Supplement to "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap"

Joseph P. Romano
Departments of Economics and Statistics
Stanford University
romano@stanford.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

September 6, 2012

Abstract

This document provides additional details and proofs for many of the results in the authors' paper "On the Asymptotic Validity of Subsampling and the Bootstrap".

S.1 Auxiliary Results

Lemma S.1.1. *Let $X^{(n)} = (X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables with distribution $P_\theta = Bernoulli(\theta)$. Denote by $J_n(x, P_\theta)$ the distribution of root $\sqrt{n}(\hat{\theta}_n - \theta)$ under P_θ, where $\hat{\theta}_n = \bar{X}_n$. Let $\hat{P}_n$ be the empirical distribution of $X^{(n)}$ or, equivalently, $P_{\hat{\theta}_n}$. Then, (11) holds for any $\epsilon > 0$ whenever ρ is a metric compatible with the weak topology.*

Proof: First, note for any $0 < \delta < 1$ and $\epsilon > 0$ that

$$\sup_{\delta<\theta<1-\delta} P_\theta \left\{ \sup_{x \in \mathbf{R}} |J_n(x, \hat{P}_n) - J_n(x, P_{\theta_n})| > \epsilon \right\} \to 0 \ . \tag{S.1}$$

To see this, it suffices to use the Berry-Esseen bound and Polya's Theorem, as in Example 11.2.2 of Lehmann and Romano (2005). Now suppose by way contradiction that (11) failed. It follows that there exists $\{\theta_n \in [0,1] : n \geq 1\}$ such that $\rho(J_n(\cdot, \hat{P}_n), J_n(\cdot, P_{\theta_n}))$ does not converge in probability to zero under P_{θ_n} and $\theta_n \to \theta^* \in [0,1]$. Since convergence with respect to the Kolmogorov metric implies weak convergence, it follows from (S.1) that we need only consider the case where $\theta^* = 0$ or $\theta^* = 1$. Suppose $\theta^* = 0$. By Chebychev's inequality, we have that $J_n(\cdot, P_{\theta_n}) \stackrel{d}{\to} \delta_0$ under P_{θ_n}, where δ_0 is the distribution that places mass one at zero. It therefore suffices to show that

$$\rho(J_n(\cdot, \hat{P}_n), \delta_0) \stackrel{P_{\theta_n}}{\to} 0 \ . \tag{S.2}$$

To establish (S.2), it suffices to consider the case where ρ is the bounded Lipschitz metric. For such ρ, we have

$$\rho(J_n(\cdot, \hat{P}_n), \delta_0) = \sup_\Psi \left| \int \Psi(x) dJ_n(x, \hat{P}_n) - \int \Psi(x) d\delta_0 \right| \ , \tag{S.3}$$

where the supremum is understood to be taken only over Lipschitz functions Ψ with Lipschitz constant equal to one. The lefthand-side of (S.3) may in turn be expressed as

$$\begin{aligned}
\sup_\Psi \left| \int \Psi(x) dJ_n(x, \hat{P}_n) - \Psi(0) \right| &\leq \sup_\Psi \int |\Psi(x) - \Psi(0)| dJ_n(x, \hat{P}_n) \\
&\leq \int |x| dJ_n(x, \hat{P}_n) \\
&\leq \left(\int x^2 dJ_n(x, \hat{P}_n) \right)^{1/2} \\
&= \sqrt{\hat{\theta}_n(1 - \hat{\theta}_n)} \\
&\stackrel{P_{\theta_n}}{\to} 0 \ ,
\end{aligned}$$

where the first inequality follows from Jensen's inequality, the second inequality follows from the fact that Ψ has Lipschitz constant equal to one, the third inequality follows from the Cauchy-Schwartz inequality, the equality follows from the definition of $J_n(x, \hat{P}_n)$, and the convergence in probability to zero follows from the fact that $\hat{\theta}_n \stackrel{P_{\theta_n}}{\to} 0$. A similar argument holds for $x < 0$. Finally, the same argument applies when $\theta^* = 1$. ■

S.2 Proof of Corollary 2.1

Note that

$$\tau_b(\hat{\theta}_b(X^{n,(b),i} - \hat{\theta}_n) = \tau_b(\hat{\theta}_b(X^{n,(b),i} - \theta(P)) - \tau_b(\hat{\theta}_n - \theta(P)) \ .$$

Therefore,

$$\begin{aligned} \hat{L}_n^{-1}(\alpha_1) &= L_n^{-1}(\alpha_1, P) - \tau_b(\hat{\theta}_n - \theta(P)) \\ \hat{L}_n^{-1}(1-\alpha_2) &= L_n^{-1}(1-\alpha_2, P) - \tau_b(\hat{\theta}_n - \theta(P)) \ . \end{aligned}$$

Hence,

$$\begin{aligned} & P\{\hat{L}_n^{-1}(\alpha_1) \le \tau_n(\hat{\theta}_n - \theta(P)) \le \hat{L}_n^{-1}(1-\alpha_2)\} \\ = \ & P\{L_n^{-1}(\alpha_1, P) \le (\tau_n + \tau_b)(\hat{\theta}_n - \theta(P)) \le L_n^{-1}(1-\alpha_2, P)\} \ , \end{aligned}$$

from which the desired result follows immediately. ■

S.3 Proof of Theorem 2.2

For $\epsilon > 0$ and $P \in \mathbf{P}$, define

$$\Delta_n(\epsilon, P) = P\left\{\sup_{x \in \mathbf{R}} |\hat{L}_n(x) - J_b(x,P)| > \epsilon\right\} \ .$$

Note that

$$\Delta_n(\epsilon, P) \le P\left\{\sup_{x \in \mathbf{R}} |\hat{L}_n(x) - L_n(x,P)| > \frac{\epsilon}{2}\right\} + P\left\{\sup_{x \in \mathbf{R}} |L_n(x,P) - J_b(x,P)| > \frac{\epsilon}{2}\right\} \ .$$

Hence, for any $\epsilon > 0$, it follows from Lemma 4.2 and (7) that

$$\sup_{P \in \mathbf{P}} \Delta_n(\epsilon, P) \to 0 \ .$$

The desired result thus follows by arguing exactly as in the proof of Theorem 2.1, but with the righthand-side of (45) replaced with $\Delta_n(\epsilon, P)$ throughout the argument. ■

S.4 Proof of Corollary 2.2

By Theorem 2.2, it suffices to show that (7) holds. Consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$. For any $\eta > 0$, note that

$$\begin{aligned}
& \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x, P_n)\} \\
\leq & \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x+\eta, P_n)\} + \sup_{x \in \mathbf{R}} \{L_n(x+\eta, P_n) - L_n(x, P_n)\} \\
\leq & \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x+\eta, P_n)\} + \sup_{x \in \mathbf{R}} \{L_n(x+\eta, P_n) - J_b(x+\eta, P_n)\} \\
& + \sup_{x \in \mathbf{R}} \{J_b(x, P_n) - L_n(x, P_n)\} + \sup_{x \in \mathbf{R}} \{J_b(x+\eta, P_n) - J_b(x, P_n)\} \ .
\end{aligned}$$

The second and third terms after the final inequality above tends to zero in probability under P_n by Lemma 4.2. By (i), $\{J_b(x, P) : b \geq 1, P \in \mathbf{P}\}$ is tight and any subsequential limiting distribution is continuous, so the last term tends to zero as $\eta \to 0$. Next, we argue that

$$\sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x+\eta, P_n)\} \leq o_{P_n}(1)$$

for any $\eta > 0$. To this end, abbreviate $\hat{\theta}_{n,b,i} = \hat{\theta}_b(X^{n,(b),i})$ and $\hat{\sigma}_{n,b,i} = \hat{\sigma}_b(X^{n,(b),i})$. First, we show that, for any $\eta > 0$,

$$\frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\left|\frac{\hat{\sigma}_{n,b,i}}{\sigma(P_n)} - 1\right| > \eta\right\} \stackrel{P_n}{\to} 0 \ . \tag{S.4}$$

Indeed, the expectation of each term in the average is

$$P_n\left\{\left|\frac{\hat{\sigma}_{n,b,i}}{\sigma(P_n)} - 1\right| > \eta\right\} \ ,$$

which tends to zero by condition (ii). The conclusion (S.4) thus follows from Lemma 4.2. Similarly, using condition (i) and the requirement that $\tau_b/\tau_n \to 0$, it follows that for any $\eta > 0$

$$\frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\frac{\tau_b|\hat{\theta}_n - \theta(P_n)|}{\sigma(P_n)} > \eta\right\} \stackrel{P_n}{\to} 0 \ . \tag{S.5}$$

It follows from (S.4) and (S.5) that for any $\eta > 0$

$$\frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\frac{\tau_b|\hat{\theta}_n - \theta(P_n)|}{\hat{\sigma}_{n,b,i}} > \eta\right\} \stackrel{P_n}{\to} 0 \ . \tag{S.6}$$

Note that

$$\begin{aligned}
\hat{L}_n(x) &= \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\frac{\tau_b(\hat{\theta}_{n,b,i} - \theta(P_n))}{\hat{\sigma}_{n,b,i}} \leq x + \frac{\tau_b(\hat{\theta}_n - \theta(P_n))}{\hat{\sigma}_{n,b,i}}\right\} \\
&\leq L_n(x+\eta, P_n) + \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\frac{\tau_b|\hat{\theta}_n - \theta(P_n)|}{\hat{\sigma}_{n,b,i}} > \eta\right\} \ .
\end{aligned}$$

From (S.6), we see that the last average tends in probability to zero under P_n. Moreover, it does not depend on x, so the desired conclusion follows. A similar argument establishes that

$$\sup_{x \in \mathbf{R}} \{L_n(x, P_n) - \hat{L}_n(x)\} \leq o_{P_n}(1) ,$$

from which the desired result follows. ∎

S.5 Proof of Theorem 2.3

Lemma S.5.1. *Let $\{G_n : n \geq 1\}$ and $\{F_n : n \geq 1\}$ be sequences of c.d.f.s on $\mathbf{R}$. Suppose $X_n \sim F_n$. Then, the following statements are true:*

(i) If $\limsup_{n\to\infty} \sup_{x \in \mathbf{R}} \{G_n(x) - F_n(x)\} > \epsilon$ for some $\epsilon > 0$, then there exists $0 \leq \alpha_2 < 1$ and $\delta > \epsilon/2$ such that

$$\liminf_{n\to\infty} P\{X_n \leq G_n^{-1}(1-\alpha_2)\} \leq 1 - (\alpha_2 + \delta) .$$

(ii) If $\limsup_{n\to\infty} \sup_{x \in \mathbf{R}} \{F_n(x) - G_n(x)\} > \epsilon$ for some $\epsilon > 0$, then there exists $0 \leq \alpha_1 < 1$ and $\delta > \epsilon/2$ such that

$$\liminf_{n\to\infty} P\{X_n \geq G_n^{-1}(\alpha_1)\} \leq 1 - (\alpha_1 + \delta) .$$

(iii) If $\limsup_{n\to\infty} \sup_{x \in \mathbf{R}} |G_n(x) - F_n(x)| > \epsilon$ for some $\epsilon > 0$, then there exists $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ with $0 \leq \alpha_1 + \alpha_2 < 1$ and $\delta > \epsilon/2$ such that

$$\liminf_{n\to\infty} P\{G_n^{-1}(\alpha_1) \leq X_n \leq G_n^{-1}(1-\alpha_2)\} \leq 1 - (\alpha_1 + \alpha_2 + \delta) .$$

Proof: We prove only (i). Analogous arguments establish (ii) and (iii). Choose a subsequence n_k and x_{n_k} such that $G_{n_k}(x_{n_k}) > F_{n_k}(x_{n_k}) + \epsilon > F_{n_k}(x_{n_k})$. By considering a further subsequence if necessary, choose $0 \leq \alpha_2 < 1$ and $\delta > \epsilon/2$ such that $G_{n_k}(x_{n_k}) > 1 - \alpha_2 > 1 - \alpha_2 - \delta > F_{n_k}(x_{n_k})$. To see that this is possible, consider the intervals $I_{n_k} = [F_{n_k}(x_{n_k}), G_{n_k}(x_{n_k})] \subseteq [0,1]$ and choose a subsequence along which the endpoints of I_{n_k} converge. The desired conclusion follows because each I_{n_k} has length at least $\epsilon > 0$. Next, note that by right-continuity of F_{n_k}, we may choose $x'_{n_k} > x_{n_k}$ such that $F_{n_k}(x'_{n_k}) < 1 - \alpha_2 - \delta$. Thus, $F_{n_k}^{-1}(1 - \alpha_2 - \delta) \geq x'_{n_k} > x_{n_k}$. Hence, $G_{n_k}^{-1}(1-\alpha_2) = F_{n_k}^{-1}(1-\alpha_2-\delta) - \eta_{n_k}$ for some $\eta_{n_k} > 0$. It follows that $P\{X_{n_k} \leq G_{n_k}^{-1}(1-\alpha_2)\} = P\{X_{n_k} \leq F_{n_k}^{-1}(1-\alpha_2-\delta) - \eta_{n_k}\} < 1-(\alpha_2+\delta)$, where the final inequality follows from the definition of $F_{n_k}^{-1}(1-\alpha_2-\delta)$. The desired result thus follows. ∎

Lemma S.5.2. *Let $\{G_n : n \geq 1\}$ and $\{F_n : n \geq 1\}$ be sequences of c.d.f.s on $\mathbf{R}$. Let $\{\hat{G}_n : n \geq 1\}$ be a (random) sequence of c.d.f.s on $\mathbf{R}$ such that for all $\eta > 0$ we have that $P\{\sup_{x\in\mathbf{R}} |\hat{G}_n(x) - G_n(x)| > \eta\} \to 0$. Suppose $X_n \sim F_n$. Then, the following statements are true:*

(i) If $\limsup_{n\to\infty} \sup_{x\in\mathbf{R}}\{G_n(x) - F_n(x)\} > \epsilon$ for some $\epsilon > 0$, then there exists $0 \leq \alpha_2 < 1$ such that

$$\liminf_{n\to\infty} P\{X_n \leq \hat{G}_n^{-1}(1-\alpha_2)\} < 1 - \alpha_2 \ .$$

(ii) If $\limsup_{n\to\infty} \sup_{x\in\mathbf{R}}\{F_n(x) - G_n(x)\} > \epsilon$ for some $\epsilon > 0$, then there exists $0 \leq \alpha_1 < 1$ such that

$$\liminf_{n\to\infty} P\{X_n \geq \hat{G}_n^{-1}(\alpha_1)\} < 1 - \alpha_1 \ .$$

(iii) If $\limsup_{n\to\infty} \sup_{x\in\mathbf{R}} |G_n(x) - F_n(x)| > \epsilon$ for some $\epsilon > 0$, then there exists $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ with $0 \leq \alpha_1 + \alpha_2 < 1$ such that

$$\liminf_{n\to\infty} P\{\hat{G}_n^{-1}(\alpha_1) \leq X_n \leq \hat{G}_n^{-1}(1-\alpha_2)\} < 1 - \alpha_1 - \alpha_2 \ .$$

Proof: We prove only (i). Analogous arguments establish (ii) and (iii). Let $E_n = \{\sup_{x\in\mathbf{R}} |\hat{G}_n(x) - G(x)| \leq \eta\}$ for some $0 < \eta < \epsilon/2$. By part (i) of Lemma S.5.1, choose $0 \leq \alpha_2 < 1$ and $\delta > \epsilon/2$ so that $\liminf_{n\to\infty} \sup_{x\in\mathbf{R}} P\{X_n \leq G_n^{-1}(1-\alpha_2+\eta)\} \leq 1 - \alpha_2 + \eta - \delta$. Note that by part (i) of Lemma 4.1, E_n implies that $\hat{G}_n^{-1}(1-\alpha_2) \leq G_n^{-1}(1-\alpha_2+\eta)$. Hence, $P\{X_n \leq \hat{G}_n^{-1}(1-\alpha_2)\} = P\{X_n \leq \hat{G}_n^{-1}(1-\alpha_2) \cap E_n\} + P\{X_n \leq \hat{G}_n^{-1}(1-\alpha_2) \cap E_n^c\} \leq P\{X_n \leq G_n^{-1}(1-\alpha_2+\eta)\} + P\{E_n^c\}$. The desired conclusion now follows from the fact that $P\{E_n^c\} \to 0$. ■

Proof of Theorem 2.3: We prove only (i). Analogous arguments establish (ii) and (iii). Choose $\{P_n \in \mathbf{P} : n \geq 1\}$ and $\epsilon > 0$ such that $\limsup_{n\to\infty} \sup_{x\in\mathbf{R}}\{J_b(x, P_n) - J_n(x, P_n)\} > \epsilon$. We apply part (i) of Lemma S.5.2 with $\hat{G}_n(x) = L_n(x, P_n)$, $F_n(x) = J_n(x, P_n)$ and $G_n(x) = J_b(x, P_n)$. The desired conclusion therefore follows provided that for any $\eta > 0$ we have that

$$P_n\left\{\sup_{x\in\mathbf{R}} |L_n(x, P_n) - J_b(x, P_n)| > \eta\right\} \to 0 \ , \tag{S.7}$$

which is ensured by Lemma 4.2. Note further that (7) and Lemma 4.2 show that (S.7) holds with $\hat{L}_n(x)$ in place of $L_n(x, P_n)$. Thus, the same argument with $\hat{L}_n(x)$ in place of $L_n(x, P_n)$ shows that (i) holds with $\hat{L}_n^{-1}(\cdot)$ in place of $L_n^{-1}(x, P)$. ■

S.6 Proof of Theorem 3.1

Lemma S.6.1. *Consider any sequence* $\{P_n \in \tilde{\mathbf{P}} : n \geq 1\}$, *where* $\tilde{\mathbf{P}}$ *satisfies* (12). *Let* $X_{n,i}, i = 1, \ldots, n$ *be an i.i.d. sequence of real-valued random variables with distribution* P_n. *Then,*

$$\frac{S_n^2}{\sigma^2(P_n)} \xrightarrow{P_n} 1 \ .$$

PROOF: First assume without loss of generality that $\mu(P_n) = 0$ and $\sigma(P_n) = 1$ for all $n \geq 1$. Next, note that

$$S_n^2 = \frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i}^2 - \bar{X}_n^2 \ .$$

By Lemma 11.4.3 of Lehmann and Romano (2005), we have that

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i}^2 \xrightarrow{P_n} 1 \ .$$

By Lemma 11.4.2 of Lehmann and Romano (2005), we have further that

$$\bar{X}_n \xrightarrow{P_n} 0 \ .$$

The desired result therefore follows from the continuous mapping theorem. ■

PROOF OF THEOREM 3.1: We argue that

$$\sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} |J_b(x, P) - J_n(x, P)| \to 0 \ . \tag{S.8}$$

Suppose by way of contradiction that (S.8) fails. It follows that there exists a subsequence n_ℓ such that $\Omega(P_{n_\ell}) \to \Omega^*$ and either

$$\sup_{x \in \mathbf{R}} |J_{n_\ell}(x, P_{n_\ell}) - \Phi_{\Omega^*}(x, \ldots, x)| \not\to 0 \tag{S.9}$$

or

$$\sup_{x \in \mathbf{R}} |J_{b_{n_\ell}}(x, P_{n_\ell}) - \Phi_{\Omega^*}(x, \ldots, x)| \not\to 0 \ . \tag{S.10}$$

To see that neither (S.9) nor (S.10) can hold, let

$$K_n(x, P) = P\left\{ \frac{\sqrt{n}(\bar{X}_{1,n} - \mu_1(P))}{S_{1,n}} \leq x_1, \ldots, \frac{\sqrt{n}(\bar{X}_{k,n} - \mu_k(P))}{S_{k,n}} \leq x_k \right\} \tag{S.11}$$

$$\tilde{K}_n(x, P) = P\left\{ \frac{\sqrt{n}(\bar{X}_{1,n} - \mu_1(P))}{\sigma_1(P)} \leq x_1, \ldots, \frac{\sqrt{n}(\bar{X}_{k,n} - \mu_k(P))}{\sigma_k(P)} \leq x_k \right\} \ . \tag{S.12}$$

Since

$$\Phi_{\Omega(P_{n_\ell})}(\cdot) \xrightarrow{d} \Phi_{\Omega^*}(\cdot) \ ,$$

it follows from the uniform central limit theorem established by Lemma 3.3.1 of Romano and Shaikh (2008) that $\tilde{K}_{n_\ell}(\cdot, P_{n_\ell}) \stackrel{d}{\to} \Phi_{\Omega^*}(\cdot)$. From Lemma S.6.1, we have for $1 \leq j \leq k$ that

$$\frac{S_{j,n_\ell}}{\sigma_j(P_{n_\ell})} \stackrel{P_{n_\ell}}{\to} 1 .$$

Hence, by Slutsky's Theorem, $K_{n_\ell}(\cdot, P_{n_\ell}) \stackrel{d}{\to} \Phi_{\Omega^*}(\cdot)$. By Polya's Theorem, we thus see that (S.9) can not hold. A similar argument establishes that (S.10) can not hold. Hence, (S.8) holds. For any fixed $P \in \mathbf{P}$, we also have that $J_n(x, P)$ tends in distribution to a continuous limiting distribution. The desired conclusion (14) therefore follows from Theorem 2.1 and Remark 2.1.

To show that (14) holds when $L_n^{-1}(\cdot, P)$ is replaced by $\hat{L}_n^{-1}(\cdot)$, it suffices by Theorem 2.2 to show that (7) holds. Consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$. For any $\eta > 0$, note that

$$\begin{aligned}
& \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x, P_n)\} \\
\leq \; & \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x + \eta, P_n)\} + \sup_{x \in \mathbf{R}} \{L_n(x + \eta, P_n) - L_n(x, P_n)\} \\
\leq \; & \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x + \eta, P_n)\} + \sup_{x \in \mathbf{R}} \{L_n(x + \eta, P_n) - J_b(x + \eta, P_n)\} \\
& + \sup_{x \in \mathbf{R}} \{J_b(x, P_n) - L_n(x, P_n)\} + \sup_{x \in \mathbf{R}} \{J_b(x + \eta, P_n) - J_b(x, P_n)\} .
\end{aligned}$$

The second and third terms after the final inequality above tends to zero in probability under P_n by Lemma 4.2. Since $\{J_b(x, P) : b \geq 1, P \in \mathbf{P}\}$ is tight and any subsequential limiting distribution is continuous, the last term tends to zero as $\eta \to 0$. Next, we argue that

$$\sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x + \eta, P_n)\} \leq o_{P_n}(1)$$

for any $\eta > 0$. To this end, for $1 \leq j \leq k$, let $\bar{X}_{j,n,b,i}$ be $\bar{X}_{j,b}$ evaluated at $X^{n,(b),i}$ and define $S^2_{j,n,b,i}$ analogously. Arguing as in the proof of Corollary 2.2, it is possible to show that

$$\frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\max_{1 \leq j \leq k} \frac{\sqrt{b}(\bar{X}_{n,j} - \mu_j(P_n))}{S_{n,b,i,j}} > \eta\right\} \stackrel{P_n}{\to} 0 . \tag{S.13}$$

Note that

$$\begin{aligned}
\hat{L}_n(x) &= \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\max_{1 \leq j \leq k} \frac{\sqrt{b}(\bar{X}_{n,b,i,j} - \bar{X}_{n,j})}{S_{n,b,i,j}} \leq x\right\} \\
&\leq \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\max_{1 \leq j \leq k} \frac{\sqrt{b}(\bar{X}_{n,b,i,j} - \mu_j(P_n))}{S_{n,b,i,j}} \leq x + \max_{1 \leq j \leq k} \frac{\sqrt{b}(\bar{X}_{n,j} - \mu_j(P_n))}{S_{n,b,i,j}}\right\} \\
&\leq L_n(x + \eta, P_n) + \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\max_{1 \leq j \leq k} \frac{\sqrt{b}(\bar{X}_{n,j} - \mu_j(P_n))}{S_{n,b,i,j}} > \eta\right\} .
\end{aligned}$$

From (S.13), we see that the last average tends in probability to zero under P_n. Moreover, it does not depend on x, so the desired conclusion follows. A similar argument establishes that

$$\sup_{x \in \mathbf{R}} \{L_n(x, P_n) - \hat{L}_n(x)\} \le o_{P_n}(1) ,$$

from which the desired result follows. ■

S.7 Proof of Theorem 3.2

Lemma S.7.1. *Let* $\mathbf{P}$ *be defined as in Theorem 3.1. Consider any sequence* $\{P_n \in \mathbf{P} : n \ge 1\}$. *Let* $X_{n,i}, i = 1, \dots, n$ *be an i.i.d. sequence of random variables with distribution* P_n *and denote by* $\hat{P}_n$ *the empirical distribution of* $X_{n,i}, i = 1, \dots, n$. *Then,*

$$||\Omega(\hat{P}_n) - \Omega(P_n)|| \stackrel{P_n}{\to} 0 ,$$

where the norm $|| \cdot ||$ *is the component-wise maximum of the absolute value of all elements.*

Proof: First assume without loss of generality that $\mu_j(P_n) = 0$ and $\sigma_j(P_n) = 1$ for all $1 \le j \le k$ and $n \ge 1$. Next, note that we may write the (j, ℓ) element of $\Omega(\hat{P}_n)$ as

$$\frac{1}{S_{j,n}} \frac{1}{S_{\ell,n}} \left(\frac{1}{n} \sum_{1 \le i \le n} X_{n,i,j} X_{n,i,\ell} - \bar{X}_{j,n} \bar{X}_{\ell,n} \right) . \tag{S.14}$$

From Lemma S.6.1, we have that

$$S_{j,n} S_{\ell,n} \stackrel{P_n}{\to} 1 . \tag{S.15}$$

From Lemma 11.4.2 of Lehmann and Romano (2005), we have that

$$\bar{X}_{j,n} \bar{X}_{\ell,n} = o_{P_n}(1) . \tag{S.16}$$

Let $Z_{n,i} = X_{n,i,j} X_{n,i,\ell}$. From the inequality

$$|a||b| I\{|a||b| > \lambda\} \le |a|^2 I\{|a| > \sqrt{\lambda}\} + |b|^2 I\{|b| > \sqrt{\lambda}\} ,$$

we see that

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} E_{P_n}[|Z_{n,i}| I\{|Z_{n,i}| > \lambda\}] = 0 .$$

Moreover, since $|E_{P_n}[Z_{n,i}]| \le 1$, we have further that

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} E_{P_n}[|Z_{n,i} - E_{P_n}[Z_{n,i}]| I\{|Z_{n,i} - E_{P_n}[Z_{n,i}]| > \lambda\}] = 0 .$$

Hence, by Lemma 11.4.2 of Lehmann and Romano (2005), we have that

$$\frac{1}{n}\sum_{1\le i\le n} X_{n,i,j}X_{n,i,\ell} = E_{P_n}[X_{n,i,j}X_{n,i,\ell}] + o_{P_n}(1) \ . \tag{S.17}$$

The desired result follows from (S.14) - (S.17) and the observation that $|E_{P_n}[X_{n,i,j}X_{n,i,\ell}]| \le 1$. ■

Proof of Theorem 3.2: The proof closely follows the one given for Theorem 3.1, so we provide only a sketch.

To prove (i), we again argue that (S.8) holds. To this end, suppose by way of contradiction that (S.8) fails. It follows that there exists a sequence $\{P_n \in \mathbf{P} : n \ge 1\}$ along which

$$\sup_{x\in\mathbf{R}} |J_b(x,P_n) - J_n(x,P_n)| \not\to 0 \ . \tag{S.18}$$

By considering a further subsequence if necessary, we may assume without loss of generality that $Z_n(P_n) \stackrel{d}{\to} Z^*$ under P_n and $\Omega(P_n) \to \Omega^*$ with $Z^* \sim N(0,\Omega^*)$. Lemma S.7.1 thus implies that $\hat\Omega_n \stackrel{P_n}{\to} \Omega^*$. By assumption, we therefore have that (17) and (18) hold for all $x \in \mathbf{R}$. This convergence is therefore uniform in $x \in \mathbf{R}$. It therefore follows from the triangle inequality that (S.18) can not hold, establishing the claim.

To prove (ii), we again show that (7) holds and apply Theorem 2.2. Consider any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$. For any $\eta > 0$, note that

$$\begin{aligned}
& \sup_{x\in\mathbf{R}}\{\hat L_n(x) - L_n(x,P_n)\} \\
\le\ & \sup_{x\in\mathbf{R}}\{\hat L_n(x) - L_n(x+\eta,P_n)\} + \sup_{x\in\mathbf{R}}\{L_n(x+\eta,P_n) - L_n(x,P_n)\} \\
\le\ & \sup_{x\in\mathbf{R}}\{\hat L_n(x) - L_n(x+\eta,P_n)\} + \sup_{x\in\mathbf{R}}\{L_n(x+\eta,P_n) - J_b(x+\eta,P_n)\} \\
& + \sup_{x\in\mathbf{R}}\{J_b(x,P_n) - L_n(x,P_n)\} + \sup_{x\in\mathbf{R}}\{J_b(x+\eta,P_n) - J_b(x,P_n)\} \ .
\end{aligned}$$

For any $\eta > 0$, the second and third terms after the final inequality above tend in probability to zero under P_n by Lemma 4.2. Since $\{J_b(x,P) : b \ge 1, P \in \mathbf{P}\}$ is tight and any subsequential limiting distribution is continuous, we see that the last term tends to zero as $\eta \to 0$. Next, we argue that

$$\sup_{x\in\mathbf{R}}\{\hat L_n(x) - L_n(x+\eta,P_n)\} \le o_{P_n}(1)$$

for any $\eta > 0$. To this end, let $Z_{n,b,i}$ equal $Z_b(P_n)$ evaluated at $X^{n,(b),i}$ and let $Z'_{n,b,i}$ equal $Z_{n,b,i}$ except with $\mu(P_n)$ replaced by $\bar X_n$. Similarly, let $\Omega_{n,b,i}$ equal $\hat\Omega_b$ evaluated at $X^{n,(b),i}$. In this notation, $L_n(\cdot,P_n)$ is the empirical c.d.f. of the values $f(Z_{n,b,i},\Omega_{n,b,i})$ and $\hat L_n(\cdot)$ is the empirical c.d.f. of the values $f(Z'_{n,b,i},\Omega_{n,b,i})$. From Lemma 3.3.1 of Romano and Shaikh (2008), we see that the distributions of both $Z_{n,b,i}$ and $Z'_{n,b,i}$ under P_n are tight. Hence, there exists a compact set K

such that $P_n\{Z_{n,b,i} \notin K\} < \epsilon/3$ and $P_n\{Z'_{n,b,i} \notin K\} < \epsilon/3$. Moreover, with the first argument of f restricted to K, f is uniformly continuous (since the second argument already lies on a compact set as correlations are bounded in absolute value by one). It follows that for any $\eta > 0$ there exists $\delta > 0$ such that $|f(z, \omega) - f(z', \omega)| < \eta$ if $|z - z'| < \delta$ and z and z' both lie in K. Hence,

$$\hat{L}_n(x) \leq L_n(x + \eta, P_n) + \frac{1}{N_n} \sum_{1 \leq i \leq N_n} \left(I\{|Z_{n,b,i} - Z'_{n,b,i}| > \delta\} + I\{Z_{n,b,i} \notin K\} + I\{Z'_{n,b,i} \notin K\} \right) .$$

Aruging, for example, as in the proof of Lemma 4.2, we see that the final term above equals

$$P_n\{|Z_{n,b,i} - Z'_{n,b,i}| > \delta\} + P_n\{Z_{n,b,i} \notin K\} + P_n\{Z'_{n,b,i} \notin K\} + o_{P_n}(1) .$$

Since

$$|Z_{n,b,i} - Z'_{n,b,i}| \stackrel{P_n}{\to} 0 ,$$

we see that

$$\hat{L}_n(x) \leq L_n(x + \eta, P_n) + \epsilon$$

with probability tending to one under P_n. As ϵ does not depend on x, the desired conclusion follows by letting $\epsilon \to 0$. A similar argument establishes that

$$\sup_{x \in \mathbf{R}} \{L_n(x + \eta, P_n) - \hat{L}_n(x)\} \leq o_{P_n}(1) ,$$

from which the desired result follows. Finally, it follows from Remark 2.1 that we may replace $\liminf_{n\to\infty}$ and $\geq$ by $\lim_{n\to\infty}$ and $=$, respectively.

S.8 Proof of Theorem 3.3

We argue that

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_0} \sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \leq 0 .$$

Note that

$$J_n(x, P) = P\left\{ \max_{1 \leq j \leq k} \left\{ Z_{j,n}(P) + \frac{\sigma_j(P)}{S_{j,n}} \frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)} \right\} \leq x \right\} ,$$

where

$$Z_{j,n}(P) = \frac{\sqrt{n}(\bar{X}_{n,j} - \mu_j(P))}{S_{j,n}} .$$

For $\delta > 0$, define

$$E_n(\delta, P) = \left\{ \max_{1 \leq j \leq k} \left| \frac{\sigma_j(P)}{S_{j,b}} - 1 \right| < \delta \right\} .$$

Note that

$$\begin{aligned} J_b(x,P) &\le P\left\{\max_{1\le j\le k}\left\{Z_{j,b}(P)+\frac{\sigma_j(P)}{S_{j,b}}\frac{\sqrt{b}\mu_j(P)}{\sigma_j(P)}\right\}\le x\cap E_n(\delta,P)\right\}+P\{E_n(\delta,P)^c\} \\ &\le P\left\{\max_{1\le j\le k}\left\{Z_{j,b}(P)+(1+\delta)\frac{\sqrt{b}\mu_j(P)}{\sigma_j(P)}\right\}\le x\right\}+P\{E_n(\delta,P)^c\} \\ &= P\left\{\max_{1\le j\le k}\left\{Z_{j,b}(P)+(1-\delta)\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}+\Delta_{j,n}(P)\right\}\le x\right\}+P\{E_n(\delta,P)^c\}\ , \end{aligned}$$

where

$$\Delta_{j,n}(P)=\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\left((1+\delta)\frac{\sqrt{b}}{\sqrt{n}}-(1-\delta)\right)\ .$$

Note that for all n sufficiently large, $\Delta_{j,n}(P)\ge 0$ for all $1\le j\le k$. Hence, for all such n, we have that

$$J_b(x,P)\le P\left\{\max_{1\le j\le k}\left\{Z_{j,b}(P)+(1-\delta)\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\right\}\le x\right\}+P\{E_n(\delta,P)^c\}\ .$$

We also have that

$$\begin{aligned} J_n(x,P) &\ge P\left\{\max_{1\le j\le k}\left\{Z_{j,n}(P)+\frac{\sigma_j(P)}{S_{j,n}}\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\right\}\le x\cap E_n(\delta,P)\right\} \\ &\ge P\left\{\max_{1\le j\le k}\left\{Z_{j,n}(P)+(1-\delta)\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\right\}\le x\cap E_n(\delta,P)\right\} \\ &\ge P\left\{\max_{1\le j\le k}\left\{Z_{j,n}(P)+(1-\delta)\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\right\}\le x\right\}-P\{E_n(\delta,P)^c\}\ . \end{aligned}$$

Therefore,

$$\begin{aligned} J_b(x,P)-J_n(x,P) &\le P\left\{\max_{1\le j\le k}\left\{Z_{j,b}(P)+(1-\delta)\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\right\}\le x\right\} \\ &\quad -P\left\{\max_{1\le j\le k}\left\{Z_{j,n}(P)+(1-\delta)\frac{\sqrt{n}\mu_j(P)}{\sigma_j(P)}\right\}\le x\right\}+2P\{E_n(\delta,P)^c\}\ . \end{aligned}$$

Note that Lemma S.6.1 implies that

$$\sup_{P\in\mathbf{P}_0}P\{E_n(\delta,P)^c\}\to 0\ .$$

It therefore suffices to show that for any sequence $\{P_n\in\mathbf{P}_0 : n\ge 1\}$ that

$$\begin{aligned} &P_n\left\{\max_{1\le j\le k}\left\{Z_{j,b}(P_n)+(1-\delta)\frac{\sqrt{n}\mu_j(P_n)}{\sigma_j(P_n)}\right\}\le x\right\} \\ &\qquad -P_n\left\{\max_{1\le j\le k}\left\{Z_{j,n}(P_n)+(1-\delta)\frac{\sqrt{n}\mu_j(P_n)}{\sigma_j(P_n)}\right\}\le x\right\} \end{aligned}$$

tends to zero uniformly in x. But this follows by simply absorbing the terms $(1-\delta)\frac{\sqrt{n}\mu_j(P_n)}{\sigma_j(P_n)}$ into the x and arguing as in the proof of Theorem 3.1. ■

S.9 Proof of Theorem 3.4

By arguing as in Romano and Wolf (2005), we see that

$$FWER_P \leq P\left\{\max_{j \in K_0(P)} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} > L_n^{-1}(1-\alpha, K_0(P))\right\} ,$$

where

$$K_0(P) = \{1 \leq j \leq k : \mu_j(P) \leq 0\} .$$

The desired conclusion now follows immediately from Theorem 3.3. ■

S.10 Proof of Theorem 3.5

We begin with some preliminaries. First note that

$$\begin{aligned} J_n(x, P) &= P\left\{\sup_{t \in \mathbf{R}} |B_n(P\{(-\infty, t]\})| \leq x\right\} \\ &= P\left\{\sup_{t \in R(P)} |B_n(t)| \leq x\right\} , \end{aligned}$$

where B_n is the uniform empirical process and

$$R(P) = \text{cl}(\{P\{(-\infty, t]\} : t \in \mathbf{R}\}) . \tag{S.19}$$

By Theorem 3.85 of Aliprantis and Border (2006), the set of all nonempty closed subsets of $[0,1]$ is a compact metric space with respect to the Hausdorff metric

$$d_H(U, V) = \inf\{\eta > 0 : U \subseteq V^\eta, V \subseteq U^\eta\} . \tag{S.20}$$

Here,

$$U^\eta = \bigcup_{u \in U} A_\eta(u) ,$$

where $A_\eta(u)$ is the open ball with center u and radius η. Thus, for any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$, there is a subsequence n_ℓ and a closed set $R \subseteq [0,1]$ along which

$$d_H(R(P_{n_\ell}), R) \to 0 . \tag{S.21}$$

Finally, denote by B the standard Brownian bridge process. By the almost sure representation theorem, we may choose B_n and B so that

$$\sup_{0 \leq t \leq 1} |B_n(t) - B(t)| \to 0 \text{ a.s.} \tag{S.22}$$

We now argue that

$$\sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} |J_b(x,P) - J_n(x,P)| \to 0 \ . \tag{S.23}$$

Suppose by way of contradiction that (S.23) fails. It follows that there exists a subsequence n_ℓ and a closed subset $R \subseteq [0,1]$ such that $d_H(R(P_{n_\ell}), R) \to 0$ and either

$$\sup_{x \in \mathbf{R}} |J_{n_\ell}(x, P_{n_\ell}) - J^*(x)| \not\to 0 \tag{S.24}$$

or

$$\sup_{x \in \mathbf{R}} |J_{b_{n_\ell}}(x, P_{n_\ell}) - J^*(x)| \not\to 0 \ , \tag{S.25}$$

where

$$J^*(x) = P\left\{\sup_{t \in R} |B(t)| \leq x\right\} \ .$$

Moreover, by the definition of $\mathbf{P}$, it must be the case that R contains some point different from zero and one. To see that neither (S.24) or (S.25) can hold, note that

$$\left| \sup_{t \in R(P_{n_\ell})} |B_n(t)| - \sup_{t \in R} |B(t)| \right| \leq \sup_{t \in R(P_{n_\ell})} |B_n(t) - B(t)| + \left| \sup_{t \in R(P_{n_\ell})} |B(t)| - \sup_{t \in R} |B(t)| \right| \ . \tag{S.26}$$

By (S.22), we see that the first term on the righthand-side of (S.26) tends to zero a.s. By the a.s. uniform continuity of $B(t)$ and (S.21), we see that the second term on the righthand-side of (S.26) tends to zero a.s. Thus,

$$\sup_{t \in R(P_{n_\ell})} |B_n(t)| \stackrel{d}{\to} \sup_{t \in R} |B(t)| \ .$$

Since R contains some point different from zero and one, we see from Theorem 11.1 Davydov et al. (1998) that $\sup_{t \in R} B(t)$ is continuously distributed. By Polya's Theorem, we therefore have that (S.24) holds. A similar argument establishes that (S.25) can not hold. Hence, (S.23) holds. For any fixed $P \in \mathbf{P}$, we also have that $J_n(x,P)$ tends in distribution to a continuous limiting distribution. The desired conclusion (28) now follows from Theorem 2.1 and Remark 2.1.

To show the same result for the feasible estimator $\hat{L}_n$, we apply Theorem 2.2. To do this, let P_n be any sequence of distributions, and denote by F_n its corresponding c.d.f. Also, let $\hat{F}_n$ be the empirical c.d.f. of $X^{(n)}$, and let $\hat{F}_{n,b,i}$ denote the empirical c.d.f. of $X^{n,(b),i}$. For any $\eta > 0$, note that

$$\begin{aligned}
& \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x, P_n)\} \\
\leq \ & \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x + \eta, P_n)\} + \sup_{x \in \mathbf{R}} \{L_n(x + \eta, P_n) - L_n(x, P_n)\} \\
\leq \ & \sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x + \eta, P_n)\} + \sup_{x \in \mathbf{R}} \{L_n(x + \eta, P_n) - J_b(x + \eta, P_n)\} \\
& + \sup_{x \in \mathbf{R}} \{J_b(x, P_n) - L_n(x, P_n)\} + \sup_{x \in \mathbf{R}} \{J_b(x + \eta, P_n) - J_b(x, P_n)\} \ .
\end{aligned}$$

For any $\eta > 0$, the second and third terms after the final inequality above tend in probability to zero under P_n by Lemma 4.2. Arguing as above, we see that $\{J_b(x, P) : b \geq 1, P \in \mathbf{P}\}$ is tight and any subsequential limiting distribution is continuous. Hence, the last term tends to zero as $\eta \to 0$. Next, we argue that

$$\sup_{x \in \mathbf{R}} \{\hat{L}_n(x) - L_n(x + \eta, P_n)\} \leq o_{P_n}(1)$$

for any $\eta > 0$. To this end, note that for any $\eta > 0$ we have by the triangle inequality that

$$\begin{aligned} \hat{L}_n(x) &= \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\sqrt{b} \sup_{t \in \mathbf{R}} |\hat{F}_{n,b,i,}(t) - \hat{F}_n(t)| \leq x\right\} \\ &\leq \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\left\{\sqrt{b} \sup_{t \in \mathbf{R}} |\hat{F}_{n,b,i,}(t) - F_n(t)| \leq x + \eta\right\} + I\left\{\sqrt{b} \sup_{t \in \mathbf{R}} |\hat{F}_n(t) - F_n(t)| > \eta\right\} \\ &= L_n(x + \eta, P_n) + I\left\{\sqrt{b} \sup_{t \in \mathbf{R}} |\hat{F}_n(t) - F_n(t)| > \eta\right\} . \end{aligned}$$

The second term is independent of x and, by the Dvoretsky-Kiefer-Wolfowitz inequality, tends to 0 in probability under any P_n, from which the desired conclusion follows. A similar argument establishes that

$$\sup_{x \in \mathbf{R}} \{L_n(x, P_n) - \hat{L}_n(x)\} \leq o_{P_n}(1) ,$$

and the result follows. ■

S.11 Proof of Theorem 3.6

Lemma S.11.1. *Let h be a symmetric kernel of degree m. Denote by $J_n(x, P)$ the distribution of $R_n(X^{(n)}, P)$ defined in* (30)*. Suppose* $\mathbf{P}$ *satisfies* (33) *and* (34)*. Then,*

$$\lim_{n \to \infty} \sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} |J_n(x, P) - \Phi(x/\sigma(P))| = 0 .$$

Proof: Let $\{P_n \in \mathbf{P} : n \geq 1\}$ be given and denote by $X_{n,i}, i = 1, \ldots, n$ an i.i.d. sequence of random variables with distribution P_n. Define

$$\Delta_n(P_n) = n^{1/2}(\hat{\theta}_n - \theta(P_n)) - mn^{-1/2} \sum_{1 \leq i \leq n} g(X_{n,i}, P_n) , \tag{S.27}$$

where $g(x, P)$ is defined as in (31). Note that $\Delta_n(P_n)$ is itself a mean zero, degenerate U-statistic. By Lemma A on p. 183 of Serfling (1980), we therefore see that

$$\mathrm{Var}_{P_n}[\Delta_n(P_n)] = n\binom{n}{m}^{-1} \sum_{2 \leq c \leq m} \binom{m}{c}\binom{n-m}{m-c}\zeta_c(P_n) , \tag{S.28}$$

where the terms $\zeta_c(P_n)$ are nondecreasing in c and thus

$$\zeta_c(P_n) \le \zeta_m(P_n) = Var_{P_n}[h(X_1, \ldots, X_m)] \ .$$

Hence,

$$\text{Var}_{P_n}[\Delta_n(P_n)] \le n \binom{n}{m}^{-1} \sum_{2 \le c \le m} \binom{m}{c} \binom{n-m}{m-c} \text{Var}_{P_n}[h(X_1, \ldots, X_m)] \ .$$

It follows that

$$\text{Var}_{P_n}\left[\frac{\Delta_n(P_n)}{\sigma(P_n)}\right] \le n \binom{n}{m}^{-1} \sum_{2 \le c \le m} \binom{m}{c} \binom{n-m}{m-c} \frac{\text{Var}_{P_n}[h(X_1, \ldots, X_m)]}{\sigma(P_n)} \ . \tag{S.29}$$

Since

$$n \binom{n}{m}^{-1} \sum_{c=2}^{m} \binom{m}{c} \binom{n-m}{m-c} \to 0 \ ,$$

it follows from (34) that the lefthand-side of (S.29) tends to zero. Therefore, by Chebychev's inequality, we see that

$$\frac{\Delta_n(P_n)}{\sigma(P_n)} \stackrel{P_n}{\to} 0 \ .$$

Next, note that from Lemma 11.4.1 of Lehmann and Romano (2005) we have that

$$mn^{-1/2} \sum_{1 \le i \le n} \frac{g(X_{n,i}, P_n)}{\sigma(P_n)} \stackrel{d}{\to} \Phi(x) \tag{S.30}$$

under P_n. We therefore have further from Slutsky's Theorem that

$$\frac{n^{1/2}(\hat{\theta}_n - \theta(P_n))}{\sigma(P_n)} \stackrel{d}{\to} \Phi(x)$$

under P_n. An appeal to Polya's Theorem establishes the desired result. ■

Proof of Theorem 3.6: From the triangle inequality and Lemma S.11.1, we see immediately that

$$\sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} |J_b(x, P) - J_n(x, P)| \to 0 \ .$$

For any fixed $P \in \mathbf{P}$, we also have that $J_n(x, P)$ tends in distribution to a continuous limiting distribution. The desired conclusion (35) therefore follows from Theorem 2.1 and Remark 2.1.

Finally, from (S.30) and Remark 2.4, it follows that the same results hold for the feasible estimator $\hat{L}_n$. ■

S.12 Proof of Theorem 3.7

Lemma S.12.1. *Let* $J_n(x, P)$ *be the distribution of the root* (13)*. Let* $\mathbf{P}$ *be defined as in Theorem 3.7. Let* $\mathbf{P}'$ *be the set of all distributions on* $\mathbf{R}^k$*. Finally, for* $(Q, P) \in \mathbf{P}' \times \mathbf{P}$*, define*

$$\rho(Q,P) = \max\left\{\max_{1\leq j\leq k}\left\{\int_0^\infty |r_j(\lambda,Q) - r_j(\lambda,P)|\exp(-\lambda)d\lambda\right\}, ||\Omega(Q) - \Omega(P)||\right\} ,$$

where $\Omega(P)$ *is the correlation matrix of* P*,*

$$r_j(\lambda, P) = E_P\left[\left(\frac{X_j - \mu_j(P)}{\sigma_j(P)}\right)^2 I\left\{\left|\frac{X_j - \mu_j(P)}{\sigma_j(P)}\right| > \lambda\right\}\right] , \quad \text{(S.31)}$$

where the norm $||\cdot||$ *is the component-wise maximum of the absolute value of all elements. Then, for all sequences* $\{Q_n \in \mathbf{P}' : n \geq 1\}$ *and* $\{P_n \in \mathbf{P} : n \geq 1\}$ *satisfying* $\rho(Q_n, P_n) \to 0$

$$\limsup_{n\to\infty}\sup_{x\in\mathbf{R}} |J_n(x, Q_n) - J_n(x, P_n)| = 0 . \quad \text{(S.32)}$$

Proof: Consider sequences $\{Q_n \in \mathbf{P}' : n \geq 1\}$ and $\{P_n \in \mathbf{P} : n \geq 1\}$ satisfying $\rho(Q_n, P_n) \to 0$. We first argue that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} r_j(\lambda, P_n) = 0 \quad \text{(S.33)}$$

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty} r_j(\lambda, Q_n) = 0 \quad \text{(S.34)}$$

for all $1 \leq j \leq k$. Since $P_n \in \mathbf{P}$ for all $n \geq 1$, we have immediately that (S.33) holds for all $1 \leq j \leq k$. To see that (S.34) holds for all $1 \leq j \leq k$ as well, suppose by way of contradiction that it fails for some $1 \leq j \leq k$. It follows that there exists $\epsilon > 0$ such that for all λ' there exists $\lambda'' > \lambda'$ for which $r_j(\lambda'', Q_n) > 2\epsilon$ infinitely often. Since (S.33) holds, we have that there exists λ' such that $r_j(\lambda', P_n) < \epsilon$ for all n sufficiently large. Hence, there exists $\lambda'' > \lambda'$ such that $r_j(\lambda'', Q_n) > 2\epsilon$ and $r_j(\lambda', P_n) < \epsilon$ infinitely often. It follows that

$$|r_j(\lambda, P_n) - r_j(\lambda, Q_n)| > \epsilon$$

for all $\lambda \in (\lambda', \lambda'')$ infinitely often. Therefore, $\rho(P_n, Q_n) \not\to 0$, from which the desired conclusion follows.

We now establish (S.32). Suppose by way of contradiction that (S.32) fails. It follows that there exists a subsequence n_ℓ such that $\Omega(P_{n_\ell}) \to \Omega^*$, $\Omega(Q_{n_\ell}) \to \Omega^*$ and either

$$\sup_{x\in\mathbf{R}} |J_{n_\ell}(x, P_{n_\ell}) - \Phi_{\Omega^*}(x, \ldots, x)| \not\to 0 \quad \text{(S.35)}$$

or

$$\sup_{x \in \mathbf{R}} |J_{n_\ell}(x, Q_{n_\ell}) - \Phi_{\Omega^*}(x, \ldots, x)| \not\to 0 \ . \tag{S.36}$$

Let $\tilde{K}_n(x, P)$ be defined as in (S.12). Since

$$\Phi_{\Omega(P_{n_\ell})}(\cdot) \stackrel{d}{\to} \Phi_{\Omega^*}(\cdot) \ ,$$

it follows from (S.33) and the uniform central limit theorem established by Lemma 3.3.1 of Romano and Shaikh (2008) that $\tilde{K}_{n_\ell}(\cdot, P_{n_\ell}) \stackrel{d}{\to} \Phi_{\Omega^*}(\cdot)$. From Lemma S.6.1, we have for $1 \leq j \leq k$ that

$$\frac{S_{j,n_\ell}}{\sigma_j(P_{n_\ell})} \stackrel{P_{n_\ell}}{\to} 1 \ .$$

Hence, by Slutsky's Theorem, $K_{n_\ell}(\cdot, P_{n_\ell}) \stackrel{d}{\to} \Phi_{\Omega^*}(\cdot)$. By Polya's Theorem, we therefore see that (S.35) can not hold. A similar argument using (S.34) establishes that (S.36) can not hold. The desired claim is thus established. ■

Lemma S.12.2. *Let* $\mathbf{P}$ *is defined as in Theorem 3.7. Consider any sequence* $\{P_n \in \mathbf{P} : n \geq 1\}$*. Let* $X_{n,i}, i = 1, \ldots, n$ *be an i.i.d. sequence of random variables with distribution* P_n *and denote by* $\hat{P}_n$ *the empirical distribution of* $X_{n,i}, i = 1, \ldots, n$*. Then,*

$$\int_0^\infty |r_j(\lambda, \hat{P}_n) - r_j(\lambda, P_n)| \exp(-\lambda) d\lambda \stackrel{P_n}{\to} 0 \tag{S.37}$$

for all $1 \leq j \leq k$*, where* $r_j(\lambda, P)$ *is defined as in* (S.31)*.*

Proof: First assume without loss of generality that $\mu_j(P_n) = 0$ and $\sigma_j(P_n) = 1$ for all $1 \leq j \leq k$ and $n \geq 1$. Next, let $1 \leq j \leq k$ be given and note that $r_j(\lambda, \hat{P}_n) = A_n - 2B_n + C_n$, where

$$\begin{aligned} A_n &= \frac{1}{S_{j,n}^2} \frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i,j}^2 I\{|X_{n,i,j} - \bar{X}_{j,n}| > \lambda S_{j,n}\} \\ B_n &= \frac{\bar{X}_{j,n}}{S_{j,n}^2} \frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i,j} I\{|X_{n,i,j} - \bar{X}_{j,n}| > \lambda S_{j,n}\} \\ C_n &= \frac{\bar{X}_{j,n}^2}{S_{j,n}^2} \frac{1}{n} \sum_{1 \leq i \leq n} I\{|X_{n,i,j} - \bar{X}_{j,n}| > \lambda S_{j,n}\} \ . \end{aligned}$$

From Lemma 11.4.2 of Lehmann and Romano (2005), we see that

$$\bar{X}_{j,n} \stackrel{P_n}{\to} 0 \tag{S.38}$$

and

$$\frac{1}{n} \sum_{1 \leq i \leq n} |X_{n,i,j}| \stackrel{P_n}{\to} E_{P_n}[|X_{n,i,j}|] \leq 1 \ ,$$

where the inequality follows from the Cauchy-Schwartz inequality. From Lemma S.6.1, we see that

$$S_{j,n}^2 \stackrel{P_n}{\to} 1 \ . \tag{S.39}$$

Since

$$|B_n| \leq \frac{|\bar{X}_{j,n}|}{S_{j,n}^2} \frac{1}{n} \sum_{1 \leq i \leq n} |X_{n,i,j}| \ ,$$

we therefore see that $B_n = o_{P_n}(1)$ uniformly in λ. A similar argument establishes that $C_n = o_{P_n}(1)$ uniformly in λ and

$$A_n = \frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i,j}^2 I\{|X_{n,i,j} - \bar{X}_{j,n}| > \lambda S_{j,n}\} + o_{P_n}(1)$$

uniformly in λ. In summary,

$$r(\lambda, \hat{P}_n) = \frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i,j}^2 I\{|X_{n,i,j} - \bar{X}_{j,n}| > \lambda S_{j,n}\} + \Delta_n \tag{S.40}$$

uniformly in λ, where

$$\Delta_n = o_{P_n}(1) \ . \tag{S.41}$$

For $\epsilon > 0$, define the events

$$\begin{aligned} E_n(\epsilon) &= \{|\bar{X}_n| < \epsilon \cap 1 - \epsilon < S_n < 1 + \epsilon\} \\ E_n'(\epsilon) &= \left\{ \sup_{t \in \mathbf{R}} \left| \frac{1}{n} \sum_{1 \leq i \leq n} X_i^2 I\{|X_i| > t\} - E_{P_n}[X_i^2 I\{|X_i| > t\}] \right| < \epsilon \right\} \\ E_n''(\epsilon) &= \{|\Delta_n| < \epsilon\} \ . \end{aligned}$$

We first argue that

$$P_n\{E_n(\epsilon) \cap E_n'(\epsilon) \cap E_n''(\epsilon)\} \to 1 \ . \tag{S.42}$$

From (S.38) - (S.39) and (S.41), it suffices to argue that $P_n\{E_n'(\epsilon)\} \to 1$. To see this, first note that the class of functions

$$\{x^2 I\{|x| > t\} : t \in \mathbf{R}\} \tag{S.43}$$

is a VC class of functions. Therefore, by Theorem 2.6.7 and Theorem 2.8.1 of van der Vaart and Wellner (1996), we see that the class of functions (S.43) is Glivenko-Cantelli uniformly over $\mathbf{P}$.

Next, note that the event $E_n(\epsilon)$ implies that

$$I\{|X_i| > t^+(\lambda, \epsilon)\} \leq I\{|X_i - \bar{X}_n| > \lambda S_n\} \leq I\{|X_i| > t^-(\lambda, \epsilon)\}$$

for all λ, where

$$t^+(\lambda,\epsilon) = (1+\epsilon)\lambda + \epsilon$$
$$t^-(\lambda,\epsilon) = (1-\epsilon)\lambda - \epsilon \ .$$

The event $E_n(\epsilon) \cap E_n'(\epsilon)$ therefore implies that the first term on the right-hand side of (S.40) falls in the interval

$$[E_{P_n}[X_i^2 I\{|X_i| > t^+(\lambda,\epsilon)\}] - \epsilon, E_{P_n}[X_i^2 I\{|X_i| > t^-(\lambda,\epsilon)\}] + \epsilon]$$

for all λ. Hence, $E_n(\epsilon) \cap E_n'(\epsilon) \cap E_n''(\epsilon)$ implies that $r(\lambda, \hat{P}_n)$ falls in the interval

$$[E_{P_n}[X_i^2 I\{|X_i| > t^+(\lambda,\epsilon)\}] - 2\epsilon, E_{P_n}[X_i^2 I\{|X_i| > t^-(\lambda,\epsilon)\}] + 2\epsilon]$$

for all λ. Since, $\lambda \in [t^-(\lambda,\epsilon), t^+(\lambda,\epsilon)]$ for all $\lambda \geq 0$, it follows that $E_n(\epsilon) \cap E_n'(\epsilon) \cap E_n''(\epsilon)$ implies that

$$|r(\lambda, \hat{P}_n) - r(\lambda, P_n)| \leq r(t^-(\lambda,\epsilon), P_n) - r(t^+(\lambda,\epsilon), P_n) + 4\epsilon$$

for all $\lambda \geq 0$.

Since (S.42) holds for any $\epsilon > 0$, it follows that there exists $\epsilon_n \to 0$ such that (S.42) holds with ϵ_n in place of ϵ. Let ϵ_n be such a sequence. We have w.p.a. 1 under P_n that the left-hand side of (S.37) is bounded from above by

$$\int_0^\infty (r(t^-(\lambda,\epsilon_n), P_n) - r(t^+(\lambda,\epsilon_n), P_n) + 4\epsilon_n) \exp(-\lambda) d\lambda \ . \tag{S.44}$$

To complete the argument, it suffices to show that (S.44) tends to zero. Suppose by way of contradiction that this is not the case. Since (S.44) is bounded, it follows that there exists a subsequence along which it converges to $\delta > 0$. Since the sequence $\{P_n : n \geq 1\}$ is tight, along such a subsequence there exists a further subsequence n_ℓ such that P_{n_ℓ} converges weakly to P. Since $t^-(\lambda, \epsilon_n) \to \lambda$ and $t^+(\lambda,\epsilon_n) \to \lambda$, we have that

$$r(t^-(\lambda,\epsilon_{n_\ell}), P_{n_\ell}) - r(t^+(\lambda,\epsilon_{n_\ell}), P_{n_\ell}) + 4\epsilon_{n_\ell} \to r(\lambda, P) - r(\lambda, P) = 0$$

for all λ in a dense subset of the real line. Hence, by dominated convergence, (S.44) converges along the subsequence n_ℓ to zero instead of δ. This contradiction establishes that (S.44) tends to zero, from which (S.37) follows. ■

PROOF OF THEOREM 3.7: Let $\mathbf{P}'$ be the set of all distributions on $\mathbf{R}^k$. For $(Q, P) \in \mathbf{P}' \times \mathbf{P}$, define $\rho(Q, P)$ as in Lemma S.12.1. Consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$. Trivially,

$$P_n\{\hat{P}_n \in \mathbf{P}'\} \to 1 \ .$$

From Lemma S.7.1 and Lemma S.12.2, we see that $\rho(\hat{P}_n, P_n) \overset{P_n}{\to} 0$. Finally, for any sequences $\{Q_n \in \mathbf{P}' : n \geq 1\}$ and $\{P_n \in \mathbf{P} : n \geq 1\}$ satisfying $\rho(Q_n, P_n) \to 0$, we have by Lemma S.12.1 that

$$\limsup_{n\to\infty} \sup_{x \in \mathbf{R}} |J_n(x, Q_n) - J_n(x, P_n)| = 0 \ .$$

The desired conclusion (36) therefore follows from Theorem 2.4 and Remark 2.6. ■

S.13 Proof of Theorem 3.8

Let $\mathbf{P}'$ be the set of all distributions on $\mathbf{R}^k$. For $(Q, P) \in \mathbf{P}' \times \mathbf{P}$, define $\rho(Q, P)$ as in Lemma S.12.1. Consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$. Trivially,

$$P_n\{\hat{P}_n \in \mathbf{P}'\} \to 1 \ .$$

From Lemma S.7.1 and Lemma S.12.2, we see that $\rho(\hat{P}_n, P_n) \overset{P_n}{\to} 0$. To complete the argument, we establish that

$$\lim_{n\to\infty} \sup_{x \in \mathbf{R}} |J_n(x, Q_n) - J_n(x, P_n)| = 0 \tag{S.45}$$

for any sequences $\{Q_n \in \mathbf{P}' : n \geq 1\}$ and $\{P_n \in \mathbf{P} : n \geq 1\}$ satisfying $\rho(Q_n, P_n) \to 0$. To this end, suppose by way of contradiction that (S.45) fails. Then, there exists a subsequence n_ℓ and $\eta > 0$ such that

$$\sup_{x \in \mathbf{R}} |J_{n_\ell}(x, Q_{n_\ell}) - J_{n_\ell}(x, P_{n_\ell})| \to \eta \ .$$

By choosing a further subsequence if necessary, we may assume that $\Omega(P_{n_\ell}) \to \Omega^*$, $\Omega(Q_{n_\ell}) \to \Omega^*$. From Lemma S.7.1, it follows that $\hat{\Omega}_{n_\ell} \overset{P_{n_\ell}}{\to} \Omega^*$ and $\hat{\Omega}_{n_\ell} \overset{Q_{n_\ell}}{\to} \Omega^*$. By choosing an even further subsequence if neccesary, we may, again by arguing as in the proof of Lemma S.12.1, assume that $Z_{n_\ell}(P_{n_\ell}) \overset{d}{\to} Z^* \sim \Phi_{\Omega^*}(x)$ under P_{n_ℓ} and $Z_{n_\ell}(Q_{n_\ell}) \overset{d}{\to} Z^* \sim \Phi_{\Omega^*}(x)$ under Q_{n_ℓ}. Hence, by the continuous mapping theorem, we see that $f(Z_{n_\ell}(P_{n_\ell}), \hat{\Omega}_{n_\ell}) \overset{d}{\to} f(Z^*, \Omega^*)$ under P_{n_ℓ} and $f(Z_{n_\ell}(Q_{n_\ell}), \hat{\Omega}_{n_\ell}) \overset{d}{\to} f(Z^*, \Omega^*)$ under Q_{n_ℓ}. It follows from Lemma 3 on p. 260 of Chow and Teicher (1978) that $P_{n_\ell}\{f(Z_{n_\ell}(P_{n_\ell}), \hat{\Omega}_{n_\ell}) \leq x\}$ and $Q_{n_\ell}\{f(Z_{n_\ell}(Q_{n_\ell}), \hat{\Omega}_{n_\ell}) \leq x\}$ both converge uniformly to $P\{f(Z^*, \Omega^*) \leq x\}$. From this, we reach a contradiction to (S.45). The desired conclusion (39) therefore follows from Theorem 2.4. ■

S.14 Proof of Theorem 3.9

Note that

$$T_n(X^{(n)}) = \inf_{t \in \mathbf{R}^k : t \leq 0} (Z_n(P) - r(t, P))' \tilde{\Omega}_n^{-1} (Z_n(P) - r(t, P)) \ ,$$

where

$$r(t, P) = \left(\frac{\sqrt{n}(\mu_1(P) - t_1)}{S_{1,n}}, \ldots, \frac{\sqrt{n}(\mu_k(P) - t_k)}{S_{k,n}} \right)' .$$

It follows that

$$T_n(X^{(n)}) = \inf_{t \in \mathbf{R}^k : t \leq \tilde{r}(P)} (Z_n(P) - t)' \tilde{\Omega}_n^{-1} (Z_n(P) - t) ,$$

where

$$\tilde{r}(P) = -\left(\frac{\sqrt{n}\mu_1(P)}{S_{1,n}}, \ldots, \frac{\sqrt{n}\mu_k(P)}{S_{k,n}} \right)' .$$

Therefore, for any $P \in \mathbf{P}_0$, we have that $T_n(X^{(n)}) \leq R_n(X^{(n)}, P)$. It follows that for any such P,

$$P\{T_n(X^{(n)}) > J_n^{-1}(1 - \alpha, \hat{P}_n)\} \leq P\{R_n(X^{(n)}, P) > J_n^{-1}(1 - \alpha, \hat{P}_n)\} .$$

To complete the argument, it suffices to apply Theorem 3.8 with $f(Z_n(P), \hat{\Omega}_n) = R_n(X^{(n)}, P)$. The function f defined in this way is clearly continuous. It therefore only remains to verify the conditions (37) and (38). To this end, consider a sequence $\{P_n \in \mathbf{P} : n \geq 1\}$ such that $Z_n(P_n) \stackrel{d}{\to} Z$ under P_n and $\hat{\Omega}_n \stackrel{P_n}{\to} \Omega$, where $Z \sim N(0, \Omega)$. Since f is non-negative by construction, (37) holds trivially for $x < 0$ and (38) holds trivially for $x \leq 0$. By the continuous mapping theorem, $f(Z_n(P_n), \hat{\Omega}_n) \stackrel{d}{\to} f(Z, \Omega)$ under P_n. Since $P\{f(Z, \Omega) \leq x\}$ is continuous at $x > 0$, it follows that (37) and (38) also hold for $x > 0$. It remains to verify (37) for $x = 0$. To this end, note that

$$\begin{aligned} P_n\{f(Z_n(P_n), \hat{\Omega}_n) \leq 0\} &= P\{Z_n(P_n) \leq 0\} \\ &\to P\{Z \leq 0\} \\ &= P\{f(Z, \Omega) \leq 0\} , \end{aligned}$$

where the first equality follows from the fact that $\tilde{\Omega}_n$ defined in (40) is strictly positive definite, the convergence follows from the assumed convergence in distribution of $Z_n(P_n)$ to Z under P_n, and the second equality follows from the fact that $\max\{\epsilon - \det(\Omega), 0\} I_k + \Omega$ is strictly positive definite.

S.15 Proof of Theorem 3.10

By arguing as in Romano and Wolf (2005), we see that

$$FWER_P \leq P\left\{ \max_{j \in K_0(P)} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}} > J_n^{-1}(1 - \alpha, K_0(P), \hat{P}_n) \right\} , \tag{S.46}$$

where

$$K_0(P) = \{1 \leq j \leq k : \mu_j(P) \leq 0\} .$$

Furthermore, the righthand-side of (S.46) is bounded from above by

$$P\left\{\max_{j\in K_0(P)} \frac{\sqrt{n}(\bar{X}_{j,n}-\mu_j(P))}{S_{j,n}} > J_n^{-1}(1-\alpha, K_0(P), \hat{P}_n)\right\} .$$

The desired conclusion now follows immediately from Theorem 3.7. ■

S.16 Proof of Theorem 3.11

As in the proof of Theorem 3.5, it is useful to begin with some preliminaries. Recall that that

$$\begin{aligned} J_n(x,P) &= P\left\{\sup_{t\in\mathbf{R}} |B_n(P\{(-\infty,t]\})| \leq x\right\} \\ &= P\left\{\sup_{t\in R(P)} |B_n(t)| \leq x\right\} , \end{aligned}$$

where B_n is the uniform empirical process and $R(P)$ is defined as in (S.19). By Theorem 3.85 of Aliprantis and Border (2006), the set of all nonempty closed subsets of $[0,1]$ is a compact metric space with respect to the Hausdorff metric (S.20). Thus, for any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$, there is a subsequence n_ℓ and a closed set $R \subseteq [0,1]$ along which (S.21) holds. Finally, denote by B the standard Brownian bridge process. By the almost sure representation theorem, we may choose B_n and B so that (S.22) holds.

Let $\mathbf{P}'$ be the set of all distributions on $\mathbf{R}$. For $(Q,P) \in \mathbf{P}' \times \mathbf{P}$, let

$$\rho(Q,P) = \sup_{t\in\mathbf{R}} |Q\{(-\infty,t]\} - P\{(-\infty,t]\}| .$$

Consider sequences $\{P_n \in \mathbf{P} : n \geq 1\}$ and $\{Q_n \in \mathbf{P}' : n \geq 1\}$ such that $\rho(Q_n, P_n) \to 0$. We now argue that

$$\limsup_{n\to\infty} \sup_{x\in\mathbf{R}} |J_n(x,Q_n) - J_n(x,P_n)| = 0 . \tag{S.47}$$

Suppose by way of contradiction that (S.47) fails. It follows that there exists a subsequence n_ℓ and a closed subset $R \subseteq [0,1]$ such that $d_H(R(P_{n_\ell}), R) \to 0$ and either

$$\sup_{x\in\mathbf{R}} |J_{n_\ell}(x,P_{n_\ell}) - J^*(x)| \not\to 0 \tag{S.48}$$

or

$$\sup_{x\in\mathbf{R}} |J_{n_\ell}(x,Q_{n_\ell}) - J^*(x)| \not\to 0 , \tag{S.49}$$

where

$$J^*(x) = P\left\{\sup_{t\in R} |B(t)| \leq x\right\} .$$

Moreover, by the definition of $\mathbf{P}$, it must be the case that R contains some point different from zero and one. Since $\rho(Q_n, P_n) \to 0$, we have further that $d_H(R(Q_{n_\ell}), R) \to 0$ as well. It now follows from the same argument used to establish that neither (S.24) or (S.25) can hold that neither (S.48) or (S.49) can hold. Thus, (S.47) holds. Next, consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$. Trivially,

$$P_n\{\hat{P}_n \in \mathbf{P}'\} \to 1 \ .$$

By an exponential inequality used in the proof of the generalized Glivenko-Cantelli theorem (see, e.g., Pollard (1984)), we also have that

$$\rho(\hat{P}_n, P_n) \stackrel{P_n}{\to} 0 \ .$$

The desired conclusion therefore follows from Theorem 2.4 and Remark 2.6. ■

S.17 Proof of Theorem 3.12

Lemma S.17.1. *Let $g(x, P)$ be defined as in* (31)*. Then,*

$$E_P[|g(X, P)|^p] \leq E_P[|h(X_1, \ldots, X_m) - \theta_h(P)|^p]$$

for any $p \geq 1$.

Proof: Note that

$$E_P[h(X_1, \ldots, X_m)|X_1] - \theta_h(P) = g(X_1, P) \ .$$

Apply Jensen's inequality (conditional on X_1) to the function $|x|^p$ to obtain

$$|g(X_1, P)|^p \leq E_P[|h(X_1, \ldots, X_m) - \theta_h(P)|^p | X_1] \ . \tag{S.50}$$

The desired conclusion follows by taking expectations of both sides of (S.50). ■

Lemma S.17.2. *Let h be a symmetric kernel of degree m. Denote by $J_n(x, P)$ the distribution of $R_n(X^{(n)}, P)$ defined in* (30)*. Suppose*

$$\mathbf{P} \subseteq \mathbf{P}_{h,2+\delta,B} \cap \mathbf{S}_{h,\delta}$$

for some $\delta > 0$ and $B > 0$, where $\mathbf{P}_{h,2+\delta,B}$ and $\mathbf{S}_{h,\delta}$ are defined as in Example 3.11. Then,

$$\lim_{n \to \infty} \sup_{P \in \mathbf{P}} \sup_x |J_n(x, P) - \Phi(x/\sigma(P))| = 0 \ .$$

PROOF: It follows from Lemma S.17.1 and the definition of $\mathbf{P}_{h,2+\delta,B}$ that

$$E_P[|g(X,P)|^{2+\delta}] \le B \tag{S.51}$$

for all $P \in \mathbf{P}$, where g is defined as in (31). From the definitions of $\mathbf{P}_{h,2+\delta,B}$ and $\mathbf{S}_{h,\delta}$ and (S.51), we see that the conditions of Lemma S.11.1 hold, from which the desired conclusion follows. ■

Lemma S.17.3. *(Uniform Weak Law of Large Numbers for U-Statistics) Let h be a kernel of degree m. Consider any sequence $\{P_n \in \mathbf{P}_{h,1+\delta,B} : n \ge 1\}$ for some $\delta > 0$ and $B > 0$, where $\mathbf{P}_{h,1+\delta,B}$ is defined as in Example 3.11. Let $X_{n,i}, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P_n. Define*

$$\hat{\theta}_n = \frac{(n-m)!}{n!} \sum_p h(X_{n,i_1}, \ldots, X_{n,i_m}) \ . \tag{S.52}$$

Here, $\sum_p$ denotes summation over all $\binom{n}{m}$ subsets $\{i_1, \ldots, i_m\}$ of $\{1, \ldots, n\}$ together with each of the $m!$ permutations of each such subset. Then,

$$E_{P_n}[|\hat{\theta}_n - \theta_h(P_n)|^{1+\delta}] \to 0 \ ,$$

so

$$\hat{\theta}_n - \theta_h(P_n) \xrightarrow{P_n} 0 \ .$$

PROOF: Let $k = k_n$ be the greatest integer less than or equal to n/m. Compare $\hat{\theta}_n$ with the estimator $\tilde{\theta}_n$ defined by

$$\tilde{\theta}_n = k_n^{-1} \sum_{i=1}^{k_n} h(X_{n,m(i-1)+1}, X_{n,m(i-1)+2}, \ldots, X_{n,mi}) \ .$$

Note that $\tilde{\theta}_n$ is an average of k_n i.i.d. random variables. Furthermore,

$$E_{P_n}[\tilde{\theta}_n | \mathcal{F}_n] = \hat{\theta}_n \ ,$$

where $\mathcal{F}_n$ is the symmetric σ-field containing the set of observations $X_1, \ldots, X_n$ without regard to ordering. Since the function $|x|^{1+\delta}$ is convex, it follows from the Rao-Blackwell Theorem that

$$E_{P_n}[|\hat{\theta}_n - \theta_h(P_n)|^{1+\delta}] \le E_{P_n}[|\tilde{\theta}_n - \theta_h(P_n)|^{1+\delta}] \ .$$

By an extension of the Marcinkiewcz-Zygmund inequality (see, for instance, p. 361 of Chow and Teicher (1978)) and the definition of $\mathbf{P}_{h,1+\delta,B}$, the righthand-side of the last expression is bounded above by

$$A_\delta k_n^{-\delta} E_{P_n}[|h(X_{n,1}, \ldots, X_{n,m}) - \theta_h(P_n)|^{1+\delta}] \le A_\delta k_n^{-\delta} B \ ,$$

where A_δ is a universal constant. Since $k_n \to \infty$, the desired result follows. ■

Lemma S.17.4. *(Uniform Weak Law of Large Numbers for V-Statistics) Let h be a kernel of degree m. Consider any sequence $\{P_n \in \bar{\mathbf{P}}_{h,1+\delta,B} : n \geq 1\}$, where $\bar{\mathbf{P}}_{h,1+\delta,B}$ is defined as in Example 3.11. Let $X_{n,i}, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P_n. Define*

$$\bar{\theta}_n = \frac{1}{n^m} \sum_{1 \leq i_1 \leq n} \cdots \sum_{1 \leq i_m \leq n} h(X_{n,i_1}, \ldots, X_{n,i_m}) \ . \tag{S.53}$$

Then,

$$\bar{\theta}_n - \theta_h(P_n) \xrightarrow{P_n} 0 \ .$$

Proof: Note that

$$\bar{\theta}_n = \delta_n \hat{\theta}_n + (1 - \delta_n) S_n \ , \tag{S.54}$$

where S_n is the average of $h(X_{n,i_1}, \ldots, X_{n,i_m})$ over indices $\{i_1, \ldots, i_m\}$ where at least one i_j equals i_k for $j \neq k$ and

$$\delta_n = \frac{n(n-1)\cdots(n-m+1)}{n^m} = 1 - O(n^{-1}) \ .$$

It therefore suffices by Lemma S.17.3 to show that

$$S_n = O_{P_n}(1) \ .$$

To see this, apply Lemma S.17.3 to S_n by separating out terms with similar configurations of duplicates. Note that in the case where $i_1, \ldots, i_m$ are not all distinct, $|E_{P_n}[h(X_{n,i_1}, \ldots, X_{n,i_m})] - \theta_h(P_n)|$ need not be zero, but it is nevertheless bounded above by

$$E_{P_n}[|h(X_{n,i_1}, \ldots, X_{n,i_m}) - \theta_h(P_n)|] \leq B^{\frac{1}{1+\delta}}$$

by Hölder's inequality. The desired result follows. ■

Lemma S.17.5. *Let h be a symmetric kernel of degree m. Define the kernel h' of degree $2m$ according to* (44)*. Consider any sequence $\{P_n \in \bar{\mathbf{P}}_{h',1+\delta,B} : n \geq 1\}$, where $\bar{\mathbf{P}}_{h',1+\delta,B}$ is defined as in Example 3.11. Let $X_{n,i}, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution P_n. Denote by $\hat{P}_n$ the empirical distribution of $X_{n,i}, i = 1, \ldots, n$. Then $\sigma^2(P)$ defined by* (32) *satisfies*

$$\sigma^2(\hat{P}_n) - \sigma^2(P_n) \xrightarrow{P_n} 0 \ ,$$

so

$$P_n\{\hat{P}_n \in \mathbf{S}_{h,\delta'}\} \to 1$$

for any $0 < \delta' < \delta$, where $\mathbf{S}_{h,\delta'}$ is defined as in Example 3.11.

PROOF: Note that

$$g(x,\hat{P}_n) = \frac{1}{n^{m-1}} \sum_{1\le i_2\le n} \cdots \sum_{1\le i_m \le n}^{n} h(x, X_{i_2},\dots,X_{i_m}) - \theta(\hat{P}_n) \ ,$$

so

$$m^{-2}\sigma^2(\hat{P}_n) = \frac{1}{n}\sum_{1\le i_1\le n}\left[\frac{1}{n^{m-1}}\sum_{1\le i_2\le n}\cdots\sum_{1\le i_m\le n} h(X_{i_1},\dots,X_{i_m}) - \theta(\hat{P}_n)\right]^2 \ .$$

Since

$$\theta^2(\hat{P}_n) = n^{-2m}\sum_{i_1=1}^{n}\cdots\sum_{i_m=1}^{n}\sum_{j_1=1}^{n}\cdots\sum_{j_m=1}^{n} h(X_{i_1},\dots,X_{i_m})h(X_{j_1},\dots,X_{j_m}) \ ,$$

we have that

$$m^{-2}\sigma^2(\hat{P}_n) = \frac{1}{n^{2m}}\sum_{i_1=1}^{n}\cdots\sum_{i_m=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_m=1}^{n} h'(X_{i_1},\dots,X_{i_m},X_{j_1},\dots,X_{j_m}) \ . \tag{S.55}$$

Applying Lemma S.17.4 to the righthand-side of (S.55), we see that

$$m^{-2}\sigma^2(\hat{P}_n) - \theta_{h'}(P_n) \overset{P_n}{\to} 0 \ .$$

Next, note that

$$\begin{aligned}\text{Var}_{P_n}[g(X,P_n)] &= \text{Var}_{P_n}[E_{P_n}[h(X_1,\dots,X_m)|X_1]] \\ &= E_{P_n}[h(X_1,\dots,X_m)h(X_1,X_{m+2},\dots,X_{2m})] \\ &\qquad -E_{P_n}[h(X_1,\dots,X_m)]E_{P_n}[h(X_{m+1},\dots,X_{2m})] \ .\end{aligned}$$

Thus, $\theta_{h'}(P_n) = m^{-2}\sigma^2(P_n)$, which completes the proof. ■

Lemma S.17.6. *Let h be a symmetric kernel of degree m. Consider any sequence $\{P_n \in \bar{\mathbf{P}}_{h,2+\delta,B} : n \ge 1\}$, where $\bar{\mathbf{P}}_{h,2+\delta,B}$ is defined as in Example 3.11. Let $X_{n,i}, i = 1,\dots,n$ be an i.i.d. sequence of random variables with distribution P_n. Denote by $\hat{P}_n$ the empirical distribution of $X_{n,i}, i = 1,\dots,n$. Then there exists $\delta' > 0$ and $B' > 0$ such that*

$$P_n\{\hat{P}_n \in \mathbf{P}_{h,2+\delta',B'}\} \to 1 \ .$$

PROOF: Choose $0 < \delta' < \delta$ and note that

$$A_n \equiv E_{\hat{P}_n}[|h(X_1,\dots,X_m) - \theta_h(\hat{P}_n)|^{2+\delta'}] = \frac{1}{n^m}\sum_{1\le i_1\le n}\cdots\sum_{1\le i_m\le n}|h(X_{n,i_1},\dots,X_{n,i_m}) - \theta_h(\hat{P}_n)|^{2+\delta'} \ .$$

It suffices to show that there exists $B' > 0$ such that $A_n \le B'$ with probability approaching one under P_n. By Minkowski's inequality,

$$A_n^{\frac{1}{2+\delta'}} \le \left[\frac{1}{n^m}\sum_{1\le i_1\le n}\cdots\sum_{1\le i_m\le n}|h(X_{n,i_1},\dots,X_{n,i_m}) - \theta_h(P_n)|^{2+\delta'}\right]^{\frac{1}{2+\delta'}} + |\theta_h(\hat{P}_n) - \theta_h(P_n)| \ . \tag{S.56}$$

To analyze the first term on the lefthand-side of (S.56), we apply Lemma S.17.4 with the kernel

$$\tilde{h}(x_1 \ldots, x_m) = |h(x_1, \ldots, x_m) - \theta_h(P_n)|^{2+\delta'} . \tag{S.57}$$

To see that the lemma is applicable, we verify that

$$D_n \equiv E_{P_n}[|\tilde{h}(X_{n,i_1}, \ldots, X_{n,i_m}) - \theta_{\tilde{h}}(P_n)|^{1+\epsilon}] \leq C$$

for some $\epsilon > 0$ and $C > 0$. By Minkowski's inequality, we have that

$$D_n^{\frac{1}{1+\epsilon}} \leq E_{P_n}[|\tilde{h}(X_{i_1}, \ldots, X_{i_m})|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} + |E_{P_n}[\tilde{h}(X_1, \ldots, X_m)]| . \tag{S.58}$$

Choose $\epsilon > 0$ so that $(1+\epsilon)(2+\delta') = 2+\delta$. By (S.57) and the definition of $\bar{\mathbf{P}}_{h,2+\delta,B}$, the first and second terms in (S.58) are both bounded from above by $B^{\frac{1}{1+\epsilon}}$. It therefore suffices to take $C = 2^{1+\epsilon}B$. It follows that the first term on the lefthand-side of (S.56) may be expressed as

$$[E_{P_n}[|h(X_{n,i_1}, \ldots, X_{n,i_m}) - \theta_h(P_n)|^{2+\delta'}] + o_{P_n}(1)]^{\frac{1}{2+\delta'}} .$$

By Lemma S.17.4, we have that the second term on the lefthand-side of (S.56) is $o_{P_n}(1)$. Hence, the lefthand-side of (S.56) may be expressed as

$$[E_{P_n}[|h(X_{n,i_1}, \ldots, X_{n,i_m}) - \theta_h(P_n)|^{2+\delta'}] + o_{P_n}(1)]^{\frac{1}{2+\delta'}} + o_{P_n}(1) .$$

From the definition of $\bar{\mathbf{P}}_{h,2+\delta,B}$, the desired result follows by setting

$$B' = ((B^{\frac{2+\delta'}{2+\delta}} + \epsilon)^{\frac{1}{2+\delta'}} + \epsilon)^{2+\delta'}$$

for some $\epsilon > 0$. ■

Proof of Theorem 3.12: Choose $\delta' > 0$ and $B' > 0$ according to Lemma S.17.6. Let $\mathbf{P}'$ be any set of distributions on $\mathbf{R}$ such that

$$\mathbf{P}' \subseteq \mathbf{P}_{h,2+\delta',B'} \cap \mathbf{S}_{h,\delta'} .$$

For $(Q, P) \in \mathbf{P}' \times \mathbf{P}$, define

$$\rho(Q, P) = |\sigma^2(Q) - \sigma^2(P)| .$$

Consider any sequence $\{P_n \in \mathbf{P} : n \geq 1\}$. From Lemma S.17.5 and Lemma S.17.6, we have that

$$P_n\{\hat{P}_n \in \mathbf{P}'\} \to 1 .$$

From Lemma S.17.5, we have that

$$\rho(\hat{P}_n, P_n) \xrightarrow{P_n} 0 .$$

Finally, for all sequences $\{Q_n \in \mathbf{P}' : n \geq 1\}$ and $\{P_n \in \mathbf{P} : n \geq 1\}$ satisfying $\rho(Q_n, P_n) \to 0$, we have from Lemma S.17.2 and the triangle inequality that

$$\limsup_{n\to\infty} \sup_{x\in\mathbf{R}} |J_n(x, Q_n) - J_n(x, P_n)| = 0 .$$

The desired conclusion therefore follows from Theorem 2.4 and Remark 2.6. ■

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, New York.

Chow, Y. S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability and Martingales*. Springer, New York.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society, Providence, RI.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer, New York.

Pollard, D. (1984). *Convergence of Stochastic Proceses*. Springer, New York.

Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson*, **138** 2786–2807.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, **100** 94–108.

Serfling, R. A. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.