*Biometrika* (201X), **00**, 0, *pp*. 1–16
Advance Access publication on XX XXXX 201X


# Supplementary Material for "Scalar-on-Image Regression via the Soft-Thresholded Gaussian Process"

BY JIAN KANG

*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.*
jiankang@umich.edu

BRIAN J. REICH AND ANA-MARIA STAICU

*Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.*
bjreich@ncsu.edu astaicu@ncsu.edu

SUMMARY

This Supplementary Material is organized as follows. Sections 1, 2, 4 and 9 present the proofs for Lemmas 1, 3, 4 and 5. Sections 3, 5–8 show the proofs for Lemmas A1–A5. Sections 10 and 11 present proofs for Theorems 2 and 3. Section 12 presents Conditions S1–S3, Lemmas S1–S4 and the proof for Theorem 4. Section 13 includes Definition S1 and Theorem S1 which establishes the theoretical properties of model representations. Section 14 shows the additional simulation results in Tables S1 and S2.

## 1. PROOF OF LEMMA 1

*Proof.* Consider three cases:
In Case 1: $|x_1| \leq \lambda$ and $|x_2| \leq \lambda$,

$$|g_\lambda(x_1) - g_\lambda(x_2)| = 0 \leq |x_1 - x_2|.$$

In Case 2: $\max\{|x_1|, |x_2|\} > \lambda$ and $\min\{|x_1|, |x_2|\} \leq \lambda$,

$$\begin{aligned}&|g_\lambda(x_1) - g_\lambda(x_2)| \\ &= \max\{|x_1|, |x_2|\} - \lambda \leq \max\{|x_1|, |x_2|\} - \min\{|x_1|, |x_2|\} \leq |x_1 - x_2|.\end{aligned}$$

In Case 3: $|x_1| > \lambda$ and $|x_2| > \lambda$,

$$|g_\lambda(x_1) - g_\lambda(x_2)| = |\mathrm{sgn}(x_1)(|x_1| - \lambda) - \mathrm{sgn}(x_2)(|x_2| - \lambda)|.$$

When $\mathrm{sgn}(x_1)\mathrm{sgn}(x_2) = 1$,

$$|\mathrm{sgn}(x_1)(|x_1| - \lambda) - \mathrm{sgn}(x_2)(|x_2| - \lambda)| = |(|x_1| - \lambda) - (|x_2| - \lambda)| \leq |x_1 - x_2|.$$

When $\mathrm{sgn}(x_1)\mathrm{sgn}(x_2) = -1$,

$$\begin{aligned}&|\mathrm{sgn}(x_1)(|x_1| - \lambda) - \mathrm{sgn}(x_2)(|x_2| - \lambda)| \\ &= |(|x_1| - \lambda) + (|x_2| - \lambda)| = |x_1| + |x_2| - 2\lambda = |x_1 - x_2| - 2\lambda \leq |x_1 - x_2|.\end{aligned}$$ □

## 2. PROOF OF LEMMA 3

*Proof.* Note that

$$\pi_{n,i}(Z_{n,i};\beta) = \pi(Y_i \mid W_i, X_i, \alpha_0^{\mathrm{v}}, \beta, \sigma_0)\pi(X_i)\pi(W_i),$$

where $Y_i$ follows model (1). Let $\widetilde{Y}_i = Y_i - W_i^{\mathrm{T}}\alpha_0^{\mathrm{v}}$. Then $\widetilde{Y}_i \sim \mathrm{N}\left\{p_n^{-1/2}\sum_{j=1}^{p_n}\beta(s_j)X_{i,j}, \sigma_0^2\right\}$. It is straightforward to see that

$$\begin{aligned}
&\Lambda_{n,i}(Z_{n,i};\beta_0,\beta)\\
&= \frac{1}{2\sigma_0^2}\left[\left\{\widetilde{Y}_i - p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\beta(s_j)\right\}^2 - \left\{\widetilde{Y}_i - p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\beta_0(s_j)\right\}^2\right]\\
&= \frac{1}{2\sigma_0^2}\left(\left[p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\left\{\beta_0(s_j)-\beta(s_j)\right\}\right]\left[2\widetilde{Y}_i - p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\left\{\beta_0(s_j)+\beta(s_j)\right\}\right]\right).
\end{aligned}$$

Given any $\zeta > 0$, let $B(\zeta) = \{\beta(s) : \sup_{s\in\mathcal{B}}|\beta(s)-\beta_0(s)| < \zeta\}$, then by Condition A3.2, for any $\beta \in B$,

$$\begin{aligned}
&K_{n,i}(\beta_0,\beta)\\
&= \frac{1}{2\sigma_0^2}E\left(\left[\sum_{j=1}^{p_n}p_n^{-1/2}X_{i,j}\left\{\beta_0(s_j)-\beta(s_j)\right\}\right]E\left[2\widetilde{Y}_i - \sum_{j=1}^{p_n}p_n^{-1/2}X_{i,j}E\left\{\beta_0(s_j)+\beta(s_j)\right\}\,\middle|\, X_i\right]\right)\\
&= \frac{1}{2\sigma_0^2}E\left[p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\left\{\beta_0(s_j)-\beta(s_j)\right\}\right]^2 \le \frac{c_{\max}\zeta^2}{2\sigma_0^2},
\end{aligned}$$

and

$$\begin{aligned}
&V_{n,i}(\beta_0,\beta)\\
&= \frac{1}{4\sigma_0^4}E\left(\left[\sum_{j=1}^{p_n}p_n^{-1/2}X_{i,j}\left\{\beta_0(s_j)-\beta(s_j)\right\}\right]^2\mathrm{var}\left[2\widetilde{Y}_i - \sum_{j=1}^{p_n}p_n^{-1/2}X_{i,j}\left\{\beta_0(s_j)+\beta(s_j)\right\}\,\middle|\, X_i\right]\right)\\
&= \frac{1}{\sigma_0^2}E\left[p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\left\{\beta_0(s_j)-\beta(s_j)\right\}\right]^2 \le \frac{c_{\max}\zeta^2}{\sigma_0^2}.
\end{aligned}$$

Thus,

$$\frac{1}{n^2}\sum_{i=1}^{n}V_{i,n}(\beta_0,\beta) \le \frac{1}{n^2}\sum_{i=1}^{n}\frac{c_{\max}\zeta^2}{\sigma_0^2} \to 0,\ n\to\infty.$$

Consider $\beta \in B(\zeta)$, then $n^{-1}\sum_{i=1}^{n}K_{i,n}(\beta_0,\beta) \le c_{\max}(2\sigma_0^2)^{-1}\left\{\sup_{s\in\mathcal{B}}|\beta(s)-\beta_0(s)|\right\}^2$. By Theorem 1, for any $0 < \varepsilon < \zeta^2$,

$$\Pi\left[B(\zeta)\cap\left\{\frac{1}{n}\sum_{i=1}^{n}K_{i,n}(\beta_0,\beta) < \varepsilon\right\}\right] \ge \Pi\left\{\sup_{s\in B}|\beta(s)-\beta_0(s)| < \frac{2\sigma_0^2\sqrt{\varepsilon}}{c_{\max}}\right\} > 0. \qquad \square$$

## 3. PROOF OF LEMMA A1

*Proof.* Define

$$\Theta_{n,k} = \left\{ \beta(s) \in \mathcal{C}^{\rho}(\overline{\mathcal{R}}_k) : \sup_{s \in \mathcal{R}_k} |D^{\tau}\beta(s)| < p_n^{1/(2d)}, 0 \le \|\tau\|_1 \le \rho \right\},$$

for $k = -1, 0, 1$. By Lemma 2 of Ghosal & Roy (2006), there exists $c_k > 0$ such that

$$N(\varepsilon, \Theta_{n,k}, \|\cdot\|_{\infty}) \le c_k p_n^{1/(2\rho)} \varepsilon^{-d/\rho}.$$

Since

$$\Theta_n \subseteq \Theta_{n,-1} \times \Theta_{n,0} \times \Theta_{n,1}.$$

Then

$$N(\varepsilon, \Theta_n, \|\cdot\|_{\infty}) \le \prod_{k=-1}^{1} N_k(\varepsilon, \Theta_n, \|\cdot\|_{\infty}).$$

and

$$\log N(\varepsilon, \Theta_n, \|\cdot\|_{\infty}) \le \sum_{k=-1}^{1} \log N(\varepsilon, \Theta_{n,k}, \|\cdot\|_{\infty}) \le C p_n^{1/(2\rho)} \varepsilon^{-d/\rho},$$

where $C = c_{-1} + c_0 + c_1$. □

## 4. PROOF OF LEMMA 4

*Proof.* By (2), we have $\beta(s) = g_{\lambda_0}\{\tilde{\beta}^{\mathrm{v}}(s)\}$ with $\tilde{\beta}^{\mathrm{v}}(s) \sim \mathcal{GP}(0, \kappa)$. Let $\mathcal{R}_1 = \{s : \beta(s) > 0\}$ and $\mathcal{R}_{-1} = \{s : \beta(s) < 0\}$. Then $\tilde{\beta}^{\mathrm{v}}(s) = \beta(s) + \lambda_0$ for $s \in \mathcal{R}_1$ and $\tilde{\beta}^{\mathrm{v}}(s) = \beta(s) - \lambda_0$ for $s \in \mathcal{R}_{-1}$, For any $\tau$ with $0 < \|\tau\|_1 \le \rho$, we have $D^{\tau}\tilde{\beta}^{\mathrm{v}}(s) = D^{\tau}\beta(s)$ for any $s \in \mathcal{B}$. Consider

$$\Pi[\Theta_n^C] \le \Pi\left\{ \sup_{s \in \mathcal{R}_1} |\tilde{\beta}^{\mathrm{v}}(s) - \lambda_0| > p_n^{1/(2d)} \right\} + \Pi\left\{ \sup_{s \in \mathcal{R}_{-1}} |\tilde{\beta}^{\mathrm{v}}(s) + \lambda_0| > p_n^{1/(2d)} \right\}$$
$$+ \sum_{\tau : 1 \le \|\tau\|_1 \le \rho} \Pi\left[ \sup_{s \in \mathcal{R}_1 \cup \mathcal{R}_{-1}} |D^{\tau}\tilde{\beta}^{\mathrm{v}}(s)| > p_n^{1/(2d)} \right].$$

Note that $\tilde{\beta}^{\mathrm{v}}(s) + \lambda_0 \sim \mathcal{GP}(\lambda_0, \kappa)$ and $\tilde{\beta}^{\mathrm{v}}(s) - \lambda_0 \sim \mathcal{GP}(-\lambda_0, \kappa)$. By Theorem 5 of Ghosal & Roy (2006), there exist positive constants $K_{-1}$, $K_1$, $K_{\tau}$, $b_{-1}$, $b_1$ and $b_{\tau}$ such that,

$$\Pi\left\{ \sup_{s \in \mathcal{R}_{-1}} |\tilde{\beta}^{\mathrm{v}}(s) + \lambda_0| > p_n^{1/(2d)} \right\} \le K_{-1} \exp(-b_{-1}\rho^{1/d}),$$
$$\Pi\left\{ \sup_{s \in \mathcal{R}_1} |\tilde{\beta}^{\mathrm{v}}(s) + \lambda_0| > p_n^{1/(2d)} \right\} \le K_1 \exp(-b_1\rho^{1/d}),$$
$$\Pi\left\{ \sup_{s \in \mathcal{R}_1 \cup \mathcal{R}_{-1}} |D^{\tau}\tilde{\beta}^{\mathrm{v}}(s)| > p_n^{1/(2d)} \right\} \le K_{\tau} \exp(-b_{\tau}\rho^{1/d}).$$

□

Take $K = K_{-1} + K_1 + \sum_{\tau : 0 < |\tau| \le \rho} K_{\tau}$ and $b = \min\{b_{-1}, b_1, \min_{1 \le |\tau| \le \rho} b_{\tau}\}$.

## 5. PROOF OF LEMMA A2

*Proof.* Given that $\gamma$ is continuous, $\mathcal{V}_{\varepsilon,\gamma}$ is an open set and there are a countable collection of disjoint open regions, denoted $\mathcal{V}_1, \mathcal{V}_2, \ldots$, such that

$$\mathcal{V}_{\varepsilon,\gamma} = \bigcup_{k=1}^{\infty} \mathcal{V}_k,$$

where open region $\mathcal{V}_k$ with volume $\lambda(\mathcal{V}_k) > 0$. By Condition 3, we have $\mathcal{B} = \bigcup_{j=1}^{p_n} \mathcal{B}_j$. For $j = 1, \ldots, p_n$, define $A_j = \{k : s_j \in \mathcal{V}_k\}$, $\mathcal{U}_j = \bigcup_{k \in A_j} \mathcal{V}_k$ and $\mathcal{U} = \bigcup_{j=1}^{p_n} \mathcal{U}_j$. Then $\mathcal{U} \supset \mathcal{V}_{\varepsilon,\gamma}$. It covers the same set of spatial locations $\{s_j\}_{j=1}^{p_n}$ as $\mathcal{V}_{\varepsilon,\gamma}$ and $\lambda(\mathcal{U}) \geq \lambda(\mathcal{V}_{\varepsilon,\gamma})$. Also, given $n, \mathcal{U}$ is equal to the union of a finite number of disjoint open regions $\mathcal{V}_k$, rewrite them as $\mathcal{W}_1, \ldots, \mathcal{W}_m$. Then $\mathcal{U} = \bigcup_{k=1}^{m} \mathcal{W}_k$. For each $k = 1, \ldots, m$, $\mathcal{W}_k$ contains at least a number of $\lfloor \lambda(\mathcal{W}_k) K p_n \rfloor$ spatial locations since the maximum volume of the disjoints $\mathcal{B}_j$ which contains and only contains $s_j$'s is smaller than $1/(Kp_n)$ by Condition A4. Let $s_k^*$ be the spatial location that is most close to $\mathcal{W}_k$ among all $s_j$'s not in $\mathcal{U}$. Since $\mathcal{D}$ is dense in $\mathcal{B}$, there exist $\gamma(s)$ differentiable points in $\mathcal{W}_k$, and let $t_k$ be one of such points, then $\|s_k^* - t_k\|_\infty < \{1/(Kp_n)\}^{1/d}$. By Taylor expansion, we have

$$|\gamma(s_k^*)| + \left| \frac{\partial \gamma}{\partial s}(t_k)(s_k^* - t_k) \right| + |R_2(s_k^*, t_k)| \geq |\gamma(t_k)| > \varepsilon.$$

Thus,

$$|\gamma(s_k^*)| > \varepsilon - \left\{ \frac{p_n^{1/(2d)} + \upsilon}{d(Kp_n)^{1/d}} + \frac{p_n^{1/d} + \upsilon}{2d^2(Kp_n)^{2/d}} \right\}.$$

Let $N$ be a large integer such that for all $n > N$,

$$\frac{p_n^{1/(2d)} + \upsilon}{d(Kp_n)^{1/d}} + \frac{p_n^{1/d} + \upsilon}{2d^2(Kp_n)^{2/d}} < \frac{\varepsilon}{2}.$$

Therefore, when $n > N$, $|\gamma(s_k^*)| > \varepsilon/2$. Let $\mathcal{W}_k^* = \mathcal{W}_k \cup \{s_k^*\}$. Then $\mathcal{W}_k^*$ contains at least a number of $\lfloor \lambda(\mathcal{W}_k) K p_n \rfloor + 1$ or $\lceil \lambda(\mathcal{W}_k) K p_n \rceil$ spatial locations $s_j$'s such that $|\gamma(s_j)| > \varepsilon/2$. Since none of spatial locations were counted more than once, by adding over all $k$, at least $\lambda(\mathcal{V}_{\varepsilon,\gamma}) K p_n$ spatial locations fall in $\mathcal{V}_{\varepsilon/2,\gamma}$. □

## 6. PROOF OF LEMMA A3

*Proof.* Let $\upsilon = \max\left\{ \sup_{\beta_0 \in \mathcal{R}_1 \cup \mathcal{R}_{-1}} |D^\tau \beta_0|, 0 \leq \|\tau\|_1 \leq 1 \right\}$. Let $0 < \delta < \varepsilon/\lambda(\mathcal{B})$ and $\mathcal{D}_i = \{s : (i-1)\delta < |\beta(s) - \beta_0(s)| < i\delta\}$, then

$$\sum_{i=1}^{\lceil \{p_n^{1/(2d)} + \upsilon\}/\delta \rceil} i\delta \lambda(\mathcal{D}_i) \geq \|\beta - \beta_0\|_1 > \varepsilon. \tag{S1}$$

Let $\ell(s) = |\beta(s) - \beta_0(s)|$ and $\ell_k(s) = \min\{k\delta, \ell(s)\}$, for $k = 0, \ldots, p_n$. Note that $\ell_k(s) = \ell(s)$ when $k = \lceil \{p_n^{1/(2d)} + \upsilon\}/\delta \rceil$ in that $\ell(s) \leq \|\beta\|_\infty + \|\beta_0\|_\infty < p_n^{1/2d} + \upsilon$.

For $k = 1, \ldots, p_n$, define $\mathcal{E}_k = \{s : \ell_k(s) > (2k-1)\delta/2\}$. Note that $\ell_{k-1} \leq (k-1)\delta$. Then for all $s \in \mathcal{E}_k$, we have $\ell_k(s) - \ell_{k-1}(s) > (2k-1)\delta/2 - (k-1)\delta = \delta/2$. Note that $\beta(s)$ and $\beta_0(s)$ are both in $\Theta$, the differentiable points for $\ell_k(s) - \ell_{k-1}(s)$ are dense in $\mathcal{B}$. By Lemma 5, there exists $N$ such that for all $n > N$,

$$\sum_{j=1}^{p_n} \{\ell_k(s_j) - \ell_{k-1}(s_j)\} > \frac{\lambda(\mathcal{E}_k) K \delta p_n}{4},$$

Write

$$\sum_{j=1}^{p_n} |\beta(s_j) - \beta_0(s_j)| = \sum_{j=1}^{p_n} \ell_{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil}(s_j)$$
$$= \sum_{j=1}^{p_n} \sum_{k=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \{\ell_k(s_j) - \ell_{k-1}(s_j)\} \geq \sum_{k=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \frac{\lambda(\mathcal{E}_k) K \delta p_n}{4}.$$

In addition, for $k = 1, \ldots, p_n$,

$$\mathcal{E}_k \supset \bigcup_{i=k+1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \mathcal{D}_i,$$

which implies that

$$\lambda(\mathcal{E}_k) \geq \sum_{i=k+1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \lambda(\mathcal{D}_i)$$
$$\sum_{k=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \lambda(\mathcal{E}_k) \geq \sum_{k=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \sum_{i=k+1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \lambda(\mathcal{D}_i)$$
$$= \sum_{i=2}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} (i-1)\lambda(\mathcal{D}_i)$$
$$= \sum_{i=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} i\lambda(\mathcal{D}_i) - \sum_{i=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \lambda(\mathcal{D}_i)$$
$$\geq \sum_{i=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} i\lambda(\mathcal{D}_i) - \lambda(\mathcal{B})$$

By (S1),

$$\sum_{k=1}^{\lceil \{p_n^{1/(2d)}+v\}/\delta \rceil} \delta\lambda(\mathcal{E}_k) \geq \varepsilon - \lambda(\mathcal{B})\delta.$$

Taking $r = K\{\varepsilon - \lambda(\mathcal{B})\delta\}/4$ completes the proof. □

## 7. PROOF OF LEMMA A4

*Proof.* Let

$$B_{n,\varepsilon} = \left\{ (x_1, \ldots, x_{p_n})^{\mathrm{T}} \;\middle|\; p_n^{-1/2} \left| \sum_{j=1}^{p_n} x_j \{\beta(s_j) - \beta_0(s_j)\} \right| > \varepsilon \right\}.$$

Recall that $X_i = (X_{i,1}, \ldots, X_{i,p_n})^{\mathrm{T}}$ are independent copies of $X(s_1), \ldots, X(s_{p_n})$. Then $\Pi[X_i \in B_{n,\varepsilon}] > \varepsilon$ for all $i = 1, \ldots, n$. Let $W_{n,\varepsilon} = n - \sum_{i=1}^{n} I(X_i \in B_{n,\varepsilon})$ and notice that $W_{n,\varepsilon}$ follows an binomial distribution with number of trails $n$ and probability $1 - \Pi(X_i \in B_{n,\varepsilon})$. Let $U_{n,\varepsilon}$ have a binomial distribution with number of trails $n$ and probability $1 - \varepsilon$. Then $U_{n,\varepsilon}$ is stochastically dominates $W_{n,\varepsilon}$. For any $t > 0$,

$$\begin{aligned}
\Pi\left(A_n^C\right) &= \Pi\left[\sum_{i=1}^{n} p_n^{-1/2} \left| \sum_{j=1}^{p_n} X_{i,j} \{\beta(s_j) - \beta_0(s_j)\} \right| < nr \right] \\
&\leq \Pi\left[\varepsilon \left\{ \sum_{i=1}^{n} I(X_i \in B_{n,\varepsilon}) \right\} < nr \right] = \Pi\left\{W_{n,\varepsilon} > n(1 - r/\varepsilon)\right\} \\
&\leq \Pi\left\{U_{n,\varepsilon} > n(1 - r/\varepsilon)\right\} \\
&= \Pi\left[\exp(tU_{n,\varepsilon}) > \exp\{tn(1 - r/\varepsilon)\}\right] \\
&\leq \{\varepsilon + (1 - \varepsilon)\exp(t)\}^n \exp\{-tn(1 - r/\varepsilon)\}.
\end{aligned}$$

Take $t = \log\{(\varepsilon^2 - r\varepsilon)/(r - r\varepsilon)\}$ and $D = t\varepsilon/(\varepsilon - r)$,

$$\Pi\left[A_n^C\right] \leq \frac{\varepsilon^2}{r} \exp(-Dn).$$

Applying Borel–Cantelli lemma completes the proof. □

## 8. PROOF OF LEMMA A5

*Proof.* Condition A4.3 implies that, for any $r > 0$, there exists $\delta > 0$, if $\beta_1$ and $\beta_0$ satisfy with

$$p_n^{-1} \sum_{i=1}^{p_n} |\beta_1(s_j) - \beta_0(s_j)| > r,$$

then

$$\Pi\left[p_n^{-1/2} \left| \sum_{j=1}^{p_n} X(s_j)\{\beta_1(s_j) - \beta_0(s_j)\} \right| > \delta \right] > \delta.$$

By Lemma A4, for any $0 < r_0 < \delta^2$, with probability one that there exists $N_0 > 1$ such that for all $n > N_0$,

$$\sum_{i=1}^{n} |\eta_{i,1} - \eta_{i,0}| \geq nr_0.$$

Now consider,

$$E_{\beta_0}\{\Psi_n(\beta_0,\beta_1)\} = \Pi_{\beta_0}\left\{\sum_{i=1}^{n}\delta_i\left(\frac{Y_i-\eta_{i,0}}{\sigma_0}\right) > 2n^{\tau+1/2}\right\}$$
$$= \Pi_{\beta_0}\left\{n^{-1/2}\sum_{i=1}^{n}\delta_i\left(\frac{Y_i-\eta_{i,0}}{\sigma_0}\right) > 2n^{\tau}\right\}$$
$$= 1-\Phi(2n^{\tau}) \le \frac{\phi(2n^{\tau})}{2n^{\tau}} = 2^{-1}(2\pi)^{-1/2}\frac{\exp(-2n^{2\tau})}{n^{\tau}}.$$

Let

$$\eta_i = \sum_{k=1}^{q}\alpha_{0,k}W_{i,k} + p_n^{-1/2}\sum_{j=1}^{p_n}X_{i,j}\beta(s_j).$$

By Condition A3·2, for any $\beta$ such that $\|\beta-\beta_1\|_\infty \le r_0/(4c_{\max}^{1/2})$,

$$|\eta_i-\eta_{i,1}| = p_n^{-1/2}\left|\sum_{j=1}^{p_n}X_{i,j}\{\beta(s_j)-\beta_1(s_j)\}\right| \le c_{\max}^{1/2}\frac{r_0}{4c_{\max}^{1/2}} = \frac{r_0}{4}, \quad i=1,\ldots,n.$$

Then

$$E_{\beta}\{1-\Psi_n(\beta_0,\beta_1)\} = \Pi_{\beta}\left\{\sum_{i=1}^{n}\delta_i\left(\frac{Y_i-\eta_{i,0}}{\sigma_0}\right) \le 2n^{\tau+1/2}\right\}$$
$$= \Pi_{\beta}\left\{n^{-1/2}\sum_{i=1}^{n}\delta_i\left(\frac{Y_i-\eta_i}{\sigma_0}\right) + n^{-1/2}\sum_{i=1}^{n}\delta_i\left(\frac{\eta_i-\eta_{i,1}}{\sigma_0}\right) + n^{-1/2}\sum_{i=1}^{n}\left|\frac{\eta_{i,1}-\eta_{i,0}}{\sigma_0}\right| \le 2n^{\tau}\right\}$$
$$\le \Pi_{\beta}\left\{n^{-1/2}\sum_{i=1}^{n}\delta_i\left(\frac{Y_i-\eta_i}{\sigma_0}\right) \le \frac{r_0n^{1/2}}{4\sigma_0} - \frac{r_0n^{1/2}}{\sigma_0} + 2n^{\tau}\right\}.$$

There exists $N > N_0$ such that for all $n > N$, $n^{\tau} < n^{1/2}r_0(4\sigma_0)^{-1}$ since $\tau < 1/2$. This further implies that

$$E_{\beta}\{1-\Psi_n(\beta_0,\beta_1)\} \le \Pi\left\{n^{-1/2}\sum_{i=1}^{n}\delta_i\left(\frac{Y_i-\eta_i}{\sigma_0}\right) \le -\frac{r_0n^{1/2}}{4\sigma_0}\right\}$$
$$= \Phi\left(-\frac{r_0n^{1/2}}{4\sigma_0}\right) \le \frac{4\sigma_0}{r_0(2\pi n)^{1/2}}\exp\left(-\frac{nr_0^2}{32\sigma_0^2}\right).$$

Thus, taking $C_0 = \max\left\{2^{-1}(2\pi)^{-1/2}, 4\sigma_0 r_0^{-1}(2\pi)^{-1/2}\right\}$ and $C_1 = r_0^2/(32\sigma_0^2)$ completes the proof. □

## 9. PROOF OF LEMMA 5

*Proof.* Let $r_0$ be the same number in Lemma A5. Let $t = \min\{4^{-1}c_{\max}^{-1/2}r_0, \varepsilon/2\}$. Let $N_t$ be the $t$ covering number of $\Theta_n$ in the supremum norm. Let $\beta^1,\ldots,\beta^{N_t} \in \Theta_n$ be such that for each $\beta \in \Theta_n$ there exist at least one $l$ such that $\|\beta-\beta^l\|_\infty < t$. For any $\beta \in \Theta_n$, define

$$\Psi_n = \max_{1\le l\le N_t}\Psi_n(\beta_0,\beta^l).$$

If $\|\beta-\beta_0\|_\infty > \varepsilon$, then $\|\beta^l-\beta_0\|_\infty > \varepsilon/2$ for $\beta^l$ that satisfies with $\|\beta-\beta^l\|_\infty < t \le \varepsilon/2$. By Lemma A3, there exist $N_0^*$ and $r>0$ such that $\sum_{j=1}^{p_n} |\beta^l(s_j)-\beta_0(s_j)| > rp_n$. By Lemma A5, $r_0$ can be chosen such that for $l=1,\ldots,N_t$, if $\|\beta-\beta^l\|_\infty < t \le 4^{-1}c_{\max}^{-1/2}$, for all $n>N_0^*$,

$$E_{\beta_0}\{\Psi_n(\beta_0,\beta^l)\} \le C_0\exp(-2n^{2\upsilon}),$$
$$E_{\beta}\{1-\Psi_n(\beta_0,\beta^l)\} \le C_0\exp(-C_1 n),$$

where $\tau$, $C_0$ and $C_1$ are the same values in Lemma A5. Furthermore,

$$\begin{aligned}
E_{\beta_0}(\Psi_n) &\le \sum_{l=1}^{N_t}\Psi_n(\beta_0,\beta^l) \le C_0 N_t \exp(-2n^{2\upsilon}) = C_0\exp(\log N_t - 2n^{2\upsilon}) \\
&\le C_0 \exp\left\{Cp_n^{1/(2\rho)}t^{-d/\rho} - 2n^{2\upsilon}\right\} \\
&\le C_0 \exp\left(Cn^{\upsilon_0}t^{-d/\rho} - 2n^{2\upsilon}\right) \\
&= C_0 \exp\left\{-(2-Cn^{\upsilon_0-2\upsilon}t^{-d/\rho})n^{2\upsilon}\right\}.
\end{aligned}$$

When $Ct^{-d/\rho} < 2$,

$$E_{\beta_0}(\Psi_n) \le C_0\exp\left\{-(2-Ct^{-d/\rho})n^{2\upsilon}\right\}.$$

When $Ct^{-d/\rho} \ge 2$, since $\upsilon_0 - 2\upsilon < 0$, there exists $N_1^*$ such that for all $n > N_1^*$, $Cn^{\upsilon_0-2\upsilon}t^{-d/\rho} < 1$, then

$$E_{\beta_0}(\Psi_n) \le C_0\exp\left\{-n^{2\upsilon}\right\}.$$

In addition,

$$E_\beta(1-\Psi_n) = E_\beta\left[\min_{1\le l\le N_t}\{1-\Psi_n(\beta_0,\beta^l)\}\right] \le E_\beta\left[\{1-\Psi_n(\beta_0,\beta^l)\}\right] \le C_0\exp(-C_1 n).$$

Thus, taking $C_2 = (2-Ct^{-d/\rho})I(Ct^{-d/\rho}<2) + I(Ct^{-d/\rho}\ge 2) > 0$ and $N = \max\{N_1^*, N_0^*\}$ completes the proof. □

## 10. PROOF OF THEOREM 2

*Proof.* The proof can be done by verifying the conditions in Theorem A·1 of Choudhuri et al. (2004). Specifically, we have the condition on prior positivity of neighborhoods by Lemma 3. By Lemma 4, Lemma A5 and Condition A1, as $n\to\infty$,

$$\begin{aligned}
E_{\beta_0}(\Psi_n) &\to 0, \\
\sup_{\beta\in\mathcal{U}_\varepsilon^C\cap\Theta_n} E_\beta(1-\Psi_n) &\le C_0\exp(-C_1 n), \\
\Pi(\Theta_n^C) &\le K\exp(-bp_n^{1/d}) \le K\exp(-C_3 n).
\end{aligned}$$

Thus, conditions on the existence of tests are established. □

## 11. PROOF OF THEOREM 3

*Proof.* Define $\mathcal{U}_\varepsilon = \{\beta\in\Theta : \|\beta-\beta_0\|_1 < \varepsilon\}$. Let $\mathcal{R}_0 = \{s : \beta_0(s) = 0\}$, $\mathcal{R}_1 = \{s : \beta_0(s) > 0\}$ and $\mathcal{R}_{-1.} = \{s : \beta_0(s) < 0\}$.

For any $\mathcal{A} \subseteq \mathcal{B}$ and any integer $m \geq 1$, define

$$\mathcal{F}_m(\mathcal{A}) = \left\{ \beta \in \Theta : \int_{\mathcal{A}} |\beta(s) - \beta_0(s)| \mathrm{d}s < \frac{1}{m} \right\}.$$

Then $\mathcal{F}_{m+1}(\mathcal{A}) \subseteq \mathcal{F}_m(\mathcal{A})$ for all $m$ and $\mathcal{F}_m(\mathcal{B}) \subseteq \mathcal{F}_m(\mathcal{A})$.

Consider

$$\mathcal{F}_m(\mathcal{R}_0) = \left\{ \beta \in \Theta : \int_{\mathcal{R}_0} |\beta(s)| \mathrm{d}s < \frac{1}{m} \right\}.$$

By Theorem 2 and the fact that $\mathcal{U}_{1/m} = \mathcal{F}_m(\mathcal{B})$,

$$\Pi\{\mathcal{F}_m(\mathcal{R}_0) \mid D_n\} \geq \Pi(\mathcal{U}_{1/m} \mid D_n) \to 1,\ n \to \infty.$$

In addition,

$$\{\beta(s) = 0,\ \text{for all } s \in \mathcal{R}_0\} = \left\{ \int_{\mathcal{R}_0} |\beta(s)| \mathrm{d}s = 0 \right\} = \bigcap_{m=1}^{\infty} \mathcal{F}_m(\mathcal{R}_0).$$

By the monotone continuity of probability measure,

$$\Pi\left\{\beta(s) = 0,\ \text{for all } s \in \mathcal{R}_0 \mid D_n\right\} = \lim_{m\to\infty} \Pi\{\mathcal{F}_m(\mathcal{R}_0) \mid D_n\} = 1, n \to \infty. \tag{S2}$$

By Condition A2·3, for any $s_0 \in \mathcal{R}_1$ and any integer $m \geq 1$, there exists $\delta_0 > 0$, such that

$$|\beta(s_1) - \beta(s_0)| < \frac{1}{2m},$$

for any $s_1 \in B(s_0, \delta_0) = \{s : \|s_1 - s_0\|_1 < \delta_0\}$. By Definition 2, $\mathcal{R}_1$ is an open set. There exists $\delta_1 > 0$, such that $B(s_0, \delta_1) \subseteq \mathcal{R}_1$. Taking $\delta = \min\{\delta_1, \delta_0\} > 0$, we have that

$$\begin{aligned}
&\left\{ \beta(s_0) > -\frac{1}{m},\ \text{for all } s_0 \in \mathcal{R}_1 \right\} \\
&\supseteq \left\{ \beta(s_0) > \beta(s_1) - \frac{1}{2m} \text{ and } \beta(s_1) > -\frac{1}{2m}, \text{for some } s_1 \in B(s_0, \delta),\ \text{for all } s_0 \in \mathcal{R}_1 \right\} \\
&\supseteq \left\{ \int_{B(s_0,\delta)} \beta(s) ds > -\frac{1}{2m},\ \text{for all } s_0 \in \mathcal{R}_1 \right\} \\
&\supseteq \left\{ \int_{B(s_0,\delta)} \beta(s) ds > \int_{B(s_0,\delta)} \beta_0(s) ds - \frac{1}{2m},\ \text{for all } s_0 \in \mathcal{R}_1 \right\} \\
&\supseteq \mathcal{F}_{2m}[B(s_0, \delta)] \supseteq \mathcal{U}_{1/2m}.
\end{aligned}$$

Thus,

$$\Pi\left\{ \beta(s_0) > -\frac{1}{m},\ \text{for all } s_0 \in \mathcal{R}_1 \mid D_n \right\} \geq \Pi(\mathcal{U}_{1/2m} \mid D_n) \to 1,\ n \to \infty.$$

By the monotone continuity of probability measure,

$$\begin{aligned}
&\Pi\left\{\beta(s) > 0,\ \text{for all } s \in \mathcal{R}_1 \mid D_n\right\} \\
&= \lim_{m\to\infty} \Pi\left\{ \beta(s_0) > -\frac{1}{m},\ \text{for all } s_0 \in \mathcal{R}_1 \mid D_n \right\} \to 1, \quad n \to \infty.
\end{aligned} \tag{S3}$$

Similar arguments can be made to show

$$\Pi\left\{\beta(s)<0,\ \text{for all } s\in\mathcal{R}_{-1}\mid D_n\right\}\to 1,\ n\to\infty. \tag{S4}$$

Combing (S2) - (S4) completes the proof. □

## 12. PROOF OF THEOREM 4

To prove the posterior consistency for model (6) in the main text, we need to introduce the additional conditions and lemmas:

*Condition S*1. Let $\sigma_{\mathcal{B}}^2=\sup_{s\in\mathcal{B}}\operatorname{var}\{X(s)\}<\infty$; and for any $\lambda>0$ and any $p_n\geq 1$,

$$\Pi\left\{p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)\right|>\lambda\right\}\leq 2\exp\left(-\frac{\lambda^2}{2\sigma_{\mathcal{B}}^2}\right).$$

This condition holds for many stochastic processes. For example, $X(s)$ is a centered Gaussian process with uncorrelated covariance structure by Adler (1990).

*Condition S*2. There exists $M_w$, $0<M_w<\infty$ such that $\max_{1\leq i\leq n}\{\max_{1\leq k\leq q}|W_{i,k}|\}<M_w$, for any $n\geq 1$, with probability one.

*Condition S*3. For any $\alpha^{\mathrm{v}}\in\mathbb{R}^q$ and $\beta\in\Theta$, let $\eta=\sum_{k=1}^q\alpha_k W_k+p_n^{-1/2}\sum_{j=1}^{p_n}\beta(s_j)X(s_j)$ and $\pi_\eta(x;\alpha^{\mathrm{v}},\beta)$ denote its density function. There exists a normal density function as the envelop function of $\pi_\eta(x;\alpha^{\mathrm{v}},\beta)$. That is, there exists $M_\pi$, $0<M_\pi<\infty$, $\mu_\pi\in\mathbb{R}$ and $\sigma_\pi>0$, such that

$$\pi_\eta(x;\alpha^{\mathrm{v}},\beta)\leq\frac{M_\pi}{\sigma_\pi}\phi\left(\frac{x-\mu_\pi}{\sigma_\pi}\right),$$

where $M_\pi$, $\mu_\pi$ and $\sigma_\pi$ may depend on $\alpha^{\mathrm{v}}$ and $\beta$.

This condition holds when $W_{i,k}$ is bounded and $X(s)$ is a Gaussian process.

LEMMA S1. *Denote by $\pi_{n,i}(\cdot;\beta)$ the density function of $Z_{n,i}=(Y_i,W_i,X_i)$ in model (6) and suppose Conditions A4, S1–S3 hold for $X_i$. Define*

$$\begin{aligned}\Lambda_{n,i}(\cdot;\beta_0,\beta)&=\log\pi_{n,i}(\cdot;\beta)-\log\pi_{n,i}(\cdot;\beta_0),\\ K_{n,i}(\beta_0,\beta)&=E_{\beta_0}\{\Lambda_{n,i}(Z_{n,i};\beta_0,\beta)\},\\ V_{n,i}(\beta_0,\beta)&=\operatorname{var}_{\beta_0}\{\Lambda_{n,i}(Z_{n,i};\beta_0,\beta)\}.\end{aligned}$$

*There exists a set $B$ with $\Pi(B)>0$ such that, for any $\varepsilon>0$,*

$$\liminf_{n\to\infty}\Pi\left[\left\{\beta\in B, n^{-1}\sum_{i=1}^n K_{n,i}(\beta_0,\beta)<\varepsilon\right\}\right]>0,\quad n^{-2}\sum_{i=1}^n V_{n,i}(\beta_0,\beta)\to 0,\ \textit{for all}\ \beta\in B.$$

*Proof.* Given any $\zeta>0$, let $B=\{\beta(s):\sup_{s\in\mathcal{B}}|\beta(s)-\beta_0(s)|<\zeta\}$. For any $\beta\in B$, let $\eta_i=\sum_{k=1}^q\alpha_k W_{i,k}+p_n^{-1/2}\sum_{j=1}^{p_n}\beta(s_j)X_{i,j}$. Let $\eta_{i,0}=\sum_{k=1}^q\alpha_k W_{i,k}+p_n^{-1/2}\sum_{j=1}^{p_n}\beta_0(s_j)X_{i,j}$. Let $M_\alpha=\max_k|\alpha_k|$ and $M_\beta=\max\{\sup_s|\beta(s)|,\sup_s|\beta_0(s)|\}$. Note that

$$\pi_{n,i}(Z_{n,i};\beta)=\pi(Y_i\mid W_i,X_i,\alpha_0^{\mathrm{v}},\beta)\pi(X_i)\pi(W_i),$$

where $Y_i$ follows model (6). Then $Y_i \sim \text{Bernoulli}\{\Phi(\eta_i)\}$. It is straightforward to see that

$$\pi(Y_i \mid W_i, X_i, \alpha_0^{\mathrm{v}}, \beta) = \Phi(\eta_i)^{Y_i}\{1-\Phi(\eta_i)\}^{1-Y_i},$$
$$\pi(Y_i \mid W_i, X_i, \alpha_0^{\mathrm{v}}, \beta_0) = \Phi(\eta_{i,0})^{Y_i}\{1-\Phi(\eta_{i,0})\}^{1-Y_i}.$$

Then

$$\Lambda_{n,i}(Z_{n,i};\beta_0,\beta) = Y_i \log\left\{\frac{\Phi(\eta_i)}{\Phi(\eta_{i,0})}\right\} + (1-Y_i)\log\left\{\frac{1-\Phi(\eta_i)}{1-\Phi(\eta_{i,0})}\right\}.$$

Then

$$\begin{aligned} K_{n,i}(\beta_0,\beta) &= E\{\Lambda_{n,i}(Z_{n,i};\beta_0,\beta)\} \\ &= E\left[\Phi(\eta_{i,0})\log\left\{\frac{\Phi(\eta_i)}{\Phi(\eta_{i,0})}\right\} + \{1-\Phi(\eta_{i,0})\}\log\left\{\frac{1-\Phi(\eta_i)}{1-\Phi(\eta_{i,0})}\right\}\right]. \end{aligned}$$

Let $g(\eta_{i,0},\eta_i) = \Phi(\eta_{i,0})\log\left\{\frac{\Phi(\eta_i)}{\Phi(\eta_{i,0})}\right\} + \{1-\Phi(\eta_{i,0})\}\log\left\{\frac{1-\Phi(\eta_i)}{1-\Phi(\eta_{i,0})}\right\}$. For any $0 < \varepsilon_0 < 1/2$,

$$\begin{aligned} &K_{n,i}(\beta_0,\beta) \\ &= E\left[g(\eta_{i,0},\eta_i)I\{\max(|\eta_{i,0}|,|\eta_i|) \le -\Phi^{-1}(\varepsilon_0)\}\right] + E\left[g(\eta_{i,0},\eta_i)I\{\max(|\eta_{i,0}|,|\eta_i|) > -\Phi^{-1}(\varepsilon_0)\}\right]. \end{aligned}$$

By Lemma 5 in Ghosal & Roy (2006), there exists $L$ depending only on $\varepsilon_0$, such that

$$\begin{aligned} &E\left[g(\eta_{i,0},\eta_i)I\{\max(|\eta_{i,0}|,|\eta_i|) \le -\Phi^{-1}(\varepsilon_0)\}\right] \\ &\le LE\left[\{\Phi(\eta_i)-\Phi(\eta_{i,0})\}^2 \cdot I\{\max(|\eta_{i,0}|,|\eta_i|) \le -\Phi^{-1}(\varepsilon_0)\}\right] \\ &\le LE\left[\{\Phi(\eta_i)-\Phi(\eta_{i,0})\}^2\right] \\ &< LE(\eta_i-\eta_{i,0})^2 \\ &\le Lc_{\max}\zeta^2. \end{aligned}$$

By Cauchy–Schwarz inequality,

$$\begin{aligned} &E\left[g(\eta_{i,0},\eta_i)I\{\max(|\eta_{i,0}|,|\eta_i|) > -\Phi^{-1}(\varepsilon_0)\}\right] \\ &\le \quad \left[E\left\{g^2(\eta_{i,0},\eta_i)\right\}\Pi\left\{\max(|\eta_{i,0}|,|\eta_i|) > -\Phi^{-1}(\varepsilon_0)\right\}\right]^{1/2}. \end{aligned}$$

where

$$\begin{aligned} &E\{g^2(\eta_{i,0},\eta_i)\} \\ &\le 2E\left[\Phi(\eta_{i,0})^2\log^2\left\{\frac{\Phi(\eta_i)}{\Phi(\eta_{i,0})}\right\} + \{1-\Phi(\eta_{i,0})\}^2\log^2\left\{\frac{1-\Phi(\eta_i)}{1-\Phi(\eta_{i,0})}\right\}\right] \\ &\le 2E\left[\log^2\left\{\frac{\Phi(\eta_i)}{\Phi(\eta_{i,0})}\right\} + \log^2\left\{\frac{1-\Phi(\eta_i)}{1-\Phi(\eta_{i,0})}\right\}\right] \\ &\le 4E\left[\log^2\Phi(\eta_i) + \log^2\Phi(\eta_{i,0}) + \log^2\{1-\Phi(\eta_i)\} + \log^2\{1-\Phi(\eta_{i,0})\}\right]. \end{aligned}$$

By Condition S3, $E\left\{g^2(\eta_{i,0},\eta_i)\right\}$ is bounded. In particular,

$$E\{\log^2\Phi(\eta_i)\} \le M_\pi \int_0^1 \log^2[\Phi\{\sigma_\pi\Phi^{-1}(x)+\mu_\pi\}]dx.$$

By L'Hospital's Rule,

$$\lim_{x\to 0}\frac{\log(x)}{\log[\Phi\{\sigma_\pi\Phi^{-1}(x)+\mu_\pi\}]} = \frac{1}{\sigma_\pi^2},$$

for any $\sigma_\pi > 0$ and $\mu_\pi \in \mathbb{R}$. Since $\int_0^1 \log^2(x)dx = 2$, then $E\{\log^2 \Phi(\eta_i)\} < \infty$. Similarly, we can show that $E\{\log^2\{1-\Phi(\eta_i)\}\} < \infty$, $E\{\log^2 \Phi(\eta_{i,0})\} < \infty$ and $E[\log^2\{1-\Phi(\eta_{i,0})\}] < \infty$.

By Conditions S1 and S2,

$$\lim_{\varepsilon_0 \to 0} \Pi\left\{\max(|\eta_{i,0}|, |\eta_i|) > -\Phi^{-1}(\varepsilon_0)\right\} = 0.$$

Thus, take $\varepsilon_0$ such that,

$$E\left[g(\eta_{i,0}, \eta_i) I\{\max(|\eta_{i,0}|, |\eta_i|) > -\Phi^{-1}(\varepsilon_0)\}\right] < L c_{\max} \zeta^2.$$

Hence, there exists $L > 0$, such that for any $\beta \in B$,

$$n^{-1} \sum_{i=1}^n K_{n,i}(\beta_0, \beta) < 2 c_{\max} L \left\{\sup_{s \in \mathcal{B}} |\beta(s) - \beta_0(s)|\right\}^2.$$

By Theorem 1, for any $0 < \varepsilon < \zeta^2$,

$$\Pi\left[B \cap \left\{n^{-1} \sum_{i=1}^n K_{i,n}(\beta_0, \beta) < \varepsilon\right\}\right] \geq \Pi\left\{\sup_{s \in B} |\beta(s) - \beta(s)| < \varepsilon^{1/2}/(2 L c_{\max})\right\} > 0.$$

In addition,

$$\begin{aligned}
&V_{n,i}(\beta_0, \beta) = \mathrm{var}_{\beta_0}\{\Lambda_{n,i}(Z_{n,i}; \beta_0, \beta)\} \\
&= E\left(\mathrm{var}\left[Y_i \log\left\{\frac{\Phi(\eta_i)}{\Phi(\eta_{i,0})}\right\} + (1-Y_i)\log\left\{\frac{1-\Phi(\eta_i)}{1-\Phi(\eta_{i,0})}\right\} \mid X_i, W_i\right]\right) + \mathrm{var}\left\{g(\eta_{i,0}, \eta_i)\right\} \\
&= E\left(\Phi(\eta_{i,0})\{1-\Phi(\eta_{i,0})\}\left[\log\left\{\frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}\right\} - \log\left\{\frac{\Phi(\eta_{i,0})}{1-\Phi(\eta_{i,0})}\right\}\right]^2\right) + \mathrm{var}\{g(\eta_{i,0}, \eta_i)\} \\
&\leq \frac{1}{2} E\left[\log^2\left\{\frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}\right\} + \log^2\left\{\frac{\Phi(\eta_{i,0})}{1-\Phi(\eta_{i,0})}\right\}\right] + E\{g^2(\eta_{i,0}, \eta_i)\} < \infty. \qquad \square
\end{aligned}$$

Then

$$\lim_{n \to \infty} n^{-2} \sum_{i=1}^n V_{n,i}(\beta_0, \beta) = 0.$$

LEMMA S2. *For any sufficiently small $\varepsilon > 0$ and $0 < r < \varepsilon^2$, let*

$$A_n = \left\{\sum_{i=1}^n |\Phi(\eta_i) - \Phi(\eta_{i,0})| \geq nr\right\},$$

*where $\eta_i = \sum_{k=1}^q \alpha_{0,k} W_{i,k} + p_n^{-1/2} \sum_{j=1}^{p_n} \beta(s_j) X_{i,j}$ and $\eta_{i,0} = \sum_{k=1}^q \alpha_{0,k} W_{i,k} + p_n^{-1/2} \sum_{j=1}^{p_n} \beta_0(s_j) X_{i,j}$ for any $\alpha_0^{\mathrm{v}} = (\alpha_{0,1}, \ldots, \alpha_{0,q})^{\mathrm{T}} \in \mathbb{R}^q$. There exist $N$ and $D > 0$ such that if for all $n > N$ and for all $\beta \in \Theta_n$,*

$$\Pi\left[p_n^{-1/2} \left|\sum_{j=1}^{p_n} X(s_j)\left\{\beta(s_j) - \beta_0(s_j)\right\}\right| > \varepsilon\right] > \varepsilon,$$

*then*

$$\Pi\left(A_n^C\right) \le \frac{\varepsilon^2}{r}\exp(-Dn), \quad \Pi\left(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} A_m\right) = 1.$$

*Proof.* Let $\eta_0 = \sum_{k=1}^{q}\alpha_k W_k + p_n^{-1/2}\sum_{j=1}^{p_n}\beta_0(s_j)X(s_j)$ and $\eta = \sum_{k=1}^{q}\alpha_k W_k + p_n^{-1/2}\sum_{j=1}^{p_n}\beta(s_j)X(s_j)$, where $(W_1,\ldots,W_q)$ has the same joint distribution with $(W_{i,1},\ldots,W_{i,q})$ for $i=1,\ldots,n$. Note that $\eta_i$ and $\eta_{i,0}$ are independent copies of $\eta$ and $\eta_0$, respectively. Let $M_\alpha = \max_k|\alpha_k|$ and $M_\beta = \max\{\sup_s|\beta(s)|, \sup_s|\beta_0(s)|\}$. By the mean value theorem, we have

$$|\Phi(\eta)-\Phi(\eta_0)| = \phi(\xi)p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)\{\beta(s_j)-\beta_0(s_j)\}\right|,$$

where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $\min\{\eta,\eta_0\} < \xi < \max\{\eta,\eta_0\}$. Thus, $\phi(\xi) > \phi(\max\{|\eta|,|\eta_0|\})$. Then by Conditions S1 and S2, for any $\lambda > 0$,

$$\begin{aligned}
&\Pi\{|\Phi(\eta)-\Phi(\eta_0)|\le\varepsilon\}\\
&\le \Pi\left[\phi\{\max(|\eta|,|\eta_0|)\}p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)\{\beta(s_j)-\beta_0(s_j)\}\right| \le \varepsilon\right]\\
&\le \Pi\left(p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)[\beta(s_j)-\beta_0(s_j)]\right| \le \varepsilon(2\pi)^{1/2}\exp\left[\frac{1}{2}\left\{p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)\right|M_\beta + qM_\alpha M_w\right\}^2\right]\right)\\
&\le \Pi\left[p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)\{\beta(s_j)-\beta_0(s_j)\}\right| \le \varepsilon(2\pi)^{1/2}\exp\{1/2(\lambda M_\beta + qM_\alpha M_w)^2\}\right]\\
&\qquad +\Pi\left\{p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)\right| > \lambda\right\}\\
&\le 1-\varepsilon(2\pi)^{1/2}\exp\{1/2(\lambda M_\beta + qM_\alpha M_w)^2\} + \exp\{-\lambda^2/(2\sigma_{\mathcal{B}}^2)\}.
\end{aligned}$$

□

Take $\lambda = \{-2\sigma_{\mathcal{B}}^2\log(\varepsilon)\}^{1/2}$, then we have

$$\Pi\{|\Phi(\eta)-\Phi(\eta_0)|>\varepsilon\} > \varepsilon(2\pi)^{1/2}-\varepsilon > \varepsilon.$$

Applying the similar approach in the proof for Lemma A4 completes the proof.

Suppose $\alpha_0^{\mathrm{v}} = (\alpha_{0,1},\ldots,\alpha_{0,q})^{\mathrm{T}}$ is known. For any $r>0$, we consider $\beta_0$ and $\beta_1$ that satisfy $\sum_{j=1}^{p_n}|\beta_1(s_j)-\beta_0(s_j)| > rp_n$. Let $\eta_{i,m} = \sum_{k=1}^{q}\alpha_{0,k}W_{i,k} + p_n^{-1/2}\sum_{j=1}^{p_n}\beta_m(s_j)X_{i,j}$, for $m=0,1$. Let $\delta_i^+ = I[\Phi(\eta_{i,1}) > \Phi(\eta_{i,0})+r_0]$ and $\delta_i^- = I[\Phi(\eta_{i,1}) < \Phi(\eta_{i,0})-r_0]$. By Condition A4·3, there exists $\delta>0$, $\Pi\left[p_n^{-1/2}\left|\sum_{j=1}^{p_n}X(s_j)(\beta_1(s_j)-\beta_0(s_j))\right| > \delta\right] > \delta$. By Lemma S2, there exists $0<r_0<\delta^2$, $\sum_{i=1}^{n}|\Phi(\eta_{i,1})-\Phi(\eta_{i,0})| > nr_0 > (n+1)\frac{r_0}{2}$ with probability one for sufficiently large $n$. Fixing $n$ and $\{s_j\}_{j=1}^{p_n}$, we can consider $\eta = \eta(X) = \sum_{k=1}^{q}\alpha_{0,k}W_{i,k} + p_n^{-1/2}\sum_{j=1}^{p_n}\beta_m(s_j)X(s_j)$ as a function of $X = [X(s_1),\ldots X(s_{p_n})]$. By Lemma 3 in Ghosal & Roy (2006), there exists $K_0>0$ such that either $n^+ = \sum_{i=1}^{n}\delta_i^+ > K_0 n$ or $n^- = \sum_{i=1}^{n}\delta_i^- > K_0 n$.

LEMMA S3. *Suppose $n^+ > K_0 n$. We construct the test statistic*

$$\Psi_n[\beta_0, \beta_1] = I\left[\sum_{i=1}^n \delta_i^+\{Y_i - \Phi(\eta_{i,0})\} > n^+ r_0/2\right]$$

*for the hypothesis testing problem*

$$H_0 : \beta = \beta_0 \in \Theta, \quad H_1 : \beta = \beta_1 \in \Theta,$$

*such that for any $n > N$, we have*

$$E_{\beta_0}\{\Psi_n(\beta_0, \beta_1)\} \le \exp(-r_0^2 K_0 n/2),$$

*and for any $\beta$ with $\|\beta - \beta_1\|_\infty < r_0/\{4c_{\max}^{1/2}\}$,*

$$E_\beta\{1 - \Psi_n(\beta_0, \beta_1)\} \le \exp(-r_0^2 K_0 n/8),$$

*for some $r_0$ and $N$.*

*Proof.* By Hoeffding's inequality (Hoeffding, 1963, Theorem 1),

$$\begin{aligned} E_{\beta_0}\left\{\Psi_n(\beta_0, \beta_1)\right\} = \Pi_{\beta_0}\left\{\frac{\sum_{i=1}^n \delta_i^+ Y_i}{n^+} - E_{\beta_0}\left(\frac{\sum_{i=1}^n \delta_i^+ Y_i}{n^+}\right) > r_0/2\right\} \\ \le \exp(-n^+ r_0^2/2) \le \exp(-K_0 r_0^2 n/2). \end{aligned}$$

Let

$$\eta_i = \sum_{k=1}^q \alpha_{0,k} W_{i,k} + p_n^{-1/2} \sum_{j=1}^{p_n} X_{i,j}\beta(s_j).$$

By Condition A3·2, for any $\beta$ such that $\|\beta - \beta_1\|_\infty \le r_0 4^{-1} c_{\max}^{-1/2}$, for any $i = 1, \ldots, n$,

$$|\Phi(\eta_i) - \Phi(\eta_{i,1})| \le |\eta_i - \eta_{i,1}| = p_n^{-1/2}\left|\sum_{j=1}^{p_n} X_{i,j}(\beta(s_j) - \beta_1(s_j))\right| \le (c_{\max})^{1/2}\frac{r_0}{4(c_{\max})^{1/2}} = \frac{r_0}{4}.$$

Then

$$\begin{aligned} &E_\beta\{1 - \Psi_n(\beta_0, \beta_1)\} = \Pi_\beta\left[\sum_{i=1}^n \delta_i^+\left\{Y_i - \Phi(\eta_{i,0})\right\} \le n^+ r_0/2\right] \\ &= \Pi_\beta\left[\sum_{i=1}^n \delta_i^+\{Y_i - \Phi(\eta_i)\} + \sum_{i=1}^n \delta_i^+\{\Phi(\eta_i) - \Phi(\eta_{i,1})\} + \sum_{i=1}^n \delta_i^+\{\Phi(\eta_{i,1}) - \Phi(\eta_{i,0})\} \le n^+ r_0/2\right] \\ &\le \Pi_\beta\left[\frac{1}{n^+}\sum_{i=1}^n \delta_i^+\left\{Y_i - \Phi(\eta_i)\right\} \le -r_0/4\right] \le \exp(-n^+ r_0^2/8) \le \exp(-K_0 r_0^2 n/8), \end{aligned}$$

which completes the proof. □

Then applying the similar argument for Lemma 5, we can construct uniform consistent tests for model (6) in the following lemma,

LEMMA S4. *For any $\varepsilon > 0$, there exist $N$, $C > 0$ such that for all $n > N$ and all $\beta \in \Theta_n$, if $\|\beta - \beta_0\|_1 > \varepsilon$, a test function $\Psi_n$ can be constructed such that*

$$E_{\beta_0}(\Psi_n) \le \exp\left(-Cn\right), \quad E_\beta(1 - \Psi_n) \le \exp\left(-Cn\right).$$

Similar to Theorem 2, the proof of the posterior consistency for model (6) is achieved by verifying the conditions in Theorem A.1 of Choudhuri et al. (2004). In particular, we have the conditions on prior positivity of neighborhoods by Lemma S1. To establish the existence of tests, we have the upper bound of the tail probability based on Lemma 4, uniform consistent tests from Lemma S4 and Condition A1. The sign consistency holds due to the nature of the soft-thresholded Gaussian process prior.

## 13. Theoretical Properties for Model Representation

In this section, we discuss theoretical properties for the model representation and prior specifications in Section 4.1 for posterior computation. In particular, when $L \to \infty$,

$$\tilde{\beta}(s) = \int \tilde{K}(\|s-t\|)a(t)dt,$$

where $a \sim \mathcal{GP}(0, \kappa_a)$, $\tilde{K}(\|s-t\|) = K(\|s-t\|)/w(s)$ and

$$w(s) = \left\{ \int \int K(\|s-t\|)\kappa_a(t,t')K(\|s-t'\|)dtdt' \right\}^{1/2}.$$

For any finite $L > 1$ and any knots $\{t_l\}_{l=1}^{L}$, we have $a(t_l) = a_l$ and the covariance kernel $\kappa_a$ is constructed by the conditional autoregressive model (7): $[\kappa_a(t_l, t_{l'})]_{L\times L} = \sigma_a^2(M - \vartheta A)^{-1}$. This implies that $\tilde{\beta} \sim \mathcal{GP}(0, \kappa_\beta)$ with $\kappa_\beta(s,s') = \int\int \tilde{K}(\|s-t\|)\kappa_a(t,t')\tilde{K}(\|s'-t'\|)dtdt'$ and $\kappa_\beta(s,s) = 1$. Of note, to establish large prior support properties in Theorem 1, we need Condition A5. Although $\kappa_\beta$ does not satisfy Condition A5 in general, we still can establish the large support properties for the soft-thresholding prior model for a sub space of the true coefficient functional space $\Theta$. In particular, we need the following definition:

Definition S1. *Define*

$$\bar{\Theta} = \left[ g_\lambda\{\tilde{\beta}(s)\} : \tilde{\beta}(s) = \int \tilde{K}(\|s-t\|)a(t)dt, \lambda > 0, a \in RKHS(\kappa_a) \right],$$

*where $RKHS(\kappa_a)$ is the reproducing kernel Hilbert space of $\kappa_a$; see (2.3) in Ghosal & Roy (2006).*

Now we establish the large support in $\bar{\Theta}$ by the following theorem.

Theorem S1. *For a function $\beta_0 \in \bar{\Theta}$, there exists $\lambda_0 > 0$ such that the soft-thresholded Gaussian process prior $\beta(s) = g_{\lambda_0}\{\tilde{\beta}(s)\}$ with $\tilde{\beta}(s) = \int \tilde{K}(\|s-t\|)a(t)dt$ and $a \sim \mathcal{GP}(0, \kappa_a)$ satisfies*

$$\Pi\left(\|\beta - \beta_0\|_\infty < \varepsilon\right) > 0, \qquad \textit{for all } \varepsilon > 0.$$

*Proof.* Based on the definition of $\Theta$, we have $\bar{\Theta} \subset \Theta$. If $\beta_0 \in \bar{\Theta}$, then there exists a smooth function $a_0(t) \in \mathcal{C}^\rho(\mathcal{B})$ and $\lambda_0 > 0$, such that $\beta_0(s) = g_\lambda\{\tilde{\beta}_0(s)\}$ and $\tilde{\beta}_0(s) = \int \tilde{K}(\|s-$

$t\|)a_0(t)dt$. Then we have

$$
\begin{aligned}
&\Pi\left\{\sup_{s\in\mathcal{B}}|\beta(s)-\beta_0(s)|<\varepsilon\right\}\\
&=\Pi\left[\sup_{s\in\mathcal{B}}|g_{\lambda_0}\{\tilde{\beta}(s)\}-g_{\lambda_0}\{\tilde{\beta}_0(s)\}|<\varepsilon\right]\\
&\geq\Pi\left\{\sup_{s\in\mathcal{B}}|\tilde{\beta}(s)-\tilde{\beta}_0(s)|<\varepsilon\right\}\\
&=\Pi\left\{\sup_{s\in\mathcal{B}}\left|\int\tilde{K}(\|s-t\|)a(t)dt-\int\tilde{K}(|s-t|)a_0(t)dt\right|<\varepsilon\right\}\\
&=\Pi\left\{\int\tilde{K}(\|s-t\|)\,|a(t)-a_0(t)|\,dt<\varepsilon\right\}\\
&\geq\Pi\left\{\sup_{t\in\mathcal{B}}|a(t)-a_0(t)|<\varepsilon/M_K\right\}>0,
\end{aligned}
$$

where $M_K=\sup_{s\in\mathcal{B}}\int\tilde{K}(\|s-t\|)dt<\infty$. The last inequality is based on Theorem 4 in (Ghosal & Roy, 2006). □

Given the large support properties, using similar arguments in Theorems 2, 3 and 4, we can establish the posterior consistency and sign consistency for models (1) and (6) for any $\beta_0\in\bar{\Theta}$ without using Condition A5.

## 14. ADDITIONAL SIMULATION STUDY RESULTS

Table S1 presents the functional principal component analysis (Xiao et al., 2013) and the fused lasso results for different values of the tuning parameters. Table S2 shows the selection accuracy of the Gaussian process approach.

## REFERENCES

ADLER, R. J. (1990). An introduction to continuity, extrema, and related topics for general Gaussian processes. *Lecture Notes-Monograph Series* **12**, 1–155.

CHOUDHURI, N., GHOSAL, S. & ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Am. Statist. Assoc.* **99**, 1050–1059.

GHOSAL, S. & ROY, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Statist.* **34**, 2413–2429.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Am. Statist. Assoc.* **58**, 13–30.

XIAO, L., LI, Y. & RUPPERT, D. (2013). Fast bivariate P-splines: the sandwich smoother. *J. R. Statist. Soc.* B **75**, 577–599.

[*Received August* 2016. *Revised XXXX* 201$X$]

Table S1. *Simulation study for the linear model fitting by fused lasso with three choices of the tuning parameter $\tilde{\gamma}$ and the functional principal component analysis with different percent variation explained*

| | | | | | Mean squared error for $\beta^{\mathrm{v}}$, multiplied by $10^4$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Signal | $n$ | $\tau$ | $\sigma$ | $\vartheta_x$ | FL, 1/5 | FL, 1 | FL, 5 | FPCA, 80% | FPCA, 90% | FPCA, 95% |
| Five peaks | 100 | 0 | 5 | 3 | 26 (1) | 22 (0) | 191 (4) | 36 (0) | 36 (0) | 37 (0) |
| | 100 | 0 | 5 | 6 | 30 (1) | 22 (0) | 31 (3) | 33 (0) | 34 (0) | 34 (0) |
| | 100 | 0 | 2 | 3 | 17 (0) | 11 (0) | 45 (1) | 26 (0) | 24 (1) | 26 (1) |
| | 250 | 0 | 5 | 3 | 19 (0) | 13 (0) | 17 (0) | 27 (1) | 30 (0) | 30 (0) |
| Triangle | 100 | 0 | 5 | 3 | 10 (0) | 9 (0) | 180 (4) | 17 (0) | 18 (0) | 18 (0) |
| | 100 | 0 | 5 | 6 | 10 (0) | 8 (0) | 67 (10) | 15 (0) | 16 (0) | 17 (0) |
| | 100 | 0 | 2 | 3 | 6 (0) | 5 (0) | 38 (1) | 12 (0) | 13 (0) | 13 (0) |
| | 250 | 0 | 5 | 3 | 7 (0) | 6 (0) | 8 (0) | 14 (0) | 15 (0) | 15 (0) |
| | 100 | 2 | 5 | 3 | 7 (0) | 8 (0) | 678 (14) | 10 (0) | 10 (0) | 11 (0) |
| | 100 | 4 | 5 | 3 | 6 (0) | 6 (0) | 694 (14) | 4 (0) | 4 (0) | 4 (0) |
| Waves | 100 | 0 | 5 | 3 | 323 (9) | 250 (5) | 578 (9) | 227 (5) | 190 (6) | 188 (7) |
| | 100 | 0 | 5 | 6 | 349 (7) | 233 (4) | 664 (10) | 246 (5) | 145 (3) | 126 (2) |

| | | | | | Type I error (%) | | |
|---|---|---|---|---|---|---|---|
| Signal | $n$ | $\tau$ | $\sigma$ | $\vartheta_x$ | FL, 1/5 | FL, 1 | FL, 5 |
| Five peaks | 100 | 0 | 5 | 3 | 66 (2) | 14 (1) | 19 (0) |
| | 100 | 0 | 5 | 6 | 81 (1) | 37 (1) | 9 (0) |
| | 100 | 0 | 2 | 3 | 72 (1) | 24 (1) | 17 (0) |
| | 250 | 0 | 5 | 3 | 70 (1) | 19 (1) | 3 (0) |
| Triangle | 100 | 0 | 5 | 3 | 24 (2) | 5 (0) | 20 (0) |
| | 100 | 0 | 5 | 6 | 37 (2) | 8 (1) | 6 (1) |
| | 100 | 0 | 2 | 3 | 28 (2) | 7 (1) | 19 (0) |
| | 250 | 0 | 5 | 3 | 22 (2) | 5 (0) | 1 (0) |
| | 100 | 2 | 5 | 3 | 7 (1) | 1 (0) | 20 (0) |
| | 100 | 4 | 5 | 3 | 2 (1) | 1 (0) | 20 (0) |

| | | | | | Power (%) | | |
|---|---|---|---|---|---|---|---|
| Signal | $n$ | $\tau$ | $\sigma$ | $\vartheta_x$ | FL, 1/5 | FL, 1 | FL, 5 |
| Five peaks | 100 | 0 | 5 | 3 | 88 (2) | 55 (2) | 35 (1) |
| | 100 | 0 | 5 | 6 | 96 (1) | 80 (1) | 47 (1) |
| | 100 | 0 | 2 | 3 | 97 (0) | 84 (1) | 56 (1) |
| | 250 | 0 | 5 | 3 | 96 (1) | 77 (1) | 46 (1) |
| Triangle | 100 | 0 | 5 | 3 | 95 (1) | 85 (1) | 50 (1) |
| | 100 | 0 | 5 | 6 | 98 (0) | 91 (1) | 67 (2) |
| | 100 | 0 | 2 | 3 | 98 (0) | 95 (0) | 80 (1) |
| | 250 | 0 | 5 | 3 | 97 (0) | 92 (1) | 81 (1) |
| | 100 | 2 | 5 | 3 | 86 (1) | 77 (1) | 38 (1) |
| | 100 | 4 | 5 | 3 | 84 (1) | 82 (1) | 38 (1) |

Type I error: proportion of times estimating zero coefficients to be nonzero; Power: proportion of times estimating nonzero coefficients to be nonzero; FL, $\tilde{\gamma}$: the fused lasso approach defined in (9) in the main text with tuning parameter $\tilde{\gamma}$ fixed and $\tilde{\lambda}$ selected based on Bayesian information criterion; FPCA, $\alpha$ %: the functional principal component analysis approach (Xiao et al., 2013) using eigen vectors that explain $\alpha$% of variation;

Table S2. *Selection accuracy by the Gaussian process approach*

Linear regression

| Signal | $n$ | $\tau$ | $\sigma$ | $\vartheta_x$ | Type I error, % | Power, % |
|---|---|---|---|---|---|---|
| Five peaks | 100 | 0 | 5 | 3 | 3 (0) | 48(1) |
| | 100 | 0 | 5 | 6 | 3 (0) | 42(1) |
| | 100 | 0 | 2 | 3 | 7(0) | 77(1) |
| | 250 | 0 | 5 | 3 | 6 (0) | 74(1) |
| Triangle | 100 | 0 | 5 | 3 | 3 (0) | 80(1) |
| | 100 | 0 | 5 | 6 | 2 (0) | 77(1) |
| | 100 | 0 | 2 | 3 | 5 (0) | 96(0) |
| | 250 | 0 | 5 | 3 | 4 (0) | 95(0) |
| | 100 | 2 | 5 | 3 | 2 (0) | 70(1) |
| | 100 | 4 | 5 | 3 | 2 (0) | 78(1) |

Binary regression

| Signal | Type I error, % | Power, % |
|---|---|---|
| Five peaks | 10 (0) | 55 (1) |
| Triangle | 8 (0) | 98 (0) |

Type I error: proportion of times the 95% posterior credible intervals of zero coefficients including zero; Power: proportion of times the 95% posterior credible intervals of nonzero coefficients excluding zero.