Supplementary material

Penalized robust estimating equation and variable selection in a partially linear single-index varying-coefficient model

Gaorong Li[1], Liugen Xue[2] and Riquan Zhang[3]

[1]*School of Statistics, Beijing Normal University, Beijing 100875, China*

[2]*School of Mathematics and Statistics, Henan University, Kaifeng 475004, China*

[3]*School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai 201620, China*

Appendix B. Proofs of Lemma A.1

The following three lemmas are useful for proving Lemmas A.1 given in the Appendix. Lemma B.1 is the lemma A.1 of Wang et al. (2010), and it can also be used when the variable t is removed. The proof of Lemma B.2 is similar to the proofs of Lemmas 1 and 2 in Zhu and Xue (2006). Lemma B.3 is the Theorem 1 of Xue (2023). Using Theorem 2 of Einmahl and Mason (2005), we can derive the first equation of Lemma B.4. The proof of the second equation of Lemma B.4 is similar to the proofs of Lemma 3 in Zhu and Xue (2006). Therefore, we omit their proofs. Let c represents a positive constant that does not depend on n, which may take a different value for each appearance.

Lemma B.1. Assume that $\{\xi_i(t,\beta), 1 \le i \le n\}$ are random variables, and satisfy the following two conditions:

$$\frac{1}{n}\sum_{i=1}^{n}|\xi_i(t,\beta)-\xi_i(t_0,\beta_0)| \le cn^a(|t-t_0|+\|\beta-\beta_0\|) \tag{B.1}$$

for some constants $c>0$, $a\ge 0$, t_0 and β_0;

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(t,\beta)\right|>\varepsilon_n\right)\le\frac{1}{2} \tag{B.2}$$

for $\beta\in\mathcal{B}_n$ and $\varepsilon_n>0$ depend only on n. Then

$$P\left(\sup_{(t,\beta)\in\mathcal{T}_\beta\times\mathcal{B}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(t,\beta)\right|>\varepsilon_n\right)\le cn^{2pa}\varepsilon_n^{-2p}E\left\{\sup_{(t,\beta)\in\mathcal{T}_\beta\times\mathcal{B}_n}\exp\left(\frac{-n^2\varepsilon_n^2/128}{\sum_{i=1}^{n}\xi_i^2(t,\beta)}\right)\wedge 1\right\},$$

[0]Corresponding author: Liugen Xue

[0]E-mail address: lgxue@bjut.edu.cn

where $\mathcal{B}_n = \{\beta \,|\, \beta \in R^p, \|\beta - \beta_0\| \le c_1 n^{-1/2}\}$ for a positive constant c_1, and $\mathcal{T}_\beta$ is defined in condition (C1),

Lemma B.2. Assume that conditions (C1), (C2) and (C5)–(C7) hold. Then, we have, uniformly over $1 \le i \le n$,

$$E\left\{I_{\mathcal{T}_\beta}(\beta^T X_i)\left\|\sum_{j=1}^n W_{nj}(\beta^T X_i;\beta)\boldsymbol{g}^T(\beta^T X_j)Z_i - \boldsymbol{g}(\beta^T X_i)\right\|^2\right\} = O(n^{-4/5}),$$

$$E\left\{I_{\mathcal{T}_\beta}(\beta^T X_i)\left\|\sum_{j=1}^n \widetilde{W}_{nj}(\beta^T X_i;\beta)\boldsymbol{g}^T(\beta^T X_j)Z_i - \boldsymbol{g}'(\beta^T X_i)\right\|^2\right\} = O(n^{-2/5}),$$

$$E\left\{I_{\mathcal{T}_\beta}(\beta^T X_i)\sum_{j=1}^n \|W_{nj}(\beta^T X_i;\beta)\|^2\right\} = O(n^{-4/5})$$

and

$$E\left\{I_{\mathcal{T}_\beta}(\beta^T X_i)\sum_{j=1}^n \|\widetilde{W}_{nj}(\beta^T X_i;\beta)\|^2\right\} = O(n^{-2/5}),$$

where $W_{nj}(\cdot;\beta)$, $\widetilde{W}_{nj}(\cdot;\beta)$ and $I_{\mathcal{T}_\beta}(\cdot)$ are defined in (2.3), (2.4) and (2.7), respectively.

Lemma B.3. Assume that conditions (C1)–(C3) and (C5)–(C7) in Section 3 hold. Then

$$\sup_{t\in\mathcal{T}_\beta,(\beta,\theta)\in\mathcal{B}_n^*\times\Theta_n} \|\hat{\boldsymbol{g}}(t;\beta,\theta) - \boldsymbol{g}(t)\| = O_P\left(n^{-2/5}\sqrt{\log n}\right)$$

and

$$\sup_{t\in\mathcal{T}_\beta,(\beta,\theta)\in\mathcal{B}_n^*\times\Theta_n} \|\hat{\boldsymbol{g}}'(t;\beta,\theta) - \boldsymbol{g}'(t)\| = O_P\left(n^{-1/5}\sqrt{\log n}\right),$$

where $\mathcal{B}_n^* = \{\beta| \,\|\beta - \beta_0\| \le c_1 n^{-1/2}\}$ for a positive constant c_1, $\hat{\boldsymbol{g}}(t;\beta,\theta)$, $\hat{\boldsymbol{g}}'(t;\beta,\theta)$, $\mathcal{T}_\beta$ and Θ_n are defined in (2.1), (2.7) and (2.10), respectively, and $\boldsymbol{g}'(t) = (\dot{g}_1(t),\ldots,\dot{g}_q(t))^T$.

Lemma B.4. Assume that conditions (C4)–(C7) hold. Then

$$\sup_{(t,z)\in\mathcal{T}_\beta\times\mathcal{Z},\beta\in\mathcal{B}_n^*} \|\hat{m}_l(t,z;\beta) - m_l(t,z)\| = O_P\left(n^{-2/5}\sqrt{\log n}\right), \quad l = 1,2$$

and uniformly over $1 \le i \le n$,

$$E\left[I_{\mathcal{T}_\beta}(\beta^T X_i)I_{\mathcal{Z}}(Z_i)|\hat{m}_{lk}(\beta^T X_i, Z_i;\beta) - m_{lk}(\beta^T X_i, Z_i)|^2\right] = O(n^{-4/5}), \tag{B.3}$$

where $\hat{m}_{lk}$ and m_{lk} are the kth component of $\hat{m}_l$ and m_l respectively, $\hat{m}_1$, $\hat{m}_2$, $\mathcal{T}_\beta$, $\mathcal{Z}$ and $\mathcal{B}_n^*$ are defined in (2.4), (2.5), (2.7) and Lemma B.3, respectively.

Proof of Lemma A.1. Note that the second derivative of $\psi(t)$ is 0. Therefore, by using the Taylor's formula, it can be obtained that for any point (β,θ) in the neighborhood of (β_0,θ_0),

$$\begin{aligned}&\psi(Y_i - \theta^T U_i - \hat{\boldsymbol{g}}^T(\beta^T X_i;\beta,\theta)Z_i) - \psi(\varepsilon_i)\\ &= \psi'(\varepsilon_i)\left[(\theta_0-\theta)^T U_i + \{\boldsymbol{g}(\beta_0^T X_i) - \hat{\boldsymbol{g}}(\beta^T X_i;\beta,\theta)\}^T Z_i\right].\end{aligned} \tag{B.4}$$

Using (B.4), Lemma B.3, the first equation of Lemma B.4, Taylor's formula and the law of large numbers, we can obtain that uniformly for $\beta^{(r)} \in \mathcal{B}_n$ and $\theta \in \Theta_n$,

$$\widehat{Q}(\beta^{(r)}, \theta) = Q_n(\beta^{(r)}, \theta) + H_1(\beta^{(r)}) + H_2(\beta^{(r)}) + H_3(\beta^{(r)}) + o_P(n^{-1/2}), \tag{B.5}$$

where $H_l(\beta^{(r)}) = (H_{l1}^T(\beta^{(r)}), H_{l2}^T(\beta^{(r)}))^T$, $l = 1, 2$, $H_3(\beta^{(r)}) = (H_{31}^T(\beta^{(r)}), 0)^T$,

$$\begin{aligned}
H_{11}(\beta^{(r)}) &= \frac{1}{n}\sum_{i=1}^n I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) J_{\beta^{(r)}}^T \{m_1(\beta^T X_i, Z_i) - \hat{m}_1(\beta^T X_i, Z_i; \beta)\} \boldsymbol{g}'^T(\beta^T X_i) Z_i \psi(\varepsilon_i), \\
H_{12}(\beta^{(r)}) &= \frac{1}{n}\sum_{i=1}^n I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) \{m_2(\beta^T X_i, Z_i) - \hat{m}_2(\beta^T X_i, Z_i; \beta)\} \psi(\varepsilon_i), \\
H_{21}(\beta^{(r)}) &= \frac{1}{n}\sum_{i=1}^n I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) \{\boldsymbol{g}(\beta^T X_i) - \hat{\boldsymbol{g}}(\beta^T X_i; \beta, \theta)\}^T Z_i \boldsymbol{g}'^T(\beta^T X_i) Z_i J_{\beta^{(r)}}^T \widetilde{X}_i \psi'(\varepsilon_i), \\
H_{22}(\beta^{(r)}) &= \frac{1}{n}\sum_{i=1}^n I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) \{\boldsymbol{g}(\beta^T X_i) - \hat{\boldsymbol{g}}(\beta^T X_i; \beta, \theta)\}^T Z_i \widetilde{U}_i \psi'(\varepsilon_i), \\
H_{31}(\beta^{(r)}) &= \frac{1}{n}\sum_{i=1}^n I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) \{\boldsymbol{g}'(\beta^T X_i) - \hat{\boldsymbol{g}}'(\beta^T X_i; \beta^{(r)}, \hat{\theta})\}^T Z_i J_{\beta^{(r)}}^T \widetilde{X}_i \psi(\varepsilon_i),
\end{aligned}$$

$\widetilde{X}_i = X_i - m_1(\beta^T X_i, Z_i)$ and $\widetilde{U}_i = U_i - m_2(\beta^T X_i, Z_i)$.

We first deal with $H_{11}(\beta^{(r)})$. Let $m_{1k}(\beta^T X_i, Z_i)$ and $\hat{m}_{1k}(\beta^T X_i, Z_i; \beta)$ represent the kth component of $m_1(\beta^T X_i, Z_i)$ and $\hat{m}_1(\beta^T X_i, Z_i; \beta)$, and let

$$\xi_{ni}(\beta^{(r)}) = \sqrt{n} I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) \{m_{1k}(\beta^T X_i, Z_i) - \hat{m}_{1k}(\beta^T X_i, Z_i; \beta)\} \boldsymbol{g}'^T(\beta^T X_i) Z_i \psi(\varepsilon_i).$$

Using Lemma B.1, we have to verify (B.1) and (B.2). A simple calculation yields (B.1), so we now verify (B.2). From the second equation of Lemma B.4, conditions (C2), (C5) and (C7), and a simple calculation, we have

$$\begin{aligned}
&E\left\{\frac{1}{n}\sum_{i=1}^n \xi_{ni}(\beta^{(r)})\right\}^2 \\
&\le 2n^{-1}\sum_{i=1}^n E[I_{\mathcal{T}_\beta}(\beta^T X_i) I_{\mathcal{Z}}(Z_i) \{m_{1k}(\beta^T X_i, Z_i) - \hat{m}_{1k}(\beta^T X_i, Z_i; \beta)\}^2 \\
&\quad \times \{\boldsymbol{g}'^T(\beta^T X_i) Z_i\}^2 E\{\varepsilon_i^2 | X_i, Z_i\}] \le c n^{-4/5}.
\end{aligned}$$

Hence, by Chebyshev's inequality, for arbitrary $\epsilon > 0$, we have

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n \xi_{ni}(\beta^{(r)})\right| > \epsilon\right\} \le c n^{-4/5} < \frac{1}{2}$$

when n large enough. This verifies that (B.2) of Lemma B.1 is satisfied. By the second equation of Lemma B.4 and conditions (C2), (C5) and (C7), when n large enough, we have

$$n^{-2}\sum_{i=1}^n E\{\xi_{ni}^2(\beta^{(r)})\} = O(n^{-4/5}).$$

This implies that $n^{-2}\sum_{i=1}^{n}\xi_{ni}^2(\beta^{(r)}) = O_P(n^{-4/5})$. Thus, from Lemma B.1 we have

$$P\left\{\sup_{\beta^{(r)}\in\mathcal{B}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{ni}(\beta^{(r)})\right| > \epsilon\right\} \le cn^{2pa}\exp\left(-cn^{4/5}\right) \longrightarrow 0, \quad \forall\epsilon > 0.$$

This proves that

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}\|H_{11}(\beta^{(r)})\| = o_P(n^{-1/2}). \tag{B.6}$$

Similarly, we can prove that

$$\sup_{\beta^{(r)}\in\mathcal{B}_n}\|H_{12}(\beta^{(r)})\| = o_P(n^{-1/2}). \tag{B.7}$$

We now deal with $H_{21}(\beta^{(r)})$. Let $\widetilde{X}_{ik}$ denotes the kth component of $\widetilde{X}_i$, and let

$$\tilde{\xi}_{ni}(\beta^{(r)}) = \sqrt{n}I_{\mathcal{T}_\beta}(\beta^T X_i)I_{\mathcal{Z}}(Z_i)\{\boldsymbol{g}(\beta^T X_i) - \hat{\boldsymbol{g}}(\beta^T X_i;\beta,\theta_0)\}^T Z_i\boldsymbol{g}'^T(\beta^T X_i)Z_i\widetilde{X}_{ik}\psi'(\varepsilon_i).$$

We yet use Lemma B.1. This need to verify that (B.1) and (B.2) of Lemma B.1 are satisfied. A simple calculation yields (B.1). We now verify (B.2). From Lemma B.2 and conditions (C2), (C5) and (C7), we have

$$\begin{aligned}
&E\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_{ni}(\beta^{(r)})\right\}^2\\
&\le 2n^{-1}\sum_{i=1}^{n}E\Bigg[\Bigg|I_{\mathcal{T}_\beta}(\beta^T X_i)\Big\{\boldsymbol{g}(\beta^T X_i) - \sum_{j=1}^{n}W_{nj}(\beta^T X_i;\beta)\boldsymbol{g}^T(\beta^T X_j)Z_j\Big\}^T Z_i\Bigg|^2\\
&\quad\times\{\boldsymbol{g}'^T(\beta^T X_i)Z_i\}^2 E\{\widetilde{X}_{ik}^2\big|\beta^T X_i, Z_i\}\Bigg] + o(1)\\
&\quad + 2n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E[I_{\mathcal{T}_\beta}(\beta^T X_i)\{\boldsymbol{g}'^T(\beta^T X_i)Z_i\}^2 W_{nj}^2(\beta^T X_i;\beta)\widetilde{X}_{ik}^2\varepsilon_j^2]\\
&\le cn^{-4/5} + o(1) + cn^{-1}\sum_{i=1}^{n}E\Big\{I_{\mathcal{T}_\beta}(\beta^T X_i)\sum_{j=1}^{n}W_{nj}^2(\beta^T X_i;\beta)\Big\}\\
&\le cn^{-4/5} + o(1) \longrightarrow 0.
\end{aligned}$$

Thereupon, we can obtain that

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_{ni}(\beta^{(r)})\right| > \epsilon\right\} \le cn^{-4/5} + o(1) < \frac{1}{2}$$

when n large enough. This verifies that (B.2) of Lemma B.1 is satisfied. By Lemma B.3 and conditions (C2), (C5) and (C7), we have

$$\frac{1}{n^2}\sum_{i=1}^{n}\tilde{\xi}_{ni}^2(\beta^{(r)}) = O_P(1)\sup_{t\in\mathcal{T}_\beta,(\beta^{(r)},\theta)\in\mathcal{B}_n\times\Phi_n}\|\boldsymbol{g}(t) - \hat{\boldsymbol{g}}(t;\beta,\theta)\|^2 = O_P(n^{-4/5}\log n).$$

Therefore, from Lemma B.2 we can obtain that

$$P\left\{\sup_{\beta^{(r)}\in\mathcal{B}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{\xi}_{ni}(\beta^{(r)})\right| > \epsilon\right\} \le cn^{2pa}\exp\left(-cn^{4/5}\right) \longrightarrow 0.$$

This implies that

$$\sup_{\beta^{(r)}\in\mathcal{B}_n} \|H_{21}(\beta^{(r)})\| = o_P(n^{-1/2}) \tag{B.8}$$

Similarly, we can proves that

$$\sup_{\beta^{(r)}\in\mathcal{B}_n} \|H_{22}(\beta^{(r)})\| = o_P(n^{-1/2}) \tag{B.9}$$

and

$$\sup_{\beta^{(r)}\in\mathcal{B}_n} \|H_{31}(\beta^{(r)})\| = o_P(n^{-1/2}) \tag{B.10}$$

This, together with (B.5)–(B.10), completes the proof of Lemma A.1.

References

Einmahl, U., and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.*, **33**, 1380–1403.

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

Xue, L. G. (2023). Two-stage estimation and bias-corrected empirical likelihood in a partially linear single-index varying-coefficient model. *J. Roy. Statist. Soc.* B **85**, 1299–1325.

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

Appendix C. Simulation results of Example 1

Table C. Simulation results of Example 1. Estimates of the regression parameters and variable selection under different loss functions and penalty methods when $n = 200$.

		$\hat{\beta}^T\beta_0$ and average number 0's					$\hat{\theta}$ and average number 0's		
Loss	Method	Mean	Bias	SD	C	I	GMSE	C	I
PSL	FLEI	0.9978	0.0022	0.0013	4.9540	0.0020	0.1051	2.9300	0.0080
	SCAD	0.9964	0.0026	0.0017	4.8580	0.0040	0.1185	2.8700	0.0100
	ALASSO	0.9963	0.0037	0.0016	4.8600	0.0040	0.1080	2.8620	0.0100
	Oracle	0.9974	0.0021	0.0011	5	0	0.1047	3	0
HL	FLEI	0.9728	0.0272	0.1298	4.7720	0.0040	0.2678	2.6740	0.0100
	SCAD	0.9663	0.0337	0.1551	4.7240	0.0060	0.2801	2.6680	0.0120
	ALASSO	0.9568	0.0432	0.1658	4.6860	0.0060	0.2772	2.6680	0.0120
	Oracle	0.9723	0.0269	0.1293	5	0	0.2672	3	0
SL	FLEI	0.9649	0.0351	0.1399	4.7200	0.0060	0.2721	2.6300	0.0110
	SCAD	0.9627	0.0373	0.1584	4.6400	0.0070	0.2810	2.6220	0.0130
	ALASSO	0.9538	0.0462	0.1602	4.6280	0.0070	0.2760	2.6220	0.0130
	Oracle	0.9643	0.0346	0.1392	5	0	0.2715	3	0