SUPPLEMENTAL MATERIAL TO "MODEL SELECTION AND STRUCTURE SPECIFICATION IN ULTRA-HIGH DIMENSIONAL GENERALISED SEMI-VARYING COEFFICIENT MODELS"

By Degui Li , Yuan Ke and Wenyang Zhang*

University of York

In this supplemental material, we provide the detailed proofs of the main results stated in Section 3 of the main document as well as some technical lemmas which are useful in our proofs.

APPENDIX B: PROOFS OF THE MAIN RESULTS

In this appendix, we give the detailed proofs of the main theoretical results developed in Section 3.

Proof of Proposition 3.1 (i). Recall that

$$\widetilde{\mathbf{a}}_k = \left[\ \widetilde{a}_1(U_k), \cdots, \widetilde{a}_{d_n}(U_k)\ \right]^{\mathrm{T}}, \quad \widetilde{\mathbf{b}}_k = \left[\ \widetilde{\dot{a}}_1(U_k), \cdots, \widetilde{\dot{a}}_{d_n}(U_k)\ \right]^{\mathrm{T}}.$$

The basic idea used in the proof of this proposition is similar to that in Bickel *et al* (2009) and Lian (2012). However, as the kernel-based smoothing method is used, we need to derive the uniform convergence rates for the kernel-based quantities, which makes the technical argument more complicated than that in Bickel *et al* (2009) and Lian (2012).

We start with the proof that with probability approaching one, uniformly for $k = 1, \cdots, n$,

$$\max\Big\{ \sum_{j=s_{n2}+1}^{d_n} |d_{jk}|, \sum_{j=s_{n1}+1}^{d_n} |\dot{d}_{jk}| \Big\} \leq b\Big(\sum_{j=1}^{s_{n2}} |d_{jk}| + \sum_{j=1}^{s_{n1}} |\dot{d}_{jk}| \Big), \tag{B.1}$$

where $b = \max\{\lambda_1/\lambda_2, \lambda_2/\lambda_1\} + \delta$ for any small $\delta > 0$, where

$$d_{jk} = \widetilde{a}_j(U_k) - a_j(U_k) \quad \text{and} \quad \dot{d}_{jk} = h\big[\widetilde{\dot{a}}_j(U_k) - \dot{a}_j(U_k)\big]$$

for $j = 1, \cdots, d_n$ and $k = 1, \cdots, n$.

*Correspondent author. Email: wenyang.zhang@york.ac.uk

By the definitions of $\widetilde{\mathbf{a}}_k$ and $\widetilde{\mathbf{b}}_k$, we readily have

$$\mathcal{Q}_{nk}(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k) \geq \mathcal{Q}_{nk}\big(\mathbf{a}_{k0}, \mathbf{b}_{k0}\big), \tag{B.2}$$

where $\mathbf{a}_{k0}$ and $\mathbf{b}_{k0}$ are defined in Section 2 of the main document. From (B.2), we have

$$\begin{aligned} & \mathcal{L}_{nk}(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k) - \mathcal{L}_{nk}\big(\mathbf{a}_{k0}, \mathbf{b}_{k0}\big) \\ \geq\ & \lambda_1\Big[\sum_{j=1}^{d_n}|\widetilde{a}_j(U_k)| - \sum_{j=1}^{d_n}|a_j(U_k)|\Big] + \lambda_2\Big[\sum_{j=1}^{d_n}|\dot{\widetilde{a}}_j(U_k)| - \sum_{j=1}^{d_n}|\dot{a}_j(U_k)|\Big]. \end{aligned} \tag{B.3}$$

By the concavity condition of $\ell(\cdot,\cdot)$ (c.f., Assumption 2(ii)), we may show that

$$\mathcal{L}_{nk}(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k) - \mathcal{L}_{nk}\big(\mathbf{a}_{k0}, \mathbf{b}_{k0}\big) \leq \mathbf{d}_k^{\mathrm{T}}\dot{\mathcal{L}}_{nk}, \tag{B.4}$$

where

$$\begin{aligned} \dot{\mathcal{L}}_{nk} =& \frac{1}{n}\sum_{i=1}^{n} q_1\big[\sum_{j=1}^{d_n} a_j(U_k) + \dot{a}_j(U_k)(U_i - U_k)x_{ij}, y_i\big]\left(\begin{array}{c} X_i \\ \frac{U_i-U_k}{h}\cdot X_i \end{array}\right)\cdot \\ & K_h(U_i - U_k) \end{aligned}$$

and $\mathbf{d}_k = (d_{1k}, \cdots, d_{d_nk}, \dot{d}_{1k}, \cdots, \dot{d}_{d_nk})^{\mathrm{T}}$. By Lemma C.1 which is given in Appendix C, we may show that

$$\begin{aligned} & \max_{1\leq j\leq d_n}\sup_{1\leq k\leq n}\left|\frac{1}{n}\sum_{i=1}^{n} q_1\big[\sum_{j_1=1}^{d_n} a_{j_1}(U_i)x_{ij_1}, y_i\big]x_{ij}K_h(U_i-U_k)\right| \\ =& O_P\left(\sqrt{\frac{\log h^{-1}}{nh}}\right) \end{aligned} \tag{B.5}$$

and

$$\begin{aligned} & \max_{1\leq j\leq d_n}\sup_{1\leq k\leq n}\left|\frac{1}{n}\sum_{i=1}^{n} q_1\big[\sum_{j_1=1}^{d_n} a_{j_1}(U_i)x_{ij_1}, y_i\big]x_{ij}\big(\frac{U_i-U_k}{h}\big)K_h(U_i-U_k)\right| \\ =& O_P\left(\sqrt{\frac{\log h^{-1}}{nh}}\right). \end{aligned} \tag{B.6}$$

Then, by (B.5), (B.6), the standard calculation in kernel-based smoothing and the argument in the proof of Lemma C.1, we may show that

$$\mathbf{d}_k^{\mathrm{T}}\dot{\mathcal{L}}_{nk} \le O_P\Big(\sqrt{\frac{\log h^{-1}}{nh}} + s_{n2}h^2\Big)\cdot\big(\sum_{j=1}^{d_n}|d_{jk}| + \sum_{j=1}^{d_n}|\dot{d}_{jk}|\big) \tag{B.7}$$

uniformly for $k = 1, \cdots, n$.

On the other hand, by the triangle inequality, we may prove that

$$\begin{aligned} &\lambda_1\Big[\sum_{j=1}^{d_n}|\widetilde{a}_j(U_k)| - \sum_{j=1}^{d_n}|a_j(U_k)|\Big] \\ = \;& \lambda_1\sum_{j=1}^{s_{n2}}\big(|\widetilde{a}_j(U_k)| - |a_j(U_k)|\big) + \lambda_1\sum_{j=s_{n2}+1}^{d_n}|\widetilde{a}_j(U_k)| \\ \ge \;& -\lambda_1\sum_{j=1}^{s_{n2}}|d_{jk}| + \lambda_1\sum_{j=s_{n2}+1}^{d_n}|d_{jk}|. \end{aligned} \tag{B.8}$$

Similarly, we also have

$$\lambda_2\Big[\sum_{j=1}^{d_n}|\widetilde{\dot{a}}_j(U_k)| - \sum_{j=1}^{d_n}|\dot{a}_j(U_k)|\Big] \ge -\lambda_2\sum_{j=1}^{s_{n1}}|\dot{d}_{jk}| + \lambda_2\sum_{j=s_{n1}+1}^{d_n}|\dot{d}_{jk}|. \tag{B.9}$$

By (B.3), (B.4), (B.7)–(B.9) and the condition that $\sqrt{\frac{\log h^{-1}}{nh}} + s_{n2}h^2 = o(\lambda_1+\lambda_2)$ and $\lambda_1 \propto \lambda_2$ (c.f., Assumption 5), we can complete the proof of (B.1).

Let

$$\mathbf{u}_1 = \big(u_{11}, \cdots, u_{1d_n}\big)^{\mathrm{T}} \quad\text{and}\quad \mathbf{u}_2 = \big(u_{21}, \cdots, u_{2d_n}\big)^{\mathrm{T}}$$

be two d_n-dimensional column vectors and define

$$\begin{aligned} \mathbf{\Omega}(C_0) \;=\; & \Big\{\big(\mathbf{u}_1^{\mathrm{T}}, \mathbf{u}_2^{\mathrm{T}}\big)^{\mathrm{T}} :\; \|\mathbf{u}_1\|^2 = \|\mathbf{u}_2\|^2 = C_0, \\ & \sum_{j=1}^{d_n}\big(|u_{1j}| + |u_{2j}|\big) \le 2(1+b)\sum_{j=1}^{s_{n2}}\big(|u_{1j}| + |u_{2j}|\big)\Big\}, \end{aligned}$$

where C_0 is a positive constant which could be sufficiently large. By the concavity of $\ell(\cdot,\cdot)$, we only need to prove that there exists a local maximiser $(\widetilde{\mathbf{a}}_k, h\widetilde{\mathbf{b}}_k)$ in the interior of $\big\{(\mathbf{a}_{k0}+\gamma_n\mathbf{u}_1, h\mathbf{b}_{k0}+\gamma_n\mathbf{u}_2) :\; (\mathbf{u}_1^{\mathrm{T}},\; \mathbf{u}_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbf{\Omega}(C_0)\big\}$, where $\gamma_n = \sqrt{s_{n2}}\lambda_1$.

Observe that

$$\mathcal{Q}_{nk}\big[\mathbf{a}_{k0}+\gamma_n\mathbf{u}_1,\mathbf{b}_{k0}+\gamma_n\mathbf{u}_2/h\big]-\mathcal{Q}_{nk}\big(\mathbf{a}_{k0},\mathbf{b}_{k0}\big)=\sum_{l=1}^{3}\mathcal{I}_{nk}(l), \tag{B.10}$$

where

$$\begin{aligned}
\mathcal{I}_{nk}(1) &= \mathcal{L}_{nk}\big(\mathbf{a}_{k0}+\gamma_n\mathbf{u}_1,\mathbf{b}_{k0}+\gamma_n\mathbf{u}_2/h\big)-\mathcal{L}_{nk}\big(\mathbf{a}_{k0},\mathbf{b}_{k0}\big),\\
\mathcal{I}_{nk}(2) &= -\lambda_1\Big(\sum_{j=1}^{d_n}|a_j(U_k)+\gamma_n u_{1j}|-\sum_{j=1}^{d_n}|a_j(U_k)|\Big),\\
\mathcal{I}_{nk}(3) &= -\lambda_2\Big(\sum_{j=1}^{d_n}|h\dot{a}_j(U_k)+\gamma_n u_{2j}|-\sum_{j=1}^{d_n}|h\dot{a}_j(U_k)|\Big).
\end{aligned}$$

We first consider $\mathcal{I}_{nk}(1)$. Letting $\mathbf{u}=(\mathbf{u}_1^{\mathrm{T}},\mathbf{u}_2^{\mathrm{T}})^{\mathrm{T}}$ and by the definition of $\mathcal{L}_{nk}(\cdot,\cdot)$ in Section 2, we have

$$\mathcal{I}_{nk}(1)\overset{P}{\sim}\gamma_n\mathbf{u}^{\mathrm{T}}\dot{\mathcal{L}}_{nk}+\frac{1}{2}\gamma_n^2\mathbf{u}^{\mathrm{T}}\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k^*,\mathbf{b}_k^*)\mathbf{u}, \tag{B.11}$$

where $a_n\overset{P}{\sim}b_n$ denotes that $a_n=b_n(1+o_P(1))$, $(\mathbf{a}_k^*,\mathbf{b}_k^*)$ lies between $\big(\mathbf{a}_{k0}+\gamma_n\mathbf{u}_1,\mathbf{b}_{k0}+\gamma_n\mathbf{u}_2/h\big)$ and $(\mathbf{a}_{k0},\mathbf{b}_{k0})$,

$$\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k,\mathbf{b}_k)=\begin{bmatrix}\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k,\mathbf{b}_k,0) & \ddot{\mathcal{L}}_{nk}(\mathbf{a}_k,\mathbf{b}_k,1)\\ \ddot{\mathcal{L}}_{nk}(\mathbf{a}_k,\mathbf{b}_k,1) & \ddot{\mathcal{L}}_{nk}(\mathbf{a}_k,\mathbf{b}_k,2)\end{bmatrix}$$

with

$$\begin{aligned}
\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k,\mathbf{b}_k,l) &= \frac{1}{n}\sum_{i=1}^{n}q_2\left\{\sum_{j=1}^{d_n}\big[\alpha_{jk}+\beta_{jk}(U_i-U_k)\big]x_{ij},\ y_i\right\}\left(\frac{U_i-U_k}{h}\right)^l\cdot\\
&\quad X_iX_i^{\mathrm{T}}K_h(U_i-U_k)
\end{aligned}$$

for $l=0,1,2$, where $\mathbf{a}_k=\big(\alpha_{1k},\cdots,\alpha_{d_nk}\big)^{\mathrm{T}}$ and $\mathbf{b}_k=\big(\beta_{1k},\cdots,\beta_{d_nk}\big)^{\mathrm{T}}$.

Note that for $\mathbf{u}\in\mathbf{\Omega}(C_0)$,

$$\sum_{j=1}^{d_n}(|u_{1j}|+|u_{2j}|)\le 2(1+b)\sum_{j=1}^{s_{n2}}(|u_{1j}|+|u_{2j}|). \tag{B.12}$$

Using Lemma C.1 in Appendix C, the Cauchy-Schwarz inequality and (B.12), we can show that uniformly for $k=1,\cdots,n$,

$$\gamma_n\mathbf{u}^{\mathrm{T}}\dot{\mathcal{L}}_{nk}=o_P(\gamma_n^2)\cdot\|\mathbf{u}\|. \tag{B.13}$$

On the other hand, note that

$$\begin{aligned} \text{(B.14)} \qquad & \frac{1}{2}\gamma_n^2 \mathbf{u}^{\mathrm{T}} \ddot{\mathcal{L}}_{nk}(\mathbf{a}_k^*, \mathbf{b}_k^*)\mathbf{u} \\ = \ & \frac{1}{2}\gamma_n^2 \mathbf{u}^{\mathrm{T}} \big[\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k^*, \mathbf{b}_k^*) - \ddot{\mathcal{L}}_n(U_k)\big]\mathbf{u} + \frac{1}{2}\gamma_n^2 \mathbf{u}^{\mathrm{T}} \ddot{\mathcal{L}}_n(U_k)\mathbf{u}, \end{aligned}$$

where $\ddot{\mathcal{L}}_n(U_k)$ is defined at the beginning of Appendix A. By Assumption 2(iii), we readily have

$$\text{(B.15)} \qquad \frac{1}{2}\gamma_n^2 \mathbf{u}^{\mathrm{T}} \ddot{\mathcal{L}}_n(U_k)\mathbf{u} \le -\frac{1}{2}\rho_1 \gamma_n^2 \|\mathbf{u}\|^2 < 0$$

uniformly for $k = 1, \cdots, n$. By (B.12), Assumptions 2(ii), the condition $s_{n2}^2 \lambda_1 = o(1)$ in Assumption 5 and following the proof of Lemma C.1, we may prove that uniformly for $k = 1, \cdots, n$,

$$\begin{aligned} \text{(B.16)} \qquad & \gamma_n^2 \mathbf{u}^{\mathrm{T}} \big[\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k^*, \mathbf{b}_k^*) - \ddot{\mathcal{L}}_n(U_k)\big]\mathbf{u} \\ = \ & O_P\Big(\gamma_n^3 \Big[\sum_{j=1}^{d_n} (|u_{1j}| + |u_{2j}|)\Big]^3\Big) = O_P\Big(\gamma_n^3 \Big[\sum_{j=1}^{s_{n2}} (|u_{1j}| + |u_{2j}|)\Big]^3\Big) \\ = \ & O_P\big(\gamma_n^3 s_{n2}^{3/2} \|\mathbf{u}\|^3\big) = o_P(\gamma_n^2) \cdot \big(\|\mathbf{u}\|^2\big). \end{aligned}$$

Hence, by (B.11) and (B.13)–(B.16), when n is sufficiently large and C_0 is large enough, we have

$$\text{(B.17)} \qquad \mathcal{I}_{nk}(1) \overset{P}{\sim} \frac{1}{2}\gamma_n^2 \mathbf{u}^{\mathrm{T}} \ddot{\mathcal{L}}_{nk}(U_k)\mathbf{u}.$$

We next consider $\mathcal{I}_{nk}(2)$ and $\mathcal{I}_{nk}(3)$. It is easy to show that

$$\begin{aligned} \text{(B.18)}\ \mathcal{I}_{nk}(2) \ = \ & -\lambda_1 \Big[\sum_{j=1}^{d_n} |a_j(U_k) + \gamma_n u_{1j}| - \sum_{j=1}^{d_n} |a_j(U_k)|\Big] \\ \le \ & \lambda_1 \sum_{j=1}^{s_{n2}} \big[|a_j(U_k)| - |a_j(U_k) + \gamma_n u_{1j}|\big] - \lambda_1 \sum_{j=s_{n2}+1}^{d_n} |\gamma_n u_{1j}| \\ = \ & O_P(\gamma_n^2) \cdot \|\mathbf{u}_1\| - \lambda_1 \sum_{j=s_{n2}+1}^{d_n} |\gamma_n u_{1j}|. \end{aligned}$$

Similarly, noting that $\lambda_1 \propto \lambda_2$ we also have

$$\text{(B.19)} \qquad \mathcal{I}_{nk}(3) = O_P(\gamma_n^2) \cdot \|\mathbf{u}_2\| - \lambda_2 \sum_{j=s_{n1}+1}^{d_n} |\gamma_n u_{2j}|.$$

Hence, by (B.10) and (B.17)–(B.19), we can prove that the leading term of $\mathcal{I}_{nk}(1)+\mathcal{I}_{nk}(2)+\mathcal{I}_{nk}(3)$ is negative in probability uniformly in k by choosing sufficiently large C_0. Hence, we may find a local maximiser $(\widetilde{\mathbf{a}}_k, h\widetilde{\mathbf{b}}_k)$ in the interior of $\big\{(\mathbf{a}_{k0}+\gamma_n\mathbf{u}_1, h\mathbf{b}_{k0}+\gamma_n\mathbf{u}_2):\ (\mathbf{u}_1^{\mathrm{T}},\ \mathbf{u}_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbf{\Omega}(C_0)\big\}$, which completes the proof of Proposition 3.1(i). $\square$

Proof of Proposition 3.1 (ii). The proof is similar to that in the proof of Proposition 3.1(i) with the role of Lemma C.1 replaced by Lemma C.2 (given in Appendix C). $\square$

Proof of Theorem 3.1. We start with the proof of the convergence rates for the biased oracle estimators $\overline{\mathcal{A}}_n^{bo}$ and $\overline{\mathcal{B}}_n^{bo}$. According to the definition, we have

$$(\overline{\mathcal{A}}_n^{bo}, \overline{\mathcal{B}}_n^{bo}) = \arg\max \mathcal{Q}_n^2(\mathcal{A}^o,\ \mathcal{B}^o), \tag{B.20}$$

where $\mathcal{A}^o$ and $\mathcal{B}^o$ are defined as in Section 3. Recall that $\mathcal{A}_0$ and $\mathcal{B}_0$ are the vectors of the true functional coefficients and their derivative functions, and denote

$$\mathcal{U}_1 = \big[\mathbf{u}_1^{\mathrm{T}}(1), \cdots, \mathbf{u}_1^{\mathrm{T}}(n)\big]^{\mathrm{T}}, \quad \mathcal{U}_2 = \big[\mathbf{u}_2^{\mathrm{T}}(1), \cdots, \mathbf{u}_2^{\mathrm{T}}(n)\big]^{\mathrm{T}},$$

where both $\mathbf{u}_1(k)$ and $\mathbf{u}_2(k)$ are d_n-dimensional column vectors, $k = 1, \cdots, n$, the last $d_n - s_{n2}$ elements of $\mathbf{u}_1(k)$ and the last $d_n - s_{n1}$ elements of $\mathbf{u}_2(k)$ are zeroes. Define

$$\mathbf{\Omega}_n^*(C_*) = \big\{(\mathcal{U}_1^{\mathrm{T}},\ \mathcal{U}_2^{\mathrm{T}})^{\mathrm{T}}:\ \|\mathcal{U}_1\|^2 = \|\mathcal{U}_2\|^2 = nC_*\big\},$$

where C_* is a positive constant which can be sufficiently large.

For $(\mathcal{U}_1^{\mathrm{T}},\ \mathcal{U}_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbf{\Omega}_n^*(C_*)$, observe that

$$\mathcal{Q}_n^2\big(\mathcal{A}_0+\gamma_n^*\mathcal{U}_1, \mathcal{B}_0+\gamma_n^*\mathcal{U}_2/h\big) - \mathcal{Q}_n^2(\mathcal{A}_0, \mathcal{B}_0) = \mathcal{I}_n(1)+\mathcal{I}_n(2)+\mathcal{I}_n(3), \tag{B.21}$$

where $\gamma_n^* = \sqrt{s_{n2}/nh}$,

$$\begin{aligned}
\mathcal{I}_n(1) &= \mathcal{L}_n^{\diamond}\big(\mathcal{A}_0 + \gamma_n^*\mathcal{U}_1, \mathcal{B}_0 + \gamma_n^*\mathcal{U}_2/h\big) - \mathcal{L}_n^{\diamond}(\mathcal{A}_0, \mathcal{B}_0),\\
\mathcal{I}_n(2) &= \sum_{j=1}^{d_n} \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|)\|\boldsymbol{\alpha}_{j0}\| - \sum_{j=1}^{d_n} \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|)\|\boldsymbol{\alpha}_{j0} + \gamma_n^*\mathbf{u}_{1j}\|,\\
\mathcal{I}_n(3) &= \sum_{j=1}^{d_n} \dot{p}_{\lambda_4^*}(\widetilde{D}_j)\|h\boldsymbol{\beta}_{j0}\| - \sum_{j=1}^{d_n} \dot{p}_{\lambda_4^*}(\widetilde{D}_j)\|h\boldsymbol{\beta}_{j0} + \gamma_n^*\mathbf{u}_{2j}\|,
\end{aligned}$$

in which $\boldsymbol{\alpha}_{j0} = \left[a_j(U_1), \cdots, a_j(U_n)\right]^{\mathrm{T}}$, $\boldsymbol{\beta}_{j0} = \left[\dot{a}_j(U_1), \cdots, \dot{a}_j(U_n)\right]^{\mathrm{T}}$, $\mathbf{u}_{1j} = \left[u_{1j}(1), \cdots, u_{1j}(n)\right]^{\mathrm{T}}$, $\mathbf{u}_{2j} = \left[u_{2j}(1), \cdots, u_{2j}(n)\right]^{\mathrm{T}}$, $u_{1j}(k)$ and $u_{2j}(k)$ are the j-th component of vectors $\mathbf{u}_1(k)$ and $\mathbf{u}_2(k)$, respectively.

For $\mathcal{I}_n(1)$, by the definition of $\mathcal{L}_n^\diamond(\cdot,\cdot)$ in Section 2, we have

$$\mathcal{I}_n(1) = \mathcal{I}_n(4) + \mathcal{I}_n(5) + o_P\big((\gamma_n^*)^2\big) \cdot \big(\|\mathcal{U}_1\|^2 + \|\mathcal{U}_2\|^2\big), \tag{B.22}$$

where

$$\begin{aligned}
\mathcal{I}_n(4) &= \gamma_n^* \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\dot{\mathcal{L}}_n(\mathcal{A}_0,\mathcal{B}_0),\\
\mathcal{I}_n(5) &= \frac{1}{2}(\gamma_n^*)^2 \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\mathcal{V}_n(\mathcal{U}_1,\mathcal{U}_2).
\end{aligned}$$

The detailed proof of (B.22) will be provided in Appendix C below. By some elementary but tedious calculations, we can show that

$$\mathcal{I}_n(4) = O_P\big((\gamma_n^*)^2 n^{1/2}\big) \cdot \big(\|\mathcal{U}\| + \|\mathcal{V}\|\big). \tag{B.23}$$

The detailed proof of (B.23) will be also given in Appendix C below. For $\mathcal{I}_n(5)$, note that

$$\begin{aligned}
\mathcal{I}_n(5) &= \frac{1}{2}(\gamma_n^*)^2 \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\Big[\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n) - \ddot{\mathcal{L}}_n(\mathcal{A}_0,\mathcal{B}_0)\Big]\mathcal{V}_n(\mathcal{U}_1,\mathcal{U}_2) + \\
&\quad \frac{1}{2}(\gamma_n^*)^2 \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\ddot{\mathcal{L}}_n(\mathcal{A}_0,\mathcal{B}_0)\mathcal{V}_n(\mathcal{U}_1,\mathcal{U}_2)\\
&\equiv \mathcal{I}_n(6) + \mathcal{I}_n(7).
\end{aligned} \tag{B.24}$$

By Assumption 2(iii) and the definitions of $\mathcal{U}_1$ and $\mathcal{U}_2$, we may show that

$$\mathcal{I}_n(7) \le -\frac{1}{2}\rho_1(\gamma_n^*)^2\big(\|\mathcal{U}_1\|^2 + \|\mathcal{U}_2\|^2\big) < 0. \tag{B.25}$$

By Assumption 2(ii) and using Proposition 3.1, we can prove that

$$\mathcal{I}_n(6) = o_P\big((\gamma_n^*)^2\big) \cdot \big(\|\mathcal{U}_1\|^2 + \|\mathcal{U}_2\|^2\big), \tag{B.26}$$

which, together with (B.22)–(B.25), implies that $\mathcal{I}_n(7)$ is the leading term of $\mathcal{I}_n(1)$. Hence, when n is sufficiently large, by taking C_* large enough, we have

$$\mathcal{I}_n(1) \overset{P}{\sim} \frac{1}{2}(\gamma_n^*)^2 \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\ddot{\mathcal{L}}_n(\mathcal{A}_0,\mathcal{B}_0)\mathcal{V}_n(\mathcal{U}_1,\mathcal{U}_2). \tag{B.27}$$

We next consider $\mathcal{I}_n(2)$. Noting that $\mathbf{u}_{1j} = \mathbf{0}$ for $j = s_{n2+1}, \cdots, d_n$, we have

$$\begin{aligned}\mathcal{I}_n(2) &= \sum_{j=1}^{d_n} \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|)\|\boldsymbol{\alpha}_{j0}\| - \sum_{j=1}^{d_n} \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|)\|\boldsymbol{\alpha}_{j0} + \gamma_n^* \mathbf{u}_{1j}\| \\ &= \sum_{j=1}^{s_{n2}} \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|)\big(\|\boldsymbol{\alpha}_{j0}\| - \|\boldsymbol{\alpha}_{j0} + \gamma_n^* \mathbf{u}_{1j}\|\big).\end{aligned}$$

By Proposition 3.1 and (A.4) in Assumption 6, we may show that with probability approaching one,

$$\min_{1\leq j\leq s_{n2}} \|\widetilde{\boldsymbol{\alpha}}_j\| > \frac{1}{2} b_\diamond n^{1/2},$$

which together with the condition of $\lambda_4 = o(n^{1/2})$ and the SCAD structure, implies that

$$\mathcal{I}_n(2) = o_P\big((\gamma_n^*)^2\big) \cdot \|\mathcal{U}_1\|^2. \tag{B.28}$$

Similarly, we may also show that

$$\mathcal{I}_n(3) = o_P\big((\gamma_n^*)^2\big) \cdot \|\mathcal{U}_2\|^2, \tag{B.29}$$

by noting that

$$\min_{1\leq j\leq s_{n1}} \widetilde{D}_j > \frac{1}{2} b_\diamond n^{1/2}.$$

Hence, by (B.21) and (B.27)–(B.29), we can prove that the leading term of $\mathcal{I}_n(1) + \mathcal{I}_n(2) + \mathcal{I}_n(3)$ is negative in probability, which indicates that for any $\epsilon > 0$, there exists a sufficiently large $C_* > 0$ such that

$$\mathsf{P}\Big\{ \sup_{(\mathcal{U}_1,\mathcal{U}_2)\in\boldsymbol{\Omega}_n^*(C_*)} \mathcal{Q}_n^2\big(\mathcal{A}_0+\gamma_n^*\mathcal{U}_1, \mathcal{B}_0+\gamma_n^*\mathcal{U}_2/h\big) < \mathcal{Q}_n^2(\mathcal{A}_0, \mathcal{B}_0)\Big\} \geq 1-\epsilon \tag{B.30}$$

for large n. Therefore, we may show that

$$\frac{1}{n}\big\|\overline{\mathcal{A}}_n^{bo} - \mathcal{A}_0\big\|^2 = \frac{s_{n2}}{nh}, \quad \frac{1}{n}\big\|\overline{\mathcal{B}}_n^{bo} - \mathcal{B}_0\big\|^2 = \frac{s_{n2}}{nh^3}, \tag{B.31}$$

which is (3.2) in Theorem 3.1.

We next complete the proof of Theorem 3.1. Define

$$\mathcal{M}_{\boldsymbol{\alpha}} = \big(\boldsymbol{\alpha}_j : \ 1 \leq j \leq s_{n2}\big) \ \text{ and } \ \mathcal{M}_{\boldsymbol{\beta}} = \big(h\boldsymbol{\beta}_j : \ 1 \leq j \leq s_{n1}\big), \tag{B.32}$$

which correspond the non-zero components in $\mathcal{A}_0$ and $\mathcal{B}_0$, respectively. Let $\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\mathcal{M}_{\boldsymbol{\alpha}})$, $\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\mathcal{M}_{\boldsymbol{\beta}})$, $\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\boldsymbol{\alpha}_j)$ and $\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|h\boldsymbol{\beta}_j)$ be the gradient vectors of $\mathcal{L}_n^\diamond(\mathcal{A},\mathcal{B})$ with respect to $\mathcal{M}_{\boldsymbol{\alpha}}$, $\mathcal{M}_{\boldsymbol{\beta}}$, $\boldsymbol{\alpha}_j$ and $h\boldsymbol{\beta}_j$, respectively. Define the sub-gradient of the penalty terms as

$$\begin{aligned}
\mathcal{P}(\mathcal{M}_{\boldsymbol{\alpha}}) &= \Big[\dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_1\|)\frac{\alpha_{11}}{\|\boldsymbol{\alpha}_1\|},\cdots,\dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_{s_{n2}}\|)\frac{\alpha_{s_{n2}1}}{\|\boldsymbol{\alpha}_{s_{n2}}\|},\cdots,\\
&\qquad \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_1\|)\frac{\alpha_{1n}}{\|\boldsymbol{\alpha}_1\|},\cdots,\dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_{s_{n2}}\|)\frac{\alpha_{s_{n2}n}}{\|\boldsymbol{\alpha}_{s_{n2}}\|}\Big]^{\mathrm{T}},\\
\mathcal{P}(\mathcal{M}_{\boldsymbol{\beta}}) &= \Big[\dot{p}_{\lambda_4}(\widetilde{D}_1)\frac{\beta_{11}}{\|\boldsymbol{\beta}_1\|},\cdots,\dot{p}_{\lambda_4}(\widetilde{D}_{s_{n1}})\frac{\beta_{s_{n1}1}}{\|\boldsymbol{\beta}_{s_{n1}}\|},\cdots,\\
&\qquad \dot{p}_{\lambda_4}(\widetilde{D}_1)\frac{\beta_{1n}}{\|\boldsymbol{\beta}_1\|},\cdots,\dot{p}_{\lambda_4}(\widetilde{D}_{s_{n1}})\frac{\beta_{s_{n1}n}}{\|\boldsymbol{\beta}_{s_{n1}}\|}\Big]^{\mathrm{T}}.
\end{aligned}$$

Following the proof of Theorem 1 in Fan *et al* (2014) (see also the proof of Theorem 1 in Fan and Lv, 2011), the objective function $\mathcal{Q}_n^2(\mathcal{A},\mathcal{B})$ has a unique maximiser $\big(\overline{\mathcal{A}}_n^{bo},\overline{\mathcal{B}}_n^{bo}\big)$ if

$$\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\mathcal{M}_{\boldsymbol{\alpha}})-\mathcal{P}(\mathcal{M}_{\boldsymbol{\alpha}})=\mathbf{0}_{ns_{n2}}, \tag{B.33}$$

$$\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\mathcal{M}_{\boldsymbol{\beta}})-\mathcal{P}(\mathcal{M}_{\boldsymbol{\beta}})=\mathbf{0}_{ns_{n1}}, \tag{B.34}$$

$$\max_{s_{n2}+1\le j\le d_n}\|\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\boldsymbol{\alpha}_j)\| < \min_{s_{n2}+1\le j\le d_n}\dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|), \tag{B.35}$$

$$\max_{s_{n1}+1\le j\le d_n}\|\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|h\boldsymbol{\beta}_j)\| < \min_{s_{n1}+1\le j\le d_n}\dot{p}_{\lambda_4^*}(\widetilde{D}_j) \tag{B.36}$$

hold at $\mathcal{A}=\overline{\mathcal{A}}_n^{bo}$ and $\mathcal{B}=\overline{\mathcal{B}}_n^{bo}$. Hence, we next only need to prove (B.33)–(B.36).

By the definition of the biased oracle estimators $\overline{\mathcal{A}}_n^{bo}$ and $\overline{\mathcal{B}}_n^{bo}$, it is easy to verify (B.33) and (B.34). We next only show the proof of (B.35) as the proof of (B.36) is analogous. By Proposition 3.1 and the condition of $(ns_{n2})^{1/2}\lambda_1 = o(\lambda_4)$, we may show that

$$\min_{s_{n2}+1\le j\le d_n}\dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_j\|)=\lambda_4 \tag{B.37}$$

with probability approaching one. On the other hand, for the left hand side of (B.35), we can prove that

(B.38)

$$\max_{s_{n2}+1\le j\le d_n}\|\dot{\mathcal{L}}_n^\diamond(\mathcal{A},\mathcal{B}|\boldsymbol{\alpha}_j)\| = O_P\Big(h^{-1/2}\big[(\log h^{-1})^{1/2}+s_{n2}^{1/2}+(nh)^{1/2}s_{n2}^2\lambda_1^2\big]\Big)$$

when $\mathcal{A} = \overline{\mathcal{A}}_n^{bo}$ and $\mathcal{B} = \overline{\mathcal{B}}_n^{bo}$. The detailed proof of (B.38) will be given in Appendix C below. Using (B.37), (B.38) and (A.3) in Assumption 6, we may prove (B.35). Then, the proof of Theorem 3.1 is completed. $\square$

Proof of Theorem 3.2. The proof is similar to the proof of Theorem 2 in Wang and Xia (2009) with some modifications. Recall that $\overline{a}_j^{bo}(U_k)$, $j = 1, \cdots, s_{n2}$, $k = 1, \cdots, n$, are the biased oracle estimators of $a_j(U_k)$ which are obtained by maximising the objective function $\mathcal{Q}_n^2(\mathcal{A}^o,\ \mathcal{B}^o)$.

Let

$$\overline{\mathbf{D}}_n^o = \Big(\max_{1\le k\le n} \left|\overline{a}_1^{bo}(U_k) - a_1^{uo}(U_k)\right|,\ \cdots,\ \max_{1\le k\le n} \left|\overline{a}_{s_{n1}}^{bo}(U_k) - a_{s_{n1}}^{uo}(U_k)\right|\Big)^{\mathrm{T}},$$

and

$$\overline{\mathbf{C}}_n^{bo} = \left(\overline{c}_{s_{n1}+1}^{bo}, \cdots, \overline{c}_{s_{n2}}^{bo}\right)^{\mathrm{T}}, \text{ where } \overline{c}_j^{bo} = \frac{1}{n}\sum_{k=1}^{n} \overline{a}_j^{bo}(U_k), \quad j = s_{n1}+1, \cdots, s_{n2}.$$

By Theorem 3.1, in order to prove (3.3) and (3.4), we only need to show that

$$\sqrt{nh}\mathbf{B}_n^{\mathrm{T}}\overline{\mathbf{D}}_n^o = o_P(1), \quad \sqrt{n}\mathbf{A}_n^{\mathrm{T}}\left(\overline{\mathbf{C}}_n^{bo} - \mathbf{C}_n^{uo}\right) = o_P(1). \tag{B.39}$$

For $k = 1, \cdots, n$, denote

$$\begin{aligned} \mathbf{a}^{uo}(U_k) &= \left[a_1^{uo}(U_k), \cdots, a_{s_{n2}}^{uo}(U_k), 0, \cdots, 0\right]^{\mathrm{T}}, \\ \overline{\mathbf{a}}^{bo}(U_k) &= \left[\overline{a}_1^{bo}(U_k), \cdots, \overline{a}_{s_{n2}}^{bo}(U_k), 0, \cdots, 0\right]^{\mathrm{T}}, \end{aligned}$$

where the last $d_n - s_{n2}$ elements in the above two vectors are zeros, and let $\mathbf{b}^{uo}(U_k)$ and $\overline{\mathbf{b}}^{bo}(U_k)$ be defined analogously. Then, using the first-order condition, we may show that the unbiased oracle estimates satisfy the following equation:

$$\mathbf{0}_{s_{n2}} = \mathcal{R}_{s_{n2}}\dot{\mathcal{L}}_{nk}\big(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k\big) + \mathcal{R}_{s_{n2}}\ddot{\mathcal{L}}_{nk}\big(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k\big) \begin{bmatrix} \mathbf{a}^{uo}(U_k) - \widetilde{\mathbf{a}}_k \\ h\mathbf{b}^{uo}(U_k) - h\widetilde{\mathbf{b}}_k \end{bmatrix} \tag{B.40}$$

uniformly for $1 \le k \le n$, where $\mathcal{R}_{s_{n2}} = \left[I_{s_{n2}},\ N_{s_{n2}\times(2d_n - s_{n2})}\right]$ with I_s being an $s \times s$ identity matrix and $N_{r\times s}$ being a $r \times s$ null matrix.

Following the proof of Theorem 3.1, we can also show that the biased oracle estimates satisfy the following equation:
(B.41)

$$\mathbf{0}_{s_{n2}} = \mathcal{R}_{s_{n2}}\dot{\mathcal{L}}_{nk}\big(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k\big) + \mathcal{R}_{s_{n2}}\ddot{\mathcal{L}}_{nk}\big(\widetilde{\mathbf{a}}_k, \widetilde{\mathbf{b}}_k\big) \begin{bmatrix} \overline{\mathbf{a}}^{bo}(U_k) - \widetilde{\mathbf{a}}_k \\ h\overline{\mathbf{b}}^{bo}(U_k) - h\widetilde{\mathbf{b}}_k \end{bmatrix} - \mathcal{P}^*(U_k)$$

uniformly for $1 \le k \le n$, where

$$\mathcal{P}^*(U_k) = \Big[\dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_1\|)\frac{\overline{a}_1^{bo}(U_k)}{\|\overline{\boldsymbol{\alpha}}_1^{bo}\|}, \cdots, \dot{p}_{\lambda_4}(\|\widetilde{\boldsymbol{\alpha}}_{s_{n2}}\|)\frac{\overline{a}_{s_{n2}}^{bo}(U_k)}{\|\overline{\boldsymbol{\alpha}}_{s_{n2}}^{bo}\|}\Big]^{\mathrm{T}},$$

$\overline{\boldsymbol{\alpha}}_j^{bo} = \big[\overline{a}_j^{bo}(U_1), \cdots, \overline{a}_j^{bo}(U_n)\big]^{\mathrm{T}}$. By Proposition 3.1 and (A.4) in Assumption 6, we may show that

$$\min_{1\le j\le s_{n2}} \|\widetilde{\boldsymbol{\alpha}}_j\| \ge \min_{1\le j\le s_{n2}} \|\boldsymbol{\alpha}_{j0}\| - \max_{1\le j\le s_{n2}} \|\widetilde{\boldsymbol{\alpha}}_j - \boldsymbol{\alpha}_{j0}\| \ge \frac{1}{2} b_\diamond \sqrt{n}$$

with probability approaching one, which together with the SCAD structure, indicates that the penalty term $\mathcal{P}^*(U_k)$ in (B.41) can be asymptotically ignored. Then, by (B.40), (B.41) and the standard argument, we may complete the proof of (B.39). □

APPENDIX C: PROOFS OF SOME TECHNICAL LEMMAS

Define

$$Z_{ij}(u, l) = Q_{i1} x_{ij} \big(\frac{U_i - u}{h}\big)^l K_h(U_i - u), \quad u \in [0, 1] \tag{C.1}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, d_n$, $l = 0, 1, 2, \cdots$, where

$$Q_{i1} = q_1\big[\sum_{j_1=1}^{d_n} a_{j_1}(U_i) x_{ij_1}, y_i\big].$$

Under different moment conditions on the random element $Q_{i1}x_{ij}$, in Lemmas C.1 and C.2 below, we give the uniform consistency results of the nonparametric kernel-based estimators in the ultra-high dimensional case, which are of independent interest. Analogous uniform consistency results also hold when $Q_{i1}x_{ij}$ in (C.1) is replaced by $Q_{i2}x_{ij}x_{ik}$ or $M(X_i, U_i, y_i)x_{ij}x_{ik}x_{il}$, where Q_{i2} and $M(X_i, U_i, y_i)$ are defined in Appendix A of the main document.

Lemma C.1. *Suppose that Assumptions 1 and 3 in Appendix A are satisfied. Moreover, suppose that the dimension* $d_n \propto n^{\tau_1}$ *with* $0 \le \tau_1 < \infty$, $\mathsf{E}\big(Q_{i1}\big|X_i, U_i\big) = 0$ *a.s., the moment condition (A.1) in Appendix A holds for some* $m_0 > 2$*, and*

$$h \propto n^{-\delta_1} \text{ with } 0 < \delta_1 < 1, \quad \frac{nh}{(nd_n)^{2/m_0} \log h^{-1}} \to \infty. \tag{C.2}$$

Then we have, as $n \to \infty$,

$$\max_{1\le j\le d_n}\sup_{u\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n Z_{ij}(u,l)\Big| = O_P\Big(\big(\frac{\log h^{-1}}{nh}\big)^{1/2}\Big) \tag{C.3}$$

for any $l = 0, 1, 2, \cdots$.

Proof. For simplicity, let $\xi_n = \big(\frac{\log h^{-1}}{nh}\big)^{1/2}$. The main idea of proving (C.3) is to consider covering the interval $[0,\ 1]$ by a finite number of subsets $U(k)$ which are centered at u_k with radius $r_n = \xi_n h^2$. Letting $\mathcal{N}_n$ be the total number of such subsets $U(k)$, $\mathcal{N}_n = O(r_n^{-1})$. It is easy to show that

$$\begin{aligned} & \max_{1\le j\le d_n}\sup_{u\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n Z_{ij}(u,l)\Big| \\ \le\ & \max_{1\le j\le d_n}\max_{1\le k\le \mathcal{N}_n}\Big|\frac{1}{n}\sum_{i=1}^n Z_{ij}(u_k,l)\Big| + \\ & \max_{1\le j\le d_n}\max_{1\le k\le \mathcal{N}_n}\sup_{u\in U(k)}\Big|\frac{1}{n}\sum_{i=1}^n Z_{ij}(u,l) - \frac{1}{n}\sum_{i=1}^n Z_{ij}(u_k,l)\Big| \\ \equiv\ & \Pi_{n1} + \Pi_{n2}. \end{aligned} \tag{C.4}$$

By the continuity condition on $K(\cdot)$ in Assumption 1 and using the definition of r_n, we readily have

$$\Pi_{n2} = O_P\big(\frac{r_n}{h^2}\big) = O_P(\xi_n). \tag{C.5}$$

For Π_{n1}, we apply the truncation technique and the Bernstein inequality for i.i.d. random variables (c.f., Lemma 2.2.9 in van der Vaart and Wellner, 1996) to obtain the convergence rate. Let $M_n = M_2(nd_n)^{1/m_0}$,

$$\begin{aligned} \overline{Z}_{ij}(u,l) =& Z_{ij}(u,l) I\Big\{\big|Q_{i1}x_{ij}\big| \le M_n\Big\}, \\ \widetilde{Z}_{ij}(u,l) =& Z_{ij}(u,l) - \overline{Z}_{ij}(u,l), \end{aligned}$$

where $I\{\cdot\}$ is an indicator function and M_2 is some positive constant. Hence we have

$$\begin{aligned} \Pi_{n1} \le\ & \max_{1\le j\le d_n}\max_{1\le k\le \mathcal{N}_n}\Big|\frac{1}{n}\sum_{i=1}^n\big\{\overline{Z}_{ij}(u_k,l) - \mathsf{E}[\overline{Z}_{ij}(u_k,l)]\big\}\Big| + \\ & \max_{1\le j\le d_n}\max_{1\le k\le \mathcal{N}_n}\Big|\frac{1}{n}\sum_{i=1}^n\big\{\widetilde{Z}_{ij}(u_k,l) - \mathsf{E}[\widetilde{Z}_{ij}(u_k,l)]\big\}\Big| \\ \equiv\ & \Pi_{n3} + \Pi_{n4}. \end{aligned} \tag{C.6}$$

Note that for $M_3 > 0$ and any $\varepsilon > 0$, by (A.1), (C.2) and the Markov inequality,

$$\begin{aligned} \mathsf{P}\left(\Pi_{n4} > M_3\xi_n\right) &\leq \sum_{j=1}^{d_n}\sum_{i=1}^{n}\mathsf{P}\left(\left|Q_{i1}x_{ij}\right| > M_n\right) \\ &\leq M_2^{-m_0}\mathsf{E}\left[\left|Q_{i1}x_{ij}\right|^{m_0}\right] < \varepsilon, \end{aligned}$$

if we choose $M_2 > \mathsf{E}\left[\left|Q_{i1}x_{ij}\right|^{m_0}\right]^{1/m_0}\varepsilon^{-1/m_0}$. Then, by letting ε be arbitrarily small, we can show that

$$\Pi_{n4} = O_P(\xi_n). \tag{C.7}$$

Note that

$$\left|\overline{Z}_{ij}(u_k,l) - \mathsf{E}[\overline{Z}_{ij}(u_k,l)]\right| \leq \frac{M_4 M_n}{h} \tag{C.8}$$

and

$$\mathrm{Var}\left[\overline{Z}_{ij}(u_k,l)\right] \leq \frac{M_4}{h} \tag{C.9}$$

for some $M_4 > 0$. By (C.2), (C.7), (C.8) and Lemma 2.2.9 in van der Vaart and Wellner (1996), we have

$$\begin{aligned} \mathsf{P}(\Pi_{n3} > M_3\xi_n) &\leq 2d_n\mathcal{N}_n\exp\left\{\frac{-n^2M_3^2\xi_n^2}{2nM_4/h + 2M_4M_3n\xi_nM_n/(3h)}\right\} \\ &\leq 2d_n\mathcal{N}_n\exp\left\{-M_3\log h^{-1}\right\} = o(1), \end{aligned} \tag{C.10}$$

where M_3 is chosen such that

$$M_3 > 3M_4, \quad d_n\mathcal{N}_n\exp\left\{-M_3\log h^{-1}\right\} = o(1),$$

which are possible as d_n is diverging with a polynomial rate. Hence we have

$$\Pi_{n3} = O_P(\xi_n). \tag{C.11}$$

In view of (C.4)–(C.7) and (C.11), we have shown (C.3), completing the proof of Lemma C.1. ◻

Lemma C.2. *Suppose that Assumptions 1 and 3 in Appendix A are satisfied. Moreover, suppose that the dimension* $d_n \propto \exp\{(nh)^{\tau_2}\}$ *with* $0 \leq \tau_2 < 1$, $\mathsf{E}\left(Q_{i1}\middle|X_i,U_i\right) = 0$ *a.s., the moment condition (A.2) in Appendix A holds for all* $m \geq 2$*, and* $h \propto n^{-\delta_1}$ *with* $0 < \delta_1 < 1$*. Then we have, as* $n \to \infty$,

$$\max_{1\leq j\leq d_n}\sup_{u\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}Z_{ij}(u,l)\right| = o_P\Big(\big(\frac{\log h^{-1}}{nh}\big)^{\tau_3/2}\Big) \tag{C.12}$$

for any $l = 0, 1, 2, \cdots,\ 0 < \tau_3 < 1 - \tau_2$.

Proof. The proof of (C.12) is similar to the proof of (C.3) in Lemma C.1. The major difference is the way of dealing with Π_{n1}. Because of the stronger moment condition in (A.2), we may directly use a different exponential inequality (Lemma 2.2.11 in van der Vaart and Wellner, 1996) and do not need to apply the truncation method. By replacing ξ_n by $\xi_n(\tau_3) \equiv \left(\frac{\log h^{-1}}{nh}\right)^{\tau_3/2}$, we may re-define $r = o\big(\xi_n(\tau_3)h^2\big)$ and thus $\mathcal{N}_n = O\big(r^{-1}\big)$, where r is the radius used in the finite covering technique (c.f., the proof of Lemma C.1).

Note that there exists a positive constant M_5 such that

$$\mathsf{E}\left[\left|Z_{ij}(u,l)\right|^m\right] \le \frac{M_5}{2h} m!(h^{-1})^{m-2} \tag{C.13}$$

for all $m \ge 2$, by using the moment condition (A.2). Then, by (C.13) and Lemma 2.2.11 in van der Vaart and Wellner (1996) with $M = h^{-1}$ and $v_i = M_5/h$, we can show that for any $\epsilon > 0$

$$\begin{aligned}
\mathsf{P}\big(\Pi_{n1} > \epsilon\xi_n(\tau_3)\big) &\le 2d_n\mathcal{N}_n \exp\left\{\frac{-n^2\epsilon^2\xi_n^2(\tau_3)}{2nM_5/h + 2n\epsilon\xi_n(\tau_3)/h}\right\} \\
&\le 2d_n\mathcal{N}_n \exp\left\{-\frac{\epsilon^2(\log h^{-1})^{\tau_3}}{3M_5}(nh)^{1-\tau_3}\right\} \\
&= 2\mathcal{N}_n \exp\left\{(nh)^{\tau_2} - \frac{\epsilon^2\delta_1^{\tau_3}(\log n)^{\tau_3}}{3M_5}(nh)^{1-\tau_3}\right\} \\
&= o(1)
\end{aligned} \tag{C.14}$$

as $1 - \tau_3 > \tau_2$. The remaining proof is the same as that in the proof of Lemma C.1. Hence details are omitted here to save space. □

Proof of (B.22). To simplify the presentation, we let

$$\widetilde{\mathcal{V}}_n = \mathcal{V}_n\big(\mathcal{A}_0 - \widetilde{\mathcal{A}}_n, h(\mathcal{B}_0 - \widetilde{\mathcal{B}}_n)\big)$$

and

$$\widetilde{\mathcal{V}}_n(\mathcal{U}_1, \mathcal{U}_2) = \mathcal{V}_n\big(\mathcal{A}_0 - \widetilde{\mathcal{A}}_n + \gamma_n^*\mathcal{U}_1,\ h(\mathcal{B}_0 - \widetilde{\mathcal{B}}_n) + \gamma_n^*\mathcal{U}_2\big).$$

Note that

$$\mathcal{I}_n(1) = \mathcal{L}_n^\diamond\big(\mathcal{A}_0 + \gamma_n^*\mathcal{U}_1, \mathcal{B}_0 + \gamma_n^*\mathcal{U}_2/h\big) - \mathcal{L}_n^\diamond\big(\mathcal{A}_0, \mathcal{B}_0\big) \equiv \mathcal{I}_n(1,1) + \mathcal{I}_n(1,2)$$

where

$$\begin{aligned}
\mathcal{I}_n(1,1) &= \gamma_n^*\mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\dot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\ \widetilde{\mathcal{B}}_n), \\
\mathcal{I}_n(1,2) &= \frac{1}{2}\Big[\widetilde{\mathcal{V}}_n^{\mathrm{T}}(\mathcal{U}_1, \mathcal{U}_2)\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\ \widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n(\mathcal{U}_1, \mathcal{U}_2) - \widetilde{\mathcal{V}}_n^{\mathrm{T}}\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\ \widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n\Big].
\end{aligned}$$

By Taylor's expansion, we have

$$\begin{aligned}
\mathcal{I}_n(1,1) &= \gamma_n^* \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\dot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\ \widetilde{\mathcal{B}}_n) \\
&= \gamma_n^* \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\dot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0) - \gamma_n^* \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\ddot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0)\widetilde{\mathcal{V}}_n + \\
&\quad o_P\big((\gamma_n^*)^2\big) \cdot \big(\|\mathcal{U}_1\|^2 + \|\mathcal{U}_2\|^2\big),
\end{aligned}$$

where $\mathcal{V}_n(\mathcal{U}_1,\ \mathcal{U}_2)$ is defined in Section 2.2 of the main document. On the other hand, by some elementary calculations, we also have

$$\begin{aligned}
\mathcal{I}_n(1,2) &= \frac{1}{2}\Big[\widetilde{\mathcal{V}}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n(\mathcal{U}_1,\mathcal{U}_2) - \widetilde{\mathcal{V}}_n^{\mathrm{T}}\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n(\mathcal{U}_1,\mathcal{U}_2)\Big] + \\
&\quad \frac{1}{2}\Big[\widetilde{\mathcal{V}}_n^{\mathrm{T}}\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n(\mathcal{U}_1,\mathcal{U}_2) - \widetilde{\mathcal{V}}_n^{\mathrm{T}}\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n\Big] \\
&= \frac{\gamma_n^*}{2}\mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\widetilde{\mathcal{V}}_n(\mathcal{U}_1,\mathcal{U}_2) + \\
&\quad \frac{\gamma_n^*}{2}\widetilde{\mathcal{V}}_n^{\mathrm{T}}\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\mathcal{V}_n(\mathcal{U}_1,\ \mathcal{U}_2) \\
&= \frac{1}{2}(\gamma_n^*)^2\mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\mathcal{V}_n(\mathcal{U}_1,\ \mathcal{U}_2) + \\
&\quad \gamma_n^*\widetilde{\mathcal{V}}_n^{\mathrm{T}}\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\mathcal{V}_n(\mathcal{U}_1,\ \mathcal{U}_2) \\
&= \frac{1}{2}(\gamma_n^*)^2\mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n,\widetilde{\mathcal{B}}_n)\mathcal{V}_n(\mathcal{U}_1,\ \mathcal{U}_2) + \\
&\quad \gamma_n^*\mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\ \mathcal{U}_2)\ddot{\mathcal{L}}_n(\mathcal{A}_0,\mathcal{B}_0)\widetilde{\mathcal{V}}_n + o_P\big((\gamma_n^*)^2\big) \cdot \big(\|\mathcal{U}_1\|^2 + \|\mathcal{U}_2\|^2\big).
\end{aligned}$$

We can easily prove (B.22) by using the above two results on the asymptotic expansion for $\mathcal{I}_n(1,1)$ and $\mathcal{I}_n(1,2)$. □

Proof (B.23). Recall that

$$\mathcal{I}_n(4) = \gamma_n^* \mathcal{V}_n^{\mathrm{T}}(\mathcal{U}_1,\mathcal{U}_2)\dot{\mathcal{L}}_n(\mathcal{A}_0,\mathcal{B}_0). \tag{C.15}$$

By Taylor's expansion for $q_1(\cdot,\cdot)$ and Assumption 4, when $|U_i - U_k| = O(h)$, we have

$$\begin{aligned}
& q_1\Big\{\sum_{j=1}^{d_n}\big[a_j(U_k) + \dot{a}_j(U_k)(U_i - U_k)\big]x_{ij},\ y_i\Big\} \\
=\ & q_1\Big\{\sum_{j=1}^{s_{n2}}\big[a_j(U_k) + \dot{a}_j(U_k)(U_i - U_k)\big]x_{ij},\ y_i\Big\} \\
=\ & q_1\Big[\sum_{j=1}^{s_{n2}} a_j(U_i)x_{ij},\ y_i\Big] + O_P(s_{n2}h^2) \\
\equiv\ & Q_{i1} + O_P(s_{n2}h^2), \qquad \text{(C.16)}
\end{aligned}$$

which implies that

$$
\begin{aligned}
\mathcal{I}_n(4) &= \frac{\gamma_n^*}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} Q_{i1}X_i^{\mathrm{T}}\mathbf{u}_1(k)K_h(U_i-U_k) + \\
&\quad \frac{\gamma_n^*}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} Q_{i1}X_i^{\mathrm{T}}\mathbf{u}_2(k)\Big(\frac{U_i-U_k}{h}\Big)K_h(U_i-U_k) + \\
&\quad O_P(\gamma_n^* s_{n2}^{3/2} n^{1/2} h^2)\cdot\big(\|\mathcal{U}_1\|+\|\mathcal{U}_2\|\big). \qquad \text{(C.17)}
\end{aligned}
$$

Note that (U_i, X_i, y_i), $i = 1, \cdots, n$, are independent and identically distributed. By Assumptions 1, 2(i) and 3 in Appendix A, and the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
& \mathsf{E}\Big[\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} Q_{i1}X_i^{\mathrm{T}}\mathbf{u}_1(k)K_h(U_i-U_k)\Big]^2 \\
\leq\ & \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\Big[\sum_{i=1}^{n} Q_{i1}X_i^{\mathrm{T}}\mathbf{u}_1(k)K_h(U_i-U_k)\Big]^2 \\
=\ & \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\Big(\mathsf{E}\Big\{\Big[\sum_{i=1}^{n} Q_{i1}X_i^{\mathrm{T}}\mathbf{u}_1(k)K_h(U_i-U_k)\Big]^2\Big|U_k\Big\}\Big) \\
=\ & \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n}\mathsf{E}\Big\{\mathsf{E}\Big[Q_{i1}^2\mathbf{u}_1^{\mathrm{T}}(k)X_iX_i^{\mathrm{T}}\mathbf{u}_1(k)K_h^2(U_i-U_k)\Big|U_k\Big]\Big\} \\
=\ & O(s_{n2}h^{-1})\cdot\|\mathcal{U}_1\|^2.
\end{aligned}
$$

Similarly, we can also show that

$$
\mathsf{E}\Big[\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{n} Q_{i1}X_i^{\mathrm{T}}\mathbf{u}_2(k)\Big(\frac{U_i-U_k}{h}\Big)K_h(U_i-U_k)\Big]^2 = O(s_{n2}h^{-1})\cdot\|\mathcal{U}_2\|^2.
$$

Noting that $s_{n2}h^2 \propto (nh)^{-1/2}$, we have

$$
\mathcal{I}_n(4) = O_P\big((\gamma_n^*)^2 n^{1/2}\big)\cdot\big(\|\mathcal{U}_1\|+\|\mathcal{U}_2\|\big), \qquad \text{(C.18)}
$$

which completes the proof of (B.23). □

Proof (B.38). Let

$$
\dot{\mathcal{L}}_n(\mathcal{A},\mathcal{B}|\boldsymbol{\alpha}_j) = \Big[\dot{\mathcal{L}}_{n1}(\mathbf{a}_1,\mathbf{b}_1,j),\cdots,\dot{\mathcal{L}}_{nn}(\mathbf{a}_n,\mathbf{b}_n,j)\Big]^{\mathrm{T}},
$$

where $\dot{\mathcal{L}}_{nk}(\mathbf{a}_k, \mathbf{b}_k, j)$ is the j-th element of $\dot{\mathcal{L}}_{nk}(\mathbf{a}_k, \mathbf{b}_k)$ defined in Section 2.2 of the main document; and let

$$\ddot{\mathcal{L}}_n(\mathcal{A}, \mathcal{B}|\boldsymbol{\alpha}_j) = \mathsf{diag}\left\{\ddot{\mathcal{L}}_{n1}(\mathbf{a}_1, \mathbf{b}_1, j), \cdots, \ddot{\mathcal{L}}_{nn}(\mathbf{a}_n, \mathbf{b}_n, j)\right\},$$

where $\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k, \mathbf{b}_k, j)$ is the j-th row of $\ddot{\mathcal{L}}_{nk}(\mathbf{a}_k, \mathbf{b}_k)$ defined in Section 2.2. Observe that

$$\dot{\mathcal{L}}_n^{\diamond}(\mathcal{A}, \mathcal{B}|\boldsymbol{\alpha}_j) = \dot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j) + \ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\mathcal{A}, h\mathcal{B}) - \mathcal{V}_n(\widetilde{\mathcal{A}}_n, h\widetilde{\mathcal{B}}_n)\right]. \tag{C.19}$$

By Taylor's expansion of $q_1(\cdot,\cdot)$ and Proposition 3.1, and following the argument in the proof of (B.22) above, we have

$$\begin{aligned}\dot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j) &= \dot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0|\boldsymbol{\alpha}_j) + \ddot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\widetilde{\mathcal{A}}_n, h\widetilde{\mathcal{B}}_n) - \mathcal{V}_n(\mathcal{A}_0, h\mathcal{B}_0)\right] \\ &\quad + O_P(s_{n2}^2\lambda_1^2)\cdot I_n, \end{aligned} \tag{C.20}$$

where I_n is an $n \times n$ identity matrix. Similarly, we may also show that

$$\begin{aligned} &\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\mathcal{A}, h\mathcal{B}) - \mathcal{V}_n(\widetilde{\mathcal{A}}_n, h\widetilde{\mathcal{B}}_n)\right] \\ = &\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\mathcal{A}, h\mathcal{B}) - \mathcal{V}_n(\mathcal{A}_0, h\mathcal{B}_0)\right] - \\ &\ddot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\widetilde{\mathcal{A}}_n, h\widetilde{\mathcal{B}}_n) - \mathcal{V}_n(\mathcal{A}_0, h\mathcal{B}_0)\right] + O_P(s_{n2}^2\lambda_1^2)\cdot I_n. \end{aligned} \tag{C.21}$$

By (C.19)–(C.21), we may show that

$$\begin{aligned}\dot{\mathcal{L}}_n^{\diamond}(\mathcal{A}, \mathcal{B}|\boldsymbol{\alpha}_j) &= \dot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0|\boldsymbol{\alpha}_j) + \ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\mathcal{A}, h\mathcal{B}) - \mathcal{V}_n(\mathcal{A}_0, h\mathcal{B}_0)\right] \\ &\quad + O_P(s_{n2}^2\lambda_1^2)\cdot I_n. \end{aligned} \tag{C.22}$$

By (B.5) and the standard argument in the kernel-based smoothing, we have

$$\max_{s_{n2}+1\le j\le d_n}\left\|\dot{\mathcal{L}}_n(\mathcal{A}_0, \mathcal{B}_0|\boldsymbol{\alpha}_j)\right\| = O_P\big(h^{-1/2}\sqrt{\log h^{-1}}\big). \tag{C.23}$$

By (A.5) in Assumption 6(ii) and (B.31), we may also show that

$$\max_{s_{n2}+1\le j\le d_n}\left\|\ddot{\mathcal{L}}_n(\widetilde{\mathcal{A}}_n, \widetilde{\mathcal{B}}_n|\boldsymbol{\alpha}_j)\left[\mathcal{V}_n(\mathcal{A}, h\mathcal{B}) - \mathcal{V}_n(\mathcal{A}_0, h\mathcal{B}_0)\right]\right\| = O_P\big(h^{-1/2}s_{n2}^{1/2}\big) \tag{C.24}$$

when $\mathcal{A} = \overline{\mathcal{A}}_n^{bo}$ and $\mathcal{B} = \overline{\mathcal{B}}_n^{bo}$.

Using (C.22)–(C.24), we may complete the proof of (B.38). □

REFERENCES

Bickel, P., Ritov, Y. and Tsybakov, A. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, **37**, 1705–1732.

Fan, J. and Lv, J. (2011). Non-concave penalized likelihood with NP-Dimensionality. *IEEE: Information Theory*, **57**, 5467–5484.

Fan, J., Xue, L. and Zou, H. (2014). Strong oracle optimality of folded concave penalized estimation. *The Annals of Statistics*, **42**, 819–849.

Lian, H. (2012). Variable selection for high-dimensional generalized varying-coefficient models. *Statistica Sinica*, **22**, 1563–1588.

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

Wang, H. and Xia, Y. (2009). Shrinkage estimation of the varying-coefficient model. *Journal of the American Statistical Association*, **104**, 747–757.

Department of Mathematics
The University of York
Heslington
York YO10 5DD
The United Kingtom
E-mail: degui.li@york.ac.uk

Department of Mathematics
The University of York
Heslington
York YO10 5DD
The United Kingtom
E-mail: yk612@york.ac.uk

Department of Mathematics
The University of York
Heslington
York YO10 5DD
The United Kingtom
E-mail: wenyang.zhang@york.ac.uk