Supplementary Document
for
Functional Linear Model with Zero-value Coefficient Function at Sub-regions

Jianhui Zhou, Nae-Yuh Wang, and Naisyin Wang

University of Virginia, Johns Hopkins University, and University of Michigan

Algorithm for the Refinement Stage

We present a practical algorithm here to implement the null region refinment and function estimation stage in Section 2.2 with $D = 1$.

Knots Placement. Denote the initial estimate of $\mathcal{T}$ by $\hat{\mathcal{T}}^{(0)} = \bigcup_{j=1}^{J}[a_j, c_j]$, which is the union of the identified subintervals in Section 2.1.

KP.1 Remove the initial knots within $[a_j, c_j]$.

KP.2 On $\hat{\mathcal{T}}^{(0),c}$, evenly-spaced knots are placed, and the total number of this set of knots is $k_{1,n} + 1$ with $k_{1,n} < k_{0,n}$. Denote this new set of knots by $\mathcal{A}$.

Working Null-region with the One-step Group SCAD Estimator. An iteration process is carried out in this step.

WN.1 Let $l = 0$.

WN.2 Take the working null region $\mathcal{T}_l = \bigcup_{j=1}^{J}[a_j + l\delta_n, c_j - l\delta_n]$ when $a_1 \neq 0$ and $c_J \neq T$. When $a_1 = 0$ or $c_J = T$, the interval $[0, c_1 - l\delta_n]$ or $[a_J + l\delta_n, T]$ are counted into the working null region.

WN.3 The current knots on $[0, T]$ contains the knots in $\mathcal{A}$ and the boundaries of working null regions $\mathcal{T}_k$ for $k = 0, ..., l$. Using this set of knots, compute the variables in the approximate model (2.4).

WN.4 Get the initial value $\tilde{\boldsymbol{b}}_1$ by least squares, and divide $\tilde{\boldsymbol{b}}_1$ into $\tilde{\boldsymbol{b}}_{1N,l}$ and $\tilde{\boldsymbol{b}}_{1S,l}$ according to their association to $\mathcal{T}_l$.

WN.5 Estimate $\boldsymbol{b}_1$ by minimizing $Q_n(\mathcal{T}_l, \lambda, \boldsymbol{b})$ by LARS algorithm, where λ is selected by the crition $C(\mathcal{T}_l, \lambda)$ to be discussed below.

WN.6 Let $l = l + 1$ and repeat WN.2-WN.5 until one interval $[a_j, c_j]$ shrinks to the empty set.

The criterion $C(\mathcal{T}_l, \lambda)$ can be generalized cross validation criterion (GCV), Akaike's information criterion (AIC), the Bayesian information criterion (BIC; Schwarz) and the residual information criterion (RIC). They are defined as

$$\begin{aligned} GCV(\mathcal{T}_l, \lambda) &= RSS/[n\{1 - d(\lambda)/n\}^2], \\ AIC(\mathcal{T}_l, \lambda) &= nlog(RSS/n) + 2d(\lambda), \\ BIC(\mathcal{T}_l, \lambda) &= nlog(RSS/n) + log(n)d(\lambda), \\ RIC(\mathcal{T}_l, \lambda) &= \{n - d(\lambda)\}log(\tilde{\sigma}^2) + d(\lambda)\{log(n) - 1\} + 4/\{n - d(\lambda) - 2\}, \end{aligned}$$

where RSS is the residual sum of squares, $d(\lambda)$ is the number of non-zero estimated coefficients when the tuning parameter is chosen to be λ, and $\tilde{\sigma}^2 = RSS/\{n - d(\lambda)\}$.

Final Determination of the Refined Estimation of $\mathcal{T}$ and $\beta(t)$. Identify the l_f that reaches the smallest criterion value across l and λ.

FD.1 Let $l_f = \arg\min_l C(\mathcal{T}_l, \arg\min_{\lambda>0} C(\mathcal{T}_l, \lambda))$. The refined estimate of the null region is $\hat{\mathcal{T}} = \mathcal{T}_{l_f}$.

FD.2 Let $\hat{\boldsymbol{b}}_1 = \arg\min_{\boldsymbol{b}} Q_n(\hat{\mathcal{T}}, \arg\min_{\lambda>0} C(\hat{\mathcal{T}}, \lambda), \boldsymbol{b})$. The refined estimate of $\beta(t)$ is $\hat{\beta}(t) = \boldsymbol{B}_1^T(t)\hat{\boldsymbol{b}}_1$, where $\boldsymbol{B}_1^T(t)$ are the B-spline basis function generated in Step 2.3 using the knots in $\mathcal{A}$ and the boundaries of working null regions $\mathcal{T}_k$ for $k = 0, ..., l_f$.

Proofs

We use $a_n > O_p(b_n)$ and $a_n \geq O_p(b_n)$ to denote that, as $n \to \infty$ with probability tending to 1, $b_n/a_n \to 0$ and b_n/a_n is bounded from above, respectively. We need the following lemma.

<u>LEMMA</u> **1** *Let* $\boldsymbol{b}_0(n) = (b_{0,1}(n), b_{0,2}(n), ..., b_{0,k_{0,n}+h}(n))^T$ *and assume that* $\beta(t)$ *has* r*th bounded derivative on* $[0, T]$ *where* $r \geq 3$*. There exists a constant* M_0 *such that for all* $b_{0,j}(n)$ *which are associated with* $\mathcal{T}$, $\max |b_{0,j}(n)| \leq M_0 k_{0,n}^{-r}$.

Lemma 1 is a direct result of the local property of the B-spline basis functions. The proof of Lemma 1 is straightforward, and is thus omitted.

Proof of the convergence rate of the initial estimator by least squares: We first prove the convergence rate of the initial estimator $\tilde{\boldsymbol{b}}_1(n)$ of $\boldsymbol{b}_1(n)$ by least squares in the refinement stage.

Define $\boldsymbol{\epsilon}_1(n) = (\epsilon_{1,1}, ..., \epsilon_{1,n})^T$ and $\boldsymbol{e}(n) = (e_1, ..., e_n)^T$. Let $L_n\{\boldsymbol{b}(n)\} = \sum_{i=1}^n (Y_i - \boldsymbol{z}_{1,i}\boldsymbol{b}(n))^2$. Given $\tilde{\boldsymbol{b}}_1(n)$ is the minimizer of $L_n\{\boldsymbol{b}(n)\}$, we have

$$\begin{aligned} & L_n\{\tilde{\boldsymbol{b}}_1(n)\} - L_n\{\boldsymbol{b}_1(n)\} \\ = \; & [\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T \boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n)[\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] - 2(\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n))[\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\ \leq \; & 0. \end{aligned}$$

Given A_8, we have $[\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T \boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n)[\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \geq c_1'(n/k_{1,n})||\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2}^2$. Since the approximation error $e_1(t)$ is bounded below $Ck_{1,n}^{-r}$ in absolute value for some constant C, A_2 ensures that $\sup |\epsilon_{1,i} - e_i| \leq M'Ck_{1,n}^{-1}$. Thus, the term $||\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n)||_{l_2} = ||\boldsymbol{Z}_1^T(n)\boldsymbol{e}(n) + \boldsymbol{Z}_1^T(n)(\boldsymbol{\epsilon}_1(n) - \boldsymbol{e}(n))||_{l_2}$ is dominated by $||\boldsymbol{Z}_1^T(n)\boldsymbol{e}(n)||_{l_2}$. Given $\boldsymbol{e}(n) \sim N(0, I_n)$, we have $n^{-1/2}(\boldsymbol{Z}_1^T(n)\boldsymbol{e}(n)) \sim N(0, n^{-1}\boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n))$, which indicates $(n^{-1}\boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n))^{-1/2} n^{-1/2}(\boldsymbol{Z}_1^T(n)\boldsymbol{e}(n)) \sim N(0, I_{k_{1,n}+h})$, where $h+1$ is the B-spline basis function order. Therefore we have $||(n^{-1}\boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n))^{-1/2} n^{-1/2}\boldsymbol{Z}_1^T(n)\boldsymbol{e}(n))||_{l_2}^2 \sim \chi^2(k_{1,n} + h)$. Given A_8, we have

$$||\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n)||_{l_2} = O_p(n^{1/2}). \tag{1}$$

Therefore,

$$\begin{aligned} & c_1'(n/k_{1,n})||\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2}^2 \\ \leq \; & [\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T \boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n)[\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\ \leq \; & 2(\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n))^T[\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\ \leq \; & 2||\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n)||_{l_2}||\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2} \\ = \; & O_p(n^{1/2})||\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2}, \end{aligned}$$

which indicates $||\tilde{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2} = O_p(n^{-1/2}k_{1,n})$. $\square$

Proof of Theorem 1, Part (iii):

Assuming A_6, with probability tending to 1, the coefficients $b_{0,j}(n)$ that are associated with $\mathcal{T}$ are identified correctly with the threshold value d_n, and, thus, the subintervals I_j that are in $\mathcal{T}$ are identified correctly into $\hat{\mathcal{T}}^{(0)}$. For a subinterval $I_j \subseteq \{t \in [0,T] : |\beta(t)| \geq k_{0,n}^{-r+2}\}$, the associated coefficients are $b_{0,j}(n), ..., b_{0,j+h}(n)$. Taking $t_0 \in I_j$, we have $\beta(t_0) = \sum_{k=0}^{h} B_{0,j+k}(t_0)b_{0,j+k}(n) + e_0(t_0)$, where $|e_0(t)| \leq ck_{0,n}^{-r}$ is the approximation error. Given the B-spline basis functions are all bounded between 0 and 1, we have that

$$\sum_{k=0}^{h} |b_{0,j+k}(n)| \geq |\sum_{k=0}^{h} B_{0,j+k}(t_0)b_{0,j+k}(n)| = |\beta(t_0) - e_0(t_0)| \geq k_{0,n}^{-r+2} - ck_{0,n}^{-r}.$$

Thus, we have that, when $k_{0,n}$ is large enough, $\sum_{k=0}^{h} |b_{0,j+k}(n)| \geq (1/2)k_{0,n}^{-r+2}$, and at least one of the coefficients $b_{0,j}(n)$ associated with I_j is larger than $(1/2)(h+1)^{-1}k_{0,n}^{-r+2}$ in absolute value. Given A_5, with probability tending to 1, at least one of the estimated coefficients $\tilde{b}_{0,j}(n)$ associated with I_j is larger than $(1/4)(h+1)^{-1}k_{0,n}^{-r+2}$ in absolute value as $k_{0,n}$ goes to infinity. By A_6, the subinterval $I_j \subseteq \{t \in [0,T] : |\beta(t)| \geq k_{0,n}^{-r+2}\}$ is identified correctly into $\hat{\mathcal{T}}^{(0),c}$ with probability tending to 1.

In summary, we have that the subintervals I_j in $\mathcal{T}$ are identified into $\hat{\mathcal{T}}^{(0)}$ and the subintervals I_j in $\{t \in [0,T] : |\beta(t)| \geq k_{0,n}^{-r+2}\}$ are identified into $\hat{\mathcal{T}}^{(0),c}$ with probability tending to 1. As a result, when the length of I_j goes to 0 as $k_{0,n}$ goes to ∞, we have $\mathcal{T} \subseteq \hat{\mathcal{T}}^{(0)}$ and $\hat{\mathcal{T}}^{(0)} \cap \mathcal{T}^c \subseteq \Omega(k_{0,n})$ with probability tending to 1, where $\Omega(k_{0,n}) = \{t \in [0,T] : 0 < |\beta(t)| < k_{0,n}^{-r+2}\}$ as defined in Theorem 1. The sub-region $\Omega(k_{0,n})$ converges to the empty region as $k_{0,n} \to \infty$. Part (iii) is proved.

Proof of Theorem 2: First we prove that $||\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2} \leq O_p(n^{-1/2}k_{1,n}^{3/2})$. This is a non-optimal bound for the convergence rate of $\hat{\boldsymbol{b}}_1(n)$, but it is sufficient to use to show the following Oracle property.

For the coefficient $b_{1,j}(n)$ associated with $\mathcal{T}$, given the construction of the $k_{1,n}+1$ adaptive knots, the results of Lemma 1 applies, i.e. $|b_{1,j}(n)| \leq Ck_{1,n}^{-r}$ for some constant C. Assume the coefficient $b_{1,j}(n)$ is associated with the region $\Omega(k_{0,n})$. The construction of the $k_{1,n}+1$ adaptive knots indicates that the knots are evenly-spaced on $\Omega(k_{0,n})$. Since $|\beta(t)| < k_{0,n}^{-r+2}$ when $t \in \Omega(k_{0,n})$, as in Lemma 1, given A_5, it is true that $|b_{1,j}(n)| < C' k_{0,n}^{-r+2}$ for $b_{1,j}(n)$ associated with $\Omega(k_{0,n})$, where C' is a constant. Recall that $\boldsymbol{b}_{1N}(n)$ and $\boldsymbol{b}_{1S}(n)$ are the division of $\boldsymbol{b}_1(n)$ according to $\hat{\mathcal{T}}^{(0)}$. Since $\boldsymbol{b}_{1N}(n)$ contains the coefficients associated with $\hat{\mathcal{T}}^{(0)}$, given the results in Theorem 1 (iii), these coefficients are either associated with $\mathcal{T}$ or with $\Omega(k_{0,n})$. Also, there are only a finite number of coefficients in $\boldsymbol{b}_{1N}(n)$ according to our method to place the $k_{1,n}$ knots. Thus, given A_5, we have that $||\boldsymbol{b}_{1N}(n)||_{l_1} = O_p(k_{0,n}^{-r+2})$. Let M be the maximum of $|\beta(t)|$ on $\mathcal{T}^c$. Following the proofs of Part (iii) of Theorem 1, we have that there is at least one coefficient in $\boldsymbol{b}_{1S}(n)$ that is greater than $M/[2(h+1)]$ in absolute value, where $h+1$ is the fixed spline order. Thus, $||\boldsymbol{b}_{1S}(n)||_{l_1} \geq O_p(1)$.

Recall that $\tilde{\boldsymbol{b}}_{1N}(n)$ and $\tilde{\boldsymbol{b}}_{1S}(n)$ are the division of $\tilde{\boldsymbol{b}}_1(n)$ according to $\hat{\mathcal{T}}^{(0)}$. Given $||\tilde{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_1} \leq C||\tilde{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} = O_p(n^{-1/2}k_{1,n})$, $||\boldsymbol{b}_{1N}(n)||_{l_1} = O_p(k_{0,n}^{-r+2})$ and A_5, we have that $||\tilde{\boldsymbol{b}}_{1N}(n)||_{l_1} = O_p(k_{0,n}^{-r+2})$ and $||\tilde{\boldsymbol{b}}_{1S}(n)||_{l_2} \geq O_p(1)$. Given A_7, with probability tending to 1, we have that $p'_{\lambda_n}(||\tilde{\boldsymbol{b}}_{1N}(n)||_{l_1}) = \lambda_n$ and $p'_{\lambda_n}(||\tilde{\boldsymbol{b}}_{1S}(n)||_{l_1}) = 0$. Since $\hat{\boldsymbol{b}}_1(n)$ minimizes $Q_n\{\boldsymbol{b}(n)\}$, with probability tending to 1, we have

$$
\begin{aligned}
0 &\geq Q_n\{\hat{\boldsymbol{b}}_1(n)\} - Q_n\{\boldsymbol{b}_1(n)\} \\
&= [\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T \boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n)[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] - 2(\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n))^T[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\
&+ n\lambda_n(||\hat{\boldsymbol{b}}_{1N}(n)||_{l_1} - ||\boldsymbol{b}_{1N}(n)||_{l_1}) \\
&\geq [\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T \boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n)[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] - 2(\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n))^T[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\
&+ n\lambda_n(||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_1} - 2||\boldsymbol{b}_{1N}(n)||_{l_1}),
\end{aligned}
$$

where $\hat{\boldsymbol{b}}_{1N}(n), \hat{\boldsymbol{b}}_{1S}(n)$ and $\boldsymbol{b}_{1N}(n), \boldsymbol{b}_{1S}(n)$ are the divisions of $\hat{\boldsymbol{b}}_1(n)$ and $\boldsymbol{b}_1(n)$, respectively, according to their association with $\hat{\mathcal{T}}^{(0)}$.

We first show that $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} \leq O_p(n^{-1/2}k_{1,n}^{3/2})$. Suppose that this is not true and that $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} > O_p(n^{-1/2}k_{1,n}^{3/2})$, which indicates $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_1} > O_p(n^{-1/2}k_{1,n}^{3/2})$. Since $||\boldsymbol{b}_{1N}(n)||_{l_1} = O_p(k_{0,n}^{-r+2})$, given A_5, we have $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_1} - 2||b_{1N}(n)||_{l_1} > 0$ with probability tending to 1.

Given $Q_n\{\hat{\boldsymbol{b}}_1(n)\} - Q_n\{\boldsymbol{b}_1(n)\} \leq 0$ and A_8, we have, with probability tending to,

$$\begin{aligned} & c_1'(n/k_{1,n})||\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2}^2 \\ \leq \; & [\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T \boldsymbol{Z}_1^T(n)\boldsymbol{Z}_1(n)[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\ \leq \; & 2(\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n))^T[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)] \\ \leq \; & 2||\boldsymbol{Z}_1^T(n)\boldsymbol{\epsilon}_1(n)||_{l_2}||\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2}. \end{aligned}$$

Given (1), we have $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} \leq ||\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2} = O_p(n^{-1/2}k_{1,n})$, which is contradictive to the assumption $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} > O_p(n^{-1/2}k_{1,n}^{3/2})$. Therefore we have

$$||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} \leq O_p(n^{-1/2}k_{1,n}^{3/2}). \tag{2}$$

Next, we show that $||\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)||_{l_2} = O_p(n^{-1/2}k_{1,n}^{3/2})$. We first define

$$Q_{n,S}\{(\boldsymbol{b}_S(n)\} = Q_n\{\boldsymbol{b}(n)|\boldsymbol{b}_N(n) = \hat{\boldsymbol{b}}_{1N}(n)\}.$$

Since $\hat{\boldsymbol{b}}_1(n)$ minimizes $Q_n\{\boldsymbol{b}(n)\}$, we have that $\hat{\boldsymbol{b}}_{1S}(n)$ is the minimizer of $Q_{n,S}\{\boldsymbol{b}_S(n)\}$. Therefore, when n is large,

$$\begin{aligned} 0 \; \geq \; & Q_{n,S}\{\hat{\boldsymbol{b}}_{1S}(n)\} - Q_{n,S}\{\boldsymbol{b}_{1S}(n)\} \\ = \; & [\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)]^T \boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)[\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)] \\ - \; & 2[\boldsymbol{Z}_{1S}^T(n)\boldsymbol{\epsilon}_1(n) - \boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)(\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n))]^T[\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)]. \end{aligned}$$

Given A_8, we have

$$\begin{aligned} & c_1'(n/k_{1,n})||\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)||_{l_2}^2 \\ \leq \; & [\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)]^T \boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)[\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)] \\ \leq \; & 2[\boldsymbol{Z}_{1S}^T(n)\boldsymbol{\epsilon}_1(n) - \boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)(\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n))]^T[\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)] \\ \leq \; & 2||\boldsymbol{Z}_{1S}^T(n)\boldsymbol{\epsilon}_1(n) - \boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)(\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n))||_{l_2}||\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)||_{l_2} \\ \leq \; & 2\{||\boldsymbol{Z}_{1S}^T(n)\boldsymbol{\epsilon}_1(n)||_{l_2} + ||\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)(\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n))||_{l_2}\}||\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)||_{l_2}. \end{aligned}$$

Following the steps to show (1), we obtain that $||\boldsymbol{Z}_{1S}^T(n)\boldsymbol{\epsilon}_1(n)||_{l_2} = O_p(n^{1/2})$. Since $||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2} = O_p(n^{-1/2}k_{1,n}^{3/2})$, given A_8, we have

$$\begin{aligned} & ||\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)(\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n))||_{l_2}^2 \\ = \; & [\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)]^T \boldsymbol{Z}_{1N}^T(n)\boldsymbol{Z}_{1S}(n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)[\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)] \\ \leq \; & c_3(n/k_{1,n})||\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n)||_{l_2}^2 \\ = \; & O_p(k_{1,n}^2). \end{aligned}$$

Thus, we have $||\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)(\hat{\boldsymbol{b}}_{1N}(n) - \boldsymbol{b}_{1N}(n))||_{l_2} = O_p(k_{1,n})$, and

$$||\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)||_{l_2} \quad \leq \quad O_p(n^{-1/2}k_{1,n}^{3/2}). \tag{3}$$

Given (2) and (3), we have

$$||\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2} \quad \leq \quad O_p(n^{-1/2}k_{1,n}^{3/2}).$$

Finally, we prove the oracle property of the proposed estimator.

We first show that $\hat{b}_{1,j}(n) = 0$, with probability tending to 1, for any $\hat{b}_{1,j}(n)$ associated with $\hat{\mathfrak{T}}^{(0)}$. We take the partial derivative of $Q_n\{\boldsymbol{b}(n)\}$ at $\boldsymbol{b}(n) = \hat{\boldsymbol{b}}_1(n)$ with respect to $b_{1,j}(n)$ in $\boldsymbol{b}_{1N}(n)$. As shown above, we have $p'_{\lambda_n}(||\tilde{\boldsymbol{b}}_{1N}(n)||_{l_1}) = \lambda_n$ and $p'_{\lambda_n}(||\tilde{\boldsymbol{b}}_{1S}(n)||_{l_1}) = 0$ with probability tending to 1. The partial derivative is then

$$\begin{aligned}
& \frac{\partial Q_n\{\boldsymbol{b}(n)\}}{\partial b_j(n)}|_{\boldsymbol{b}(n)=\hat{\boldsymbol{b}}_1(n)} \\
= & \sum_{i=1}^n 2[Y_i - \boldsymbol{z}_{1,i}\hat{\boldsymbol{b}}_1(n)](-z_{1,i,j}) + n\lambda_n sign[\hat{b}_{1,j}(n)] \\
= & \sum_{i=1}^n 2\{Y_i - \boldsymbol{z}_{1,i}\boldsymbol{b}_1(n) + \boldsymbol{z}_{1,i}[\boldsymbol{b}_1(n) - \hat{\boldsymbol{b}}_1(n)]\}(-z_{1,i,j}) + n\lambda_n sign[\hat{b}_{1,j}(n)] \\
= & -2\boldsymbol{Z}_{1,j}^T(n)\boldsymbol{\epsilon}_1(n) + 2[\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)]^T\boldsymbol{Z}_1^T(n)\boldsymbol{Z}_{1,j}(n) + n\lambda_n sign[\hat{b}_{1,j}(n)] \\
= & -I - II + III,
\end{aligned}$$

where $\boldsymbol{Z}_{1,j}(n)$ is the jth column of the matrix $\boldsymbol{Z}_1(n)$.

Given A_2 and the uniformly bounded B-spline approximation error, we have $\sup|\epsilon_{1,i} - e_i| \leq M'Ck_{1,n}^{-1}$ for some constant C. Thus, the term $\boldsymbol{Z}_{1,j}^T(n)\boldsymbol{\epsilon}_1(n)$ is dominated by $\boldsymbol{Z}_{1,j}(n)\boldsymbol{e}_n$. Since $\boldsymbol{e}(n) \sim N(0, I_n)$, we have

$$(k_{1,n}/n)^{1/2}\boldsymbol{Z}_{1,j}^T(n)\boldsymbol{e}(n) \quad \sim \quad N[0, (k_{1,n}/n)\boldsymbol{Z}_{1,j}^T(n)\boldsymbol{Z}_{1,j}(n)].$$

Given A_8, we know that $(k_{1,n}/n)\boldsymbol{Z}_{1,j}^T(n)\boldsymbol{Z}_{1,j}(n)$ is between the constants c_1' and c_2'. Therefore,

$$(k_{1,n}/n)^{1/2}I \quad = \quad N[0, (k_{1,n}/n)\boldsymbol{Z}_{1,j}^T(n)\boldsymbol{Z}_{1,j}(n))] + o_p(1).$$

By A_8, we have $||\boldsymbol{Z}_1^T(n)\boldsymbol{Z}_{1,j}(n)||_{l_2} = O_p(nk_{1,n}^{-1})$. Thus, we have

$$\begin{aligned} |(k_{1,n}/n)^{1/2}II| &\leq 2(k_{1,n}/n)^{1/2}||\hat{\boldsymbol{b}}_1(n) - \boldsymbol{b}_1(n)||_{l_2}||\boldsymbol{Z}_1^T(n)\boldsymbol{Z}_{1,j}(n)||_{l_2} \\ &= 2(k_{1,n}/n)^{1/2}O_p(n^{-1/2}k_{1,n}^{3/2})O_p(nk_{1,n}^{-1}) \\ &= O_p(k_{1,n}). \end{aligned}$$

We also have

$$(k_{1,n}/n)^{1/2}III = n^{1/2}\lambda_n k_{1,n}^{1/2}.$$

Since $Q_n\{\boldsymbol{b}(n)\}$ minimizes at $\hat{\boldsymbol{b}}_1(n)$, we have that

$$I + II = III.$$

Given A_5 and A_7, we have $|I/III| = o_p(1)$ and $|II/III| = o_p(1)$. Therefore,

$$Pr(\hat{b}_{1,j}(n) \neq 0) \leq Pr(I + II = III) \to 0,$$

indicating that, with probability tending to 1, $\hat{b}_{1,j}(n) = 0$ for any $\hat{b}_{1,j}(n)$ associated with $\hat{\mathfrak{T}}^{(0)}$. Since $\mathfrak{T} \subseteq \hat{\mathfrak{T}}^{(0)}$, with probability tending to 1, as shown in Theorem 1, we have that $\hat{\beta}(t) = 0$ for $t \in \mathfrak{T}$ with probability tending to 1. Part (i) is proved.

Next, we show the asymptotic distribution of $\hat{\beta}(t)$ for $t \in \mathfrak{T}^c$. We first define

$$P_n(\boldsymbol{b}') = \sum_{i=1}^{n}(Y_i - \boldsymbol{z}_{1S,i}\boldsymbol{b}')^2,$$

where $\boldsymbol{z}_{1S,i}$ are the elements of $\boldsymbol{z}_{1,i}$ that correspond to the coefficients in $\boldsymbol{b}_S(n)$.

With probability tending to 1, $\hat{\boldsymbol{b}}_{1N}(n) = \boldsymbol{0}$ and $p'_{\lambda_n}(||\tilde{\boldsymbol{b}}_{1S}(n)||_{l_1}) = 0$ as shown above. Since $\hat{\boldsymbol{b}}_1(n)$ minimizes $Q_n\{\boldsymbol{b}(n)\}$, we know that $\hat{\boldsymbol{b}}_{1S}(n)$ is the minimizer of $P_n(\boldsymbol{b}')$ and $\nabla P_n\{\hat{\boldsymbol{b}}_{1S}(n)\} = \boldsymbol{0}$, with probability tending to 1. Using the Taylor expansion of $\nabla P_n\{\hat{\boldsymbol{b}}_{1S}(n)\}$ at $\boldsymbol{b}_{1S}(n)$, we have

$$\nabla P_n\{\hat{\boldsymbol{b}}_{1S}(n)\} = \nabla P_n\{\boldsymbol{b}_{1S}(n)\} + \nabla^2 P_n(\boldsymbol{b}^*)[\hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n)],$$

where $\boldsymbol{b}^*$ is a point between $\hat{\boldsymbol{b}}_{1S}(n)$ and $\boldsymbol{b}_{1S}(n)$. Thus, we have

$$\begin{aligned} \hat{\boldsymbol{b}}_{1S}(n) - \boldsymbol{b}_{1S}(n) &= -(\nabla^2 P_n(\boldsymbol{b}^*))^{-1}\nabla P_n\{\boldsymbol{b}_{1S}(n)\} \\ &= (\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n))^{-1}\boldsymbol{Z}_{1S}^T(n)[\boldsymbol{\epsilon}_1(n) + \boldsymbol{Z}_{1N}(n)\boldsymbol{b}_{1N}(n)], \end{aligned}$$

where $\boldsymbol{Z}_{1N}(n)$ and $\boldsymbol{Z}_{1S}(n)$ are sub-matrices of $\boldsymbol{Z}_1(n)$ corresponding to the coefficients in $\boldsymbol{b}_{1N}(n)$ and $\boldsymbol{b}_{1S}(n)$, respectively. Recall that $\boldsymbol{B}_1(n,t)$ are the B-spline basis functions evaluated at t. Let $\boldsymbol{B}_{1N}(n,t)$ and $\boldsymbol{B}_{1S}(n,t)$ be the partitioning of $\boldsymbol{B}_1(n,t)$ according to $\boldsymbol{b}_{1N}(n)$ and $\boldsymbol{b}_{1S}(n)$.

By Theorem 1, we have $\hat{\mathcal{T}}^{(0)} \cap \mathcal{T}^c \subseteq \Omega(k_{0,n})$, where $\Omega(k_{0,n}) = \{t \in [0,T] : 0 < |\beta(t)| < k_{0,n}^{-r+2}\}$. For $t \in \mathcal{T}^c$, when n is large enough, we have $|\beta(t)| > k_{0,n}^{-r+2}$. Thus, we have that $t \in \hat{\mathcal{T}}^{(0),c}$ when n is large enough. As a results, when n is large enough, we have

$$
\begin{aligned}
& (n/k_{1,n})^{1/2}(\hat{\beta}(t)-\beta(t)) \\
= \;& (n/k_{1,n})^{1/2}\boldsymbol{B}_{1S}^T(n,t)[\hat{\boldsymbol{b}}_{1S}(n)-\boldsymbol{b}_{1S}(n)] + (n/k_{1,n})^{1/2}[\boldsymbol{B}_1^T(n,t)\boldsymbol{b}_1(n)-\beta(t)] \\
= \;& \boldsymbol{B}_{1S}^T(n,t)[(k_{1,n}/n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)]^{-1}\{(n/k_{1,n})^{-1/2}\boldsymbol{Z}_{1S}^T(n)[\boldsymbol{\epsilon}_1(n)+\boldsymbol{Z}_{1N}^T(n)\boldsymbol{b}_{1N}(n)]\} \\
+ \;& (n/k_{1,n})^{1/2}[\boldsymbol{B}_1^T(n,t)\boldsymbol{b}_1(n)-\beta(t)] \\
= \;& \boldsymbol{B}_{1S}^T(n,t)[(k_{1,n}/n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)]^{-1}[(n/k_{1,n})^{-1/2}\boldsymbol{Z}_{1S}^T(n)\boldsymbol{e}(n)] \\
+ \;& \boldsymbol{B}_{1S}^T(n,t)[(k_{1,n}/n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)]^{-1}[(n/k_{1,n})^{-1/2}\boldsymbol{Z}_{1S}^T(n)(\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n))] \\
+ \;& \boldsymbol{B}_{1S}^T(n,t)[(k_{1,n}/n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)]^{-1}[(n/k_{1,n})^{-1/2}\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)\boldsymbol{b}_{1N}(n)]] \\
+ \;& (n/k_{1,n})^{1/2}[\boldsymbol{B}_1^T(n,t)\boldsymbol{b}_1(n)-\beta(t)] \\
= \;& U_n(t) + (n/k_{1,n})^{1/2}\mathcal{B}'{}_n(t) + (n/k_{1,n})^{1/2}\mathcal{B}''_n(t) + (n/k_{1,n})^{1/2}\mathcal{W}_n(t)
\end{aligned}
$$

By Huang (1998), $U_n(t)$ is the variance component, $\mathcal{B}_n(t) = \mathcal{B}'_n(t) + \mathcal{B}''_n(t)$ is the estimation bias, and $W_n(t)$ is the approximation error.

Given that $\boldsymbol{e}(n) \sim N(0, I_n)$, we have that, for $t \in \mathcal{T}^c$,

$$U_n(t) \xrightarrow{\mathcal{D}} N[0, \sigma^2(t)]$$

where $\sigma^2(t) = \lim_{n\to\infty} \boldsymbol{B}_{1S}^T(n,t)[(k_{1,n}/n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1S}(n)]^{-1}\boldsymbol{B}_{1S}(n,t)$.

Given A_8, we have that $\lambda_{max}((k_{1,n}/n)\boldsymbol{Z}_{1S}(n)\boldsymbol{Z}_{1S}^T(n)) \leq c'_2$. As shown above, we have $\sup|\epsilon_{1,i} - e_i| \leq M'Ck_{1,n}^{-r}$ for some constant C. Thus, we have that

$$
\begin{aligned}
& (n/k_{1,n})^{-1}(\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n))^T\boldsymbol{Z}_{1S}(n)\boldsymbol{Z}_{1S}^T(n)(\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n)) \\
= \;& (\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n))^T[(k_{1,n}/n)\boldsymbol{Z}_{1S}(n)\boldsymbol{Z}_{1S}^T(n)](\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n)) \\
\leq \;& c'_2(\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n))^T(\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n)) \\
\leq \;& c'_2(M'C)^2 n k_{1,n}^{-2r}.
\end{aligned}
$$

Thus, we have $||(n/k_{1,n})^{-1/2}\boldsymbol{Z}_{1S}^T(n)(\boldsymbol{\epsilon}_1(n)-\boldsymbol{e}(n))||_{l_2} \le C'n^{1/2}k_{1,n}^{-r}$ for some constant C'. Since $\boldsymbol{B}_{1S}(n,t)$ are bounded and at most h of them are nonzero, given A_8, we have

$$(n/k_{1,n})^{1/2}|\mathcal{B}_n'(t)| = O_p(n^{1/2}k_{1,n}^{-r}).$$

Given A_8, we have

$$\begin{aligned} & (n/k_{1,n})^{-1}\boldsymbol{b}_{1N}^T(n)\boldsymbol{Z}_{1N}^T(n)\boldsymbol{Z}_{1S}(n)\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)\boldsymbol{b}_{1N}(n) \\ \le\; & c_2'^2||\boldsymbol{b}_{1N}(n)||_{l_2}^2 \end{aligned}$$

Given A_5, each coefficient in $\boldsymbol{b}_{1N}(n)$ is bounded by $C'k_{0,n}^{-r+2}$ for some constant C' when n is large enough, as shown in the proof above, and there are a finite number of coefficients in $\boldsymbol{b}_{1N}(n)$. Thus, we obtain that $||\boldsymbol{b}_{1N}(n)||_{l_2}^2 = O_p(k_{0,n}^{-2r+4})$ and $||(n/k_{1,n})^{-1/2}\boldsymbol{Z}_{1S}^T(n)\boldsymbol{Z}_{1N}(n)\boldsymbol{b}_{1N}(n)||_{l_2} = O_p(k_{0,n}^{-r+2})$. Given A_7, we have that $k_{0,n}^{-r+2} = o_p(1)$. Therefore,

$$(n/k_{1,n})^{1/2}|\mathcal{B}_n''(t)| \;=\; o_p(1).$$

Therefore we have

$$(n/k_{1,n})^{1/2}|\mathcal{B}_n(t)| = O_p(n^{1/2}k_{1,n}^{-r}).$$

The term $\mathcal{W}_n(t)$ is the B-spline approximation error at $\beta(t)$. Given A_1 and the B-spline approximation property, we have

$$(n/k_{1,n})^{1/2}|\mathcal{W}_n(t)| \;=\; O_p(n^{1/2}k_{1,n}^{-r-1/2}).$$

Therefore we have, for $t \in \mathcal{T}^c$,

$$(n/k_{1,n})^{1/2}[\hat{\beta}(t) - \beta(t) - \mathcal{B}_n(t) - \mathcal{W}_n(t)] \;\xrightarrow{\mathcal{D}}\; N[0,\sigma^2(t)].$$

Part (ii) is proved.

Assuming the additional stronger condition $n^{-1}k_{1,n}^{2r} \to \infty$ in A_5, it follows that $(n/k_{1,n})^{1/2}|\mathcal{B}_n(t)| = o_p(1)$ and $(n/k_{1,n})^{1/2}|\mathcal{W}_n(t)| = o_p(1)$. Therefore we have, for $t \in \mathcal{T}^c$,

$$(n/k_{1,n})^{1/2}[\hat{\beta}(t) - \beta(t)] \;\xrightarrow{\mathcal{D}}\; N[0,\sigma^2(t)].$$

Part (iii) is proved.

The proof of Theorem 2 is completed. □.

Performance of GCV, AIC, BIC and RIC in Studies 1 and 2:

Table 1: Integrated absolute biases of the least squares, the Dantzig selector, the adaptive LASSO (adpLASSO), and the one-step group SCAD (gSCAD) estimates for Study 1. Each entry is the Monte Carlo average of A_j, $j = 0$ or 1; the corresponding standard deviation is reported in parentheses.

	$\beta_1(t)$		$\beta_2(t)$	
Estimator	A_0	A_1	A_0	A_1
Oracle Estimator	-	0.157 (0.041)	-	0.166 (0.046)
Least Squares	2.205 (1.432)	3.283 (2.549)	1.963 (1.256)	4.088 (2.716)
Dantizig Selector	0.006 (0.013)	0.692 (0.094)	0.006 (0.010)	0.821 (0.132)
adpLASSO GCV	0.039 (0.031)	0.196 (0.059)	0.034 (0.028)	0.218 (0.070)
adpLASSO AIC	0.041 (0.030)	0.193 (0.059)	0.036 (0.028)	0.214 (0.069)
adpLASSO BIC	0.031 (0.031)	0.212 (0.059)	0.025 (0.029)	0.240 (0.074)
adpLASSO RIC	0.030 (0.031)	0.213 (0.059)	0.024 (0.028)	0.241 (0.074)
gSCAD GCV	0.016 (0.026)	0.141 (0.038)	0.015 (0.023)	0.154 (0.046)
gSCAD AIC	0.024 (0.033)	0.143 (0.038)	0.024 (0.030)	0.155 (0.048)
gSCAD BIC	0.004 (0.013)	0.140 (0.037)	0.003 (0.009)	0.154 (0.049)
gSCAD RIC	0.003 (0.011)	0.140 (0.037)	0.002 (0.007)	0.155 (0.049)

Table 2: Null region estimates for Study 1. Each entry is the Monte Carlo average of estimated boundary of the null region; the corresponding standard deviation is reported in parentheses.

	$\beta_1(t)$		$\beta_2(t)$	
Estimator	lower	upper	lower	upper
Dantzig Selector	0.008 (0.064)	6.230 (0.175)	0.002 (0.038)	7.123 (0.202)
gSCAD GCV	0.010 (0.082)	5.926 (0.268)	0.003 (0.051)	6.818 (0.292)
gSCAD AIC	0.011 (0.091)	5.773 (0.479)	0.004 (0.063)	6.666 (0.528)
gSCAD BIC	0.010 (0.082)	6.058 (0.171)	0.003 (0.051)	6.951 (0.181)
gSCAD RIC	0.010 (0.082)	6.067 (0.168)	0.003 (0.051)	6.960 (0.179)

Table 3: Monte Carlo bias, standard deviation (SD), mean squared error (MSE), and empirical coverage probability (CP) of 95% pointwise confidence intervals of group SCAD (gSCAD) estimates for Study 1. Each entry is the average over the selected points in the non-null region of $\beta_1(t)$ or $\beta_2(t)$; the corresponding standard deviation is reported in parentheses.

	$\beta_1(t)$			
Estimator	Ave. MC Bias	Ave. MC SD	Ave. MC MSE	CP
gSCAD GCV	0.003 (0.013)	0.198 (0.213)	0.083 (0.328)	0.932 (0.059)
gSCAD AIC	0.004 (0.013)	0.201 (0.213)	0.085 (0.331)	0.932 (0.047)
gSCAD BIC	-0.001 (0.019)	0.195 (0.218)	0.084 (0.339)	0.928 (0.094)
gSCAD RIC	-0.001 (0.022)	0.194 (0.218)	0.084 (0.338)	0.927 (0.101)
	$\beta_2(t)$			
Estimator	Ave. MC Bias	Ave. MC SD	Ave. MC MSE	CP
gSCAD GCV	-0.007 (0.033)	0.221 (0.244)	0.107 (0.386)	0.925 (0.067)
gSCAD AIC	-0.006 (0.031)	0.224 (0.247)	0.110 (0.394)	0.924 (0.053)
gSCAD BIC	-0.012 (0.043)	0.221 (0.242)	0.107 (0.378)	0.915 (0.098)
gSCAD RIC	-0.013 (0.044)	0.221 (0.242)	0.107 (0.379)	0.912 (0.105)

Table 4: Integrated absolute biases of the least squares, the Dantzig selector, the adaptive LASSO (adpLASSO), and the one-step group SCAD (gSCAD) estimates for Study 2. Each entry is the Monte Carlo average of A_j, $j = 0$ or 1; the corresponding standard deviation is reported in parentheses.

Estimator	A_0	A_1
Oracle Estimator	-	0.257 (0.054)
Least Squares	0.246 (0.060)	0.240 (0.054)
Dantzig Selector	0.006 (0.007)	0.485 (0.069)
adpLASSO GCV	0.064 (0.062)	0.246 (0.063)
adpLASSO AIC	0.066 (0.063)	0.246 (0.063)
adpLASSO BIC	0.023 (0.041)	0.278 (0.079)
adpLASSO RIC	0.018 (0.034)	0.288 (0.084)
gSCAD GCV	0.034 (0.071)	0.230 (0.054)
gSCAD AIC	0.038 (0.076)	0.230 (0.054)
gSCAD BIC	0.009 (0.020)	0.226 (0.056)
gSCAD RIC	0.009 (0.019)	0.226 (0.056)

Table 5: Null region estimates for Study 2. Each entry is the Monte Carlo average of estimated boundary of the null region; the corresponding standard deviation is reported in parentheses.

	$[0.000, 0.200]$		$[0.486, 0.771]$	
Estimator	lower	upper	lower	upper
Dantzig Selector	0.001 (0.009)	0.199 (0.016)	0.502 (0.014)	0.749 (0.008)
gSCAD GCV	0.001 (0.009)	0.194 (0.020)	0.507 (0.019)	0.744 (0.015)
gSCAD AIC	0.001 (0.009)	0.194 (0.021)	0.507 (0.019)	0.744 (0.016)
gSCAD BIC	0.001 (0.009)	0.199 (0.016)	0.502 (0.014)	0.749 (0.008)
gSCAD RIC	0.001 (0.009)	0.199 (0.016)	0.502 (0.014)	0.749 (0.008)

Table 6: Monte Carlo bias, standard deviation (SD), mean squared error (MSE), and empirical coverage probability (CP) of 95% pointwise confidence intervals of group SCAD (gSCAD) estimates for Study 2. Each entry is the average over the selected points in the non-null region of $\beta_1(t)$ or $\beta_2(t)$; the corresponding standard deviation is reported in parentheses.

	$\beta_1(t)$			
Estimator	Ave. MC Bias	Ave. MC SD	Ave. MC MSE	CP
gSCAD GCV	-0.013 (0.058)	0.295 (0.174)	0.119 (0.266)	0.951 (0.016)
gSCAD AIC	-0.012 (0.055)	0.296 (0.173)	0.120 (0.265)	0.950 (0.016)
gSCAD BIC	-0.020 (0.072)	0.286 (0.183)	0.120 (0.272)	0.951 (0.020)
gSCAD RIC	-0.020 (0.072)	0.286 (0.183)	0.120 (0.272)	0.951 (0.020)

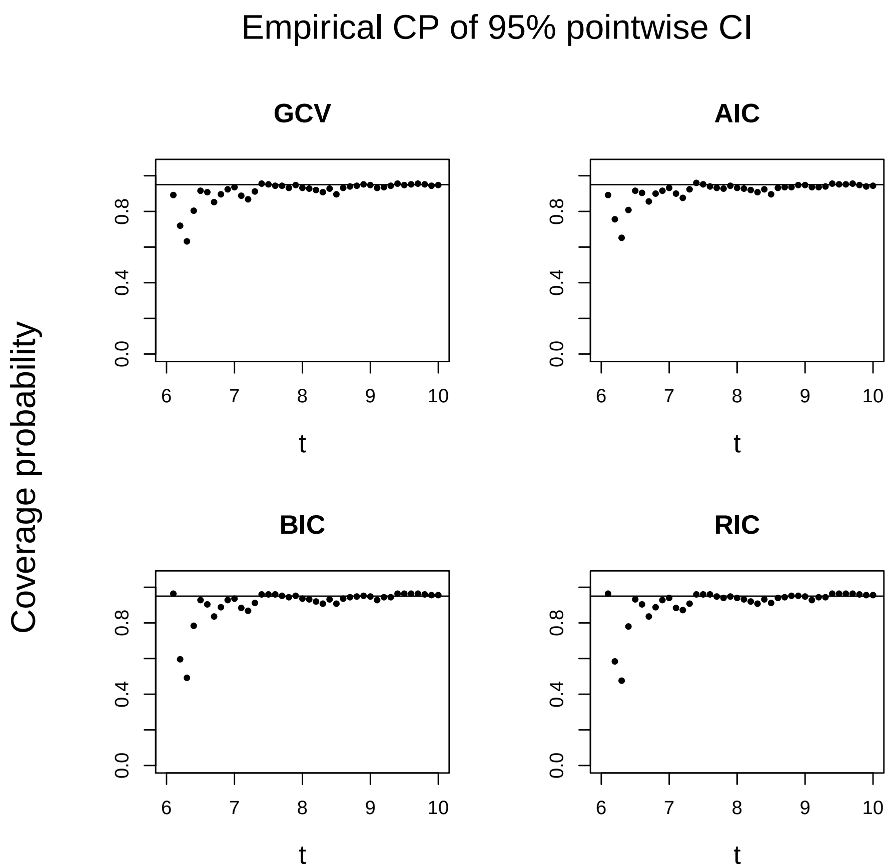

Figure 1: Empirical coverage probabilities (CP) of 95% pointwise confidence intervals for coefficient estimate over non-null region of $\beta_1(t)$ for Study 1, by GCV, AIC, BIC and RIC, respectively. The points are taken at $t = 6.1, 6.2, \cdots, 10.0$.

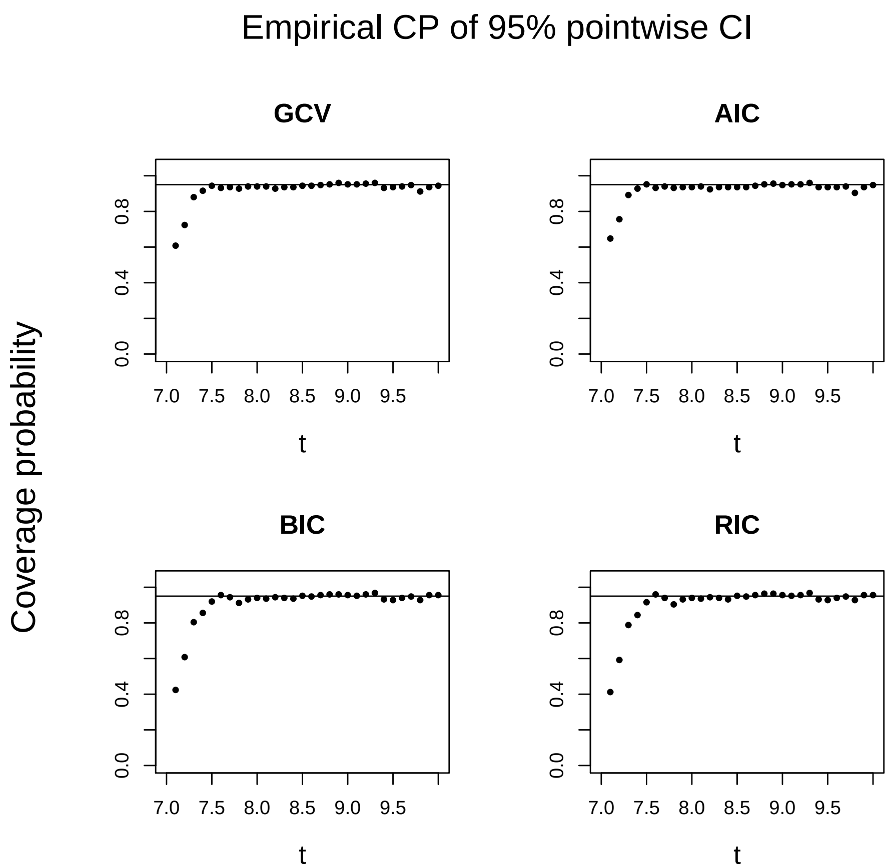

Figure 2: Empirical coverage probabilities (CP) of 95% pointwise confidence intervals for coefficient estimate over non-null region of $\beta_2(t)$ for Study 1, by GCV, AIC, BIC and RIC, respectively. The points are taken at $t = 7.1, 7.2, \cdots, 10.0$.

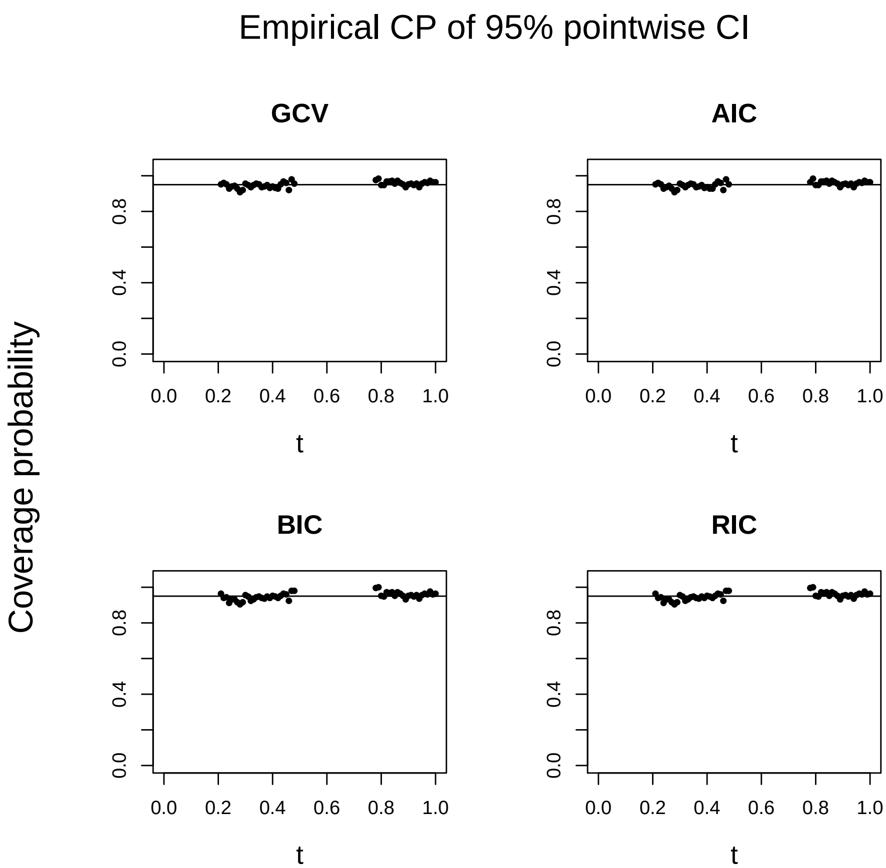

Figure 3: Empirical coverage probabilities (CP) of 95% pointwise confidence intervals for coefficient estimate over non-null region of $\beta(t)$ for Study 2, by GCV, AIC, BIC and RIC, respectively. The points are taken at $t = 0.21, 0.22, \cdots, 0.48, 0.78, 0.79, \cdots, 0.99, 1.00$..