Feature Screening in Ultrahigh Dimensional Generalized Varying-coefficient Models

Guangren Yang[1], Songshan Yang[2] and Runze Li[2]

[1]*Jinan University,* [2]*Pennsylvania State University*

Proof of Theorem 1. It follows by the Taylor expansion for the quasi-likelihood function $\ell(\boldsymbol{\gamma})$ at $\boldsymbol{\beta}$ lying within a neighbor of $\boldsymbol{\gamma}$ that

$$\ell(\boldsymbol{\gamma}) = \ell(\boldsymbol{\beta}) + (\boldsymbol{\gamma}-\boldsymbol{\beta})^T\ell'(\boldsymbol{\beta}) + \frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\beta})^T\ell''(\tilde{\boldsymbol{\beta}})(\boldsymbol{\gamma}-\boldsymbol{\beta}),$$

where $\tilde{\boldsymbol{\beta}}$ lies between $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. For $(\boldsymbol{\gamma}-\boldsymbol{\beta})^T\ell''(\tilde{\boldsymbol{\beta}})(\boldsymbol{\gamma}-\boldsymbol{\beta})$ term, we have

$$\begin{aligned}
& (\boldsymbol{\gamma}-\boldsymbol{\beta})^T\{-\ell''(\tilde{\boldsymbol{\beta}})\}(\boldsymbol{\gamma}-\boldsymbol{\beta}) \\
= \; & (\boldsymbol{\gamma}-\boldsymbol{\beta})^T W^{1/2}(\boldsymbol{\beta})W^{-1/2}(\boldsymbol{\beta})\{-\ell''(\tilde{\boldsymbol{\beta}})\}W^{-1/2}(\boldsymbol{\beta})W^{1/2}(\boldsymbol{\beta})(\boldsymbol{\gamma}-\boldsymbol{\beta}) \\
\leq \; & \lambda_{\max}[W^{-1/2}(\boldsymbol{\beta})\{-\ell''(\tilde{\boldsymbol{\beta}})\}W^{-1/2}(\boldsymbol{\beta})](\boldsymbol{\gamma}-\boldsymbol{\beta})^T W(\boldsymbol{\beta})(\boldsymbol{\gamma}-\boldsymbol{\beta}),
\end{aligned}$$

where $W(\boldsymbol{\beta})$ is a block diagonal matrix with $W_j(\boldsymbol{\beta})$ being a $d_{nj}\times d_{nj}$ matrix. Since $-\ell''(\boldsymbol{\beta})$ is non-negative definite, $\lambda_{\max}[W^{-1/2}(\boldsymbol{\beta})\{-\ell''(\tilde{\boldsymbol{\beta}})\}W^{-1/2}(\boldsymbol{\beta})] \geq 0$ Thus, if

$$u > \lambda_{\max}[W^{-1/2}(\boldsymbol{\beta})\{-\ell''(\tilde{\boldsymbol{\beta}})\}W^{-1/2}(\boldsymbol{\beta})],$$

then

$$\ell(\boldsymbol{\gamma}) \geq \ell(\boldsymbol{\beta}) + (\boldsymbol{\gamma}-\boldsymbol{\beta})^T\ell'(\boldsymbol{\beta}) - \frac{u}{2}(\boldsymbol{\gamma}-\boldsymbol{\beta})^T W(\boldsymbol{\beta})(\boldsymbol{\gamma}-\boldsymbol{\beta}) = h(\boldsymbol{\gamma}|\boldsymbol{\beta}).$$

Thus it follows that $\ell(\boldsymbol{\gamma}) \geq h(\boldsymbol{\gamma}|\boldsymbol{\beta})$ and $\ell(\boldsymbol{\beta}) = h(\boldsymbol{\beta}|\boldsymbol{\beta})$ by the definition of $h(\boldsymbol{\gamma},\boldsymbol{\beta})$. The solution of $\partial h(\boldsymbol{\gamma}|\boldsymbol{\beta})/\partial\boldsymbol{\gamma} = 0$ is $\boldsymbol{\gamma} = \boldsymbol{\beta} + u^{-1}W(\boldsymbol{\beta})\ell'(\boldsymbol{\beta})$. Hence, under the conditions of Theorem 1, it follows that

$$\ell(\boldsymbol{\beta}^{*(t+1)}) \geq h(\boldsymbol{\beta}^{*(t+1)}|\boldsymbol{\beta}^{(t)}) \geq h(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}) = \ell(\boldsymbol{\beta}^{(t)}).$$

The second inequality is due to the fact that $\tau(\{j : \|\boldsymbol{\beta}_j^{*(t+1)}\|_2 > 0\}) = \tau(\{j : \|\boldsymbol{\beta}_j^{(t)}\|_2 > 0\}) = m$, and $\boldsymbol{\beta}^{*(t+1)} = \arg\max_{\boldsymbol{\gamma}} h(\boldsymbol{\gamma}|\boldsymbol{\beta}^{(t)})$ subject to $\tau(\{j : \|\boldsymbol{\gamma}_j\|_2 > 0\}) \leq m$. By definition of $\boldsymbol{\beta}^{(t+1)}$, $\ell(\boldsymbol{\beta}^{(t+1)}) \geq \ell(\boldsymbol{\beta}^{*(t+1)})$ and $\tau(\{j : \|\boldsymbol{\beta}_j^{(t+1)}\|_2 > 0\}) = m$. This proves Theorem 1. □

Proof of Theorem 2. For a given model s, a subset of $\{1,\ldots,p\}$, let $\hat{\boldsymbol{\alpha}}_s(\cdot)$ be the unrestricted maximum likelihood estimation of $\boldsymbol{\alpha}_s(\cdot)$ based on the spline approximation. It suffices to show that

$$Pr\left[\max_{s\in S_-^m}\ell\{\hat{\boldsymbol{\alpha}}_s(U)\} \geq \min_{s\in S_+^m}\ell\{\hat{\boldsymbol{\alpha}}_s(U)\}\right] \longrightarrow 0, \qquad \text{(A.1)}$$

as $n \to \infty$.

We approximate $\alpha_j(U)$ by

$$\alpha_{nj}(U) = \sum_{k=1}^{d_n} \beta_{jk}\psi_{jk}(U) = \boldsymbol{\beta}_j^T \boldsymbol{\psi}_j(U), \quad j = 1, \cdots, p, \tag{A.2}$$

where $\psi_{jk}(U)$, $k = 1, \ldots, d_n$, are basis functions and d_n is the number of basis functions, which is allowed to increase with the sample size n.

Let $\mathcal{S}_j$ denote all functions that have the form $\sum_{k=1}^{d_n} \beta_{jk}\psi_{jk}(U)$ for a given set of basis $\{\psi_{jk}, k = 1, \ldots, d_n\}$. For $\alpha_{nj}(U)$, define the approximation error by

$$\rho_j(U) = \alpha_j(U) - \alpha_{nj}(U) = \alpha_j(U) - \sum_{k=1}^{d_n} \beta_{jk}\psi_{jk}(U), \quad j = 1, \ldots, p.$$

Let $\mathrm{dist}(\alpha_j(\cdot), \mathcal{S}_j) = \inf_{\alpha_{nj}(U)\in\mathcal{S}_j} \sup_{U\in[a,b]} \|\rho_j(U)\|_2$, and take $\rho = \max_{1\le j\le p} \mathrm{dist}(\alpha_j(\cdot), \mathcal{S}_j)$. Let $\boldsymbol{\alpha}_n(U) = (\alpha_{n1}(U), \ldots, \alpha_{np}(U))^T$ and $\boldsymbol{\alpha}(U) = (\alpha_1(U), \ldots, \alpha_p(U))^T$. For any s,

$$\begin{aligned}
\boldsymbol{\alpha}_s(U) &= \begin{pmatrix} \boldsymbol{\psi}_1(U) & & \\ & \ddots & \\ & & \boldsymbol{\psi}_s(U) \end{pmatrix}_{s\times sd_n} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_s \end{pmatrix}_{sd_n\times 1} + \begin{pmatrix} \rho_1(U) \\ \vdots \\ \rho_s(U) \end{pmatrix} \\
&\hat{=} \Psi_s(U)\boldsymbol{\beta}_s + \rho_s(U),
\end{aligned}$$

where $\Psi_s(U) = \mathrm{diag}(\boldsymbol{\psi}_1(U), \ldots, \boldsymbol{\psi}_s(U))$ with $\boldsymbol{\psi}_j(U) = (\psi_{j1}(U), \ldots, \psi_{jd_n}(U))$ and $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jd_n})^T$, $j = 1, \ldots, s$.

For any $s \in S_-^m$, define $s' = s \cup s^* \in S_+^{2m}$. So, we have

$$\begin{aligned}
& \ell\{\boldsymbol{\alpha}_{s'}(U)\} - \ell\{\boldsymbol{\alpha}_{s'}^*(U)\} \\
=\ & \ell\{\Psi_{s'}(U)\boldsymbol{\beta}_{s'} + \rho_{s'}(U)\} - \ell\{\Psi_{s'}(U)\boldsymbol{\beta}_{s'}^* + \rho_{s'}^*(U)\} \\
=\ & \ell\{\Psi_{s'}(U)\boldsymbol{\beta}_{s'}\} + \ell'\{\Psi_{s'}(U)\tilde{\boldsymbol{\beta}}_{s'}\}\rho_{s'}(U) - \ell\{\Psi_{s'}(U)\boldsymbol{\beta}_{s'}^*\} - \ell'\{\Psi_{s'}(U)\tilde{\boldsymbol{\beta}}_{s'}^*\}\rho_{s'}^*(U),
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_{s'}$ and $\tilde{\boldsymbol{\beta}}_{s'}^*$ are two immediate values. Denote

$$\Delta_1 = \ell(\boldsymbol{\beta}_{s'}) - \ell(\boldsymbol{\beta}_{s'}^*), \quad \Delta_2 = \ell'(\tilde{\boldsymbol{\beta}}_{s'})\rho_{s'}(U), \quad \Delta_3 = \ell'(\tilde{\boldsymbol{\beta}}_{s'}^*)\rho_{s'}^*(U).$$

Thus,

$$\ell\{\boldsymbol{\alpha}_{s'}(U)\} - \ell\{\boldsymbol{\alpha}_{s'}^*(U)\} = \Delta_1 + \Delta_2 - \Delta_3.$$

For Δ_2, by the Cauchy-Schwartz inequality, we have

$$E|\Delta_2| = E|\ell'(\tilde{\boldsymbol{\beta}}_{s'})\rho_{s'}(U)| \le \sqrt{E\|\ell'(\tilde{\boldsymbol{\beta}}_{s'})\|^2}\sqrt{E\|\rho_{s'}(U)\|^2}.$$

According to the property of quasi-likelihood, we have

$$E\|\ell'(\tilde{\boldsymbol{\beta}}_{s'})\|^2 = \mathrm{tr}E\{\ell'(\tilde{\boldsymbol{\beta}}_{s'})\ell'(\tilde{\boldsymbol{\beta}}_{s'})^T\} = -\mathrm{tr}E\ell''(\tilde{\boldsymbol{\beta}}_{s'}).$$

By condition (C6) and Corollary 1 in Wei, Huang, and Li (2011), it follows $\Delta_2 = o_p(1)$. Similarly Δ_2, we have $\Delta_3 = o_p(1)$.

Next, we consider Δ_1. By Wedderburn (Part 5, 1974), the quasi-score function of $\boldsymbol{\beta}_s$ is given by

$$S_n(\boldsymbol{\beta}_s) = \frac{\partial \ell(\boldsymbol{\beta}_s)}{\partial \boldsymbol{\beta}_s} = \sum_{i=1}^{n} \frac{\mu'(\mathbf{z}_{is}^T\boldsymbol{\beta}_s)}{V(\mathbf{z}_{is}^T\boldsymbol{\beta}_s)}[Y_i - E(Y_i|\mathbf{z}_i)]\mathbf{z}_{is},$$

where $\mu'(t)$ is the first-order derivative of $\mu(t)$. Let $H_n(\boldsymbol{\beta}_s) = -\partial^2\ell(\boldsymbol{\beta}_s)/\partial\boldsymbol{\beta}_s\partial\boldsymbol{\beta}_s^T$ be the Hessian matrix of $\ell(\boldsymbol{\beta}_s)$ corresponding to $\boldsymbol{\beta}_s$.

Under (C3), we consider $\boldsymbol{\beta}_{s'}$ close to $\boldsymbol{\beta}^*_{s'}$ such that $\|\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}\| = w_1 d_n n^{-\tau_1}$ for some $w_1, \tau_1 > 0$. Clearly, when n is sufficiently large, $\boldsymbol{\beta}_{s'}$ falls into a neighborhood of $\boldsymbol{\beta}^*_{s'}$, so that condition (C6) becomes applicable. Thus, it follows by Condition (C6) and the Cauchy-Schwarz inequality that, we have

$$\begin{aligned}
\Delta_1 &= \ell(\boldsymbol{\beta}_{s'}) - \ell(\boldsymbol{\beta}^*_{s'}) \\
&= [\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}]^T S_n(\boldsymbol{\beta}^*_{s'}) - (1/2)[\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}]^T H_n(\tilde{\boldsymbol{\beta}}_{s'})[\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}] \\
&\leq [\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}]^T S_n(\boldsymbol{\beta}^*_{s'}) - (C_1/2) n d_n^{-1} \|\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}\|_2^2 \\
&\leq w_1 d_n n^{-\tau_1} \|S_n(\boldsymbol{\beta}^*_{s'})\|_2 - (C_1/2) d_n^{-1} w_1^2 d_n^2 n^{1-2\tau_1}, \qquad (A.3)
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_{s'}$ is an intermediate value between $\boldsymbol{\beta}_{s'}$ and $\boldsymbol{\beta}^*_{s'}$. Thus, we have

$$\begin{aligned}
Pr\{\ell(\boldsymbol{\beta}_{s'}) - \ell(\boldsymbol{\beta}^*_{s'}) \geq 0\} &\leq Pr\{\|S_n(\boldsymbol{\beta}^*_{s'})\|_2 \geq (C_1 w_1/2) n^{1-\tau_1}\} \\
&\leq \sum_{j \in s'} Pr\{S_{nj}^2(\boldsymbol{\beta}^*_{s'}) \geq (2m)^{-1}(C_1 w_1/2)^2 n^{2-2\tau_1}\},
\end{aligned}$$

where

$$S_{nj}(\boldsymbol{\beta}^*_{s'}) = \sum_{i=1}^{n} \frac{\mu'(\mathbf{z}_{is'}^T\boldsymbol{\beta}_{s'})}{V(\mathbf{z}_{is'}^T\boldsymbol{\beta}_{s'})}[Y_i - E(Y_i|\mathbf{z}_i)]z_{ij}.$$

Let $t_{ni} = z_{ij}(\sum_{i=1}^n z_{ij}^2)^{-1/2}$ such that $\sum_{i=1}^n t_{ni}^2 = 1$, and $\mu'(\mathbf{z}_{is'}^T\boldsymbol{\beta}_{s'})/V(\mathbf{z}_{is'}^T\boldsymbol{\beta}_{s'})$ is bounded by constant M under condition (C5). Under Condition (C6), we have $\max_i\{t_{ni}^2\} = O_P(n^{-1})$. By condition (C3), we have $m \leq w_2 n^{\tau_2}$. These conditions give the exponential bounds for sums of bounded variable probability inequality (Lin and Bai, 2009, Page 74), we have

$$
\begin{aligned}
& Pr\{S_{nj}(\boldsymbol{\beta}^*_{s'}) \geq (C_1 w_1/2)(2m)^{-1/2} n^{1-\tau_1}\} \\
\leq\ & Pr\{S_{nj}(\boldsymbol{\beta}^*_{s'}) \geq (C_1 w_1/2)(2w_2)^{-1/2} n^{-0.5\tau_2} n^{1-\tau_1}\} \\
\leq\ & Pr\left\{\sum_{i=1}^{n} t_{ni}[Y_i - E(Y_i|\mathbf{z}_i)] > c n^{0.5(1-2\tau_1-\tau_2)}\right\} \\
\leq\ & \exp\left(-(c^2/2) n^{1-2\tau_1-\tau_2}\right), \qquad (A.4)
\end{aligned}
$$

where $c = C_1 w_1/(2M\sqrt{2w_2})$. Also, by the same arguments, we have

$$
Pr\{S_{nj}(\boldsymbol{\beta}^*_{s'}) \leq -(C_1 w_1/2)(2m)^{-1/2} n^{1-\tau_1}\} \leq \exp\left(-(c^2/2) n^{1-2\tau_1-\tau_2}\right), \qquad (A.5)
$$

The inequalities (A.4) and (A.5) imply that,

$$
Pr\{\ell(\boldsymbol{\beta}_{s'}) \geq \ell(\boldsymbol{\beta}^*_{s'})\} \leq 4m \exp\left(-(c^2/2) n^{1-2\tau_1-\tau_2}\right).
$$

So, under condition (C4), we have

$$
\begin{aligned}
& Pr\left\{\max_{s \in S^m_-} \ell(\boldsymbol{\beta}_{s'}) \geq \ell(\boldsymbol{\beta}^*_{s'})\right\} \\
\leq\ & \sum_{s \in S^m_-} Pr\{\ell(\boldsymbol{\beta}_{s'}) \geq \ell(\boldsymbol{\beta}^*_{s'})\} \\
\leq\ & 4mp^m \exp\{-0.5c^2 n^{1-2\tau_1-\tau_2}\} \\
=\ & 4\exp\{\log m + m\log p - 0.5c^2 n^{1-2\tau_1-\tau_2}\} \\
\leq\ & 4\exp\{\log w_2 + \tau_2 \log n + w_2 n^{\tau_2} \log p - 0.5c^2 n^{1-2\tau_1-\tau_2}\} \\
=\ & 4w_2 \exp\{\tau_2 \log n + w_2 n^{\tau_2} \log p - 0.5c^2 n^{1-2\tau_1-\tau_2}\} \\
=\ & o(1) \quad \text{as} \quad n \to \infty. \qquad (A.6)
\end{aligned}
$$

By Condition (C6), $\ell(\boldsymbol{\beta}_{s'})$ is concave in $\boldsymbol{\beta}_{s'}$, (A.6) holds for any $\boldsymbol{\beta}_{s'}$ such that $\|\boldsymbol{\beta}_{s'} - \boldsymbol{\beta}^*_{s'}\| = w_1 d_n n^{-\tau_1}$.

For any $s \in S^m_-$, let $\breve{\boldsymbol{\beta}}_{s'}$ be $\hat{\boldsymbol{\beta}}_s$ augmented with zeros corresponding to the elements in $s' \backslash s^*$ (i.e. $s' = \{s \cup (s^* \backslash s)\} \cup (s' \backslash s^*)$). By Condition (C1), it is seen that $\|\breve{\boldsymbol{\beta}}_{s'} - \boldsymbol{\beta}^*_{s'}\|_2 = \|\breve{\boldsymbol{\beta}}_{s^* \cup (s' \backslash s^*)} - \boldsymbol{\beta}^*_{s^* \cup (s' \backslash s^*)}\|_2 = \|\breve{\boldsymbol{\beta}}_{s^* \cup (s' \backslash s^*)} - \boldsymbol{\beta}^*_{s^*}\|_2 \geq \|\boldsymbol{\beta}^*_{s^* \cup (s' \backslash s^*)} - \boldsymbol{\beta}^*_{s^*}\|_2 \geq \|\boldsymbol{\beta}^*_{s' \backslash s^*}\|_2 = w_1 d_n n^{-\tau_1}$. Consequently,

$$
Pr\left\{\max_{s \in S^m_-} \ell(\hat{\boldsymbol{\beta}}_s) \geq \min_{s \in S^m_+} \ell(\hat{\boldsymbol{\beta}}_s)\right\} \leq Pr\left\{\max_{s \in S^m_-} \ell_p(\breve{\boldsymbol{\beta}}_{s'}) \geq \ell_p(\boldsymbol{\beta}^*_{s'})\right\} = o(1).
$$

So, we have shown that

$$
Pr\left[\max_{s \in S^m_-} \ell\{\hat{\boldsymbol{\alpha}}_s(U)\} \geq \min_{s \in S^m_+} \ell\{\hat{\boldsymbol{\alpha}}_s(U)\}\right] \longrightarrow 0,
$$

as $n \to \infty$. The theorem is proved. $\square$

Proof of Theorem 3. According to the definition of HBIC, for any model s, $HBIC(\tau(s)) \leq HBIC(q)$ implies that

$$\begin{aligned} \ell(\hat{\boldsymbol{\beta}}_s) - \ell(\hat{\boldsymbol{\beta}}_{s^*}) &\geq d_n\{\tau(s) - q\}\frac{C_n \log(d_n p)}{2n} \\ &\geq -d_n q\frac{C_n \log(d_n p)}{2n}. \end{aligned} \tag{A.7}$$

We show that the probability that (A.7) occurs at any $s \in S_-^m$ goes to 0. For any $s \in S_-^m$, let $\tilde{s} = s \cup s^*$. To consider those $\boldsymbol{\beta}_{\tilde{s}}$ near $\boldsymbol{\beta}_{\tilde{s}}^*$, we have

$$\ell(\boldsymbol{\beta}_{\tilde{s}}) - \ell(\boldsymbol{\beta}_{\tilde{s}}^*) = \{\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\}^T \ell'(\boldsymbol{\beta}_{\tilde{s}}^*) - \frac{1}{2}\{\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\}^T[-\ell''(\tilde{\boldsymbol{\beta}}_{\tilde{s}}^*)]\{\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\},$$

for some $\tilde{\boldsymbol{\beta}}_{\tilde{s}}^*$ between $\boldsymbol{\beta}_{\tilde{s}}$ and $\boldsymbol{\beta}_{\tilde{s}}^*$. By Condition (C6),

$$\{\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\}^T[-\ell''(\tilde{\boldsymbol{\beta}}_{\tilde{s}}^*)]\{\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\} \geq C_1 d_n^{-1} n \|\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\|^2.$$

Therefore,

$$\ell(\boldsymbol{\beta}_{\tilde{s}}) - \ell(\boldsymbol{\beta}_{\tilde{s}}^*) \leq \{\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\}^T \ell'(\boldsymbol{\beta}_{\tilde{s}}^*) - \frac{C_1}{2} d_n^{-1} n \|\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\|^2.$$

Hence, for any $\boldsymbol{\beta}_{\tilde{s}}$ such that $\|\boldsymbol{\beta}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\| = w_1 d_n n^{-\tau_1}$, we have

$$\ell(\boldsymbol{\beta}_{\tilde{s}}) - \ell(\boldsymbol{\beta}_{\tilde{s}}^*) \leq w_1 d_n n^{-\tau_1} \|\ell'(\boldsymbol{\beta}_{\tilde{s}}^*)\| - \frac{C_1}{2} d_n^{-1} n (w_1 d_n n^{-\tau_1})^2.$$

By (A.4), (A.5) and (A.6), we can get

$$Pr\left\{\sup_{s \in S_-^m} \ell(\boldsymbol{\beta}_{\tilde{s}}) \geq \ell(\boldsymbol{\beta}_{\tilde{s}}^*)\right\} = o(1).$$

Now let $\breve{\boldsymbol{\beta}}_{\tilde{s}}$ be $\hat{\boldsymbol{\beta}}_s$ augmented with zeros corresponding to the elements in $\tilde{s} \backslash s$. It can be seen that

$$\|\breve{\boldsymbol{\beta}}_{\tilde{s}} - \boldsymbol{\beta}_{\tilde{s}}^*\| \geq \|\boldsymbol{\beta}_{s^* \backslash s}^*\| = w_1 d_n n^{-\tau_1},$$

by (C3). Therefore, uniformly over $s \in S_-^m$ and with probability tending to 1,

$$Pr\left\{\sup_{s \in S_-^m} \ell(\hat{\boldsymbol{\beta}}_{\tilde{s}}) \geq \ell(\boldsymbol{\beta}_{\tilde{s}}^*)\right\} \leq Pr\left\{\sup_{s \in S_-^m} \ell(\breve{\boldsymbol{\beta}}_{\tilde{s}}) \geq \ell(\boldsymbol{\beta}_{\tilde{s}}^*)\right\} = o(1).$$

Hence, the probability that (A.7) occurs at any $s \in S_-^m$ tends to 0 which is (2.13).

On the other hand, for $s \in S_+^m$, let $k = \tau(s) - q$. It suffices to consider a fixed k, since k takes only the values $1, \ldots, m-q$. By definition, $HBIC(\tau(s)) \leq HBIC(q)$ if and only if

$$\ell(\hat{\boldsymbol{\beta}}_s) - \ell(\hat{\boldsymbol{\beta}}_{s^*}) \geq k d_n \frac{C_n \log(d_n p)}{2n}.$$

We show that, uniformly in $s \in S_+^m$ with $\tau(s) = k + q$, this inequality does not occur. For large n, by condition (C6),

$$\begin{aligned}
\ell(\hat{\boldsymbol{\beta}}_s) - \ell(\hat{\boldsymbol{\beta}}_{s^*}) &\leq \ell(\hat{\boldsymbol{\beta}}_s) - \ell(\boldsymbol{\beta}_s^*) \\
&\leq \{\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*\}^T \ell'(\boldsymbol{\beta}_s^*) - \frac{1}{2}\{\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*\}^T [-\ell''(\tilde{\boldsymbol{\beta}}_s^*)]\{\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*\} \\
&\leq \{\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*\}^T \ell'(\boldsymbol{\beta}_s^*) - \frac{1}{2} C_1 d_n^{-1} n \{\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*\}^T \{\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*\}.
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_s^*$ lies between $\hat{\boldsymbol{\beta}}_s$ and $\hat{\boldsymbol{\beta}}_s^*$. Denote $\Delta = \hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^*$, and define

$$f(\Delta) = \Delta^T \ell'(\boldsymbol{\beta}_s^*) - \frac{1}{2} C_1 d_n^{-1} n \Delta^T \Delta.$$

So, we have

$$\frac{\partial f(\Delta)}{\partial \Delta} = \ell'(\boldsymbol{\beta}_s^*) - C_1 d_n^{-1} n \Delta = 0.$$

This implies that $f(\Delta)$ reaches its maximum at $\Delta = d_n \ell'(\hat{\boldsymbol{\beta}}_s^*)/(C_1 n)$. Thus,

$$\ell(\hat{\boldsymbol{\beta}}_s) - \ell(\hat{\boldsymbol{\beta}}_{s^*}) \leq \frac{1}{2}(C_1 n d_n^{-1})^{-1} \ell'(\boldsymbol{\beta}_s^*)^T \ell'(\boldsymbol{\beta}_s^*).$$

Hence, we show that, uniformly over $s \in S_+^m$ with $\tau(s) = k + q$,

$$\frac{1}{2}(C_1 n d_n^{-1})^{-1} \ell'(\boldsymbol{\beta}_s^*)^T \ell'(\boldsymbol{\beta}_s^*) \geq k d_n \frac{C_n \log(d_n p)}{2n},$$

occurs with diminishing probability. Thus, under conditions (C4) and (C6), by Markov inequality, for each $s \in S_+^m$, we have

$$\begin{aligned}
& Pr\left[\frac{1}{2}(C_1 n d_n^{-1})^{-1} \ell'(\boldsymbol{\beta}_s^*)^T \ell'(\boldsymbol{\beta}_s^*) \geq k d_n \frac{C_n \log(d_n p)}{2n}\right] \\
= \; & Pr\left[\ell'(\boldsymbol{\beta}_s^*)^T \ell'(\boldsymbol{\beta}_s^*) \geq C_1 k C_n \log(d_n p)\right] \\
\leq \; & \frac{E[\ell'(\boldsymbol{\beta}_s^*)^T \ell'(\boldsymbol{\beta}_s^*)]}{C_1 k C_n \log(d_n p)} = \frac{E[\ell'(\boldsymbol{\beta}_s^*)^T \ell'(\boldsymbol{\beta}_s^*)]}{C_1 k C_n (\log(d_n) + n^\kappa)} \longrightarrow 0.
\end{aligned}$$

the number of models in S_+^m is lower than p^κ, we have shown that

$$Pr\left[\frac{1}{2}(C_1 n d_n^{-1})^{-1}\ell'(\boldsymbol{\beta}_s^*)^T\ell'(\boldsymbol{\beta}_s^*) \geq k d_n \frac{C_n \log(d_n p)}{2n}, \forall s \in S_+^m\right] \longrightarrow 0,$$

This completes the proof. □

Table 1: The proportions of $\mathcal{P}_j$s and $\mathcal{P}_a$ for continuous response with $\Sigma = \Sigma_1$

				CC-SIS					New (SJS)				
n	p	ρ	$\boldsymbol{\alpha}(\cdot)$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$
200	1000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0	0	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	0.995	1	1	0.992	0.987	1	1	1	1	1
200	1000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	0.015	0.015	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	994	0.999	0.996	0.979	0.970	1	1	1	1	1
200	1000	2/3	$\boldsymbol{\alpha}_1$	0.995	0.997	0.995	0.302	0.297	1	1	0.999	1	0.999
			$\boldsymbol{\alpha}_2$	0.976	0.995	0.984	0.942	0.909	1	1	1	1	1
200	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0.001	0.001	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	0.992	0.999	0.998	0.989	0.979	1	1	1	1	1
200	2000	1/2	$\boldsymbol{\alpha}_1$	0.999	0.997	0.998	0.008	0.008	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	0.991	0.998	0.994	0.973	0.958	1	1	1	1	1
200	2000	2/3	$\boldsymbol{\alpha}_1$	0.989	0.987	0.985	0.284	0.274	1	1	0.993	1	0.993
			$\boldsymbol{\alpha}_2$	0.974	0.999	0.976	0.932	0.892	1	1	1	1	1
400	1000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0	0	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	1000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	0.023	0.023	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	1000	2/3	$\boldsymbol{\alpha}_1$	1	1	1	0.623	0.623	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0	0	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	2000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	0.011	0.011	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	2000	2/3	$\boldsymbol{\alpha}_1$	1	1	1	0.549	0.549	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1

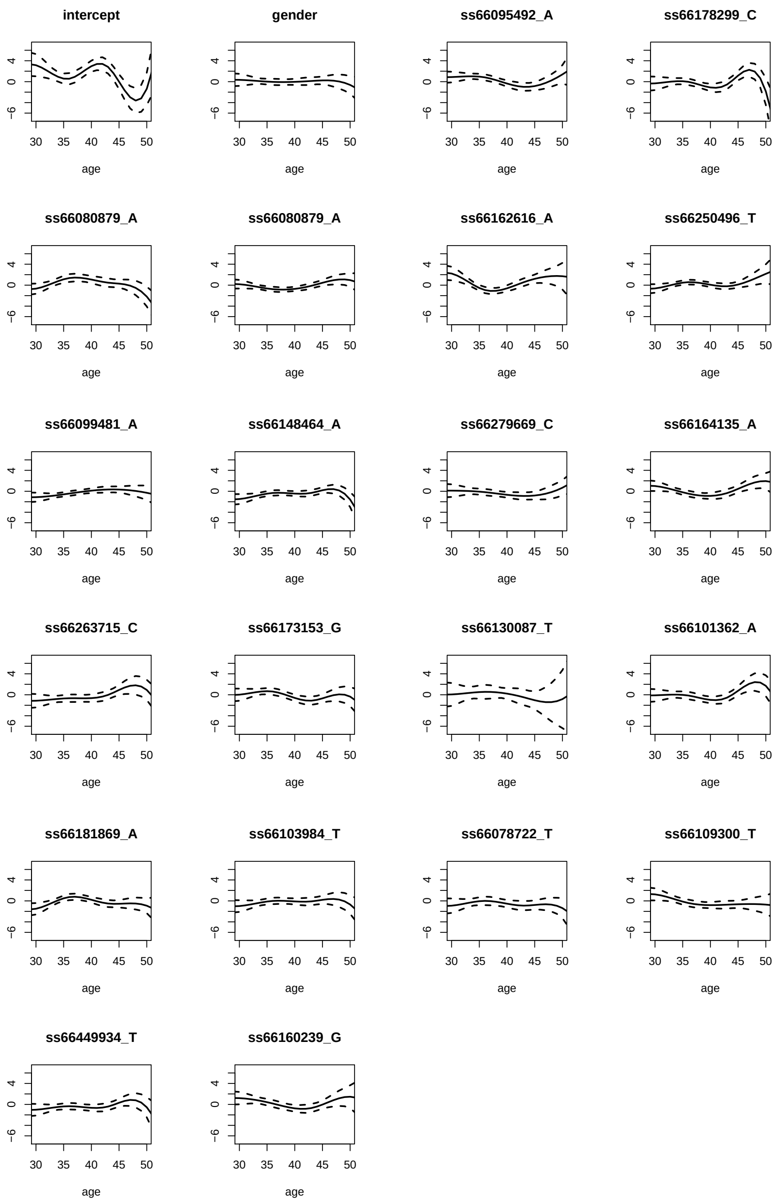

Figure 1: Estimated Coefficient Functions for selected by the HBIC tuning parameter selector.

Table 2: The proportions of $\mathcal{P}_j$s and $\mathcal{P}_a$ for continuous response with $\Sigma = \Sigma_2$

				CC-SIS					New (SJS)				
n	p	ρ	$\boldsymbol{\alpha}(\cdot)$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$
200	1000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0.644	0.644	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
200	1000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	0.887	0.887	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	0.996	0.999	0.995	1	1	1	1	1
200	1000	2/3	$\boldsymbol{\alpha}_1$	1	1	0.741	0.990	0.731	1	1	0.952	1	0.952
			$\boldsymbol{\alpha}_2$	1	0.745	0.999	1	0.744	1	1	0.998	1	0.998
200	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0.551	0.551	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
200	2000	1/2	$\boldsymbol{\alpha}_1$	1	1	0.997	0.858	0.855	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	0.991	0.999	1	0.990	1	1	1	1	1
200	2000	2/3	$\boldsymbol{\alpha}_1$	1	1	0.678	0.991	0.669	1	1	0.903	1	0.903
			$\boldsymbol{\alpha}_2$	0.999	0.693	0.999	1	0.692	1	1	0.996	1	0.996
400	1000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0.982	0.982	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	1000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	1000	2/3	$\boldsymbol{\alpha}_1$	1	1	0.993	1	0.993	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	0.996	1	1	0.996	1	1	1	1	1
400	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	0.951	0.951	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	2000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	0.999	0.999	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
400	2000	2/3	$\boldsymbol{\alpha}_1$	1	1	0.991	1	0.991	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	0.986	1	1	0.986	1	1	1	1	1

Table 3: Computing times (Seconds) and the Number of Iterations for Continuous Response

	Σ_1				Σ_2			
	$\boldsymbol{\alpha}_1$		$\boldsymbol{\alpha}_2$		$\boldsymbol{\alpha}_1$		$\boldsymbol{\alpha}_2$	
ρ	Time	Iterations	Time	Iterations	Time	Iterations	Time	Iterations
	$(n,p) = (200, 1000)$							
1/3	3.97(0.17)	10(0)	4.10(0.36)	10(0)	4.13(0.45)	10(0)	3.90(0.20)	10(0)
1/2	4.22(0.24)	10(0)	5.03(0.87)	10(0)	3.98(0.83)	10(0)	4.25(0.37)	10(0)
2/3	3.93(0.11)	10(0)	4.08(0.83)	10(0)	4.25(0.36)	10(0)	4.21(0.32)	10(0)
	$(n,p) = (200, 2000)$							
1/3	7.87(0.47)	10(0)	7.37(0.63)	10(0)	8.04(0.70)	10(0)	7.24(0.20)	10(0)
1/2	7.91(0.59)	10(0)	8.40(0.53)	10(0)	7.98(0.53)	10(0)	7.25(0.21)	10(0)
2/3	7.75(0.61)	10(0)	7.03(0.64)	10(0)	8.05(0.35)	10(0)	7.15(0.39)	10(0)
	$(n,p) = (400, 1000)$							
1/3	2.73(0.37)	5(1)	2.03(0.3)	4(1)	2.98(0.41)	5(1)	2.89(0.46)	5(0)
1/2	2.20(0.21)	4(0)	1.44(0.10)	3(0)	2.91(0.40)	5(1)	2.86(0.46)	5(1)
2/3	1.98(0.30)	4(1)	1.50(0.22)	3(0)	2.42(0.39)	5(1)	2.58(0.33)	5(1)
	$(n,p) = (400, 2000)$							
1/3	4.87(0.67)	5(1)	3.73(0.47)	4(0)	4.87(0.57)	5(1)	6.01(0.98)	5(1)
1/2	3.69(0.29)	4(0)	3.34(0.55)	3(0)	5.97(1.05)	5(1)	6.03(0.93)	5(1)
2/3	3.18(0.43)	4(0)	2.34(0.68)	3(0)	4.67(0.68)	5(1)	6.54(1.72)	5(1)

Table 4: The proportions of $\mathcal{P}_j$s and $\mathcal{P}_a$ for binary response

				$\Sigma = \Sigma_1$					$\Sigma = \Sigma_2$				
n	p	ρ	$\boldsymbol{\alpha}(\cdot)$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$
300	1000	1/3	$\boldsymbol{\alpha}_1$	0.999	0.998	1	1	0.997	1	1	0.998	0.994	0.992
			$\boldsymbol{\alpha}_2$	0.999	1	1	1	0.999	1	1	1	1	1
300	1000	1/2	$\boldsymbol{\alpha}_1$	0.983	0.987	0.987	1	0.958	1	1	0.984	1	0.984
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	0.996	1	0.996
300	1000	2/3	$\boldsymbol{\alpha}_1$	0.925	0.928	0.946	1	0.813	1	1	0.896	0.996	0.894
			$\boldsymbol{\alpha}_2$	0.995	1	0.996	0.994	0.988	1	0.997	0.976	1	0.973
300	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	0.998	0.99	0.988
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
300	2000	1/2	$\boldsymbol{\alpha}_1$	0.974	0.98	0.984	1	0.941	0.998	1	0.955	0.999	0.952
			$\boldsymbol{\alpha}_2$	0.999	1	1	0.998	0.997	1	1	0.994	1	0.994
300	2000	2/3	$\boldsymbol{\alpha}_1$	0.898	0.903	0.923	1	0.75	0.998	0.999	0.821	0.994	0.816
			$\boldsymbol{\alpha}_2$	0.991	1	0.996	0.99	0.979	1	0.99	0.952	1	0.943
500	1000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	1000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	1000	2/3	$\boldsymbol{\alpha}_1$	0.998	0.998	0.998	1	0.994	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	2000	1/2	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	2000	2/3	$\boldsymbol{\alpha}_1$	0.987	0.995	0.998	1	0.980	1	1	0.998	1	0.998
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1

Table 5: Computing times (Seconds) and the number of iterations for binary response

	Σ_1				Σ_2			
	$\boldsymbol{\alpha}_1$		$\boldsymbol{\alpha}_2$		$\boldsymbol{\alpha}_1$		$\boldsymbol{\alpha}_2$	
ρ	Time	Iterations	Time	Iterations	Time	Iterations	Time	Iterations
	$(n,p) = (300, 1000)$							
1/3	15.65(2.51)	5(1)	13.18(2.37)	4(1)	12.36(1.69)	4(1)	14.52(2.62)	4(0)
1/2	17.39(2.56)	4(0)	8.17(0.28)	3(0)	14.70(2.39)	4(1)	14.48(2.67)	4(0)
2/3	15.44(2.39)	4(0)	9.19(1.75)	3(0)	14.55(1.98)	4(1)	16.76(3.19)	4(1)
	$(n,p) = (300, 2000)$							
1/3	23.63(4.09)	5(1)	19.80(3.31)	4(1)	17.76(3.55)	4(1)	16.93(3.21)	4(1)
1/2	17.70(1.08)	4(0)	13.54(0.39)	3(0)	22.61(4.13)	5(1)	18.79(3.60)	4(1)
2/3	16.94(1.94)	4(0)	13.46(0.64)	3(0)	22.24(3.89)	5(1)	21.50(3.56)	4(1)
	$(n,p) = (500, 1000)$							
1/3	75.23(11.43)	5(0)	50.36(8.00)	4(0)	55.09(8.95)	5(1)	55.03(7.53)	5(1)
1/2	64.40(8.98)	4(0)	33.64(3.32)	3(0)	62.36(8.52)	5(1)	56.10(9.03)	5(1)
2/3	55.52(8.34)	4(0)	31.63(3.18)	3(0)	63.35(8.16)	5(1)	56.07(9.19)	5(1)
	$(n,p) = (500, 2000)$							
1/3	112.07(18.07)	5(0)	57.70(4.09)	4(0)	70.14(12.46)	5(1)	71.20(10.52)	5(1)
1/2	75.85(13.67)	4(0)	49.28(7.43)	3(0)	69.76(11.67)	5(1)	70.23(12.71)	5(1)
2/3	78.53(11.51)	4(0)	44.31(3.67)	3(0)	79.09(13.66)	5(1)	72.74(11.21)	5(1)

Table 6: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ for count response

				$\Sigma = \Sigma_1$					$\Sigma = \Sigma_2$				
n	p	ρ	$\boldsymbol{\alpha}(\cdot)$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_a$
300	1000	1/3	$\boldsymbol{\alpha}_1$	0.982	0.976	0.978	0.983	0.942	0.998	0.998	0.983	0.989	0.975
			$\boldsymbol{\alpha}_2$	0.998	0.999	1	0.997	0.996	1	0.998	0.998	0.998	0.995
300	1000	1/2	$\boldsymbol{\alpha}_1$	0.945	0.941	0.928	0.989	0.842	0.999	1	0.884	0.994	0.883
			$\boldsymbol{\alpha}_2$	0.982	0.988	0.994	0.98	0.95	1	0.981	0.979	0.999	0.968
300	1000	2/3	$\boldsymbol{\alpha}_1$	0.815	0.848	0.808	0.979	0.554	0.993	0.998	0.622	0.994	0.617
			$\boldsymbol{\alpha}_2$	0.866	0.917	0.894	0.852	0.626	1	0.825	0.793	0.997	0.703
300	2000	1/3	$\boldsymbol{\alpha}_1$	0.965	0.966	0.956	0.973	0.895	0.998	1	0.966	0.97	0.955
			$\boldsymbol{\alpha}_2$	0.987	0.994	0.997	0.989	0.976	1	0.99	0.99	0.999	0.987
300	2000	1/2	$\boldsymbol{\alpha}_1$	0.897	0.895	0.88	0.994	0.739	0.996	0.997	0.811	0.991	0.806
			$\boldsymbol{\alpha}_2$	0.962	0.982	0.985	0.964	0.909	0.999	0.95	0.938	0.997	0.913
300	2000	2/3	$\boldsymbol{\alpha}_1$	0.744	0.743	0.748	0.986	0.421	0.992	0.99	0.489	0.988	0.479
			$\boldsymbol{\alpha}_2$	0.811	0.879	0.858	0.806	0.534	1	0.694	0.676	0.995	0.54
500	1000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	1000	1/2	$\boldsymbol{\alpha}_1$	0.999	0.999	1	1	0.998	0.999	1	0.991	1	0.990
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	1000	2/3	$\boldsymbol{\alpha}_1$	0.989	0.983	0.991	1	0.965	0.999	1	0.958	1	0.958
			$\boldsymbol{\alpha}_2$	0.996	1	1	0.993	0.989	1	0.996	0.997	1	0.994
500	2000	1/3	$\boldsymbol{\alpha}_1$	1	1	1	1	1	1	1	1	1	1
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	2000	1/2	$\boldsymbol{\alpha}_1$	0.999	1	0.999	1	0.998	1	1	0.988	1	0.988
			$\boldsymbol{\alpha}_2$	1	1	1	1	1	1	1	1	1	1
500	2000	2/3	$\boldsymbol{\alpha}_1$	0.981	0.976	0.972	1	0.933	1	1	0.929	1	0.929
			$\boldsymbol{\alpha}_2$	0.988	0.995	0.996	0.994	0.974	1	0.987	0.979	1	0.973

Table 7: Computing times (Seconds) and the number of iterations for count response

	Σ_1				Σ_2			
	$\boldsymbol{\alpha}_1$		$\boldsymbol{\alpha}_2$		$\boldsymbol{\alpha}_1$		$\boldsymbol{\alpha}_2$	
ρ	Time	Iterations	Time	Iterations	Time	Iterations	Time	Iterations
	$(n,p) = (300, 1000)$							
1/3	13.62(2.44)	4(1)	11.10(2.10)	4(1)	16.17(2.40)	5(1)	11.86(2.39)	4(1)
1/2	10.51(2.23)	4(1)	12.61(2.03)	3(1)	12.90(2.46)	5(1)	15.39(2.65)	5(1)
2/3	9.76(0.67)	3(0)	11.15(1.51)	3(0)	12.84(2.46)	5(1)	13.04(2.44)	5(1)
	$(n,p) = (300, 2000)$							
1/3	17.24(3.16)	4(1)	18.50(3.96)	4(1)	22.47(3.79)	5(1)	20.40(3.48)	5(1)
1/2	17.12(3.23)	4(1)	16.64(2.84)	4(1)	20.38(3.67)	5(1)	20.53(3.61)	5(1)
2/3	13.84(0.62)	3(0)	13.67(0.51)	3(0)	19.84(3.73)	5(1)	21.20(3.98)	5(1)
	$(n,p) = (500, 1000)$							
1/3	56.39(9.94)	4(1)	43.94(6.90)	4(1)	54.58(8.08)	5(1)	63.15(9.99)	5(1)
1/2	43.14(6.40)	4(0)	39.69(6.17)	4(1)	51.78(9.01)	5(1)	52.92(8.86)	5(1)
2/3	47.08(7.45)	4(1)	29.25(1.14)	3(0)	51.12(9.04)	5(1)	52.86(8.80)	5(1)
	$(n,p) = (500, 2000)$							
1/3	77.70(11.08)	4(1)	53.43(10.93)	4(1)	70.14(12.30)	5(1)	71.47(12.31)	5(1)
1/2	61.36(8.73)	4(0)	52.00(11.15)	4(1)	70.80(12.03)	5(1)	74.42(10.20)	5(1)
2/3	50.81(11.06)	4(1)	50.32(8.40)	3(0)	70.83(11.98)	5(1)	76.46(11.58)	6(1)

Table 8: Comparing AIC, BIC and HBIC (mean and sd)

		Continuous response		Binary response		Count response	
		p=1000	p=2000	p=1000	p=2000	p=1000	p=2000
AIC	P	0.100	0.060	0.055	0.020	0.420	0.370
	C	4(0)	4(0)	4(0.100)	4(0)	4(0)	4(0.141)
	I	10.200(7.366)	9.850(7.262)	11.425(6.889)	13.63(6.030)	1.64(2.242)	2.030(2.901)
BIC	P	0.745	0.715	0.760	0.710	0.665	0.570
	C	4(0)	4(0)	4(0.571)	4(0)	4(0.262)	4(0.278)
	I	0.305(0.560)	0.325(0.549)	0.300(0.481)	0.220(0.503)	0.530(0.956)	0.720(1.161)
HBIC	P	0.970	0.975	0.915	0.710	0.700	0.620
	C	4(0)	4(0)	3.73(0.954)	4(0)	4(0)	4(0)
	I	0.030(0.171)	0.025(0.157)	0.005(0.171)	0.320(0.509)	0.600(1.143)	0.660(1.002)