On A Principal Varying Coefficient Model

APPENDIX: TECHNICAL DETAILS

Qian Jiang, Hansheng Wang, Yingcun Xia and Guohua Jiang

To establish the asymptotic theory for the proposed estimation methods, we need the following technical assumptions.

(C.1) (*The Index Variable*). The *index* variable U has a bounded compact support $\mathcal{D}$ and a probability density function $f(u)$, which is Lipschitz continuous and bounded away from 0 on $\mathcal{D}$.

(C.2) (*Smoothness Assumptions*). Every component of $W(u) = E(XX^\top|U = u)$ and $L(u) = E(XY^\top|U = u)$ is Lipschitz continuous. In addition to that, we assume $\beta_0(u)$ has continuous second order derivatives in $u \in \mathcal{D}$. The matrix $W(u)$ is positive definite for all $u \in \mathcal{D}$.

(C.3) (*Moment Conditions*). There exist $s > 2$ and $\delta < 2 - s^{-1}$, such that $E\|X\|^s < \infty$ with $n^{2\delta-1}h \to \infty$, where $\|\cdot\|$ stands for a typical L_2 norm.

(C.4) (*The Kernel and Bandwidth*). We assume that the kernel function $K(\cdot)$ is a

Qian Jiang is a PhD student, the Department of Statistics, National University of Singapore. Hansheng Wang is a Professor of Statistics, Guanghua School of Management, Peking University, China. Yingcun Xia is a Professor of Statistics, Department of statistics, National University of Singapore, and a Professor of Finance, Nanjing University, China. Guohua Jiang is a Professor of Accounting, Guanghua School of Management, Peking University, China. Wang's work was supported in part by National Natural Science Foundation of China (11131002, 11271032), Fox Ying Tong Education Foundation, and the Center for Statistical Science at Peking University. Xia's work is partially supported by a grand from the National University of Singapore (R-155-000-121-112). Guohua Jiang's work was supported in part by the National Natural Science Foundation of China (71132004). The authors are very grateful for the helpful comments of 3 referees, the associate editor, and the editor.

symmetric density function with a compact support. Moreover, we assume $h \propto n^{-c}$ with $c > 0$ such that $\sqrt{n}h^2 \to 0$ and $nh/\log n \to \infty$.

We remark that the above regularity conditions are rather standard. Similar assumptions have been used in, for example, Zhang et al (2002) and Fan and Huang (2005). Let $\mu_k = \int t^k K(t)$. Then by (C.4) we have $\mu_0 = 1$ and $\mu_1 = 0$. For ease of exposition, we further standardize $K(.)$ such that $\mu_2 = 1$ in the following proofs. In addition, we denote $U_i - u$ by U_{iu} and $U_i - U_j$ by U_{ij} in the following proofs.

Lemma 1. *Under the regularity conditions (C.1)-(C.4), for the estimator defined below Theorem 1, we have the following expansion*

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}(u|B,\theta) &= \boldsymbol{\gamma}_0(u) + \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2 + \{B^\top W(u)B\}^{-1}\{nf(u)\}^{-1}B^\top \sum_{i=1}^n K_h(U_{iu})X_i\varepsilon_i \\
&\quad + \{B^\top W(u)B\}^{-1}B^\top W(u)(B_0 - B)\boldsymbol{\gamma}_0(u) + \{B^\top W(u)B\}^{-1}B^\top W(u)(\theta_0 - \theta) \\
&\quad + O_p(h^3 + h\delta_n + \delta_n^2)
\end{aligned}$$

uniformly for any $u \in \mathcal{D}$ and $(\theta, B) \in \Theta_n$.

Proof. Write $Y_i - X_i^\top\theta = \varepsilon_i + X_i^\top B\boldsymbol{\gamma}_0(U_i) + X_i^\top(B_0 - B)\boldsymbol{\gamma}_0(U_i) + X_i^\top(\theta_0 - \theta)$. Thus

$$\begin{aligned}
\sum_{i=1}^n K_h(U_{iu})X_i\{Y_i - X_i^\top\theta\} &= \sum_{i=1}^n K_h(U_{iu})X_i\varepsilon_i + \sum_{i=1}^n K_h(U_{iu})X_iX_i^\top B\boldsymbol{\gamma}_0(U_i) \qquad \text{(A.1)} \\
&\quad + \sum_{i=1}^n K_h(U_{iu})X_iX_i^\top(B_0 - B)\boldsymbol{\gamma}_0(U_i) + S_n(u)(\theta_0 - \theta).
\end{aligned}$$

Let $s_n(u) = \sum_{i=1}^n K_h(U_{iu})$. By Mack and Silverman (1982), we have uniformly for $u \in \mathcal{D}$, $s_n^{-1}(u) = (nf(u))^{-1}(1 + O_p(h^2 + \delta_n))$, and

$$\frac{1}{n}\sum_{i=1}^n K_h(U_{iu})X_iX_i^\top = f(u)W(u)(1 + O_p(h^2 + \delta_n)), \quad \frac{1}{n}\sum_{i=1}^n K_h(U_{iu})X_i\varepsilon_i = O_p(\delta_n).$$

Thus,

$$s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top} = W(u) + O_p(h^2+\delta_n),$$

$$s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}\boldsymbol{\gamma}_0(U_i) = W(u)\boldsymbol{\gamma}_0(u) + O_p(h^2+\delta_n),$$

$$s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_i\varepsilon_i = \{nf(u)\}^{-1}\sum_{i=1}^{n} K_h(U_{iu})X_i\varepsilon_i + O_p(h^2\delta_n+\delta_n^2),$$

and

$$s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}(B_0-B)\boldsymbol{\gamma}_0(U_i) = W(u)(B_0-B)\boldsymbol{\gamma}_0(u) + \|B_0-B\|O_p(h^2+\delta_n)$$

uniformly for $u \in \mathcal{D}$. Combining the above results yields that uniformly in $u \in \mathcal{D}$,

$$\begin{aligned}
& s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}B\boldsymbol{\gamma}_0(U_i) \\
= \; & s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}B\boldsymbol{\gamma}_0(u) + s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}B\{\boldsymbol{\gamma}_0(U_i)-\boldsymbol{\gamma}_0(u)\} \\
= \; & s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}B\boldsymbol{\gamma}_0(u) \\
& + s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}B\{\boldsymbol{\gamma}_0'(u)(U_{iu}) + \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)(U_{iu})^2 + O_p(U_{iu}^3)\} \\
= \; & s_n^{-1}(u)\sum_{i=1}^{n} K_h(U_{iu})X_iX_i^{\top}B\boldsymbol{\gamma}_0(u) + \{f^{-1}(u)f'(u)W'(u)B\boldsymbol{\gamma}_0'(u) + \frac{1}{2}\mu_2 W(u)B\boldsymbol{\gamma}_0''(u)\}h^2 \\
& + O_p(h^3).
\end{aligned}$$

For $(\theta, B) \in \Theta_n$, we have

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}(u|B,\theta) &= (B^\top S_n(u)B)^{-1}B^\top \sum_{i=1}^n K_h(U_{iu})X_i\{Y_i - X_i^\top\theta\} \\
&= (B^\top s_n^{-1}(u)S_n(u)B)^{-1}B^\top \Big(s_n^{-1}(u)\sum_{i=1}^n K_h(U_{iu})X_i\{Y_i - X_i^\top\theta\}\Big) \\
&= \boldsymbol{\gamma}_0(u) + \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2 + \{B^\top W(u)B\}^{-1}\{nf(u)\}^{-1}B^\top \sum_{i=1}^n K_h(U_{iu})X_i\varepsilon_i \\
&\quad +\{B^\top W(u)B\}^{-1}B^\top W(u)(B_0 - B)\boldsymbol{\gamma}_0(u) + \{B^\top W(u)B\}^{-1}B^\top W(u)(\theta_0 - \theta) \\
&\quad +O_p(h^3 + h\delta_n + \delta_n^2).
\end{aligned}$$

As a special case,

$$\begin{aligned}
\hat{\boldsymbol{\gamma}}(u|B_0,\theta_0) &= \boldsymbol{\gamma}_0(u) + \frac{1}{2}\mu_2\boldsymbol{\gamma}_0''(u)h^2 + \{B_0^\top W(u)B_0\}^{-1}\{nf(u)\}^{-1}B_0^\top \sum_{i=1}^n K_h(U_{iu})X_i\varepsilon_i \\
&\quad +O_p(h^3 + h\delta_n + \delta_n^2).
\end{aligned}$$

We have completed the proof. $\square$

Proof of Theorems 1. By Theorem 1 of Fan and Zhang (2000b) or Lemma 1, we have

$$\sup_{u\in\mathcal{D}} |\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)| = O_p(h^2 + \delta_n), \tag{A.2}$$

where $\delta_n = \{nh/\log(n)\}^{-1/2}$. Theorems 1 follows immediately from (A.2). $\square$

Proof of Theorem 2. Let $\alpha = (\theta^\top, vec(B)^\top)^\top$, $\alpha_0 = (\alpha_{0,1}, ..., \alpha_{0,p(d_0+1)})^\top = (\theta_0^\top, vec(B_0)^\top)^\top$, $\hat{\alpha} = (\hat{\theta}^\top, vec(\hat{B})^\top)^\top$ and $Q(\alpha) = Q(\theta, B)$. By Taylor expansion about α_0, we have

$$0 = \frac{\partial Q(\hat{\alpha})}{\partial \alpha} = \frac{\partial Q(\alpha_0)}{\partial \alpha} + \frac{\partial^2 Q(\alpha^*)}{\partial\alpha\partial\alpha^\top}(\hat{\alpha} - \alpha_0),$$

where α^* lies on the line segment between α_0 and $\hat{\alpha}$. Let $\Delta_i(\alpha) = Y_i - X_i^\top\theta - X_i^\top B\tilde{\boldsymbol{\gamma}}(U_i)$, $\eta_i(\alpha) = Y_i - X_i^\top\theta - X_i^\top B\boldsymbol{\gamma}_0(U_i)$, then $\Delta_i(\alpha) = \eta_i(\alpha) - X_i^\top B(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i))$, $\eta_i(\alpha_0) = \varepsilon_i$, and

$$Q(\alpha) = \sum_{i=1}^n \Delta_i^2(\alpha).$$

Let $Q_0(\alpha) = \sum_{i=1}^n \eta_i^2(\alpha)$. From Lemma 1, when $||\alpha - \alpha_0|| = O_p(h^2 + \delta_n)$ we have

$$\sup_{u\in\mathcal{D}} \|\tilde{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u)\| = O_p(h^2 + \delta_n) = o_p(1).$$

Thus $\Delta_i(\alpha) = \eta_i(\alpha) - X_i^\top B(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i)) = \eta_i(\alpha) + o_p(1)$, $\partial\Delta_i(\alpha)/\partial\alpha = \partial\eta_i(\alpha)/\partial\alpha + o_p(1)$. It follows that

$$\begin{aligned}
\frac{1}{2n}\frac{\partial^2 Q(\alpha)}{\partial\alpha\partial\alpha^\top} &= \frac{1}{n}\sum_{i=1}^n \frac{\partial\Delta_i(\alpha)}{\partial\alpha}\frac{\partial\Delta_i(\alpha)}{\partial\alpha^\top} + \frac{1}{n}\sum_{i=1}^n \Delta_i(\alpha)\frac{\partial^2\Delta_i(\alpha)}{\partial\alpha\partial\alpha^\top} \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\partial\eta_i(\alpha)}{\partial\alpha}\frac{\partial\eta_i(\alpha)}{\partial\alpha^\top} + \frac{1}{n}\sum_{i=1}^n \eta_i(\alpha)\frac{\partial^2\eta_i(\alpha)}{\partial\alpha\partial\alpha^\top} + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\partial\eta_i(\alpha_0)}{\partial\alpha}\frac{\partial\eta_i(\alpha_0)}{\partial\alpha^\top} + \frac{1}{n}\sum_{i=1}^n \eta_i(\alpha_0)\frac{\partial^2\eta_i(\alpha_0)}{\partial\alpha\partial\alpha^\top} + o_p(1) \\
&\to E\Big\{\frac{\partial\eta_1(\alpha_0)}{\partial\alpha}\frac{\partial\eta_1(\alpha_0)}{\partial\alpha^\top}\Big\} \\
&= E\Big\{\begin{pmatrix} X \\ \boldsymbol{\gamma}_0(U)\otimes X\end{pmatrix}\begin{pmatrix} X \\ \boldsymbol{\gamma}_0(U)\otimes X\end{pmatrix}^\top\Big\} = \Sigma_0, \quad \text{in probability.}
\end{aligned}$$

In the last step, $\partial^2\eta_i(\alpha_0)/(\partial\alpha\partial\alpha^\top) = 0$ is used. Write

$$\frac{1}{2\sqrt{n}}\frac{\partial Q(\alpha_0)}{\partial\alpha} = \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\eta_i(\alpha_0) - X_i^\top B_0(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i))\}\{\frac{\partial\eta_i(\alpha_0)}{\partial\alpha} + \frac{\partial\Delta_i(\alpha_0)}{\partial\alpha} - \frac{\partial\eta_i(\alpha_0)}{\partial\alpha}\}. \tag{A.3}$$

Let $Z_{n0} = Z_{n1} + Z_{n2}$ with $Z_{n1} = n^{-1/2}\sum_{i=1}^n \eta_i(\alpha_0)\partial\eta_i(\alpha_0)/\partial\alpha$ and

$$Z_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_i(\alpha_0)\Big(\frac{\partial\Delta_i(\alpha_0)}{\partial\alpha} - \frac{\partial\eta_i(\alpha_0)}{\partial\alpha}\Big) - \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^\top B_0(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i))\frac{\partial\eta_i(\alpha_0)}{\partial\alpha}.$$

By Lemma 1, we have

$$\begin{aligned}\Big|\frac{1}{2\sqrt{n}}\frac{\partial Q(\alpha_0)}{\partial\alpha} - Z_{n0}\Big| &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^\top B_0(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i))\Big(\frac{\partial\Delta_i(\alpha_0)}{\partial\alpha} - \frac{\partial\eta_i(\alpha_0)}{\partial\alpha}\Big)\Big| \\ &\le \sqrt{n}\max_{1\le i\le n}|X_i^\top B_0(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i))| \max_{1\le i\le n}\Big\|\frac{\partial\Delta_i(\alpha_0)}{\partial\alpha} - \frac{\partial\eta_i(\alpha_0)}{\partial\alpha}\Big\| \\ &= \sqrt{n}O_p(h^2+\delta_n)O_p(h^2+\delta_n) = o_p(1).\end{aligned}$$

It is easy to check that

$$Z_{n1} = -n^{-1/2}\sum_{i=1}^n \begin{pmatrix} X_i \\ \boldsymbol{\gamma}_0(U_i)\otimes X_i\end{pmatrix}\varepsilon_i.$$

Let $\ell(U) = (1, \boldsymbol{\gamma}_0(U)^\top)^\top \otimes W(U)$ and $\bar{\ell} = E\ell(U)$. Write $Z_{n2} = E_{n1} - E_{n2}$, where

$$E_{n1} = n^{-1/2}\sum_{i=1}^n \eta_i(\alpha_0)\Big(\frac{\partial\Delta_i(\alpha_0)}{\partial\alpha} - \frac{\partial\eta_i(\alpha_0)}{\partial\alpha}\Big),$$

$$E_{n2} = n^{-1/2}\sum_{i=1}^n X_i^\top B_0(\tilde{\boldsymbol{\gamma}}(U_i) - \boldsymbol{\gamma}_0(U_i))\frac{\partial\eta_i(\alpha_0)}{\partial\alpha}.$$

Under assumptions (C.1)-(C.4), we can show that

$$E_{n1} = o_p(1) \tag{A.4}$$

and

$$E_{n2} = \frac{1}{2}E\{(\ell(U)-\bar{\ell})B_0\boldsymbol{\gamma}_0''(U)\}n^{1/2}h^2 + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(U_j)-\bar{\ell})V(U_j)X_j\varepsilon_j$$
$$+\bar{\ell}B_0\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\gamma}_0(U_i) + o_p(1). \tag{A.5}$$

Thus, we have

$$Z_{n2} = \frac{1}{2}E\{(\ell(U)-\bar{\ell})B_0\boldsymbol{\gamma}_0''(U)\}n^{1/2}h^2 + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(U_j)-\bar{\ell})V(U_j)X_j\varepsilon_j$$
$$+\begin{pmatrix} EW(U) \\ E\{\boldsymbol{\gamma}_0(U)\otimes W(U)\}\end{pmatrix} B_0\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\gamma}_0(U_i) + o_p(1),$$

where $W(u)$ and $V(u)$ are defined in Theorem 2. By the central limit theorem (CLT), we have

$$Z_{n1} + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\ell(U_j)-\bar{\ell})V(U_j)X_j\varepsilon_j \to N(0,\Sigma_1),$$

where Σ_1 is given in Theorem 2. On the other hand, since $E\boldsymbol{\gamma}_0(U)=0$, we have

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\gamma}_0(U_i) \to N\Big(0, E\{\boldsymbol{\gamma}_0(U)\boldsymbol{\gamma}_0^\top(U)\}\Big).$$

Theorem 2 follows from last three equations and (A.3).

Now, we turn to prove (A.4) and (A.5). We only give the details for the latter. Decompose E_{n2} into two terms.

$$E_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i^\top B_0(\hat{\boldsymbol{\gamma}}(U_i)-\boldsymbol{\gamma}_0(U_i))\frac{\partial\eta_i(\alpha_0)}{\partial\alpha} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i^\top B_0\bar{\boldsymbol{\gamma}}\frac{\partial\eta_i(\alpha_0)}{\partial\alpha} \triangleq E_{n2}^1 - E_{n2}^2, \tag{A.6}$$

where $\hat{\boldsymbol{\gamma}}(U_i) = \hat{\boldsymbol{\gamma}}(U_i|\theta_0, B_0)$ and $\bar{\boldsymbol{\gamma}} = n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\gamma}}(U_i)$. From Lemma 1, we have

$$E_{n2}^1 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} X_i \\ \boldsymbol{\gamma}_0(U_i)\otimes X_i \end{pmatrix} X_i^\top B_0\{\frac{1}{2}\boldsymbol{\gamma}_0''(U_i)h^2 + R_n(U_i) + O_p(h^3 + h\delta_n + \delta_n^2)\},$$

where $R_n(U_i) = \{nf(U_i)B_0^\top W(U_i)B_0\}^{-1}B_0^\top \sum_{j=1}^n K_h(U_{ij})X_j\varepsilon_j$. It follows from the laws of large numbers

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} X_i \\ \boldsymbol{\gamma}_0(U_i)\otimes X_i \end{pmatrix} X_i^\top B_0\boldsymbol{\gamma}_0''(U_i)h^2 = E\{\ell(U)B_0\boldsymbol{\gamma}_0''(U)\}n^{1/2}h^2 + o_p(1). \quad \text{(A.7)}$$

As $f(u)$ is bounded away from 0, we then have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} X_i \\ \boldsymbol{\gamma}_0(U_i)\otimes X_i \end{pmatrix} X_i^\top B_0R_n(U_i) &= \frac{1}{\sqrt{n}}\sum_{j=1}^n \Big\{\sum_{i=1}^n \begin{pmatrix} X_i \\ \boldsymbol{\gamma}_0(U_i)\otimes X_i \end{pmatrix} X_i^\top V(U_i) \\
&\quad \times \frac{1}{nf(U_i)}K_h(U_{ij})\Big\}X_j\varepsilon_j \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^n \ell(U_j)V(U_j)X_j\varepsilon_j + \Delta_n, \qquad \text{(A.8)}
\end{aligned}$$

where

$$\Delta_n = \frac{1}{\sqrt{n}}\sum_{j=1}^n \Big\{\sum_{i=1}^n \begin{pmatrix} X_i \\ \boldsymbol{\gamma}_0(U_i)\otimes X_i \end{pmatrix} X_i^\top V(U_i)\frac{1}{nf(U_i)}K_h(U_{ij}) - \ell(U_j)V(U_j)\Big\}X_j\varepsilon_j.$$

By simple calculation, we have $Var(\Delta_n) = O\{(h^2+\delta_n)^2\}$ and thus

$$\Delta_n = O_p(h^2 + \delta_n). \qquad \text{(A.9)}$$

For E_{n2}^2, by Lemma 1 we have $\bar{\boldsymbol{\gamma}} = O_p(h^2 + \delta_n)$,

$$\bar{\boldsymbol{\gamma}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\gamma}_0(U_i) + \frac{1}{2}E\boldsymbol{\gamma}_0''(U)h^2 + \frac{1}{n}\sum_{i=1}^{n}(B_0^\top W(U_i)B_0)^{-1}B_0^\top X_i\varepsilon_i + o_p(n^{-1/2})$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\eta_i(\alpha_0)}{\partial\alpha}X_i^\top = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} X_iX_i^\top \\ \boldsymbol{\gamma}_0(U_i)\otimes X_iX_i^\top\end{pmatrix} = \bar{\ell} + O_p(n^{-1/2}).$$

It follows from Lemma 1 that

$$E_{n2}^2 = \bar{\ell}\Big\{B_0\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\gamma}_0(U_i) + \frac{1}{2}B_0E\boldsymbol{\gamma}_0''(U)\sqrt{n}h^2 + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}V(U_i)X_i\varepsilon_i\Big\} + o_p(1). \quad \text{(A.10)}$$

Equation (A.5) follows from (A.6)-(A.10) and the following fact

$$\bar{\ell}B_0\frac{1}{2}E\boldsymbol{\gamma}_0''(U)h^2 - E\{\ell(U)B_0\frac{1}{2}\boldsymbol{\gamma}_0''(U)\}n^{-1/2}h^2 = -\frac{1}{2}E\{(\ell(U)-\bar{\ell})B_0\boldsymbol{\gamma}_0''(U)\}h^2.$$

This completes the proof. □

Proof of Theorem 4. For any fixed d, denote the estimators of θ_0, B_0 and $\boldsymbol{\gamma}_0(u)$ by $\hat{\theta}_d$, $\hat{B}_d$ and $\hat{\boldsymbol{\gamma}}_d(u)$ respectively.

Case 1. ($d < d_0$, underfitted model) By the proof of Theorem 1, $\hat{\theta}_d - \theta_d = O_p(h^2 + \delta_n)$ and that there exist nonrandom matrix B_d and function $\boldsymbol{\gamma}_d(u)$ such that

$$\hat{B}_d - B_d = O_p(h^2 + \delta_n), \quad \hat{\boldsymbol{\gamma}}_d(u) - \boldsymbol{\gamma}_d(u) = O_p(h^2 + \delta_n)$$

uniformly for $u \in \mathcal{D}$. By the definition of d_0, if $d < d_0$ then $E||B_0\boldsymbol{\gamma}_0(U) - B_d\boldsymbol{\gamma}_d(U)|| > 0$.

It is easy to see by the above facts and the CLT that

$$
\begin{aligned}
\hat{\sigma}_d^2 &= n^{-1}\sum_{i=1}^n \{Y_i - (\hat{\theta}_d + \hat{B}_d\hat{\boldsymbol{\gamma}}_d(U_i))^\top X_i\}^2 \\
&= n^{-1}\sum_{i=1}^n \{Y_i - (\theta_d + B_d\boldsymbol{\gamma}_d(U_i))^\top X_i\}^2 + O_p(h^2+\delta_n) \\
&= n^{-1}\sum_{i=1}^n \{\varepsilon_i + (B_0\boldsymbol{\gamma}_0(U_i) - B_d\boldsymbol{\gamma}_d(U_i))^\top X_i\}^2 + O_p(h^2+\delta_n) \\
&= n^{-1}\sum_{i=1}^n \varepsilon_i^2 + 2n^{-1}\sum_{i=1}^n \varepsilon_i (B_0\boldsymbol{\gamma}_0(U_i) - B_d\boldsymbol{\gamma}_d(U_i))^\top X_i \\
&\quad + n^{-1}\sum_{i=1}^n \{(B_0\boldsymbol{\gamma}_0(U_i) - B_d\boldsymbol{\gamma}_d(U_i))^\top X_i\}^2 + O_p(h^2+\delta_n) \\
&= \sigma^2 + E\{(B_0\boldsymbol{\gamma}_0(U) - B_d\boldsymbol{\gamma}_d(U))^\top X\}^2 + O_p(h^2+\delta_n+n^{-1/2}). \qquad \text{(A.11)}
\end{aligned}
$$

Therefore, as a special case we have $\hat{\sigma}_{d_0}^2 = \sigma^2 + O_p(h^2+\delta_n+n^{-1/2})$. Note that

$$
\begin{aligned}
E\{(B_0\boldsymbol{\gamma}_0(U) - B_d\boldsymbol{\gamma}_d(U))^\top X\}^2 &= E\{(B_0\boldsymbol{\gamma}_0(U) - B_d\boldsymbol{\gamma}_d(U))^\top W(U)(B_0\boldsymbol{\gamma}_0(U) - B_d\boldsymbol{\gamma}_d(U))\} \\
&\geq \lambda_1(W(u))E||B_0\boldsymbol{\gamma}_0(U) - B_d\boldsymbol{\gamma}_d(U)|| \stackrel{def}{=} c_0 > 0.
\end{aligned}
$$

Therefore, for $d < d_0$ we have $\hat{\sigma}_d^2 \geq \sigma_{d_0}^2 + c_0 + O_p(h^2+\delta_n+n^{-1/2})$. Therefore

$$
P\left\{\mathrm{BIC}(d) > \mathrm{BIC}(d_0)\right\} \to 1 \text{ for any } d < d_0. \qquad \text{(A.12)}
$$

Case 2. ($d \geq d_0$, overfitted model) By the definition of d_0, if $d \geq d_0$ then $B_d\boldsymbol{\gamma}_d(u) = B_0\boldsymbol{\gamma}_0(u)$. For ease of exposition, we only consider the case that ε_i is independent of (X_i, U_i). If $d > d_0$, following the same argument of Theorem 2 and Lemma 1 we have

$\hat{\theta}_d - \theta_0 = O_p(n^{-1/2})$ and

$$\begin{aligned} B_d\boldsymbol{\gamma}_d(u) - B_0\boldsymbol{\gamma}_0(u) &= \frac{1}{2}\mu_2 B_d\boldsymbol{\gamma}_d''(u)h^2 + B_d\{nf(u)B_d^\top W(u)B_d\}^{-1}B_d^\top \sum_{i=1}^n K_h(U_{iu})X_i\varepsilon_i \\ &\quad + O_p(n^{-1/2} + h^3 + h\delta_n + \delta_n^2). \end{aligned}$$

where $O_p(n^{-1/2} + h^3 + h\delta_n + \delta_n^2)$ are independent of ε_i. Thus, by CLT we have

$$\begin{aligned} \hat{\sigma}_d^2 &= n^{-1}\sum_{j=1}^n \Big(\varepsilon_j - \big(\frac{1}{2}\mu_2 B_d\boldsymbol{\gamma}_d''(U_j)h^2 + B_d\{nf(U_j)B_d^\top W(U_j)B_d\}^{-1}B_d^\top \sum_{i=1}^n K_h(U_{ij})X_i\varepsilon_i\big)^\top X_j\Big)^2 \\ &\quad + O_p\{n^{-1/2}(n^{-1/2} + h^3 + h\delta_n + \delta_n^2)\} \\ &= n^{-1}\sum_{i=1}^n \varepsilon_i^2 - 2n^{-1}\sum_{j=1}^n (B_d\{nf(U_j)B_d^\top W(U_j)B_d\}^{-1}B_d^\top \sum_{i=1}^n K_h(U_{ij})X_i\varepsilon_i)^\top X_j\varepsilon_j \\ &\quad + \frac{1}{4}\mu_2^2 E\{(B_d\boldsymbol{\gamma}_d''(U))^\top W(U)(B_d\boldsymbol{\gamma}_d''(U))\}h^4 + O_p((nh)^{-1} + n^{-1/2}h^2 + n^{-1}). \end{aligned}$$

It is easy to see that

$$Var(n^{-1}\sum_{j=1}^n (B_d\{nf(U_j)B_d^\top W(U_j)B_d\}^{-1}B_d^\top \sum_{i=1}^n K_h(U_{ij})X_i\varepsilon_i)^\top X_j\varepsilon_j) = O(\frac{1}{n^2h}).$$

Note that $B_d\boldsymbol{\gamma}_d''(U)$ are the same for different $d \geq d_0$. Thus, we have

$$\hat{\sigma}_d^2 = \hat{\sigma}_{d_0}^2 + O_p\{(nh)^{-1} + n^{-1/2}h^2\}.$$

It follows that $\log\hat{\sigma}_d^2 - \log\hat{\sigma}_{d_0}^2 = O_p\{(nh)^{-1} + n^{-1/2}h^2\}$. As a consequence, we have

$$\mathrm{BIC}(d) - \mathrm{BIC}(d_0) = (d - d_0)\frac{\log(nh)}{nh} + O_p\{(nh)^{-1} + n^{-1/2}h^2\},$$

where the first term on the right hand side dominates under the condition (C.4). Hence,

$$P\Big\{\text{BIC}(d) > \text{BIC}(d_0)\Big\} \to 1 \text{ for any } d > d_0. \tag{A.13}$$

Equations (A.12) and (A.13) together imply that $P\{\text{BIC}(d) > \text{BIC}(d_0)\} \to 1$. This further implies that $P(\hat{d} = d_0) = 1$. $\square$

Proof of Theorem 5. The proof is an adaption to our case of Zou (2006). We first prove the second part of Theorem 5.

Let $\tilde{\alpha}^{(n)} = \alpha_0 + u/\sqrt{n}$ where $u = (u_1, \ldots, u_S)^\top \in \mathcal{R}^S$, the objective function of our adaptive LASSO can be written as a function of u as

$$\tilde{Q}_n(u) = Q_n(\alpha_0 + \frac{u}{\sqrt{n}}) + \lambda_n \sum_{s=1}^{S} \hat{w}_s |\alpha_{0,s} + \frac{u}{\sqrt{n}}|.$$

Let $\tilde{u} = arg\min_{u\in\mathcal{R}^S} \tilde{Q}_n(u)$ and obviously $\tilde{Q}_n(u)$ is minimized at $\tilde{u}_n = \sqrt{n}(\tilde{\alpha}^{(n)} - \alpha_0)$. Next, write

$$\begin{aligned} D_n(u) &= \tilde{Q}_n(u) - \tilde{Q}_n(0) \\ &= \big(Q_n(\alpha_0 + \frac{u}{\sqrt{n}}) - Q_n(\alpha_0)\big) + \lambda_n \sum_{s=1}^{S} \hat{w}_s \big(|\alpha_{0,s} + \frac{u_s}{\sqrt{n}}| - |\alpha_{0,s}|\big) \\ &\equiv I_{1,n}(u) + I_{2,n}(u), \end{aligned}$$

where $I_{1,n}(u) = Q_n(\alpha_0 + \frac{u}{\sqrt{n}}) - Q_n(\alpha_0)$ is due to the loss function and $I_{2,n}(u)$ is due to the penalty term. From the proof of theorem 2, we know that

$$\begin{aligned} \frac{1}{2n}\frac{\partial^2 Q(\alpha_0)}{\partial\alpha\partial\alpha^\top} &\to \Sigma_0 \text{ in probability}, \\ \frac{1}{2}n^{-\frac{1}{2}}\frac{\partial Q(\alpha_0)}{\partial\alpha} &\xrightarrow{D} Z = N(0, \Sigma_1 + \Sigma_2). \end{aligned}$$

Thus the loss function term

$$I_{1,n}(u) = \frac{1}{\sqrt{n}} u^\top \frac{\partial Q(\alpha_0)}{\partial \alpha} + \frac{1}{2n} u^\top \frac{\partial^2 Q(\alpha_0)}{\partial \alpha \partial \alpha^\top} u(1 + o_p(1)) \xrightarrow{D} 2u^\top Z + u^\top \Sigma_0 u.$$

Now, we consider the limiting behavior of the penalty term $I_{2,n}(u)$. If $s \in \mathcal{A}$, that is $\alpha_{0,s} \neq 0$, then $\hat{w}_s \to |\alpha_{0,s}|^{-\tau}$ in probability and $\sqrt{n}(|\alpha_{0,s} + u_s/\sqrt{n}| - |\alpha_{0,s}|) \to u_s \mathrm{sgn}(\alpha_{0,s})$. Since $\lambda_n/\sqrt{n} \to 0$, we have

$$\frac{\lambda_n}{\sqrt{n}} \hat{w}_s \sqrt{n}(|\alpha_{0,s} + u_s/\sqrt{n}| - |\alpha_{0,s}|) \to 0.$$

If $s \notin \mathcal{A}$ then $\sqrt{n}(|\alpha_{0,s} + u_s/\sqrt{n}| - |\alpha_{0,s}|) = |u_s|$. Since $\sqrt{n}\hat{\alpha}_n = O_p(1)$ and $\lambda_n n^{\frac{\tau-1}{2}} \to \infty$, we have $\frac{\lambda_n}{\sqrt{n}} \hat{w}_s = \lambda_n n^{\frac{\tau-1}{2}} |\sqrt{n}\hat{\alpha}_s^{(n)}|^{-\tau} \to \infty$ in probability. It follows that

$$D_n(u) \Rightarrow D(u) = \begin{cases} 2(u_{\mathcal{A}})^\top Z_{\mathcal{A}} + (u_{\mathcal{A}})^\top (\Sigma_0)_{\mathcal{A}} (u_{\mathcal{A}}), & \text{if } u_s = 0, \forall s \notin \mathcal{A} \\ \infty, & \text{otherwise ,} \end{cases}$$

where $u_{\mathcal{A}}$ and $Z_{\mathcal{A}}$ are the j-th ($j \in \mathcal{A}^c$) elements deleted from u and Z respectively. Note that $D_n(u)$ is convex, and the unique minimum of $D(u)$ is

$$u_{min} = \begin{pmatrix} -\Big((\Sigma_0)_{\mathcal{A}}\Big)^{-1} Z_{\mathcal{A}} \\ 0 \end{pmatrix},$$

where 0 denotes a vector of zeros. Following the epi-convergence result of Geyer (1994), we have

$$\tilde{\alpha}_{\mathcal{A}}^{(n)} \xrightarrow{D} \Big((\Sigma_0)_{\mathcal{A}}\Big)^{-1} Z_{\mathcal{A}} = N\left(0, \Big((\Sigma_0)_{\mathcal{A}}\Big)^{-1} (\Sigma_1 + \Sigma_2)_{\mathcal{A}} \Big((\Sigma_0)_{\mathcal{A}}\Big)^{-1}\right) \tag{A.14}$$

and $\tilde{\alpha}_{\mathcal{A}^c}^{(n)} \to 0$. Now we prove the consistency part. It suffices to show that $\forall s \in \mathcal{A}^c$,

$P(s \in \mathcal{A}_n) \to 0$. By the KKT optimality conditions,

$$\frac{1}{\sqrt{n}} \frac{\partial Q_n(\tilde{\alpha}^{(n)})}{\partial \alpha_s} + \frac{\lambda_n}{\sqrt{n}} \hat{w}_s \mathrm{sgn}(\tilde{\alpha}_s^{(n)}) = 0.$$

If $s \in \mathcal{A}^c$, then

$$\frac{\lambda_n}{\sqrt{n}} \hat{w}_s = \lambda_n n^{\frac{\tau-1}{2}} |\sqrt{n} \hat{\alpha}_s^{(n)}|^{-\tau} \to \infty$$

in probability, whereas

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{\partial Q_n(\tilde{\alpha}^{(n)})}{\partial \alpha_s} &= \frac{1}{\sqrt{n}} \frac{\partial Q_n(\tilde{\alpha}^{(n)})}{\partial \alpha_s} + \frac{1}{n} \frac{\partial^2 Q_n(\tilde{\alpha}^{(n)})}{\partial \alpha_s^2} \sqrt{n}(\tilde{\alpha}_s^{(n)} - \alpha_{0,s})(1 + o_p(1)) \\ &\stackrel{D}{\to} \text{some normal distribution} \end{aligned}$$

by (A.14) and Slutsky's theorem. Thus, for $s \in \mathcal{A}^c$,

$$P(s \in \mathcal{A}^{(n)}) \leq P\Big(|\frac{1}{\sqrt{n}} \frac{\partial Q_n(\tilde{\alpha}^{(n)})}{\partial \alpha_s}| = \frac{\lambda_n}{\sqrt{n}} \hat{w}_s \Big) \to 0.$$

We have completed the proof. $\square$

REFERENCES

Fan, J. and Huang, T. (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models, *Bernoulli*, **11**, 1031-1057.

— (2000) Simultaneous confidence bands and hypotheses testing in varying-coefficient models, *Scandinavian Journal of Statistics*, **27**, 715-731.

Geyer, C. (1994) On the asymptotics of constrained M-estimation, *Annals of Statistics*, **22**, 1993-2010.

Mack, Y. P. and Silverman, B. W. (1982) Weak and strong uniform consistency of kernel regression estimates, *Z. Wahrasch. Verw. Gebiete*, **61**, 405-415.

Zhang, W., Lee, S. Y., and Song, X. (2002) Local polynomial fitting in semivarying coefficient model, *Journal of Multivariate Analysis*, **82**, 166-188.

Zou, H. (2006) The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association,* **101**, 1418-1429.