ONLINE SUPPLEMENT TO A NONPARAMETRIC TEST OF SIGNIFICANT VARIABLES IN GRADIENTS

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
Email: feng.yao@mail.wvu.edu
Voice: +1 304 293 7867

Taining Wang

International School of Economics and Management,
Capital University of Economics and Business,
Beijing, 100070, China.
Email: taining.wang@cueb.edu.cn

Last update: August 16, 2020

Abstract

This online supplement provides detailed proofs and simulation results for "A nonparametric test of significant variables in gradients." In Appendix 2, we provide proofs of Lemma 1, parts (2) and (3) in Theorem 1, Corollary 1, and Theorem 3. We also state Lemma 2 for reference. In Appendix 3, Section C.1 offers simulation results of our centered and un-centered bootstrap tests in the bivariate regression. We also check the robustness of the tests' performance by varying the scaling factor of the bandwidth. Section C.2 provides the tests' performance with bandwidths selected through a data-driven cross-validation criterion. Section C.3 reports simulation results in the trivariate regression with different scaling factors of the bandwidth. Finally, Section C.4 provides an extension of our tests to a partially linear model, for which the *curse of dimensionality* issue is less severe.

Keywords: Gradients; Nonparametric significance test; Local polynomial regression.

JEL Classifications: C14, C21.

B Appendix 2

Lemma 1. *Uniformly for all $w \in \mathcal{W}$, the support of W, which is a compact subset of $\Re^d$,*

$$\hat{g}_X(w) - g_X(w) = \frac{1}{nh^{d+1}f(w)}\sum_{i=1}^{n} SK(\frac{W_i - w}{h})(\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(w+\lambda(W_i - w))(\frac{W_i - w}{h})^{\mathbf{k}} + \epsilon_i)(1+o_p(1)).$$

Proof. The first order conditions from minimizing the multivariate weighted least square criterion lead to the following set of equations for $0 \leq |\mathbf{j}| \leq p$,

$$\frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{W_i - w}{h}\right)\left(\frac{W_i - w}{h}\right)^{\mathbf{j}} Y_i \equiv t_{n,\mathbf{j}}(w) = \sum_{0\le|\mathbf{k}|\le p} \hat{a}_{\mathbf{k}} h^{|\mathbf{k}|} S_{n,\mathbf{j}+\mathbf{k}}(w), \tag{B.1}$$

where $S_{n,\mathbf{j}}(w) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{\mathbf{j}}$.

From the first order condition in equation (B.1), we arrange the $N_{|\mathbf{j}|}$ values of $t_{n,\mathbf{j}}(w)$ in a column vector $\tau_{n,|\mathbf{j}|}(w)$, with the k-th element being $(\tau_{n,|\mathbf{j}|})_k(w) = t_{n,G_{|\mathbf{j}|}(k)}(w)$. For $N = \sum_{i=0}^{p} N_i$, we define $\tau_n(w) = \begin{bmatrix} \tau_{n,0}(w) \\ \tau_{n,1}(w) \\ \vdots \\ \tau_{n,p}(w) \end{bmatrix}$. We arrange the distinct values of $h^{|\mathbf{k}|}\hat{a}_{\mathbf{k}}(w)$ for $0 \le |\mathbf{k}| \le p$ as an $N \times 1$ column vector $\hat{\alpha}_n(w) = \begin{bmatrix} \hat{\alpha}_{n,0}(w) \\ \hat{\alpha}_{n,1}(w) \\ \vdots \\ \hat{\alpha}_{n,p}(w) \end{bmatrix}$, where $(\hat{\alpha}_{n,|\mathbf{j}|})_k(w) = h^{|\mathbf{j}|}\hat{a}_{n,G_{|\mathbf{j}|}(k)}(w)$. For the true values, $\alpha(w) = \begin{bmatrix} \alpha_0(w) \\ \alpha_1(w) \\ \vdots \\ \alpha_p(w) \end{bmatrix}$, where $(\alpha_{|\mathbf{j}|})_k(w) = h^{|\mathbf{j}|}a_{G_{|\mathbf{j}|}(k)}(w)$. Since we arrange $w = (x, z')'$ with x being the first element, $g_X(w) = a_{G_1(d)}(w)$. Next we arrange the possible values of $S_{n,\mathbf{j}+\mathbf{k}}(w)$ by a matrix $S_{n,|\mathbf{j}|,|\mathbf{k}|}(w)$ in a lexicographical order with the $(l,m)th$ element being $[S_{n,|\mathbf{j}|,|\mathbf{k}|}(w)]_{l,m} = S_{n,G_{|\mathbf{j}|}(l)+G_{|\mathbf{k}|}(m)}(w)$. So $S_{n,|\mathbf{j}|,|\mathbf{k}|}(w)$ has dimension $N_{|\mathbf{j}|} \times N_{|\mathbf{k}|}$. Define $S_n(w) = \begin{bmatrix} S_{n,0,0}(w) & S_{n,0,1}(w) & \cdots & S_{n,0,p}(w) \\ S_{n,1,0}(w) & S_{n,1,1}(w) & \cdots & S_{n,1,p}(w) \\ \vdots & \vdots & \cdots & \vdots \\ S_{n,p,0}(w) & S_{n,p,1}(w) & \cdots & S_{n,p,p}(w) \end{bmatrix}$.

The set of equations in (B.1) are $\tau_n(w) = S_n(w)\hat{\alpha}_n(w)$. Assuming that $S_n(w)$ is positive definite, the solution is expressed as $\hat{\alpha}_n(w) = S_n^{-1}(w)\tau_n(w)$. Specifically, let $e_{N,1+d}$ be an $N \times 1$ vector of zeros, except one at its $(1+d)th$ position. Then

$$h\hat{g}_X(w) = e'_{N,1+d} S_n^{-1}(w)\tau_n(w).$$

We note that $S_{n,G_{|\mathbf{j}|}(l)+G_{|\mathbf{k}|}(m)}(w) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{G_{|\mathbf{j}|}(l)+G_{|\mathbf{k}|}(m)}$. By assumption A3(2), for $w^* = \lambda W_i + (1-\lambda)w$ with $\lambda \in (0,1)$,

$$Y_i = m(W_i) + \epsilon_i = \sum_{0\le|\mathbf{k}|\le p} \frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w)(W_i - w)^{\mathbf{k}} + \sum_{|\mathbf{k}|=p+1} \frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i - w)^{\mathbf{k}} + \epsilon_i.$$

$$
\begin{aligned}
t_{n,\mathbf{j}}(w) = & \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{\mathbf{j}}\Big[\sum_{0\le|\mathbf{k}|\le p}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w)(W_i-w)^{\mathbf{k}}+\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i-w)^{\mathbf{k}}+\epsilon_i\Big] \\
= & \sum_{0\le|\mathbf{k}|\le p}a_{\mathbf{k}}h^{|\mathbf{k}|}S_{n,\mathbf{j}+\mathbf{k}}(w)+\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{\mathbf{j}}\Big[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i-w)^{\mathbf{k}}+\epsilon_i\Big] \\
= & \sum_{0\le|\mathbf{k}|\le p}\hat{a}_{\mathbf{k}}h^{|\mathbf{k}|}S_{n,\mathbf{j}+\mathbf{k}}(w),
\end{aligned}
$$

where the last equality is from equation (B.1). Thus,

$$
\sum_{0\le|\mathbf{k}|\le p}h^{|\mathbf{k}|}(\hat{a}_{\mathbf{k}}(w)-a_{\mathbf{k}}(w))S_{n,\mathbf{j}+\mathbf{k}}(w)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{\mathbf{j}}\Big[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i-w)^{\mathbf{k}}+\epsilon_i\Big]. \tag{B.2}
$$

Then we let $\kappa_{i,|\mathbf{j}|}(\frac{W_i-w}{h})$ be an $N_{|\mathbf{j}|}$ dimensional subvector whose k-th element is $[\kappa_{i,|\mathbf{j}|}(\frac{W_i-w}{h})]_k = K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{G_{|\mathbf{j}|}(k)}$. Furthermore, $\kappa_i(\frac{W_i-w}{h}) = \begin{bmatrix} \kappa_{i,0}(\frac{W_i-w}{h}) \\ \kappa_{i,1}(\frac{W_i-w}{h}) \\ \vdots \\ \kappa_{i,p}(\frac{W_i-w}{h}) \end{bmatrix}$. Thus we express the equations in (B.2) in a matrix format

$S_n(w)(\hat{\alpha}_n(w)-\alpha(w)) = \frac{1}{nh^d}\sum_{i=1}^{n}\kappa_i(\frac{W_i-w}{h})[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i-w)^{\mathbf{k}}+\epsilon_i]$. So we have

$$
h(\hat{g}_X(w)-g_X(w)) = e'_{N,1+d}S_n^{-1}(w)\frac{1}{nh^d}\sum_{i=1}^{n}\kappa_i\left(\frac{W_i-w}{h}\right)\Big[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i-w)^{\mathbf{k}}+\epsilon_i\Big]. \tag{B.3}
$$

We note that the elements of S are simply multivariate moments of $K(\cdot)$, corresponding to the element of S_n (we suppress its dependence on w here). Consider a typical element of $S_n - f(w)S$ as $[S_{n,i,j}-f(w)S_{i,j}]_{l,m} = S_{n,G_i(l)+G_j(m)} - ES_{n,G_i(l)+G_j(m)} + ES_{n,G_i(l)+G_j(m)} - \mu_{k,G_i(l)+G_j(m)}f(w) = I_1 + I_2$. $I_1 = \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{G_i(l)+G_j(m)} - E\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{W_i-w}{h}\right)\left(\frac{W_i-w}{h}\right)^{G_i(l)+G_j(m)}$. With A1(2), A2 and A4(1), we easily obtain $\sup_{w\in\mathcal{W}}|I_1| = O_p((\frac{nh^d}{lnn})^{-\frac{1}{2}})$ (see Lemma 1 in Martins-Filho et al. (2018)). $I_2 = \int K(\Psi)\Psi^{G_i(l)+G_j(m)}(f(w+h\Psi)-f(w))d\Psi = h\int K(\Psi)\Psi^{G_i(l)+G_j(m)}\Psi' f^{(1)}(w^*)d\Psi = O(h)$ uniformly $\forall w\in\mathcal{W}$, by A1(2) and A2(2). So $\sup_{w\in\mathcal{W}}|[S_{n,i,j}-f(w)S_{i,j}]_{l,m}| = O_p(h+(\frac{nh^d}{lnn})^{-\frac{1}{2}}) = o_p(1)$. Furthermore, given A5 and A1(2), we obtain $\sup_{w\in\mathcal{W}}||S_n - Sf(w)|| = o_p(1)$, where $||\cdot||$ refers to the Euclidean norm. Since S is positive definite in A5, the smallest eigenvalue of S is greater than zero, then we have $\sup_{w\in\mathcal{W}}||S_n^{-1}-\frac{1}{f(w)}S^{-1}|| = o_p(1)$.

From equation (B.3), we have

$$h(\hat{g}_X(w) - g_X(w)) = \frac{1}{f(w)}[S^{-1}]_{(1+d),\cdot}\frac{1}{nh^d}\sum_{i=1}^{n}\kappa_i(\frac{W_i - w}{h})[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(w^*)(W_i - w)^{\mathbf{k}} + \epsilon_i](1 + o_p(1)),$$
(B.4)

where $[S^{-1}]_{(1+d),\cdot}$ refers to the $(1+d)th$ row of S^{-1}. Recall that

$$SK(\Psi) = \sum_{0\leq|\mathbf{j}|\leq p}\sum_{i=1}^{N_{|\mathbf{j}|}}[S^{-1}]_{1+d,\sum_{i'=0}^{|\mathbf{j}|-1}N_{i'}+i}K(\Psi)\Psi^{G_{|\mathbf{j}|}(i)}. \tag{B.5}$$

With the definition of $SK(\cdot)$ in equation (B.5), we obtain the claimed result.

□

In our proof, we have made repeated use of the following Lemma 2, which is the same as Theorem 1 in Yao and Martins-Filho (2015). Let $\{Q_i\}_{i=1}^{n}$ be a sequence of independent and identically distributed (iid) random variables and $\phi_n(Q_1, \cdots, Q_k)$ be a symmetric function with $k < n$. We call $\phi_n(Q_1, \cdots, Q_k)$ a kernel function that depends on n and a U-statistic u_n of degree k is defined as

$$u_n = \begin{pmatrix} n \\ k \end{pmatrix}^{-1}\sum_{(n,k)}\phi_n(Q_{i_1}, \cdots, Q_{i_k}), \tag{B.6}$$

where $\sum_{(n,k)}$ denotes the sum over all subsets $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ of $\{1, 2, \cdots, n\}$. Now, let $\phi_{cn}(q_1, \cdots, q_c) = E(\phi_n(Q_1, \cdots, Q_c, Q_{c+1}, \cdots, Q_k)|Q_1 = q_1, Q_2 = q_2, \cdots, Q_c = q_c)$, $\sigma_{cn}^2 = Var(\phi_{cn}(Q_1, \cdots, Q_c))$ and $\theta_n = E(\phi_n(Q_1, \cdots, Q_k))$. In addition, recursively define $h_n^{(1)}(q_1) = \phi_{1n}(q_1) - \theta_n, \cdots, h_n^{(c)}(q_1, \cdots, q_c) = \phi_{cn}(q_1, \cdots, q_c) - \sum_{j=1}^{c-1}\sum_{(c,j)}h_n^{(j)}(q_{i_1}, \cdots, q_{i_j}) - \theta_n$ for $c = 2, \cdots, k$, where the sum $\sum_{(c,j)}$ is over all subsets $1 \leq i_1 < \cdots < i_j \leq c$ of $\{1, \cdots, c\}$. By Hoeffding's H-decomposition we have

$$u_n = \theta_n + \begin{pmatrix} n \\ k \end{pmatrix}^{-1}\sum_{j=1}^{k}\begin{pmatrix} n-j \\ k-j \end{pmatrix}\sum_{(n,j)}h_n^{(j)}(Q_{v_1}, \cdots, Q_{v_j}) = \theta_n + \sum_{j=1}^{k}\begin{pmatrix} k \\ j \end{pmatrix}H_n^{(j)}(Q_{v_1}, \cdots, Q_{v_j}),$$

where $H_n^{(j)}(Q_{v_1}, \cdots, Q_{v_j}) = \begin{pmatrix} n \\ j \end{pmatrix}^{-1}\sum_{(n,j)}h_n^{(j)}(Q_{v_1}, \cdots, Q_{v_j})$. Since u_n can be written as a finite sum of $H_n^{(j)}$, its magnitude can be determined by studying $H_n^{(j)}$. The following result shows that the magnitude of $H_n^{(j)}$ is determined by n and the leading variance σ_{jn}^2 defined above.

Lemma 2. *Let $\{Q_i\}_{i=1}^{n}$ be an iid sequence and u_n be defined as in equation (B.6) such that*

$$u_n = \theta_n + \sum_{j=1}^{k}\begin{pmatrix} k \\ j \end{pmatrix}H_n^{(j)}(Q_{v_1}, \cdots, Q_{v_j}).$$

Then,

(a) $Var\left(H_n^{(j)}\right) = O\left(n^{-j}\sum_{c=1}^{j}\sigma_{cn}^2\right) = O\left(n^{-j}\sigma_{jn}^2\right)$ *and* $H_n^{(j)} = O_p\left((n^{-j}\sigma_{jn}^2)^{\frac{1}{2}}\right)$;

(b) for $1 \le c \le c' \le k$, *we have* $\frac{\sigma_{cn}^2}{c} \le \frac{\sigma_{c'n}^2}{c'}$.

Proof of Theorem 1

To complete the proof, we only need to show that

$$(2)\quad T_{23} = -\frac{2}{nh^{2+d-d_1}}B_{3n} + o_p(nh^{2+\frac{d}{2}}).$$
$$(3)\quad T_{22} = \frac{1}{nh^{2+d-d_1}}B_{2n} + o_p(nh^{2+\frac{d}{2}}).$$

Proof of (2): Recall that $T_{23} = T_{231n} + T_{232n} + T_{233n} + T_{234n}$. Continuing the proof in the paper, we only need to show (a)-(d) below.

(a) $T_{231n} = -\frac{2}{nh^{2+d-d_1}}B_{3n} + o_p(n^{-1}h^{-\frac{d}{2}-2})$ from results (i)-(iii) given below.

(i) When $i \neq j \neq l \neq t$, we apply Lemma 2 and the second equality in (A.3) to have

$$\begin{aligned} T_{231n} = &\ -\frac{2}{n^4}\sum_{\substack{i=1 j=1 l=1 t=1\\ i\neq j\neq l\neq t}}^{n}\sum^{n}\sum^{n}\sum^{n}\frac{1}{h^{2d+2}f(W_i)f(W_l^r;W_i^c)}SK(\frac{W_j-W_i}{h})SK(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_j\epsilon_t(1+o_p(1)) \\ = &\ -\frac{2}{n^4}\sum_{\substack{i=1 j=1 l=1 t=1\\ i\neq j\neq l\neq t}}^{n}\sum^{n}\sum^{n}\sum^{n}\psi_{nijlt}(1+o_p(1)) = T_{231}(1+o_p(1)). \end{aligned}$$

$$T_{231} = -\frac{2}{24}\begin{pmatrix} n \\ 4 \end{pmatrix}^{-1}\sum_{\substack{i=1 j=1 l=1 t=1\\ i\neq j\neq l\neq t}}^{n}\sum^{n}\sum^{n}\sum^{n}\psi_{nijlt}(1+o_p(1)) = -\frac{2}{24}\begin{pmatrix} n \\ 4 \end{pmatrix}^{-1}\sum_{\substack{i=1 j=1 l=1 t=1\\ i<j<l<t}}^{n}\sum^{n}\sum^{n}\sum^{n}\phi_{nijlt}(1+o_p(1)),\ \text{where}$$

$$\begin{aligned} \phi_{nijlt} = &\ \psi_{nijlt} + \psi_{nijtl} + \psi_{niltj} + \psi_{niljt} + \psi_{nitjl} + \psi_{nitlj} + \psi_{njilt} + \psi_{njitl} + \psi_{njlti} + \psi_{njlit} + \psi_{njtli} + \psi_{njtil} \\ &+ \psi_{nlijt} + \psi_{nlitj} + \psi_{nljti} + \psi_{nljit} + \psi_{nltji} + \psi_{nltij} + \psi_{ntijl} + \psi_{ntilj} + \psi_{ntjli} + \psi_{ntjil} + \psi_{ntlij} + \psi_{ntlji} \end{aligned}$$

present a permutation of ψ_{nijlt} across the indices, so that ϕ_{nijlt} is symmetric in the indices. Again, we write $T_{231} = -\frac{2}{24}[\theta_n + \sum_{j=1}^{4}\begin{pmatrix} 4 \\ j \end{pmatrix}H_n^{(j)}(Q_{v_1},\cdots,Q_{v_j})]$. $\theta_n = 0$ and $H_n^{(1)}(Q_{v_1}) = 0$. $V(H_n^{(2)}(Q_{v_1},Q_{v_2})) = O(n^{-2}\sigma_{2n}^2)$. $\sigma_{2n}^2 = V(\phi_{2n}(Q_1,Q_2))$, $\phi_{2n}(Q_j,Q_t) = E(\phi_{nijlt}|Q_j,Q_t) = E(\psi_{nijlt}+\psi_{nljit}+\psi_{nitlj}+\psi_{nltij}|Q_j,Q_t)$. Since $E(\psi_{nijlt}|Q_j,Q_t) = \epsilon_t\epsilon_j\int\frac{1}{h^{2d+2}f(W_i)f(W_l^r;W_i^c)}SK(\frac{W_j-W_i}{h})SK(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})f(W_i)f(W_l)dW_idW_l$, $E[E^2(\psi_{nijlt}|Q_j,Q_t)] = O(h^{-4-(d-d_1)})$. Similar arguments apply to the other terms and we obtain $\sigma_{2n}^2 = O(h^{-4-(d-d_1)})$. So we obtain that $H_n^{(2)}(Q_{v_1},Q_{v_2}) = O_p(n^{-1}h^{-2-\frac{(d-d_1)}{2}}) = o_p(n^{-1}h^{-2-\frac{d}{2}})$, since $d_1 \ge 1$. $V(H_n^{(3)}(Q_{v_1},Q_{v_2},Q(v_3))) = O(n^{-3}\sigma_{3n}^2)$. $\sigma_{3n}^2 = V(\phi_{3n}(Q_1,Q_2,Q_3))$, $\phi_{3n}(Q_i,Q_j,Q_l) = E(\phi_{nijlt}|Q_i,Q_j,Q_l) = E(\psi_{ntjli}+\psi_{ntjil}+\psi_{ntlji}+\psi_{ntlij}+\psi_{ntijl}+\psi_{ntilj}+\psi_{nijtl}+\psi_{njitl}+\psi_{njlti}+\psi_{nljti}+\psi_{niltj}+\psi_{nlitj}|Q_i,Q_j,Q_l)$. Since $E(\psi_{ntjli}|Q_i,Q_j,Q_l) = \epsilon_i\epsilon_j\int\frac{1}{h^{2d+2}f(W_t)f(W_l^r;W_t^c)}SK(\frac{W_j-W_t}{h})SK(\frac{W_i^r-W_l^r}{h};\frac{W_i^c-W_t^c}{h})f(W_t)dW_t$, we obtain that $E[E^2(\psi_{ntjli}|Q_i,Q_j,Q_l)] = O(h^{-4-d})$. Similar arguments apply to the other terms and we obtain

$\sigma^2_{3n} = O(h^{-4-d})$. So $H^{(3)}_n(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}}h^{-2-\frac{d}{2}}) = o_p(n^{-1}h^{-2-\frac{d}{2}})$. $V(H^{(4)}_n(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4})) = O(n^{-4}\sigma^2_{4n})$, $\sigma^2_{4n} \le CE\psi^2_{nijlt} = O_p(h^{-4-2d})$ by A2, A3(1), and A1(2), so we obtain $H^{(4)}_n(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4}) = O_p(n^{-2}h^{-2-d}) = o_p(n^{-1}h^{-2-\frac{d}{2}})$. So in all, $T_{231} = o_p(n^{-1}h^{-2-\frac{d}{2}})$ and $T_{231n} = o_p(n^{-1}h^{-2-\frac{d}{2}})$ when $i \neq j \neq l \neq t$.

(ii) When $i \neq j = l \neq t$,

$$\begin{aligned} T_{231n} = & -\frac{2}{n^4}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j\neq t}}^{n}\sum_{t=1}^{n}\frac{\epsilon_t\epsilon_j}{h^{2d+2}f(W_i)f(W^r_j;W^c_i)}SK(\frac{W_j-W_i}{h})SK(\frac{W^r_t-W^r_j}{h};\frac{W^c_t-W^c_i}{h})(1+o_p(1)) = O_p(n^{-1}h^{-2}) \\ = & \; o_p(n^{-1}h^{-2-\frac{d}{2}}). \end{aligned}$$

(iii) When $i \neq j = t \neq l$,

$$T_{231n} \equiv T_{231B} = -\frac{2}{(nh^{2+d-d_1})}\frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j\neq l}}^{n}\sum_{l=1}^{n}\frac{1}{h^{d+d_1}}\hat{SK}(\frac{W_j-W_i}{h})\hat{SK}(\frac{W^r_j-W^r_l}{h};\frac{W^c_j-W^c_i}{h})\epsilon^2_j = -\frac{2}{nh^{2+d-d_1}}B_{3n}.$$

With similar but lengthy arguments, we show that

$$\begin{aligned} \text{(b)}\, T_{232n} = & -\frac{2}{n^4}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j, t\neq l, t\neq i, l\neq i}}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h^{2d+2}f(W_i)f(W^r_l;W^c_i)}SK(\frac{W_j-W_i}{h})SK(\frac{W^r_t-W^r_l}{h};\frac{W^c_t-W^c_i}{h})\epsilon_t \\ & \times\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i))(\frac{W_j-W_i}{h})^{\mathbf{k}}(1+o_p(1)) \\ = & \; o_p(n^{-1}h^{-2-\frac{d}{2}}). \end{aligned}$$

$$\begin{aligned} \text{(c)}\, T_{233n} = & -\frac{2}{n^4}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j, t\neq l, t\neq i, l\neq i}}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h^{2d+2}f(W_i)f(W^r_l;W^c_i)}SK(\frac{W_j-W_i}{h})SK(\frac{W^r_t-W^r_l}{h};\frac{W^c_t-W^c_i}{h})\epsilon_j \\ & \times\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W^r_l+\lambda(W^r_t-W^r_l);W^c_i+\lambda(W^c_t-W^c_i))(\frac{W^r_t-W^r_l}{h};\frac{W^c_t-W^c_i}{h})^{\mathbf{k}}(1+o_p(1)) \\ = & \; o_p(n^{-1}h^{-2-\frac{d}{2}}). \end{aligned}$$

(d) T_{234n}

$$\begin{aligned} = & -\frac{2}{n^4}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j, t\neq l, t\neq i, l\neq i}}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h^{2d+2}f(W_i)f(W^r_l;W^c_i)}SK(\frac{W_j-W_i}{h})SK(\frac{W^r_t-W^r_l}{h};\frac{W^c_t-W^c_i}{h})\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_i+\lambda(W_j-W_i)) \\ & \times(\frac{W_j-W_i}{h})^{\mathbf{k}}\sum_{|\mathbf{k}'|=p+1}\frac{h^{p+1}}{\mathbf{k}'!}(D^{\mathbf{k}'}m)(W^r_l+\lambda(W^r_t-W^r_l);W^c_i+\lambda(W^c_t-W^c_i))(\frac{W^r_t-W^r_l}{h};\frac{W^c_t-W^c_i}{h})^{\mathbf{k}'}(1+o_p(1)) \\ = & \; o_p(n^{-1}h^{-2-\frac{d}{2}}). \end{aligned}$$

The claim in (2) follows from (2)(a)-(d) above.

Proof of (3): Following (2) above, and again $\frac{1}{n-1}-\frac{1}{n} = O(n^{-2})$, we let $D^{\mathbf{k}}m_{tj;ti} = (D^{\mathbf{k}}m)(W^r_j+\lambda(W^r_t-W^r_j);W^c_i+\lambda(W^c_t-W^c_i))(\frac{W^r_t-W^r_j}{h};\frac{W^c_t-W^c_i}{h})^{\mathbf{k}}$, and write

$$
\begin{aligned}
T_{22} = & \frac{1}{n}\sum_{i=1}^{n}[\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}(\hat{g}_X(W_j^r;W_i^c)-g_X(W_j^r;W_i^c))]^2\\
= & \frac{1}{n}\sum_{i=1}^{n}[\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{nh^{d+1}}\sum_{\substack{t=1\\ t\neq j,t\neq i}}^{n}\hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})(\sum_{|\mathbf{k}|=p+1}D^{\mathbf{k}}m_{tj;ti}\frac{h^{p+1}}{\mathbf{k}!}+\epsilon_t)]^2\\
= & \frac{1}{n^3(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\sum_{\substack{m=1\\ t\neq j\neq i,m\neq l\neq i}}^{n}\frac{1}{h^{2d+2}}\hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\hat{SK}(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})[\epsilon_t\epsilon_m\\
& +\epsilon_t\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{ml;mi}+\epsilon_m\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{tj;ti}+\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{tj;ti}\sum_{|\mathbf{k}'|=p+1}\frac{h^{p+1}}{\mathbf{k}'!}D^{\mathbf{k}'}m_{ml;mi}]\\
= & T_{221n}+T_{222n}+T_{223n}+T_{224n}.
\end{aligned}
$$

(a) $T_{221n}\equiv T_{221B}=\frac{1}{nh^{2+d-d_1}}B_{2n}+o_p(n^{-1}h^{-2-\frac{d}{2}})$ from results (i)-(iii) given below.

(i) When $i\neq j\neq t\neq l\neq m$,

$$
\begin{aligned}
& T_{221n}\\
= & \frac{1}{n^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\sum_{\substack{m=1\\ i\neq j\neq t\neq l\neq m}}^{n}\underbrace{\frac{1}{h^{2d+2}f(W_j^r;W_i^c)f(W_l^r;W_i^c)}SK(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})SK(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})\epsilon_t\epsilon_m}_{\psi_{nijtlm}}\\
& \times(1+o_p(1))=T_{221}(1+o_p(1)).
\end{aligned}
$$

We apply Lemma 2 to perform Hoeffding's H-decomposition to have $T_{221}=\theta_n+\sum_{j=1}^{5}\begin{pmatrix}5\\ j\end{pmatrix}H_n^{(j)}(Q_{v_1},\cdots,Q_{v_j})$. We can show that $\theta_n=0$, $H_n^{(1)}(Q_{v_1})=0$, $H_n^{(2)}(Q_{v_1},Q_{v_2})=O_p(n^{-1}\sigma_{2n})=O_p(n^{-1}h^{-2-\frac{d}{2}+\frac{d_1}{2}})=o_p(n^{-1}h^{-2-\frac{d}{2}})$ since $\sigma_{2n}^2=O(h^{-4-(d-d_1)})$, $H_n^{(3)}(Q_{v_1},Q_{v_2},Q_{v_3})=O_p(n^{-\frac{3}{2}}\sigma_{3n})=O_p(n^{-\frac{3}{2}}(h^{-2-(d-d_1)}+h^{-2-\frac{d}{2}}))=o_p(n^{-1}h^{-2-\frac{d}{2}})$ since $\sigma_{3n}^2=O(h^{-4-(2d-2d_1)}+h^{-4-d})$, $H_n^{(4)}(Q_{v_1},Q_{v_2},Q_{v_3},Q_{v_4})=O_p(n^{-2}\sigma_{4n})=O_p(n^{-2}(h^{-2-\frac{(d+d_1)}{2}}+h^{-2-d+\frac{d_1}{2}}))=o_p(n^{-1}h^{-2-\frac{d}{2}})$ since $\sigma_{4n}^2=O(h^{-4-(d+d_1)}+h^{-4-(2d-d_1)})$, $H_n^{(5)}(Q_{v_1},Q_{v_2},Q_{v_3},Q_{v_4},Q_{v_5})=O_p(n^{-\frac{5}{2}}\sigma_{5n})=O_p(n^{-\frac{5}{2}}(h^{-2-d}))=o_p(n^{-1}h^{-2-\frac{d}{2}})$ since $\sigma_{5n}^2=O(h^{-4-2d})$. Thus, $T_{221n}=o_p(n^{-1}h^{-2-\frac{d}{2}})$.

(ii) When $i\neq j=l\neq t=m$, or $i\neq j=m\neq t=l$, or $i\neq j=l\neq t\neq m$, $i\neq j=m\neq t\neq l$, $i\neq j\neq t=l\neq m$, we can show that $T_{221n}=o_p(n^{-1}h^{-2-\frac{d}{2}})$.

(iii) When $i\neq j\neq t=m\neq l$, we can show that with similar arguments as in T_{211B} and T_{231B} that

$$
\begin{aligned}
T_{221n}= & \frac{1}{nh^{2+d-d_1}}\frac{1}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{\substack{l=1\\ i\neq j\neq t\neq l}}^{n}\frac{\epsilon_t^2}{h^{d+d_1}}\hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\hat{SK}(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})\\
\equiv & T_{221B}=\frac{1}{nh^{2+d-d_1}}B_{2n}.
\end{aligned}
$$

(b)
$$
\begin{aligned}
T_{222n}= & \frac{1}{n^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\sum_{\substack{m=1\\ t\neq j\neq i,m\neq l\neq i}}^{n}\frac{1}{h^{2d+2}f(W_j^r;W_i^c)f(W_l^r;W_i^c)}SK(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})SK(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})\epsilon_t\\
& \times\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{ml;mi}(1+o_p(1)).
\end{aligned}
$$

$$
\begin{aligned}
T_{223n}= & \frac{1}{n^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\sum_{\substack{m=1\\ t\neq j\neq i,m\neq l\neq i}}^{n}\frac{1}{h^{2d+2}f(W_j^r;W_i^c)f(W_l^r;W_i^c)}SK(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})SK(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})\epsilon_m\\
& \times\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{tj;ti}(1+o_p(1)).
\end{aligned}
$$

$$T_{224n} = \frac{1}{n^5}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^n\sum_{l=1}^n\sum_{\substack{m=1\\ t\neq j\neq i, m\neq l\neq i}}^n \frac{1}{h^{2d+2}f(W_j^r;W_i^c)f(W_l^r;W_i^c)} SK(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})SK(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})$$
$$\sum_{|\mathbf{k}|=p+1}\frac{h^{p+1}}{\mathbf{k}!}D^{\mathbf{k}}m_{tj;ti}\sum_{|\mathbf{k}'|=p+1}\frac{h^{p+1}}{\mathbf{k}'!}D^{\mathbf{k}'}m_{ml;mi}(1+o_p(1)).$$

We show similarly that $T_{22in} = o_p(n^{-1}h^{-2-\frac{d}{2}})$ for $i = 2, 3$, and 4. The claim in (3) follows from (3)(a) and (b) above. □

Proof of Corollary 1

For the claim to be valid, we only need to show

(i) $\hat{B}_{1n} - B_{1n} = o_p(h^{\frac{d}{2}})$.
(ii) $\hat{B}_{in} - B_{in} = o_p(h^{\frac{d}{2}-d_1})$ for $i = 2$, and 3.
(iii) $\hat{\Omega} - \Omega = o_p(1)$.

The claim that $\hat{T}_c \xrightarrow{d} \mathcal{N}(0,1)$ follows from Theorem 1 and (i)-(iii) above.

Proof of (i): As $\hat{\epsilon}_j^2 = \epsilon_j^2 + 2\epsilon_j(m(W_j) - \hat{m}(W_j)) + (m(W_j) - \hat{m}(W_j))^2$, with the additional assumption A1(3), we have $\sup_{w\in\mathcal{W}}|\hat{m}(w) - m(w)| = O_p(L_n)$ where $L_n = (\frac{lnn}{nh^d})^{\frac{1}{2}} + h^{p+1}$. Then,

$$\hat{B}_{1n} - B_{1n} = \frac{1}{n^2}\sum_{i=1}^n\sum_{\substack{j=1\\ i\neq j}}^n \frac{1}{h^d}\hat{SK}^2(\frac{W_j-W_i}{h})[2\epsilon_j(m(W_j) - \hat{m}(W_j)) + (m(W_j) - \hat{m}(W_j))^2]$$
$$= \frac{1}{n^2}\sum_{i=1}^n\sum_{\substack{j=1\\ i\neq j}}^n \frac{1}{h^d f^2(W_i)}SK^2(\frac{W_j-W_i}{h})[2\epsilon_j(m(W_j) - \hat{m}(W_j)) + O_p(L_n^2)](1+o_p(1)).$$

By $A4(2)$ and $A4(3)$, we have $O_p(L_n^2) = o_p(h^{\frac{d}{2}})$. We define

$$SK_m(\Psi) = \sum_{0\leq|\mathbf{j}|\leq p}\sum_{i=1}^{N_{|\mathbf{j}|}}[S^{-1}]_{1,\sum_{i'=0}^{|\mathbf{j}|-1}N_{i'}+i}K(\Psi)\Psi^{G_{|\mathbf{j}|}(i)}, \quad (B.7)$$

Then $\hat{m}(W_j) - m(W_j) = \frac{1}{f(W_j)}\frac{1}{nh^d}\sum_{\substack{t=1\\ t\neq j}}^n SK_m(\frac{W_t-W_j}{h})[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_{tj}^*)(W_t-W_j)^{\mathbf{k}} + \epsilon_t](1+o_p(1))$, where $W_{tj}^* = W_j + \lambda(W_t - W_j)$.

$$\hat{B}_{1n} - B_{1n} = -\frac{2}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{\substack{t=1\\ i\neq j, t\neq j}}^n \frac{1}{h^{2d}f^2(W_i)f(W_j)}SK^2(\frac{W_j-W_i}{h})SK_m(\frac{W_t-W_j}{h})\epsilon_j[\sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_{tj}^*)(W_t-W_j)^{\mathbf{k}}$$
$$+\epsilon_t] \times (1+o_p(1)) + o_p(h^{\frac{d}{2}}) = -2[B_{1n1} + B_{1n2}](1+o_p(1)) + o_p(h^{\frac{d}{2}}).$$

$$B_{1n2} = \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{\substack{t=1\\ i\neq j, t\neq j}}^n \frac{1}{h^{2d}f^2(W_i)f(W_j)}SK^2(\frac{W_j-W_i}{h})SK_m(\frac{W_t-W_j}{h})\epsilon_j\epsilon_t = \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{\substack{t=1\\ i\neq j, t\neq j}}^n \psi_{nijt}.$$

(a) For the case that $i \neq j \neq t$, let $\phi_{nijt} = \psi_{nijt} + \psi_{nitj} + \psi_{njit} + \psi_{njti} + \psi_{ntij} + \psi_{ntji}$, which is symmetric in i, j and t. Then $B_{1n2} = \frac{1}{6}\begin{pmatrix} n \\ 3 \end{pmatrix}^{-1}\sum_{i=1}^n\sum_{j=1}^n\sum_{\substack{t=1\\ i<j<t}}^n \phi_{nijt}(1+o_p(1))$. We apply Lemma 2. $\theta_n = 0$, $H_n^{(1)}(Q_{v_1}) = 0$. $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}\sigma_{2n}) = O_p(n^{-1}h^{-\frac{d}{2}})$, since $\sigma_{2n}^2 = E(E^2(\phi_{nijt}|Q_j, Q_t)) =$

$O_p(h^{-d})$, as $E(\psi_{nijt}|Q_j,Q_t) = \frac{\epsilon_j\epsilon_t}{h^d f(W_j)} SK_m(\frac{W_t-W_j}{h}) E(\frac{1}{h^d f^2(W_i)} SK^2(\frac{W_j-W_i}{h})|W_j)$. $H_n^{(3)}(Q_{v_1},Q_{v_2},Q_{v_3}) = O_p(n^{-\frac{3}{2}}\sigma_{3n}) = O_p(n^{-\frac{3}{2}}h^{-d})$. So in all, $B_{1n2} = O_p(n^{-1}h^{-\frac{d}{2}}) = o_p(h^{\frac{d}{2}})$.

(b) For the case $i = t \neq j$,

$B_{1n2} = \frac{1}{n}\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} \frac{1}{h^{2d} f^2(W_i)f(W_j)} SK^2(\frac{W_j-W_i}{h}) SK_m(\frac{W_i-W_j}{h})\epsilon_j\epsilon_i = O_p(n^{-2}h^{-\frac{3}{2}d}) = o_p(h^{\frac{d}{2}})$. So in all, $B_{1n2} = o_p(h^{\frac{d}{2}})$ based on (a) and (b) above.

$$\begin{aligned} B_{1n1} = & \frac{1}{n^3}\sum_{\substack{i=1\\i\neq j,t\neq j}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n} \frac{1}{h^{2d} f^2(W_i)f(W_j)} SK^2(\frac{W_j-W_i}{h}) SK_m(\frac{W_t-W_j}{h})\epsilon_j \sum_{|\mathbf{k}|=p+1}\frac{1}{\mathbf{k}!}(D^{\mathbf{k}}m)(W_{tj}^*)(W_t-W_j)^{\mathbf{k}} \\ = & \ o_p(h^{\frac{d}{2}}) \text{ can be show similarly.} \end{aligned}$$

So we have the claimed result that $\hat{B}_{1n} - B_{1n} = o_p(h^{\frac{d}{2}})$.

Proof of (ii): $\hat{B}_{in} - B_{in} = o_p(h^{\frac{d}{2}-d_1})$ for $i = 2$, and 3 can be shown similarly.

Proof of (iii): We only need to show that $\frac{1}{n^2h^d}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\frac{\hat{\epsilon}_i^2\hat{\epsilon}_j^2}{\hat{f}(W_i)\hat{f}(W_j)}K(\frac{W_i-W_j}{h}) - \int\sigma^4(W)DW = o_p(1)$.

Since $\sup_{W_i\in\mathcal{W}}|\frac{1}{\hat{f}(W_i)} - \frac{1}{f(W_i)}| = o_p(1)$, and $\sup_{w\in\mathcal{W}}|\hat{m}(w)-m(w)| = o_p(1)$, $\frac{1}{n^2h^d}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\frac{\hat{\epsilon}_i^2\hat{\epsilon}_j^2}{\hat{f}(W_i)\hat{f}(W_j)}K(\frac{W_i-W_j}{h}) = \frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\underbrace{\frac{\epsilon_i^2\epsilon_j^2}{h^d f(W_i)f(W_j)}K(\frac{W_i-W_j}{h})}_{\psi_{nij}}(1+o_p(1)) = I_\Omega(1+o_p(1))$. We show that $I_\Omega \xrightarrow{p} \int\sigma^4(W)DW$, which will give the claim in (iii). Since ψ_{nij} is symmetric in i,j, $I_\Omega = \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\psi_{nij}(1+o_p(1))$. By Lemma 2, $\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\psi_{nij} = \theta_n + \sum_{j=1}^{2}\begin{pmatrix} 2 \\ j \end{pmatrix}H_n^{(j)}(Q_{v_1},\cdots,Q_{v_j})$. $\theta_n = E\frac{\epsilon_i^2\epsilon_j^2}{h^d f(W_i)f(W_j)}K(\frac{W_i-W_j}{h}) \to \int\sigma^4(W)dW$. $H_n^{(1)}(Q_{v_1}) = O_p(n^{-\frac{1}{2}}\sigma_{1n}) = O_p(n^{-\frac{1}{2}})$ as $\sigma_{1n}^2 = E(E^2(\psi_{nij}|Q_i)) = O(1)$. $H_n^{(2)}(Q_{v_1},Q_{v_2}) = O_p(n^{-1}\sigma_{2n}) = O_p(n^{-1}h^{-\frac{d}{2}})$ as $\sigma_{2n}^2 = E(\psi_{nij}^2) = O(h^{-d})$. So in all, $I_\Omega \xrightarrow{p} \int\sigma^4(W)DW$. □

Proof of Theorem 3

The expressions of $\hat{B}_{jn}^*$ for $j = 1, 2$ and 3, and $\hat{\Omega}^*$ in Theorem 3 are

$\hat{B}_{1n}^* = \frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\frac{\epsilon_j^{*2}}{h^d}\hat{SK}^2(\frac{W_j-W_i}{h})$, $\hat{B}_{3n}^* = \frac{1}{n^2(n-1)}\sum_{\substack{i=1\\i\neq j\neq l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\frac{\epsilon_j^{*2}}{h^{d+d_1}}\hat{SK}(\frac{W_j-W_i}{h})\hat{SK}(\frac{W_j^r-W_l^r}{h};\frac{W_j^c-W_i^c}{h})$,

$\hat{B}_{2n}^* = \frac{1}{n^2(n-1)^2}\sum_{\substack{i=1\\i\neq j\neq t\neq l}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\epsilon_t^{*2}}{h^{d+d_1}}\hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\hat{SK}(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})$.

$\hat{\Omega}^* = \frac{2}{n^2h^d}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\frac{(\epsilon_i^*)^2(\epsilon_j^*)^2}{\hat{f}(W_i)\hat{f}(W_j)}K(\frac{W_i-W_j}{h})[\int(\int SK(\Psi_1+\Psi)SK(\Psi)d\Psi)^2d\Psi_1]$.

(i) Following the arguments in Lemma 1, we define $t_{n,\mathbf{j}}^*(W) = \frac{1}{nh^d}\sum_{i=1}^{n}K(\frac{W_i-W}{h})(\frac{W_i-W}{h})^{\mathbf{j}}\epsilon_i^*$, $(\tau_{n,|\mathbf{j}|}^*)_k =$

$t^*_{n,G_{|\mathbf{j}|}(k)}$, and $\tau^*_n(W) = \begin{bmatrix} \tau^*_{n,0}(W) \\ \tau^*_{n,1}(W) \\ \vdots \\ \tau^*_{n,p}(W) \end{bmatrix}$, then we obtain from the first order condition in equation (B.1) that

$$h\hat{g}^*_X(W) = e'_{N,1+d}S_n(W)^{-1}\tau^*_n(W) = \frac{1}{f(W)}[S^{-1}]_{1+d,\cdot}\tau^*_n(W)(1+o_p(1)).$$

From these, we obtain

$$\hat{g}^*_X(W_i) = \tfrac{1}{nh^{d+1}}\sum_{\substack{j=1\\ j\neq i}}^n \hat{SK}(\tfrac{W_j-W_i}{h})\epsilon^*_j = \tfrac{1}{f(W_i)nh^{d+1}}\sum_{\substack{j=1\\ j\neq i}}^n SK(\tfrac{W_j-W_i}{h})\epsilon^*_j(1+o_p(1)), \text{ and}$$

$$\begin{aligned}\hat{g}^*_X(W^r_l;W^c_i) = &\ \tfrac{1}{nh^{d+1}}\sum_{\substack{j=1\\ j\neq l, j\neq i}}^n \hat{SK}(\tfrac{W^r_j-W^r_l}{h};\tfrac{W^c_j-W^c_i}{h})\epsilon^*_j \\ = &\ \tfrac{1}{f(W^r_l;W^c_i)nh^{d+1}}\sum_{\substack{j=1\\ j\neq l, j\neq i}}^n SK(\tfrac{W^r_j-W^r_l}{h};\tfrac{W^c_j-W^c_i}{h})\epsilon^*_j(1+o_p(1)).\end{aligned}$$

$$\hat{T}^* = \tfrac{1}{n}\sum_{i=1}^n(\hat{g}^*_X(W_i))^2 + \tfrac{1}{n}\sum_{i=1}^n[\tfrac{1}{n-1}\sum_{\substack{l=1\\ l\neq i}}^n\hat{g}^*_X(W^r_l;W^c_i)]^2 - \tfrac{2}{n(n-1)}\sum_{i=1}^n\sum_{\substack{l=1\\ l\neq i}}^n\hat{g}^*_X(W_i)\hat{g}^*_X(W^r_l;W^c_i) = T^*_1+T^*_2+T^*_3.$$

We show below that conditional on $Q_{(n)} = \{W_i, Y_i\}^n_{i=1}$,

$$\begin{aligned}&(1)\quad nh^{2+\frac{d}{2}}(T^*_1 - \tfrac{1}{nh^{2+d}}\hat{B}^*_{1n})/\sqrt{\hat{\Omega}^*} \xrightarrow{d} \mathcal{N}(0,1).\\ &(2)\quad T^*_3 = -\tfrac{2}{nh^{2+d-d_1}}\hat{B}^*_{3n} + o_p((nh^{2+d})^{-1}).\\ &(3)\quad T^*_2 = \tfrac{1}{nh^{2+d-d_1}}\hat{B}^*_{2n} + o_p((nh^{2+d})^{-1}).\end{aligned}$$

The claim of (i) in Theorem 3 follows from (1)-(3).

Proof of (1): $T^*_1 = \tfrac{1}{n}\sum_{i=1}^n[\tfrac{1}{nh^{d+1}}\sum_{\substack{j=1\\ j\neq i}}^n\hat{SK}(\tfrac{W_j-W_i}{h})\epsilon^*_j]^2 = \tfrac{1}{n^3}\sum_{\substack{i=1\\ i\neq j, i\neq l}}^n\sum_{j=1}^n\sum_{l=1}^n\tfrac{1}{h^{2(d+1)}}\hat{SK}(\tfrac{W_j-W_i}{h})\hat{SK}(\tfrac{W_l-W_i}{h})\epsilon^*_j\epsilon^*_l.$

(a) When $i\neq j\neq l$, we note that the U-statistic result in Lemma 2 can not be applied here as we do not have the iid assumption conditioning on the data $Q_{(n)}$.

$$T^*_1 = \tfrac{1}{n^3}\sum_{\substack{i=1\\ i\neq j\neq l}}^n\sum_{j=1}^n\sum_{l=1}^n\underbrace{\frac{1}{f^2(W_i)h^{2(d+1)}}SK(\frac{W_j-W_i}{h})SK(\frac{W_l-W_i}{h})\epsilon^*_j\epsilon^*_l}_{\psi_{nijl}}(1+o_p(1)).$$

We let $\phi_{nijl} = \psi_{nijl}+\psi_{njil}+\psi_{nlji}$, which is symmetric in i,j,l. We rewrite T^*_1 as

$$\begin{aligned}T^*_1 = &\ (1+o_p(1))\{\tfrac{1}{3}\begin{pmatrix} n\\ 3\end{pmatrix}^{-1}\begin{pmatrix} n-2\\ 1\end{pmatrix}\sum_{\substack{i=1\\ i<j}}^n\sum_{j=1}^n\phi_{2nij} + \tfrac{1}{3}\begin{pmatrix} n\\ 3\end{pmatrix}^{-1}[\sum_{\substack{i=1\\ i<j<l}}^n\sum_{j=1}^n\sum_{l=1}^n\phi_{nijl} - \begin{pmatrix} n-2\\ 1\end{pmatrix}\sum_{\substack{i=1\\ i<j}}^n\sum_{j=1}^n\phi_{2nij}]\}\\ = &\ (1+o_p(1))\{T^*_{1a}+T^*_{1b}\},\end{aligned}$$

where $\phi_{2nij} = \int\psi_{nlji}f(W_l)dW_l = \frac{\epsilon^*_j\epsilon^*_i}{h^{2(d+1)}}\int\frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l.$

(I) Claim: $T^*_{1b} = o_p(n^{-1}h^{-2-\frac{d}{2}})$.

$$T^*_{1b} = \tfrac{1}{3}\begin{pmatrix} n\\ 3\end{pmatrix}^{-1}\sum_{\substack{i=1\\ i<j<l}}^n\sum_{j=1}^n\sum_{l=1}^n[\phi_{nijl}-\phi_{2nij}-\phi_{2nil}-\phi_{2njl}] = \tfrac{1}{3}\begin{pmatrix} n\\ 3\end{pmatrix}^{-1}\sum_{\substack{i=1\\ i<j<l}}^n\sum_{j=1}^n\sum_{l=1}^n\Phi_{nijl}.$$

Note that $E^*(\epsilon_i^*) \equiv E(\epsilon_i^*|Q_{(n)}) = 0$, so $E(T_{1b}^*|Q_{(n)}) = 0$.

$$\begin{aligned} V(T_{1b}^*|Q_{(n)}) = & \frac{1}{9}\left(\begin{array}{c} n \\ 3 \end{array}\right)^{-2} \sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{\substack{i'=1 \\ i'<j'<l'}}^{n}\sum_{j'=1}^{n}\sum_{l'=1}^{n} E(\Phi_{nijl}\Phi_{ni'j'l'}|Q_{(n)}) \\ = & \frac{1}{9}\left(\begin{array}{c} n \\ 3 \end{array}\right)^{-2} \sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} E(\Phi_{nijl}^2|Q_{(n)}) + V_{1b1}. \end{aligned}$$

First, consider V_{1b1}. Note that if (i,j,l) are each distinct from (i',j',l'), since ϵ_i^* is independent conditioning on $Q_{(n)}$, then $E(\Phi_{nijl}\Phi_{ni'j'l'}|Q_{(n)}) = E(\Phi_{nijl}|Q_{(n)})E(\Phi_{ni'j'l'}|Q_{(n)}) = 0$. Similarly, if only one index in (i,j,l) is the same as that in (i',j',l'), $E(\Phi_{nijl}\Phi_{ni'j'l'}|Q_{(n)}) = 0$. So in V_{1b1}, we only consider the case that two of the indices in (i,j,l) are the same as that in (i',j',l'). Due to the symmetry in Φ_{nijl}, we can just consider any two indices. Consider $i = i' \neq j = j' \neq l \neq l'$, $E(\Phi_{nijl}\Phi_{nijl'}|Q_{(n)}) = E[(\psi_{nlji} - \phi_{2nij})(\psi_{nl'ji} - \phi_{2nij})|Q_{(n)}]$. Thus,

$$\begin{aligned} V_{1b1} = & O(\frac{1}{9}\left(\begin{array}{c} n \\ 3 \end{array}\right)^{-2} \sum_{\substack{i=1 \\ i\neq j\neq l\neq l'}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{l'=1}^{n} E[(\psi_{nlji} - \phi_{2nij})(\psi_{nl'ji} - \phi_{2nij})|Q_{(n)}]) \\ = & O(\frac{1}{9}\left(\begin{array}{c} n \\ 3 \end{array}\right)^{-2} \sum_{\substack{i=1 \\ i\neq j\neq l\neq l'}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{l'=1}^{n} \frac{\hat{\epsilon}_i^2\hat{\epsilon}_j^2}{h^{4(d+1)}}[\frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h}) \\ & -\int \frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l] \\ & \times[\frac{1}{f^2(W_{l'})}SK(\frac{W_j-W_{l'}}{h})SK(\frac{W_i-W_{l'}}{h}) - \int \frac{1}{f^2(W_{l'})}SK(\frac{W_j-W_{l'}}{h})SK(\frac{W_i-W_{l'}}{h})f(W_{l'})dW_{l'}]) \\ = & O(\frac{1}{9}\left(\begin{array}{c} n \\ 3 \end{array}\right)^{-2} \sum_{\substack{i=1 \\ i\neq j\neq l\neq l'}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{l'=1}^{n} \frac{\epsilon_i^2\epsilon_j^2}{h^{4(d+1)}}[\frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h}) \\ & -\int \frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l] \\ & \times[\frac{1}{f^2(W_{l'})}SK(\frac{W_j-W_{l'}}{h})SK(\frac{W_i-W_{l'}}{h}) - \int \frac{1}{f^2(W_{l'})}SK(\frac{W_j-W_{l'}}{h})SK(\frac{W_i-W_{l'}}{h})f(W_{l'})dW_{l'}](1+o_p(1))) \\ = & O(\frac{1}{9}\left(\begin{array}{c} n \\ 3 \end{array}\right)^{-2} \sum_{\substack{i=1 \\ i\neq j\neq l\neq l'}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{l'=1}^{n} \psi_{nvijll'}(1+o_p(1))) = O(V_{1b11}), \end{aligned}$$

where the third equality is from the fact that $\hat{\epsilon}_i = y_i - \hat{m}(W_i) = \epsilon_i + (m(W_i) - \hat{m}(W_i)) = \epsilon_i + o_p(1)$ uniformly $\forall W_i \in \mathcal{W}$. The claim that $\sup_{W\in\mathcal{W}} |\hat{m}(W) - m(W)| = o_p(1)$ follows by applying Lemma 3 of Martins-Filho et al. (2018) or Masry (1996). Note that assumption A3(1) implies that $E|\epsilon|^s < C$ for some $s > 2$. This observation, together with assumptions A1(1)-(3), A2, A3(1), (2) and A4(1), enable us to apply Lemma 3 of Martins-Filho et al. (2018) to obtain the uniform convergence result. We apply Lemma 2 on V_{1b11} below.

Define

$$\begin{aligned} \phi_{nvijll'} = & \psi_{nvijll'} + \psi_{nvijl'l} + \psi_{nviljl'} + \psi_{nvill'j} + \psi_{nvil'jl} + \psi_{nvil'lj} \\ & +\psi_{nvjill'} + \psi_{nvjil'l} + \psi_{nvjlil'} + \psi_{nvjll'i} + \psi_{nvjl'il} + \psi_{nvjl'li} \\ & +\psi_{nvlijl'} + \psi_{nvlil'j} + \psi_{nvljl'i} + \psi_{nvljil'} + \psi_{nvll'ij} + \psi_{nvll'ji} \\ & +\psi_{nvl'ijl} + \psi_{nvl'ilj} + \psi_{nvl'jil} + \psi_{nvl'jli} + \psi_{nvl'lij} + \psi_{nvl'lji}, \end{aligned}$$

$V_{1b11} = \frac{1}{9}\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\begin{pmatrix} n \\ 4 \end{pmatrix}\begin{pmatrix} n \\ 4 \end{pmatrix}^{-1}\sum_{\substack{i=1 \\ i<j<l<l'}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{l'=1}^{n}\phi_{nvijll'} = O(n^{-2})[\theta_n + \sum_{j=1}^{4}\begin{pmatrix} 4 \\ j \end{pmatrix} H_n^{(j)}(Q_{v_1}, \cdots, Q_{v_j})]$.

Since $\int[\frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h}) - \int\frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l]f(W_l)dW_l = 0$, $\theta_n = 0$, and $H_n^{(1)}(Q_{v_1}) = 0$. $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}\sigma_{2n}) = O_p(n^{-1}h^{-4-\frac{3d}{2}})$, since $\sigma_{2n}^2 = E[E^2(\phi_{nvijll'}|Q_l, Q_{l'})] = O(h^{-8-3d})$. $H_n^{(3)}(Q_{v_1}, Q_{v_2}, Q_{v_3}) = O_p(n^{-\frac{3}{2}}\sigma_{3n}) = O_p(n^{-\frac{3}{2}}h^{-4-2d})$, since $\sigma_{3n}^2 = E[E^2(\phi_{nvijll'}|Q_i, Q_j, Q_l)] = O(h^{-8-4d})$. $H_n^{(4)}(Q_{v_1}, Q_{v_2}, Q_{v_3}, Q_{v_4}) = O_p(n^{-2}\sigma_{4n}) = O_p(n^{-2}h^{-4-\frac{5}{2}d})$, since $\sigma_{4n}^2 = E[\phi_{nvijll'}] = O(h^{-8-5d})$. So $V_{1b11} = O_p(n^{-2}[n^{-1}h^{-4-\frac{3d}{2}} + n^{-\frac{3}{2}}h^{-4-2d} + n^{-2}h^{-4-\frac{5}{2}d}]) = O_p(n^{-3}h^{-4-\frac{3}{2}d})$, and $V_{1b1} = O_p(n^{-3}h^{-4-\frac{3}{2}d})$.

Next, we note that by c-r inequality, and since $E(\phi_{2nij}^2|Q_{(n)})$ is of smaller order,

$$
\begin{aligned}
&\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}E(\Phi_{nijl}^2|Q_{(n)}) \\
\leq\ & C\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}[E(\phi_{nijl}^2|Q_{(n)}) + E(\phi_{2nij}^2|Q_{(n)}) + E(\phi_{2nil}^2|Q_{(n)}) + E(\phi_{2njl}^2|Q_{(n)})] \\
=\ & C\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}E(\phi_{nijl}^2|Q_{(n)})(1+o_p(1)) \\
=\ & C\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}[E(\psi_{nijl}^2|Q_{(n)}) + E(\psi_{njil}^2|Q_{(n)}) + E(\psi_{nlji}^2|Q_{(n)})](1+o_p(1)) \\
=\ & O_p(n^{-3}h^{-4-2d}),
\end{aligned}
$$

where the last equality follows from the observation that

$$
\begin{aligned}
&\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}E(\psi_{nijl}^2|Q_{(n)}) \\
=\ & \begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\frac{\hat{\epsilon}_j^2\hat{\epsilon}_l^2}{f^4(W_i)h^{4(d+1)}}SK^2(\frac{W_j-W_i}{h})SK^2(\frac{W_l-W_i}{h}) \\
=\ & \begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\frac{\epsilon_j^2\epsilon_l^2}{f^4(W_i)h^{4(d+1)}}SK^2(\frac{W_j-W_i}{h})SK^2(\frac{W_l-W_i}{h})(1+o_p(1)) \\
=\ & O_p(n^{-3}h^{-4-2d}).
\end{aligned}
$$

So $V(T_{1b}^*|Q_{(n)}) = \frac{1}{9}\begin{pmatrix} n \\ 3 \end{pmatrix}^{-2}\sum_{\substack{i=1 \\ i<j<l}}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}E(\Phi_{nijl}^2|Q_{(n)}) + V_{1b1} = O_p(n^{-3}h^{-4-2d}) + O_p(n^{-3}h^{-4-\frac{3}{2}d})$. Thus, $T_{1b}^* = O_p(n^{-\frac{3}{2}}h^{-2-d}) = o_p(n^{-1}h^{-2-\frac{d}{2}})$, as claimed.

(II) Claim: $nh^{2+\frac{d}{2}}T_{1a}^*/\sqrt{\hat{\Omega}^*} \xrightarrow{d} \mathcal{N}(0,1)$.

$$
\begin{aligned}
nh^{2+\frac{d}{2}}T_{1a}^* = &\ nh^{2+\frac{d}{2}}\tfrac{1}{3}\begin{pmatrix} n \\ 3 \end{pmatrix}^{-1}\begin{pmatrix} n-2 \\ 1 \end{pmatrix}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\phi_{2nij} \\
= &\ \sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n} nh^{2+\frac{d}{2}}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\tfrac{\epsilon_i^*\epsilon_j^*}{h^{2(d+1)}}\int\tfrac{1}{f^2(W_l)}SK(\tfrac{W_j-W_l}{h})SK(\tfrac{W_i-W_l}{h})f(W_l)dW_l \\
= &\ \sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\psi_{n1ij}.
\end{aligned}
$$

Note that ψ_{n1ij} is symmetric in i, j, $E(\psi_{n1ij}|Q_{(n)}, \epsilon_j^*) = 0$, so conditioning on $Q_{(n)}$, $nh^{2+\frac{d}{2}}T_{1a}^*$ is a degenerate second order U-statistic. Defining $(S_n^*)^2 \equiv E((\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\psi_{n1ij})^2|Q_{(n)}) = V(T_{1a}^*|Q_{(n)})$, we apply Proposition 3.2 of de Jong (1987) to obtain

$$
(S_n^*)^{-1}nh^{2+\frac{d}{2}}T_{1a}^* \xrightarrow{d} \mathcal{N}(0,1),
$$

if G_I, G_{II}, G_{IV} are each of order $o_p((S_n^*)^4)$, where $G_I = \sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}E(\psi_{n1ij}^4|Q_{(n)})$,

$$
G_{II} = \sum_{\substack{i=1\\i<j<t}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}[E(\psi_{n1ij}^2\psi_{n1it}^2|Q_{(n)}) + E(\psi_{n1ji}^2\psi_{n1jt}^2|Q_{(n)}) + E(\psi_{n1ti}^2\psi_{n1tj}^2|Q_{(n)})] = G_{II1} + G_{II2} + G_{II3},
$$

$$
\begin{aligned}
G_{IV} = &\ \sum_{\substack{i=1\\i<j<t<l}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}[E(\psi_{n1ij}\psi_{n1it}\psi_{n1lj}\psi_{n1lt}|Q_{(n)}) + E(\psi_{n1ij}\psi_{n1il}\psi_{n1tj}\psi_{n1tl}|Q_{(n)}) + E(\psi_{n1it}\psi_{n1il}\psi_{n1jt}\psi_{n1jl}|Q_{(n)})] \\
= &\ G_{IV1} + G_{IV2} + G_{IV3}.
\end{aligned}
$$

Consider $(S_n^*)^2$. Since $E(\epsilon_i^*\epsilon_j^*\epsilon_t^*\epsilon_m^*|Q_{(n)}) = \hat{\epsilon}_i^2\hat{\epsilon}_j^2$ when $i = t < j = m$, and zero in the other cases, we have

$$
\begin{aligned}
(S_n^*)^2 = &\ n^2h^{4+d}\begin{pmatrix} n \\ 2 \end{pmatrix}^{-2}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\tfrac{\hat{\epsilon}_i^2\hat{\epsilon}_j^2}{h^{4(d+1)}}[\int\tfrac{1}{f^2(W_l)}SK(\tfrac{W_j-W_l}{h})SK(\tfrac{W_i-W_l}{h})f(W_l)dW_l]^2 \\
= &\ \tfrac{4}{(n-1)^2}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\tfrac{\epsilon_i^2\epsilon_j^2}{h^{3d}}[\int\tfrac{1}{f^2(W_l)}SK(\tfrac{W_j-W_l}{h})SK(\tfrac{W_i-W_l}{h})f(W_l)dW_l]^2(1+o_p(1)) \\
= &\ (1+o_p(1))2\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\tfrac{\epsilon_i^2\epsilon_j^2}{h^{3d}}[\int\tfrac{1}{f^2(W_l)}SK(\tfrac{W_j-W_l}{h})SK(\tfrac{W_i-W_l}{h})f(W_l)dW_l]^2 \\
= &\ (1+o_p(1))2\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\psi_{nsij},
\end{aligned}
$$

where the second equality follows since $n^2\begin{pmatrix} n \\ 2 \end{pmatrix}^{-2} = \frac{4}{(n-1)^2}$, and $\hat{\epsilon}_i = \epsilon_i + o_p(1)$ uniformly. The third follows since $\frac{2}{(n-1)^2} - \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} = O(n^{-3})$. We then apply Lemma 2 to obtain that

$$
\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\psi_{nsij} = \theta_n + \sum_{j=1}^{2}\begin{pmatrix} 2 \\ j \end{pmatrix}H_n^{(j)}(Q_{v_1},\cdots,Q_{v_j}) = \frac{\Omega}{2} + O_p(n^{-1}h^{-\frac{d}{2}}),
$$

since $\theta_n = E\psi_{nsij} \to \int\sigma^4(W_i)dW_i\int[\int SK(\Psi_1+\Psi)SK(\Psi)d\Psi]^2d\Psi_1 = \frac{\Omega}{2}$, $H_n^{(1)}(Q_{v_1}) = O_p(n^{-1})$, and $H_n^{(2)}(Q_{v_1}, Q_{v_2}) = O_p(n^{-1}h^{-\frac{d}{2}})$. Thus, $(S_n^*)^2 = \Omega + o_p(1)$.

$$G_I = \sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} h^{8+2d}\frac{16}{(n-1)^4}\frac{E((\epsilon_i^*\epsilon_j^*)^4|Q_{(n)})}{h^{8(d+1)}}[\int \frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l]^4 = O(n^{-2}h^{-d}) = o_p((S_n^*)^4),$$

since $(S_n^*)^4 = O(1)$.

$$\begin{aligned} G_{II1} = & \sum_{\substack{i=1 \\ i<j<t}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n} h^{8+2d}\frac{16}{(n-1)^4}h^{-8-8d}E((\epsilon_i^*)^4(\epsilon_j^*)^2(\epsilon_t^*)^2|Q_{(n)})[\int \frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l]^2 \\ & \times[\int \frac{1}{f^2(W_{l'})}SK(\frac{W_t-W_{l'}}{h})SK(\frac{W_i-W_{l'}}{h})f(W_{l'})dW_{l'}]^2 \\ = & \ O(n^{-1}). \end{aligned}$$

Similarly, $G_{II2} = O(n^{-1})$, $G_{II3} = O(n^{-1})$, thus we conclude that $G_{II} = O(n^{-1}) = o((S_n^*)^4)$.

$$\begin{aligned} & G_{IV1} \\ = & \sum_{\substack{i=1 \\ i<j<t<l}}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{16}{(n-1)^4}h^{-2d}E((\epsilon_i^*)^2(\epsilon_j^*)^2(\epsilon_t^*)^2(\epsilon_l^*)^2|Q_{(n)})[\frac{1}{h^d}\int \frac{1}{f^2(W_l)}SK(\frac{W_j-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l] \\ & \times[\frac{1}{h^d}\int \frac{1}{f^2(W_l)}SK(\frac{W_t-W_l}{h})SK(\frac{W_i-W_l}{h})f(W_l)dW_l][\frac{1}{h^d}\int \frac{1}{f^2(W_{l'})}SK(\frac{W_j-W_{l'}}{h})SK(\frac{W_l-W_{l'}}{h})f(W_{l'})dW_{l'}] \\ & \times[\frac{1}{h^d}\int \frac{1}{f^2(W_{l'})}SK(\frac{W_t-W_{l'}}{h})SK(\frac{W_l-W_{l'}}{h})f(W_{l'})dW_{l'}](1+o_p(1)) \\ = & \ O(h^d). \end{aligned}$$

Similarly, $G_{IV2} = O(h^d)$, and $G_{IV3} = O(h^d)$, and thus we have $G_{IV} = O(h^d) = o((S_n^*)^4)$. Since we obtain that $(S_n^*)^2 \xrightarrow{p} \Omega$, we conclude that $nh^{2+\frac{d}{2}}T_{1a}^* \xrightarrow{d} \mathcal{N}(0,\Omega)$.

We show that $\hat{\Omega}^* - \Omega = o_p(1)$, which implies the claim that $nh^{2+\frac{d}{2}}T_{1a}^*/\sqrt{\hat{\Omega}^*} \xrightarrow{d} \mathcal{N}(0,1)$. Since $\hat{\Omega}^* - \Omega = \hat{\Omega}^* - \hat{\Omega} + \hat{\Omega} - \Omega = \hat{\Omega}^* - \hat{\Omega} + o_p(1)$ as Corollary 1. We only need to show $\hat{\Omega}^* - \hat{\Omega} = o_p(1)$. To this end, it is sufficient to show $I_{\hat{\Omega}^*} \equiv \frac{1}{n^2h^d}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{\epsilon_i^{*2}\epsilon_j^{*2}}{\hat{f}(W_i)\hat{f}(W_j)}K(\frac{W_i-W_j}{h}) = \frac{1}{n^2h^d}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{\hat{\epsilon}_i^2\hat{\epsilon}_j^2}{\hat{f}(W_i)\hat{f}(W_j)}K(\frac{W_i-W_j}{h}) + o_p(1)$. Given $E(\epsilon_i^{*2}|Q_{(n)}) = \hat{\epsilon}_i^2$, $E(I_{\hat{\Omega}^*}|Q_{(n)}) = \frac{1}{n^2h^d}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{\hat{\epsilon}_i^2\hat{\epsilon}_j^2}{\hat{f}(W_i)\hat{f}(W_j)}K(\frac{W_i-W_j}{h})$. We show $V(I_{\hat{\Omega}^*}|Q_{(n)}) = o_p(1)$, so that by Chebyshev's inequality, we have the claim. With the wild bootstrap, we have $E(\epsilon_i^{*4}|Q_{(n)}) = 2\hat{\epsilon}_i^4$, thus

$$\begin{aligned} V(I_{\hat{\Omega}^*}|Q_{(n)}) = & \ E([I_{\hat{\Omega}^*} - E(I_{\hat{\Omega}^*}|Q_{(n)})]^2|Q_{(n)}) \\ = & \ \frac{1}{n^4h^{2d}}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\sum_{\substack{l=1 \\ l\neq m}}^{n}\sum_{m=1}^{n}\frac{K(\frac{W_i-W_j}{h})K(\frac{W_l-W_m}{h})}{\hat{f}(W_i)\hat{f}(W_j)\hat{f}(W_l)\hat{f}(W_m)}E((\epsilon_i^{*2}\epsilon_j^{*2} - \hat{\epsilon}_i^2\hat{\epsilon}_j^2)(\epsilon_l^{*2}\epsilon_m^{*2} - \hat{\epsilon}_l^2\hat{\epsilon}_m^2)|Q_{(n)}). \end{aligned}$$

When $(i,j) \neq (l,m)$, above is zero. Due to symmetry, we only need to consider the case $i = l \neq j = m$. So

$$\begin{aligned} V(I_{\hat{\Omega}^*}|Q_{(n)}) = & \ \frac{1}{n^2h^d}\frac{1}{n^2}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{K^2(\frac{W_i-W_j}{h})}{h^d\hat{f}(W_i)\hat{f}(W_j)\hat{f}(W_l)\hat{f}(W_m)}E((\epsilon_i^{*2}\epsilon_j^{*2} - \hat{\epsilon}_i^2\hat{\epsilon}_j^2)^2|Q_{(n)}) \\ = & \ \frac{1}{n^2h^d}\frac{1}{n^2}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{K^2(\frac{W_i-W_j}{h})}{h^d\hat{f}(W_i)\hat{f}(W_j)\hat{f}(W_l)\hat{f}(W_m)}3\hat{\epsilon}_i^4\hat{\epsilon}_j^4 = O_p((n^2h^d)^{-1}) = o_p(1). \end{aligned}$$

Thus, we have the claim that $\hat{\Omega}^* - \Omega = o_p(1)$.

(I) and (II) imply that $nh^{2+\frac{d}{2}}T_1^*/\sqrt{\hat{\Omega}} \xrightarrow{d} \mathcal{N}(0,1)$, the case with $i \neq j \neq l$.

(b) When $i \neq j = l$, $T_1^* \equiv T_{1B}^* = \frac{1}{nh^{2+d}}\frac{1}{n^2}\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n}\frac{1}{h^d}\hat{SK}^2(\frac{W_j-W_i}{h})\epsilon_j^{*2} = \frac{1}{nh^{2+d}}\hat{B}_{1n}^*$. Combining results (a) and (b) above, we obtain the claim in (1).

Proof of (2): $$T_3^* = -\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{\substack{l=1\\l\neq i}}^n \frac{1}{nh^{d+1}}\sum_{\substack{j=1\\j\neq i}}^n \hat{SK}(\frac{W_j-W_i}{h})\epsilon_j^* \frac{1}{nh^{d+1}}\sum_{\substack{t=1\\t\neq l,t\neq i}}^n \hat{SK}(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_t^*$$
$$= -\frac{2}{n^3(n-1)}\sum_{\substack{i=1\\i\neq j,t\neq l\neq i}}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{t=1}^n \frac{1}{h^{2(d+1)}}\hat{SK}(\frac{W_j-W_i}{h})\hat{SK}(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_j^*\epsilon_t^*.$$

(a) When $i\neq j\neq l\neq t$, with $\frac{1}{n-1}-\frac{1}{n}=O(n^{-2})$,

$$T_3^* = -\frac{2}{n^4}\sum_{\substack{i=1\\i\neq j\neq l\neq t}}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{t=1}^n \frac{1}{f(W_i)f(W_l^r;W_i^c)h^{2(d+1)}}SK(\frac{W_j-W_i}{h})SK(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_j^*\epsilon_t^*(1+o_p(1))$$
$$= -\frac{2}{n^4}\sum_{\substack{i=1\\i\neq j\neq l\neq t}}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{t=1}^n \psi_{nijlt}(1+o_p(1)) = T_{3n}^*(1+o_p(1)).$$

$E(T_{3n}^*|Q_{(n)})=0$ as $E(\epsilon_j^*|Q_{(n)})=0$, and $V(T_{3n}^*|Q_{(n)}) = \frac{4}{n^8}\sum_{\substack{i=1\\i\neq j\neq l\neq t}}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{t=1}^n\sum_{\substack{i'=1\\i'\neq j'\neq l'\neq t'}}^n\sum_{j'=1}^n\sum_{l'=1}^n\sum_{t'=1}^n E(\psi_{nijlt}\psi_{ni'j'l't'}|Q_{(n)})$.

We can show that $E(|V(T_{3n}^*|Q_{(n)})|)=o(n^{-2}h^{-4-d})$ with different combinations of indices in the summation, thus we apply Chebyshev's inequality to conclude that $T_3^*=o_p(n^{-1}h^{-2-\frac{d}{2}})$.

(b) When $i\neq j=l\neq t$,

$$T_3^* = -\frac{2}{n^4}\sum_{\substack{i=1\\i\neq j\neq t}}^n\sum_{j=1}^n\sum_{t=1}^n \frac{1}{f(W_i)f(W_j^r;W_i^c)h^{2(d+1)}}SK(\frac{W_j-W_i}{h})SK(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_j^*\epsilon_t^*(1+o_p(1)) = T_{3n}^*(1+o_p(1)).$$

Again $E(T_{3n}^*|Q_{(n)})=0$, and we can show that $V(T_{3n}^*|Q_{(n)})=O_p(n^{-4}h^{-4-d})$, and thus $T_3^*=o_p(n^{-1}h^{-2-\frac{d}{2}})$.

(c) When $i\neq j=t\neq l$,

$$T_3^* \equiv T_{3B}^* = -\frac{2}{nh^{2+(d-d_1)}}\frac{1}{n^2(n-1)}\sum_{\substack{i=1\\i\neq j\neq l}}^n\sum_{j=1}^n\sum_{l=1}^n \frac{1}{h^{d+d_1}}\hat{SK}(\frac{W_j-W_i}{h})\hat{SK}(\frac{W_j^r-W_l^r}{h};\frac{W_j^c-W_i^c}{h})(\epsilon_j^*)^2 = -\frac{2}{nh^{2+(d-d_1)}}\hat{B}_{3n}^*.$$

The claim in (2) follows from (2)(a)-(2)(c) above.

Proof of (3): we perform similar arguments to obtain that

$$T_2^* = \frac{1}{n}\sum_{i=1}^n[\frac{1}{n-1}\sum_{\substack{j=1\\j\neq i}}^n\frac{1}{nh^{d+1}}\sum_{\substack{t=1\\t\neq j,t\neq i}}^n \hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_t^*][\frac{1}{n-1}\sum_{\substack{l=1\\l\neq i}}^n\frac{1}{nh^{d+1}}\sum_{\substack{m=1\\m\neq l,m\neq i}}^n \hat{SK}(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})\epsilon_m^*]$$
$$= \frac{1}{n^3(n-1)^2}\sum_{\substack{i=1\\i\neq j\neq t,i\neq l\neq m}}^n\sum_{j=1}^n\sum_{t=1}^n\sum_{l=1}^n\sum_{m=1}^n \frac{1}{h^{2(d+1)}}\hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\hat{SK}(\frac{W_m^r-W_l^r}{h};\frac{W_m^c-W_i^c}{h})\epsilon_t^*\epsilon_m^*$$
$$= \frac{1}{nh^{2+d-d_1}}\frac{1}{n^2(n-1)^2}\sum_{\substack{i=1\\i\neq j\neq t\neq l}}^n\sum_{j=1}^n\sum_{t=1}^n\sum_{l=1}^n \frac{1}{h^{d+d_1}}\hat{SK}(\frac{W_t^r-W_j^r}{h};\frac{W_t^c-W_i^c}{h})\hat{SK}(\frac{W_t^r-W_l^r}{h};\frac{W_t^c-W_i^c}{h})\epsilon_t^{*2} + o_p(n^{-1}h^{-2-\frac{d}{2}})$$
$$= \frac{1}{nh^{2+d-d_1}}\hat{B}_{2n}^* + o_p(n^{-1}h^{-2-\frac{d}{2}}).$$

Thus, we obtain the claim in (3), which concludes the proof of Theorem 3 (i).

(ii) Recall that z_α^* is the $(1-\alpha)$ quantile of the empirical distribution of $\{\hat{T}_{c,b}^*\}_{b=1}^B$. Let $\bar{z}_\alpha^*$ be the $(1-\alpha)$

conditional quantile of $\hat{T}_c^*$ given $Q_{(n)}$, i.e., $P(\hat{T}_c^* > \bar{z}_\alpha^*) = \alpha$. With B large enough, the approximation error of z_α^* to $\bar{z}_\alpha^*$ is arbitrarily small in probability and negligible. By (i), $\bar{z}_\alpha^* \xrightarrow{p} z_\alpha$, where z_α is the $(1-\alpha)$ quantile of the standard normal distribution. Then by Corollary 1, $\lim_{n\to\infty} P(\hat{T}_c \geq z_\alpha^*) = \lim_{n\to\infty} P(\hat{T}_c \geq z_\alpha) = \alpha$ under H_0.

(iii) Under H_A, as shown in Theorem 2, $P(\hat{T}_c \geq z) \to 1$ for any given z. Given that $z_\alpha^* \xrightarrow{p} z_\alpha$, $\lim_{n\to\infty} P(\hat{T}_c \geq z_\alpha^*) = \lim_{n\to\infty} P(\hat{T}_c \geq z_\alpha) = 1$. □

C Appendix 3

In this section, we study extensively the finite sample performance of our centered bootstrap test $\hat{T}_c^*$ and un-centered bootstrap test $\hat{T}^*$ by evaluating their empirical size and power in the bivariate regression ($d = 2$) and the trivariate regression ($d = 3$). With $d = 2$, we report testing results with three scaling factors for the *rule-of-thumb* bandwidth in Section C.1, and with a cross-validation least-square (CVLS) bandwidth in Section C.2. We report results for $d = 3$ with three scaling factors for the *rule-of-thumb* bandwidth in Section C.3, and for a partially linear model with a CVLS bandwidth in Section C.4. Unless stated otherwise, in each of the simulation studies we perform 1000 repetitions for each design, and perform 299 repetitions for the bootstrap. We report the empirical relative rejection frequency for $\hat{T}_c^*$ and $\hat{T}^*$(in parentheses).

C.1 Bivariate Case: Rule-of-Thumb Bandwidth Selection

We first consider bivariate regressions ($d = 2$) with $W = [X, Z_1]'$, and test three simple null hypotheses on the significant variables in $g_X(W)$. The first null, denoted by Case 1, is that $W^r = Z_1$ is insignificant in $g_X(W)$. It is easy to infer that the regression model is additive, i.e., $m(W) = m_1(X) + m_2(Z_1)$; The second null, Case 2, specifies that $W^r = X$ is insignificant in $g_X(W)$, and a varying coefficient structure satisfies this hypothesis, i.e., $m(W) = Xm_1(Z_1) + m_2(Z_1)$; The third null, Case 3, is that $W^r = [X, Z_1]'$ are insignificant in $g_X(W)$, which corresponds to the structure where X enters the model linearly with a constant coefficient and X is additively separable from Z_1, i.e., a partially linear model $m(W) = X\beta + m_1(Z_1)$ exhibits this structure. We consider the following three data-generating processes (DGPs) for $i = 1, \cdots, n$:

$$DGP_1 : Y_i = 0.5 + X_i + \delta X_i^2 + Z_{1i} + Z_{1i}^2 + \delta_1 X_i Z_{1i} + \epsilon_i$$

$$DGP_2 : Y_i = 5 + 2X_i - \delta e^{1.1X_i} + Z_{1i}^3 + 2\delta_1 X_i \sin(Z_{1i}) + \epsilon_i$$

$$DGP_3 : Y_i = 1 + X_i + \delta X_i^3 + 0.4Z_{1i}^2 - \delta_1 X_i e^{Z_{1i}} + \epsilon_i$$

where X_i and Z_{1i} are each iid and drawn independently from a uniform distribution $U(-2, 2)$, and $\epsilon_i \sim \mathcal{N}(0, 1)$ is the error term. With a nonzero δ, the three DGPs exhibit nonlinearity in X. With a nonzero δ_1, we introduce an interaction term between X and Z_1, in which the impact of X is always linear across the three DGPs, but the impact of Z_1 is linear only in DGP_1. DGP_2 contains a high frequency function ($sin(Z_1)$) in the interaction, and is a modified version from Wang and Carriere (2011) which tests additivity. DGP_3 is adapted from Yang et al. (2006), which test for a constant coefficient against a varying coefficient model.

We investigate the size and power of our test under Cases 1-3 with different choices of (δ, δ_1). For all three DGPs in Case 1, we investigate the size performance by letting $\delta_1 = 0$, i.e., X does not interact with Z_1 in H_0. We simply set $\delta = 1$ to allow for nonlinearity in X. In Case 2, we examine the size by letting $\delta = 0$, i.e., X enters the model linearly in H_0. We simply set $\delta_1 = 1$ to allow for the presence of interaction effects. In Case 3, we set $(\delta, \delta_1) = (0, 0)$ to investigate the size under H_0. Different values of δ_1, δ and (δ, δ_1) other than those chosen above allow us to explore the power performance. Here, we simply illustrate the empirical power performance by letting $\delta_1 = 1$ in Case 1, $\delta = 1$ in Case 2, and $(\delta, \delta_1) = (1, 1)$ in Case 3 to save space.

We utilize the Gaussian kernel function as $\mathcal{K}(\psi) = \frac{1}{\sqrt{2\pi}} exp(-\psi^2/2)$, and choose a *rule-of-thumb* bandwidth $h_\xi = C_h \hat{\sigma}_\xi n^{-\frac{1}{2p+d}}$, where C_h is the scaling factor and $\hat{\sigma}_\xi$ is the sample standard deviation of the variable ξ, which is either X or Z_1. We consider three sample sizes $n = (50, 100, 200)$, and set $C_h = (0.5, 1.0, 1.5)$ to check for the sensitivity of the test performance to bandwidths.

Table C.1 reports the empirical relative rejection frequency for $\hat{T}_c^*$ ($\hat{T}^*$) with $C_h = 1.0$, for significant levels $\alpha = (0.10, 0.05, 0.01)$. For all three cases, both tests exhibit fairly reasonable size performance, generally

Table C.1: Empirical Size and Power for $\hat{T}_c^*(\hat{T}^*)$ from Bivariate Regressions with $d = 2$ ($C_h = 1.0$)

Case 1		H_0: W^r=Z_1 (i.e., an additive model with δ=1)								
			DGP_1			DGP_2			DGP_3	
	δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.114	0.106	0.099	0.083	0.092	0.096	0.085	0.094	0.103
		(0.136)	(0.088)	(0.094)	(0.060)	(0.067)	(0.082)	(0.073)	(0.082)	(0.088)
	1.0	0.983	1.000	1.000	0.921	1.000	1.000	1.000	1.000	1.000
		(0.986)	(1.000)	(1.000)	(0.938)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.056	0.052	0.048	0.037	0.044	0.053	0.039	0.048	0.052
		(0.058)	(0.045)	(0.052)	(0.029)	(0.038)	(0.044)	(0.032)	(0.046)	(0.051)
	1.0	0.974	1.000	1.000	0.861	1.000	1.000	1.000	1.000	1.000
		(0.979)	(1.000)	(1.000)	(0.888)	(0.997)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.017	0.012	0.010	0.005	0.007	0.008	0.006	0.009	0.011
		(0.018)	(0.015)	(0.011)	(0.003)	(0.004)	(0.006)	(0.005)	(0.009)	(0.012)
	1.0	0.908	1.000	1.000	0.789	0.982	1.000	0.999	1.000	1.000
		(0.911)	(1.000)	(1.000)	(0.802)	(0.992)	(1.000)	(0.997)	(1.000)	(1.000)
Case 2		H_0: W^r=X (i.e., a varying coefficient type model with δ_1=1)								
			DGP_1			DGP_2			DGP_3	
	δ	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.118	0.108	0.103	0.086	0.094	0.099	0.122	0.112	0.104
		(0.147)	(0.115)	(0.103)	(0.075)	(0.083)	(0.108)	(0.169)	(0.114)	(0.103)
	1.0	1.000	1.000	1.000	0.960	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.969)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.065	0.058	0.053	0.039	0.048	0.052	0.074	0.059	0.050
		(0.086)	(0.063)	(0.055)	(0.030)	(0.045)	(0.049)	(0.093)	(0.067)	(0.054)
	1.0	1.000	1.000	1.000	0.943	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.962)	(1.000)	(1.000)	(0.997)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.025	0.019	0.009	0.008	0.012	0.013	0.015	0.011	0.008
		(0.036)	(0.027)	(0.013)	(0.006)	(0.009)	(0.015)	(0.019)	(0.022)	(0.016)
	1.0	0.982	1.000	1.000	0.929	1.000	1.000	0.984	1.000	1.000
		(0.996)	(1.000)	(1.000)	(0.935)	(1.000)	(1.000)	(0.990)	(1.000)	(1.000)
Case 3		H_0: W^r=[X,Z_1] (i.e., a partially linear type model)								
			DGP_1			DGP_2			DGP_3	
	δ=δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.104	0.096	0.098	0.118	0.106	0.102	0.125	0.107	0.098
		(0.112)	(0.126)	(0.102)	(0.123)	(0.108)	(0.097)	(0.144)	(0.115)	(0.101)
	1.0	1.000	1.000	1.000	0.985	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.994)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.057	0.051	0.048	0.072	0.058	0.052	0.063	0.055	0.047
		(0.068)	(0.064)	(0.056)	(0.081)	(0.044)	(0.051)	(0.070)	(0.062)	(0.054)
	1.0	1.000	1.000	1.000	0.974	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.980)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.015	0.011	0.008	0.020	0.016	0.012	0.017	0.012	0.008
		(0.020)	(0.016)	(0.010)	(0.034)	(0.016)	(0.013)	(0.020)	(0.022)	(0.014)
	1.0	0.991	1.000	1.000	0.959	1.000	1.000	0.984	1.000	1.000
		(0.998)	(1.000)	(1.000)	(0.964)	(1.000)	(1.000)	(0.998)	(1.000)	(1.000)

Note: Empirical size and power are calculated based on 1000 simulations with 299 bootstrap repetitions. The *rule-of-thumb* bandwidths have a scaling factor $C_h = 1.0$, and α is the significance level.

oversized in smaller samples under DGP_1 and DGP_3, and undersized in DGP_2. As the sample size increases, the size of the tests generally improves toward its nominal level across all DGPs and three cases. For the chosen parameters, the empirical power of the tests in Case 1-3 rises quickly to one as n increases, with

Table C.2: Empirical Size and Power for $\hat{T}_c^*(\hat{T}^*)$ from Bivariate Regressions with $d = 2$ ($C_h = 0.5$)

Case 1	H_0: W^r=Z_1 (i.e., an additive model with δ=1)									
		DGP_1			DGP_2			DGP_3		
	δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.123	0.114	0.105	0.077	0.085	0.095	0.076	0.091	0.097
		(0.154)	(0.129)	(0.109)	(0.053)	(0.072)	(0.093)	(0.065)	(0.096)	(0.104)
	1.0	0.971	1.000	1.000	0.878	0.984	1.000	1.000	1.000	1.000
		(0.979)	(1.000)	(1.000)	(0.885)	(0.996)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.062	0.058	0.053	0.032	0.042	0.047	0.034	0.043	0.048
		(0.076)	(0.069)	(0.046)	(0.023)	(0.041)	(0.048)	(0.029)	(0.040)	(0.046)
	1.0	0.944	1.000	1.000	0.827	0.982	1.000	0.985	1.000	1.000
		(0.951)	(1.000)	(1.000)	(0.834)	(0.988)	(1.000)	(0.994)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.020	0.014	0.008	0.006	0.009	0.012	0.005	0.007	0.013
		(0.021)	(0.014)	(0.008)	(0.008)	(0.012)	(0.007)	(0.007)	(0.008)	(0.014)
	1.0	0.868	1.000	1.000	0.682	0.965	1.000	0.970	1.000	1.000
		(0.872)	(1.000)	(1.000)	(0.709)	(0.977)	(0.997)	(0.988)	(1.000)	(1.000)
Case 2	H_0: W^r=X (i.e., a varying coefficient type model with δ_1=1)									
		DGP_1			DGP_2			DGP_3		
	δ	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.119	0.105	0.098	0.089	0.096	0.102	0.111	0.107	0.097
		(0.137)	(0.122)	(0.106)	(0.081)	(0.092)	(0.095)	(0.136)	(0.105)	(0.094)
	1.0	1.000	1.000	1.000	0.940	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.945)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.061	0.054	0.047	0.030	0.044	0.054	0.068	0.054	0.047
		(0.069)	(0.058)	(0.050)	(0.025)	(0.039)	(0.044)	(0.075)	(0.057)	(0.053)
	1.0	0.997	1.000	1.000	0.925	1.000	1.000	0.984	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.933)	(1.000)	(1.000)	(0.992)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.017	0.013	0.007	0.005	0.007	0.102	0.014	0.009	0.008
		(0.018)	(0.022)	(0.017)	(0.003)	(0.007)	(0.009)	(0.017)	(0.006)	(0.007)
	1.0	0.980	1.000	1.000	0.895	0.998	1.000	0.976	1.000	1.000
		(0.987)	(1.000)	(1.000)	(0.902)	(1.000)	(1.000)	(0.987)	(1.000)	(1.000)
Case 3	H_0: W^r=$[X,Z_1]$ (i.e., a partially linear type model)									
		DGP_1			DGP_2			DGP_3		
	δ=δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.120	0.111	0.105	0.121	0.112	0.106	0.129	0.112	0.106
		(0.145)	(0.127)	(0.117)	(0.128)	(0.105)	(0.095)	(0.155)	(0.127)	(0.107)
	1.0	1.000	1.000	1.000	0.964	1.000	1.000	0.991	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.975)	(1.000)	(1.000)	(0.997)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.074	0.062	0.057	0.073	0.061	0.055	0.074	0.064	0.056
		(0.085)	(0.071)	(0.066)	(0.074)	(0.051)	(0.060)	(0.089)	(0.057)	(0.046)
	1.0	1.000	1.000	1.000	0.936	1.000	1.000	0.989	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.942)	(1.000)	(1.000)	(0.992)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.022	0.015	0.012	0.022	0.015	0.007	0.020	0.015	0.013
		(0.029)	(0.020)	(0.007)	(0.023)	(0.009)	(0.014)	(0.024)	(0.013)	(0.008)
	1.0	0.985	1.000	1.000	0.901	1.000	1.000	0.940	1.000	1.000
		(0.994)	(1.000)	(1.000)	(0.912)	(1.000)	(1.000)	(0.953)	(1.000)	(1.000)

Table C.3: Empirical Size and Power for $\hat{T}_c^*(\hat{T}^*)$ from Bivariate Regressions with $d = 2$ ($C_h = 1.5$)

Case 1		H_0: W^r=Z_1 (i.e., an additive model with δ=1)								
		DGP_1			DGP_2			DGP_3		
	δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.120	0.111	0.104	0.084	0.094	0.097	0.091	0.109	0.105
		(0.149)	(0.115)	(0.107)	(0.079)	(0.086)	(0.104)	(0.084)	(0.093)	(0.097)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.069	0.060	0.055	0.036	0.042	0.048	0.048	0.060	0.056
		(0.097)	(0.074)	(0.060)	(0.031)	(0.045)	(0.053)	(0.045)	(0.057)	(0.057)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.024	0.017	0.014	0.006	0.008	0.011	0.006	0.014	0.008
		(0.035)	(0.020)	(0.007)	(0.004)	(0.007)	(0.015)	(0.003)	(0.006)	(0.007)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
Case 2		H_0: W^r=X (i.e., a varying coefficient type model with δ_1=1)								
		DGP_1			DGP_2			DGP_3		
	δ	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.118	0.110	0.106	0.092	0.103	0.097	0.118	0.112	0.106
		(0.121)	(0.129)	(0.109)	(0.086)	(0.092)	(0.115)	(0.149)	(0.120)	(0.108)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.065	0.058	0.053	0.042	0.054	0.048	0.060	0.056	0.046
		(0.057)	(0.061)	(0.056)	(0.039)	(0.049)	(0.056)	(0.062)	(0.055)	(0.047)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.998)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.019	0.015	0.012	0.008	0.013	0.008	0.018	0.014	0.009
		(0.013)	(0.019)	(0.009)	(0.008)	(0.008)	(0.013)	(0.023)	(0.015)	(0.014)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.982	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.993)	(1.000)	(1.000)
Case 3		H_0: W^r=$[X,Z_1]$ (i.e., a partially linear type model)								
		DGP_1			DGP_2			DGP_3		
	δ=δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.112	0.107	0.103	0.118	0.108	0.096	0.117	0.105	0.096
		(0.128)	(0.115)	(0.097)	(0.125)	(0.114)	(0.107)	(0.128)	(0.107)	(0.098)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.056	0.044	0.047	0.069	0.060	0.048	0.062	0.055	0.044
		(0.061)	(0.056)	(0.052)	(0.071)	(0.064)	(0.056)	(0.060)	(0.054)	(0.042)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.015	0.007	0.007	0.020	0.015	0.011	0.018	0.014	0.008
		(0.017)	(0.011)	(0.008)	(0.029)	(0.017)	(0.007)	(0.015)	(0.012)	(0.009)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)

Table C.4: Empirical Size and Power for $\hat{T}_c^*(\hat{T}^*)$ from Bivariate Regressions with $d = 2$ (CVLS)

Case 1	H_0: W^r=Z_1 (i.e., an additive model with δ=1)									
			DGP_1			DGP_2			DGP_3	
	δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.077	0.084	0.094	0.068	0.075	0.090	0.088	0.107	0.096
		(0.082)	(0.078)	(0.086)	(0.046)	(0.079)	(0.095)	(0.074)	(0.083)	(0.094)
	1.0	0.938	0.989	1.000	0.917	1.000	1.000	0.976	1.000	1.000
		(0.943)	(0.996)	(1.000)	(0.930)	(1.000)	(1.000)	(0.989)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.038	0.044	0.056	0.034	0.039	0.044	0.040	0.056	0.048
		(0.043)	(0.039)	(0.045)	(0.028)	(0.039)	(0.048)	(0.033)	(0.042)	(0.054)
	1.0	0.861	0.941	1.000	0.884	0.931	1.000	0.912	1.000	1.000
		(0.871)	(0.957)	(1.000)	(0.891)	(0.945)	(1.000)	(0.927)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.005	0.007	0.014	0.005	0.007	0.009	0.006	0.014	0.012
		(0.016)	(0.011)	(0.008)	(0.004)	(0.008)	(0.016)	(0.004)	(0.009)	(0.012)
	1.0	0.829	0.910	0.986	0.802	0.887	0.936	0.873	0.984	1.000
		(0.848)	(0.929)	(0.998)	(0.816)	(0.895)	(0.959)	(0.885)	(0.991)	(1.000)
Case 2	H_0: W^r=X (i.e., a varying coefficient type model with δ_1=1)									
			DGP_1			DGP_2			DGP_3	
	δ	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.125	0.114	0.106	0.082	0.094	0.104	0.129	0.118	0.106
		(0.154)	(0.123)	(0.104)	(0.072)	(0.087)	(0.096)	(0.153)	(0.127)	(0.105)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.999)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.072	0.063	0.057	0.041	0.046	0.056	0.070	0.063	0.054
		(0.086)	(0.074)	(0.056)	(0.031)	(0.042)	(0.055)	(0.082)	(0.062)	(0.055)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.988	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.997)	(1.000)	(1.000)	(0.994)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.024	0.017	0.014	0.005	0.008	0.014	0.016	0.013	0.009
		(0.039)	(0.022)	(0.016)	(0.005)	(0.008)	(0.012)	(0.021)	(0.017)	(0.013)
	1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.977	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.996)	(1.000)	(1.000)	(0.986)	(1.000)	(1.000)
Case 3	H_0: W^r=$[X,Z_1]$ (i.e., a partially linear type model)									
			DGP_1			DGP_2			DGP_3	
	δ=δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.108	0.104	0.095	0.117	0.109	0.104	0.126	0.115	0.107
		(0.115)	(0.125)	(0.113)	(0.127)	(0.114)	(0.102)	(0.157)	(0.124)	(0.112)
	1.0	1.000	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.985)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.068	0.055	0.048	0.064	0.058	0.053	0.067	0.060	0.056
		(0.078)	(0.067)	(0.055)	(0.078)	(0.068)	(0.052)	(0.079)	(0.064)	(0.057)
	1.0	1.000	1.000	1.000	0.965	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.974)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.019	0.014	0.009	0.024	0.018	0.012	0.024	0.017	0.014
		(0.025)	(0.019)	(0.014)	(0.039)	(0.028)	(0.014)	(0.031)	(0.026)	(0.017)
	1.0	0.986	1.000	1.000	0.941	1.000	1.000	0.986	1.000	1.000
		(0.996)	(1.000)	(1.000)	(0.953)	(1.000)	(1.000)	(0.993)	(1.000)	(1.000)

Note: Empirical size and power are calculated based on 1000 simulations with 299 bootstrap repetitions. The *cross-validation least-square* (CVLS) bandwidths are employed for regression estimation, and α is the significance level.

that in Case 1 and DGP_2 increasing in a slightly slower rate. We observe that $\hat{T}_c^*$ outperforms $\hat{T}^*$ in terms of size, whereas $\hat{T}^*$ exhibits empirical power slightly higher than $\hat{T}_c^*$, particularly with small sample size $n = (50, 100)$. When we change the bandwidth magnitude with $C_h = 0.5$ in Table C.2 and $C_h = 1.5$ in Table

C.3, most of the observations reached above remain intact, but the performance of the tests is sensitive to the choice of C_h in the bandwidth. The size of the tests is influenced differently by the magnitude of C_h across DGP_{1-3} in three cases. The power of the test increases with the constant C_h across all cases and DGPs, and reaches one when $n = 200$ except for DGP_2 in Case 1 when $C_h = 0.5$. Overall, the empirical results confirm the validity of our proposed tests.

C.2 Bivariate Case: CVLS Bandwidth Selection

Here we investigate the performance of $\hat{T}_c^*$ and $\hat{T}^*$ by selecting the bandwidth h_{cvls} with a CVLS approach. Specifically,

$$h_{cvls} = \underset{h}{argmin}\ \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}_{-i}(W_i)\right)^2,$$

where $\hat{m}_{-i}(W_i)$ is the *leave-one-out* local linear estimate (i.e., $p' = p - 1 = 1$ for $p = 2$). We adopt the same experiment design as in the bivariate case above, and present the results in Table C.4. With small samples, i.e., $n = (50, 100)$, both tests are slightly undersized in Case 1 and DGP_2 of Case 2, and oversized otherwise. Compared to the results with *rule-of-thumb* bandwidths in Table C.1-C.3, the power in Case 2 and 3 is similar or larger relative to that with $C_h = 0.5$, but slightly smaller than that with $C_h = (1.0, 1.5)$. Nonetheless, the size (power) approaches the nominal level (unity) as n reaches 200, a similar observation made in Tables C.1-C.3. The results suggest that our tests with h_{cvls}, the optimal regression bandwidth, perform reasonably well, and do not exhibit significant changes in both size and power across all DGPs, especially with a large sample. This is consistent with our arguments on the use of the optimal regression bandwidth when $d < 4$ (see Section 5.1).

C.3 Trivariate Case

We further explore trivariate regressions ($d = 3$) with $W = [X, Z']'$, and $Z = [Z_1, Z_2]'$. We test the null, with $W^r = Z$ in Case 1, $W^r = X$ in Case 2, and $W^r = [X, Z']'$ in Case 3. In addition, the trivariate regression model allows us to test the null with $W^r = Z_s$ in Case 1.1, and with $W^r = [X, Z_s]'$ in Case 2.1, for $s = 1, 2$. Correspondingly, the null is satisfied by $m(W) = m_1(X, Z_{-s}) + m_2(Z)$, an overlapping additive model for Case 1.1, and $m(W) = Xm_1(Z_{-s}) + m_2(Z)$, an overlapping varying coefficient model for Case 2.1, with Z_{-s}

Table C.5: Empirical Size and Power from Trivariate Regressions for $\hat{T}_c^*$ ($\hat{T}^*$) with $d = 3$ and $C_h = 1.0$

Case 1		H_0: W^r=$[Z_1, Z_2]'$ (i.e., an additive model with δ=1)																	
		DGP_4						DGP_5						DGP_6					
	$\delta_1 = \delta_2$	50		100		200		50		100		200		50		100		200	
$\alpha = 0.10$	0.0	0.136	(0.160)	0.124	(0.174)	0.108	(0.112)	0.135	(0.158)	0.123	(0.134)	0.111	(0.117)	0.158	(0.192)	0.136	(0.142)	0.113	(0.126)
	1.0	0.958	(0.966)	1.000	(1.000)	1.000	(1.000)	0.910	(0.925)	1.000	(1.000)	1.000	(1.000)	0.947	(0.955)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.05$	0.0	0.086	(0.062)	0.076	(0.078)	0.066	(0.072)	0.074	(0.106)	0.065	(0.077)	0.057	(0.042)	0.097	(0.130)	0.074	(0.088)	0.057	(0.058)
	1.0	0.931	(0.940)	1.000	(1.000)	1.000	(1.000)	0.846	(0.854)	0.946	(0.954)	1.000	(1.000)	0.915	(0.920)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.01$	0.0	0.024	(0.018)	0.018	(0.020)	0.015	(0.014)	0.035	(0.078)	0.028	(0.032)	0.016	(0.017)	0.035	(0.052)	0.020	(0.026)	0.016	(0.015)
	1.0	0.871	(0.886)	1.000	(1.000)	1.000	(1.000)	0.769	(0.787)	0.934	(0.948)	1.000	(1.000)	0.824	(0.834)	0.990	(0.997)	1.000	(1.000)
Case 1.1		H_0: W^r=Z_1 (i.e., an overlapping additive model with $\delta = \delta_2$=1)																	
		DGP_4						DGP_5						DGP_6					
	δ_1	50		100		200		50		100		200		50		100		200	
$\alpha = 0.10$	0.0	0.124	(0.106)	0.116	(0.120)	0.107	(0.113)	0.157	(0.206)	0.139	(0.134)	0.119	(0.080)	0.138	(0.166)	0.125	(0.110)	0.111	(0.080)
	1.0	0.912	(0.920)	0.991	(0.996)	1.000	(1.000)	0.809	(0.828)	0.921	(0.932)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.05$	0.0	0.075	(0.042)	0.065	(0.072)	0.060	(0.062)	0.090	(0.112)	0.078	(0.074)	0.062	(0.034)	0.084	(0.106)	0.077	(0.060)	0.061	(0.042)
	1.0	0.815	(0.827)	0.979	(0.984)	1.000	(1.000)	0.717	(0.735)	0.857	(0.869)	0.990	(0.998)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.01$	0.0	0.030	(0.018)	0.021	(0.016)	0.015	(0.010)	0.059	(0.092)	0.035	(0.074)	0.021	(0.032)	0.041	(0.060)	0.028	(0.020)	0.017	(0.006)
	1.0	0.683	(0.694)	0.940	(0.958)	1.000	(1.000)	0.588	(0.607)	0.748	(0.779)	0.946	(0.954)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)
Case 2		H_0: W^r=X (i.e., a varying coefficient type model with $\delta_1 = \delta_2$=1)																	
		DGP_4						DGP_5						DGP_6					
	δ	50		100		200		50		100		200		50		100		200	
$\alpha = 0.10$	0.0	0.156	(0.186)	0.133	(0.142)	0.109	(0.104)	0.141	(0.164)	0.123	(0.078)	0.106	(0.080)	0.152	(0.182)	0.125	(0.146)	0.105	(0.098)
	1.0	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)	0.948	(0.953)	1.000	(1.000)	1.000	(1.000)	0.864	(0.875)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.05$	0.0	0.099	(0.160)	0.071	(0.082)	0.056	(0.048)	0.080	(0.086)	0.066	(0.068)	0.055	(0.044)	0.106	(0.130)	0.078	(0.074)	0.054	(0.054)
	1.0	0.995	(1.000)	1.000	(1.000)	1.000	(1.000)	0.914	(0.925)	1.000	(1.000)	1.000	(1.000)	0.825	(0.832)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.01$	0.0	0.032	(0.043)	0.026	(0.038)	0.013	(0.008)	0.024	(0.036)	0.015	(0.014)	0.008	(0.013)	0.058	(0.082)	0.036	(0.045)	0.014	(0.017)
	1.0	0.974	(0.989)	1.000	(1.000)	1.000	(1.000)	0.896	(0.903)	1.000	(1.000)	1.000	(1.000)	0.688	(0.712)	1.000	(1.000)	1.000	(1.000)
Case 2.1		H_0: W^r=$[X, Z_1]'$ (i.e., a overlapping varying coefficient type model with δ_2=1)																	
		DGP_4						DGP_5						DGP_6					
	$\delta = \delta_1$	50		100		200		50		100		200		50		100		200	
$\alpha = 0.10$	0.0	0.128	(0.122)	0.115	(0.108)	0.105	(0.104)	0.148	(0.172)	0.122	(0.128)	0.105	(0.094)	0.125	(0.146)	0.112	(0.122)	0.105	(0.108)
	1.0	0.980	(0.996)	1.000	(1.000)	1.000	(1.000)	0.903	(0.911)	1.000	(1.000)	1.000	(1.000)	0.990	(0.995)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.05$	0.0	0.067	(0.058)	0.061	(0.044)	0.055	(0.042)	0.074	(0.088)	0.065	(0.061)	0.056	(0.044)	0.070	(0.076)	0.060	(0.056)	0.048	(0.050)
	1.0	0.974	(0.981)	1.000	(1.000)	1.000	(1.000)	0.855	(0.869)	1.000	(1.000)	1.000	(1.000)	0.966	(0.979)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.01$	0.0	0.022	(0.018)	0.015	(0.010)	0.007	(0.011)	0.025	(0.036)	0.017	(0.014)	0.013	(0.012)	0.029	(0.036)	0.017	(0.018)	0.008	(0.011)
	1.0	0.957	(0.968)	1.000	(1.000)	1.000	(1.000)	0.794	(0.805)	1.000	(1.000)	1.000	(1.000)	0.942	(0.958)	1.000	(1.000)	1.000	(1.000)
Case 3		H_0: W^r=$[X, Z_1, Z_2]$ (i.e., a partially linear type model)																	
		DGP_4						DGP_5						DGP_6					
	$\delta = \delta_1 = \delta_2$	50		100		200		50		100		200		50		100		200	
$\alpha = 0.10$	0.0	0.135	(0.169)	0.110	(0.138)	0.102	(0.104)	0.128	(0.162)	0.119	(0.124)	0.106	(0.096)	0.122	(0.154)	0.108	(0.120)	0.099	(0.114)
	1.0	0.992	(0.998)	1.000	(1.000)	1.000	(1.000)	0.874	(0.887)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.05$	0.0	0.075	(0.089)	0.066	(0.072)	0.053	(0.054)	0.078	(0.098)	0.061	(0.064)	0.056	(0.060)	0.072	(0.084)	0.056	(0.075)	0.048	(0.052)
	1.0	0.988	(0.994)	1.000	(1.000)	1.000	(1.000)	0.822	(0.839)	1.000	(1.000)	1.000	(1.000)	0.991	(0.999)	1.000	(1.000)	1.000	(1.000)
$\alpha = 0.01$	0.0	0.021	(0.035)	0.015	(0.022)	0.012	(0.009)	0.026	(0.036)	0.018	(0.025)	0.013	(0.014)	0.021	(0.036)	0.015	(0.023)	0.008	(0.015)
	1.0	0.968	(0.979)	1.000	(1.000)	1.000	(1.000)	0.793	(0.806)	1.000	(1.000)	1.000	(1.000)	0.942	(0.955)	1.000	(1.000)	1.000	(1.000)

Table C.6: Empirical Size and Power from Trivariate Regressions for $\hat{T}_c^*$ ($\hat{T}^*$) with $d = 3$ and $C_h = 0.5$

Case 1	H_0: W^r=$[Z_1, Z_2]'$ (i.e., an additive model with δ=1)									
		DGP_4			DGP_5			DGP_6		
	$\delta_1 = \delta_2$	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.153 (0.179)	0.124 (0.138)	0.109 (0.120)	0.122 (0.141)	0.113 (0.112)	0.105 (0.090)	0.155 (0.182)	0.131 (0.155)	0.116 (0.134)
	1.0	0.915 (0.927)	0.986 (0.997)	1.000 (1.000)	0.805 (0.818)	0.917 (0.934)	1.000 (1.000)	0.927 (0.946)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.081 (0.089)	0.065 (0.068)	0.057 (0.059)	0.077 (0.094)	0.059 (0.064)	0.048 (0.060)	0.090 (0.166)	0.074 (0.093)	0.057 (0.044)
	1.0	0.864 (0.889)	0.944 (0.952)	1.000 (1.000)	0.725 (0.732)	0.889 (0.900)	0.925 (0.944)	0.889 (0.914)	0.995 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.036 (0.024)	0.021 (0.019)	0.015 (0.016)	0.042 (0.062)	0.025 (0.023)	0.008 (0.014)	0.046 (0.067)	0.028 (0.034)	0.014 (0.017)
	1.0	0.817 (0.823)	0.890 (0.906)	0.991 (0.997)	0.506 (0.604)	0.864 (0.887)	0.904 (0.920)	0.800 (0.811)	0.984 (0.995)	1.000 (1.000)
Case 1.1	H_0: W^r=Z_1 (i.e., an overlapping additive model with $\delta = \delta_2$=1)									
		DGP_4			DGP_5			DGP_6		
	δ_1	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.144 (0.165)	0.123 (0.126)	0.107 (0.124)	0.177 (0.218)	0.146 (0.126)	0.108 (0.106)	0.135 (0.142)	0.118 (0.127)	0.108 (0.106)
	1.0	0.901 (0.906)	0.974 (0.969)	1.000 (1.000)	0.784 (0.789)	0.889 (0.867)	0.992 (0.988)	0.957 (0.969)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.079 (0.069)	0.064 (0.054)	0.059 (0.068)	0.104 (0.125)	0.071 (0.085)	0.056 (0.044)	0.072 (0.084)	0.065 (0.068)	0.055 (0.044)
	1.0	0.809 (0.815)	0.935 (0.929)	1.000 (1.000)	0.721 (0.729)	0.819 (0.807)	0.984 (0.976)	0.933 (0.941)	1.000 (0.991)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.037 (0.026)	0.025 (0.018)	0.018 (0.019)	0.058 (0.088)	0.032 (0.070)	0.020 (0.028)	0.025 (0.034)	0.015 (0.018)	0.009 (0.007)
	1.0	0.671 (0.677)	0.869 (0.867)	1.000 (0.997)	0.579 (0.582)	0.694 (0.688)	0.956 (0.942)	0.864 (0.879)	0.997 (0.973)	1.000 (1.000)
Case 2	H_0: W^r=X (i.e., a varying coefficient type model with $\delta_1 = \delta_2$=1)									
		DGP_4			DGP_5			DGP_6		
	δ	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.157 (0.174)	0.134 (0.125)	0.108 (0.111)	0.152 (0.177)	0.125 (0.112)	0.109 (0.106)	0.157 (0.185)	0.129 (0.155)	0.107 (0.106)
	1.0	0.896 (0.904)	0.980 (0.988)	1.000 (1.000)	0.806 (0.824)	0.918 (1.000)	1.000 (1.000)	0.752 (0.768)	0.982 (0.989)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.109 (0.147)	0.068 (0.061)	0.054 (0.055)	0.081 (0.085)	0.064 (0.062)	0.046 (0.040)	0.109 (0.145)	0.079 (0.069)	0.057 (0.059)
	1.0	0.794 (0.806)	0.972 (0.978)	1.000 (1.000)	0.704 (0.726)	0.846 (1.000)	1.000 (1.000)	0.700 (0.708)	0.948 (0.954)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.047 (0.042)	0.022 (0.015)	0.015 (0.014)	0.027 (0.045)	0.018 (0.027)	0.006 (0.007)	0.066 (0.098)	0.035 (0.064)	0.018 (0.023)
	1.0	0.675 (0.687)	0.933 (0.944)	1.000 (1.000)	0.584 (0.592)	0.736 (1.000)	0.994 (0.998)	0.674 (0.689)	0.811 (0.826)	1.000 (0.999)
Case 2.1	H_0: W^r=$[X, Z_1]'$ (i.e., a overlapping varying coefficient type model with δ_2=1)									
		DGP_4			DGP_5			DGP_6		
	$\delta = \delta_1$	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.134 (0.149)	0.115 (0.129)	0.104 (0.108)	0.135 (0.163)	0.116 (0.136)	0.110 (0.116)	0.127 (0.156)	0.115 (0.136)	0.106 (0.112)
	1.0	0.979 (0.988)	1.000 (1.000)	1.000 (1.000)	0.854 (0.866)	0.958 (0.973)	1.000 (1.000)	0.985 (0.989)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.069 (0.062)	0.060 (0.043)	0.048 (0.046)	0.074 (0.079)	0.059 (0.054)	0.053 (0.046)	0.074 (0.089)	0.062 (0.061)	0.056 (0.048)
	1.0	0.944 (0.951)	0.991 (0.998)	1.000 (1.000)	0.806 (0.818)	0.951 (0.965)	1.000 (1.000)	0.911 (0.924)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.035 (0.049)	0.019 (0.028)	0.008 (0.014)	0.027 (0.031)	0.018 (0.020)	0.012 (0.017)	0.034 (0.046)	0.021 (0.022)	0.015 (0.017)
	1.0	0.908 (0.914)	0.975 (0.982)	1.000 (1.000)	0.711 (0.722)	0.924 (0.935)	1.000 (1.000)	0.866 (0.876)	0.991 (0.998)	1.000 (1.000)
Case 3	H_0: W^r=$[X,Z_1,Z_2]$ (i.e., a partially linear type model)									
		DGP_4			DGP_5			DGP_6		
	$\delta = \delta_1 = \delta_2$	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.130 (0.164)	0.114 (0.149)	0.105 (0.121)	0.135 (0.149)	0.114 (0.115)	0.096 (0.110)	0.129 (0.171)	0.108 (0.132)	0.097 (0.109)
	1.0	0.986 (0.997)	1.000 (1.000)	1.000 (1.000)	0.872 (0.885)	1.000 (1.000)	1.000 (1.000)	0.980 (0.987)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.081 (0.098)	0.067 (0.081)	0.052 (0.045)	0.082 (0.090)	0.065 (0.063)	0.045 (0.055)	0.082 (0.104)	0.063 (0.089)	0.046 (0.064)
	1.0	0.968 (0.970)	1.000 (1.000)	1.000 (1.000)	0.820 (0.831)	1.000 (1.000)	1.000 (1.000)	0.968 (0.971)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.024 (0.047)	0.018 (0.031)	0.007 (0.015)	0.030 (0.028)	0.019 (0.012)	0.007 (0.008)	0.029 (0.035)	0.015 (0.021)	0.008 (0.017)
	1.0	0.922 (0.936)	1.000 (1.000)	1.000 (1.000)	0.755 (0.761)	0.928 (0.939)	1.000 (1.000)	0.937 (0.954)	1.000 (1.000)	1.000 (1.000)

Table C.7: Empirical Size and Power from Trivariate Regressions for $\hat{T}_c^*$ ($\hat{T}^*$) with $d = 3$ and $C_h = 1.5$

Case 1		H_0: W^r=$[Z_1, Z_2]'$ (i.e., an additive model with δ=1)								
			DGP_4			DGP_5			DGP_6	
	$\delta_1 = \delta_2$	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.148 (0.166)	0.120 (0.127)	0.110 (0.114)	0.117 (0.125)	0.108 (0.109)	0.095 (0.106)	0.145 (0.174)	0.121 (0.125)	0.109 (0.108)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.956 (0.967)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.074 (0.072)	0.059 (0.064)	0.055 (0.053)	0.067 (0.078)	0.054 (0.066)	0.046 (0.058)	0.084 (0.095)	0.070 (0.067)	0.055 (0.063)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.981 (0.989)	1.000 (1.000)	1.000 (1.000)	0.906 (0.918)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.028 (0.015)	0.018 (0.023)	0.012 (0.006)	0.024 (0.044)	0.011 (0.018)	0.007 (0.007)	0.034 (0.046)	0.024 (0.013)	0.012 (0.009)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.968 (0.975)	1.000 (1.000)	1.000 (1.000)	0.841 (0.852)	1.000 (1.000)	1.000 (1.000)
Case 1.1		H_0: W^r=Z_1 (i.e., an overlapping additive model with $\delta = \delta_2$=1)								
			DGP_4			DGP_5			DGP_6	
	δ_1	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.112 (0.136)	0.105 (0.115)	0.094 (0.105)	0.152 (0.174)	0.135 (0.138)	0.116 (0.118)	0.120 (0.125)	0.112 (0.108)	0.106 (0.092)
	1.0	0.934 (0.942)	1.000 (1.000)	1.000 (1.000)	0.839 (0.846)	0.959 (0.970)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.072 (0.074)	0.057 (0.059)	0.044 (0.054)	0.088 (0.106)	0.062 (0.068)	0.047 (0.055)	0.068 (0.060)	0.059 (0.056)	0.047 (0.048)
	1.0	0.844 (0.852)	0.990 (0.996)	1.000 (1.000)	0.774 (0.789)	0.924 (0.940)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.025 (0.019)	0.015 (0.013)	0.006 (0.006)	0.039 (0.071)	0.024 (0.054)	0.009 (0.017)	0.023 (0.015)	0.013 (0.011)	0.007 (0.008)
	1.0	0.691 (0.703)	0.975 (0.986)	1.000 (1.000)	0.672 (0.684)	0.907 (0.922)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
Case 2		H_0: W^r=X (i.e., a varying coefficient type model with $\delta_1 = \delta_2$=1)								
			DGP_4			DGP_5			DGP_6	
	δ	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.121 (0.156)	0.110 (0.131)	0.107 (0.108)	0.134 (0.154)	0.122 (0.119)	0.107 (0.091)	0.143 (0.168)	0.128 (0.126)	0.105 (0.088)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.986 (0.988)	1.000 (1.000)	1.000 (1.000)	0.947 (0.956)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.085 (0.122)	0.064 (0.059)	0.056 (0.047)	0.075 (0.072)	0.061 (0.055)	0.054 (0.048)	0.089 (0.124)	0.064 (0.062)	0.056 (0.044)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.957 (0.965)	1.000 (1.000)	1.000 (1.000)	0.897 (0.903)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.037 (0.028)	0.021 (0.014)	0.015 (0.007)	0.024 (0.028)	0.016 (0.012)	0.008 (0.009)	0.054 (0.068)	0.033 (0.036)	0.014 (0.012)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.905 (0.913)	1.000 (1.000)	1.000 (1.000)	0.833 (0.843)	1.000 (1.000)	1.000 (1.000)
Case 2.1		H_0: W^r=$[X, Z_1]'$ (i.e., a overlapping varying coefficient type model with δ_2=1)								
			DGP_4			DGP_5			DGP_6	
	$\delta = \delta_1$	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.139 (0.157)	0.118 (0.131)	0.106 (0.089)	0.129 (0.148)	0.107 (0.116)	0.095 (0.092)	0.121 (0.138)	0.111 (0.104)	0.105 (0.092)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.072 (0.065)	0.065 (0.057)	0.054 (0.053)	0.075 (0.071)	0.054 (0.058)	0.044 (0.052)	0.070 (0.063)	0.059 (0.048)	0.048 (0.045)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.987 (0.995)	1.000 (1.000)	1.000 (1.000)	0.971 (0.983)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.038 (0.034)	0.020 (0.009)	0.008 (0.007)	0.034 (0.028)	0.018 (0.009)	0.008 (0.008)	0.025 (0.018)	0.015 (0.014)	0.009 (0.006)
	1.0	0.991 (0.999)	1.000 (1.000)	1.000 (1.000)	0.959 (0.970)	1.000 (1.000)	1.000 (1.000)	0.954 (0.967)	1.000 (1.000)	1.000 (1.000)
Case 3		H_0: W^r=$[X,Z_1,Z_2]$ (i.e., a partially linear type model)								
			DGP_4			DGP_5			DGP_6	
	$\delta = \delta_1 = \delta_2$	50	100	200	50	100	200	50	100	200
$\alpha = 0.10$	0.0	0.125 (0.141)	0.108 (0.124)	0.098 (0.106)	0.124 (0.142)	0.108 (0.121)	0.095 (0.091)	0.121 (0.138)	0.105 (0.118)	0.094 (0.096)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.05$	0.0	0.076 (0.062)	0.052 (0.059)	0.047 (0.048)	0.075 (0.086)	0.058 (0.061)	0.046 (0.043)	0.071 (0.065)	0.059 (0.056)	0.048 (0.044)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.984 (0.996)	0.964 (0.975)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
$\alpha = 0.01$	0.0	0.017 (0.022)	0.011 (0.018)	0.007 (0.006)	0.025 (0.032)	0.015 (0.019)	0.008 (0.016)	0.027 (0.022)	0.017 (0.013)	0.006 (0.007)
	1.0	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	0.961 (0.975)	1.000 (1.000)	1.000 (1.000)	0.984 (0.992)	1.000 (1.000)	1.000 (1.000)

denoting variables in Z excluding Z_s. To accommodate the nonlinearity of X and its interaction with Z_1 and Z_2, we consider the following

$$DGP_4 : Y_i = 0.5 + X_i + \delta X_i^2 + \delta_1 X_i Z_{1i} + \delta_2 X_i Z_{2i} + Z_{1i}^2 + Z_{2i}^2 + Z_{1i} Z_{2i} + \epsilon_i$$

$$DGP_5 : Y_i = 5 + 2X_i - \delta e^{1.1X_i} + 2\delta_1 X_i \sin(Z_{1i}) + \delta_2 X_i \cos(-Z_{2i}) + Z_{1i}^3 + Z_{2i}^3 + \epsilon_i$$

$$DGP_6 : Y_i = 1 + X_i + \delta X_i^3 - \delta_1 X_i e^{Z_{1i}} + \delta_2 X_i cos(\pi Z_{2i}) + 0.4(Z_{1i}^2 + Z_{2i}^2) + \epsilon_i,$$

where Z_{2i} is iid and generated from $U(-2, 2)$, and all other variables are generated as in the bivariate study. Note that δ controls for the degree of nonlinearity of X, δ_1 for the interaction between X and Z_1, and δ_2 for the interaction between X and Z_2. Note that Z_s can be either Z_1 or Z_2 in Cases 1.1 and 2.1, and to save space, we focus on $Z_s = Z_1$ below for illustration. Under DGP_{4-6}, we investigate the size by setting $(\delta, \delta_1, \delta_2) = (1, 0, 0)$ in Case 1, $(\delta, \delta_1, \delta_2) = (1, 0, 1)$ in Case 1.1 ($W^r = Z_1$), $(\delta, \delta_1, \delta_2) = (0, 1, 1)$ in Case 2, $(\delta, \delta_1, \delta_2) = (0, 0, 1)$ in Case 2.1 ($W^r = [X, Z_1]'$), and $(\delta, \delta_1, \delta_2) = (0, 0, 0)$ in Case 3. We explore the power performance by simply setting $(\delta, \delta_1, \delta_2) = (1, 1, 1)$ in each case.

We summarize the simulation results with $C_h = 1.0$ for $\hat{T}_c^*$ ($\hat{T}^*$) in Table C.5. The results with $C_h = 0.5$ and $C_h = 1.5$ are reported in Table C.6 and C.7, respectively. Due to the *curse of dimensionality*, we expect deterioration of size and power performance relative to the results in Section C.1. Indeed, the size across different choices of C_h deviates slightly more away from the nominal level, relative to the corresponding bivariate DGPs, and the exception appears to be Case 3. Tables C.5, C.6, and C.7 show that the tests are generally over-sized, at least in small samples ($n = 50$, or 100). However, the size improves rapidly towards the nominal level as the sample size increases, regardless of the choice of C_h. For the small sample ($n = 50$) across all cases, the tests in the trivariate DGP_{4-6} in Tables C.5, C.6, and C.7 generally exhibit lower empirical power for Cases 1-3 relative to those in the bivariate DGPs. Similar to bivariate cases, a larger constant C_h leads to a higher empirical power of both tests, which approaches one as n increases. The large sample results are still reasonable. At $n = 200$, for instance, the size of the two tests in most cases across DGPs is fairly close the target nominal level, and the power is almost one.

Table C.8: Empirical Size and Power for $\hat{T}_c^*(\hat{T}^*)$ in PLM with $d = 3$ (CVLS)

Case 1		H_0: W^r=Z_1 (i.e., an additive model with δ=1)								
			DGP_1			DGP_2			DGP_3	
	δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.127	0.112	0.106	0.075	0.088	0.096	0.078	0.086	0.095
		(0.085)	(0.075)	(0.088)	(0.043)	(0.077)	(0.090)	(0.074)	(0.080)	(0.092)
	1.0	0.942	0.987	1.000	0.915	0.982	1.000	0.969	1.000	1.000
		(0.951)	(0.994)	(1.000)	(0.927)	(0.998)	(1.000)	(0.987)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.075	0.064	0.057	0.034	0.047	0.053	0.036	0.040	0.047
		(0.046)	(0.039)	(0.044)	(0.026)	(0.037)	(0.046)	(0.030)	(0.039)	(0.045)
	1.0	0.859	0.946	1.000	0.876	0.929	1.000	0.909	0.979	1.000
		(0.872)	(0.961)	(1.000)	(0.886)	(0.942)	(1.000)	(0.920)	(0.999)	(1.000)
$\alpha = 0.01$	0.0	0.032	0.022	0.013	0.005	0.008	0.013	0.004	0.007	0.011
		(0.015)	(0.012)	(0.007)	(0.004)	(0.007)	(0.012)	(0.003)	(0.006)	(0.009)
	1.0	0.842	0.914	1.000	0.801	0.881	0.943	0.870	0.964	1.000
		(0.852)	(0.938)	(1.000)	(0.814)	(0.891)	(0.952)	(0.882)	(0.988)	(1.000)
Case 2		H_0: W^r=X (i.e., a varying coefficient type model with δ_1=1)								
			DGP_1			DGP_2			DGP_3	
	δ	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.112	0.104	0.097	0.085	0.095	0.104	0.125	0.113	0.102
		(0.151)	(0.125)	(0.109)	(0.070)	(0.084)	(0.092)	(0.158)	(0.128)	(0.107)
	1.0	1.000	1.000	1.000	0.987	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.994)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.066	0.053	0.048	0.034	0.045	0.054	0.072	0.054	0.048
		(0.089)	(0.075)	(0.060)	(0.030)	(0.039)	(0.045)	(0.085)	(0.066)	(0.056)
	1.0	1.000	1.000	1.000	0.974	1.000	1.000	0.988	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.991)	(1.000)	(1.000)	(0.992)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.019	0.012	0.008	0.006	0.008	0.012	0.019	0.013	0.008
		(0.042)	(0.026)	(0.017)	(0.004)	(0.007)	(0.014)	(0.025)	(0.018)	(0.014)
	1.0	1.000	1.000	1.000	0.965	0.987	1.000	0.971	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.987)	(0.997)	(1.000)	(0.983)	(1.000)	(1.000)
Case 3		H_0: W^r=[X,Z_1] (i.e., a partially linear type model)								
			DGP_1			DGP_2			DGP_3	
	δ=δ_1	$n = 50$	100	200	$n = 50$	100	200	$n = 50$	100	200
$\alpha = 0.10$	0.0	0.108	0.101	0.096	0.124	0.114	0.105	0.132	0.116	0.104
		(0.124)	(0.125)	(0.116)	(0.130)	(0.119)	(0.104)	(0.159)	(0.122)	(0.113)
	1.0	1.000	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.981)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.05$	0.0	0.065	0.054	0.047	0.075	0.064	0.054	0.076	0.062	0.053
		(0.081)	(0.069)	(0.057)	(0.081)	(0.069)	(0.055)	(0.078)	(0.068)	(0.060)
	1.0	1.000	1.000	1.000	0.966	1.000	1.000	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(0.972)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
$\alpha = 0.01$	0.0	0.021	0.012	0.008	0.024	0.016	0.008	0.023	0.016	0.011
		(0.039)	(0.020)	(0.015)	(0.042)	(0.033)	(0.015)	(0.032)	(0.028)	(0.018)
	1.0	0.978	1.000	1.000	0.932	1.000	1.000	0.981	1.000	1.000
		(0.981)	(1.000)	(1.000)	(0.944)	(1.000)	(1.000)	(0.990)	(1.000)	(1.000)

Note: Empirical size and power are calculated based on 1000 simulations with 299 bootstrap repetitions. The *cross-validation least-square* (CVLS) bandwidths are employed in estimating β and constructing $\hat{T}_c^*$ and $\hat{T}^*$, and α is the significance level.

C.4 Extension to a Partially Linear Model

Our tests are constructed on the basis of a fully nonparametric regression, thus limiting its empirical applicability due to the *curse of dimensionality*. In this section, we consider applying our tests to a semiparametric partially linear model, where the imposed structure allows a wider application possibility. The fully nonparametric regression with an iid sample $\{Y_i, X_i, Z_{1i}, Z_{2i}\}_{i=1}^n$ is

$$Y_i = m(X_i, Z_{1i}, Z_{2i}) + \epsilon_i, \;\; i = 1, ..., n, \tag{C.1}$$

where $X \in \Re$, $Z_1 \in \Re^{q_1}$, $Z_2 \in \Re^{q_2}$, $Z = [Z_1, Z_2]'$ and $q_1 + q_2 = d - 1$. Suppose we know that $Z_2 \in \mathbb{R}^{q_2}$ enter the model linearly, but are uncertain of the impact of $W_1 = [X, Z_1]'$, where ideally a relatively low dimension of W_1 can be empirically fruitful. So we consider a partially linear model (PLM)

$$Y_i = m(X_i, Z_{1i}) + Z_{2i}'\beta + \epsilon_i, \tag{C.2}$$

where the dependent variable Y is influenced parametrically by $Z_2'\beta$ and nonparametrically through the unknown smooth function $m(\cdot)$. Clearly, the PLM is more practical than a fully nonparametric regression for empirical applications, since only $q_1 + 1$ variables enter (C.2) nonparametrically. Here our interests lie in inferring certain aspects of the model structure in $m(X, Z_1)$. If β were known, we can easily construct a new dependent variable $\tilde{Y}_i = Y_i - Z_{2i}'\beta$ and employ our test in the regression $\tilde{Y}_i = m(X_i, Z_{1i}) + u_i$. In practice, β were unknown so we replace them with the $\sqrt{n}$-consistent estimate $\hat{\beta}$ by Robinson (1988). To evaluate the performance of $\hat{T}_c^*$ and $\hat{T}^*$ in (C.2), we consider the following three DGPs

$$DGP_7: Y_i = 0.75Z_{2i} + X_i + \delta X_i^2 + Z_{1i} + Z_{1i}^2 + \delta_1 X_i Z_{1i} + \epsilon_i$$

$$DGP_8: Y_i = 1.5Z_{2i} + 2X_i - \delta e^{1.1X_i} + Z_{1i}^3 + 2\delta_1 X_i \sin(Z_{1i}) + \epsilon_i$$

$$DGP_9: Y_i = 0.35Z_{2i} + X_i + \delta X_i^3 + 0.4Z_{1i}^2 - \delta_1 X_i e^{Z_{1i}} + \epsilon_i$$

where they are adapted from DGP_{1-3}, by including a linear component of Z_2, which is iid and generated from $U(-2, 2)$. We perform 1000 repetitions in the simulation with 299 repetitions for the bootstrap, for Cases 1-3 in DGP_{7-9}. Here, we simply select bandwidths in estimating β and in constructing our test through the CVLS criterion.

Table C.8 reports the results. With β being estimated, we expect some minor distortions in the size and power relative to our bivariate studies in Section C.1. With small sample sizes ($n = 50$ or 100), $\hat{T}_c^*$ and $\hat{T}^*$ generally are under-sized in Case 1, and over-sized in Cases 2-3. Nonetheless, the size approaches its nominal level quickly as n reaches 200, similar to the observation in Section C.1. When each DGP deviates from its null, both tests have their power rising rapidly with increasing sample sizes, being unity or very close to unity when $n = 200$ across the designs. In all, our tests show promising finite sample performance in the partially linear model with bandwidth selected with a data-driven tuning strategy. The results suggest that our tests in semiparametric models are suitable for empirical studies.

References

de Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.

Martins-Filho, C., F. Yao, & M. Torero (2018) Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* 34 (1), 23-67.

Masry, E. (1996) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.

Robinson, P. M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.

Wang, X. & K. C. Carriere (2011) Assessing additivity in nonparametric models – a kernel-based method. *Canadian Journal of Statistics* 39 (4), 632–655.

Yang, L., B. U. Park, L. Xue, & W. Härdle (2006) Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association* 101, 1212–1227.

Yao, F. & C. Martins-Filho (2015) An asymptotic characterization of finite degree U-statistics with sample size-dependent kernels: Applications to nonparametric estimators and test statistics. *Communications in Statistics -Theory and Methods* 44 (15), 3251-3265.