Web-based Supplementary Materials for "On selection of spatial linear models for lattice data"

Jun Zhu,[1,2,*] **Hsin-Cheng Huang**[3,4] **and Perla E. Reyes**[2]

[1]Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA

[2]Department of Statistics, University of Wisconsin–Madison, Madison, WI 53706 USA

[3]Institute of Statistics, Academia Sinica, Taipei, 115 Taiwan

[4]Institute of Statistics, National Chiao Tung University, 1001 University Road, Hsinchu, 30010 Taiwan

December 28, 2009

Appendix A: Assumptions

The following regularity conditions are made.

(A.1) $\boldsymbol{X}$ has full rank p and $\boldsymbol{\eta} \in \Omega$ where Ω is an open subset of $\mathbb{R}^{p+q+1}$ that contains $\boldsymbol{\eta}^0$.

(A.2) $\boldsymbol{\Gamma}_{\boldsymbol{\gamma}}$ is twice differentiable with respect to $\boldsymbol{\gamma}$ with continuous second-order derivatives and is positive definite.

(A.3) $\mathcal{I}(\boldsymbol{\eta})^{-1/2}\Big\{-\frac{\partial^2\ell(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\Big\}\mathcal{I}(\boldsymbol{\eta})^{-1/2} \xrightarrow{P} \boldsymbol{I}_{p+q+1}$, as $n \to \infty$.

(A.4) $n^{-1}\mathcal{I}(\boldsymbol{\beta}) \to \boldsymbol{J}(\boldsymbol{\beta})$ and $n^{-1}\mathcal{I}(\boldsymbol{\gamma}) \to \boldsymbol{J}(\boldsymbol{\gamma})$, as $n \to \infty$.

Since $\partial^2\ell(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}$ is non-random, sufficient conditions for (A.3) are,

$$E\left[\left\|\mathcal{I}(\boldsymbol{\beta})^{-1/2}\Big\{-\frac{\partial^2\ell(\boldsymbol{\eta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}'}\Big\}\mathcal{I}(\boldsymbol{\gamma})^{-1/2}\right\|^2\right] \to 0, E\left[\left\|\mathcal{I}(\boldsymbol{\gamma})^{-1/2}\Big\{-\frac{\partial^2\ell(\boldsymbol{\eta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\Big\}\mathcal{I}(\boldsymbol{\gamma})^{-1/2} - \boldsymbol{I}_{q+1}\right\|^2\right] \to 0,$$

where $\|\cdot\|^2$ is the Euclidean matrix norm. Some sufficient conditions of (A.3) and (A.4) can be found in Mardia and Marshall (1984).

* *email:* jzhu@stat.colostate.edu

Appendix B: Proof of Theorem 1

Proof. We first show that,

$$n^{-1/2}\frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}^0)). \tag{1}$$

For any $\boldsymbol{w} \in \mathbf{R}^{p+q+1}$, let, $\boldsymbol{\eta}^* \equiv \boldsymbol{\eta}^0 + \mathcal{I}(\boldsymbol{\eta}^0)^{-1/2}\boldsymbol{w}$. Then, $\boldsymbol{\eta}^* \to \boldsymbol{\eta}^0$, as $n \to \infty$. Write,

$$\ell(\boldsymbol{\eta}^*) = \ell(\boldsymbol{\eta}^0) + (\boldsymbol{\eta}^* - \boldsymbol{\eta}^0)'\frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} + \frac{1}{2}(\boldsymbol{\eta}^* - \boldsymbol{\eta}^0)'\frac{\partial \ell(\boldsymbol{\eta}^{**})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}(\boldsymbol{\eta}^* - \boldsymbol{\eta}^0),$$

where $\boldsymbol{\eta}^{**} = \alpha\boldsymbol{\eta}^0 + (1-\alpha)\boldsymbol{\eta}^*$ for some random variable $\alpha \in [0, 1]$. It follows after rearranging and taking exponential on both sides that,

$$\exp\left\{\boldsymbol{w}'\mathcal{I}(\boldsymbol{\eta}^0)^{-1/2}\frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}}\right\}p^0(\boldsymbol{y}) = \exp\left(\frac{1}{2}\boldsymbol{w}'\boldsymbol{V}\boldsymbol{w}\right)p^*(\boldsymbol{y}), \tag{2}$$

where $p^0(\boldsymbol{y})$ and $p^*(\boldsymbol{y})$ are the probability density functions of $\boldsymbol{y}$ under $\boldsymbol{\eta}^0$ and $\boldsymbol{\eta}^*$, respectively, and,

$$\boldsymbol{V} = \mathcal{I}(\boldsymbol{\eta}^0)^{-1/2}\left\{-\frac{\partial \ell(\boldsymbol{\eta}^{**})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\right\}\mathcal{I}(\boldsymbol{\eta}^0)^{-1/2},$$

which by (A.3) tends to $\boldsymbol{I}_{p+q+1}$ as $n \to \infty$. Integrating (2) with respect to $\boldsymbol{y}$ gives,

$$E^0\left[\exp\left\{\boldsymbol{w}'\mathcal{I}(\boldsymbol{\eta}^0)^{-1/2}\frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}}\right\}\right] = E^*\left\{\exp\left(\frac{1}{2}\boldsymbol{w}'\boldsymbol{V}\boldsymbol{w}\right)\right\} \to \exp\left(\frac{1}{2}\boldsymbol{w}'\boldsymbol{w}\right), \tag{3}$$

as $n \to \infty$, for any $\boldsymbol{w} \in \mathbb{R}^{p+q+1}$, where E^0 and E^* denote expectation under $\boldsymbol{\eta}^0$ and $\boldsymbol{\eta}^*$, respectively. It follows from (3) that $\mathcal{I}(\boldsymbol{\eta}^0)^{-1/2}\frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{I}_{p+q+1})$ by the uniqueness of the Laplace transform. This together with (A.4) imply (1).

The remainder arguments bear similarity to the proofs of Theorems 1–2 of Fan and Li (2001), but the details differ, as Fan and Li (2001) did not consider adaptive Lasso nor dependent errors in the regression. Let $c_n = n^{-1/2} + a_n$, $\boldsymbol{u} = (u_1, \ldots, u_p)' \in \mathbb{R}^p$,

$\boldsymbol{v} = (v_1, \ldots, v_{q+1})' \in \mathbb{R}^{q+1}$, and $\boldsymbol{w} = (\boldsymbol{u}', \boldsymbol{v}')'$. Then $\{\boldsymbol{\eta}^0 + c_n\boldsymbol{w} : \|\boldsymbol{w}\| \leq \delta\}$ denote the ball centered around $\boldsymbol{\eta}^0$ with radius $c_n\delta$. Write,

$$\begin{aligned} Q(\boldsymbol{\eta}^0 + c_n\boldsymbol{w}) - Q(\boldsymbol{\eta}^0) &= \ell(\boldsymbol{\eta}^0 + c_n\boldsymbol{w}) - \ell(\boldsymbol{\eta}^0) - n\sum_{j=1}^{p} \lambda_j(|\beta_j^0 + c_n u_j| - |\beta_j^0|) \\ &\quad - n\sum_{k=1}^{q} \tau_k(|\theta_k^0 + c_n v_k| - |\theta_k^0|). \end{aligned}$$

By Taylor's expansion and (1),

$$\ell(\boldsymbol{\eta}^0 + c_n\boldsymbol{w}) - \ell(\boldsymbol{\eta}^0) = c_n\boldsymbol{w}'\frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} - \frac{1}{2}nc_n^2\boldsymbol{w}'\boldsymbol{J}(\boldsymbol{\eta})\boldsymbol{w}\{1 + o_p(1)\},$$

where the first term of is of order $O_p(n^{1/2}c_n\delta)$ and the second term $O_p(nc_n^2\delta^2)$. Also note that, since $\beta_j^0 = 0$ for $j = s+1, \ldots, p$, and $\theta_k^0 = 0$ for $k = t+1, \ldots, q$,

$$\begin{aligned} & n\sum_{j=1}^{p} \lambda_j(|\beta_j^0 + c_n u_j| - |\beta_j^0|) + n\sum_{k=1}^{q} \tau_k(|\theta_k^0 + c_n v_k| - |\theta_k^0|) \\ &\geq n\sum_{j=1}^{s} \lambda_j(|\beta_j^0 + c_n u_j| - |\beta_j^0|) + n\sum_{k=1}^{t} \tau_k(|\theta_k^0 + c_n v_k| - |\theta_k^0|) \\ &\geq nc_n\sum_{j=1}^{s} \lambda_j|u_j| + nc_n\sum_{k=1}^{t} \tau_k|v_k| \geq nc_n^2 s\delta + nc_n^2 t\delta = nc_n^2(s+t)\delta, \end{aligned}$$

which is of order $O(nc_n^2\delta)$. Thus, for a given $\epsilon > 0$, there exists δ such that,

$$P\left\{\sup_{\|\boldsymbol{w}\|=\delta} Q(\boldsymbol{\eta}^0 + c_n\boldsymbol{w}) < Q(\boldsymbol{\eta}^0)\right\} > 1 - \epsilon.$$

Thus with probability at least $1-\epsilon$, there exists a local maximizer in the ball $\{\boldsymbol{\eta}^0 + c_n\boldsymbol{w} : \|\boldsymbol{w}\| \leq \delta\}$. This completes the proof of (i).

For $j = s+1, \ldots, p$,

$$\begin{aligned} \frac{\partial Q(\hat{\boldsymbol{\eta}})}{\partial \beta_j} &= \frac{\partial \ell(\hat{\boldsymbol{\eta}})}{\partial \beta_j} - n\lambda_j \mathrm{sgn}(\beta_j) \\ &= \frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \beta_j} + \frac{\partial^2 \ell(\boldsymbol{\eta}^0)}{\partial \beta_j^2}(\hat{\beta}_j - \beta_j^0)\{1 + o_p(1)\} - n\lambda_j \mathrm{sgn}(\beta_j), \end{aligned}$$

where by (1) and Theorem 1(i), the three terms are of order $O_p(n^{1/2})$, $O_p(n^{1/2})$, and $n^{1/2}O(n^{1/2}b_n)$, respectively. Since $n^{1/2}b_n \to \infty$, the last term dominates. Thus the sign of β_j determines the sign of $\partial Q(\hat{\boldsymbol{\eta}})/\partial\beta_j$ and $P(\hat{\beta}_j = 0) \to 1$ as $n \to \infty$ for $j = s+1, \ldots, p$. By similar arguments, $P(\hat{\theta}_k = 0) \to 1$ as $n \to \infty$ for $k = t+1, \ldots, q$. This completes the proof of (ii).

Since

$$\left.\frac{\partial Q(\boldsymbol{\eta})}{\partial \boldsymbol{\beta}_1}\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}} = \mathbf{0},$$

we have,

$$\mathbf{0} = \frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\beta}_1} + \frac{\partial^2 \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0)\{1 + o_p(1)\} - n(\lambda_1 \mathrm{sgn}(\beta_1^0), \ldots, \lambda_s \mathrm{sgn}(\beta_s^0))'.$$

Since $a_n = o(n^{-1/2})$, $n^{1/2}\lambda_j \mathrm{sgn}(\beta_j^0) = o(1)$ for $j = 1, \ldots, s$, thus by (1), we have, $n^{1/2}\boldsymbol{J}(\boldsymbol{\beta}_1^0)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\beta}_1^0))$. That is, $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\beta}_1^0)^{-1})$. By similar arguments, we have, $n^{1/2}(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1^0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\gamma}_1^0)^{-1})$. This completes the proof of (iii) and thus the theorem.

Appendix C: Proof of Theorem 2

Proof. Let $\hat{\boldsymbol{\eta}}^{(0)} = \hat{\boldsymbol{\eta}}_{\mathrm{MLE}}$, which maximizes $Q(\boldsymbol{\eta})$ under $\lambda_1 = \cdots = \lambda_p = 0$ and $\tau_1 = \cdots = \tau_q = 0$. Since

$$\mathbf{0} = \frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} + \frac{\partial^2 \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}(\hat{\boldsymbol{\eta}}^{(0)} - \boldsymbol{\eta}^0)\{1 + o_p(1)\},$$

by (1), (A.3) and (A.4), we have $n^{1/2}\boldsymbol{J}(\boldsymbol{\eta}^0)(\hat{\boldsymbol{\eta}}^{(0)} - \boldsymbol{\eta}^0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}^0))$. That is, $n^{1/2}(\hat{\boldsymbol{\eta}}^{(0)} - \boldsymbol{\eta}^0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}^0)^{-1})$.

We shall prove the oracle properties of $\hat{\boldsymbol{\eta}}^{(m)}$ by induction on $m \in \mathbb{N}$. For each m, we use arguments similar to those in the proof of Theorems 4–5 of Zou and Li (2008) for a one-step update under nonconcave penalized likelihood. For $\boldsymbol{\eta} \in \mathbb{R}^{p+q+1}$,

$$
\begin{aligned}
Q^*(\boldsymbol{\eta}) - Q^*(\boldsymbol{\eta}^0) &= (\boldsymbol{\eta} - \boldsymbol{\eta}^0)' \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(m-1)})}{\partial \boldsymbol{\eta}} - \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}^0)' \mathcal{I}(\hat{\boldsymbol{\eta}}^{(m-1)})(\boldsymbol{\eta} - \boldsymbol{\eta}^0) \\
&\quad -(\boldsymbol{\eta} - \boldsymbol{\eta}^0)' \mathcal{I}(\hat{\boldsymbol{\eta}}^{(m-1)})(\boldsymbol{\eta}^0 - \hat{\boldsymbol{\eta}}^{(m-1)}) - n \sum_{j=1}^{p} \lambda_j (|\beta_j| - |\beta_j^0|) \\
&\quad -n \sum_{k=1}^{q} \tau_k (|\theta_k| - |\theta_k^0|).
\end{aligned}
$$

Let $\boldsymbol{\eta}^* \equiv n^{1/2}(\boldsymbol{\eta} - \boldsymbol{\eta}^0)$. Rewrite $Q^*(\boldsymbol{\eta}) - Q^*(\boldsymbol{\eta}^0)$ as a function of $\boldsymbol{\eta}^*$, we obtain,

$$
\begin{aligned}
R(\boldsymbol{\eta}^*) &\equiv n^{-1/2} \boldsymbol{\eta}^{*'} \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(m-1)})}{\partial \boldsymbol{\eta}}) - \frac{1}{2} \boldsymbol{\eta}^{*'} \Big\{ \frac{1}{n} \mathcal{I}(\hat{\boldsymbol{\eta}}^{(m-1)}) \Big\} \boldsymbol{\eta}^* + \boldsymbol{\eta}^* \Big\{ \frac{1}{n} \mathcal{I}(\hat{\boldsymbol{\eta}}^{(m-1)}) \Big\} n^{1/2} \big(\hat{\boldsymbol{\eta}}^{(m-1)} - \boldsymbol{\eta}^0 \big) \\
&\quad -n \sum_{j=1}^{p} \lambda_j \big(|\beta_j^0 + n^{-1/2} \beta_j^*| - |\beta_j^0| \big) - n \sum_{k=1}^{q} \tau_k \big(|\theta_k^0 + n^{-1/2} \theta_k^*| - |\theta_k^0| \big).
\end{aligned}
$$

From (1), (A.3), $n^{1/2} b_n \to \infty$, and $a_n = o(n^{-1/2})$, we have,

$$
\begin{aligned}
\boldsymbol{\eta}^{*'} \Big\{ \frac{1}{n} \mathcal{I}(\hat{\boldsymbol{\eta}}^{(m-1)}) \Big\} \boldsymbol{\eta}^* &\xrightarrow{P} \boldsymbol{\eta}^{*'} \boldsymbol{J}(\boldsymbol{\eta}^0) \boldsymbol{\eta}^*, \\
\boldsymbol{\eta}^{*'} \Big\{ \frac{1}{n} \mathcal{I}(\hat{\boldsymbol{\eta}}^{(m-1)}) \Big\} n^{1/2} \big(\hat{\boldsymbol{\eta}}^{(m-1)} - \boldsymbol{\eta}^0 \big) &\xrightarrow{D} \boldsymbol{\eta}^{*'} \boldsymbol{J}(\boldsymbol{\eta}^0) \boldsymbol{\xi}, \\
n\lambda_j \big(|\beta_j^0 + n^{-1/2} \beta_j^*| - |\beta_j^0| \big) &\to \begin{cases} 0; & \text{if } j = 1, \dots, s, \\ \infty I\big(\beta_j^* \neq 0\big); & \text{if } j = s+1, \dots, p, \end{cases} \\
n\tau_k \big(|\theta_k^0 + n^{-1/2} \theta_k^*| - |\theta_k^0| \big) &\to \begin{cases} 0; & \text{if } k = 1, \dots, t, \\ \infty I\big(\theta_k^* \neq 0\big); & \text{if } k = t+1, \dots, q, \end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
n^{-1/2} \boldsymbol{\eta}^{*'} \frac{\partial \ell(\hat{\boldsymbol{\eta}}^{(m-1)})}{\partial \boldsymbol{\eta}}) &= n^{-1/2} \boldsymbol{\eta}^{*'} \Big\{ \frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} + \frac{\partial^2 \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \big(\hat{\boldsymbol{\eta}}^{(m-1)} - \boldsymbol{\eta}^0 \big) \Big\} \{1 + o_p(1)\} \\
&\xrightarrow{D} \boldsymbol{\eta}^{*'} \big(\boldsymbol{\zeta} - \boldsymbol{J}(\boldsymbol{\eta}^0) \boldsymbol{\xi} \big),
\end{aligned}
$$

where $\boldsymbol{\xi} \equiv n^{1/2}(\hat{\boldsymbol{\eta}}^{(m-1)} - \boldsymbol{\eta}^0) \sim N\big(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}^0)^{-1}\big)$, and $n^{-1/2} \frac{\partial \ell(\boldsymbol{\eta}^0)}{\partial \boldsymbol{\eta}} \xrightarrow{D} \boldsymbol{\zeta} = (\boldsymbol{\zeta}_1', \boldsymbol{\zeta}_2')' \sim N\big(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}^0)\big)$. It follows by Slutsky's theorem that,

$$
R(\boldsymbol{\eta}^*) \xrightarrow{D} R^\infty(\boldsymbol{\eta}^*) \equiv \Big\{ \boldsymbol{\eta}_1^{*'} \boldsymbol{\zeta}_1 - \frac{1}{2} \boldsymbol{\eta}_1^{*'} \boldsymbol{J}(\boldsymbol{\eta}_1^0) \boldsymbol{\eta}_1^* \Big\} I(\boldsymbol{\eta}_2^* = \mathbf{0}) - \infty I(\boldsymbol{\eta}_2^* \neq \mathbf{0}).
$$

Thus,

$$\arg\max_{\boldsymbol{\eta}^*} R^{\infty}(\boldsymbol{\eta}) = \begin{pmatrix} \boldsymbol{\zeta}_1 \\ \mathbf{0} \end{pmatrix}.$$

Let $\hat{\boldsymbol{\eta}}^{*(m)} = \big(\hat{\boldsymbol{\eta}}_1^{*(m)'}, \hat{\boldsymbol{\eta}}_2^{*(m)'}\big)' \equiv \arg\max_{\boldsymbol{\eta}^*} R(\boldsymbol{\eta}^*)$. Since $R(\cdot)$ is convex, by epiconvergence (Geyer, 1994), we obtain $\hat{\boldsymbol{\eta}}_1^{(m)*} \xrightarrow{D} \boldsymbol{J}(\boldsymbol{\eta}_1^0)^{-1}\boldsymbol{\zeta}_1 \sim N\big(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}_1^0)^{-1}\big)$ and $\hat{\boldsymbol{\eta}}_2^{(m)*} \xrightarrow{P} \mathbf{0}$ as $n \to \infty$, which implies,

$$n^{1/2}\hat{\boldsymbol{\eta}}_2^{(m)} \xrightarrow{P} \mathbf{0}. \tag{4}$$

Note that $\hat{\boldsymbol{\eta}}^{(m)} = \arg\max_{\boldsymbol{\eta}} \big\{Q^*(\boldsymbol{\eta}) - Q^*(\boldsymbol{\eta}^0)\big\} = n^{-1/2}\Big\{\arg\max_{\boldsymbol{\eta}^*} R(\boldsymbol{\eta}^*)\Big\} + \boldsymbol{\eta}^0$. It follows that,

$$n^{1/2}\big(\hat{\boldsymbol{\eta}}_1^{(m)} - \boldsymbol{\eta}_1^0\big) \xrightarrow{D} N\big(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\eta}_1^0)^{-1}\big). \tag{5}$$

We now show that $P\big(\hat{\boldsymbol{\eta}}_2^{(m)} = \mathbf{0}\big) \to 1$ as $n \to \infty$. Suppose that $\hat{\eta}_j^{(m)} \neq 0$ for some $j = s+1, \ldots, p$. From $Q^*(\boldsymbol{\eta})$

$$\frac{\partial \ell\big(\hat{\boldsymbol{\eta}}^{(m-1)}\big)}{\partial \eta_j} - \Big\{\mathcal{I}\big(\hat{\boldsymbol{\eta}}^{(m-1)}\big)\big(\hat{\boldsymbol{\eta}}^{(m)} - \hat{\boldsymbol{\eta}}^{(m-1)}\big)\Big\}_j = n\lambda_j \operatorname{sgn}(\hat{\eta}_j^{(m)}),$$

which implies,

$$\begin{aligned} &\frac{\partial \ell\big(\boldsymbol{\eta}^0\big)}{\partial \eta_j} + \frac{\partial^2 \ell\big(\boldsymbol{\eta}^0\big)}{\partial \eta_j^2}(\hat{\eta}_j^{(m-1)} - \eta_j^0)\{1 + o_p(1)\} \\ &\qquad + \mathcal{I}\big(\hat{\boldsymbol{\eta}}^{(m-1)}\big)\mathcal{I}\big(\boldsymbol{\eta}^0\big)^{-1}\Big\{\mathcal{I}\big(\boldsymbol{\eta}^0\big)\big(\hat{\boldsymbol{\eta}}^{(m)} - \boldsymbol{\eta}^0\big) - \mathcal{I}\big(\boldsymbol{\eta}^0\big)\big(\hat{\boldsymbol{\eta}}^{(m-1)} - \boldsymbol{\eta}^0\big)\Big\}_j \\ &= \quad n\lambda_j \operatorname{sgn}(\hat{\eta}_j^{(m)}). \end{aligned} \tag{6}$$

By Theorem 1, (1), (4), (5), and (A.3), the left-hand side of the equality of (6) is of order $O_p(n^{1/2})$, whereas the right-hand side is of order $n^{1/2}O(n^{1/2}b_n)$. Since $n^{1/2}b_n \to \infty$, it follows that $P(\hat{\eta}_j^{(m)} = 0) \to 1$ as $n \to \infty$ for $j = s+1, \ldots, p$. By similar arguments, $P(\hat{\eta}_{p+k}^{(m)} = 0) \to 1$ as $n \to \infty$ for $k = t+1, \ldots, q$. Thus $P\big(\hat{\boldsymbol{\eta}}_2^{(m)} = \mathbf{0}\big) \to 1$ as $n \to \infty$. This completes the proof.

Appendix D: Additional Simulation Results

See Tables 1–2.

Table 1
Mean and standard deviation (SD) of approximate penalized maximum likelihood estimates of model parameters using the proposed $LARS_m$ with either one or two regularization parameters. The sample sizes are $n = 25, 100, 225$ and the models are either conditional autoregressive (CAR) or simultaneous autoregressive (SAR) model.

Model		SAR						CAR					
Method	n	β_1	β_2	β_3	β_4	θ_1	σ^2	β_1	β_2	β_3	β_4	θ_1	σ^2
LARS_m	25	4.03	2.86	2.10	0.98	0.37	0.33	4.00	2.98	2.06	0.99	0.35	0.35
2-regular	(SD)	0.33	0.37	0.45	0.40	0.24	0.23	0.34	0.33	0.48	0.33	0.29	0.23
	100	3.99	3.00	2.00	1.01	0.20	0.90	3.99	3.01	2.00	1.01	0.21	0.90
	(SD)	0.14	0.16	0.15	0.19	0.02	0.13	0.14	0.17	0.12	0.18	0.03	0.16
	225	4.00	2.98	2.03	0.99	0.20	0.98	4.00	2.98	2.02	1.00	0.20	0.98
	(SD)	0.08	0.09	0.12	0.12	0.01	0.11	0.08	0.08	0.10	0.09	0.02	0.10
LARS_m	25	4.04	2.88	2.08	0.95	0.32	0.31	4.00	2.98	2.02	0.98	0.19	0.95
1-regular	(SD)	0.34	0.41	0.47	0.40	0.26	0.20	0.09	0.08	0.10	0.11	0.04	0.10
	100	3.99	3.00	2.00	1.01	0.21	0.87	3.99	2.99	2.00	1.01	0.20	0.87
	(SD)	0.14	0.16	0.15	0.21	0.04	0.14	0.15	0.16	0.12	0.16	0.08	0.14
	225	4.00	2.98	2.03	0.99	0.20	0.96	4.00	2.98	2.02	1.00	0.20	0.96
	(SD)	0.08	0.08	0.12	0.12	0.02	0.11	0.09	0.09	0.10	0.11	0.03	0.11
	Truth	4	3	2	1	0.2	1	4	3	2	1	0.2	1

Appendix E: Additional Data Analysis Results

See Table 3.

Table 2

Mean and standard deviation (SD) of approximate penalized maximum likelihood estimates of model parameters using three alternative methods: $LARS_1$ with either one or two regularization parameters, local quadratic approximation (LQA), and selection of regression coefficients only. The sample sizes are $n = 25, 100, 225$ and the models are either conditional autoregressive (CAR) or simultaneous autoregressive (SAR) model.

Model		SAR						CAR					
Method	n	β_1	β_2	β_3	β_4	θ_1	σ^2	β_1	β_2	β_3	β_4	θ_1	σ^2
LARS_1	25	4.00	2.94	2.03	1.02	0.25	0.34	4.02	2.94	2.03	0.93	0.28	0.34
2-regular	(SD)	0.49	0.41	0.46	0.38	0.22	0.22	0.34	0.32	0.34	0.31	0.34	0.22
	100	3.99	2.98	2.00	0.98	0.20	0.88	3.99	2.99	2.01	0.98	0.19	0.85
	(SD)	0.12	0.17	0.14	0.18	0.03	0.19	0.12	0.14	0.12	0.17	0.08	0.11
	225	4.01	2.99	2.01	0.97	0.19	0.95	4.00	2.98	2.02	0.97	0.18	0.95
	(SD)	0.10	0.11	0.09	0.11	0.02	0.08	0.09	0.08	0.10	0.11	0.03	0.10
LARS_1	25	4.03	2.87	2.07	0.92	0.31	0.30	4.02	2.94	2.03	0.93	0.18	0.33
1-regular	(SD)	0.34	0.43	0.53	0.40	0.28	0.20	0.33	0.33	0.34	0.36	0.25	0.21
	100	3.99	2.99	2.02	0.97	0.21	0.85	4.02	2.94	2.03	0.93	0.18	0.33
	(SD)	0.14	0.15	0.16	0.20	0.04	0.13	0.33	0.33	0.34	0.36	0.25	0.21
	225	3.99	2.98	2.02	0.98	0.20	0.95	4.00	2.98	2.02	0.98	0.19	0.95
	(SD)	0.08	0.09	0.12	0.13	0.02	0.11	0.09	0.08	0.10	0.11	0.04	0.10
LQA	25	4.01	2.94	2.04	1.02	0.24	0.41	3.99	2.96	1.99	1.03	0.10	0.52
	(SD)	0.45	0.37	0.46	0.30	0.18	0.28	0.38	0.32	0.40	0.34	0.18	0.30
	100	3.99	2.98	2.00	0.98	0.20	0.91	3.99	2.99	2.00	0.98	0.18	0.92
	(SD)	0.12	0.16	0.13	0.16	0.03	0.17	0.13	0.17	0.13	0.17	0.07	0.15
	225	4.01	2.98	2.01	0.97	0.19	0.96	4.01	2.98	2.01	0.97	0.18	0.97
	(SD)	0.10	0.10	0.10	0.11	0.01	0.08	0.08	0.10	0.11	0.08	0.04	0.09
LARS_m	25	4.03	2.89	2.10	0.95	0.23	0.48	4.01	2.93	2.04	0.94	0.15	0.47
regression-	(SD)	0.30	0.32	0.46	0.36	0.17	0.41	0.34	0.29	0.42	0.29	0.06	0.31
only	100	3.99	3.00	2.00	1.00	0.20	0.93	4.00	2.99	2.01	0.99	0.18	0.90
	(SD)	0.16	0.17	0.15	0.18	0.02	0.12	0.14	0.14	0.12	0.16	0.06	0.13
	225	4.00	2.98	2.03	0.99	0.20	0.98	4.00	2.98	2.02	0.98	0.19	0.98
	(SD)	0.08	0.10	0.12	0.12	0.01	0.10	0.09	0.09	0.11	0.11	0.02	0.10
LARS_1	25	4.02	2.89	2.08	0.92	0.24	0.43	4.00	2.98	2.02	0.98	0.19	0.98
regression-	(SD)	0.32	0.34	0.47	0.34	0.18	0.30	0.09	0.09	0.11	0.11	0.02	0.10
only	100	3.99	2.99	2.02	0.98	0.20	0.90	4.00	2.99	2.01	1.01	0.18	0.92
	(SD)	0.15	0.18	0.14	0.18	0.02	0.12	0.16	0.16	0.11	0.17	0.05	0.15
	225	3.99	2.98	2.03	0.98	0.20	0.98	4.00	2.98	2.02	1.00	0.19	0.98
	(SD)	0.08	0.09	0.12	0.12	0.01	0.09	0.09	0.08	0.10	0.10	0.03	0.10
	Truth	4	3	2	1	0.2	1	4	3	2	1	0.2	1

Table 3

Nonzero approximate penalized maximum likelihood estimate (APMLE) and standard deviation (SD) of model parameters in a simultaneous autoregressive model A using either $LARS_1$ or LQA for the mountain pine beetle data example.

Model	LARS_1			LQA		
Variable	Parameter	APMLE	SD	Parameter	APMLE	SD
Covariates						
Elevation	β_1	–	–	β_1	–	–
Temp min	β_2	–	–	β_2	–	–
Temp max	β_3	-4.18	13.68	β_3	-4.81	14.40
Temp mean	β_4	5.00	10.94	β_4	5.37	11.35
August temp mean	β_5	–	–	β_5	–	–
DD	β_6	–	–	β_6	–	–
DDEG	β_7	–	–	β_7	–	–
Precip	β_8	0.89	0.74	β_8	1.09	0.78
Order of neighborhoods						
1st	θ_1	0.17	0.04	θ_1	0.16	0.03
2nd	θ_2	–	–	θ_2	–	–
3rd	θ_3	0.02	0.04	θ_3	0.05	0.04
4th	θ_4	–	–	θ_4	–	–
5th	θ_5	–	–	θ_5	–	–
Variance						
	σ^2	8.70	1.29	σ^2	8.78	1.30
BIC	357.73			353.81		

References

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

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

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–146.

Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *Annals of Statistics* **36**, 1509–1566.