Omnibus Model Checks of Linear Assumptions through Distance Covariance

Kai Xu and Daojiang He

Anhui Normal University

Supplementary Material

The supplementary file contains proofs of Theorems 2–3 and the second statement of Theorem 1 as well as more numerical results on some aspects of limiting distributions and a real data set.

S1 More technical proofs

Proof of Theorem 1. The proof of the first statement is available in the main part of the submitted paper. We here prove the second assertion.

Let $\varepsilon_i = m(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_i, i = 1, \cdots, n$. By (3.7),

$$E\left[\mathbf{g}(\mathbf{x})\{m(\mathbf{x}) - \mathbf{g}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta}_0\}\right] = 0.$$

In the light of the independence between η and $\mathbf{x}$, $E\{\mathbf{g}(\mathbf{x})\eta\} = 0$. Thus, $E\{\mathbf{g}(\mathbf{x})\varepsilon\} = E\left[\mathbf{g}(\mathbf{x})\{m(\mathbf{x}) - \mathbf{g}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta}_0\}\right] + E\{\mathbf{g}(\mathbf{x})\eta\} = 0$. This, together with Slutsky's theorem and the moment condition $E\{\|\mathbf{g}(\mathbf{x})\|^2 + \varepsilon^2\} < \infty$, entails

$$n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) = n^{-1/2}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\varepsilon_i + o_p(1). \tag{S1.1}$$

Recall that $\eta_{in} - \varepsilon_i = Y_i - \mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\beta}_n - \varepsilon_i = Y_i - \mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\beta}_n - \{m(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_i\} = -(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{g}(\mathbf{x}_i)$. Let $\mathfrak{z}_i = (\varepsilon_i, \mathbf{x}_i)$. By the analog of (6.24),

$$U_n \stackrel{\text{def}}{=} U_{3n} + (\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)^{\mathrm{T}}U_{4n} + U_{51n} + U_{52n}, \tag{S1.2}$$

where $U_{3n}, U_{4n}, U_{51n}, U_{52n}$ are defined exactly in a similar manner to U_{0n}, U_{1n}, U_{21n}, U_{22n} except that we replace $\mathbf{z}_i = (\eta_i, \mathbf{x}_i)$ by $\mathfrak{z}_i = (\varepsilon_i, \mathbf{x}_i)$. Since the indicator function is bounded, employing arguments exactly similar to treating (6.28), $nU_{51n} = o_p(1)$, which indicates $n^{1/2}U_{51n} = o_p(1)$. From Condition D1 and (6.29), uniformly over $1 \le s, t \le n$,

$$\begin{aligned} &\int_0^{|\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)|}\{F_{\varepsilon^{(2)}_{(1)}}(z) - F_{\varepsilon^{(2)}_{(1)}}(0)\}dz \\ = \ & O_p(1)(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_s) - \mathbf{g}(\mathbf{x}_t)\}\{\mathbf{g}(\mathbf{x}_s) - \mathbf{g}(\mathbf{x}_t)\}^{\mathrm{T}}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0). \end{aligned} \tag{S1.3}$$

By (S1.1), $n^{1/2}\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_0\| = O_p(1)$ and $\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_0\| = o_p(1)$. This together with (S1.3) and Slutsky's theorem implies

$$n^{1/2}U_{52n} = O_p(1)\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}U^{\natural}_{2n}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) = o_p(1).$$

Therefore, $n^{1/2}U_{5n} = n^{1/2}U_{51n} + n^{1/2}U_{52n} = o_p(1)$. Combination of (S1.2) leads to

$$n^{1/2}U_n = n^{1/2}U_{3n} + \{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}} U_{4n} + o_p(1).$$

Obviously, U_{3n} and U_{4n} are the U-statistics with the kernels $h_0(\mathfrak{z}_i, \mathfrak{z}_j, \mathfrak{z}_k, \mathfrak{z}_l)$ and $h_1(\mathfrak{z}_i, \mathfrak{z}_j, \mathfrak{z}_k, \mathfrak{z}_l)$, respectively. Under the alternative (1.4), both U_{3n} and U_{4n} are non-degenerate. Invoking technical appendix 1.1 in Yao, Zhang and Shao (2018), we have

$$\begin{aligned} h_0^{(1)}(\mathfrak{z}_1) &\stackrel{\text{def}}{=} 4E[\{h_0(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_4, \mathfrak{z}_4) - dCov^2(\varepsilon_1, \mathbf{x}_1)\}|\mathfrak{z}_1] \\ &= 2E\{C_\varepsilon(\varepsilon_1, \varepsilon_2) C_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2)|\mathfrak{z}_1\} - 2dCov^2(\varepsilon_1, \mathbf{x}_1), \quad \text{(S1.4)} \end{aligned}$$

where $C_\varepsilon(\cdot,\cdot)$ and $C_{\mathbf{x}}(\cdot,\cdot)$ are defined as in (6.33). As a result, $n^{1/2}\{U_{3n} - dCov^2(\varepsilon_1, \mathbf{x}_1)\} = n^{1/2}\sum_{i=1}^{n} h_0^{(1)}(\mathfrak{z}_i)$. Tedious calculation yields

$$E\{h_1(\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_4, \mathfrak{z}_4)\} = -2E[\{\mathbf{g}(\mathbf{x}_1) - \mathbf{g}(\mathbf{x}_2)\} C_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2) I(\varepsilon_1 > \varepsilon_2)] \stackrel{\text{def}}{=} 2\varrho_1 \in \mathbb{R}^d \quad \text{(S1.5)}$$

By the law of large numbers for U-statistics, it follows $U_{4n} \to 2\varrho_1$, in probability. Consequently, combination of these, (S1.1) and Slutsky's theorem leads to

$$\begin{aligned} & n^{1/2}\{U_n - dCov^2(\varepsilon_1, \mathbf{x}_1)\} \\ = \;& n^{1/2}\{U_{3n} - dCov^2(\varepsilon_1, \mathbf{x}_1)\} + 2\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}} \varrho_1 + o_p(1) \\ = \;& 2n^{-1/2}\sum_{i=1}^{n}\{L(\varepsilon_i, \mathbf{x}_i) + \varrho_1^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{g}(\mathbf{x}_i)\varepsilon_i\} + o_p(1), \end{aligned}$$

where

$$L(\varepsilon_1, \mathbf{x}_1) = E\{C_\varepsilon(\varepsilon_1, \varepsilon_2) C_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2) | \mathfrak{z}_1\} - dCov^2(\varepsilon_1, \mathbf{x}_1).$$

By the central limit theorem, $n^{1/2}\{U_n - dCov^2(\varepsilon_1, \mathbf{x}_1)\}$ has an asymptotic normal distribution with mean zero and variance $4\text{var}\{L(\varepsilon_i, \mathbf{x}_i) + \varrho_1^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{g}(\mathbf{x}_i)\varepsilon_i\}$. Recalling the classical theory of non-degenerate statistics of U-type and V-type and combining the expression (6.43), the desired result on V_n holds and hence the related details are omitted to avoid repetition.

□

Proof of Theorem 2. We only report the analysis of the statistic U_n in that the derivations on the statistics U_n and V_n are parallel. Under the local alternatives (4.14), we have $\eta_{in} - \eta_i = -(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)^{\mathrm{T}} \mathbf{g}(\mathbf{x}_i) + n^{-1/2}\ell(\mathbf{x}_i)$. Apparently,

$$\begin{aligned} & |\eta_{in} - \eta_{jn}| \\ = \; & |\eta_i - \eta_j| - \{\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_j)\}^{\mathrm{T}}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\{I(\eta_i > \eta_j) - I(\eta_i < \eta_j)\} \\ & + n^{-1/2}\{\ell(\mathbf{x}_i) - \ell(\mathbf{x}_j)\}\{I(\eta_i > \eta_j) - I(\eta_i < \eta_j)\} \\ & + 2\int_0^{\{\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_j)\}^{\mathrm{T}}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) - n^{-1/2}\{\ell(\mathbf{x}_i) - \ell(\mathbf{x}_j)\}} \{I(\eta_i - \eta_j \le z) - I(\eta_i \le \eta_j)\} dz. \end{aligned} \tag{S1.6}$$

By careful calculation, we obtain

$$nU_n = nU_{0n} + \{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}\{n^{1/2} U_{1n}\} + nU_{2n}^{(\ell)} + U_{30n}^{(\ell)},$$

where U_{0n} and U_{1n} are defined as in (6.24), $U_{2n}^{(\ell)} = \{c(n,4)\}^{-1} \sum_{i<j<k<l} h_2^{(\ell)}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{z}_l)$ with

$$h_2^{(\ell)}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{z}_l) = 6^{-1} \sum_{s<t,u<v}^{(i,j,k,l)} \delta_{2st}^{(\ell)}(\|\mathbf{x}_{(s)}^{(t)}\| + \|\mathbf{x}_{(u)}^{(v)}\|) - 12^{-1} \sum_{(s,t,u)}^{(i,j,k,l)} \delta_{2st}^{(\ell)} \|\mathbf{x}_{(s)}^{(u)}\|$$

and $\delta_{2st}^{(\ell)} = \int_0^{\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)-n^{-1/2}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}}\{I(\eta_s - \eta_t \le z) - I(\eta_s \le \eta_t)\}dz + \int_0^{\{\mathbf{g}(\mathbf{x}_t)-\mathbf{g}(\mathbf{x}_s)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)-n^{-1/2}\{\ell(\mathbf{x}_t)-\ell(\mathbf{x}_s)\}} \{I(\eta_t - \eta_s \le z) - I(\eta_t \le \eta_s)\}dz$, and $U_{30n}^{(\ell)} = \{c(n,4)\}^{-1} \sum_{i<j<k<l} h_{30}^{(\ell)}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{z}_l)$ with

$$h_{30}^{(\ell)}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{z}_l) = 6^{-1} \sum_{s<t,u<v}^{(i,j,k,l)} \delta_{30st}^{(\ell)}(\|\mathbf{x}_{(s)}^{(t)}\| + \|\mathbf{x}_{(u)}^{(v)}\|) - 12^{-1} \sum_{(s,t,u)}^{(i,j,k,l)} \delta_{30st}^{(\ell)} \|\mathbf{x}_{(s)}^{(u)}\|$$

and $\delta_{30st}^{(\ell)} = \{\ell(\mathbf{x}_s) - \ell(\mathbf{x}_t)\}\{I(\eta_s > \eta_t) - I(\eta_s < \eta_t)\}$. Clearly,

$$E\{h_{30}^{(\ell)}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4)\} = 2E[\{\ell(\mathbf{x}_1) - \ell(\mathbf{x}_2)\}C_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2)I(\eta_1 > \eta_2)] = 0,$$

where we use the independence between η and $\mathbf{x}$ and the fact $E\{\ell(\mathbf{x}_1)C_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2)\} = E\{\ell(\mathbf{x}_2)\ C_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2)\}$. By the law of large numbers for U-statistics, we have

$$U_{30n}^{(\ell)} \to 0, \tag{S1.7}$$

in probability. Recalling that the terms U_{0n} and U_{1n} have been considered in (6.35), we now need to focus on the rest term $U_{2n}^{(\ell)}$.

For convenience, write $\delta_{2ij}^{(\ell)} = \delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)$ and $h_2^{(\ell)}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{z}_l) = h_2^{(\ell)}(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{z}_l; \boldsymbol{\beta}_n)$. Apparently, $\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0) = \int_0^{-n^{-1/2}\{\ell(\mathbf{x}_i)-\ell(\mathbf{x}_j)\}}\{I(\eta_i - \eta_j \le z) - I(\eta_i \le$

$\eta_j)\}dz+\int_0^{-n^{-1/2}\{\ell(\mathbf{x}_i)-\ell(\mathbf{x}_j)\}}\{I(\eta_j-\eta_i\le z)-I(\eta_j\le\eta_i)\}dz$. That is, $h_2^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l;\boldsymbol{\beta}_0)\neq 0$ and Lemma 2 can not be used directly. To this end, we decompose $\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)$ into three parts

$$\begin{aligned}
&\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)\\
=\;&\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0)-E\{\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0)|\mathbf{x}_i,\mathbf{x}_j\}+E\{\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)|\mathbf{x}_i,\mathbf{x}_j\}\\
&+\Big[\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)-\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0)-E\{\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)-\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0)|\mathbf{x}_i,\mathbf{x}_j\}\Big]\\
\stackrel{\text{def}}{=}\;&\delta_{20ij}^{(\ell)}(\boldsymbol{\beta}_0)+\delta_{21ij}^{(\ell)}(\boldsymbol{\beta}_n)+\delta_{22ij}^{(\ell)}(\boldsymbol{\beta}_n),
\end{aligned}$$

where $\delta_{20ij}^{(\ell)}(\boldsymbol{\beta}_0)=\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0)-E\{\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_0)|\mathbf{x}_i,\mathbf{x}_j\}$, $\delta_{21ij}^{(\ell)}(\boldsymbol{\beta}_n)=E\{\delta_{2ij}^{(\ell)}(\boldsymbol{\beta}_n)|\mathbf{x}_i,\mathbf{x}_j\}$ and $\delta_{22ij}^{(\ell)}(\boldsymbol{\beta}_n)$ is denoted in an obvious way. By the exact analog of dealing with (6.28), $n^{-3}\sum_{i\neq j\neq k\neq l}\delta_{22st}^{(\ell)}(\boldsymbol{\beta}_n)(\|\mathbf{x}_{(s)}^{(t)}\|+\|\mathbf{x}_{(u)}^{(v)}\|-2\|\mathbf{x}_{(s)}^{(u)}\|)=o_p(1)$ uniformly over $\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|\le Cn^{-1/2}$. Using Slutsky's theorem gives

$$nU_{2n}^{(\ell)}=nU_{20n}^{(\ell)}+nU_{21n}^{(\ell)}+o_p(1),$$

where $U_{20n}^{(\ell)}=\{c(n,4)\}^{-1}\sum_{i<j<k<l}h_{20}^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)$ with

$$h_{20}^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)=\sum_{s<t,u<v}^{(i,j,k,l)}\delta_{20st}^{(\ell)}(\boldsymbol{\beta}_0)(\|\mathbf{x}_{(s)}^{(t)}\|+\|\mathbf{x}_{(u)}^{(v)}\|)-12^{-1}\sum_{(s,t,u)}^{(i,j,k,l)}\delta_{20st}^{(\ell)}(\boldsymbol{\beta}_0)\|\mathbf{x}_{(s)}^{(u)}\|,$$

and

$$h_{21}^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)=\sum_{s<t,u<v}^{(i,j,k,l)}\delta_{21st}^{(\ell)}(\boldsymbol{\beta}_n)(\|\mathbf{x}_{(s)}^{(t)}\|+\|\mathbf{x}_{(u)}^{(v)}\|)-12^{-1}\sum_{(s,t,u)}^{(i,j,k,l)}\delta_{21st}^{(\ell)}(\boldsymbol{\beta}_n)\|\mathbf{x}_{(s)}^{(u)}\|.$$

It is observed that $U_{20n}^{(\ell)}$ is degenerate, namely, $E\{h_{20}^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)|\mathbf{z}_i\}=0$ implied by $E\{\delta_{20ij}^{(\ell)}(\boldsymbol{\beta}_0)|\mathbf{z}_i\}=0$ and the independence between η and $\mathbf{x}$, and

$E\{I(\eta_1-\eta_2\le z)-I(\eta_1\le\eta_2)\}^2=|F_{\eta_{(1)}^{(2)}}(z)-F_{\eta_{(1)}^{(2)}}(0)|$. Besides, condition $E\{|\ell(\mathbf{x})|^{2+\gamma}\}<\infty$ implies $\max_{1\le i,j\le n}|\ell(\mathbf{x}_i)-\ell(\mathbf{x}_j)|\le 2\max_{1\le i\le n}|\ell(\mathbf{x}_i)|=o_p(n^{1/2})$. Based on these observations, we have

$$\mathrm{var}\{nU_{20n}^{(\ell)}\}\le CE\int_0^{|n^{-1/2}\{\ell(\mathbf{x}_1)-\ell(\mathbf{x}_2)\}|}|F_{\eta_{(1)}^{(2)}}(z)-F_{\eta_{(1)}^{(2)}}(0)|dz=O(n^{-1}),$$

which leads to $nU_{2n}^{(\ell)}=nU_{21n}^{(\ell)}+o_p(1)$. Consequently,

$$nU_n \;=\; nU_{0n}+\{n^{1/2}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)\}^{\mathrm{T}}\{n^{1/2}U_{1n}\}+nU_{21n}^{(\ell)}+o_p(1),$$

Additionally, $\max_{1\le s,t\le n}|\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)+n^{-1/2}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}|\le 2\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|\max_{1\le i\le n}\|\mathbf{g}(\mathbf{x}_i)\|+2n^{-1/2}\max_{1\le i\le n}|\ell(\mathbf{x}_i)|=o_p(n^{1/2})\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|+o_p(1)=o_p(1)$. Combining this and Taylor's expansion, we have uniformly over $1\le s,t\le n$,

$$\begin{aligned}
&\delta_{21st}^{(\ell)}(\boldsymbol{\beta}_n) \qquad\qquad (\mathrm{S1.8})\\
=\;&\int_0^{\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)-n^{-1/2}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)}\{F_{\eta_{(1)}^{(2)}}(z)-F_{\eta_{(1)}^{(2)}}(0)\}dz\\
&+\int_0^{\{\mathbf{g}(\mathbf{x}_t)-\mathbf{g}(\mathbf{x}_s)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)-n^{-1/2}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)}\{F_{\eta_{(1)}^{(2)}}(z)-F_{\eta_{(1)}^{(2)}}(0)\}dz\\
=\;&\{1+o_p(1)\}f_{\eta_{(1)}^{(2)}}(0)\left[(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}-n^{-1/2}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}\right]^2\\
=\;&\{1+o_p(1)\}f_{\eta_{(1)}^{(2)}}(0)(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}^{\mathrm{T}}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)\\
&-2\{1+o_p(1)\}n^{-1/2}f_{\eta_{(1)}^{(2)}}(0)(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}\\
&+\{1+o_p(1)\}n^{-1}f_{\eta_{(1)}^{(2)}}(0)\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}^2,
\end{aligned}$$

from which it follows

$$\begin{aligned} nU_{21n}^{(\ell)} &= f_{\eta_{(1)}^{(2)}}(0)\{n^{1/2}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)\}^{\mathrm{T}}U_{2n}^{\natural}\{n^{1/2}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)\} \\ &\quad -2f_{\eta_{(1)}^{(2)}}(0)\{n^{1/2}(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)\}^{\mathrm{T}}U_{3n}^{(\ell)}+f_{\eta_{(1)}^{(2)}}(0)U_{4n}^{(\ell)}, \qquad \text{(S1.9)} \end{aligned}$$

where $U_{2n}^{\natural}$ is described as in (6.31), $U_{3n}^{(\ell)}=\{c(n,4)\}^{-1}\sum_{i<j<k<l}h_3^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)$ with

$$h_3^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)=\sum_{s<t,u<v}^{(i,j,k,l)}\delta_{3st}^{(\ell)}(\|\mathbf{x}_{(s)}^{(t)}\|+\|\mathbf{x}_{(u)}^{(v)}\|)-12^{-1}\sum_{(s,t,u)}^{(i,j,k,l)}\delta_{3st}^{(\ell)}\|\mathbf{x}_{(s)}^{(u)}\|,$$

and $\delta_{3st}^{(\ell)}=\{\mathbf{g}(\mathbf{x}_s)-\mathbf{g}(\mathbf{x}_t)\}\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}$, and $U_{4n}^{(\ell)}=\{c(n,4)\}^{-1}\sum_{i<j<k<l}h_4^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)$ with

$$h_4^{(\ell)}(\mathbf{z}_i,\mathbf{z}_j,\mathbf{z}_k,\mathbf{z}_l)=\sum_{s<t,u<v}^{(i,j,k,l)}\delta_{4st}^{(\ell)}(\|\mathbf{x}_{(s)}^{(t)}\|+\|\mathbf{x}_{(u)}^{(v)}\|)-12^{-1}\sum_{(s,t,u)}^{(i,j,k,l)}\delta_{4st}^{(\ell)}\|\mathbf{x}_{(s)}^{(u)}\|,$$

and $\delta_{4st}^{(\ell)}=\{\ell(\mathbf{x}_s)-\ell(\mathbf{x}_t)\}^2$. By direct calculations, $E\{h_3^{(\ell)}(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3,\mathbf{z}_4)\}=E[\{\mathbf{g}(\mathbf{x}_1)-\mathbf{g}(\mathbf{x}_2)\}\{\ell(\mathbf{x}_1)-\ell(\mathbf{x}_2)\}C_{\mathbf{x}}(\mathbf{x}_1,\mathbf{x}_2)]=-2E\{\mathbf{g}(\mathbf{x}_1)-E\mathbf{g}(\mathbf{x}_1)\}\{\ell(\mathbf{x}_2)-E\ell(\mathbf{x}_2)\}\|\mathbf{x}_1-\mathbf{x}_2\|\stackrel{\text{def}}{=}2\varrho_{1\ell}\in\mathbb{R}^d$ and $E\{h_4^{(\ell)}(\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3,\mathbf{z}_4)\}=E[\{\ell(\mathbf{x}_1)-\ell(\mathbf{x}_2)\}^2C_{\mathbf{x}}(\mathbf{x}_1,\mathbf{x}_2)]=-2E\{\ell(\mathbf{x}_1)-E\ell(\mathbf{x}_1)\}\{\ell(\mathbf{x}_2)-E\ell(\mathbf{x}_2)\}\|\mathbf{x}_1-\mathbf{x}_2\|\stackrel{\text{def}}{=}2\varrho_{2\ell}$, where we also employ the fact $E\{C_{\mathbf{x}}(\mathbf{x}_1,\mathbf{x}_2)|\mathbf{x}_1\}=0$. Under the local alternatives (4.14), $\boldsymbol{\beta}_n - \boldsymbol{\beta}_0$ can be expressed as

$$\begin{aligned}
& n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0) \qquad \text{(S1.10)}\\
= \; & n^{-1/2}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\eta_i + \left[\{n^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\}^{-1} - \boldsymbol{\Sigma}^{-1}\right] n^{-1/2}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\eta_i \\
& + n^{-1/2}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\ell(\mathbf{x}_i) + \left[\{n^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\}^{-1} - \boldsymbol{\Sigma}^{-1}\right] n^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\ell(\mathbf{x}_i) \\
= \; & n^{-1/2}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\eta_i + \boldsymbol{\Sigma}^{-1}E\{\mathbf{g}(\mathbf{x})\ell(\mathbf{x})\} + o_p(1).
\end{aligned}$$

Write $\varrho_{0\ell} = E\{\mathbf{g}(\mathbf{x})\ell(\mathbf{x})\} \in \mathbb{R}^d$. By employing (S1.10) and Slutsky's theorem,

$$\begin{aligned}
& nU_n \\
= \; & nU_{0n} + \{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}\{n^{1/2}U_{1n}\} + f_{\eta_{(1)}^{(2)}}(0)\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}U_{2n}^{\natural}\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\} \\
& -2f_{\eta_{(1)}^{(2)}}(0)\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}U_{3n}^{(\ell)} + f_{\eta_{(1)}^{(2)}}(0)U_{4n}^{(\ell)} + o_p(1) \\
= \; & nU_{0n} + \{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}\{n^{1/2}U_{1n}\} + 2f_{\eta_{(1)}^{(2)}}(0)\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}\boldsymbol{\Lambda}\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\} \\
& -4f_{\eta_{(1)}^{(2)}}(0)\{n^{1/2}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)\}^{\mathrm{T}}\varrho_{1\ell} + 2f_{\eta_{(1)}^{(2)}}(0)\varrho_{2\ell} + o_p(1) \\
= \; & \sum_{i=1}^{\infty}\lambda_i\{n^{-1/2}\sum_{j=1}^{n}\phi_i(\mathbf{z}_j)\}^2 + \{n^{-1/2}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\eta_i + \varrho_{0\ell}\}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\{n^{-1/2}\sum_{i=1}^{n}h_1^{(1)}(\mathbf{z}_i) - 4\varrho_{1\ell}f_{\eta_{(1)}^{(2)}}(0)\} \\
& +2f_{\eta_{(1)}^{(2)}}(0)\{n^{-1/2}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\eta_i + \varrho_{0\ell}\}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\{n^{-1/2}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i)\eta_i + \varrho_{0\ell}\} + 2f_{\eta_{(1)}^{(2)}}(0)\varrho_{2\ell} \\
& +o_p(1).
\end{aligned}$$

Following the arguments for (6.42), we complete the proof.

□

Proof of Theorem 3. Let $\mathbf{z}_{in}^* = (\eta_{in}^*, \mathbf{x}_{in}), i = 1, \cdots, n$, and denote by the p in o_p^*, E^* and var* the probability, expectation and variance under the bootstrapped space. Recall that given the data, $\{\mathbf{z}_{in}^*\}_{i=1}^n$ is independent and identically distributed and η_{in}^* is independent of $\mathbf{x}_{in}$. Applying the technical details of Theorem 1 to the bootstrapped space, we have

$$\begin{aligned} & nU_n^* \\ = \;& n^{-1}\sum_{i\neq j} h_{0n}^{(2)}(\mathbf{z}_{in}^*, \mathbf{z}_{jn}^*) + \{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^*\}^{\mathrm{T}}\boldsymbol{\Sigma}_n^{*-1}\{n^{-1/2}\sum_{i=1}^n h_{1n}^{(1)}(\mathbf{z}_{in}^*)\} \\ & +2f^*_{\eta^{(2)}_{(1)}}(0)\{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^*\}^{\mathrm{T}}\boldsymbol{\Sigma}_n^{*-1}\boldsymbol{\Lambda}_n^*\boldsymbol{\Sigma}_n^{*-1}\{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^*\} + o_p^*(1), \end{aligned}$$

and

$$\begin{aligned} & nV_n^* \\ = \;& n^{-1}\sum_{i=1}^n\sum_{j=1}^n h_{0n}^{(2)}(\mathbf{z}_{in}^*, \mathbf{z}_{jn}^*) + \{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^*\}^{\mathrm{T}}\boldsymbol{\Sigma}_n^{*-1}\{n^{-1/2}\sum_{i=1}^n h_{1n}^{(1)}(\mathbf{z}_{in}^*)\} \\ & +2f^*_{\eta^{(2)}_{(1)}}(0)\{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^*\}^{\mathrm{T}}\boldsymbol{\Sigma}_n^{*-1}\boldsymbol{\Lambda}_n^*\boldsymbol{\Sigma}_n^{*-1}\{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^*\} + o_p^*(1), \end{aligned}$$

where $\boldsymbol{\Sigma}_n^* = E^*\{\mathbf{g}(\mathbf{x}_{1n})\mathbf{g}(\mathbf{x}_{1n})^{\mathrm{T}}\}$, $\boldsymbol{\Lambda}_n^* = -E^*[\{\mathbf{g}(\mathbf{x}_{1n}) - E^*\mathbf{g}(\mathbf{x}_{1n})\}\{\mathbf{g}(\mathbf{x}_{2n}) - E^*\mathbf{g}(\mathbf{x}_{2n})\}^{\mathrm{T}}\|\mathbf{x}_{1n} - \mathbf{x}_{2n}\|]$, $h_{0n}^{(2)}(\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*) = E\{h_0(\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*, \mathbf{z}_{3n}^*, \mathbf{z}_{4n}^*)|\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*\} = C_\eta^*(\eta_{1n}^*, \eta_{2n}^*)C_{\mathbf{x}}^*(\mathbf{x}_{in}, \mathbf{x}_{jn})$ with $C_{\eta^*}^*(\eta_{1n}^*, \eta_{2n}^*) = |\eta_{1n}^* - \eta_{2n}^*| - E^*(|\eta_{1n}^* - \eta_{2n}^*||\eta_{1n}^*) - E^*(|\eta_{1n}^* - \eta_{2n}^*||\eta_{2n}^*) + E^*(|\eta_{1n}^* - \eta_{2n}^*|)$ and $C_{\mathbf{x}}^*(\mathbf{x}_{1n}, \mathbf{x}_{2n}) = \|\mathbf{x}_{1n} - \mathbf{x}_{2n}\| - E^*(\|\mathbf{x}_{1n} - \mathbf{x}_{2n}\||\mathbf{x}_{1n}) - E^*(\|\mathbf{x}_{1n} - \mathbf{x}_{2n}\||\mathbf{x}_{2n}) + E^*(\|\mathbf{x}_{1n} - \mathbf{x}_{2n}\|)$, $h_{1n}^{(1)}(\mathbf{z}_{1n}^*) = E\{h_1(\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*, \mathbf{z}_{3n}^*, \mathbf{z}_{4n}^*)|\mathbf{z}_{1n}^*\} = 4\{1 - 2F_{\eta_{1n}^*}(\eta_{1n}^*))\}E^*[\{\mathbf{g}(\mathbf{x}_{1n}) - \mathbf{g}(\mathbf{x}_{2n})\}C_{\mathbf{x}}^*(\mathbf{x}_{1n},$

$\mathbf{x}_{2n})|\mathbf{x}_{1n}]$, and $f^*_{\eta^{(2)}_{(1)}}(\cdot)$ and $F_{\eta^*_{1n}}(\cdot)$ are the pdf and cdf of $\eta^*_{1n}-\eta^*_{2n}$ and η^*_{1n}, respectively.

Let $\widetilde{\mathbf{z}}_i=(\widetilde{\varepsilon}_i,\mathbf{x}_i)$, where $\widetilde{\varepsilon}_i$ has the same distribution as $\varepsilon_i=m(\mathbf{x}_i)-\mathbf{g}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_i$ and is independent of $\mathbf{x}_i$. To obtain the desired results, our next goal is to show that

$$
\begin{aligned}
\boldsymbol{\Sigma}^*_n-\boldsymbol{\Sigma} &= o^*_p(1), &\text{(S1.11)}\\
\boldsymbol{\Lambda}^*_n-\boldsymbol{\Lambda} &= o^*_p(1), &\text{(S1.12)}\\
f^*_{\eta^{(2)}_{(1)}}(\cdot)-f_{\widetilde{\varepsilon}^{(2)}_{(1)}}(\cdot) &= o^*_p(1), &\text{(S1.13)}\\
n^{-1/2}\sum_{i=1}^n\mathbf{g}(\mathbf{x}_{in})\eta^*_{in}-n^{-1/2}\sum_{i=1}^n\mathbf{g}(\mathbf{x}_i)\widetilde{\varepsilon}_i &= o^*_p(1), &\text{(S1.14)}\\
n^{-1/2}\sum_{i=1}^n h^{(1)}_{1n}(\mathbf{z}^*_{in})-n^{-1/2}\sum_{i=1}^n h^{(1)}_1(\widetilde{\mathbf{z}}_i) &= o^*_p(1), &\text{(S1.15)}\\
n^{-1}\sum_{i=1}^n h^{(2)}_{0n}(\mathbf{z}^*_{in},\mathbf{z}^*_{in})-E\{h^{(2)}_0(\widetilde{\mathbf{z}}_i,\widetilde{\mathbf{z}}_i)\} &= o^*_p(1), &\text{(S1.16)}\\
n^{-1}\sum_{i\neq j} h^{(2)}_{0n}(\mathbf{z}^*_{in},\mathbf{z}^*_{jn})-n^{-1}\sum_{i\neq j} h^{(2)}_0(\widetilde{\mathbf{z}}_i,\widetilde{\mathbf{z}}_j) &= o^*_p(1). &\text{(S1.17)}
\end{aligned}
$$

Since $\{\mathbf{x}_{in}\}_{i=1}^n$ is an independent and identically distributed sample from $\{\mathbf{x}_i\}_{i=1}^n$, $\boldsymbol{\Sigma}^*_n=n^{-2}\sum_{i=1}^n\sum_{j=1}^n\mathbf{g}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_j)^{\mathrm{T}}$ and $\boldsymbol{\Lambda}^*_n=-n^{-2}\sum_{i=1}^n\sum_{j=1}^n\mathbf{g}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_j)^{\mathrm{T}}\|\mathbf{x}_i-\mathbf{x}_j\|-n^{-4}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n\mathbf{g}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_j)^{\mathrm{T}}\|\mathbf{x}_k-\mathbf{x}_l\|+2n^{-3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\mathbf{g}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_j)^{\mathrm{T}}\|\mathbf{x}_i-\mathbf{x}_k\|$. By the law of large numbers for V-statistics, (S1.11) and (S1.12) hold.

For any $t\in\mathbb{R}$, given the data, the characteristic function of $\eta^*_1-\eta^*_2$ is $E^*[\exp\{it(\eta^*_1-\eta^*_2)\}]=n^{-2}\sum_{i,j=1}^n\exp\{it(\eta_{in}-\eta_{jn})\}$, which further equals

$n^{-2}\sum_{i,j=1}^{n}\exp\{it(\varepsilon_i-\varepsilon_i)\}+n^{-2}\sum_{i,j=1}^{n}\exp\{it(\varepsilon_i-\varepsilon_i)\}(\exp[-it(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_i)-\mathbf{g}(\mathbf{x}_j)\}]-1])$ via recalling that $\eta_{in}-\eta_{jn}-(\varepsilon_i-\varepsilon_i)=-(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_i)-\mathbf{g}(\mathbf{x}_j)\}$. Condition F yields $\max_{1\le i,j\le n}|(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_i)-\mathbf{g}(\mathbf{x}_j)\}|\le\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|\max_{1\le i,j\le n}\|\mathbf{g}(\mathbf{x}_i)-\mathbf{g}(\mathbf{x}_j)\|=o_p(1)$. Thus,

$$\begin{aligned}E^*[\exp\{it(\eta_1^*-\eta_2^*)\}] &= n^{-2}\sum_{i,j=1}^{n}\exp\{it(\varepsilon_i-\varepsilon_j)\}+o_p^*(1)\\ &= E[\exp\{it(\widetilde{\varepsilon}_1-\widetilde{\varepsilon}_2)\}]+o_p^*(1).\end{aligned}$$

Together with the Glivenko-Cantelli theorem almost surely, $\eta_1^*-\eta_2^*\to\widetilde{\varepsilon}_1-\widetilde{\varepsilon}_2$ in distribution. Combining continuity theorem and Sultsky's theorem, we obtain (S1.13).

Similarly,

$$\begin{aligned}&E^*[\exp\{it\mathbf{g}(\mathbf{x}_{1n})\eta_{1n}^*\}]\\ =\;& n^{-2}\sum_{i,j=1}^{n}\exp\{it\mathbf{g}(\mathbf{x}_i)\varepsilon_j\}+n^{-2}\sum_{i,j=1}^{n}\exp\{it\mathbf{g}(\mathbf{x}_i)\varepsilon_j\}[\exp\{-it(\boldsymbol{\beta}_n-\boldsymbol{\beta}_0)^{\mathrm{T}}\{\mathbf{g}(\mathbf{x}_j)\}-1]\\ =\;& n^{-2}\sum_{i,j=1}^{n}\exp\{it\mathbf{g}(\mathbf{x}_i)\varepsilon_j\}+o_p^*(1)=E[\exp\{it\mathbf{g}(\mathbf{x}_1)\widetilde{\varepsilon}_1\}]+o_p^*(1).\end{aligned}$$

From the Glivenko-Cantelli theorem, we have almost surely, $\mathbf{g}(\mathbf{x}_{1n})\eta_{1n}^*\to\mathbf{g}(\mathbf{x}_1)\widetilde{\varepsilon}_1$ in distribution. It is noted that $E\|\mathbf{g}(\mathbf{x}_{1n})\eta_{1n}^*\|^2=n^{-2}\sum_{i,j=1}^{n}\|\mathbf{g}(\mathbf{x}_i)\varepsilon_j\|^2+n^{-2}\sum_{i,j=1}^{n}\|\mathbf{g}(\mathbf{x}_i)\|^2(\eta_{jn}^2-\varepsilon_j^2)$ and

$$\begin{aligned}\max_{1\le j\le n}|\eta_{jn}^2-\varepsilon_j^2| &\le 2\max_{1\le j\le n}\{\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|^2\max_{1\le i,j\le n}\|\mathbf{g}(\mathbf{x}_j)\|^2+\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|\|\mathbf{g}(\mathbf{x}_j)\||\varepsilon_j|\}\\ &= \{1+o_p(1)\}\max_{1\le j\le n}\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_0\|\|\mathbf{g}(\mathbf{x}_j)\||\varepsilon_j|.\end{aligned}$$

By Condition F, $E\{\|\mathbf{g}(\mathbf{x}_j)\||\varepsilon_j|\}^{2+c} < \infty$ for some $c > 0$. Therefore, $\max_{1\le j\le n} \|\mathbf{g}(\mathbf{x}_j)\||\varepsilon_j| = o_p(n^{-1/2})$ implying $\max_{1\le j\le n} |\eta_{jn}^2 - \varepsilon_j^2| = o_p(1)$. That is, $E\|\mathbf{g}(\mathbf{x}_{1n})\eta_{1n}^*\|^2 = E\|\mathbf{g}(\mathbf{x}_1)\widetilde{\varepsilon}_1\|^2 + o_p^*(1)$. Consequently,

$$E^*\{\mathbf{g}(\mathbf{x}_{1n})\eta_{1n}^* - \mathbf{g}(\mathbf{x}_1)\widetilde{\varepsilon}_1\}^2 = o_p^*(1).$$

Also, $E^*\{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^* - n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_i)\widetilde{\varepsilon}_i\}^2 = E^*\{n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^* - E^* n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^* + E^* n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_{in})\eta_{in}^* - n^{-1/2}\sum_{i=1}^n \mathbf{g}(\mathbf{x}_i)\widetilde{\varepsilon}_i\}^2 = E^*\{\mathbf{g}(\mathbf{x}_{1n})$ $\eta_{1n}^* - \mathbf{g}(\mathbf{x}_1)\widetilde{\varepsilon}_1\}^2 = o_p^*(1)$. This shows (S1.14).

Likewise, we can show for any $t_1, t_2, t_3, t_4 \in \mathbb{R}^{d+1}$, $E^*\{\exp(i\boldsymbol{t}_1^\top\mathbf{z}_{1n}^* + i\boldsymbol{t}_2^\top\mathbf{z}_{2n}^* + i\boldsymbol{t}_3^\top\mathbf{z}_{3n}^* + i\boldsymbol{t}_4^\top\mathbf{z}_{4n}^*)\} = E\{\exp(i\boldsymbol{t}_1^\top\widetilde{\mathbf{z}}_1 + i\boldsymbol{t}_2^\top\widetilde{\mathbf{z}}_2 + i\boldsymbol{t}_3^\top\widetilde{\mathbf{z}}_3 + i\boldsymbol{t}_4^\top\widetilde{\mathbf{z}}_4)\} + o_p^*(1)$. Invoking the Glivenko-Cantelli theorem yields almost surely, $(\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*, \mathbf{z}_{3n}^*, \mathbf{z}_{4n}^*)$ $\to (\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2, \widetilde{\mathbf{z}}_3, \widetilde{\mathbf{z}}_4)$ in distribution. By the continuous mapping theorem, it follows $h_{1n}^{(1)}(\mathbf{z}_{1n}^*) \to h_1^{(1)}(\widetilde{\mathbf{z}}_1)$ $h_{0n}^{(2)}(\mathbf{z}_{1n}^*, \mathbf{z}_{1n}^*) \to h_0^{(2)}(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_1)$ and $h_{0n}^{(2)}(\mathbf{z}_{1n}^*, \mathbf{z}_{1n}^*) \to h_0^{(2)}(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2)$ in distribution. Recalling the definitions of $h_{1n}^{(1)}$ and $h_{0n}^{(2)}$ described at the beginning of the proof, $E^*\{h_{1n}^{(1)}(\mathbf{z}_{1n}^*)\}^2 = E\{h_1^{(1)}(\widetilde{\mathbf{z}}_1)\}^2$, $E^*\{h_{0n}^{(2)}$ $(\mathbf{z}_{1n}^*, \mathbf{z}_{1n}^*)\}^2 = E\{h_0^{(2)}(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_1)\}^2$ and $E^*\{h_{0n}^{(2)}(\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*)\}^2 = E\{h_0^{(2)}(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2)\}^2$, which implies $E^*\{h_{1n}^{(1)}(\mathbf{z}_{1n}^*) - h_1^{(1)}(\widetilde{\mathbf{z}}_1)\}^2 = o_p^*(1)$, $E^*\{h_{0n}^{(2)}(\mathbf{z}_{1n}^*, \mathbf{z}_{1n}^*) - h_0^{(2)}(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_1)$ $\}^2 = o_p^*(1)$ and $E^*\{h_{0n}^{(2)}(\mathbf{z}_{1n}^*, \mathbf{z}_{2n}^*) - h_0^{(2)}(\widetilde{\mathbf{z}}_1, \widetilde{\mathbf{z}}_2)\}^2 = o_p^*(1)$. Combining the uniform integrable theorem and the lemma 5.2.1A of Serfling (1980), we immediately obtain (S1.15), (S1.16) and (S1.17). Together with (S1.11)–(S1.17), the desired result easily follows from exact analogy to the proof of part I of Theorem 1.

□

S2 More numerical results

S2.1 Some aspects of limiting distributions

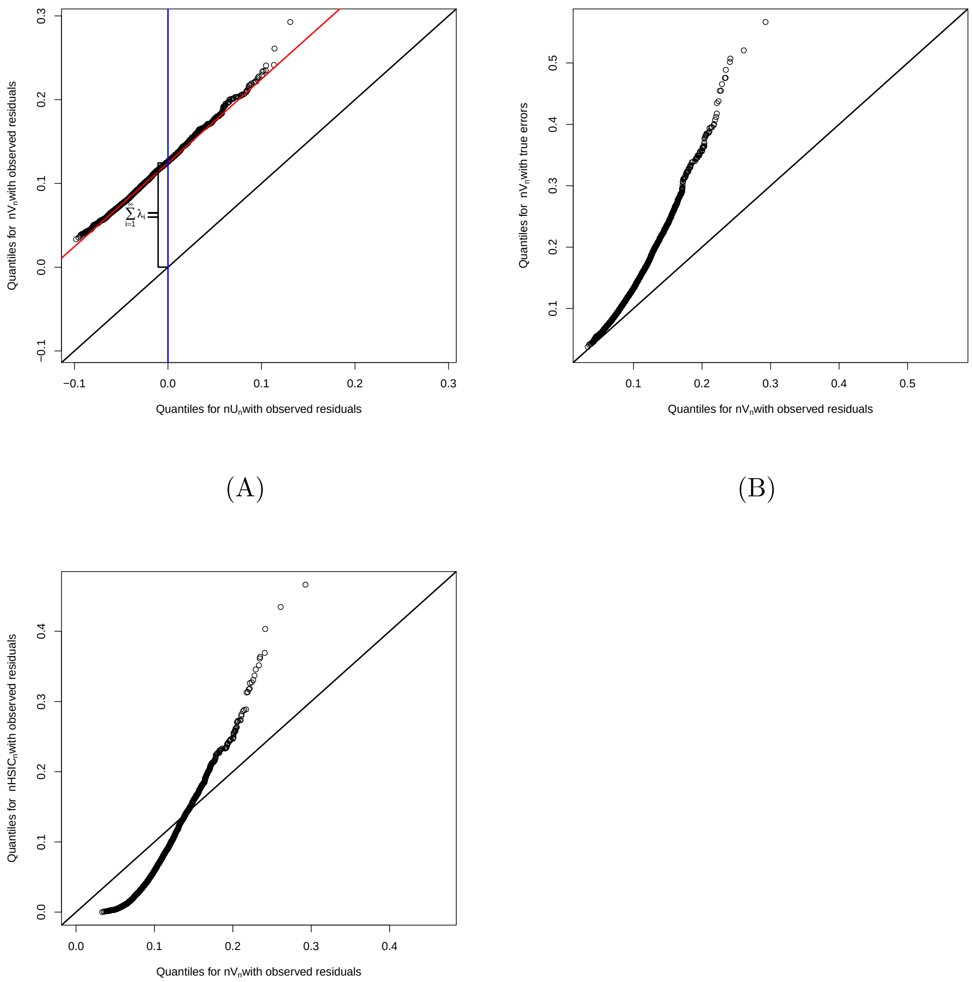
We carry out simulations to assess some aspects of limiting distributions in the main text. We use quantile-quantile plots to compare the null distributions of nU_n and nV_n in which the estimated residuals are used as well as their oracle versions in which the unobserved true errors are used. As a simple comparison, we here include the HSIC test ($nHSIC_n$ for short) developed in Sen and Sen (2014).

Consider the linear model

$$Y = 1 + 2X + \eta, \tag{S2.18}$$

where X is from the uniform distribution on the unit interval, η is normally distributed with mean 0 and variance 0.01 and η is independent of X. We consider the sample size n=100 and the predictor vector $\mathbf{x} = (1, X)$.

The Q-Q plots obtained by simulating 5000 Monte Carlo samples are shown in Figure 1. It can be seen from Figure 1(A) and Figure 1(B) that when generalizing Székely, Rizzo and Bakirov (2007)'s independent



(A) (B)

(C)

Figure 1: *Quantile-quantile plots of 5000 realizations of (A):nU_n obtained using the observed residuals versus nV_n obtained using the observed residuals in the linear model (S2.18); (B):nV_n obtained using the observed residuals versus nV_n obtained using the true unknown errors in the linear model; (C): nV_n obtained using the observed residuals versus $nHSIC_n$ obtained using the observed residuals in the same model.*

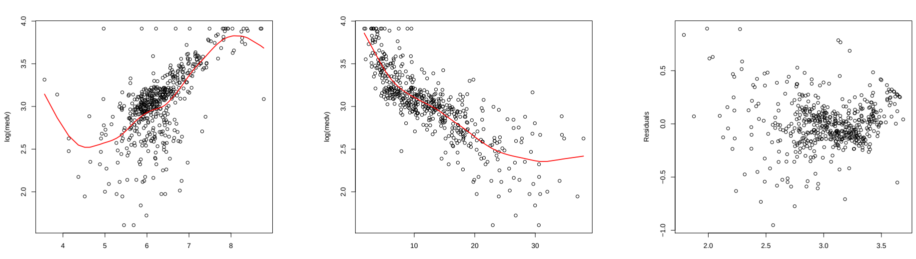
test of no-effect model to the test of the lack-of-fit of a regression model, the limiting distributions of the test statistic of Székely, Rizzo and Bakirov (2007) and its unbiased version will change. From Figure 1(A), we further observe that the black 45° reference line is parallel to the red line. This indicates that though replacing the unobserved true errors by the observed residuals turns out to have an effect, difference of asymptotic null distributions between nU_n and nV_n only relies on the constant $\sum_{i=1}^{\infty} \lambda_i = E(| \eta_1 - \eta_2 |)E(\|\mathbf{x}_1 - \mathbf{x}_2\|) \approx 0.125$. These simulation results are in line with these theoretical findings, as suggested by our Theorem 1. Additionally, the simulation results reported in Figure 1(C) illustrate clearly difference of the limiting distributions for nV_n and $nHSIC_n$. This is anticipated in that both our proposals and Sen and Sen (2014)'s test are not asymptotically distribution-free and depend in a complicated way on data generating process and kernels.

S2.2 Data illustration

We illustrate the performance of our proposals through a real data analysis on the well-known Boston housing data collected by Harrison and Rubinfeld (1978). The original data have been taken from the UCI Repository Of Machine Learning Databases at `http://www.ics.uci.edu/~mlearn/`

`MLRepository.html`. This data have been analyzed by many researchers, such as Kong and Xia (2012), Fan, Ma and Dai (2014), and Sen and Sen (2014). However, we here consider the version of the dataset that incorporates the minor corrections and is available in R package "mlbench". The corrected data consist of 506 observations on 13 variables, with each observation corresponding to one census tract. The response variable Y is medv (median value in \$1,000s of owner-occupied homes in a given area).

To preprocess the data, following the suggestion by Kong and Xia (2012), we take logarithm to the response variable. In particular, Kong and Xia (2012) revealed that among the 13 variables, the variables rm and lstat are very influential factors on prices for house. As an illustration,



(A) (B) (C)

Figure 2: *(A) Plot of* $\log(medv)$ *against factor rm. (B) Plot of* $\log(medv)$ *against factor lstat. (C) Plot of the residuals against the fitted values.*

we consider the 3-dimensional predictor vector $\mathbf{x} = (1, X_1, X_2)$, where X_1

is the variable rm and X_2 is the variable lstat. Many empirical results from the analysis of the data tend to exhibit a nonlinear trend between the response and the predictors, and an error variance heteroscedasticity; see Harrison and Rubinfeld (1978), Kong and Xia (2012) and Sen and Sen (2014). To demonstrate the non-linear dependence, we report the scatter-plots of $\log(medv)$ versus these two factors with cubic spline fit curves in Figure 2 (A) and Figure 2 (B), respectively. To demonstrate the error variance heteroscedasticity, we report the residual plot for the fitted model in Figure 2(C). These results are clearly in line with the conclusions available in the literature.

It is reasonable to assume alternative hypothesis that both the assumption of independence between the error and predictor variables and the goodness-of-fit of the parametric model are violated, to be true. Therefore, different tests that jointly check parametric specification and independence in linear regression models can be compared on the basis of their power functions. As developed in Theorems 1 and 3 in the main text, the proposed bootstrap method can also be used to provide p-values for simultaneously testing independence and goodness-of-fit in linear models. We compare our proposals U_n and V_n with HSIC-based tests $HSIC(1)$ and $mHSIC$ in that to the best of our knowledge, the only existing method can be directly used for implementing a joint test for parametric specification and independence in linear regression models. As expected, all the tests U_n, V_n, $HSIC(1)$

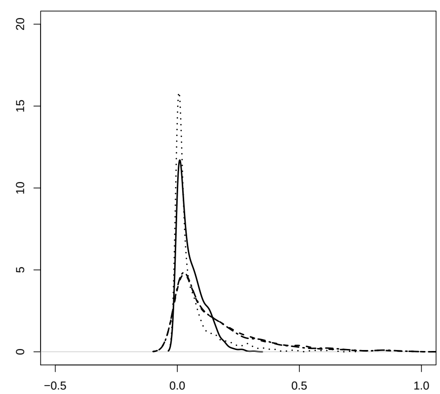


Figure 3: *The density functions of p-values of the tests* U_n *(solid line),* V_n *(dotted line),* $HSIC(1)$ *(dashed line) and* $mHSIC$ *(twodash line) in the analysis of the Boston housing data with the subset size* $N = 50$ *chosen from the whole dataset.*

and $mHSIC$ yield p-values of essentially 0. However, if we use the whole dataset for testing, any test will either reject null hypothesis or accept it. In other words, based on that single experiment, it is not possible to compare among different test procedures. Borrowing the idea from section 5 of Biswas, Mukhopadhyay and Ghosh (2014), we repeat the experiment 1000 times based on 1000 random subsets of the same size chosen from the whole data set. Let N denote the subset size and consider $N = 50$ here. The density functions of p-values of U_n, V_n, $HSIC(1)$ and $mHSIC$ with corresponding powers being 0.578, 0.668, 0.370 and 0.384 at significance level 0.05, are reported in Figure 3, from which we see that p-values of our tests fluctuate more closely around the origin.

Bibliography

Biswas, M., Mukhopadhyay, M. and Ghosh, A. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika*, 101, 913-926.

Fan, J., Ma, Y. and Dai, W. (2014). Nonparametric independence screening in sparse ultra-high dimensional varying coefficient models. *Journal of the American Statistical Association*, 109, 1270-1284.

Harrison, D. and Rubinfeld, D. (1978). Hedonic prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81-102.

Kong, E. and Xia, Y. C. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, 28, 730-768.

Sen, A. and Sen, B. (2014). Testing independence and goodness-of-fit in linear models. *Biometrika*, 101, 927-942.

Serfling, R. L. (1980). *Approximation Theorems in Mathematical Statistics.* New York: Wiley.

Székely, G., Rizzo, M., and Bakirov, N. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics,* 35, 2769-2794.

Yao, S., Zhang, X., and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of the Royal Statistical Society: Series B*, 80, 455-480.