SIMULTANEOUS CONFIDENCE BANDS AND HYPOTHESIS TESTING FOR SINGLE-INDEX MODELS

Gaorong Li[a], Heng Peng[b1], Kai Dong[b] and Tiejun Tong[b]

[a] *College of Applied Sciences, Beijing University of Technology, Beijing 100124, P. R. China*

[b] *Department of Mathematics, Hong Kong Baptist University, Hong Kong, P. R. China*

Supplementary Material

This note contains two lemmas and the proofs for Theorem 1–Theorem 5.

S1 Two lemmas

Lemma 1 *Suppose that conditions C1–C3 hold. If* $h \to 0, nh^3 \to \infty$*, we have*

$$\sup_{u\in\mathcal{U}} \left| S_{n,l}^{\boldsymbol{\beta}} - \mu_l f(u) - \mu_{l+1} f'(u) h \right| = O_P(h^2 + \delta_n), \quad l = 0, 1, 2, 3,$$

where $\delta_n = \sqrt{\frac{\log n}{nh}}$*,* $f'(u)$ *is the derivative of* $f(u)$ *and*

$$S_{n,l}^{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\boldsymbol{\beta}^T \boldsymbol{X}_i - u}{h} \right)^l K_h(\boldsymbol{\beta}^T \boldsymbol{X}_i - u), \qquad l = 0, 1, 2, 3.$$

The details of proof can be found from Martins-Filho and Yao (2007). The following lemma provides the uniform convergence rates for the estimators $\hat{\eta}$ and $\hat{\eta}'$ respectively.

Lemma 2 *Let* $\mathcal{B}_n = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le cn^{-1/2}\}$ *for some positive constant c. Suppose that conditions C1–C5 hold, we have*

$$\sup_{u\in\mathcal{U}, \boldsymbol{\beta}\in\mathcal{B}_n} |\hat{\eta}(u; \boldsymbol{\beta}) - \eta(u)| = O_P(\delta_n), \tag{S1.1}$$

and

$$\sup_{u\in\mathcal{U}, \boldsymbol{\beta}\in\mathcal{B}_n} |\hat{\eta}'(u; \boldsymbol{\beta}) - \eta'(u)| = O_P(\delta_n/h). \tag{S1.2}$$

[1]Heng Peng is the corresponding author. E-mail: hpeng@math.hkbu.edu.hk.

The proof of Lemma 2 can be found from the full version of Wang *et al.* (2010) on the web arXiv.org at arXiv:0905.2042, hence we omit the details here.

S2 Proof of Theorem 1

The proof of Theorem 1 can immediately be obtained from Carroll *et al.* (1997), Liang *et al.* (2010), or Chen, Gao and Li (2013). So we omit all the details. □

S3 Proof of Theorem 2

By (2.7) in Section 2 and some simple calculations, we have

$$\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)=\sum_{i=1}^{n}W_{ni}(u;\hat{\boldsymbol{\beta}}_0)Y_i, \tag{S3.1}$$

where

$$W_{ni}(u;\hat{\boldsymbol{\beta}}_0) = \frac{n^{-1}K_h(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i-u)[S_{n,2}^{\hat{\boldsymbol{\beta}}_0}-\{(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i-u)/h\}S_{n,1}^{\hat{\boldsymbol{\beta}}_0}]}{S_{n,0}^{\hat{\boldsymbol{\beta}}_0}S_{n,2}^{\hat{\boldsymbol{\beta}}_0}-(S_{n,1}^{\hat{\boldsymbol{\beta}}_0})^2}.$$

By Lemma 1, we have uniformly for $u\in\mathcal{U}$ and $\boldsymbol{\beta}\in\mathcal{B}_n$,

$$\begin{aligned}&S_{n,0}^{\boldsymbol{\beta}}=f(u)+O_P(h^2+\delta_n),\quad S_{n,1}^{\boldsymbol{\beta}}=O_P(h+\delta_n),\\ &S_{n,2}^{\boldsymbol{\beta}}=\mu_2 f(u)+O_P(h^2+\delta_n),\quad S_{n,3}^{\boldsymbol{\beta}}=O_P(h+\delta_n).\end{aligned} \tag{S3.2}$$

Hence, we have

$$S_{n,0}^{\boldsymbol{\beta}}S_{n,2}^{\boldsymbol{\beta}}-(S_{n,1}^{\boldsymbol{\beta}})^2=\mu_2 f^2(u)+O_P(h^2+\delta_n). \tag{S3.3}$$

By (S3.1), we have

$$\begin{aligned}\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)-\eta(u) &= \sum_{i=1}^{n}W_{ni}(u;\hat{\boldsymbol{\beta}}_0)Y_i-\eta(u)\\ &= \sum_{i=1}^{n}W_{ni}(u;\hat{\boldsymbol{\beta}}_0)\varepsilon_i+\Big[\sum_{i=1}^{n}W_{ni}(u;\hat{\boldsymbol{\beta}}_0)\eta(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i)-\eta(u)\Big]\\ &\quad -\sum_{i=1}^{n}W_{ni}(u;\hat{\boldsymbol{\beta}}_0)[\eta(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i)-\eta(\boldsymbol{\beta}_0^T\boldsymbol{X}_i)]\\ &=: I_1+I_2-I_3.\end{aligned} \tag{S3.4}$$

For I_2, by using Taylor expansion and some calculations, and from (S3.2)–(S3.3) and condition C1, we have

$$\begin{aligned} I_2 &= [S^{\hat{\boldsymbol{\beta}}_0}_{n,0}S^{\hat{\boldsymbol{\beta}}_0}_{n,2} - (S^{\hat{\boldsymbol{\beta}}_0}_{n,1})^2]^{-1}\Big\{\frac{1}{n}\sum_{i=1}^{n} K_h(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)\Big[S^{\hat{\boldsymbol{\beta}}_0}_{n,2} - \{(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)/h\}S^{\hat{\boldsymbol{\beta}}_0}_{n,1}\Big] \\ &\quad \times\Big(\eta(u) + \eta'(u)(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u) + \frac{1}{2}\eta''(u)(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)^2 + O(|\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u|^3)\Big)\Big\} - \eta(u) \\ &= \frac{1}{2}\eta''(u)h^2[S^{\hat{\boldsymbol{\beta}}_0}_{n,0}S^{\hat{\boldsymbol{\beta}}_0}_{n,2} - (S^{\hat{\boldsymbol{\beta}}_0}_{n,1})^2]^{-1}[(S^{\hat{\boldsymbol{\beta}}_0}_{n,2})^2 - S^{\hat{\boldsymbol{\beta}}_0}_{n,1}S^{\hat{\boldsymbol{\beta}}_0}_{n,3}] + o_P(1) \\ &= \frac{1}{2}\eta''(u)h^2\mu_2 + O_P(h^2 + \delta_n). \end{aligned} \tag{S3.5}$$

Now we consider I_3. By Theorem 1, we know that $\hat{\boldsymbol{\beta}}_0$ is a $\sqrt{n}$–consistent estimator, that is, $\hat{\boldsymbol{\beta}}_0$ satisfies that $\hat{\boldsymbol{\beta}}_0 \in \mathcal{B}_n$. Note that $\sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\beta}\in\mathcal{B}_n} |\eta(\boldsymbol{\beta}^T\boldsymbol{x}) - \eta(\boldsymbol{\beta}_0^T\boldsymbol{x})| = O(n^{-1/2})$. Invoking the Lemma 2 in Zhu and Xue (2006), it is easy to show that $I_3 = o_P(n^{-1/2})$.

We approximate the process I_1 as follows. Following the steps of Lemma A.2 in Xia and Li (1999), we have

$$\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}_0^T\boldsymbol{X}_i - u)\{(\boldsymbol{\beta}_0^T\boldsymbol{X}_i - u)/h\}\varepsilon_i = O_P(\delta_n). \tag{S3.6}$$

By the proof of Lemma A.1 in Xia *et al.* (2004), we can obtain that

$$\frac{1}{nf(u)}\sum_{i=1}^{n} K_h(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)\varepsilon_i = \frac{1}{nf(u)}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}_0^T\boldsymbol{X}_i - u)\varepsilon_i + O_P(h\delta_n). \tag{S3.7}$$

By (S3.2), (S3.6)–(S3.7) and using the same argument of (S3.5), it is easy to show that

$$\begin{aligned} I_1 &= [S^{\hat{\boldsymbol{\beta}}_0}_{n,0}S^{\hat{\boldsymbol{\beta}}_0}_{n,2} - (S^{\hat{\boldsymbol{\beta}}_0}_{n,1})^2]^{-1}S^{\hat{\boldsymbol{\beta}}_0}_{n,2}\Big(\frac{1}{n}\sum_{i=1}^{n} K_h(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)\varepsilon_i\Big) \\ &\quad -[S^{\hat{\boldsymbol{\beta}}_0}_{n,0}S^{\hat{\boldsymbol{\beta}}_0}_{n,2} - (S^{\hat{\boldsymbol{\beta}}_0}_{n,1})^2]^{-1}S^{\hat{\boldsymbol{\beta}}_0}_{n,1}\Big(\frac{1}{n}\sum_{i=1}^{n} K_h(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)\{(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i - u)/h\}\varepsilon_i\Big) \\ &= \frac{1}{nf(u)}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}_0\boldsymbol{X}_i - u)\varepsilon_i + O_P(\delta_n(h + \delta_n)) \\ &=: \widetilde{I}_1(u) + O_P(\delta_n(h + \delta_n)) \end{aligned}$$

uniformly for $u \in \mathcal{U}$. This also implies that $\|I_1 - \widetilde{I}_1(u)\|_\infty = O_P(\delta_n(h + \delta_n))$.

Next, we will derive the asymptotic distribution of $\widetilde{I}_1(u)$. For convenience, let

$$\begin{aligned} \{nhf(u)\sigma^{-2}\nu_0^{-1}\}^{1/2}\widetilde{I}_1(u) &= \{nhf(u)\sigma^2\nu_0\}^{-1/2}\sum_{i=1}^{n} K\left(\frac{\boldsymbol{\beta}_0\boldsymbol{X}_i - u}{h}\right)\varepsilon_i \\ &=: \{nhf(u)\sigma^2\nu_0\}^{-1/2}\xi(u). \end{aligned} \tag{S3.8}$$

Divide the interval $[b_1, b_2]$ into N subintervals $J_r = [d_{r-1}, d_r), r = 1, 2, \ldots, N-1, J_N = [d_{N-1}, b_2]$ where $d_r = b_1 + \frac{b_2 - b_1}{N} r$. Define $U_i = \boldsymbol{\beta}_0^T \boldsymbol{X}_i$ and $\widetilde{U}_i = d_r I(U_i \in J_r), r = 1, \ldots, N$, and it is obvious that $U_i - \widetilde{U}_i = O(N^{-1})$. Then, by law of large numbers for the random sequence $\{\varepsilon_i\}, i = 1, \ldots, n$, we have

$$\begin{aligned} \xi(u) &= \sum_{i=1}^n \left[K\left(\frac{U_i - u}{h}\right) - K\left(\frac{\widetilde{U}_i - u}{h}\right) \right] \varepsilon_i + \sum_{i=1}^n K\left(\frac{\widetilde{U}_i - u}{h}\right) \varepsilon_i \\ &= O_P(h^{-1}) + \sum_{i=1}^n K\left(\frac{\widetilde{U}_i - u}{h}\right) \varepsilon_i \\ &=: O_P(h^{-1}) + \widetilde{\xi}(u) \end{aligned} \tag{S3.9}$$

uniformly for $u \in [b_1, b_2]$. By the definition of $\widetilde{U}_i$, we have that

$$\widetilde{\xi}(u) = \sum_{r=1}^N K\left(\frac{d_r - u}{h}\right) \sum_{i=1}^n I(\widetilde{U}_i \in J_r) \varepsilon_i.$$

Let $\widetilde{\xi}_t = \sum_{r=1}^t \sum_{i=1}^n I(U_i \in J_r)\varepsilon_i = \sum_{i=1}^n I(b_1 \leq U_i \leq d_t)\varepsilon_i, \widetilde{\xi}_0 = 0$. Then, by Lemma 2 in Zhang, Fan and Sun (2009), for any $t = 1, \ldots, N$ and $u \in [b_1, b_2]$, we have

$$|\widetilde{\xi}_t - N^{1/2} W(G(d_t))| = O(N^{1/4} \log N) \quad a.s.,$$

where $W(\cdot)$ is a Wiener process and $G(c) = \int_{b_1}^c \sigma^2 f(v) dv$. By Abel's transform, we have

$$\widetilde{\xi}(u) = K\left(\frac{b_2 - u}{n}\right) \widetilde{\xi}_N - \sum_{r=1}^{N-1} \left[K\left(\frac{d_{r+1} - u}{h}\right) - K\left(\frac{d_r - u}{h}\right) \right] \widetilde{\xi}_r$$

and

$$\begin{aligned} &\left\| \sum_{r=1}^{N-1} \left[K\left(\frac{d_{r+1} - u}{h}\right) - K\left(\frac{d_r - u}{h}\right) \right] [\widetilde{\xi}_r - N^{1/2} W(G(d_r))] \right\|_\infty \\ &\quad \leq \left\| \max_{1 \leq r \leq N} |\widetilde{\xi}_r - N^{1/2} W(G(d_r))| \sum_{r=1}^{N-1} \left| K\left(\frac{d_{r+1} - u}{h}\right) - K\left(\frac{d_r - u}{h}\right) \right| \right\| \\ &\quad = O_P(N^{1/4} \log N). \end{aligned}$$

Hence, we have

$$\begin{aligned} \widetilde{\xi}(u) &= N^{1/2} K\left(\frac{b_2 - u}{h}\right) W(G(b_2)) \\ &\quad - N^{1/2} \sum_{r=1}^{N-1} \left[K\left(\frac{d_{r+1} - u}{h}\right) - K\left(\frac{d_r - u}{h}\right) \right] W(G(d_r)) + O_P(N^{1/4} \log N) \end{aligned} \tag{S3.10}$$

uniformly for $u \in [b_1, b_2]$.

For a Wiener process, it is known that (Csörgö and Révész 1981, Page 44)

$$\sup_{t\in[b_1,b_2]} |W(G(t+\varsigma)) - W(G(t))| = O(\{\varsigma \log(1/\varsigma)\}^{1/2}) \quad a.s.$$

when ς is any small number. Using this property and the boundness of $K(\cdot)$, we obtain that

$$\begin{aligned}&\sum_{r=1}^{N-1}\left[K\left(\frac{d_{r+1}-u}{h}\right) - K\left(\frac{d_r-u}{h}\right)\right] W(G(d_r)) \\ &\quad = \int_{b_1}^{b_2} W(G(v))dK\left(\frac{v-u}{h}\right) + O_P(\{N^{-1}\log N\}^{1/2})\end{aligned}$$

uniformly for $u \in [b_1, b_2]$. Together with (S3.9) and (S3.10), it is easy to show that

$$\begin{aligned}&\left\|(nh)^{-1/2}\xi(u) - h^{-1/2}\int_{b_1}^{b_2} K\left(\frac{v-u}{h}\right) dW(G(v))\right\|_\infty \\ &\quad = O_P\big((nh^3)^{-1/2} + (nh)^{-1/2}N^{1/4}\log N\big). \qquad \text{(S3.11)}\end{aligned}$$

Note that the order is $O_P\big((nh^3)^{-1/2} + (nh^2)^{-1/4}\log n\big)$ if N is taken as $N = O(n)$. Let

$$\begin{aligned}Z_{1n}(u) &= h^{-1/2}\int_{b_1}^{b_2} K\left(\frac{v-u}{h}\right) dW(G(v)),\\ Z_{2n}(u) &= h^{-1/2}\int_{b_1}^{b_2} K\left(\frac{v-u}{h}\right) [\sigma^2 f(v)]^{1/2} dW(v-b_1),\\ Z_{3n}(u) &= h^{-1/2}\int_{b_1}^{b_2} K\left(\frac{v-u}{h}\right) dW(v-b_1).\end{aligned}$$

For a Gaussian process, invoking Lemma 2.1–Lemma 2.5 in Claeskens and Van Keilegom (2003), we have

$$\begin{aligned}&\|Z_{1n}(u) - Z_{2n}(u)\|_\infty = O_P(h^{1/2}),\\ &\|(\sigma^2 f(u))^{-1/2} Z_{2n}(u) - Z_{3n}(u)\|_\infty = O_P(h^{1/2}). \qquad \text{(S3.12)}\end{aligned}$$

By (S3.11) and (S3.12), we have

$$\big\|(nh\sigma^2 f(u))^{-1/2}\xi(u) - Z_{3n}(u)\big\|_\infty = O_P\big((nh^3)^{-1/2} + (nh^2)^{-1/4}\log n + h^{1/2}\big).$$

This, together with (S3.8), and invoking Theorem 1 and Theorem 3.1 in Bickel and Rosenblatt (1973), when $h = O(n^{-\rho}), 1/5 < \rho < 1/3$, we have

$$\begin{aligned}&P\Big\{(-2\log\{h/(b_2-b_1)\})^{1/2}\Big(\nu_0^{-1/2}\Big\|(nh\sigma^2 f(u))^{-1/2}\xi(u)\Big\|_\infty - d_{n0}\Big) < x\Big\}\\ &\quad \longrightarrow \exp\big(-2e^{-x}\big).\end{aligned}$$

Summarizing the above results, we complete the proof of Theorem 2. □

S4 Proof of Theorem 3

We also can finish the proof of Theorem 3 along the same lines as the proof of Theorem 2, here we omit the details of proof. □

S5 Proof of Theorem 4

To prove the theorem, we need to derive the rate of convergence for the bias and variance estimators. We first consider the difference between $\widehat{\text{bias}}(\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D})$ and $2^{-1}h^2\mu_2\eta''(u)$. By using the standards as in the proof of Lemma 2, we have

$$\begin{aligned}\|\widehat{\text{bias}}(\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}) - 2^{-1}h^2\mu_2\eta''(u)\|_\infty &= O_P(h^2\{\sqrt{\log n/nh_*^5}\}) \\ &= O_P(h^2(n^{-1/7}\log^{1/2} n)), \qquad \text{(S5.1)}\end{aligned}$$

where $h_* = n^{-1/7}$ comes from the pilot estimation of $\eta''(\cdot)$.

Furthermore, by Lemma 1, we have

$$\left\|\frac{h}{n}\mathbf{X}^T\boldsymbol{W}^2\mathbf{X} - f(u)\widetilde{S}(u)\right\|_\infty = o_P(1),$$

where $\widetilde{S}(u) = \begin{pmatrix} \nu_0 & 0 \\ 0 & \nu_2 \end{pmatrix}$. For the estimator of variance σ^2 defined in Section 2.3, Tong and Wang (2005) showed that $\hat{\sigma}^2$ is a consistent estimator of σ^2 and $\text{bias}(\hat{\sigma}^2) = O(n^{-3+3\tau})$ by taking $m = cn^\tau$ with the constant $c > 0$ and $0 \le \tau \le \frac{1}{3}$.

These results, together with Theorem 1, by some simple calculations, it is easy to show that

$$\left\|nh\widehat{\text{Var}}\{\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}\} - \frac{\nu_0}{f(u)}\sigma^2\right\|_\infty = o_P(1). \qquad \text{(S5.2)}$$

By (S5.1) and (S5.2), and invoking the result of Theorem 2, we finish the proof of Theorem 4. □

S6 Proof of Theorem 5

Under the null hypothesis, model (1.1) reduces to

$$Y_i = \gamma_0 + \gamma_1(\boldsymbol{\beta}_0^T\boldsymbol{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n.$$

For the convenience, we use the matrix and vector notations in the following. Let e_n be an $n \times 1$ vector with all elements being ones, and $\mathbf{X}^* = \Gamma\mathbf{X}$ be an $n \times p$ matrix. By

Theorem 1, it is known that $\hat{\boldsymbol{\beta}}_0$ is a $\sqrt{n}-$consistent estimator of $\boldsymbol{\beta}_0$. By the definitions of $\hat{\boldsymbol{\varepsilon}}^*$ and least squares estimators of γ_0 and γ_1, we have

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}^* &= \Gamma\boldsymbol{\varepsilon} - (\hat{\gamma}_0 - \gamma_0)\Gamma e_n + \gamma_1(\Gamma\mathbf{X}\boldsymbol{\beta}_0) - \hat{\gamma}_1(\Gamma\mathbf{X}\hat{\boldsymbol{\beta}}_0) \\ &= \Gamma\boldsymbol{\varepsilon} - (\hat{\gamma}_0 - \gamma_0)\Gamma e_n - (\hat{\gamma}_1 - \gamma_1)(\mathbf{X}^*\boldsymbol{\beta}_0) \\ &\quad -(\hat{\gamma}_1 - \gamma_1)\mathbf{X}^*(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) - \gamma_1\mathbf{X}^*(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \\ &=: J_1 - J_2 - J_3 - J_4 - J_5,\end{aligned}$$

and $J_1 \sim N(0, \sigma^2 I_n)$. Let J_{ki} denote the ith components of the vectors $J_k, k = 1, \ldots, 5$, respectively. Then we have

$$\sum_{i=1}^{m} \hat{\varepsilon}_i^{*2} = \sum_{i=1}^{m}(J_{1i} - J_{2i} - J_{3i} - J_{4i} - J_{5i})^2. \tag{S6.1}$$

We first consider the orders of $J_{ki}^2, k = 1, \ldots, 5$. By the Cauchy-Schwarz inequality, we have

$$J_{3i}^2 = (\hat{\gamma}_1 - \gamma_1)^2\Big[\sum_{j=1}^{p} X_{ij}^*\beta_{0j}\Big]^2 \le (\hat{\gamma}_1 - \gamma_1)^2\|\boldsymbol{\beta}_0\|^2\sum_{j=1}^{p} X_{ij}^{*2}, \tag{S6.2}$$

where β_{0j} is the jth component of the vector $\boldsymbol{\beta}_0$. Note that $\|\boldsymbol{\beta}_0\| = 1$ and $|\hat{\gamma}_1 - \gamma_1| = O_P(n^{-1/2})$, and from the condition (C6), we have

$$\sum_{i=n_0}^{m} J_{3i}^2 \le (\hat{\gamma}_1 - \gamma_1)^2\sum_{j=1}^{p}\sum_{i=n_0}^{n/(\log\log n)^4} X_{ij}^{*2} = O_P\{(\log\log n)^{-4}\}. \tag{S6.3}$$

From $\|\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$, condition (C6) and the same argument, it is easy to show that

$$\sum_{i=n_0}^{m} J_{4i}^2 = O_P\{n^{-1}(\log\log n)^{-4}\}, \quad \sum_{i=n_0}^{m} J_{5i}^2 = O_P\{(\log\log n)^{-4}\}.$$

By $|\hat{\gamma}_0 - \gamma_0| = O_P(n^{-1/2})$ and the definition of Γ, it is easy to check that $\sum_{i=1}^m J_{2i}^2 = O_P(n^{-1}) = o_P(1)$. Note that

$$\frac{1}{m}\sum_{i=1}^{m} J_{1i}^2 \le \sigma^2 + \max_{1\le m\le n}\frac{1}{m}\sum_{i=1}^{m}(J_{1i}^2 - \sigma^2) = o_P(\log\log n). \tag{S6.4}$$

Next we will consider the order of cross-terms in (S6.1). Since J_{1i} controls the convergence rate of the other terms, we only need to consider the orders of $J_{1i}J_{ki}, k = 2, 3, 4, 5$. By the Cauchy-Schwarz inequality, (S6.3) and (S6.4), we have

$$\left|\frac{1}{\sqrt{m}}\sum_{i=n_0}^{m} J_{1i}J_{3i}\right| \le \left\{\frac{1}{m}\sum_{i=n_0}^{m} J_{1i}^2\right\}^{1/2}\left\{\sum_{i=n_0}^{m} J_{3i}^2\right\}^{1/2} = O_P\{(\log\log n)^{-3/2}\}. \tag{S6.5}$$

Similarly, we have

$$\left|\frac{1}{\sqrt{m}}\sum_{i=n_0}^{m} J_{1i}J_{2i}\right| \le \left\{\frac{1}{m}\sum_{i=n_0}^{m} J_{1i}^2\right\}^{1/2}\left\{\sum_{i=n_0}^{m} J_{2i}^2\right\}^{1/2} = O_P\{(\log\log n/n)^{1/2}\},$$

$$\left|\frac{1}{\sqrt{m}}\sum_{i=n_0}^{m} J_{1i}J_{4i}\right| \le \left\{\frac{1}{m}\sum_{i=n_0}^{m} J_{1i}^2\right\}^{1/2}\left\{\sum_{i=n_0}^{m} J_{4i}^2\right\}^{1/2} = O_P\{n^{-1/2}(\log\log n)^{-3/2}\}$$

and

$$\left|\frac{1}{\sqrt{m}}\sum_{i=n_0}^{m} J_{1i}J_{5i}\right| \le \left\{\frac{1}{m}\sum_{i=n_0}^{m} J_{1i}^2\right\}^{1/2}\left\{\sum_{i=n_0}^{m} J_{5i}^2\right\}^{1/2} = O_P\{(\log\log n)^{-3/2}\}.$$

Summarizing the above results, we have

$$\frac{1}{\sqrt{m}}\sum_{i=n_0}^{m} \hat{\varepsilon}_i^{*2} = \frac{1}{\sqrt{m}}\sum_{i=n_0}^{m} J_{1i}^2 + O_P\{(\log\log n)^{-3/2}\}. \tag{S6.6}$$

Let

$$T_n^* = \max_{1\le m\le n}\frac{1}{\sqrt{2m\sigma^4}}\sum_{i=1}^{m}(J_{1i}^2-\sigma^2).$$

Invoking Theorem 1 in Darling and Erdös (1956), we obtain that

$$P\Big(\sqrt{2\log\log n}T_n^* - \{2\log\log n + 0.5\log\log\log n - 0.5\log(4\pi)\} \le x\Big) \longrightarrow \exp(-\exp(-x)). \tag{S6.7}$$

By (S6.7), it is easy to check that

$$T_n^* = \{2\log\log n\}^{1/2}\{1+o_P(1)\}$$

and

$$T_{\log n}^* = \{2\log\log\log n\}^{1/2}\{1+o_P(1)\},$$

which implies that the maximum of T_n^* can not be achieved at $m < \log n$. By (3.5) in Section 3, we have

$$T_{AN}^* = \max_{1\le m\le n/(\log\log n)^4}\frac{1}{\sqrt{2m\sigma^4}}\sum_{i=1}^{m}(\hat{\varepsilon}_i^{*2}-\sigma^2) + o_P\{(\log\log n)^{-3/2}\}.$$

Again invoking the result of $|\hat{\sigma}^2-\sigma^2| = O(n^{-3+3\tau})$ in Tong and Wang (2005), and by (S6.6), it is easy to obtain that

$$T_{AN}^* = T_n^* + O_P\{(\log\log n)^{-3/2}\}.$$

Summarizing the above results, we complete the proof of Theorem 5. □

References

Bickel, P. L. and Rosenblatt, M. (1973). On some global measures of the derivations of density function estimates. *Ann. Statist.* **1**, 1071–1095.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of American Statistical Association* **92**, 477–489.

Chen, J., Gao, J. and Li, D. (2013). Estimation in a single-index panel data model with heterogeneous link functions. *Econometric Reviews* **33**, 928–955.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–1884.

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Darling, D. A. and Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Mathematical Journal* **23**, 143–155.

Liang, H., Liu, X., Li, R. and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38**, 3811–3836.

Martins-Filho, C. and Yao, F. (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics* **141**, 283–319.

Wang, J. L., Xue, L. G., Zhu, L. X. and Chong, Y. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38**, 246–274.

Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *Journal of the American Statistical Association* **94**, 1275–1285.

Xia, Y., Li, W. K., Tong, H. and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica* **14**, 1–39.

Zhang, W. Y., Fan, J. Q. and Sun, Y. (2009). A semiparametric model for cluster data. *Ann. Statist.* **37**, 2377–2408.

Zhu, L. X. and Xue, L. G. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Statist. Soc. B* **68**, 549–570.