Supplemental Appendix - Auxiliary Lemmas for the proof of Theorem 2.4

Throughout Appendix C, we employ the notation of Section 2.4, which emphasizes the dependence on $P \in \mathbf{P}$. Equations new to this appendix have the prefix $C.\#$, while equations in the main text are references by their number.

Lemma C.1. *Let Assumption 2.3 hold, and denote the Edgeworth expansion for $P(T_n \leq z)$ by:*

$$\mathcal{E}_n(z,P) \equiv \Phi(z) + \frac{\phi(z)\kappa(P)}{6\sigma(P)^3\sqrt{n}}(2z^2+1) - \frac{\phi(z)}{\sigma(P)^3\sqrt{n}}(c'H_0(P)^{-1}\Sigma_0(P)H_0(P)^{-1}\gamma_0(P)(z^2+1) - \gamma_1(P)\sigma^2(P)) \ . \quad \text{(C.1)}$$

If $c \neq 0$, then it follows that $\limsup_{n\to\infty}\sup_{P\in\mathbf{P}}\sup_{z\in\mathbf{R}}\sqrt{n}|P(T_n \leq z) - \mathcal{E}_n(z,P)| = 0$.

Proof: For fixed $P \in \mathbf{P}$, the validity of the Edgeworth expansion has already been established in Theorem 2.3. We establish the Lemma by showing Assumption 2.3 controls all approximation errors uniformly. Specifically, in lieu of (15) and (16) note that with $\tilde{\nu}$ in place of ν: Lemma A.2(i) holds uniformly in $P \in \mathbf{P}$ due to $\sup_{P\in\mathbf{P}} E_P[\|X\epsilon\|^{\tilde{\nu}}] < \infty$ by Assumption 2.3(ii); Lemma A.2(ii) holds uniformly in $P \in \mathbf{P}$ due to Assumptions 2.3(ii)-(iii) implying:

$$0 < \inf_{P\in\mathbf{P}} \|H_0(P)^{-1}\|_o < \sup_{P\in\mathbf{P}} \|H_0(P)^{-1}\|_F < \infty \ , \quad \text{(C.2)}$$

and $\sup_{P\in\mathbf{P}} E_P[\|XX'\|_F^{\tilde{\nu}}] < \infty$ by Assumption 2.3(ii); Lemma A.2(iii) holds uniformly in $P \in \mathbf{P}$ by (C.2); and Lemma A.2(iv) holds uniformly in $P \in \mathbf{P}$ by $E_P[\|XX'\|_F^{\tilde{\nu}}]$, $E_P[\|XX'\epsilon^2\|_F^{\tilde{\nu}}]$, $E_P[\|(c'H_0(P)^{-1}X)^2\epsilon X\|^{\frac{\tilde{\nu}}{2}}]$, $\|\gamma_0(P)\|$ and $\|\Sigma_0(P)\|_F$ being uniformly bounded in $P \in \mathbf{P}$ by Assumptions 2.3(ii)-(iii) and result (C.2). Similarly, since $\inf_{P\in\mathbf{P}} \sigma(P) > 0$ by Assumption 2.3(iii), we get by result (C.2) and $\sup_{P\in\mathbf{P}} \|\gamma_0(P)\| < \infty$ by Assumptions 2.3(ii)-(iii), that the arguments in Lemma A.3 hold uniformly in $P \in \mathbf{P}$. Therefore we obtain for any $\alpha \in [0, \frac{2\tilde{\nu}-3}{2\tilde{\nu}})$:

$$\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} \sqrt{n}P(|T_n - L_n(P)| > n^{-\alpha}) = 0 \ . \quad \text{(C.3)}$$

Let $Z \in \mathbf{R}^{d_z}$ be as in Assumption 2.3(iv), set $S_n(P) = \frac{1}{\sqrt{n}}\sum_i(Z_i - E_P[Z_i])$, $V(P) = E_P[ZZ']$ and $\Phi_{V(P)}$ to be a mean zero Gaussian measure on $\mathbf{R}^{d_z}$ with covariance $V(P)$. For $\mathcal{X}_k(S_n(P))$ the k^{th} cumulant of $S_n(P)$ under P, and P_j the Cramer-Edgeworth measures we next aim to show that for any Borel set B and all $P \in \mathbf{P}$:

$$|P(S_n(P) \in B) - \sum_{j=0}^{1}\int_B dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\})| \leq \delta_n + \Phi_{V(P)}((\partial B)^{2e^{-dn}}) \quad \text{(C.4)}$$

where $\delta_n = o(n^{-\frac{1}{2}})$ and $d > 0$ are independent of B and P. The validity of the Edgeworth expansion in (C.4) pointwise in $P \in \mathbf{P}$ is immediate from Assumption 2.3 and Theorem 20.1 in Bhattacharya and Rao (1976). Most of their error bounds can be controlled uniformly by $\sup_{P\in\mathbf{P}} E_P[\|Z\|^4] < \infty$. The only necessary modifications to their arguments is in their equation (20.22) which can be controlled uniformly due to $\inf_{P\in\mathbf{P}} \lambda(E_P[ZZ']) > 0$ by Assumption 2.3(iv), and in their equations (20.29)-(20.34), which can be controlled uniformly in $P \in \mathbf{P}$ since:

$$\sup_{\|t\| \geq \frac{\sqrt{n}}{16E_P[\|Z\|^3]}} |\xi_{Z,P}(t/\sqrt{n})| \leq \sup_{\|t\|\geq(16\sup_{P\in\mathbf{P}} E_P[\|Z\|^3])^{-1}} |\xi_{Z,P}(t)| \leq \sup_{\|t\|\geq(16\sup_{P\in\mathbf{P}} E_P[\|Z\|^3])^{-1}} F(t) < 1 \ , \quad \text{(C.5)}$$

due to Assumption 2.3(iv). The remaining arguments in establishing (C.4) are identical to their proof and therefore omitted; see also Lemma 2 in Singh and Babu (1990) for the univariate case.

Next, let $G_P : \mathbf{R}^{d_z} \to \mathbf{R}$ be such that $L_n(P) = \sqrt{n}G_P(\frac{1}{n}\sum_i Z_i)$, and note $G_P(E_P[Z]) = 0$. Further define $g_{n,P}(z) = \sqrt{n}G_P(E_P[Z] + z/\sqrt{n})$ and note $L_n(P) = g_{n,P}(S_n(P))$. Exploiting result (C.4), we aim to establish that:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{z\in\mathbf{R}} \sqrt{n}|P(L_n(P) \le z) - \mathcal{E}_n(z,P)| = \limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{z\in\mathbf{R}} \sqrt{n}|P(g_{n,P}(S_n(P)) \le z) - \mathcal{E}_n(z,P)| = 0 \quad \text{(C.6)}$$

The validity of (C.6) pointwise in P follows from Assumption 2.3 and Theorem 2 in Bhattacharya and Ghosh (1978). The arguments leading to a uniform result are similar, and we describe only the necessary modifications. To this end, let $K > 0$ satisfy $\sup_{P\in\mathbf{P}} \|E_P[ZZ']\|_F < K < \infty$, which is feasible by Assumption 2.3(ii), and define $M_n \equiv \{z \in \mathbf{R}^{d_z} : \|z\| < K\log(n)\}$. By Assumption 2.3(ii), $\{\mathcal{X}_k(S_n(P))\}_{k=1}^3$ are bounded in $P \in \mathbf{P}$, and hence:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sum_{j=0}^{1} \sqrt{n}|\int_{(M_n)^c} dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\})| = 0 \ . \quad \text{(C.7)}$$

Since in addition $\nabla g_{n,P}(\tilde{z})$ is uniformly bounded on $(\tilde{z}, P) \in M_n \times \mathbf{P}$ and n by Assumption 2.3(ii)-(iii), Lemma 2.1 in Bhattacharya and Ghosh (1978) holds uniformly in $P \in \mathbf{P}$. For each $z \in \mathbf{R}$, then define the set $A_{n,P}(z) \equiv \{\tilde{z} \in \mathbf{R}^{d_z} : g_{n,P}(\tilde{z}) \le z\}$ and note that by continuity $\partial A_{n,P}(z) \subseteq \{\tilde{z} \in \mathbf{R}^{d_z} : g_{n,P}(\tilde{z}) = z\}$. Moreover, $\nabla g_{n,P}(\tilde{z})$ being uniformly bounded on $M_n \times \mathbf{P}$ further implies that if $\tilde{z} \in \partial A_{n,P}(z) \cap M_n$, $\tilde{z}' \in M_n$, and $\|\tilde{z} - \tilde{z}'\| \le \epsilon$, then by the mean value theorem $g_{n,P}(\tilde{z}') \in z \pm M\epsilon$ for some M not depending on P, $\tilde{z}$ or n. Hence, $(\partial A_{n,P}(z))^\epsilon \cap M_n \subseteq \{\tilde{z} \in \mathbf{R}^{d_z} : g_{n,P}(\tilde{z}') \in z \pm M\epsilon\}$, and since $\sup_{P\in\mathbf{P}} \int_{M_n^c} d\Phi_{V(P)}(\tilde{z}) = o(n^{-\frac{1}{2}})$ by (C.7), we conclude:

$$\begin{aligned}\int_{(\partial A_{n,P}(z))^{2e^{-dn}}} d\Phi_{V(P)}(\tilde{z}) &= \int_{(\partial A_{n,P}(z))^{2e^{-dn}} \cap M_n} d\Phi_{V(P)}(\tilde{z}) + o(n^{-\frac{1}{2}}) \\ &\le 2\sum_{j=0}^{1} \int_{\{\tilde{z}: g_{n,P}(\tilde{z}) \in z \pm M\epsilon\}} dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\}) + o(n^{-\frac{1}{2}}) \le O(e^{-dn}) + o(n^{-\frac{1}{2}}) \ , \quad \text{(C.8)}\end{aligned}$$

where the first inequality holds for n large enough uniformly in P by arguing as in (20.37) in Bhattacharya and Rao (1976), while the second inequality holds by Lemma 2.1 in Bhattacharya and Ghosh (1978), Corollary 3.2 in Bhattacharya and Rao (1976) and Assumptions 2.3(ii)-(iv). Therefore, by (C.4) and (C.8):

$$\sup_{P\in\mathbf{P}} \sup_{z\in\mathbf{R}} |P(L_n(P) \le z) - \sum_{j=0}^{1} \int_{A_{n,P}(z)} dP_j(-\Phi_{V(P)} : \{\mathcal{X}_k(S_n(P))\})| = o(n^{-\frac{1}{2}}) \ , \quad \text{(C.9)}$$

where we have used that $L_n(P) \le z$ if and only if $S_n(P) \in A_{n,P}(z)$. Replacing equation (2.20) in Bhattacharya and Ghosh (1978) with result (C.9), claim (C.6) then follows using the same arguments in the proof of Theorem 2 in Bhattacharya and Ghosh (1978) and noting that due to Assumption 2.3(ii)-(iii) the arguments in Lemmas A.8 and A.9 hold uniformly in $P \in \mathbf{P}$. The claim of the Lemma then follows from (C.3), (C.9), Assumptions 2.3(ii)-(iii) implying the coefficients in $\mathcal{E}_n(\cdot, P)$ are bounded in $P \in \mathbf{P}$ and Lemma 5 in Andrews (2002). ■

Lemma C.2. *Let Assumptions 2.2(i)-(ii), 2.3(i)-(iii) hold and $T^*_{s,n}$ be as in (47). It then follows that for any $9 \le \zeta \le 2\tilde{\nu}$, and $\alpha \in [0, \frac{(2\omega)\wedge\zeta - 2}{(2\omega)\wedge\zeta} - \frac{1}{2(\omega\wedge\zeta)})$ there exists a deterministic sequence $\delta_n = o(n^{-\frac{1}{2}})$ and sets $A_n \subseteq \mathbf{R}^{n(d_x+1)}$ such that $P^*(|T^*_{s,n} - T^*_n| > n^{-\alpha}) \le \delta_n$ whenever $\{Y_i, X_i\}_{i=1}^n \in A_n$ and $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin A_n) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$.*

Proof: Let K_0 satisfy $\sup_{P\in\mathbf{P}}\{\|H_0(P)^{-1}\|_o^\zeta E_P[\|X\epsilon\|^\zeta]\} < K_0 < \infty$ which is possible by Assumption 2.3(ii)-(iii), and:

$$A_{0n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \frac{1}{n}\sum_{i=1}^{n} \|H_n^{-1}\|_o^\zeta \|X_i(Y_i - X_i'\hat{\beta})\|^\zeta < K_0\} \ . \quad \text{(C.10)}$$

For any $\alpha_0 \in [0, \frac{\omega\wedge\zeta-1}{2(\omega\wedge\zeta)})$, we then obtain from (41) together with (39) and (40) that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{0n}$,

$$P^*(\|\hat{\beta}^* - \hat{\beta}\| > n^{-\alpha_0}) \leq \frac{C_0 K_0}{n^{(\frac{1}{2}-\alpha_0)(\omega\wedge\zeta)}} \tag{C.11}$$

for some constant $C_0 > 0$. Similarly, let $\sup_{P\in\mathbf{P}}\{(2d_x^2)^{\frac{\zeta}{2}}\|c\|^\zeta\|H_0(P)^{-1}\|_o^\zeta E_P[\|X\|^{\frac{3\zeta}{2}}|\epsilon|^{\frac{\zeta}{2}}]\} < K_1 < \infty$, and:

$$A_{1n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \frac{(2d_x^2)^{\frac{\zeta}{2}}\|c\|^\zeta}{n}\sum_{i=1}^n \|H_n^{-1}\|_o^\zeta\|X_i\|^{\frac{3\zeta}{2}}|(Y_i - X_i'\hat{\beta})|^{\frac{\zeta}{2}} < K_1\} . \tag{C.12}$$

For $X_i^{(l)}$ the l^{th} coordinate of X_i, we obtain by (39) and (40) that for any $1 \leq j \leq k \leq d_x$ and $\alpha_1 \in [0, \frac{\omega\wedge(\zeta/2)-1}{2(\omega\wedge(\zeta/2))})$:

$$P^*(\|c\|^2\|H_n^{-1}\|_o^2\|\frac{2d_x^2}{n}\sum_{i=1}^n X_i^{(j)}X_i^{(k)}X_i\epsilon_i^*\| > n^{-\alpha_1}) \leq \frac{C_1K_1}{n^{(\frac{1}{2}-\alpha_1)(\omega\wedge(\zeta/2))}} \tag{C.13}$$

for some $C_1 > 0$ whenever $\{Y_i, X_i\}_{i=1}^n \in A_{1n}$. Set $\sup_{P\in\mathbf{P}}\{\|c\|^2\|H_0(P)^{-1}\|_o^2 E_P[\|XX'\|_F\|X\|^2]\} < K_2 < \infty$, and:

$$A_{2n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \|c\|^2\|H_n^{-1}\|_o^2\frac{1}{n}\sum_{i=1}^n\|X_iX_i'\|_F\|X_i\|^2 < K_2\} . \tag{C.14}$$

We then obtain from (42), (43), (44), together with (C.11) and (C.12) that for any $\alpha_1 \in [0, \frac{\omega\wedge(\zeta/2)-1}{2(\omega\wedge(\zeta/2))})$ there exists a constant $C_2 > 0$ (depending on K_0, K_1, K_2, ω and ζ) such that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{0n}\cap A_{1n}\cap A_{2n}$:

$$P^*(|(\hat{\sigma}^*)^2 - (\hat{\sigma}_s^*)^2| > n^{-\alpha_1}) \leq \frac{C_2}{n^{(\frac{1}{2}-\alpha_1)(\omega\wedge(\zeta/2))}} . \tag{C.15}$$

Let $\sup_{P\in\mathbf{P}}\{\|c\|^4\|H_0(P)^{-1}\|_o^4 E_P[\|XX'\epsilon^2\|_F^2]\} < K_3 < \infty$ which is possible by Assumption 2.3(ii), and define:

$$A_{3n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \|c\|^4\|H_n^{-1}\|_o^4\frac{1}{n}\sum_{i=1}^n\|X_iX_i'(Y_i - X_i\hat{\beta})^2\|_F^2 < K_3\} . \tag{C.16}$$

The inequalities (39) and (40) then imply that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{3n}$, for any $\epsilon > 0$ we obtain that:

$$P^*(|(\hat{\sigma}_s^*)^2 - \hat{\sigma}^2| > \epsilon) \leq \frac{C_3}{\epsilon^2 n} . \tag{C.17}$$

Therefore, setting $\inf_{P\in\mathbf{P}}\sigma^2(P) > \epsilon_0 > 0$, which is feasible by Assumption 2.3(iii) and letting $A_{4n} \equiv \{\{Y_i, X_i\}_{i=1}^n : \hat{\sigma}^2 > \epsilon_0\}$, we obtain from (C.17) that whenever $\{Y_i, X_i\}_{i=1}^n \in A_{3n}\cap A_{4n}$ we must have:

$$P^*((\hat{\sigma}_s^*)^2 < \epsilon_0/2) \leq \frac{2C_3}{\epsilon_0 n} . \tag{C.18}$$

Letting $A_n = \bigcap_{j=0}^4 A_{jn}$, we then obtain from (48) together with (C.11), (C.17) and (C.18) and Assumptions 2.2(ii), 2.3(ii) that the desired deterministic sequence $\delta_n = o(n^{-\frac{1}{2}})$ exists.

To conclude the proof, we next show that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_n^c) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$. To this end, note that:

$$\sup_{P\in\mathbf{P}} P(\|H_n - H_0(P)\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) \tag{C.19}$$

for any $\eta > 0$ due to (15), (16) and Assumption 2.3(ii). Moreover, since $\sup_{P\in\mathbf{P}}\|H_0(P)^{-1}\|_o > 0$ by Assumption 2.3(iii), (C.19) implies $\sup_{P\in\mathbf{P}} P(\|H_0(P)^{-1}(H_n - H_0(P))\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{2}})$, and therefore (18) and (19) yield:

$$\sup_{P\in\mathbf{P}} P(\|H_n^{-1} - H_0(P)^{-1}\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) . \tag{C.20}$$

Therefore, by (C.20) and Assumption 2.3(iii) there exists an $M_0 > 0$ such that $\sup_{P\in\mathbf{P}} P(\|H_n^{-1}\|_F > M_0) = O(n^{-\frac{\tilde{\nu}}{2}})$.

It then follows by Assumption 2.3(ii) and (15) and (16), that for any $\eta > 0$ we have:

$$\sup_{P\in\mathbf{P}} P(\|\hat{\beta}-\beta_0\| > \eta) \leq \sup_{P\in\mathbf{P}} P(\|\frac{1}{n}\sum_{i=1}^{n} X_i\epsilon_i\| > \frac{\epsilon}{M_0}) + O(n^{-\frac{\tilde{\nu}}{2}}) = O(n^{-\frac{\tilde{\nu}}{2}}) \ . \tag{C.21}$$

Since (C.20), the mean value theorem and Assumption 2.3(iii) yield $\sup_{P\in\mathbf{P}} P(|\|H_n^{-1}\|_o^{\zeta} - \|H_0(P)^{-1}\|_o^{\zeta}| > \eta) = O(n^{-\frac{\tilde{\nu}}{2}})$, and $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_iX_i'\|_F^{\zeta} - E_P[\|XX'\|_F^{\zeta}]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$ by Assumption 2.3(ii) and (15) and (16):

$$\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^{n} \|H_n^{-1}\|_o^{\zeta}\|X_i(Y_i - X_i'\hat{\beta})\|^{\zeta} - \|H_0(P)^{-1}\|_o^{\zeta} E_P[\|X\epsilon\|^{\zeta}]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}}) \tag{C.22}$$

due to (C.21). Since (C.22) holds for any $\eta > 0$, the definition of A_{0n} and the constant K_0 in turn imply that:

$$\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_{0n}^c) \leq \sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^{n} \|H_n^{-1}\|_o^{\zeta}\|X_i(Y_i - X_i'\hat{\beta})\|^{\zeta} - \|H_0(P)^{-1}\|_o^{\zeta} E_P[\|X\epsilon\|^{\zeta}]|$$

$$> K_0 - \sup_{P\in\mathbf{P}} \|H_0(P)^{-1}\|_o^{\zeta} E_P[\|X\epsilon\|^{\zeta}]) = O(n^{-\frac{\tilde{\nu}}{2\zeta}}) \ . \tag{C.23}$$

Analogously, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_i\|^{2\zeta} - E_P[\|X\|^{2\zeta}]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$ due to (15), (16) and Assumption 2.3(ii) implying $\sup_{P\in\mathbf{P}} E_P[(\|X\|^{2\zeta})^{\delta}] < \infty$ for any $\delta \leq \tilde{\nu}/\zeta$. Similarly, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_i\|^{\zeta}\|X_i\epsilon_i\|^{\frac{\zeta}{2}} - E_P[\|X\|^{\zeta}\|X\epsilon\|^{\frac{\zeta}{2}}]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\zeta}})$, and therefore from (C.20), (C.21) and arguing as in (C.22) and (C.23):

$$\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_{1n}^c) = O(n^{-\frac{\tilde{\nu}}{2\zeta}}) \ . \tag{C.24}$$

The same arguments, but bounding $\sup_{P\in\mathbf{P}} E_P[(\|XX'\|_F\|X\|^2)^{\delta}]^2 \leq \sup_{P\in\mathbf{P}}\{E_P[\|XX'\|_F^{2\delta}]E_P[\|X\|^{4\delta}]\} < \infty$ for $\delta \leq \tilde{\nu}/2$, and $\sup_{P\in\mathbf{P}} E_P[(\|XX'X\|^2)^{\delta}]^2 \leq \sup_{P\in\mathbf{P}}\{E_P[\|XX'\|^{4\delta}]E_P[\|X\|^{4\delta}]\} < \infty$ for $\delta \leq \tilde{\nu}/4$, yields:

$$\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \in A_{2n}^c) = O(n^{-\frac{\tilde{\nu}}{4}}) \qquad \sup_{P\in\mathbf{P}} \max\{P(\{Y_i, X_i\}_{i=1}^n \in A_{3n}^c), P(\{Y_i, X_i\}_{i=1}^n \in A_{4n}^c)\} = O(n^{-\frac{\tilde{\nu}}{8}}) \tag{C.25}$$

The lemma then follows from $P(\{Y_i, X_i\}_{i=1}^n \in A_n^c) \leq \sum_{j=1}^4 P(\{Y_i, X_i\}_{i=1}^n \in A_{jn}^c)$, (C.23), (C.24) and (C.25). ■

Lemma C.3. *Let Assumptions 2.2, 2.3(i)-(iii) hold, and* $(\sup_{P\in\mathbf{P}} |\kappa(P)|)/(\inf_{P\in\mathbf{P}} \sigma(P)^3) < C_0$. *In addition, denote*

$$\mathcal{E}_n^*(z) \equiv \Phi(z) + \frac{\phi(z)E[W^3]}{6\sqrt{n}}(2z^2+1) \times (\frac{|\hat{\kappa}|}{\hat{\sigma}^3} \wedge C_0) \times sign\{\hat{\kappa}\} \ , \tag{C.26}$$

then there exist a deterministic $\delta_n = o(n^{-\frac{1}{2}})$ *and sets* $A_n \subset \mathbf{R}^{n(1+d_x)}$ *such that* $\sup_{z\in\mathbf{R}} |P^*(T_n \leq z) - \mathcal{E}_n^*(z)| \leq \delta_n$ *whenever* $\{Y_i, X_i\}_{i=1}^n \in A_n$ *and in addition* $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin A_n) = O(n^{-\frac{\tilde{\nu}}{18}})$. *Additionally, for any* $\epsilon > 0$:

$$\sup_{P\in\mathbf{P}} P(|\frac{\kappa(P)}{\sigma(P)^3} - \frac{\hat{\kappa}}{\hat{\sigma}^3}| > \epsilon) = O(n^{-\frac{\tilde{\nu}}{8}}) \tag{C.27}$$

Proof: We first proceed as in Lemmas B.2 and B.3 by verifying the conditions of Theorems 3.4 in Skovgaard (1986) and 3.2 in Skovgaard (1981) respectively. Throughout, let $a_{in} \equiv c'H_n^{-1}X_i(Y_i - X_i\hat{\beta})$, $V_{in} \equiv (a_{in}W_i, a_{in}^2(W_i^2-1))$, $\Omega_n \equiv \frac{1}{n}\sum_i E^*[V_{in}V_{in}']$ and $S_n \equiv \frac{1}{\sqrt{n}}\sum_i \Omega_n^{-\frac{1}{2}}V_{in}$. We first aim to show there exist sets B_n such that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_n) = O(n^{-\frac{\tilde{\nu}}{18}})$, and that there exists a deterministic sequence $b_n = o(n^{-\frac{1}{2}})$ satisfying:

$$P^*(S_n \in B) = \sum_{j=0}^{1}\int_B dP_j(-\Phi_{I_2} : \{\mathcal{X}_k^*(S_n)\}) + b_n \ , \tag{C.28}$$

uniformly over all Borel sets B with $\int_{(\partial B)^{\epsilon}} d\Phi_{I_2}(u) \leq C\epsilon$ whenever $\{Y_i, X_i\}_{i=1}^n \in B_n$. To this end, let $a_i \equiv c'H_0(P)^{-1}X_i\epsilon_i$, $V_i \equiv (a_iW_i, a_i^2(W_i^2-1))$ and $\Omega(P) \equiv E_P[VV']$. By Assumption 2.3(ii)-(iii) and Exercise 3.8 in

Durrett (1996), there exists a $1 > \delta_0 > 0$ such that $\inf_{P\in\mathbf{P}} P(|a_i|^2 > \delta_0) > 0$, and hence by Assumption 2.3(iii):

$$\inf_{P\in\mathbf{P}} P(|a_i|^2 > \delta_0 \text{ and } \max\{\|X\epsilon\|, \|XX'\|_F\} \leq M_0) > \epsilon_0 \tag{C.29}$$

for some $M_0 < \infty$ and some $\epsilon_0 > 0$. We can now define the sequence of sets B_n, by $B_n = \bigcap_{j=0}^4 B_{jn}$, where:

$$\begin{aligned}
B_{0n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : \sup_{t\in\mathbf{R}^2} \tfrac{1}{3!\|t\|^3}|E^*[(t'S_n)^3]| \leq n^{-\frac{5}{18}}\} \\
B_{1n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : \|\Omega_n^{-\frac{1}{2}}\|_o(\tfrac{1}{n}\textstyle\sum_i\{|a_{in}|^{4.5} + |a_{in}|^9\})^{\frac{2}{4.5}} < 2\sup_{P\in\mathbf{P}}\{\|\Omega(P)^{-\frac{1}{2}}\|_o(E_P[|a_i|^{4.5}] + E_P[|a_i|^9])^{\frac{2}{4.5}}\}\} \\
B_{2n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : 2n^{-1}\|\Omega_n^{-\frac{1}{2}}\|_o^4 \textstyle\sum_i\{a_{in}^4 E[W^4] + a_{in}^8 E[(W^2-1)^4]\} \leq n^{\frac{4}{9}}\} \\
B_{3n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : \|\Omega_n^{-1}\|_o^{\frac{3}{2}} n^{-1}\textstyle\sum_i\{|a_{in}|^3 + |a_{in}|^6\} < 2\sup_{P\in\mathbf{P}}\{\|\Omega(P)^{-1}\|_o^{\frac{3}{2}} E_P[|a_i|^3 + |a_i|^6]\}\} \\
B_{4n} &\equiv \{\{Y_i, X_i\}_{i=1}^n : n^{-1}\textstyle\sum_i 1\{\min\{|a_{in}|, a_{in}^2\} \geq \delta_0/2\} > \epsilon_0/2 \text{ and } \|\Omega_n\|_o^{\frac{1}{2}} < 2\sup_{P\in\mathbf{P}}\|\Omega(P)\|_o^{\frac{1}{2}}\}
\end{aligned}$$

Then note that whenever $\{Y_i, X_i\}_{i=1}^n \in B_n$: (i) $\{Y_i, X_i\}_{i=1}^n \in B_{0n}$ implies Conditions (I) and (II) in Theorem 3.4 in Skovgaard (1986) are satisfied with $r_n \asymp n^{\frac{5}{18}}$; (ii) $\{Y_i, X_i\}_{i=1}^n \in B_{1n} \cap B_{2n}$ implies together with results (78)-(81) that Condition (IV) in Theorem 3.4 in Skovgaard (1986) is satisfied; (iii) $\{Y_i, X_i\}_{i=1}^n \in B_{3n} \cap B_{4n}$ implies by (84)-(86), together with setting $\epsilon < (\delta_0 \sup_{P\in\mathbf{P}}\{\|\Omega(P)^{-1}\|_o^{\frac{3}{2}} E_P[|a_i|^3 + |a_i|^6]\})/(2\sup_{P\in\mathbf{P}}\|\Omega(P)\|_o^{\frac{1}{2}})$ in equation (88), Assumption 2.2(ii) and (89) that Condition III" of Theorem 3.4 in Skovgaard (1986) also holds. Therefore, the existence of the desired deterministic sequence $b_n = o(n^{-\frac{1}{2}})$ follows from Theorem 3.4 in Skovgaard (1986).

We now verify $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_n) = O(n^{-\frac{\tilde{\nu}}{18}})$. To this end, let δ satisfy $1 \leq \delta \leq 9$. By result (C.20) and Assumption 2.3(iii), there exists a $0 < M_1 < \infty$ such that $\sup_{P\in\mathbf{P}} P(\|H_n^{-1}\|_F > M_1) = O(n^{-\frac{\tilde{\nu}}{2}})$. Moreover, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_iX_i\|_F^\delta - E_P[\|X_iX_i\|_F^\delta]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\delta}})$ for any $\eta > 0$ due to $\tilde{\nu} \geq 18$, and results (15) and (16). Hence, by Assumption 2.3(ii) there exists a $0 < M_2 < \infty$ such that $\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_i \|X_iX_i'\|_F^\delta > M_2) = O(n^{-\frac{\tilde{\nu}}{2\delta}})$. Combining these results, we then obtain that:

$$\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_{i=1}^n |c'H_n^{-1}X_iX_i'(\hat{\beta} - \beta_0)|^\delta > \eta) \leq \sup_{P\in\mathbf{P}} P(M_1M_2\|c\|^\delta\|\hat{\beta} - \beta_0\|^\delta > \eta) + O(n^{-\frac{\tilde{\nu}}{2\delta}}) = O(n^{-\frac{\tilde{\nu}}{2\delta}}) , \tag{C.30}$$

where the final equality follows from (C.21) and $\delta \geq 1$. Next, note that by (15), (16) and Assumption 2.3(ii) we have $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_i \|X_i\epsilon_i\|^\delta - E_P[\|X\epsilon\|^\delta]| > \eta) = O(n^{-\frac{\tilde{\nu}}{\delta}})$ for any $\eta > 0$. Therefore, by Assumption 2.3(ii), there exists a $0 < M_3 < \infty$ such that $\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_i \|X_i\epsilon_i\|^\delta > M_3) = O(n^{-\frac{\tilde{\nu}}{\delta}})$, and thus we have:

$$\sup_{P\in\mathbf{P}} P(\frac{1}{n}\sum_{i=1}^n |c'(H_n^{-1} - H_0^{-1})X_i\epsilon_i|^\delta > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) + O(n^{-\frac{\tilde{\nu}}{\delta}}) \tag{C.31}$$

due to result (C.20). Moreover, $\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n |a_i|^\delta - E_P[|a_i|^\delta]| > \eta) = O(n^{-\frac{\tilde{\nu}}{\delta}})$ for any $\eta > 0$ by the same arguments and Assumption 2.3(ii). Therefore, combining (C.30) and (C.31) we can conclude that:

$$\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n |a_{in}|^\delta - E_P[|a_i|^\delta]| > \eta) = O(n^{-\frac{\tilde{\nu}}{2\delta}}) \, . \tag{C.32}$$

Hence, result (C.32), the definition of Ω_n and $\Omega(P)$ and a_{in}, a_i being nonstochastic with respect to L^*, imply:

$$\sup_{P\in\mathbf{P}} P(\|\Omega_n - \Omega(P)\|_F > \eta) = O(n^{-\frac{\tilde{\nu}}{8}}) \tag{C.33}$$

for any $\eta > 0$. In addition, Assumptions 2.3(iii)-(iv) and $E[(W^2-1)^2] > 0$ by Assumption 2.2(i)-(ii) imply that

$\inf_{P\in\mathbf{P}} \lambda(\Omega(P)) > 0$, where $\lambda(\Omega(P))$ denotes the smallest eigenvalue of $\Omega(P)$. Hence, arguing as in (18)-(19):

$$\sup_{P\in\mathbf{P}} P(\|\Omega_n^{-1} - \Omega(P)^{-1}\|_o > \eta) = O(n^{-\frac{\tilde{\nu}}{8}}) \;, \tag{C.34}$$

for any $\eta > 0$. Therefore, employing (73)-(74) for B_{0n} and results (C.32) and (C.34) allow us to obtain the bounds:

$$\begin{array}{llll} \sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_{0n}) = & O(n^{-\frac{\tilde{\nu}}{12}}) & \sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_{1n}) & = O(n^{-\frac{\tilde{\nu}}{18}}) \\ \sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_{2n}) = & O(n^{-\frac{\tilde{\nu}}{16}}) & \sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_{3n}) & = O(n^{-\frac{\tilde{\nu}}{12}}) \end{array} . \tag{C.35}$$

Moreover, we also note by direct calculation that results (C.20) and (C.21) imply that (for M_0 as in (C.29)):

$$\sup_{P\in\mathbf{P}} P(\sup_{\max\{\|XX'\|_F, \|X\epsilon\|\}\leq M_0} |c'H_0(P)^{-1}X\epsilon - c'H_n^{-1}X(Y - X\hat{\beta})| > \eta) = O(n^{-\frac{\tilde{\nu}}{2}}) \;. \tag{C.36}$$

Hence, since $0 < \delta_0 < 1$, we obtain that on a set with probability $1 - O(n^{-\frac{\tilde{\nu}}{2}})$ (uniformly in $P \in \mathbf{P}$) we have:

$$\frac{1}{n}\sum_{i=1}^n 1\{\min\{|a_{in}|, a_{in}^2\} \geq \frac{\delta_0}{2}\} \geq \frac{1}{n}\sum_{i=1}^n 1\{a_i^2 \geq \delta_0 \text{ and } \max\{\|X_i\epsilon_i\|, \|X_iX_i'\|_F\} \leq M_0\} \;. \tag{C.37}$$

Thus, by (C.37), Bernstein's inequality and (C.29), together with (C.33) we conclude that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin B_{4n}) = O(n^{-\frac{\tilde{\nu}}{8}})$. Result (C.28) then follows by (C.35) and $P(\{Y_i, X_i\}_{i=1}^n \notin B_n) \leq \sum_{j=0}^4 P(\{Y_i, X_i\}_{i=1}^n \notin B_{jn})$.

Next, we aim to exploit result (C.28) to establish the existence of sets C_n such that $P(\{Y_i, X_i\}_{i=1}^n \notin C_n) = O(n^{-\frac{\tilde{\nu}}{18}})$ and a deterministic sequence $c_n = o(n^{-\frac{1}{2}})$ such that whenever $\{Y_i, X_i\}_{i=1}^n \in C_n$, then uniformly in $z \in \mathbf{R}$:

$$P^*(T_{s,n}^* \leq z) = \Phi(z) + \frac{\phi(z)E[W^3]}{6\sqrt{n}}(2z^2 + 1) \times \frac{\hat{\kappa}}{\hat{\sigma}^3} + c_n \;. \tag{C.38}$$

To this end, define $C_n = B_n \cap (\bigcap_{j=0}^2 C_{jn})$ where the sets C_{jn} are given by:

$$\begin{array}{ll} C_{0n} & \equiv \{\{Y_i, X_i\}_{i=1}^n : \hat{\sigma}^2 > \frac{1}{2}\inf_{P\in\mathbf{P}} \sigma^2(P) \text{ and } \|\Omega_n\|_F < \sup_{P\in\mathbf{P}} 2\|\Omega(P)\|_F\} \\ C_{1n} & \equiv \{\{Y_i, X_i\}_{i=1}^n : |E^*[(L_n^*)^2] - 1| \leq n^{-\frac{3}{4}}\} \\ C_{2n} & \equiv \{\{Y_i, X_i\}_{i=1}^n : |E^*[(L_n^*)^3] + (7E[W^3]\hat{\kappa})/(2\hat{\sigma}^3\sqrt{n})| \leq n^{-\frac{3}{4}}\} \end{array} .$$

Then note that whenever $\{Y_i, X_i\}_{i=1}^n \in C_n$: (i) $\{Y_i, X_i\}_{i=1}^n \in B_n$ and (C.28) implies condition (3.1) of Theorem 3.2 in Skovgaard (1981) is satisfied; (ii) $\{Y_i, X_i\}_{i=1}^n \in C_{0n}$ and result (100) verifies condition (3.11) of Theorem 3.2 in Skovgaard (1981), while $\{Y_i, X_i\}_{i=1}^n \in C_{0n}$ and result (101) verifies condition (3.12). The Edgeworth expansion in (C.38) then holds due to Theorem 3.2 and Remark 3.4 in Skovgaard (1981), Lemma A.7 and $\{Y_i, X_i\}_{i=1}^n \in C_{1n} \cap C_{2n}$. Moreover, by (C.33) and (C.25), $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin C_{0n}) = O(n^{-\frac{\tilde{\nu}}{8}})$, while from (56), (57) and (C.32), together with (C.25) we obtain $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin C_{1n}) = O(n^{-\frac{\tilde{\nu}}{8}})$ (note in Lemma A.8, $a_{in} = c'H_n^{-1}X_i$, and not $a_{in} = c'H_n^{-1}(Y_i - X_i\hat{\beta})$ as used in (C.32)). Finally, by direct calculation, we also obtain from (67)-(70 and (C.32), together with (C.25) that $\sup_{P\in\mathbf{P}} P(\{Y_i, X_i\}_{i=1}^n \notin C_{2n}) = O(n^{-\frac{\tilde{\nu}}{18}})$, and hence (C.38) follows.

Finally, note $\hat{\kappa} = n^{-1}\sum_i a_{in}^3$, (C.25), (C.30) and (C.31) verify (C.27), which implies $\sup_{P\in\mathbf{P}} P(\frac{|\hat{\kappa}|}{\hat{\sigma}^3} > C_0) = O(n^{-\frac{\tilde{\nu}}{8}})$. The Lemma then follows from (C.38), Lemma C.2 and Lemma 5 in Andrews (2002). ■

References

Andrews, D. W. K., 2002, Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. Econometrica 70, 119–162.

Bhattacharya, R. N. and Ghosh, J. K., 1978, On the validity of the formal edgeworth expansion. The Annals of Statistics 6, 434–451.

Bhattacharya, R. N. and Rao, R. R., 1976, Normal Approximation and Asymptotic Expansions. John Wiley & Sons, New York.

Durrett, R. 1996, Probability: Theory and Examples Second Edition. Duxbury Press, Belmont.

Singh K. and Babu G. J., 1990, On Asymptotic Optimality of the Bootstrap. Scandinavian Journal of Statistics 17, 1–9.

Skovgaard, I. M., 1981, Transformation of an edgeworth expansion by a sequence of smooth functions. Scandinavian Journal of Statistics 8, 207–217.

Skovgaard, I. M., 1986, On multivariate edgeworth expansions. International Statistical Review 54, 169–186.