Algorithm for the Refinement Stage

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

Knots Placement. Denote the initial estimate of $\mathcal{T}$ by $\hat{T}^{(0)} = \bigcup_{j=1}^{J} [a_j, c_j]$, 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 $\hat{T}^{(0,c)}$, 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 $A$.

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 $A$ and the boundaries of working null regions $\mathcal{T}_k$ for $k = 0, \ldots, l$. Using this set of knots, compute the variables in the approximate model (2.4).

WN.4 Get the initial value $\tilde{b}_l$ by least squares, and divide $\tilde{b}_l$ into $\tilde{b}_{1N,l}$ and $\tilde{b}_{1S,l}$ according to their association to $\mathcal{T}_l$. 
WN.5 Estimate $b_1$ by minimizing $Q_n(T_l, \lambda, b)$ by LARS algorithm, where $\lambda$ is selected by the criterion $C(T_l, \lambda)$ 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 $C(T_l, \lambda)$ 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

\[
\begin{align*}
GCV(T_l, \lambda) &= \frac{RSS}{n \{1 - d(\lambda)/n\}^2}, \\
AIC(T_l, \lambda) &= n \log(RSS/n) + 2d(\lambda), \\
BIC(T_l, \lambda) &= n \log(RSS/n) + \log(n)d(\lambda), \\
RIC(T_l, \lambda) &= \{n - d(\lambda)\} \log(\hat{\sigma}^2) + d(\lambda)\{\log(n) - 1\} + 4/\{n - d(\lambda) - 2\},
\end{align*}
\]

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 $\lambda$, and $\hat{\sigma}^2 = RSS/\{n - d(\lambda)\}$.

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

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

FD.2 Let $\hat{b}_1 = \arg\min_b Q_n(\hat{T}, \arg\min_{\lambda>0} C(\hat{T}, \lambda), b)$. The refined estimate of $\beta(t)$ is $\hat{\beta}(t) = B_1(t)\hat{b}_1$, where $B_1(t)$ are the B-spline basis function generated in Step 2.3 using the knots in $A$ and the boundaries of working null regions $T_k$ 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.
We first prove the convergence rate of the initial estimator \( \tilde{\mathbf{b}}_1(n) \):

**Proof of the convergence rate of the initial estimator by least squares**

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

Define \( \epsilon_1(n) = (\epsilon_{1,1}, ..., \epsilon_{1,n})^T \) and \( \mathbf{e}(n) = (\epsilon_1, ..., \epsilon_n)^T \). Let \( L_n\{\mathbf{b}(n)\} = \sum_{i=1}^{n}(Y_i - z_{1,i,n})^2 \). Given \( \tilde{\mathbf{b}}_1(n) \) is the minimizer of \( L_n\{\mathbf{b}(n)\} \), we have

\[
L_n\{\tilde{\mathbf{b}}_1(n)\} - L_n\{\mathbf{b}_1(n)\} = \left[ \tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n) \right]^T Z_1^T(n) Z_1(n) [\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)] - 2(Z_1^T(n) \epsilon_1(n)) [\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)] \leq 0.
\]

Given \( A_8 \), we have \( \left[ \tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n) \right]^T Z_1^T(n) Z_1(n) [\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)] \geq c_1(k_1,n)\|\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)\|_{I_2}^2 \). Since the approximation error \( \epsilon_1(t) \) is bounded below \( Ck_1^{-T} \) in absolute value for some constant \( C \), \( A_2 \) ensures that \( \sup\|\epsilon_{1,i} - \epsilon_i\| \leq M'Ck_1^{-1} \).

Thus, the term \( \|Z_1^T(n) \epsilon_1(n)\|_{I_2} = \|Z_1^T(n) \mathbf{e}(n) + Z_1^T(n)(\epsilon_1(n) - \mathbf{e}(n))\|_{I_2} \) is dominated by \( \|Z_1^T(n) \mathbf{e}(n)\|_{I_2} \). Given \( \mathbf{e}(n) \sim N(0, I_n) \), we have \( n^{-1/2}(Z_1^T(n) \mathbf{e}(n)) \sim N(0, n^{-1}Z_1^T(n) Z_1(n)) \), which indicates \( (n^{-1}Z_1^T(n) Z_1(n))^{-1/2} n^{-1/2} (Z_1^T(n) \mathbf{e}(n)) \sim N(0, I_{k_1,n+h}) \), where \( h+1 \) is the B-spline basis function order. Therefore we have

\[
\|Z_1^T(n) \epsilon_1(n)\|_{I_2} = O_p(n^{1/2}).
\] (1)

Therefore,

\[
c_1(k_1,n)\|\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)\|_{I_2}^2 \\
\leq \left[ \tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n) \right]^T Z_1^T(n) Z_1(n) [\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)] \\
\leq 2(Z_1^T(n) \epsilon_1(n))^T [\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)] \\
\leq 2\|Z_1^T(n) \epsilon_1(n)\|_{I_2} \|\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)\|_{I_2} \\
= O_p(n^{1/2}) \|\tilde{\mathbf{b}}_1(n) - \mathbf{b}_1(n)\|_{I_2}.
\]
which indicates $||\hat{b}_1(n) - b_1(n)||_2 = O_p(n^{-1/2}k_{1,n})$. □

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 $b_{0,j}(n), \ldots, b_{0,j+h}(n)$. 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 $|e_0(t)| \leq c k_{0,n}^{-r}$ is the approximation error. Given the B-spline basis functions are all bounded between 0 and 1, we have that

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

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 $\hat{b}_{0,j}(n)$ 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 $\infty$, 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 $k_{0,n} \rightarrow \infty$. Part (iii) is proved.

Proof of Theorem 2: First we prove that $||\hat{b}_1(n) - b_1(n)||_2 \leq O_p(n^{-1/2}k^{3/2}_{1,n})$. This is a non-optimal bound for the convergence rate of $\hat{b}_1(n)$, 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. $|b_{1,j}(n)| \leq C k_{1,n}^{-r}$ 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
Lemma 1, given \( k_{0,n} \) are evenly-spaced on \( \Omega(k_{0,n}) \), 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 \( b_{1N}(n) \) and \( b_{1S}(n) \) are the division of \( b_1(n) \) according to \( \hat{T}^{(0)} \). Since \( b_{1N}(n) \) contains the coefficients associated with \( \hat{T}^{(0)} \), 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 \( b_{1N}(n) \) according to our method to place the \( k_{1,n} \) knots. Thus, given \( A_5 \), we have that \( \|b_{1N}(n)\|_{l_1} = O_p(k_{0,n}^{-r+2}) \). 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 \( b_{1S}(n) \) that is greater than \( M/2(h+1) \) in absolute value, where \( h+1 \) is the fixed spline order. Thus, \( \|b_{1S}(n)\|_{l_1} \geq O_p(1) \).

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

\[
0 \geq Q_n \{\hat{b}_1(n)\} - Q_n \{b_1(n)\} = [\hat{b}_1(n) - b_1(n)]^T \mathbf{Z}_1^T(n) \mathbf{Z}_1(n) [\hat{b}_1(n) - b_1(n)] - 2(\mathbf{Z}_1^T(n) \mathbf{e}_1(n))^T [\hat{b}_1(n) - b_1(n)] + n\lambda_n(\|\hat{b}_{1N}(n)\|_{l_1} - \|b_{1N}(n)\|_{l_1})
\]

\[
\geq [\hat{b}_1(n) - b_1(n)]^T \mathbf{Z}_1^T(n) \mathbf{Z}_1(n) [\hat{b}_1(n) - b_1(n)] - 2(\mathbf{Z}_1^T(n) \mathbf{e}_1(n))^T [\hat{b}_1(n) - b_1(n)] + n\lambda_n(\|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_1} - 2\|b_{1N}(n)\|_{l_1}),
\]

where \( \hat{b}_{1N}(n), \hat{b}_{1S}(n) \) and \( b_{1N}(n), b_{1S}(n) \) are the divisions of \( \hat{b}_1(n) \) and \( b_1(n) \), respectively, according to their association with \( \hat{T}^{(0)} \).

We first show that \( \|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2} \leq O_p(n^{-1/2}k_{1,n}^{3/2}) \). Suppose that this is not true and that \( \|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2} > O_p(n^{-1/2}k_{1,n}^{3/2}) \), which indicates \( \|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_1} > O_p(n^{-1/2}k_{1,n}^{3/2}) \). Since \( \|b_{1N}(n)\|_{l_1} = O_p(k_{0,n}^{-r+2}) \), given \( A_5 \), we have \( \|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_1} - 2\|b_{1N}(n)\|_{l_1} > 0 \) with probability tending to 1.
Given $Q_n\{\hat{b}_1(n)\} - Q_n\{b_1(n)\} \leq 0$ and $A_8$, we have, with probability tending to,

$$c_1'(n/k_{1,n})\|\hat{b}_1(n) - b_1(n)\|_{l_2}^2$$

$$\leq \|\hat{b}_1(n) - b_1(n)\|_{l_2}^T \Z^T_{1}(n) \Z_1(n) \|\hat{b}_1(n) - b_1(n)\|_{l_2}$$

$$\leq 2(\Z_{1}(n) e_1(n))\|\hat{b}_1(n) - b_1(n)\|_{l_2}$$

$$\leq 2\|\Z^T_{1}(n) e_1(n)\|_{l_2} \|\hat{b}_1(n) - b_1(n)\|_{l_2}.$$

Given (1), we have $\|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2} \leq \|\hat{b}_1(n) - b_1(n)\|_{l_2} = O_p(n^{-1/2}k_{1,n}^3)$, which is contradictory to the assumption $\|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2} > O_p(n^{-1/2}k_{1,n}^3)$. Therefore we have

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

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

$$Q_{n,S}\{b_S(n)\} = Q_n\{b(n)\} b_{N}(n) - \hat{b}_{1N}(n).$$

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

Therefore, when $n$ is large,

$$0 \geq Q_{n,S}\{\hat{b}_{1S}(n)\} - Q_{n,S}\{b_{1S}(n)\}$$

$$= \|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}^T \Z^T_{1S}(n) \Z_{1S}(n) \|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}$$

$$- 2(\Z^T_{1S}(n) e_1(n) - \Z^T_{1S}(n) \Z_{1N}(n) (\hat{b}_{1N}(n) - b_{1N}(n)))\|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}.$$

Given $A_8$, we have

$$c_1'(n/k_{1,n})\|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}^2$$

$$\leq \|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}^T \Z^T_{1S}(n) \Z_{1S}(n) \|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}$$

$$\leq 2\|\Z^T_{1S}(n) e_1(n) - \Z^T_{1S}(n) \Z_{1N}(n) (\hat{b}_{1N}(n) - b_{1N}(n))\|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}$$

$$\leq 2\|\Z^T_{1S}(n) e_1(n)\|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2} + \|\Z^T_{1S}(n) \Z_{1N}(n) (\hat{b}_{1N}(n) - b_{1N}(n))\|\hat{b}_{1S}(n) - b_{1S}(n)\|_{l_2}$$

Following the steps to show (1), we obtain that $\|\Z^T_{1S}(n) e_1(n)\|_{l_2} = O_p(n^{1/2})$.

Since $\|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2} = O_p(n^{-1/2}k_{1,n}^3)$, given $A_8$, we have

$$\|\Z^T_{1S}(n) \Z_{1N}(n) (\hat{b}_{1N}(n) - b_{1N}(n))\|_{l_2}^2$$

$$= \|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2}^T \Z^T_{1N}(n) \Z_{1S}(n) \Z^T_{1S}(n) \Z_{1N}(n) (\hat{b}_{1N}(n) - b_{1N}(n))$$

$$\leq c_3(n/k_{1,n})\|\hat{b}_{1N}(n) - b_{1N}(n)\|_{l_2}^2$$

$$= O_p(k_{1,n}^3).$$
Thus, we have \[ ||Z_{1S}(n)Z_{1N}(n)(\hat{b}_{1N}(n) - b_{1N}(n))||_{l_2} = O_p(k_{1,n}), \]
and
\[ ||\hat{b}_{1S}(n) - b_{1S}(n)||_{l_2} \leq O_p(n^{-1/2}k_{1,n}^{3/2}). \]  

(3)

Given (2) and (3), we have
\[ ||\hat{b}_1(n) - b_1(n)||_{l_2} \leq O_p(n^{-1/2}k_{1,n}^{3/2}). \]

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 \( \hat{f}^{(0)} \). We take the partial derivative of \( Q_n(b(n)) \) at \( b(n) = b_1(n) \) with respect to \( b_{1,j}(n) \) in \( b_{1N}(n) \). As shown above, we have \( p_{\lambda_n}'(||\hat{b}_{1N}(n)||_{l_1}) = \lambda_n \) and \( p_{\lambda_n}'(||\hat{b}_{1S}(n)||_{l_1}) = 0 \) with probability tending to 1. The partial derivative is then
\[
\frac{\partial Q_n(b(n))}{\partial b_j(n)} \bigg|_{b(n) = b_1(n)} = \sum_{i=1}^{n} 2[Y_i - z_{1,i}\hat{b}_1(n)](-z_{1,i,j}) + n\lambda_n \text{sign}\hat{b}_{1,j}(n) \\
= \sum_{i=1}^{n} 2\{Y_i - z_{1,i}b_1(n) + z_{1,i}[b_1(n) - \hat{b}_1(n)]\}(-z_{1,i,j}) + n\lambda_n \text{sign}\hat{b}_{1,j}(n) \\
= -2Z_{1,j}^T(n)e_1(n) + 2[b_1(n) - \hat{b}_1(n)]^T Z_{1,j}^T(n)Z_{1,j}(n) + n\lambda_n \text{sign}\hat{b}_{1,j}(n) \\
= -I - II + III,
\]

where \( Z_{1,j}(n) \) is the \( j \)th column of the matrix \( Z_1(n) \).

Given \( A_2 \) and the uniformly bounded B-spline approximation error, we have \( \sup |e_{1,i} - e_i| \leq M'Ck_{1,n}^{-1} \) for some constant \( C \). Thus, the term \( Z_{1,j}(n)e_1(n) \) is dominated by \( Z_{1,j}(n)e_n \). Since \( e(n) \sim N(0, I_n) \), we have
\[
(k_{1,n}/n)^{1/2}Z_{1,j}^T(n)e_1(n) \sim N[0,(k_{1,n}/n)Z_{1,j}^T(n)Z_{1,j}(n)].
\]

Given \( A_8 \), we know that \( (k_{1,n}/n)Z_{1,j}^T(n)Z_{1,j}(n) \) is between the constants \( c_1' \) and \( c_2' \). Therefore,
\[
(k_{1,n}/n)^{1/2}I = N[0,(k_{1,n}/n)Z_{1,j}^T(n)Z_{1,j}(n)] + o_p(1).
\]
By $A_8$, we have $\|Z^T_1(n)Z_{1,j}(n)\|_{l_2} = O_p(nk_{1,n}^{-1})$. Thus, we have

\[
| (k_{1,n}/n)^{1/2} II | \leq 2 (k_{1,n}/n)^{1/2} \| \hat{b}_1(n) - b_1(n) \|_{l_2} \| Z^T_1(n)Z_{1,j}(n) \|_{l_2} \\
= 2 (k_{1,n}/n)^{1/2} O_p(n^{-1/2}k_{1,n}^{3/2}) O_p(nk_{1,n}^{-1}) \\
= O_p(k_{1,n}).
\]

We also have

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

Since $Q_n(b(n))$ minimizes at $\hat{b}_1(n)$, 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 $\hat{\beta}(0)$. Since $\mathcal{T} \subseteq \hat{\mathcal{T}}(0)$, with probability tending to 1, as shown in Theorem 1, we have that $\hat{\beta}(t) = 0$ for $t \in \mathcal{T}$ with probability tending to 1. Part (i) is proved.

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

\[ P_n(b') = \sum_{i=1}^n (Y_i - z_{1S,i}b')^2, \]

where $z_{1S,i}$ are the elements of $z_{1,i}$ that correspond to the coefficients in $b_S(n)$.

With probability tending to 1, $\hat{b}_{1N}(n) = 0$ and $p_{\lambda_n} (||\hat{b}_{1S}(n)||_{l_2}) = 0$ as shown above. Since $\hat{b}_1(n)$ minimizes $Q_n\{b(n)\}$, we know that $\hat{b}_{1S}(n)$ is the minimizer of $P_n(b')$ and $\nabla P_n(\hat{b}_{1S}(n)) = 0$, with probability tending to 1. Using the Taylor expansion of $\nabla P_n(\hat{b}_{1S}(n))$ at $\hat{b}_{1S}(n)$, we have

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

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

\[ \hat{b}_{1S}(n) - b_{1S}(n) = - (\nabla^2 P_n(b^*))^{-1} \nabla P_n(\hat{b}_{1S}(n)) \\
= (Z^T_{1S}(n)Z_{1S}(n))^{-1}Z^T_{1S}(n)\epsilon_1(n) + Z_{1N}(n)b_{1N}(n), \]
where $Z_{1N}(n)$ and $Z_{1S}(n)$ are sub-matrices of $Z_1(n)$ corresponding to the coefficients in $b_{1N}(n)$ and $b_{1S}(n)$, respectively. Recall that $B_1(n,t)$ are the B-spline basis functions evaluated at $t$. Let $B_{1N}(n,t)$ and $B_{1S}(n,t)$ be the partitioning of $B_1(n,t)$ according to $b_{1N}(n)$ and $b_{1S}(n)$.

By Theorem 1, we have $\mathcal{T}^{(0)} \cap \mathcal{F} \subseteq \Omega(k_0,n)$, where $\Omega(k_0,n) = \{ t \in [0,T] : 0 < |\beta(t)| < k_0^{-r+2} \}$. For $t \in \mathcal{F}^c$, when $n$ is large enough, we have $|\beta(t)| > k_0^{-r+2}$. Thus, we have that $t \in \mathcal{T}^{(0)}$. When $n$ is large enough. As a result, when $n$ is large enough, we have

$$
(n/k_1, n)^{1/2}(\hat{\beta}(t) - \beta(t))
$$

$$
= (n/k_1, n)^{1/2}B_S^T(n,t)[b_{1S}(n) - b_{1S}(n)] + (n/k_1, n)^{1/2}[B_{1N}^T(n,t)b_{1N}(n) - \beta(t)]
$$

$$
= B_S^T(n,t)[(k_1/n)Z_{1S}(n)Z_{1S}(n)]^{-1}[(n/k_1, n)^{-1/2}Z_{1S}(n)e_1(n) + Z_{1N}(n)b_{1N}(n)]
$$

$$
+ (n/k_1, n)^{1/2}[B_{1N}^T(n,t)b_{1N}(n) - \beta(t)]
$$

$$
= U(n) + (n/k_1, n)^{1/2}B_S'(n,t) + (n/k_1, n)^{1/2}B_{1N}'(n,t) + (n/k_1, n)^{1/2}W_n(t)
$$

By Huang (1998), $U(n)$ is the variance component, $B_S(n) = B'_S(n) + B_{1N}'(n)$ is the approximation error, and $W_n(t)$ is the approximation error.

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

$$
U_n(t) \overset{d}{\rightarrow} N[0, \sigma^2(t)]
$$

where $\sigma^2(t) = \lim_{n \rightarrow \infty} B_{1S}^T(n,t)[(k_1/n)Z_{1S}(n)Z_{1S}(n)]^{-1}B_{1S}(n,t)$.

Given $A_8$, we have that $\lambda_{max}((k_1/n)Z_{1S}(n)Z_{1S}^T(n)) \leq c_1^2$. As shown above, we have sup $|e_{1,i} - e_i| \leq M'Ck_1^{-r}$ for some constant $C$. Thus, we have that

$$
(n/k_1, n)^{-1}(e_1(n) - e(n))^TZ_{1S}(n)Z_{1S}^T(n)(e_1(n) - e(n))
$$

$$
= (e_1(n) - e(n))^T[(k_1/n)Z_{1S}(n)Z_{1S}^T(n)](e_1(n) - e(n))
$$

$$
\leq c_1^2(e_1(n) - e(n))^T(e_1(n) - e(n))
$$

$$
\leq c_1^2(M'C)^2nk_1^{-2r}.$$

Thus, we have \(|(n/k_{1,n})^{-1/2}Z_{1S}(n)(e_1(n) - e(n))||_2 \leq C''n^{1/2}k_{1,n}^{-r}\) for some constant \(C''\). Since \(B_{1S}(n, t)\) are bounded and at most \(h\) of them are nonzero, given \(A_8\), we have

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

Given \(A_8\), we have

\[
(n/k_{1,n})^{-1}b_{1N}(n)Z^T_{11}(n)Z_{1S}(n)Z^T_{12}(n)Z_{1N}(n)b_{1N}(n)
\leq c_2^2||b_{1N}(n)||_2^2
\]

Given \(A_5\), each coefficient in \(b_{1N}(n)\) 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 \(b_{1N}(n)\). Thus, we obtain that \(||b_{1N}(n)||_2^2 = O_p(k_{0,n}^{-2r+4})\) and \(||(n/k_{1,n})^{-1/2}Z^T_{1S}(n)Z_{1N}(n)b_{1N}(n)||_2 = O_p(k_{0,n}^{-r+2})\). Given \(A_7\), we have that \(k_{0,n}^{-r+2} = o_p(1)\). Therefore,

\[
(n/k_{1,n})^{1/2}|B''(t)| = o_p(1).
\]

Therefore we have

\[
(n/k_{1,n})^{1/2}|B_n(t)| = O_p(n^{1/2}k_{1,n}^{-r}).
\]

The term \(W_n(t)\) is the B-spline approximation error at \(\beta(t)\). Given \(A_1\) and the B-spline approximation property, we have

\[
(n/k_{1,n})^{1/2}|W_n(t)| = O_p(n^{1/2}k_{1,n}^{-r-1/2}).
\]

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

\[
(n/k_{1,n})^{1/2}[\hat{\beta}(t) - \beta(t) - B_n(t) - W_n(t)] \overset{D}{\to} N[0, \sigma^2(t)].
\]

Part (ii) is proved.

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

\[
(n/k_{1,n})^{1/2}[\hat{\beta}(t) - \beta(t)] \overset{D}{\to} N[0, \sigma^2(t)].
\]

Part (iii) is proved.

The proof of Theorem 2 is completed. \(\Box\).
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.

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

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

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.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Ave. MC Bias</th>
<th>Ave. MC SD</th>
<th>Ave. MC MSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1(t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gSCAD GCV</td>
<td>0.003 (0.013)</td>
<td>0.198 (0.213)</td>
<td>0.083 (0.328)</td>
<td>0.932 (0.059)</td>
</tr>
<tr>
<td>gSCAD AIC</td>
<td>0.004 (0.013)</td>
<td>0.201 (0.213)</td>
<td>0.085 (0.331)</td>
<td>0.932 (0.047)</td>
</tr>
<tr>
<td>gSCAD BIC</td>
<td>-0.001 (0.019)</td>
<td>0.195 (0.218)</td>
<td>0.084 (0.339)</td>
<td>0.928 (0.094)</td>
</tr>
<tr>
<td>gSCAD RIC</td>
<td>-0.001 (0.022)</td>
<td>0.194 (0.218)</td>
<td>0.084 (0.338)</td>
<td>0.927 (0.101)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Ave. MC Bias</th>
<th>Ave. MC SD</th>
<th>Ave. MC MSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2(t)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gSCAD GCV</td>
<td>-0.007 (0.033)</td>
<td>0.221 (0.244)</td>
<td>0.107 (0.386)</td>
<td>0.925 (0.067)</td>
</tr>
<tr>
<td>gSCAD AIC</td>
<td>-0.006 (0.031)</td>
<td>0.224 (0.247)</td>
<td>0.110 (0.394)</td>
<td>0.924 (0.053)</td>
</tr>
<tr>
<td>gSCAD BIC</td>
<td>-0.012 (0.043)</td>
<td>0.221 (0.242)</td>
<td>0.107 (0.378)</td>
<td>0.915 (0.098)</td>
</tr>
<tr>
<td>gSCAD RIC</td>
<td>-0.013 (0.044)</td>
<td>0.221 (0.242)</td>
<td>0.107 (0.379)</td>
<td>0.912 (0.105)</td>
</tr>
</tbody>
</table>
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.

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

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.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>lower</th>
<th>upper</th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dantzig Selector</td>
<td>0.001 (0.009)</td>
<td>0.199 (0.016)</td>
<td>0.502 (0.014)</td>
<td>0.749 (0.008)</td>
</tr>
<tr>
<td>gSCAD GCV</td>
<td>0.001 (0.009)</td>
<td>0.194 (0.020)</td>
<td>0.507 (0.019)</td>
<td>0.744 (0.015)</td>
</tr>
<tr>
<td>gSCAD AIC</td>
<td>0.001 (0.009)</td>
<td>0.194 (0.021)</td>
<td>0.507 (0.019)</td>
<td>0.744 (0.016)</td>
</tr>
<tr>
<td>gSCAD BIC</td>
<td>0.001 (0.009)</td>
<td>0.199 (0.016)</td>
<td>0.502 (0.014)</td>
<td>0.749 (0.008)</td>
</tr>
<tr>
<td>gSCAD RIC</td>
<td>0.001 (0.009)</td>
<td>0.199 (0.016)</td>
<td>0.502 (0.014)</td>
<td>0.749 (0.008)</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\beta_1(t)$</th>
<th>Ave. MC Bias</th>
<th>Ave. MC SD</th>
<th>Ave. MC MSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>gSCAD GCV</td>
<td></td>
<td>-0.013 (0.058)</td>
<td>0.295 (0.174)</td>
<td>0.119 (0.266)</td>
<td>0.951 (0.016)</td>
</tr>
<tr>
<td>gSCAD AIC</td>
<td></td>
<td>-0.012 (0.055)</td>
<td>0.296 (0.173)</td>
<td>0.120 (0.265)</td>
<td>0.950 (0.016)</td>
</tr>
<tr>
<td>gSCAD BIC</td>
<td></td>
<td>-0.020 (0.072)</td>
<td>0.286 (0.183)</td>
<td>0.120 (0.272)</td>
<td>0.951 (0.020)</td>
</tr>
<tr>
<td>gSCAD RIC</td>
<td></td>
<td>-0.020 (0.072)</td>
<td>0.286 (0.183)</td>
<td>0.120 (0.272)</td>
<td>0.951 (0.020)</td>
</tr>
</tbody>
</table>
Empirical CP of 95% pointwise CI

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 $t = 0.21, 0.22, \cdots, 0.48, 0.78, 0.79, \cdots, 0.99, 1.00$. 

Empirical CP of 95% pointwise CI

- **GCV**
- **AIC**
- **BIC**
- **RIC**