Estimation and Variable Selection for Generalized Additive Partial Linear Models

Lily Wang, Xiang Liu, Hua Liang, and Raymond J. Carroll

Supplementary Material

B.1. *Proof of Lemma A.3*

Proof. We first derive the lower and upper bound of the eigenvalue of matrix $\mathbf{W}_n$. For any vectors $\boldsymbol{\omega}_1 = \{\omega_{j,k}, j = 1, ..., N_n, k = 1, ..., d_1\} \in R^{N_n \times d_1}$ and $\boldsymbol{\omega}_2 \in R^{d_2}$, let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^{\mathrm{T}}, \boldsymbol{\omega}_2^{\mathrm{T}})^{\mathrm{T}}$, then one has

$$\boldsymbol{\omega}^{\mathrm{T}}\mathbf{W}_n\boldsymbol{\omega} = \boldsymbol{\omega}_1^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{B}_i^{\otimes 2}\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i^{\otimes 2}\boldsymbol{\omega}_2 + 2\boldsymbol{\omega}_1^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{B}_i\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\omega}_2.$$

Lemma 1 of Stone (1985) provides a constant $c > 0$ such that

$$\left\|\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_2^2 \geq c\sum_{k=1}^{d_1}\left\|\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_2^2.$$

According to Theorem 5.4.2 of DeVore and Lorentz (1993), Condition (C3) and the definition of $B_{j,k}$ in (5), there exist constants $C'_k > c'_k > 0$ such that for any $k = 1, ..., d_1$

$$c'_k\sum_{j=1}^{N_n}\omega_{j,k}^2 \leq \left\|\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_2^2 \leq C'_k\sum_{j=1}^{N_n}\omega_{j,k}^2.$$

Thus there exist constants $C_0 > c_0 > 0$ such that

$$c_0\left\|\boldsymbol{\omega}_1\right\|^2 \leq \left\|\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_2^2 \leq C_0\left\|\boldsymbol{\omega}_1\right\|^2.$$

By Lemma A.8 in Wang and Yang (2007), we have

$$\text{(S.1)}\qquad A_n \equiv \sup_{g_1,g_2\in\mathcal{G}_n}\left|\frac{\langle g_1,g_2\rangle_n - \langle g_1,g_2\rangle}{\|g_1\|_2\|g_2\|_2}\right| = O\left\{(\log(n)N_n/n)^{1/2}\right\},\ a.s.$$

It is clear to see that

$$(1-A_n)\left\|\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_2^2 \leq \boldsymbol{\omega}_1^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{B}_i^{\otimes 2}\boldsymbol{\omega}_1$$
$$= \left\|\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_{2,n}^2 \leq (1+A_n)\left\|\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\omega_{j,k}B_{j,k}\right\|_2^2.$$

Therefore,

$$c\left\|\boldsymbol{\omega}_1\right\|^2 \leq \boldsymbol{\omega}_1^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{B}_i^{\otimes 2}\boldsymbol{\omega}_1 \leq C\left\|\boldsymbol{\omega}_1\right\|^2, a.s.$$

Next,

$$\begin{aligned}
\boldsymbol{\omega}_2^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i^{\otimes 2}\boldsymbol{\omega}_2 &= \boldsymbol{\omega}_2^{\mathrm{T}} E\left(\mathbf{Z}^{\otimes 2}\right)\boldsymbol{\omega}_2 + \boldsymbol{\omega}_2^{\mathrm{T}}\left[n^{-1}\sum_{i=1}^{n}\left\{\mathbf{Z}_i^{\otimes 2} - E\left(\mathbf{Z}^{\otimes 2}\right)\right\}\right]\boldsymbol{\omega}_2 \\
&= \boldsymbol{\omega}_2^{\mathrm{T}} E\left(\mathbf{Z}^{\otimes 2}\right)\boldsymbol{\omega}_2 + \left\|\boldsymbol{\omega}_2\right\|^2 o_{a.s.}(1),
\end{aligned}$$

and according to Condition (C4), $c\left\|\boldsymbol{\omega}_2\right\|^2 \leq \boldsymbol{\omega}_2^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i^{\otimes 2}\boldsymbol{\omega}_2 \leq C\left\|\boldsymbol{\omega}_2\right\|^2$, a.s. Then $\left|\boldsymbol{\omega}_1^{\mathrm{T}} n^{-1}\sum_{i=1}^{n}\mathbf{B}_i\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\omega}_2\right| = o(\left\|\boldsymbol{\omega}_1\right\|\left\|\boldsymbol{\omega}_2\right\|)$, a.s. Thus

$$c\left\|\boldsymbol{\omega}\right\|^2 \leq \boldsymbol{\omega}^{\mathrm{T}}\mathbf{W}_n\boldsymbol{\omega} \leq C\left\|\boldsymbol{\omega}\right\|^2, a.s. \tag{S.2}$$

Let $\lambda_{\max}\left(\mathbf{W}_n\right)$ and $\lambda_{\min}\left(\mathbf{W}_n\right)$ be the maximum and minimum eigenvalues of $\mathbf{W}_n$. Algebra and (S.2) show that $\left\|\mathbf{W}_n\right\|_2 = \lambda_{\max}\left(\mathbf{W}_n\right) \leq C$ and $\left\|\mathbf{W}_n^{-1}\right\|_2 = \lambda_{\min}^{-1}\left(\mathbf{W}_n\right) \leq c^{-1}$, a.s. ■

B.2. *Proof of Lemma A.4*

Proof. Let $\widetilde{\boldsymbol{\delta}} = \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Since $\widetilde{\boldsymbol{\beta}}$ maximizes the quasi-likelihood function $n^{-1}\sum_{i=1}^{n} Q[g^{-1}\{\widetilde{\eta}(\mathbf{X}_i) + \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}\}, Y_i]$, $\widetilde{\boldsymbol{\delta}}$ maximizes

$$\widetilde{l}_n\left(\boldsymbol{\delta}\right) = \sum_{i=1}^{n}\left[Q\left\{g^{-1}\left(\widetilde{m}_{0i} + n^{-1/2}\boldsymbol{\delta}^{\mathrm{T}}\mathbf{Z}_i\right), Y_i\right\} - Q\left\{g^{-1}\left(\widetilde{m}_{0i}\right), Y_i\right\}\right].$$

By Taylor expansion, $\widetilde{l}_n\left(\boldsymbol{\delta}\right) = n^{-1/2}\sum_{i=1}^{n} q_1\left(\widetilde{m}_{0i}, Y_i\right)\boldsymbol{\delta}^{\mathrm{T}}\mathbf{Z}_i + \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{A}_n\boldsymbol{\delta}$, where the random matrix $\mathbf{A}_n = n^{-1}\sum_{i=1}^{n}\left[Y_i\rho_1'\left(\widetilde{m}_{0i} + \zeta_{ni}, Y_i\right) - \rho_3\left(\widetilde{m}_{0i} + \zeta_{ni}'\right)\right]\mathbf{Z}_i^{\otimes 2}$, with ζ_{ni} and ζ_{ni}' between 0 and $n^{-1/2}\boldsymbol{\delta}^{\mathrm{T}}\mathbf{Z}_i$, independent of Y_i, and $\rho_3\left(m\right) = g^{-1}\left(m\right)\rho_1'\left(m\right) - \rho_2\left(m\right)$. From the structure of $\mathbf{A}_n$ and Carroll et al. (1997),

$$\left\|\mathbf{A}_n - n^{-1}\sum_{i=1}^{n} q_2\left(m_{0i}, Y_i\right)\mathbf{Z}_i^{\otimes 2}\right\| = o_P\left(1\right),$$

and

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n} q_2\left(m_i, Y_i\right)\mathbf{Z}_i^{\otimes 2} &= E\left[q_2\left\{m_0\left(\mathbf{T}\right), Y\right\}\mathbf{Z}^{\otimes 2}\right] + o_P\left(1\right) \\
&= -E\left[\rho_2\left\{m_0\left(\mathbf{T}\right)\right\}\mathbf{Z}^{\otimes 2}\right] + o_P\left(1\right).
\end{aligned}$$

Thus $\mathbf{A}_n = -E\left[\rho_2\left\{m_0\left(\mathbf{T}\right)\right\}\mathbf{Z}^{\otimes 2}\right] + o_P\left(1\right) = -\mathbf{A} + o_P\left(1\right)$. As in Carroll et al. (1997),

$$
\begin{aligned}
n^{-1/2}\sum_{i=1}^{n} q_1\left(\widetilde{m}_{0i}, Y_i\right)\mathbf{Z}_i &= n^{-1/2}\sum_{i=1}^{n} q_1\left(m_{0i}, Y_i\right)\mathbf{Z}_i \\
&\quad + n^{-1/2}\sum_{i=1}^{n} q_2\left(m_{0i}, Y_i\right)\left(\widetilde{\eta} - \eta_0\right)\left(\mathbf{X}_i\right)\mathbf{Z}_i + O_P\left(n^{1/2}\left\|\widetilde{\eta} - \eta_0\right\|_\infty^2\right).
\end{aligned}
$$

By Condition (C5) and (A.2), the second term on the right hand side of the above is

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\rho_1'\left(m_{0i}\right)\mathbf{Z}_i\left\{\left(\widetilde{\eta} - \eta_0\right)\left(\mathbf{X}_i\right)\right\}\varepsilon_i + n^{-1/2}\sum_{i=1}^{n}\rho_2\left(m_{0i}\right)\mathbf{Z}_i\left\{\left(\widetilde{\eta} - \eta_0\right)\left(\mathbf{X}_i\right)\right\} \\
&\quad \leq n^{-1/2}\sum_{i=1}^{n}\left|\rho_1'\left(m_{0i}\right)\mathbf{Z}_i\varepsilon_i\right|\left\|\widetilde{\eta} - \eta_0\right\|_\infty + O_P\left(n^{1/2}N_n^{-p}\right) = o_P\left(1\right).
\end{aligned}
$$

By the convexity lemma of Pollard (1991), $\widetilde{\boldsymbol{\delta}} = \mathbf{A}^{-1}n^{-1/2}\sum_{i=1}^{n} q_1\left(m_{0i}\right)\mathbf{Z}_i + o_P\left(1\right)$, and $\operatorname{var}\left(q_1\left\{m_0\left(\mathbf{T}\right)\right\}\mathbf{Z}\right) = E\left[q_1^2\left\{m_0\left(\mathbf{T}\right)\right\}\mathbf{Z}^{\otimes 2}\right] = \boldsymbol{\Sigma}_1$, from which the result follows. ■

B.3. *Proof of Lemma A.5*

Note that

$$
\left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left.\frac{\partial^2\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}}\left(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\right),
$$

with $\overline{\boldsymbol{\theta}} = t\widehat{\boldsymbol{\theta}} + (1-t)\widetilde{\boldsymbol{\theta}}$ ($t \in [0,1]$). So

$$
\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} = -\left(\left.\frac{\partial^2\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}}\right)^{-1}\left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}.
$$

Next write

$$
\left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = \left.\left\{\left(\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\gamma}}\right)^{\mathrm{T}}, \left(\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\beta}}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} = n^{-1}\sum_{i=1}^{n} q_1\left(\widetilde{m}_i, Y_i\right)\mathbf{D}_i^{\mathrm{T}},
$$

where $\widetilde{m}_i$ is given in (A.4) and

$$
\begin{aligned}
\left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\gamma}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} &= n^{-1}\sum_{i=1}^{n} q_1\left(m_{0i}, Y_i\right)\mathbf{B}_i + n^{-1}\sum_{i=1}^{n} q_2\left(\xi_i, Y_i\right)\left\{\left(\widetilde{\eta} - \eta_0\right)\left(\mathbf{X}_i\right)\right\}\mathbf{B}_i \\
&\quad + n^{-1}\sum_{i=1}^{n} q_2\left(\xi_i, Y_i\right)\mathbf{Z}_i^{\mathrm{T}}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\mathbf{B}_i,
\end{aligned}
$$

$$
\begin{aligned}
\left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} &= \frac{1}{n}\sum_{i=1}^{n} q_1\left(m_{0i}, Y_i\right)\mathbf{Z}_i + n^{-1}\sum_{i=1}^{n} q_2\left(\xi_i, Y_i\right)\left\{\left(\widetilde{\eta} - \eta_0\right)\left(\mathbf{X}_i\right)\right\}\mathbf{Z}_i \\
&\quad + n^{-1}\sum_{i=1}^{n} q_2\left(\xi_i, Y_i\right)\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)^{\mathrm{T}}\mathbf{Z}_i^{\otimes 2}
\end{aligned}
$$

with ξ_i between m_{0i} and $\widetilde{m}_i$. Observing that

$$\left\| n^{-1}\sum_{i=1}^{n} q_1\left(m_{0i},Y_i\right)\mathbf{B}_i \right\| = \left[\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\left\{\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(m_{0i}\right)B_{j,k}\left(X_{ik}\right)\varepsilon_i\right\}^2\right]^{1/2}$$

and

$$E\left[\sum_{k=1}^{d_1}\sum_{j=1}^{N_n}\left\{\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(m_{0i}\right)B_{j,k}\left(X_{ik}\right)\varepsilon_i\right\}^2\right] \leq CN_n/n,$$

we have $\left\|n^{-1}\sum_{i=1}^{n} q_1\left(m_{0i},Y_i\right)\mathbf{B}_i\right\| = O_P\{(N_n/n)^{1/2}\}$. In addition, (A.2) and Condition (C2) imply that

$$\left\| n^{-1}\sum_{i=1}^{n} q_2\left(\xi_i,Y_i\right)\left\{\left(\widetilde{\eta}-\eta_0\right)\left(\mathbf{X}_i\right)\right\}\mathbf{B}_i \right\| = O_P\left(N_n^{1/2-p}\right),$$

and

$$\left\| n^{-1}\sum_{i=1}^{n} q_2\left(\xi_i,Y_i\right)\mathbf{Z}_i^{\mathrm{T}}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\mathbf{B}_i \right\| = O_P\{(N_n/n)^{1/2}\}.$$

Therefore,

$$\left\| \left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\gamma}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right\| = O_P\{N_n^{1/2-p}+(N_n/n)^{1/2}\}.$$

Similarly, $\left\|n^{-1}\sum_{i=1}^{n}\rho_1\left(m_{0i}\right)\mathbf{Z}_i\varepsilon_i\right\| = O_P\left(n^{-1/2}\right)$ and

$$\left\| \left.\frac{\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right\| = n^{-1}\sum_{i=1}^{n}\left\|q_1\left(\widetilde{m}_i,Y_i\right)\mathbf{Z}_i\right\| = O_P\left(N_n^{-p}+n^{-1/2}\right).$$

Thus

$$\left\| \left.\frac{\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right\| = O_P\{N_n^{1/2-p}+(N_n/n)^{-1/2}\}.$$

Next let $\bar{m}_i = \bar{m}\left(\mathbf{T}_i\right) = \overline{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{D}_i^{\mathrm{T}}$. For the second order derivative, one has

$$\mathbf{V}_n \equiv \left.\frac{\partial^2\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\overline{\boldsymbol{\theta}}} = n^{-1}\sum_{i=1}^{n} q_2\left(\bar{m}_i,Y_i\right)\mathbf{D}_i^{\mathrm{T}}\mathbf{D}_i,$$

where $\mathbf{D}_i$ is given in (A.1). According to Lemma A.3 and Condition (C2), $\left\|\mathbf{V}_n^{-1}\right\|_2 = O\left(1\right)$, a.s., thus

$$\left\|\widehat{\boldsymbol{\theta}}-\widetilde{\boldsymbol{\theta}}\right\| \leq \left\|\mathbf{V}_n^{-1}\right\|_2 \left\| \left.\frac{\partial\widehat{l}_n\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right\| = O_P\{N_n^{1/2-p}+(N_n/n)^{1/2}\}.$$

■

B.4. *Further results for simulation in subsection 5.1*

The following tables present the results for the settings of simulation, in which we generated $\mathbf{X}$ and $\mathbf{Z}$ as follows.

(i) X_1 and X_2 are independently uniformly distributed on $[0, 1]$. Z_1-Z_6 are normally distributed with mean 0, variance 0.09 and autoregressive structure with correlation coefficient $\rho = 0.5$. Z_7 and Z_8 are independently Bernoulli distributed with success probability 0.5.

(ii) All Xs and Zs except Z_5 and Z_6 are the same as those in (i). We let $Z_5 = X_1 + U_1$ and $Z_6 = X_2 + U_2$, where U_1 and U_2 are independently normal distributed with mean 0 and variance $1/48$ and $1/36$, respectively.

		Case (i)			Case (ii)		
n	method	C	I	MRME	C	I	MRME
100	ORACLE	5	0	0.29	5	0	0.12
	SCAD	4.41	0.84	0.58	4.23	0.89	0.53
	Lasso	4.07	0.71	0.51	4.09	0.74	0.34
	BIC	4.58	0.97	0.51	4.69	0.99	0.29
200	ORACLE	5	0	0.36	5	0	0.14
	SCAD	4.83	0.66	0.63	4.29	0.51	0.62
	Lasso	3.88	0.30	0.68	4.00	0.24	0.43
	BIC	4.82	0.73	0.67	4.79	0.71	0.27
400	ORACLE	5	0	0.33	5	0	0.15
	SCAD	4.74	0.21	0.56	4.53	0.15	0.44
	Lasso	3.74	0.08	0.69	3.89	0.03	0.44
	BIC	4.87	0.37	0.60	4.91	0.31	0.26

TABLE 3

Additional results from the simulation study in Section 5.1. C, I, and MRME stand for the average number of the five zero coefficients correctly set to 0, the average number of the three nonzero coefficients incorrectly set to 0, and the median of the relative model errors. The model errors are defined in (14).

REFERENCES

CARROLL, R., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, **92**, 477-489.

DE BOOR, C. (2001). *A Practical Guide to Splines*, New York: Springer-Verlag.

DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation: Polynomials and Splines Approximation*, Berlin: Springer-Verlag.

POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, **7**, 186-199.

STONE, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, **13**, 689-705.

WANG, L. and YANG, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics*, **35**, 2474-2503.