Supplementary Materials for "A Test of Homogeneity of Distributions when Observations are Subject to Measurement Errors"

DongHyuk Lee[1], Soumendra N. Lahiri[2], and Samiran Sinha[3,*]

[1] Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

[2]Department of Statistics, North Carolina State University, Raleigh, NC, USA

[3]Department of Statistics, Texas A&M University, College Station, TX, USA

*email: sinha@stat.tamu.edu

Appendix

1 Proof of Theorems 1, 2 and 3

Let $c_{jw}(t) = \cos(t\overline{W}_j)$, $d_{jw}(t) = \sin(t\overline{W}_j)$. Next define $e_{jw}(t) = M_x^{-1}\sum_{(l_1,l_2)\in\mathcal{S}_x}\cos\{(t/m_x)(W_{jl_1} - W_{jl_2})\}$. Denote the expectations by $c_{0w}(t) = E\{c_{jw}(t)\}, d_{0w}(t) = E\{d_{1w}(t)\}$, and $e_{0w}(t) = E\{e_{1w}(t)\}$. Then, $\boldsymbol{\Lambda}_{\boldsymbol{W}_j}(t) \equiv (c_{jw}(t) - c_{0w}(t), d_{jw}(t) - d_{0w}(t), e_{jw}(t) - e_{0w}(t))^T$ are iid mean zero random vectors. Similarly define $\boldsymbol{\Lambda}_{\boldsymbol{V}_j}(t)$ by replacing $\boldsymbol{W}_j$'s by $\boldsymbol{V}_j$'s in the definition of $\boldsymbol{\Lambda}_{\boldsymbol{W}_j}(t)$. Let $t_0 = \max\{|t_1|, |t_2|\}$, where recall that $\omega(t) = 0$ for all $t \notin [t_1, t_2]$. Define

$$\boldsymbol{Z}_n(t) = n_x^{-1/2}\begin{pmatrix} \sum_{j=1}^{n_x}\boldsymbol{\Lambda}_{\boldsymbol{W}_j}(t) \\ \sum_{j=1}^{n_y}\boldsymbol{\Lambda}_{\boldsymbol{V}_j}(t) \end{pmatrix}, \ |t| \le t_0.$$

Let $C, C(\cdot)$ denote generic constants with values in $(0, \infty)$ that may depend on their arguments (if any) but not on n_x, n_y. Also, let $\ell^\infty[-t_0, t_0]$ denote the set of all bounded measurable functions from $[-t_0, t_0]$ to the real line and let $\|x\|_\infty = \sup\{|x(t)| : t \in [-t_0, t_0]\}$, $x \in \ell^\infty[-t_0, t_0]$. Finally, let A^{T} denote the transpose of a matrix (vector) A.

Then we have the following result.

Lemma 1. *$\boldsymbol{Z}_n \stackrel{d}{\to} \boldsymbol{Z}$ as random elements of the space $(l^\infty[-t_0, t_0])^6$, where $\boldsymbol{Z}$ is a 6-dimensional zero-mean Gaussian process on $[-t_0, t_0]$ with the covariance function*

$$\Gamma(s,t) = \begin{bmatrix} \Gamma_w(s,t) & 0 \\ 0 & \rho^{-2}\Gamma_v(s,t) \end{bmatrix},$$

with $\Gamma_v(s,t) = E\{\boldsymbol{\Lambda}_{\boldsymbol{W}_1}(s)\boldsymbol{\Lambda}_{\boldsymbol{W}_1}(t)\}$, $\Gamma_v(s,t) = E\{\boldsymbol{\Lambda}_{\boldsymbol{V}_1}(s)\boldsymbol{\Lambda}_{\boldsymbol{V}_1}(t)\}$, for $-t_0 \le s, t \le t_0$. Further, the paths of $\boldsymbol{Z}(\cdot)$ are continuous on $[-t_0, t_0]$ with probability one.

Proof: Note that *i)* $\boldsymbol{\Lambda}_{\boldsymbol{W}_j}(t)$ and $\boldsymbol{\Lambda}_{\boldsymbol{V}_j}(t)$ are bounded random vectors, *ii)* the collection of functions $\{(\boldsymbol{\Lambda}_{\boldsymbol{w}}(t), \boldsymbol{\Lambda}_{\boldsymbol{v}}(t)); t \in [-t_0, t_0]\}$ is a VC-class, where $\boldsymbol{\Lambda}_{\boldsymbol{w}}(t) = [\cos(t\sum_{j=1}^{m_x} w_j/m_x), \sin$

$(t\sum_{j=1}^{m_x} w_j/m_x)$, $M_x^{-1}\sum_{(l_1,l_2)\in S}\cos\{t(w_{l_1}-w_{l_2})/m_x\}]$, and $\boldsymbol{\Lambda}_{\boldsymbol{v}}(t)$ is defined similarly. Hence, by using the Multivariate CLT (cf. Ch 11.1, Athreya and Lahiri, 2006), the finite dimensional distribution of the $\boldsymbol{Z}_n(\cdot)$-process converges in distribution to those of the $\boldsymbol{Z}(\cdot)$-process. Further, using the standard exponential inequalities (e.g., Hoeffding, 1963) and the chaining argument (Wellner and van der Vaart, 2006), it follows that $\boldsymbol{Z}_n \to \boldsymbol{Z}$ in distribution, where $\boldsymbol{Z}$ is a random element of $l^\infty([-t_0,t_0])]^6$ and it has continuous paths on $[-t_0,t_0]$ with probability one. □

Proof of Theorem 1. Recall the definitions of $\widehat{a}_x(t)$ and $\widehat{a}_{2x}(t)$ given in Section 2.2 of the main document. Define $a_{2x}(t)=\phi_{u_x}^{m_x}(t/m_x)$ and let $Z_{kn}(t)$ be the kth component of $\boldsymbol{Z}_n(t)$ defined in Lemma 1. Then $a_x(t)=c_{0w}(t)/a_{2x}(t)$, and

$$\begin{aligned}
\sqrt{n_x}\{\widehat{a}_x(t)-a_x(t)\} &= \sqrt{n_x}\left\{\frac{n_x^{-1}\sum_{j=1}^{n_x} c_{jw}(t)}{\widehat{a}_{2x}(t)}-\frac{c_{0w}(t)}{a_{2x}(t)}\right\}\\
&= \sqrt{n_x}\left[\frac{n_x^{-1}\sum_{j=1}^{n_x}\{c_{jw}(t)-c_{0w}(t)+c_{0w}(t)\}}{\widehat{a}_{2x}(t)}-\frac{c_{0w}(t)}{a_{2x}(t)}\right]\\
&= \sqrt{n_x}\left[\frac{n_x^{-1}\sum_{j=1}^{n_x}\{c_{jw}(t)-c_{0w}(t)\}}{\widehat{a}_{2x}(t)}+\frac{c_{0w}(t)}{\widehat{a}_{2x}}-\frac{c_{0w}(t)}{a_{2x}(t)}\right]\\
&= \frac{Z_{1n}(t)}{\widehat{a}_{2x}(t)}-\frac{c_{0w}(t)\sqrt{n_x}\{\widehat{a}_{2x}(t)-a_{2x}(t)\}}{a_{2x}(t)\widehat{a}_{2x}(t)}.
\end{aligned}$$

Now using the fact that $\widehat{a}_{2x}(t)=\{\phi_{u_x}^2(t/m_x)+Z_{3n}(t)/\sqrt{n_x}\}^{m_x/2}$, we get

$$\begin{aligned}
\sqrt{n_x}\{\widehat{a}_x(t)-a_x(t)\} &= \frac{Z_{1n}(t)}{a_{2x}(t)}-\frac{m_x c_{0w}(t)Z_{3n}(t)\phi_{u_x}^{m_x-2}(t/m_x)}{2a_{2x}^2(t)}+R_{nx}(t),\\
&\equiv A_{nx}(t)+R_{nx}(t),
\end{aligned}$$

where

$$A_{nx}(t)=\frac{Z_{1n}(t)}{a_{2x}(t)}-\frac{m_x c_{0w}(t)Z_{3n}(t)\phi_{u_x}^{m_x-2}(t/m_x)}{2a_{2x}^2(t)}$$

and where, with a suitable constant $C(m_x)\in(0,\infty)$,

$$\begin{aligned}
|R_{nx}(t)| &\leq \frac{|Z_{1n}(t)||Z_{3n}(t)|}{\sqrt{n_x}}\times\frac{m_x\{1+|Z_{3n}(t)/\sqrt{n_x}|^{m_x/2-1}\}}{2|a_{2x}(t)|\widehat{a}_{2x}(t)}+\frac{|c_{0w}(t)|\{1+|Z_{3n}(t)|^{m_x}\}}{|a_{2x}(t)|\sqrt{n_x}|\widehat{a}_{2x}(t)|}\times C(m_x)\\
&\quad+\frac{m_x|c_{0w}||Z_{3n}(t)|^2\{1+|Z_{3n}(t)/\sqrt{n_x}|^{m_x/2-1}\}}{2|a_{2x}(t)|^3\sqrt{n_x}|\widehat{a}_{2x}(t)|}.
\end{aligned}$$

Hence,

$$\int|R_{nx}(t)|^2\omega(t)dt \;\leq\; \frac{C(m_x)}{n_x}\left\{\int\frac{\omega(t)}{a_{2x}^2(t)}dt\right\}\left[||Z_{1n}||_\infty^2||Z_{3n}||_\infty^2+\frac{\{1+||Z_{3n}||_\infty^{2m_x}\}}{\alpha_x^{4m_x}}\right]$$

$$\times \frac{\{1+(||Z_{3n}||_\infty/\sqrt{n_x})^{m_x/2-1}\}^2}{(\alpha_x^2-||Z_{3n}||_\infty/\sqrt{n_x})^{m_x}},$$

where $\alpha_x = \min\{|\phi_{u_x}(t/m_x)|; |t| \leq t_0\}$. Since $||\cdot||_\infty$ is continuous on $\ell^\infty[-t_0,t_0]$, it follows that $||Z_{kn}||_\infty \xrightarrow{d} ||Z_k||_\infty$ for $k=1,\dots,6$. Hence

$$\int |R_{nx}(t)|^2\omega(t)dt \to 0 \tag{A.1}$$

in probability. Next, define $a_{2y}(t) = \phi_{u_y}^{m_y}(t/m_y)$ and write $\widehat{a}_{2y}(t) = \{\phi_{u_y}^2(t/m_y)+(\sqrt{n_x}/n_y)Z_{6n}(t)\}^{m_y/2}$. Then, using similar steps as above, we obtain

$$\sqrt{n_x}\{\widehat{a}_y(t)-a_y(t)\} \quad = \quad \frac{n_x}{n_y}\left\{A_{ny}(t)+R_{ny}(t)\right\},$$

where

$$A_{ny} = \frac{Z_{4n}(t)}{a_{2y}(t)} - \frac{m_y c_{0v}(t) Z_{6n}(t)\phi_{u_y}^{m_y-2}(t/m_y)}{2a_{2y}^2(t)},$$

and where, retracing arguments above, one can show that

$$\int |R_{ny}(t)|^2\omega(t)dt \to 0 \tag{A.2}$$

in probability. Under $H_0 : \phi_x(t) = \phi_y(t)$, that means $a_x(t) = a_y(t)$ for all t. So, under H_0,

$$\begin{aligned} I_1 &= n_x \int \{\widehat{a}_x(t)-\widehat{a}_y(t)\}^2\omega(t)dt \\ &= n_x \int [\{\widehat{a}_x(t)-a_x(t)\}-\{\widehat{a}_y(t)-a_y(t)\}]^2\omega(t)dt \\ &= I_{11}+Q_n, \end{aligned}$$

where

$$I_{11} \quad = \quad \int \left\{A_{nx}(t) - \frac{n_x}{n_y}A_{ny}(t)\right\}^2 \omega(t)dt,$$

and using the Cauchy-Schwartz inequality

$$|Q_n| \leq \int \{R_{nx}(t)+\frac{n_x}{n_y}R_{ny}(t)\}^2\omega(t)dt + 2\left[I_{11}\times\int\{R_{nx}(t)+\frac{n_x}{n_y}R_{ny}(t)\}^2\omega(t)dt\right]^{1/2}.$$

By (A.1) and (A.2), $|Q_n| \to 0$ in probability. Next applying the continuous mapping theorem, we obtain

$$I_{11} \xrightarrow{d} I_{1\infty} \equiv \int \xi_1^2(t)\omega(t)dt,$$

where $\xi_1(t) = A_x(t) - \rho^2 A_y(t)$. Repeating the arguments above with $I_2 = n_x \int \{\widehat{b}_x(t) - \widehat{b}_y(t)\}^2 \omega(t)dt$ and using the joint weak convergence result of Lemma 1, one can show that

$$\begin{aligned}
T_n &= I_1 + I_2 \\
&= I_{11} + \int \Bigg[\Bigg\{ \frac{Z_{2n}(t)}{a_{2x}(t)} - \frac{m_x d_{0v}(t) Z_{3n}(t) \phi_{u_x}^{m_x-2}(t/m_x)}{2a_{2x}^2(t)} \Bigg\} \\
&\qquad - \frac{n_x}{n_y} \Bigg\{ \frac{Z_{5n}(t)}{a_{2y}(t)} - \frac{m_y d_{0v}(t) Z_{6n}(t) \phi_{u_y}^{m_y-2}(t/m_y)}{2a_{2y}^2(t)} \Bigg\} \Bigg]^2 \omega(t)dt + o_p(1) \\
&\xrightarrow{d} I_{1\infty} + \int \Bigg[\Bigg\{ \frac{Z_2(t)}{a_{2x}(t)} - \frac{m_x d_{0v}(t) Z_3(t) \phi_{u_x}^{m_x-2}(t/m_x)}{2a_{2x}^2(t)} \Bigg\} \\
&\qquad - \rho^2 \Bigg\{ \frac{Z_5(t)}{a_{2y}(t)} - \frac{m_y d_{0v}(t) Z_6(t) \phi_{u_y}^{m_y-2}(t/m_y)}{2a_{2y}^2(t)} \Bigg\} \Bigg]^2 \omega(t)dt \\
&\equiv \int [\xi_1^2(t) + \xi_2^2(t)] \omega(t)dt.
\end{aligned}$$

This completes the proof of Theorem 1. □

Proof of Theorem 2. First suppose that $\int D_a^2(t)\omega(t)dt \neq 0$. Let $W_a(t) = \{\widehat{a}_x(t) - a_x(t)\} + \{a_y(t) - \widehat{a}_y(t)\}$, $|t| \leq t_0$. Then, it follows that

$$T_{1n_x} \equiv n_x \int \Big[\{\widehat{a}_x(t) - a_x(t)\} + \{a_x(t) - a_y(t)\} + \{a_y(t) - \widehat{a}_y(t)\} \Big]^2 \omega(t)dt \geq L_{1n_x},$$

where $L_{1n_x} = n_x \int \{a_x(t) - a_y(t)\}^2 \omega(t)dt + 2n_x \int W_a(t)\{a_x(t) - a_y(t)\}\omega(t)dt$. Now, using the steps in the proof of Theorem 1 and the continuous mapping theorem, one can show that the second term of L_{1n_x} is $O_p(\sqrt{n_x})$ while the first term diverges at the rate n_x. Thus, $L_{1n_x} = O_p(n_x)$. Hence, for

$$\mathrm{pr}(T_{1n_x} \leq r) \leq \mathrm{pr}(L_{1n_x} \leq r) \to 0 \;\text{ for any } r \in (0, \infty).$$

Next consider the case where $\int D_b^2(t)\omega(t)dt \neq 0$. Then, defining T_{2n_x} by replacing $\widehat{a}_x$, $\widehat{a}_y$, a_x, a_y in T_{1n_x} by $\widehat{b}_x$, $\widehat{b}_y$, b_x, b_y and using the arguments above, we have $\mathrm{pr}(T_{2n_x} \leq r) \to 0$ for any $r \in (0, \infty)$. Thus, if $\int [D_a^2(t) + D_b^2(t)]\omega(t)dt \neq 0$, then for any α,

$$\begin{aligned}
\mathrm{pr}(T_{n_x} > t_{n_x,\alpha}) &= 1 - \mathrm{pr}(T_{1n_x} + T_{2n_x} \leq t_{n_x,\alpha}) \\
&\geq 1 - \min\big\{\mathrm{pr}(T_{1n_x} \leq t_\alpha), \mathrm{pr}(T_{2n_x} \leq t_\alpha)\big\} \to 1 \;\text{ as } n_x \to \infty,
\end{aligned}$$

proving Theorem 2. □

Proof of Theorem 3. First we show that $\widehat{\phi}_x(t) \equiv \widehat{\phi}_{\overline{W}}(t)/\{\widehat{\phi}_{u_x}(t/m_x)\}^{m_x} = \widehat{\phi}_1(t)\phi_K(h_w t)/\{\widehat{\phi}_{u_x}(t/m_x)\}^{m_x}$ converges to $\phi_x(t)$ uniformly over $|t| \leq t_0$, almost surely. Since $h_w \to 0$, it is enough to show that

$$\sup\{|\widehat{\phi}_1(t) - \phi_1(t)| : |t| \leq t_0\} \to 0 \text{ almost surely, and} \tag{A.3}$$

$$\sup\{|\widehat{\phi}_{u_x}(t) - \phi_{u_x}(t)| : |t| \leq t_0 m_x\} \to 0 \text{ almost surely.} \tag{A.4}$$

Since $\widehat{\phi}_1(t) = n_x^{-1}\sum_{j=1}^{n_x}\exp(it\overline{W}_j)$ is an average of i.i.d., bounded random variables, one can prove (A.3) using a discretization argument and Hoeffding's inequality (Hoeffding, 1963); see, e.g., Lahiri (1994). Next, for $h > 0$, write $e_{jw}(t,h) = M_x^{-1}\sum_{(l_1,l_2)\in\mathcal{S}_x}\cos\{(t/m_x)(W_{jl_1} - W_{jl_2})\}(1 - h^2t^2)^3 I(|ht| \le 1)$ and $e_{0w}(t,h) \equiv E\{e_{jw}(t,h)\}$. Then, it is easy to check that $e_{0w}(t,h) = |\phi_{u_x}(t/m_x)|^2(1-h^2t^2)^3 I(|ht| \le 1)$, and hence, $\sup\{|e_{0w}(t,h_w) - \phi_{u_x}(t/m_x)| : |t| \le t_0 m_x\} \to 0$, as $h_w \to 0$. Further, using arguments similar to those in the proof of (A.3), one can show that $\sup\{|\widehat{\phi}_{u_x}(t) - e_{0w}(t,h_w)| : |t| \le t_0 m_x\} \to 0$, almost surely. Thus, (A.4) holds. Let A be the event where (A.3) and (A.4) hold. Then $\mathrm{pr}(A) = 1$. Next, let B be the event where

$$\sup\{|\widehat{\phi}_2(t) - \phi_2(t)| : |t| \le t_0\} \to 0, \quad \text{and}$$
$$\sup\{|\widehat{\phi}_{v_x}(t) - \phi_{v_x}(t)| : |t| \le t_0 m_y\} \to 0,$$

as $n_x \to \infty$. Then, by similar arguments, $\mathrm{pr}(B) = 1$, implying, $\mathrm{pr}(A \cap B) = 1$.

We shall now show that $T^*_{n_x}$ converges in distribution to $T_\infty \equiv \int[[\xi_1^2(t) + \xi_2^2(t)]\omega(t)dt$, i.e., the Prohorov distance between the Bootstrap probability distribution of $T^*_{n_x}$ and the the probability distribution of T_∞ goes to zero, on the set $A \cap B$. Let $\boldsymbol{Z}^*_n(t)$ be defined by replacing $(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_{n_x})$ and $(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{n_y})$ in $\boldsymbol{Z}(t)$ by the corresponding Bootstrap variables $(\boldsymbol{W}^*_1, \ldots, \boldsymbol{W}^*_{n_x})$ and $(\boldsymbol{V}^*_1, \ldots, \boldsymbol{V}^*_{n_y})$, respectively. Also, let $\hat{\Gamma}(s,t)$ denote the covaraince matrix function of $\boldsymbol{Z}^*_n(\cdot)$, i.e., $\hat{\Gamma}(s,t) = E_*\boldsymbol{Z}^*_n(t)\boldsymbol{Z}^*_n(t)^T, \quad s,t \in [-t_0, t_0]$, where E_* denotes expectation under P_*. Then, using Lemma 1, it is easy to check that on the set $A \cap B$,

$$\sup\left\{\|\hat{\Gamma}(s,t) - \Gamma(s,t)\| : s,t \in [-t_0,t_0]\right\} \to 0 \ \text{ as } n_x \to \infty.$$

As a result, for any $\omega \in A \cap B$, the finite dimensional distributions of the $\boldsymbol{Z}^*_n$-process converges to those of the $\boldsymbol{Z}$-process, and further by Hoeffding's inequality, the tightness condition continues to hold. This implies that on the set $A \cap B$, $\boldsymbol{Z}^*_n$ converges in distribution to the same limiting process $\boldsymbol{Z}$ as in Lemma 1. Further, repeating the arguments in the proof of Theorem 1 and using uniform convergence of $\widehat{\phi}_1(t)$, $\widehat{\phi}_2(t)$, $\widehat{\phi}_{u_x}(t)$ and $\widehat{\phi}_{v_y}(t)$ o their respective limits on the set $A \cap B$, one can show that, for any $\omega \in A \cap B$,

$$T^*_{n_x} \to^d T_\infty.$$

Theorem 3 now follows from Theorem 1, Polya's Theorem, and the continuity of the limiting random variable T_∞. □

2 Figures for the sensitivity analysis of Section 4

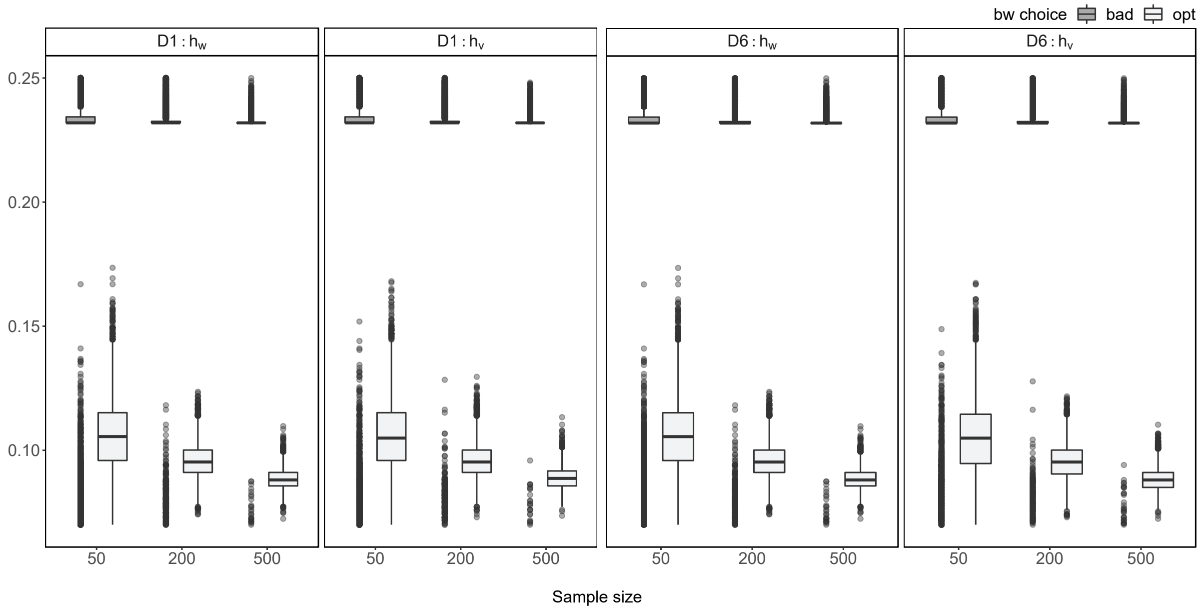

Figure 1: Boxplots of the optimal and bad choices of bandwidth (h_w, h_v) for simulation scenarios D1 and D6.

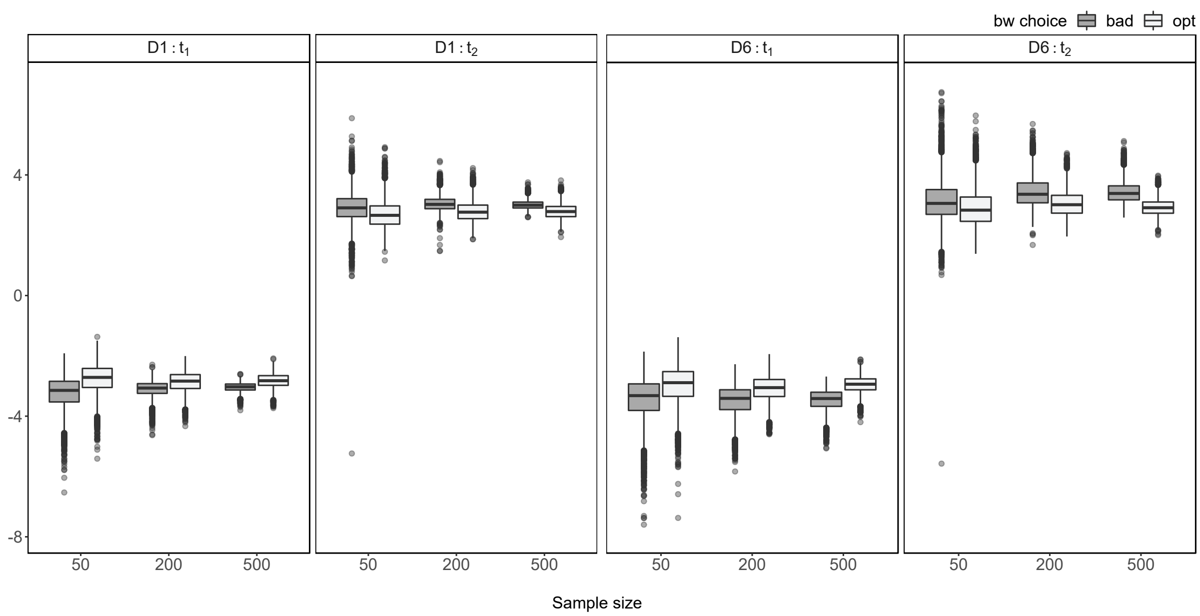

Figure 2: Boxplots of t_1 and t_2 for the optimal and bad choices of bandwidth (h_w, h_v) for simulation scenarios D1 and D6.

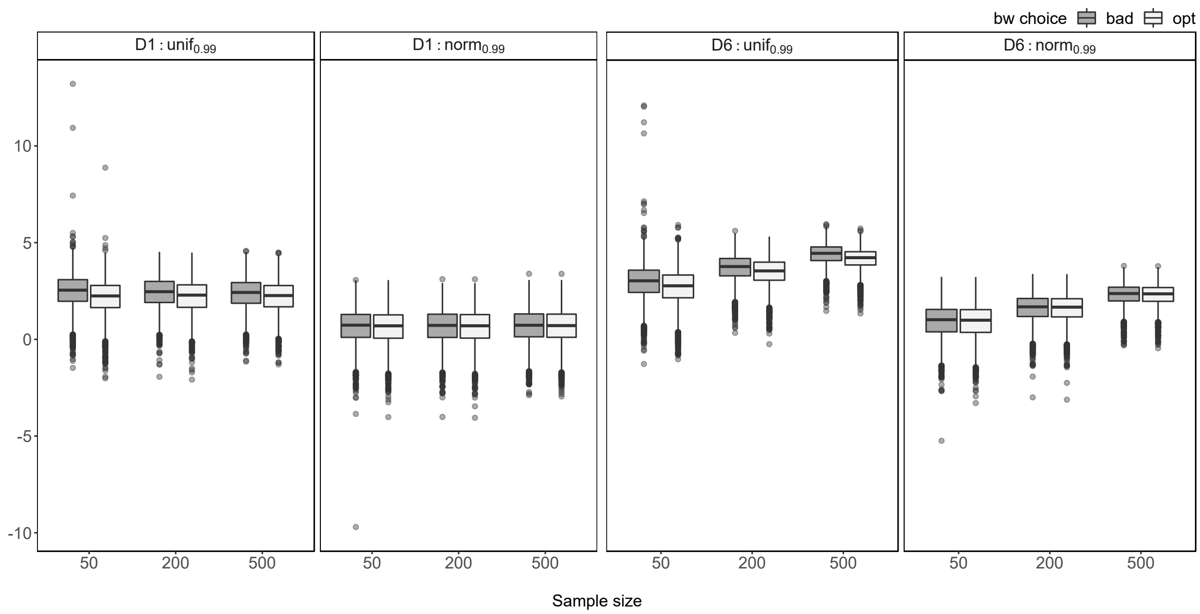

Figure 3: Boxplots of test statistics for the optimal and bad choices of bandwidth (h_w, h_v) for simulation scenarios D1 and D6.

References

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory.* Springer-Verlag: New York.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**, 13-30.

Lahiri, S. N.(1994). Two term Edgeworth expansion and bootstrap approximation for multivariate studentized M-estimators. *Sankhya, Series A* **56**, 201–226.

Wellner, J. A. and van der Vaart, A. W. (2006). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, NY.