_{21} = \Gamma_{22} = D_0^{-1}$, and thus, (2a) holds. Note further that $\Omega_2 = D_0^{-1} H_2 (H_0 D_0^{-1} H_2^T)^{-1} H_0 D_0^{-1}$. Since $\Omega_2^{1/2} \Sigma \Omega_2^{-1/2}$ is idempotent of rank $r$, by (7a), (7b), and (7c), (2b) holds. By (A.10), then $V_0^{-1/2} D_1 \Sigma D_2^T V_0^{-1/2} = A_2$, where $A_2 = I_q - V_0^{-1/2} G_0 D_0^{-1} G_0^T V_0^{-1/2}$ is idempotent of rank $q - p$. Hence (4c) holds.

Proof (Lemma 1) Assume that $H_0 : \theta_1 \in \Theta_0$ is true. Note first that

$$\Pr(\hat{V}(\hat{\theta}_1)^{-1} \leq x) = \Pr(\hat{V}(\hat{\theta}_1)^{-1} \leq x) = -n^{-1} \frac{d}{dx} \left( \text{E}(\hat{V}(\hat{\theta}_1)^{-1}) \hat{V}(\hat{\theta}_1)^{-1} \leq x) \right)$$

$$= -n^{-1} \frac{d}{dx} \left( \text{E}(\hat{V}(\hat{\theta}_1)^{-1}) \hat{V}(\hat{\theta}_1)^{-1} \hat{V}(\hat{\theta}_1)^{-1} \leq x) \right)$$

$$- n^{-3/2} \frac{d}{dx} \left( \text{E}(\hat{V}(\hat{\theta}_1)^{-1}) \hat{V}(\hat{\theta}_1)^{-1} \hat{V}(\hat{\theta}_1)^{-1} \leq x) \right)$$

$$+ (1/2) n^{-1} \frac{d^2}{dx^2} \text{E}(\hat{V}(\hat{\theta}_1)^{-1}) \hat{V}(\hat{\theta}_1)^{-1} \hat{V}(\hat{\theta}_1)^{-1} \leq x)$$

$$+ n^{-3/2} \frac{d^2}{dx^2} \text{E}(\hat{V}(\hat{\theta}_1)^{-1}) \hat{V}(\hat{\theta}_1)^{-1} \hat{V}(\hat{\theta}_1)^{-1} \leq x)$$

$$- (1/6) n^{-3/2} \frac{d^3}{dx^3} \text{E}(\hat{V}(\hat{\theta}_1)^{-1}) \hat{V}(\hat{\theta}_1)^{-1} \hat{V}(\hat{\theta}_1)^{-1} \leq x) + O(n^{-1}).$$
where \( I(E) \) is the indicator function of an event \( E \). Here and in other places, we use

\[
\Pr(R + n^{-1/2}S + n^{-1}T + n^{-3/2}U \leq x) = \Pr(R \leq x) + \int_0^x (-n^{-1/2}s - n^{-1}t - n^{-3/2}u) \, dF(s,t,u)
\]

\[
= \Pr(R \leq x) + (1/2) \int (-n^{-1/2}s - n^{-1}t - n^{-3/2}u)^2 \, dF(s,t,u)
\]

\[
+ (1/6) \int (-n^{-1/2}s - n^{-1}t - n^{-3/2}u)^3 \, dF(s,t,u) + O(\alpha^{-2})
\]

\[
= \Pr(R \leq x) - n^{-1} \frac{d}{dx} \left( E(SI(R \leq x)) - n^{-1} \frac{d}{dx} E(TI(R \leq x)) \right)
\]

\[
= \Pr(R \leq x) - n^{-1} \frac{d}{dx} E(UI(R \leq x)) + (1/2) n^{-1} \frac{d^2}{dx^2} E(S^2I(R \leq x))
\]

\[
+ n^{-3/2} \frac{d^2}{dx^2} E(STI(R \leq x)) - (1/6) n^{-3/2} \frac{d^3}{dx^3} E(S^3I(R \leq x)) + O(\alpha^{-2}),
\]

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

\[
\int I(z^2 \alpha \leq x) p(z) \phi(z) \, dz = 0,
\]

for every odd polynomial \( p(z) \) and \( A \in \mathbb{R}^{4 \times 4} \). Hence

\[
\Pr(T(\tilde{V}(\tilde{R})^{-1}) \leq x) = \int I(z^2 \gamma \leq x) (1 + n^{-1} p(z)) \phi(z) \, dz
\]

\[
- n^{-1} \frac{d}{dx} \int (z^2 \gamma \leq x) p(z) \phi(z) \, dz
\]

\[
- n^{-1} \frac{d}{dx} \int (z^2 \gamma \leq x) (z \otimes k) p(z) \phi(z) \, dz
\]

\[
+ (1/2) n^{-1} \frac{d^2}{dx^2} \int (z^2 \gamma \leq x) (z \otimes k) (z \otimes k) p(z) \phi(z) \, dz
\]

where \( F(x,r) \) denotes the distribution function of \( \chi^2_{\alpha} \). Hence,

\[
\Pr(T(\tilde{V}(\tilde{R})^{-1}) > \chi^2_{\alpha}) = 1 - F(\chi^2_{\alpha}; r) - n^{-1} R(\chi^2_{\alpha}) + O(\alpha^{-2})
\]

\[
= \alpha + n^{-1} a(\alpha) + O(\alpha^{-2}),
\]

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

**B Proofs of Section 4**

Let \( \tilde{Y}(\xi) \) be the bootstrap version of \( \tilde{Y}(\xi) \) given by \( \tilde{Y}(\xi) = (\tilde{C}(\xi, \tilde{W}^*)^T, \tilde{C}(\tilde{Y}(\xi), \tilde{V}(\xi)^{-1})^T)^T \). In the proofs of this section, we use the following notation. Let \( \tilde{T}^* \in \mathbb{R}^d \) denote a bootstrap quantity. We say that \( \tilde{T}^* \) converges to 0 conditionally in probability, and write \( \tilde{T}^* = \mathcal{O}_p(1) \), if \( \lim_{n \to \infty} \Pr(\| \tilde{T}^* \| \geq \varepsilon | X_n) = 0 \) for all \( \varepsilon > 0 \).

By definition, \( \tilde{T}^* \) converges (unconditionally) in probability to 0 if \( \Pr(\| \tilde{T}^* \| \geq \varepsilon) \to 0 \) 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 \( \tilde{T}^* \) is conditionally bounded in probability, and write \( \tilde{T}^* = \mathcal{O}_p(1) \), if for all \( \varepsilon > 0 \) there exists \( M = M(\varepsilon) \) such that \( \Pr(\| \tilde{T}^* \| \geq M | X_n) \leq \varepsilon + o(1) \leq 1 \).

By definition, \( \tilde{T}^* \) is bounded in probability if for all \( \varepsilon > 0 \) there exists \( M = M(\varepsilon) \) such that \( \Pr(\| \tilde{T}^* \| \geq M) \leq \varepsilon \). 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) \( \sup_{\theta \in \Theta}(\hat{\theta}\hat{^*}(\theta) - \theta, \hat{W}^*) \to 0, \)
(ii) \( \hat{\theta}^* \to \theta_0, \)
and (iii) \( \text{Pr}(\hat{\theta}^* = \theta_0) = 0(\delta), \)
\( \theta \to \theta_0. \)

Proof (Theorem 8) Similarly to (A.2),
\[
n^{1/2}(\hat{\theta}^* - \theta^*) = -n^{1/2}\left(-A_2^{-1}BA_1^{-1} O\right) \hat{\theta} + o_P(1).
\]

Similarly to (A.3),
\[
\hat{\theta}^* = \left(\begin{array}{c}
G_2W = (W_1 + W_2)
\end{array}\right) \hat{\theta}^* + \left(\begin{array}{c}
\mu_1 \hat{\theta}^* + \mu_2 \hat{\lambda}^*
\end{array}\right) + \text{vec}(\hat{\theta} - \hat{\theta}_1) + o_P(n^{-1/2})
\]
\[
= \Pi(\hat{\theta}^* - \hat{\theta}) + a_P(n^{-1/2}).
\]

By (31b) and (B.1), then (22a) holds. By the delta method for the bootstrap and (22a),
\[
\mathbb{L}_Q(n^{1/2}h^*(\hat{\theta}^*_1), X|n) \to N(0, H_2F_2h^*_2).
\]
and thus, (22b) holds, as stated. The SB estimator \( \hat{\theta}^*_1 \) minimizes \( \hat{\theta}^* \) subject to \( h^*(\hat{\theta}) = 0. \)

Let \( (1/2)\hat{\theta}^* (\hat{\theta}^* - \hat{\theta}_1) \) be the SB version of the Lagrangian and let \( \lambda^* \) be the SB version of \( \lambda. \) Taking the partial derivatives of the SB Lagrangian and solving for \( \theta \) and \( \lambda \) yields
\[
\begin{cases}
C^*(\hat{\theta}_1, \hat{\theta}^*_1) + H^*(\hat{\theta}_1) \lambda^* = 0,
\end{cases}
\]

Since \( \hat{\theta}^*_1 = \hat{\theta}_1 + o_P(n^{-1/2}), \) then \( \hat{\lambda}^* = o_P(n^{-1/2}). \) Taylor expansion of (B.2) about \( (\hat{\theta}_1, 0) \in \mathbb{R}^p \times \mathbb{R} \)
yields
\[
0 = \begin{pmatrix}
C^*(\hat{\theta}_1, \hat{\theta}^*_1) & H^*(\hat{\theta}_1)
\end{pmatrix}
\begin{pmatrix}
\hat{\theta}_1 - \hat{\theta}
\lambda^*
\end{pmatrix} + o_P(1).
\]

Hence
\[
n^{1/2}(\hat{\theta}^*_1 - \hat{\theta}_1) = -n^{1/2}(A_2^{-1} - A_2)C^*(\hat{\theta}_1, \hat{\theta}^*_1) + o_P(1).
\]

Similarly to (B.1),
\[
n^{1/2}C^*(\hat{\theta}_1, \hat{\theta}^*_1) = -n^{1/2}A_2(\hat{\theta}_1 - \hat{\theta}) + o_P(1),
\]

and by (B.3),
\[
n^{1/2}(\hat{\theta}^*_1 - \hat{\theta}_1) = n^{1/2}A_2C^*(\hat{\theta}_1, \hat{\theta}^*_1) + o_P(1).
\]

By (B.4), (B.5), and (22a), \( \mathbb{L}_Q(n^{1/2}h^*(\hat{\theta}^*_1), X|n) \to N(0, \Omega_2). \) Taylor expansion of \( \hat{\theta}^* (\hat{\theta}^*_1, \hat{\theta}^*_1) \) about \( \hat{\theta}^*_1 \) yields
\[
L^* (\hat{\theta}^*_1, \hat{\theta}^*_1) = n(\hat{\theta}^*_1 - \hat{\theta})J^* A_2(\hat{\theta}^*_1 - \hat{\theta}) + o_P(1),
\]

and thus, (22c) holds. Similarly,
\[
n^{-1/2}L^* (\hat{\theta}^*_1, \hat{\theta}^*_1) = 2n^{-1/2}C^*(\hat{\theta}^*_1, \hat{\theta}^*_1)J^* (\hat{\theta}^*_1 - \hat{\theta}) + o_P(1),
\]

and thus, (22d) holds. Taylor expansion of \( C^*(\hat{\theta}^*_1, \hat{\theta}^*_1) \) about \( \hat{\theta}^*_1 \) yields
\[
n^{1/2}C^*(\hat{\theta}^*_1, \hat{\theta}^*_1) = n^{1/2}A_2(\hat{\theta}^*_1 - \hat{\theta}) + o_P(1),
\]

and thus, (22e) holds. By the delta method for the bootstrap, (22f) holds. Taylor expansion of \( \hat{\theta}^* (\hat{\theta}^*_1) \) about \( \hat{\theta}^*_1 \), similarly to (A.10),
\[
\mathbb{L}_Q(n^{1/2}(\hat{\theta}^*_1) - \hat{\theta}_1) \to N(0, \Omega D_2^2),
\]
and thus, (22g) holds. Finally, write \( n^{-1}f^*(\hat{\theta}^*_1) = \psi_J(\hat{\theta}^*_1, \text{vec}(\hat{\theta}^*_1)) \). By the delta theorem for the bootstrap, then (22h) holds, as stated.
Proof (Theorem 9) By Lemma B.1, \( \tilde{\theta}_2 = \theta_2 + o_p(1) \) and \( \tilde{\theta}_4 = \theta_4 + o_p(1) \). Hence \( \bar{A}^* (\tilde{\theta}_1, \tilde{W}^*) \overset{p}{\to} A_3 \), \( \bar{B}^* (\tilde{\theta}_1, \tilde{\theta}_4) = o_p(1) \), and \( \bar{A}^* (\tilde{\theta}_1, \bar{V}^* (\bar{\theta}_4)) \overset{p}{\to} A_4. \) Similarly to (B.1),
\[
n^{1/2} (\tilde{\eta}_2 - \eta_2) = -n^{1/2} \left( A_3^{-1} O A_4^{-1} \right) \Psi^* (\tilde{\eta}_4) + o_p(1).
\]
(B.8)
Since
\[
\mathcal{L}_C (n^{1/2} \Psi^* (\bar{\eta}_4))_{|X_{1,a}} \overset{p}{\to} N \left( \begin{pmatrix} G^T W_0 V_0 W_0 G_2 A_3 \end{pmatrix} \right),
\]
then (24a) holds. By the delta method for the bootstrap, then
\[
\mathcal{L}_C (n^{1/2} h^* (\bar{\theta}_4))_{|X_{1,a}} \overset{p}{\to} N (0, H_2 A_4^{-1} H_2^T),
\]
and since \( (H_2 A_4^{-1} H_2^T)^{1/2} H_2 A_4^{-1} H_2^T (H_2 A_4^{-1} H_2^T)^{1/2} \) is idempotent of rank \( r \), then (24b) holds. Similarly to (B.5),
\[
n^{1/2} (\bar{\theta}_2 - \tilde{\theta}_2) = n^{1/2} A_4 C^* (\bar{\theta}_2, \tilde{V}^* (\tilde{\theta}_2))^{-1} + o_p(1),
\]
where \( A_4 = A_4^{-1} H_2^T (H_2 A_4^{-1} H_2^T)^{-1} H_2 A_4^{-1}. \) Hence \( \mathcal{L}_C (n^{1/2} (\bar{\theta}_2 - \tilde{\theta}_2))_{|X_{1,a}} \overset{p}{\to} N (0, A_4). \) Since the matrix \( A_4^{-1/2} A_4^{-1/2} \) is idempotent of rank \( r \), (24c) holds. Since \( g^* (\bar{\theta}_2) = o_p(1) \), then
\[
C^* (\bar{\theta}_2, \tilde{V}^* (\tilde{\theta}_2))^{-1} = C^* (\bar{\theta}_2, \tilde{V}^* (\tilde{\theta}_2))^{-1} = \tilde{V}^* (\tilde{\theta}_4) = o_p(n^{-1/2}).
\]
(B.9)
Hence \( \bar{L} (\tilde{V}^* (\tilde{\theta}_4)^{-1}) = \bar{L} (\tilde{V}^* (\tilde{\theta}_4)^{-1}) + o_p(1) \), and thus, (24d) holds. Similarly to (B.7) and by (B.9), then (24e) and (24f) hold. By (B.8), Taylor expansion of \( g^* (\bar{\theta}_4) \) about \( \tilde{\theta}_4 \) yields
\[
n^{1/2} g^* (\bar{\theta}_4) = n^{1/2} (I_q - G_2 A_4^{-1} G_2^T V_0^{-1}) g^* (\tilde{\theta}_4) + o_p(1).
\]
(B.10)
Since \( \mathcal{L}_C (g^* (\bar{\theta}_4))_{|X_{1,a}} \overset{p}{\to} N (0, V_1) \) and \( A_4 = I_q - V_0^{-1/2} G_2 A_4^{-1} G_2^T V_0^{-1/2} \) 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 \( g^* (\bar{\theta}_4) \) about \( \tilde{\theta}_4 \) and (9a), then
\[
n^{1/2} g^* (\bar{\theta}_4) = n^{1/2} (I_q - G_0 D_0^{-1} G_0^T V_0^{-1}) g^* (\tilde{\theta}_4) + o_p(1).
\]
(C.1)
Hence
\[
J^* (\tilde{V}^* (\tilde{\theta}_4)^{-1}) = n g^* (\tilde{\theta}_4) T V_0^{-1/2} A_4 V_0^{-1/2} g^* (\tilde{\theta}_4) + o_p(1).
\]
(C.2)
where recall that \( A_2 = I_q - V_0^{-1/2} G_0 D_0^{-1} G_0^T V_0^{-1/2} \). Since
\[
\mathcal{L}_S (n^{1/2} V_0^{-1/2} (\tilde{g}^* (\tilde{\theta}_4) - \tilde{g} (\tilde{\theta}_4)))_{|X_{1,a}} \overset{p}{\to} N (0, I_q)
\]
and \( n^{1/2} V_0^{-1/2} \tilde{g} (\tilde{\theta}_4) \overset{d}{\rightarrow} Z \), where \( Z \sim N (0, A_2) \), by Slutsky’s theorem in the product metric space \( \mathcal{P}^q \times \mathcal{R}^q \) (van der Vaart 1998, Theorem 18.10(v)),
\[
\mathcal{L}_S (n^{1/2} V_0^{-1/2} (\tilde{g}^* (\tilde{\theta}_4) - \tilde{g} (\tilde{\theta}_4)))_{|X_{1,a}}, n^{1/2} V_0^{-1/2} \tilde{g} (\tilde{\theta}_4) \overset{d}{\rightarrow} (N (0, I_q), Z).
\]
Let the mapping \( \varphi : \mathcal{P}^q \times \mathcal{R}^q \rightarrow \mathcal{P}^q \) be given by \( \varphi (P, x) = \frac{\tilde{F}}{x} (X + x)^T A_2 (X + x) \), where \( X \sim P, P \in \mathcal{P}^q \),
and \( x \in \mathcal{R}^q \). By Corollary 13 of Fristedt and Gray (1997, p. 414) and (C.2),
\[
\mathcal{L}_S (J^* (\tilde{V}^* (\tilde{\theta}_4)^{-1}))_{|X_{1,a}} = \varphi (\mathcal{L}_S (n^{1/2} V_0^{-1/2} (\tilde{g}^* (\tilde{\theta}_4) - \tilde{g} (\tilde{\theta}_4)))_{|X_{1,a}}, n^{1/2} V_0^{-1/2} \tilde{g} (\tilde{\theta}_4)) + o_p(1).
\]
Since \( \varphi \) is continuous, by the continuous mapping theorem,
\[
\varphi (\mathcal{L}_S (n^{1/2} V_0^{-1/2} (\tilde{g}^* (\tilde{\theta}_4) - \tilde{g} (\tilde{\theta}_4)))_{|X_{1,a}}), n^{1/2} V_0^{-1/2} \tilde{g} (\tilde{\theta}_4) \overset{d}{\rightarrow} \varphi (N (0, I_q), Z).
\]
Since \( \varphi (N (0, I_q), x) = \chi^2_{q-p} (x^T A_2 x) \) and \( Z^T A_2 Z \sim \chi^2_{q-p} \), then (9e) holds, as stated.
Proof (Corollary 1) By Theorem 2 and continuous mapping theorem, then
\[ f(\hat{\theta}_1^{(1)} - \xi_{\theta_0, \alpha}^2) \sim U^2 - \chi^2_{2, \alpha}(U^2), \]
where \( U^2 \sim \chi^2_{2, \alpha} \) and \( \chi^2_{2, \alpha}(u^2) \) is the upper \( \alpha \)-quantile of the non-central chi-squared distribution \( \chi^2_{2, \alpha}(u^2) \) with non-centrality parameter \( \alpha^2 \). Hence, \[ \Pr(f(\hat{\theta}_1^{(1)} - \xi_{\theta_0, \alpha}^2) > \xi_{\theta_0, \alpha}^2) \rightarrow \Pr(U - \chi^2_{2, \alpha}(U) > 0). \] (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)), \( U^2 - \chi^2_{2, \alpha}(U^2) \leq 0 \) with probability one for \( \alpha \leq 1/2 \). This completes the proof of the corollary.

Proof (Theorem 3) Note first that
\[ \Pr(\hat{\chi}^2(\hat{\theta}_1^{(1)})^{-1}) \leq x|X_{1,n}) = \Pr(\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x|X_{1,n}) \]
\[ - n^{-1/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz \right) \]
\[ - n^{-1} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz \right) \]
\[ - n^{-3/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz \right) \]
\[ + (1/2) n^{-1} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz \right) \]
\[ + n^{-3/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz \right) \]
\[ - (1/6) n^{-3/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz + O_P(n^{-2}). \right) \]
We further obtain
\[ \Pr(\hat{\chi}^2(\hat{\theta}_1^{(1)})^{-1}) \leq x|X_{1,n}) = \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) (1 + n^{-1/2} \hat{p}_1(z) + n^{-1} \hat{p}_2(z) + n^{-3/2} \hat{p}_3(z)) \phi(z) dz \]
\[ - n^{-1/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) (1 + n^{-1/2} \hat{p}_1(z) + n^{-1} \hat{p}_2(z) + n^{-3/2} \hat{p}_3(z)) \phi(z) \right) \]
\[ - n^{-1} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) (1 + n^{-1/2} \hat{p}_1(z) + n^{-1} \hat{p}_2(z) + n^{-3/2} \hat{p}_3(z)) \phi(z) \right) \]
\[ - n^{-3/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) (1 + n^{-1/2} \hat{p}_1(z) + n^{-1} \hat{p}_2(z) + n^{-3/2} \hat{p}_3(z)) \phi(z) \right) \]
\[ + (1/2) n^{-1} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) (1 + n^{-1/2} \hat{p}_1(z) + n^{-1} \hat{p}_2(z) + n^{-3/2} \hat{p}_3(z)) \phi(z) \right) \]
\[ + n^{-3/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) (1 + n^{-1/2} \hat{p}_1(z) + n^{-1} \hat{p}_2(z) + n^{-3/2} \hat{p}_3(z)) \phi(z) \right) \]
\[ - (1/6) n^{-3/2} \frac{d}{dx} \frac{d}{dx} \left( \int (\hat{Z}^T \tilde{\nu} \hat{Z}^* \leq x) \phi(z) dz + O_P(n^{-2}). \right) \]
Hence
\[ \Pr(\hat{\chi}^2(\hat{\theta}_1^{(1)})^{-1}) \leq x|X_{1,n}) = F(x; r) + n^{-1} \tilde{R}(x) + O_P(n^{-2}). \]
where \( \tilde{R}(x) = R(x) + O_P(n^{-1/2}) \). Let \( \tilde{\xi}_{\alpha, \theta_0} = \chi^2_{\alpha, \theta_0} + n^{-1} \alpha_1(z) + O_P(n^{-2}) \) be an empirical Cornish-Fisher expansion of the upper \( \alpha \)-quantile of \( \chi^2_{\alpha, \theta_0}(\hat{\chi}^2(\hat{\theta}_1^{(1)})) \). Taylor expansion of \( \Pr(\hat{\chi}^2(\hat{\theta}_1^{(1)})) \leq \tilde{\xi}_{\alpha, \theta_0}|X_{1,n}) \) about \( \chi^2_{\alpha, \theta_0} \) yields
\[ 1 - \alpha = \Pr(\hat{\chi}^2(\hat{\theta}_1^{(1)})) \leq \tilde{\xi}_{\alpha, \theta_0}|X_{1,n}) = 1 - \alpha + n^{-1} \tilde{F}(\tilde{\xi}_{\alpha, \theta_0}, \tilde{\nu}; \theta_0) + n^{-1} \tilde{R}(\tilde{\xi}_{\alpha, \theta_0}) + O_P(n^{-2}). \]
Hence, \( \tilde{\xi}_{\alpha, \theta_0} = \chi^2_{\alpha, \theta_0} - n^{-1} \tilde{F}(\tilde{\xi}_{\alpha, \theta_0})^{-1} \tilde{R}(\tilde{\xi}_{\alpha, \theta_0}) \). Thus, \( \Pr(\hat{\chi}^2(\hat{\theta}_1^{(1)})) > \tilde{\xi}_{\alpha, \theta_0} = \alpha + O(n^{-2}) \), 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:

\[
J_n^*(\hat{\theta}_n)^{-1} = Z^T (\hat{B}_0 + n^{-1/2} \hat{B}_1 (Z^* \otimes I_k) + n^{-1} \hat{B}_2 (Z^* \otimes I_k)) (Z^* \otimes I_k) + n^{-1/2} \hat{B}_1 (Z^* \otimes I_k) (Z^* \otimes I_k) (Z^* \otimes I_k) (Z^* \otimes I_k) O(n^{-2}).
\]

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

\[
\inf_{|\theta-\theta_0| \leq \varepsilon} \inf_{|\rho| \leq 1} \text{Pr}(\rho^T g(X, \theta) < 0) > 0.
\]

Proof (Lemma C.1) Let \( S^{p-1} = \{|\rho| = 1\} \) denote the unit sphere in \( \mathbb{R}^p \). Suppose, to the contrary, that the conclusion of the lemma is false. Then, there exist \( \theta_k \in \mathbb{R}^p \) and \( \rho_m \in S^{p-1} \) such that \( \theta_k \to \theta_0 \) and \( \text{Pr}(\rho_m^T g(X, \theta_k) < 0) < n^{-1} \). Since \( S^{p-1} \) is compact, there exists \( \rho_0 \in S^{p-1} \) and a subsequence \( m_n \) such that \( \rho_{m_n} \to \rho_0 \). Since \( g(x, \theta) \) is continuous at \( \theta_0 \), then \( \rho_{m_n}^T g(x, \theta_{m_n}) \to \rho_0^T g(x, \theta_0) \) for \( \Theta_0 \)-almost all \( x \in \mathcal{X} \).

Hence, if \( \rho_{m_n}^T g(x, \theta_{m_n}) < 0 \) then \( \rho_{m_n}^T g(x, \theta_{m_n}) < 0 \) for \( n \) sufficiently large. Thus, for \( \Theta_0 \)-almost all \( x \in \mathcal{X} \),

\[
I(\rho_{m_n}^T g(x, \theta_{m_n}) < 0) \leq \liminf_n I(\rho_{m_n}^T g(x, \theta_{m_n}) < 0).
\]

By Fatou’s lemma,

\[
\text{Pr}(\rho_0^T g(X, \theta_k) < 0) = \text{E}(I(\rho_0^T g(X, \theta_k) < 0)) \leq \text{E}(\liminf_n I(\rho_{m_n}^T g(X, \theta_{m_n}) < 0)) \leq \liminf_n I(\rho_{m_n}^T g(X, \theta_{m_n}) < 0) \leq \liminf_n m_n^{-1} = 0.
\]

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

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

\[
\text{Pr}\left\{0 \in \text{int}\left(\text{conv}\{g(X, \theta_i) : 1 \leq i \leq n\}\right)\right\} \to 1,
\]

where \( \text{int}(A) \) denotes the interior of a set \( A \) and \( \text{conv}\{g(X, \theta_i) : 1 \leq i \leq n\} \) denotes the convex hull of \( \{g(X, \theta_i) : 1 \leq i \leq n\} \).

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

\[
A_n = \left\{ 0 \notin \bigcap_{\theta \in \bar{B}(\theta_0, \varepsilon)} \text{conv}\{g(X, \theta) : 1 \leq i \leq n\} \right\}.
\]

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

\[
A_n = \left\{ \sup_{\theta \in \bar{B}(\theta_0, \varepsilon)} \sup_{|\rho| = 1} n^{-1} \sum_{i=1}^n (\rho^T g(X_i, \theta)) = 1 \right\}.
\]

Since

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

by Lemma C.1, then

\[
\sup_{\theta \in \bar{B}(\theta_0, \varepsilon)} \sup_{|\rho| = 1} \text{Pr}(\rho^T g(X, \theta) \geq 0) < 1.
\]
Thus \( \Pr(p^T g(X, \theta) > 0) < 1 \) for all \( \theta \in B(\theta_0, \epsilon) \) and \( p \in \mathbb{S}^{r-1} \). For \( (p, \theta) \in \mathbb{S}^{r-1} \times B(\theta_0, \epsilon) \), let

\[
\gamma_n(x; p, \theta) = \max_{(p', \theta') \in \partial B(p, \theta) \times \mathbb{S}^{r-1}} \rho^T g(x, \theta').
\]

Since \( g(x, \theta) \) is continuous in \( \theta \) for almost all \( x \in \mathcal{X} \), \( \gamma_n(X; p, \theta) \to \rho^T g(X, \theta) \) with probability one. Thus, for all \( \theta \in B(\theta_0, \epsilon) \) and \( p \in \mathbb{S}^{r-1} \), there exists \( m_{p, \theta} = m(p, \theta) \) such that \( \Pr(\gamma_n(X; p, \theta) > 0) < 1 \).

Since

\[
\mathbb{S}^{r-1} \times B(\theta_0, \epsilon) \subseteq \bigcup_{(p, \theta) \in \mathbb{S}^{r-1} \times B(\theta_0, \epsilon)} B((p, \theta), m_{p, \theta})
\]

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

\[
\mathbb{S}^{r-1} \times B(\theta_0, \epsilon) \subseteq \bigcup_{i=1}^h B((p_i, \theta_i), m_{p_i, \theta_i}).
\]

Therefore, with probability one,

\[
\sup_{\theta \in B(\theta_0, \epsilon)} \sup_{|p| = 1} n^{-1} \sum_{i=1}^n \left( \rho^T g(X_i, \theta) > 0 \right) \leq \max_{1 \leq i \leq h} n^{-1} \sum_{i=1}^n \left( \gamma_{n, \theta_i}(X_i; p_i, \theta_i) > 0 \right).
\]

Since

\[
n^{-1} \sum_{i=1}^n \left( \gamma_{n, \theta_i}(X_i; p_i, \theta_i) > 0 \right) \to \Pr(\gamma_{n, \theta_i}(X_i; p_i, \theta_i) > 0) < 1
\]

with probability one for all \( i = 1, \ldots, h \), we conclude that \( \Pr(A_{\beta_0}) \to 0 \), as stated.

It is known that when \( 0 \in \text{int}(\text{conv}(g(X_i, \hat{\theta}_1) : 1 \leq i \leq n)) \), the ELB weights \( \hat{v}_i \) have the following expressions (Owen 1990):

\[
\hat{v}_i = \frac{1}{n} \left( 1 + \hat{\lambda}^T g(X_i, \hat{\theta}_1) \right)
\]

for all \( i = 1, \ldots, n \), where \( \hat{\lambda} \) satisfies \( \sum_{i=1}^n \hat{v}_i g(X_i, \hat{\theta}_1) = 0 \).

\[
\text{(C.4)}
\]

**Lemma C.3** Suppose the model (1) and conditions A–E hold. Then (i) \( \hat{\lambda} = \mathcal{O}_P(n^{-1/2}) \); (ii) \( \max_{1 \leq i \leq n} \hat{v}_i = \mathcal{O}_P(n^{-1}) \); and (iii) \( \max_{1 \leq i \leq n} \hat{v}_i - n^{-1} = \mathcal{O}_P(n^{-1}) \).

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

\[
n^{-1} - \hat{v}_i = \hat{v}_i g(X_i, \hat{\theta}_1)^T \hat{\lambda}, \quad 1 \leq i \leq n.
\]

Left multiplication of (5) by \( g(X_i, \hat{\theta}_1) \) and summing over \( i \) yields \( \bar{V}(\hat{\theta}_1) \hat{\lambda} = \bar{g}(\hat{\theta}_1) \), where \( \bar{V}(\theta) = \sum_{i=1}^n \hat{v}_i g(X_i, \theta) g(X_i, \theta)^T \). Write \( \hat{\lambda} = \frac{1}{\|\hat{\lambda}\|} \hat{\rho} \), with \( \|\hat{\rho}\| = 1 \). Since \( \hat{v}_i \geq n^{-1}(1 + \|\hat{\lambda}\| M)^{-1} \), where \( M = \max_{1 \leq i \leq n} \|g(X_i, \hat{\theta}_1)\| \), we have

\[
\|\hat{\lambda}\| \|\hat{\rho}\|^T \hat{V}(\hat{\theta}_1) \hat{\rho} \leq (1 + \|\hat{\lambda}\| M) \sum_{i=1}^n \hat{v}_i |g(X_i, \hat{\theta}_1) g(X_i, \hat{\theta}_1)^T| \hat{\rho}
\]

\[
= (1 + \|\hat{\lambda}\| M) \hat{V}(\hat{\theta}_1) \hat{\lambda} = (1 + \|\hat{\lambda}\| M) \hat{\rho}^T \bar{g}(\hat{\theta}_1),
\]

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

\[
\|\hat{\lambda}\| \|\hat{\rho}\|^T \hat{V}(\hat{\theta}_1) \hat{\rho} - M \hat{\rho}^T \bar{g}(\hat{\theta}_1) \leq \hat{\rho}^T g(\hat{\theta}_1).
\]

Let \( \sigma_0 > 0 \) be the smallest eigenvalue of \( V_0 \). Since \( M = \mathcal{O}_P(n^{-1/2}) \), \( \hat{V}(\hat{\theta}_1) = V_0 + \mathcal{O}_P(1) \), and \( \bar{g}(\hat{\theta}_1) = \mathcal{O}_P(n^{-1/2}) \), then \( \|\hat{\lambda}\| (\sigma_0 + \mathcal{O}_P(1)) \leq \mathcal{O}_P(n^{-1/2}) \). Hence \( \hat{\lambda} = \mathcal{O}_P(n^{-1/2}) \), as stated at (i). Since \( \|\hat{\lambda}\| M = \mathcal{O}_P(1) \), then \( \Pr(\|\hat{\lambda}\| M < 1) \to 1 \). Note that on the event \( \|\hat{\lambda}\| M < 1 \), we have

\[
(1 + \|\hat{\lambda}\| M) \hat{\lambda} = (1 - \|\hat{\lambda}\| M)^{-1} \leq 1 - \|\hat{\lambda}\| M^{-1}.
\]

Thus, \( \max_{1 \leq i \leq n} \hat{v}_i = \mathcal{O}_P(n^{-1}) \), as stated at (ii). Finally,

\[
\max_{1 \leq i \leq n} \left| \hat{v}_i - n^{-1} \right| \leq \|\hat{\lambda}\| \max_{1 \leq i \leq n} \|g(X_i, \hat{\theta}_1)\| \leq \mathcal{O}_P(n^{-1/2}) \mathcal{O}_P(n^{-1/2}) \mathcal{O}_P(n^{-1/2}) = \mathcal{O}_P(n^{-1}),
\]

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,
\[ \mathcal{L}_E(n^{1/2} \hat{g}^* (\hat{\theta}_T) | X_{1:n} ) \overset{p}{\to} N(0, V_0). \]

Proof (Lemma C.4) Note that \( \hat{g}^* (\hat{\theta}_T) \) is the average of \( n \) conditionally i.i.d. random variables, with conditional mean
\[ \mathbb{E}_i [g^* (X^*, \hat{\theta}_T) | X_{1:n}] = \sum_{i=1}^n \hat{\psi}_i g(X_i, \hat{\theta}_T) = 0 \]
and conditional covariance matrix
\[ \mathbb{V}_i (\hat{\theta}_T) = \text{Var}_i [g^* (X^*, \hat{\theta}_T) | X_{1:n}] = \sum_{i=1}^n \hat{\psi}_i g(X_i, \hat{\theta}_T) g(X_i, \hat{\theta}_T)^T. \]

By Lemma C.3,
\[ \| \hat{V}_i (\hat{\theta}_T) - \hat{V}_G (\hat{\theta}_T) \| \leq \max_{1 \leq i \leq n} \| \hat{\psi}_i - n^{-1} \sum_{j=1}^n g(X_j, \hat{\theta}_T) \| \leq O_p(n^{-1})O_p(n) = o_p(1), \]
and since \( \hat{V}_i (\hat{\theta}_T) = V_0 + o_p(1) \), then \( \hat{V}_i (\hat{\theta}_T) = V_0 + o_p(1) \). To complete the proof, we prove that a (conditional) Lindeberg condition holds; i.e., for every \( \varepsilon > 0 \),
\[ \mathbb{E}_i (\| g^* (X^*, \hat{\theta}_T) \|^2 I(\| g^* (X^*, \hat{\theta}_T) \| \geq \varepsilon n^{1/2}) | X_{1:n} ) = o_p(1). \]

To this end, note that,
\[ \mathbb{E}_i (\| g^* (X^*, \hat{\theta}_T) \|^2 I(\| g^* (X^*, \hat{\theta}_T) \| \geq \varepsilon n^{1/2}) | X_{1:n} ) = \sum_{i=1}^n \hat{\psi}_i \| g(X_i, \hat{\theta}_T) \|^2 I(\| g(X_i, \hat{\theta}_T) \| \geq \varepsilon n^{1/2}) \]
\[ \leq \left\{ \max_{1 \leq i \leq n} \| \hat{\psi}_i \| I(M \geq \varepsilon n^{1/2}) \sum_{i=1}^n g(X_i, \hat{\theta}_T) \| \leq O_p(n^{-1})O_p(n) = o_p(1). \]

This completes the proof of the lemma.

Proof (Theorem 6) Similarly to (B.1),
\[ n^{1/2} (\hat{\eta}^* - \hat{\eta}) = -n^{1/2} \begin{pmatrix} \Lambda_1^{-1} & O \\ O & \Lambda_2 \end{pmatrix} \hat{\Psi}^*(\hat{\eta}) + o_p(1). \] (C.6)

Similarly to the proof of Lemma C.4,
\[ \mathcal{L}_E \left( n^{1/2} \left( \hat{g} (\hat{\theta}_0) - \hat{g} (\hat{\theta}_0) \right) \right) \overset{p}{\to} N \left( 0, \begin{pmatrix} V_0 & V_0 \\ V_0 & V_0 \end{pmatrix} \right). \] (C.7)

where \( \hat{g} (\theta) = \sum_{i=1}^n \hat{\psi}_i g(X_i, \theta) \). Since \( n^{1/2} \hat{g}^* (\hat{\theta}_0) \hat{\Psi}^* (\hat{\theta}_0) = o_p(1) \), by (C.6) and (C.7), then (20a) holds. By the delta method for the bootstrap, then (20b) holds. Similarly to (B.5),
\[ \mathcal{L}_E (n^{1/2} (\hat{\theta}_T - \hat{\theta}_T) | X_{1:n} ) \overset{p}{\to} N(0, \Lambda_2). \]

Similarly to (B.6) and (B.7), then (20c) and (20d) hold. Similarly to (C.1),
\[ n^{1/2} V_0^{-1/2} \hat{g} (\hat{\theta}_T) = n^{1/2} \Lambda_2 V_0^{-1/2} \hat{g} (\hat{\theta}_T) + o_p(1), \]
and by Lemma C.4, then (20e) holds, as stated.
Proof (Theorem 7) Note first that by (19), $\lim_{n \to \infty} n^{-\alpha} \mathcal{X}^2_{a, \alpha} = 1$. 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

$$\lim sup_{n \to \infty} n^{-\alpha} \log \left( \Pr \left( \left\| \mathbf{Z} \right\| \geq Cn^{\beta/2} \right) \right) \leq -c.$$

Hence, for $n$ large enough, $\Pr \left( \left\| \mathbf{Z} \right\| \leq \mathcal{X}^2_{a, \alpha} \right) \geq 1 - \exp(-cn^\beta)$. Thus, on an event of probability exponentially close to 1, the ELB weights are not modified to the uniform weights. Note further that

$$\Pr \left( \hat{\Theta} \right) \leq s|X_{1:n}) = \Pr \left( \hat{\Theta} \right) \leq s|X_{1:n}) - n^{-1/2} d \frac{d}{dx} \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n})$$

$$- n^{-1} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) - n^{-3/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) + (1/2) n^{-1} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n})$$

$$+ n^{-3/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) - (1/6) n^{-3/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) + O(n^{-1}),$$

where $\hat{\Theta}$ are the ELB versions of $\hat{\Theta}$. We further obtain

$$\Pr \left( \hat{\Theta} \right) \leq s|X_{1:n}) = \int \left( 1 + n^{-1/2} \hat{p}_1(z) + n^{-1/2} \hat{p}_2(z) + n^{-1/2} \hat{p}_3(z) \right) \phi(z) dz$$

$$- n^{-1/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) - n^{-1} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n})$$

$$- n^{-3/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) + (1/2) n^{-1} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n})$$

$$+ n^{-3/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) - (1/6) n^{-3/2} d \frac{d}{dx} \left( \hat{\Theta} \right) \leq s|X_{1:n}) + O(n^{-1}),$$

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

$$\Pr \left( \hat{\Theta} \right) \leq s|X_{1:n}) = F(x; r) + n^{-1} \hat{R}(s) + O(n^{-2}),$$

where $\hat{R}(s) = R(s) + O(n^{-1/2})$. Let $\xi_{a}^{\mathcal{X}^2_{a, \alpha}} = \mathcal{X}^2_{a, \alpha} + n^{-1} \hat{q}_1(\alpha) + O(n^{-2})$ be an empirical Cornish-Fisher expansion of the upper $\alpha$-quantile of $\xi_{a}^{\mathcal{X}^2_{a, \alpha}} \left( \hat{\Theta} \right) \left( \hat{\Theta} \right) \left( \hat{\Theta} \right) \left( \hat{\Theta} \right)$, Taylor expansion of $\Pr \left( \hat{\Theta} \right) \left( \hat{\Theta} \right) \leq s|X_{1:n})$, yields:

$$\xi_{a}^{\mathcal{X}^2_{a, \alpha}} = \mathcal{X}^2_{a, \alpha} - n^{-1} \left( \mathcal{X}^2_{a, \alpha} \right)^2 \hat{R}(\alpha) + O(n^{-2}).$$

Hence, $\xi_{a}^{\mathcal{X}^2_{a, \alpha}} = \mathcal{X}^2_{a, \alpha} - n^{-1} \left( \mathcal{X}^2_{a, \alpha} \right)^2 \hat{R}(\alpha)$. Therefore, $\Pr \left( \hat{\Theta} \right) \left( \hat{\Theta} \right) \leq s|X_{1:n}) = \alpha + O(n^{-2})$, as stated. Similar results hold for the other test statistics.
References