Supplementary Material to the article “Consistent group selection using global-local shrinkage priors in sparse normal linear regression”

Sayantan Paul, Prasenjit Ghosh and Arijit Chakrabarti

This document consists of Sections A, B, C and D. In Section A, we provide proofs of Propositions 1 and 2 along with Theorems 1-6 mentioned in Sections 2 and 3 of our main work. Section B consists of results related to the accuracy of the half-thresholding (HT) estimate of the group coefficient $\boldsymbol{\beta}_g$. Gibbs sampling algorithm is discussed in Section C. Finally, in Section D we provide additional simulation results which could not be included in the main document due to the page constraint. Now we proceed to establish the results mentioned in the main document. But before going into the technical details, let us restate the general class of one-group global-local priors that we are interested in this work:

$$\begin{aligned}
\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_n),\\
\boldsymbol{\beta}_g \mid \lambda_g^2, \sigma^2, \tau^2 &\overset{ind}{\sim} \mathcal{N}_{m_g}(\mathbf{0}, \lambda_g^2\sigma^2\tau^2(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1}), \text{ for } g = 1, \ldots, G,\\
\lambda_g^2 &\overset{ind}{\sim} \pi(\lambda_g^2) = K(\lambda_g^2)^{-a-1}L(\lambda_g^2), \text{ for } g = 1, \ldots, G, \text{ and}\\
(\tau, \sigma^2) &\sim \pi(\tau, \sigma^2).
\end{aligned} \tag{1}$$

where $K \in (0, \infty)$ is the constant of proportionality, a is a positive real number, and $L : (0, \infty) \to (0, \infty)$ is measurable non-constant slowly varying function, that is, for any $\alpha > 0$, $\frac{L(\alpha x)}{L(x)} \to 1$ as $x \to \infty$. For the theoretical development of the paper, we assumed several conditions on the slowly varying function $L(\cdot)$, the global shrinkage parameter τ, the number of active variables $G_{\mathcal{A}_n}$, the minimum value of non-null group coefficient and the singular values of the design matrix.

A Proofs of Lemmas, Propositions and Theorems

A.1 Some important Lemmas

We first state and prove Lemmas A.1 to A.4 which are crucial to proving the main theorems of our work.

Lemma A.1. *Let L be a nonnegative, measurable, slowly varying function defined over an interval unbounded to the right. Then the following results hold.*

(1) L^α *is slowly varying for all* $\alpha \in \mathbb{R}$.

(2) $\frac{\log L(x)}{\log x} \to 0$ *as* $x \to \infty$.

(3) For every $\alpha > 0$, $x^{-\alpha}L(x) \to 0$ *and* $x^\alpha L(x) \to \infty$ *as* $x \to \infty$.

(4) For $\alpha < -1$, $-\frac{\int_x^\infty t^\alpha L(t)dt}{x^{\alpha+1}L(x)} \to \frac{1}{\alpha+1}$ *as* $x \to \infty$.

(5) There exists a global constant $A_0 > 0$ such that, for any $\alpha > -1$, $\frac{\int_{A_0}^x t^\alpha L(t)dt}{x^{\alpha+1}L(x)} \to \frac{1}{\alpha+1}$ as $x \to \infty$.

Proof. See Bingham et al. (1987). ◻

Lemma A.2. *Let $L : (0,\infty) \to (0,\infty)$ be a measurable and integrable function such that for fixed $a > 0$, $\int_0^\infty t^{-a-1}L(t)dt = K^{-1}$, with $K \in (0,\infty)$. Assume $\tau_n \to 0$ as $n \to \infty$. Then*

$$\int_0^1 u^{a+\frac{m_g}{2}-1}(1-u)^{-a-1} L\left(\frac{1}{\tau_n^2}(\frac{1}{u}-1)\right) du = K^{-1}(\tau_n^2)^{-a}(1+o(1)),$$

where the $o(1)$ term is such that $\lim_{n\to\infty} o(1) = 0$.

Proof. The proof follows using the same arguments used to establish Lemma 5 of Ghosh et al. (2016). ◻

Lemma A.3. *Consider the hierarchical framework of (5) where the local shrinkage parameters are modeled with the class of priors given by (6). Suppose $\tau_n \to 0$ as $n \to \infty$. Then for any given $a \in (0,1)$, there exists $A_0 \geq 1$ such that*

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq \frac{A_0 K}{a(1-a)}(\tau_n^2)^a L\left(\frac{1}{\tau_n^2}\right) \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)(1+o(1)).$$

Assume that the slowly varying function $L(\cdot)$ satisfies Assumption 1 for some $a \geq 1$. Then

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq \frac{KM}{a}\tau_n \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)(1+o(1)).$$

The terms $o(1)$ in both inequalities above tend to zero as $n \to \infty$.

Proof. The proof for the case $a \in (0,1)$ follows using the same arguments employed by Ghosh et al. (2016) to establish Theorem 4 of their paper.

Let us now consider the case $a \geq 1$. First, note that

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) = \frac{\int_0^1 \kappa_g^{a+\frac{m_g}{2}-1}(1-\kappa_g)^{-a} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right)\exp\left\{(1-\kappa_g)\cdot\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right\}d\kappa_g}{\int_0^1 \kappa_g^{a+\frac{m_g}{2}-1}(1-\kappa_g)^{-a-1} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right)\exp\left\{(1-\kappa_g)\cdot\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right\}d\kappa_g}. \tag{2}$$

Using the transformation $s = \frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)$ in the integrals above, we obtain

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) = \tau_n^2 \frac{\int_0^\infty (1+s\tau_n^2)^{-\frac{m_g}{2}-1} s^{-a} L(s)\exp\left(\frac{s\tau_n^2}{1+s\tau_n^2}\cdot\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)ds}{\int_0^\infty (1+s\tau_n^2)^{-\frac{m_g}{2}} s^{-a-1} L(s)\exp\left(\frac{s\tau_n^2}{1+s\tau_n^2}\cdot\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)ds}. \tag{3}$$

Note that

$$\begin{aligned}\int_0^\infty (1+s\tau_n^2)^{-\frac{m_g}{2}} s^{-a-1} L(s)\exp\left(\frac{s\tau_n^2}{1+s\tau_n^2}\cdot\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)ds &\geq \int_0^\infty (1+s\tau_n^2)^{-\frac{m_g}{2}} s^{-a-1} L(s)ds \\ &= K^{-1}(1+o(1)),\end{aligned} \tag{4}$$

where the last equality above follows from the Dominated Convergence Theorem. Combining (3) and (4), we obtain

$$\begin{aligned}
&E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le K\tau_n^2 \int_0^\infty (1+s\tau_n^2)^{-\frac{m_g}{2}-1} s^{-a} L(s) \exp\left(\frac{s\tau_n^2}{1+s\tau_n^2} \cdot \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right) ds(1+o(1)) \\
&= K\tau_n^2 \left(\int_0^1 + \int_1^{\frac{1}{\tau_n}} + \int_{\frac{1}{\tau_n}}^\infty\right) (1+s\tau_n^2)^{-\frac{m_g}{2}-1} s^{-a} L(s) \exp\left(\frac{s\tau_n^2}{1+s\tau_n^2} \cdot \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right) ds(1+o(1)) \\
&= K(A_{1,\tau_n} + A_{2,\tau_n} + A_{3,\tau_n})(1+o(1)), \quad \text{say.}
\end{aligned} \tag{5}$$

Observe that for $s \in (0,1)$ and $\tau_n \in (0,1)$, $\frac{s\tau_n^2}{1+s\tau_n^2} \le \frac{1}{2}$. Also, $\int_0^\infty s^{-a-1} L(s) dt = K^{-1}$. Therefore, it follows that

$$A_{1,\tau_n} \le K^{-1} \tau_n^2 \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{4\sigma^2}\right). \tag{6}$$

Likewise, for $s \in [1, \frac{1}{\tau_n})$ and $\tau_n \in (0,1)$, using the above arguments, we obtain

$$A_{2,\tau_n} \le K^{-1} \tau_n \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{4\sigma^2}\right). \tag{7}$$

Finally, using (4) of Lemma A.1, we have

$$\begin{aligned}
A_{3,\tau_n} &\le \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right) \int_{\frac{1}{\tau}}^\infty s^{-a-1} L(s) ds \\
&= \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right) \frac{\tau_n^a}{a} L(\frac{1}{\tau_n})(1+o(1)) \\
&\le \frac{\tau_n}{a} M \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)(1+o(1)).
\end{aligned} \tag{8}$$

Combining (5)-(8), the desired result follows. □

Lemma A.4. *Consider the framework of Lemma A.3. Then under Assumption 1, for any arbitrary constants $\eta \in (0,1), q \in (0,1)$ and any fixed $\tau > 0$,*

$$P(\kappa_g > \eta | \tau, \sigma^2, \mathcal{D}) \le \frac{(a + \frac{m_g}{2})(1-\eta q)^a}{\tau^{2a} (\eta q)^{a+\frac{m_g}{2}} c_0} \exp\left(-\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g \eta(1-q)}{2\sigma^2}\right).$$

Proof. The proof follows using similar arguments used by Ghosh et al. (2016) to establish Theorem 5 in their paper. □

A.2 Proofs of Propositions

Proof of Proposition 1:

Proof. First, we prove the if part. Towards that, first, consider the case when $a \in (0,1)$. Using Lemma A.3, we obtain

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq \frac{A_0 K}{a(1-a)} (\tau_n^2)^a L(\frac{1}{\tau_n^2}) \exp\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g}{2\sigma^2}\right)(1+o(1)). \tag{9}$$

When $\tau_n \to 0$ as $n \to \infty$, using Part (3) of Lemma A.1,

$$\lim_{n\to\infty} (\tau_n^2)^a L(\frac{1}{\tau_n^2}) = \lim_{n\to\infty} (\frac{1}{\tau_n^2})^{-a} L(\frac{1}{\tau_n^2}) = 0. \tag{10}$$

Under a block-orthogonal design, from the standard theory of linear regression, the distribution of the ordinary least square estimator $\widehat{\boldsymbol{\beta}}_g$ is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g^0\right) \sim \mathcal{N}_{m_g}(\mathbf{0}, \sigma^2 \mathbf{Q}_{n,g}^{-1}).$$

Clearly, if $\boldsymbol{\beta}_g^0 = \mathbf{0}$, $\sqrt{n}\widehat{\boldsymbol{\beta}}_g \sim \mathcal{N}_{m_g}(\mathbf{0}, \sigma^2 \mathbf{Q}_{n,g}^{-1})$. Therefore, $\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} \sim \chi^2_{m_g}$, whence using (A3)

$$\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} = O_p(1), \text{ for all } n. \tag{11}$$

Combining (9) - (11), and using Slutsky's Theorem, it readily follows that

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \xrightarrow{P} 0 \text{ as } n \to \infty.$$

Next, we consider the case when $a \geq 1$. Observe that the upper bound to $E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D})$ is similar to the upper bound when $a \in (0,1)$. Hence, the proof follows using the same arguments as in the case $a \in (0,1)$.
Now, we consider the 'Only if' part. Here, we will show that if $\tau_n \nrightarrow 0$ as $n \to \infty$,
$P(E(1-\kappa_g|\tau_n, \sigma^2, \mathcal{D}) > \frac{1}{2}) \nrightarrow 0$ as $n \to \infty$.
Since, $\tau_n \nrightarrow 0$ as $n \to \infty$, implies, there exists at least one subsequence of τ_n, say τ_{n_k}, such that as $k \to \infty$, $\tau_{n_k} > \epsilon$ for some $\epsilon > 0$. Hence, the remaining calculations will be based on that subsequence of τ_n. Note that

$$\begin{aligned} E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) &= \int_0^{\frac{1}{4}} \kappa_g \pi(\kappa_g|\tau_n, \sigma^2, \mathcal{D}) d\kappa_g + \int_{\frac{1}{4}}^1 \kappa_g \pi(\kappa_g|\tau_n, \sigma^2, \mathcal{D}) d\kappa_g \\ &\leq \frac{1}{4} + P(\kappa_g > \frac{1}{4}|\tau_n, \sigma^2, \mathcal{D}). \end{aligned} \tag{12}$$

Therefore, we have,

$$P\left(E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) < \frac{1}{2}\right) \geq P\left(P(\kappa_g > \frac{1}{4}|\tau_n, \sigma^2, \mathcal{D}) < \frac{1}{4}\right). \tag{13}$$

Now, substituting $\eta = \frac{1}{4}$ in Lemma A.4, some simple algebra yields, for any fixed $\tau > 0$,

$$\begin{aligned} P\left(P(\kappa_g > \frac{1}{4}|\tau_n, \sigma^2, \mathcal{D}) < \frac{1}{4}\right) &\geq P\left(\frac{(a+\frac{1}{2})(1-\frac{1}{4}q)^a}{{\tau_n}^{2a}(\eta q)^{a+\frac{1}{2}} c_0} \exp\left(-\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g} \widehat{\boldsymbol{\beta}}_g (1-q)}{8\sigma^2}\right) < \frac{1}{4}\right) \\ &= P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > \frac{8}{(1-q)}[2a \log\left(\frac{1}{\tau}\right) + K_2]\right), \end{aligned} \tag{14}$$

where K_2 is a constant depending on a, q, c_0. Note that, for sufficiently large k, $\log\left(\frac{1}{\tau_{n_k}}\right) < \log\left(\frac{1}{\epsilon}\right)$, where the ϵ is mentioned previously. This ensures a lower bound on (14), as $k \to \infty$

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} > \frac{8}{(1-q)}[2a\log\left(\frac{1}{\tau_{n_k}}\right) + K_2]\right) \geq P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} > \frac{8}{(1-q)}[2a\log\left(\frac{1}{\epsilon}\right) + K_2]\right)$$
$$\nrightarrow 0 \text{ as } k \to \infty. \tag{15}$$

Combining (12)-(15) completes the proof of the Only if part and the proof of Proposition 1 is completed. □

Proof of Proposition 2:

Proof. It would be enough to show that $E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \xrightarrow{P} 0$ as $n \to \infty$ when $\boldsymbol{\beta}_g^0 \neq \mathbf{0}$.

Let us fix $\epsilon_0 > 0$. Then

$$\begin{aligned} E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) &= \int_0^{\frac{\epsilon_0}{2}} \kappa_g \pi(\kappa_g|\tau_n, \sigma^2, \mathcal{D}) d\kappa_g + \int_{\frac{\epsilon_0}{2}}^1 \kappa_g \pi(\kappa_g|\tau_n, \sigma^2, \mathcal{D}) d\kappa_g \\ &\leq \frac{\epsilon_0}{2} + P(\kappa_g > \frac{\epsilon_0}{2}|\tau_n, \sigma^2, \mathcal{D}) . \end{aligned} \tag{16}$$

Therefore, for a given $\epsilon_0 > 0$,

$$P\left(E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \epsilon_0\right) \leq P\left(P(\kappa_g > \frac{\epsilon_0}{2}|\tau_n, \sigma^2, \mathcal{D}) > \frac{\epsilon_0}{2}\right). \tag{17}$$

Now, substituting $\eta = \frac{\epsilon_0}{2}$ in Lemma A.4, some simple algebra yields

$$\begin{aligned} P(E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \epsilon_0) &\leq P\left(\frac{(a+\frac{m_g}{2})(1-\eta q)^a}{{\tau_n}^{2a}(\eta q)^{a+\frac{m_g}{2}} c_0} \exp\left(-\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} Q_{n,g}\widehat{\boldsymbol{\beta}}_g \eta(1-q)}{2\sigma^2}\right) > \frac{\epsilon_0}{2}\right) \\ &= P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right), \end{aligned} \tag{18}$$

where

$$d_n = \frac{4}{\epsilon_0(1-q)}\left[d' + a \cdot \log\left(\frac{1}{\tau_n^2}\right)\right],$$

d' being a constant is independent of n.
Observe now that using (A1), we have

$$\begin{aligned} P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) &\leq \sum_{j=1}^{m_g} P\left(\frac{\sqrt{n}|\widehat{\beta}_{gj}|}{\sigma\sqrt{\zeta_j}} \leq \frac{1}{\sqrt{C_1}} \cdot \sqrt{\frac{d_n}{\zeta_j}}\right) \\ &= \sum_{j=1}^{m_g} P\left(-\frac{1}{\sqrt{C_1}} \cdot \sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}} \leq \frac{\sqrt{n}(\widehat{\beta}_{gj} - \beta_{gj}^0)}{\sigma\sqrt{\zeta_j}} \leq \frac{1}{\sqrt{C_1}} \cdot \sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}\right), \end{aligned} \tag{19}$$

where ζ_j is the j^{th} diagonal element of $\mathbf{Q}_{n,g}^{-1}$. Now, the cases $\min_j \beta_{gj}^0 > 0$ and $\min_j \beta_{gj}^0 < 0$ are dealt with separately as follows.

Case-(I): Assume $\min_j \beta^0_{gj} > 0$. Using (19), we have

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) \leq \sum_{j=1}^{m_g}\left[1 - \Phi\left(\frac{\sqrt{n}\beta^0_{gj}}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}}\right)\right]$$
$$\leq \sum_{j=1}^{m_g}\frac{\phi\left(\frac{\sqrt{n}\beta^0_{gj}}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}}\right)}{\frac{\sqrt{n}\beta^0_{gj}}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}}},$$

The inequality in the last line follows using the fact that $1-\Phi(t) \leq \frac{\phi(t)}{t}$, for any $t > 0$. Under assumption (A2), we have

$$\min_j \beta^0_{gj} > m_n \text{ for all } g \in \mathcal{A}_n \text{ with } m_n \propto n^{-b} \text{ and } 0 \leq b < \frac{1}{2}. \tag{20}$$

Next we use the fact $d_n \asymp \log\left(\frac{1}{\tau_n}\right)$ and the assumption (A4). Hence, we obtain $\frac{\sqrt{d_n}}{\sqrt{n}\beta^0_{gj}} \to 0$ as $n \to \infty$. Therefore, we have

$$\frac{\sqrt{n}\beta^0_{gj}}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} = \frac{\sqrt{n}\beta^0_{gj}}{\sigma\sqrt{\zeta_j}}(1+o(1)), \tag{21}$$

where $o(1)$ term tends to zero as $n \to \infty$. Since, $\frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} = o\left(\frac{\sqrt{n}\beta^0_{gj}}{\sigma\sqrt{\zeta_j}}\right)$, we have, for for any $\epsilon > 0$,

$$\frac{\sqrt{nd_n}\beta^0_{gj}}{\sigma\zeta_j\sqrt{C_1}} < \epsilon\frac{n{\beta^0_{gj}}^2}{\sigma^2\zeta_j} \tag{22}$$

for sufficiently large n. Let $e_1, e_2, \cdots, e_{m_g}$ be the eigenvalues of $\mathbf{Q}_{n,g}$. Then $\sum_{j=1}^{m_g}\frac{1}{e_j} = \sum_{j=1}^{m_g}\zeta_j$. Next, note that, under (A1), $e_j > C_1$ for all $j = 1,\ldots,m_g$. This implies that ζ_j is bounded above for all $j = 1,\ldots,m_g$, i.e. for all $j = 1,\ldots,m_g$, $\zeta_j < C_3$ for some $0 < C_3 < \infty$. Combining (19) - (22) and using (A3) with the above observation, it follows

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) \leq \frac{\sigma s\sqrt{C_3}}{\sqrt{n}m_n}\exp\left\{-(\frac{1}{2}-\epsilon)\frac{nm_n^2}{2\sigma^2C_3}\right\}. \tag{23}$$

Using (20) and choosing $\epsilon \in (0, \frac{1}{2})$ yields,

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) = o(1), \text{ as } n \to \infty. \tag{24}$$

Case-(II): Assume $\min_j \beta^0_{gj} < 0$. This implies that some of the $\beta^0_{gj} < 0$'s are negative and the remaining are positive. Let us assume that $\beta^0_{gj} < 0$ for all $j = 1,\ldots,\tilde{m}_g$, and $\beta^0_{gj} > 0$ for all $j = \tilde{m}_g+1,\ldots,m_g$ for some $\tilde{m}_g$.

Hence again using (19) we have

$$
\begin{aligned}
&P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) \\
&\leq \sum_{j=1}^{m_g} P\left(-\frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}} \leq \frac{\sqrt{n}(\widehat{\beta}_{gj}-\beta_{gj}^0)}{\sigma\sqrt{\zeta_j}} \leq \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}\right) \\
&\leq \sum_{j=1}^{\tilde{m}_g}\left[\Phi\left(\frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}\right) - \Phi\left(-\frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}\right)\right] + \sum_{j=\tilde{m}_g+1}^{m_g}\left[1-\Phi\left(\frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}}\right)\right] \\
&\leq \sum_{j=\tilde{m}_g+1}^{m_g}\frac{\phi\left(\frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}}\right)}{\frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}}} + \sum_{j=1}^{\tilde{m}_g}\Big[P(Z > a_{1n,j}) - P(Z > a_{2n,j})\Big], \text{ say}
\end{aligned}
\tag{25}
$$

where $a_{1n,j} = -\frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}$ and $a_{2n,j} = \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} - \frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}$, $j = 1,\ldots,\tilde{m}_g$ and $Z \sim \mathcal{N}(0,1)$. For providing an upper bound to the first term on the right hand side of (25), we use arguments used in (20)-(23). It is well known

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\left(\frac{1}{z} - \frac{1}{z^3}\right) \leq P(Z > z) \leq \frac{1}{\sqrt{2\pi}z}e^{-\frac{z^2}{2}}.$$

Using this it follows from (25) that

$$
\begin{aligned}
\sum_{j=1}^{\tilde{m}_g}\Big[P(Z > a_{1n,j}) - P(Z > a_{2n,j})\Big] &\leq \sum_{j=1}^{\tilde{m}_g}\frac{1}{\sqrt{2\pi}}\left[\frac{1}{a_{1n,j}}\exp\left(-\frac{a_{1n,j}^2}{2}\right) - \left(\frac{1}{a_{2n,j}} - \frac{1}{a_{2n,j}^3}\right)\exp\left(-\frac{a_{2n,j}^2}{2}\right)\right] \\
&\leq \sum_{j=1}^{\tilde{m}_g}\frac{1}{\sqrt{2\pi}}\left[\frac{1}{a_{1n,j}}\exp\left(-\frac{a_{1n,j}^2}{2}\right) + \left(\frac{1}{a_{2n,j}} + \frac{1}{a_{2n,j}^3}\right)\exp\left(-\frac{a_{2n,j}^2}{2}\right)\right] \\
&\leq \frac{3}{\sqrt{2\pi}}\sum_{j=1}^{\tilde{m}_g}\left[\frac{1}{a_{1n,j}}\exp\left(-\frac{a_{1n,j}^2}{2}\right)\right],
\end{aligned}
\tag{26}
$$

where inequality in the last line holds due to the use of $a_{1n,j} \leq a_{2n,j}, j = 1,\ldots,\tilde{m}_g$. Note that using (A2) and the assumption (A4), we have $-\frac{\sqrt{d_n}}{\sqrt{n}\beta_{gj}^0} \to 0$ as $n \to \infty$.

Therefore, we have

$$a_{1n,j} = -\frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}} - \frac{1}{\sqrt{C_1}}\cdot\sqrt{\frac{d_n}{\zeta_j}} = -\frac{\sqrt{n}\beta_{gj}^0}{\sigma\sqrt{\zeta_j}}(1 + o(1)), \tag{27}$$

where $o(1)$ term tends to zero as $n \to \infty$. Hence, using arguments similar to that of (22) and (23), we obtain

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) = o(1), \text{ as } n \to \infty. \tag{28}$$

Hence combining (18), (24) and (28), the proof of Proposition 2 is complete.

□

A.3 Proofs of Theorems

Proof of Theorem 1:

Proof. First, we observe that

$$P(\widehat{\mathcal{A}}_n \leq \mathcal{A}_n) \leq \sum_{g\in\mathcal{A}_n} P(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) < \frac{1}{2}) + \sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \frac{1}{2}). \tag{29}$$

To prove this result, it suffices to show

$$\sum_{g\in\mathcal{A}_n} P(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) < \frac{1}{2}) = o(1), \text{ as } n \to \infty, \tag{30}$$

and

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \frac{1}{2}) = o(1), \text{ as } n \to \infty, \tag{31}$$

both when $\frac{1}{2} \leq a < 1$, and $a \geq 1$.
Proof of (30): Let us fix an arbitrary $\epsilon_0 > 0$. Now, using the arguments employed in (16) and (17) in the proof of Proposition 2 and applying Lemma A.4 with $\eta = \frac{\epsilon_0}{2}$, we have,

$$\begin{aligned}
&\sum_{g\in\mathcal{A}_n} P(E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \frac{1}{2}) \\
&\leq \sum_{g\in\mathcal{A}_n} P(\frac{\epsilon_0}{2} + P(\kappa_g > \frac{\epsilon_0}{2}|\tau, \sigma^2, \mathcal{D}) > \frac{1}{2}) \\
&= \sum_{g\in\mathcal{A}_n} P(P(\kappa_g > \frac{\epsilon_0}{2}|\tau, \sigma^2, \mathcal{D}) > \frac{1-\epsilon_0}{2}) \\
&\leq \sum_{g\in\mathcal{A}_n} P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right),
\end{aligned} \tag{32}$$

where d_n is the same as before. Since $G_{\mathcal{A}_n} \leq n$, it follows from (23) that

$$\sum_{g\in\mathcal{A}_n} P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) \leq \frac{\sigma s\sqrt{C_3}\sqrt{n}}{m_n} \exp\left\{ -(\frac{1}{2}-\epsilon)\frac{nm_n^2}{2\sigma^2 C_3}\right\}(1+o(1)) \to 0, \text{ as } n \to \infty,$$

whence again using (20),

$$\sum_{g\in\mathcal{A}_n} P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} < d_n\right) = o(1), \text{ as } n \to \infty. \tag{33}$$

Hence, (32) coupled with (33) completes the proof of (30).
Proof of (31):
Case (I): First consider the case when $a \in [\frac{1}{2}, 1)$. Using Lemma A.3 and our previous arguments, it follows that for all $g \notin \mathcal{A}_n$,

$$P\left(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \frac{1}{2}\right) \leq P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} > M_n\right)(1+o(1)), \tag{34}$$

where $M_n = 2\log\left(\frac{C_4}{(\tau_n^2)^a L(\frac{1}{\tau_n^2})}\right)$ and C_4 is a global constant that is independent of n. In (34), the $o(1)$ term is such that it is independent of any specific group g, and $\lim_{n\to\infty} o(1) = 0$.

Now, observe that for all $g \notin \mathcal{A}_n$, $\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} \sim \chi^2_{m_g}$.
We consider the two cases $m_g = 1$, and $m_g > 1$ separately.
For $m_g = 1$, we use

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > M_n\right) = P(|Z| > \sqrt{M_n})$$
$$= \sqrt{\frac{2}{\pi}} e^{-\frac{M_n}{2}} M_n^{-\frac{1}{2}}(1 + o(1)),$$

where Z denotes a standard normal random variable, and the last line follows due to Mills' ratio.
On the other hand, for $m_g \geq 2$, we first observe that a $\chi^2_{m_g}$ distribution can equivalently be regarded as a Gamma$(\frac{m_g}{2}, \frac{1}{2})$ distribution, with shape parameter $\frac{m_g}{2}$, and scale parameter $\frac{1}{2}$. Then we have

$$P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > M_n\right) = \frac{1}{2^{\frac{m_g}{2}}\Gamma(\frac{m_g}{2})} \int_{M_n}^{\infty} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}-1} du$$
$$= \frac{1}{\Gamma(\frac{m_g}{2})} \int_{M_n/2}^{\infty} e^{-u} u^{\frac{m_g}{2}-1} du, \tag{35}$$

where $\Gamma(r) = \int_0^\infty e^{-u} u^{r-1} du$ denotes the gamma function evaluated at $r > 0$.

We now state below a result due to Gabcke (2015) that is instrumental in completing the remainder of this proof. This is presented as Lemma A.5 below.

Lemma A.5. *When $r \geq 1$ and $c > r + 1$,*

$$e^{-c} c^{r-1} \leq \int_c^\infty e^{-u} u^{r-1} du \leq r e^{-c} c^{r-1},$$

that is, for sufficiently large $c > 0$,

$$\int_c^\infty e^{-u} u^{r-1} du \lesssim r e^{-c} c^{r-1}.$$

Thus, using Lemma A.5 coupled with the (35) and the fact that $M_n \to \infty$ as $n \to \infty$, we have, for all sufficiently large n, for all $g \notin \mathcal{A}_n$,

$$e^{-\frac{M_n}{2}} M_n^{\frac{m_g}{2}-1} \leq P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > M_n\right) \lesssim e^{-\frac{M_n}{2}} M_n^{\frac{s}{2}-1}, \tag{36}$$

where $s = \sup_{n\geq 1} \max_{g\in\{1,\ldots,G_n\}} m_g$, is finite. Using this observation, and combining (34)-(36), we have,

$$\sum_{g\notin\mathcal{A}_n} P\left(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > \frac{1}{2}\right) \lesssim G_n (\tau_n^2)^a L(\frac{1}{\tau_n^2}) \left[\log\left(\frac{1}{(\tau_n^2)^a L(\frac{1}{\tau_n^2})}\right)\right]^{\frac{s}{2}-1}. \tag{37}$$

Hence, for $a \in [\frac{1}{2}, 1)$ using Assumption 1 on $L(\cdot)$, the term of the right-hand side of (37) converges to 0, as $n \to \infty$ if $G_n \tau_n [\log\left(\frac{1}{\tau_n}\right)]^{\frac{s}{2}-1} \to 0$ as $n \to \infty$. Note that (B1) and (B2) imply $G_n \tau_n \lesssim G_n^{1-\frac{(1+\delta_1)(1-\epsilon_2)}{(1-\epsilon_1)}}$ and

$\big[\log\big(\frac{1}{\tau_n}\big)\big]^{\frac{s}{2}-1} \lesssim (\log G_n)^{\frac{s}{2}-1}$. Hence, using these two observations along with the choice of δ_1 given in (B2) ensures, $G_n\tau_n\big[\log\big(\frac{1}{\tau_n}\big)\big]^{\frac{s}{2}-1} \to 0$ as $n\to\infty$. This completes the proof of (31) when $\frac{1}{2}\le a<1$.

Case (II): Now we consider the situation $a\ge 1$. Using similar arguments employed to prove **Case (I)**, one can easily verify that there exists a constant C_5 independent of n, such that $M_n = 2\log\big(\frac{C_5}{\tau_n}\big)$ and

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \tau_n,\sigma^2,\mathcal{D}) > \frac{1}{2}) \lesssim G_n\tau_n\Big[\log\Big(\frac{1}{\tau_n}\Big)\Big]^{\frac{s}{2}-1}. \tag{38}$$

Again observe (B1) and (B2) imply $G_n\tau_n\big[\log\big(\frac{1}{\tau_n}\big)\big]^{\frac{s}{2}-1} \to 0$ as $n\to\infty$, for $a\ge 1$. Hence under the assumption of (B1) and (B2), the right hand side of (38) goes to 0 as $n\to\infty$, for each fixed $a\ge 1$, which establishes (31). This completes the proof of Theorem 1.

□

Proof of Theorem 2:

Proof. Define $T = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathcal{A}_n}^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}} - \boldsymbol{\beta}_{\mathcal{A}_n}^{0})$. Then $T = T_1 + T_2$, where $T_1 = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathcal{A}_n}^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n} - \boldsymbol{\beta}_{\mathcal{A}_n}^{0})$ and $T_2 = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathcal{A}_n}^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n})$. Now, it boils down to show that,

$$T_1 \xrightarrow{d} \mathcal{N}(0,\sigma^2), \text{ as } n\to\infty, \tag{39}$$

and

$$T_2 \xrightarrow{P} 0 \text{ as } n\to\infty. \tag{40}$$

First, we prove (39). Note that, due to the block-diagonal property of the design matrix $\mathbf{X}$, $\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n} = \boldsymbol{\Sigma}_{\mathcal{A}_n}^{-1}\mathbf{X}_{\mathcal{A}_n}^{\mathrm{T}}\mathbf{y}$ and using the standard theory of linear models, it readily follows that for $||\boldsymbol{\alpha}|| = 1$,

$$T_1 \sim \mathcal{N}(0,\sigma^2).$$

For the time being, let us assume (40) to be true. Then, combining (39) and (40), coupled with Slutsky's Theorem, the desired asymptotic normality of $\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}}$ follows.

We now turn our focus on establishing (40) above. Towards that, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} T_2^2 &\le (\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathcal{A}_n}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}) \\ &\le C_2 n(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}) \\ &= C_2\sum_{g\in\mathcal{A}_n}\Big\|\sqrt{n}(\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_g)\Big\|_2^2. \end{aligned} \tag{41}$$

The second inequality in (41) follows due to the assumption on the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ as given in (C1) and using the fact that $e_{max}(\mathbf{X}_{\mathcal{A}_n}^{\mathrm{T}}\mathbf{X}_{\mathcal{A}_n}) \le e_{max}(\mathbf{X}^{\mathrm{T}}\mathbf{X})$.
Next, observe that using the form of the posterior mean $\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}$ as given by (8) of the main document coupled with the definition of the half-thresholding estimator $\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}}$ given by (11), one may rewrite the difference

$\sqrt{n}(\widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \widehat{\boldsymbol{\beta}}_g)$ as

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \widehat{\boldsymbol{\beta}}_g) &= \sqrt{n}\Big[E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) I\Big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) > 0.5\Big\} - 1\Big]\widehat{\boldsymbol{\beta}}_g \\ &= -\sqrt{n}\widehat{\boldsymbol{\beta}}_g E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) - \sqrt{n}\widehat{\boldsymbol{\beta}}_g E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) I\Big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5\Big\}. \quad (42)\end{aligned}$$

Note that

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5 \text{ if and only if } E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \geq 0.5. \quad (43)$$

Thus,

$$0 \leq E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) I\Big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5\Big\} \leq E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}), \quad (44)$$

whence

$$\|\sqrt{n}\widehat{\boldsymbol{\beta}}_g E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) I\Big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5\Big\}\| \leq \|\sqrt{n}\widehat{\boldsymbol{\beta}}_g E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D})\|. \quad (45)$$

Now to establish (40), let us first define the following random variables:

$$W_{n,g} = \frac{n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2}, \text{ and } U_{n,g} = W_{n,g} E(\kappa_g^2 \mid \tau_n, \sigma^2, \mathcal{D}).$$

Combining (41) - (45) together with the triangle inequality for the ℓ_2 norm, we obtain

$$\begin{aligned}T_2^2 &\leq 4C_2 \sum_{g \in \mathcal{A}_n} n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \widehat{\boldsymbol{\beta}}_g E(\kappa_g^2 \mid \tau_n, \sigma^2, \mathcal{D}) \\ &\leq \frac{4C_2\sigma^2}{C_1} \sum_{g \in \mathcal{A}_n} \frac{n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} E(\kappa_g^2 \mid \tau_n, \sigma^2, \mathcal{D}) \\ &= \frac{4C_2\sigma^2}{C_1} \sum_{g \in \mathcal{A}_n} U_{ng}. \quad (46)\end{aligned}$$

The second inequality above follows using (C1) and the fact that for all $g \in \mathcal{A}_n, e_{min}(\frac{\mathbf{X}_g^{\text{T}}\mathbf{X}_g}{n}) \geq e_{min}(\frac{\mathbf{X}^{\text{T}}\mathbf{X}}{n})$. Thus to prove (40), it is enough to show that

$$\sum_{g \in \mathcal{A}_n} U_{ng} \xrightarrow{P} 0 \text{ as } n \to \infty. \quad (47)$$

Using the above definitions coupled with (18) of the main paper, we obtain

$$\begin{aligned}U_{n,g} &= W_{n,g} \frac{\int_0^1 \kappa_g^2 \cdot \kappa_g^{(a+\frac{m_g}{2}-1)}(1-\kappa_g)^{-a-1} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right) \exp\left(-\kappa_g \cdot \frac{W_{n,g}}{2}\right) d\kappa_g}{\int_0^1 \kappa_g^{(a+\frac{m_g}{2}-1)}(1-\kappa_g)^{-a-1} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right) \exp\left(-\kappa_g \cdot \frac{W_{n,g}}{2}\right) d\kappa_g} \\ &= J(W_{n,g}, \tau), \text{ say}. \quad (48)\end{aligned}$$

Next, we follow the arguments used in Lemma 3 of Ghosh and Chakrabarti (2017) to find an upper bound to $J(W_{n,g}, \tau)$. First note that, given any $c > 2$, we can find $\eta, q \in (0,1)$ such that $c = \frac{2}{\eta(1-q)}$. Following Ghosh and Chakrabarti (2017) there exists a non-negative measurable function

$h(W_{n,g},\tau) = h_1(W_{n,g},\tau) + h_2(W_{n,g},\tau)$, where $h_1(W_{n,g},\tau) = C_*[W_{n,g}\int_0^{\frac{W_{n,g}}{1+t_0}} \exp(-\frac{u}{2})u^{\frac{m_g}{2}+a-1}du]^{-1}$ and $h_2(W_{n,g},\tau) = C^* W_{n,g}\tau^{-2a}e^{-\frac{\eta(1-q)}{2}W_{n,g}}$ where t_0 is as in Assumption 1 and C_* and C^* are two constants which depends on $a, \eta, q, L(\cdot)$ and satisfies:
for any $W_{n,g}$,

$$J(W_{n,g},\tau) \leq h(W_{n,g},\tau), \tag{49}$$

and we also have for any $\rho > c$,

$$\lim_{\tau\to 0} \sup_{W_{n,g} > 2a\rho\frac{1}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}} h(W_{n,g},\tau) = 0. \tag{50}$$

First, we prove (49) and (50). For any arbitrary constant $\eta \in (0,1)$, we split $E(\kappa_g^2|\tau,\sigma^2,\mathcal{D})$ as

$$E(\kappa_g^2|\tau,\sigma^2,\mathcal{D}) = E(\kappa_g^2 1\{\kappa_g > \eta\}|\tau,\sigma^2,\mathcal{D}) + E(\kappa_g^2 1\{\kappa_g \leq \eta\}|\tau,\sigma^2,\mathcal{D}). \tag{51}$$

With the use of Lemma A.4, the first term in the r.h.s. of (51) can be bounded as

$$E(\kappa_g^2 1\{\kappa_g > \eta\}|\tau,\sigma^2,\mathcal{D}) \leq \frac{(a+\frac{m_g}{2})(1-\eta q)^a}{\tau^{2a}(\eta q)^{a+\frac{m_g}{2}}c_0}\exp\Bigg(-\eta(1-q)\frac{W_{n,g}}{2}\Bigg), \tag{52}$$

where $q \in (0,1)$ is any arbitrary constant.
Now, let us concentrate on the second term in the r.h.s. of (51). Using the definition of the posterior distribution of κ_g and then considering the transformation $t = \frac{1}{\tau^2}(\frac{1}{\kappa_g}-1)$, we get

$$\begin{aligned}
E(\kappa_g^2 1\{\kappa_g \leq \eta\}|\tau,\sigma^2,\mathcal{D}) &= \frac{\int_0^\eta \kappa_g^2.\kappa_g^{(a+\frac{m_g}{2}-1)}(1-\kappa_g)^{-a-1}L\Bigg(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\Bigg)\exp\Bigg(-\kappa_g\cdot\frac{n\widehat{\boldsymbol\beta}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol\beta}_g}{2\sigma^2}\Bigg)d\kappa_g}{\int_0^1 \kappa_g^{(a+\frac{m_g}{2}-1)}(1-\kappa_g)^{-a-1}L\Bigg(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\Bigg)\exp\Bigg(-\kappa_g\cdot\frac{n\widehat{\boldsymbol\beta}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol\beta}_g}{2\sigma^2}\Bigg)d\kappa_g} \\
&= \frac{\int_{\frac{1}{\tau^2}(\frac{1}{\eta}-1)}^\infty (1+t\tau^2)^{-\frac{m_g}{2}-2}t^{-a-1}L(t)\exp\Big(-\frac{W_{n,g}}{2(1+t\tau^2)}\Big)dt}{\int_0^\infty (1+t\tau^2)^{-\frac{m_g}{2}}t^{-a-1}L(t)\exp\Big(-\frac{W_{n,g}}{2(1+t\tau^2)}\Big)dt}.
\end{aligned}$$

Since, $L(\cdot)$ satisfies Assumption 1 and noting that $\frac{t_0}{\tau^2} > t_0$ as $\tau^2 < 1$, we immediately have

$$E(\kappa_g^2 1\{\kappa_g \leq \eta\}|\tau,\sigma^2,\mathcal{D}) \leq \frac{M}{c_0}\frac{\int_{\frac{1}{\tau^2}(\frac{1}{\eta}-1)}^\infty (1+t\tau^2)^{-\frac{m_g}{2}-2}t^{-a-1}\exp\Big(-\frac{W_{n,g}}{2(1+t\tau^2)}\Big)dt}{\int_0^\infty (1+t\tau^2)^{-\frac{m_g}{2}}t^{-a-1}\exp\Big(-\frac{W_{n,g}}{2(1+t\tau^2)}\Big)dt}. \tag{53}$$

Hence, considering the transformation $u = \frac{W_{n,g}}{2(1+t\tau^2)}$ to both numerator and denominator of (53),

$$\begin{aligned}
E(\kappa_g^2 1\{\kappa_g \leq \eta\}|\tau,\sigma^2,\mathcal{D}) &\leq \frac{M}{c_0}\frac{\int_0^{\eta W_{n,g}}(\frac{W_{n,g}}{u})^{-\frac{m_g}{2}-2}\Bigg(\frac{1}{\tau^2}(\frac{W_{n,g}}{u}-1)\Bigg)^{-a-1}\exp(-\frac{u}{2})u^{-2}du}{\int_0^{\frac{W_{n,g}}{1+t_0}}(\frac{W_{n,g}}{u})^{-\frac{m_g}{2}}\Bigg(\frac{1}{\tau^2}(\frac{W_{n,g}}{u}-1)\Bigg)^{-a-1}\exp(-\frac{u}{2})u^{-2}du} \\
&= \frac{M}{c_0}\cdot\frac{1}{W_{n,g}^2}\frac{\int_0^{\eta W_{n,g}}\exp(-\frac{u}{2})u^{\frac{m_g}{2}+a+1}(1-\frac{u}{W_{n,g}})^{-a-1}du}{\int_0^{\frac{W_{n,g}}{1+t_0}}\exp(-\frac{u}{2})u^{\frac{m_g}{2}+a-1}(1-\frac{u}{W_{n,g}})^{-a-1}du}.
\end{aligned}$$

Now observe that for any $0 < u < \eta W_{n,g}$, we have $1 < (1 - \frac{u}{W_{n,g}})^{-a-1} < (1-\eta)^{-a-1}$ and for any $0 < u < \frac{W_{n,g}}{1+t_0}, 1 - \frac{u}{W_{n,g}} < 1$. Using these two facts,

$$\begin{aligned} E(\kappa_g^2 1\{\kappa_g \le \eta\}|\tau, \sigma^2, \mathcal{D}) &\le \frac{M}{c_0} \cdot \frac{1}{W_{n,g}^2} \cdot (1-\eta)^{-a-1} \frac{\int_0^{\eta W_{n,g}} \exp(-\frac{u}{2}) u^{\frac{m_g}{2}+a+1} du}{\int_0^{\frac{W_{n,g}}{1+t_0}} \exp(-\frac{u}{2}) u^{\frac{m_g}{2}+a-1} du} \\ &\le \frac{M}{c_0} \cdot \frac{1}{W_{n,g}^2} \cdot (1-\eta)^{-a-1} \frac{\int_0^{\infty} \exp(-\frac{u}{2}) u^{\frac{m_g}{2}+a+1} du}{\int_0^{\frac{W_{n,g}}{1+t_0}} \exp(-\frac{u}{2}) u^{\frac{m_g}{2}+a-1} du} \\ &= \frac{M}{c_0} \cdot \frac{1}{W_{n,g}^2} \cdot (1-\eta)^{-a-1} C_* \Big[\int_0^{\frac{W_{n,g}}{1+t_0}} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}+a-1} du \Big]^{-1}, \end{aligned} \tag{54}$$

where C_* is a generic constant independent of both $W_{n,g}$ and τ. Hence, using (51), (52) and (54), we finally obtain

$$E(\kappa_g^2|\tau, \sigma^2, \mathcal{D}) \le \frac{(a+\frac{m_g}{2})(1-\eta q)^a}{\tau^{2a}(\eta q)^{a+\frac{m_g}{2}} c_0} \exp\left(-\eta(1-q)\frac{W_{n,g}}{2}\right) + \frac{M}{c_0} \cdot \frac{1}{W_{n,g}^2} \cdot (1-\eta)^{-a-1} C_* \Big[\int_0^{\frac{W_{n,g}}{1+t_0}} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}+a-1} du \Big]^{-1}. \tag{55}$$

Next, under the assumption (A3), using (55), we have $W_{n,g} E(\kappa_g^2 1\{\kappa_g \le \eta\}|\tau, \sigma^2, \mathcal{D}) \le h_1(W_{n,g}, \tau)$ and $W_{n,g} E(\kappa_g^2 1\{\kappa_g > \eta\}|\tau, \sigma^2, \mathcal{D}) \le h_2(W_{n,g}, \tau)$ with the choices of $h_k(W_{n,g}, \tau), k = 1, 2$ stated before. This proves (49). For (50), first observe that, $\int_0^{\frac{W_{n,g}}{1+t_0}} \exp(-\frac{u}{2}) u^{\frac{m_g}{2}+a-1} du$ is increasing in $W_{n,g}$, that is,

$$\sup_{W_{n,g} > 2a\rho \frac{1}{\sqrt{\tau}\log(\frac{1}{\tau})}} \Big[\int_0^{\frac{W_{n,g}}{1+t_0}} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}+a-1} du \Big]^{-1} = \Big[\int_0^{\frac{2a\rho}{(1+t_0)(\sqrt{\tau})\log(\frac{1}{\tau})}} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}+a-1} du \Big]^{-1}. \tag{56}$$

This implies $h_k(W_{n,g}, \tau), k = 1, 2$ is a decreasing function of $W_{n,g}$. Also note that

$$\lim_{\tau \to 0} \int_0^{\frac{2a\rho}{(1+t_0)(\sqrt{\tau})\log(\frac{1}{\tau})}} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}+a-1} du = \int_0^{\infty} \exp\Big(-\frac{u}{2}\Big) u^{\frac{m_g}{2}+a-1} du = \frac{\Gamma(\frac{m_g}{2}+a)}{(\frac{1}{2})^{\frac{m_g}{2}+a}} \ge \Gamma(a). \tag{57}$$

Hence, using (57) and the monotonicity of $h_1(W_{n,g}, \tau)$, we have

$$\lim_{\tau \to 0} \sup_{W_{n,g} > 2a\rho \frac{1}{\sqrt{\tau}\log(\frac{1}{\tau})}} h_1(W_{n,g}, \tau) \lesssim \lim_{\tau \to 0} \sqrt{\tau} \log\Big(\frac{1}{\tau}\Big) = 0. \tag{58}$$

On the other hand, using the monotonicity of $h_2(W_{n,g}, \tau)$, for any $\rho > \frac{2}{\eta(1-q)}$,

$$\sup_{W_{n,g} > 2a\rho \frac{1}{\sqrt{\tau}\log(\frac{1}{\tau})}} h_2(W_{n,g}, \tau) \le \frac{\tau^{-2a-\frac{1}{2}}}{\log(\frac{1}{\tau})} \exp\left(-\frac{2a}{\sqrt{\tau}\log(\frac{1}{\tau})}\right). \tag{59}$$

Next, we prove that

$$\frac{\tau^{-2a-\frac{1}{2}}}{\log(\frac{1}{\tau})} \exp\left(-\frac{2a}{\sqrt{\tau}\log(\frac{1}{\tau})}\right) = o(\sqrt{\tau}\log\Big(\frac{1}{\tau}\Big)), \text{ as } \tau \to 0. \tag{60}$$

In order to show, (60) holds, it is enough to prove that $\log(A_\tau) \to \infty$ as $\tau \to 0$, where $A_\tau = \frac{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}{\frac{\tau^{-2a-\frac{1}{2}}}{\log\left(\frac{1}{\tau}\right)}\exp\left(-\frac{2a}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}\right)} = \tau^{2a+1}\exp\left(\frac{2a}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}\right)(\log\left(\frac{1}{\tau}\right))^2$.
Hence,

$$\begin{aligned}\log(A_\tau) &= (2a+1)\log\tau + \frac{2a}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)} + 2\log\log\left(\frac{1}{\tau}\right) \\ &= \frac{2a}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}[1 - \frac{2a+1}{2a}\sqrt{\tau}(\log\left(\frac{1}{\tau}\right))^2 + \frac{2}{2a}\sqrt{\tau}\log\left(\frac{1}{\tau}\right)\log\log\left(\frac{1}{\tau}\right)] \\ &= \frac{2a}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}[1+o(1)] \to \infty \text{ as } \tau \to 0.\end{aligned}$$

This implies (60) holds, and as a result using (59) and (58),

$$\lim_{\tau\to 0}\sup_{W_{n,g}>2a\rho\frac{1}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}} h_2(W_{n,g},\tau) \le \lim_{\tau\to 0}\frac{\tau^{-2a-\frac{1}{2}}}{\log\left(\frac{1}{\tau}\right)}\exp\left(-\frac{2a}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}\right) = \lim_{\tau\to 0} o(\sqrt{\tau}\log\left(\frac{1}{\tau}\right)) = 0. \tag{61}$$

Finally, combining (58) and (61), (50) holds. Now, we prove (47). Let $\epsilon > 0$ be given. Then

$$P(\sum_{g\in\mathcal{A}_n} U_{ng} > \epsilon) \le \sum_{g\in\mathcal{A}_n} P(U_{ng} > \frac{\epsilon}{|\mathcal{A}_n|}).$$

Let us fix some $c > 2$ and any $\rho > c$. Let B_n and C_n denote the events $\{U_{n,g} > \frac{\epsilon}{|\mathcal{A}_n|}\}$ and $\{W_{n,g} > 2a\rho\frac{1}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}\}$, respectively. Then,

$$\begin{aligned}\sum_{g\in\mathcal{A}_n} P(U_{n,g} > \frac{\epsilon}{|\mathcal{A}_n|}) &= \sum_{g\in\mathcal{A}_n} P(B_n) \\ &= \sum_{g\in\mathcal{A}_n} P(B_n\cap C_n) + \sum_{g\in\mathcal{A}_n} P(B_n\cap C_n^c) \\ &\le \sum_{g\in\mathcal{A}_n} P(B_n\cap C_n) + \sum_{g\in\mathcal{A}_n} P(C_n^c).\end{aligned} \tag{62}$$

Using (48) and (49) along with the form of $h_k(W_{n,g},\tau), k=1,2$, it follows that using (56) and (57) for sufficiently large n, under the assumption (C3),

$$\sup_{W_{n,g}>2a\rho\frac{1}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}} h_1(W_{n,g},\tau) \le \sqrt{\tau}\log\left(\frac{1}{\tau}\right) = o(\frac{1}{|\mathcal{A}_n|}). \tag{63}$$

Also note that for all sufficiently large n, using (59), (60) and (63),

$$\sup_{W_{n,g}>2a\rho\frac{1}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}} h_2(W_{n,g},\tau) \le \frac{\tau^{-2a-\frac{1}{2}}}{\log\left(\frac{1}{\tau}\right)} e^{-\frac{a\rho\eta(1-q)}{\sqrt{\tau}\log\left(\frac{1}{\tau}\right)}} = o(\sqrt{\tau}\log\left(\frac{1}{\tau}\right)) = o(\frac{1}{|\mathcal{A}_n|}). \tag{64}$$

Observe that, with the definition of B_n, and using (63) and (64) we have, for sufficiently large n,

$$\text{for all } g\in\mathcal{A}_n, P(B_n\cap C_n) = 0. \tag{65}$$

This implies the first term in the right-hand side of (62) goes to zero and we are left with the second term only. For the second term, under the assumption of (C2) and (C3), using similar set of arguments used in (19)-(33), we can show that

$$\lim_{n\to\infty}\sum_{g\in\mathcal{A}_n} P(C_n^c) = 0. \tag{66}$$

Since $\epsilon > 0$ is arbitrary, combining (48)-(66) ensures (47) holds. This completes the proof of Theorem 2. □

Proof of Theorem 3:

Proof. As noted before,

$$P(\widehat{\mathcal{A}}_n^{(\mathrm{EB})} \neq \mathcal{A}_n) \leq \sum_{g\in\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) < \frac{1}{2}) + \sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}). \tag{67}$$

To prove (67), it suffices to show

$$\sum_{g\in\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) < \frac{1}{2}) = o(1), \text{ as } n\to\infty, \tag{68}$$

and

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}) = o(1), \text{ as } n\to\infty, \tag{69}$$

both when $0.5 < a < 1$, and $a \geq 1$.
Note that, for fixed $\mathcal{D} = \{\mathbf{y}\}$ and σ^2, $E(\kappa_g \mid \tau, \sigma^2, \mathcal{D})$ is a non-increasing function of τ. Moreover, $\widehat{\tau}^{\mathrm{EB}} \geq \gamma_n$, where $\gamma_n = \frac{1}{G_n}$, for $n \geq 1$. Combining these two facts, we obtain

$$\sum_{g\in\mathcal{A}_n} P(E(\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}) \leq \sum_{g\in\mathcal{A}_n} P(E(\kappa_g \mid \gamma_n, \sigma^2, \mathcal{D}) > \frac{1}{2}). \tag{70}$$

Observe that, the sequence $\gamma_n = G_n^{-1}$ for $n \geq 1$ satisfies $\log\left(\frac{1}{\tau_n}\right) \asymp \log(G_n)$. Therefore, under assumptions (A1)-(A3) of Theorem 1, using the same set of arguments employed in proving Theorem 1 (when τ was a tuning parameter), one has

$$\lim_{n\to\infty}\sum_{g\in\mathcal{A}_n} P(E(\kappa_g \mid \gamma_n, \sigma^2, \mathcal{D}) > \frac{1}{2}) = 0. \tag{71}$$

Therefore, (70) and (71) together yield

$$\lim_{n\to\infty}\sum_{g\in\mathcal{A}_n} P(E(\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}) = 0,$$

which completes the proof of (68). Now, we are left to prove (69). Note that, for any $\alpha_n > 0$,

$$\begin{aligned} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}) &= P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2},\ \widehat{\tau}^{\mathrm{EB}} \leq 2\alpha_n) + \\ &\quad P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2},\ \widehat{\tau}^{\mathrm{EB}} > 2\alpha_n). \end{aligned} \tag{72}$$

To complete the proof, we now appropriately choose $\{\alpha_n\}_{n\geq 1} > 0$ such that $\alpha_n \to 0$ as $n \to \infty$ so that both

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} > 2\alpha_n) = o(1), \text{ as } n \to \infty \tag{73}$$

and

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) = o(1), \text{ as } n \to \infty. \tag{74}$$

For studying (73), we define

$$\widehat{\tau}_1 = \frac{1}{G_n}, \text{ and } \widehat{\tau}_2 = \frac{1}{c_2 G_n} \sum_{g=1}^{G_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\},$$

where $c_1 \geq 2$, and $c_2 \geq 1$.

Clearly,

$$\widehat{\tau}^{\text{EB}} = \max\{\widehat{\tau}_1, \widehat{\tau}_2\}.$$

Therefore we have,

$$\begin{aligned} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \ \widehat{\tau}^{\text{EB}} > 2\alpha_n) &\leq P(\widehat{\tau}^{\text{EB}} > 2\alpha_n) \\ &\leq P(\widehat{\tau}_1 > 2\alpha_n) + P(\widehat{\tau}_2 > 2\alpha_n). \end{aligned} \tag{75}$$

We note that if $\alpha_n > 0$ is such that

$$\frac{1}{G_n} \leq 2\alpha_n, \text{ for all sufficiently large } n, \tag{76}$$

whence

$$P(\widehat{\tau}_1 > 2\alpha_n) = 0, \text{ for all sufficiently large } n. \tag{77}$$

Thus, (75) coupled with (77) yields

$$P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \epsilon_0, \ \widehat{\tau}^{\text{EB}} > 2\alpha_n) \leq P(\widehat{\tau}_2 > 2\alpha_n), \tag{78}$$

for all sufficiently large n for α_n satisfying (76).
Let us define

$$\widehat{\tau}_3 = \frac{1}{c_2 G_n} \sum_{g \in \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\},$$

and

$$\widehat{\tau}_4 = \frac{1}{c_2 G_n} \sum_{g \notin \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\},$$

so that

$$\widehat{\tau}_2 = \widehat{\tau}_3 + \widehat{\tau}_4.$$

Clearly,

$$P(\widehat{\tau}_2 > 2\alpha_n) \leq P(\widehat{\tau}_3 > \alpha_n) + P(\widehat{\tau}_4 > \alpha_n). \tag{79}$$

Observe that

$$\widehat{\tau}_3 = \frac{1}{c_2 G_n} \sum_{g \in \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\} \leq \frac{1}{c_2 G_n} \sum_{g \in \mathcal{A}_n} 1 = \frac{G_{\mathcal{A}_n}}{c_2 G_n}. \tag{80}$$

We now observe that if $\alpha_n > 0$ is chosen such that

$$\frac{G_{\mathcal{A}_n}}{c_2 G_n} \leq \alpha_n, \text{ for all sufficiently large } n, \tag{81}$$

then

$$P(\widehat{\tau}_3 > \alpha_n) = 0, \text{ for all sufficiently large } n. \tag{82}$$

Note that condition (81) implies condition (76) and therefore (77) is automatically satisfied for this choice of α_n. For bounding the second term in the right-hand side of (79), we consider the generalized version of the Chernoff-Hoeffding bound for an independent but non-i.i.d. sequence of random variables, which is stated in the next lemma.

Lemma A.6. *Let $Z_1, Z_2, \cdots, Z_m$ be m independent $0-1$ random variables with $\mathbb{E}(Z_i) = p_i$, $i = 1, \ldots, m$. Let $Z = \sum_{i=1}^m Z_i$, $\mu = \mathbb{E}(Z) = \sum_{i=1}^m p_i$ and $p = \frac{\mu}{m}$. Then*

$$\mathbb{P}(Z \geq \mu + \lambda) \leq \exp\left\{ -\{mH_p(p + \frac{\lambda}{m})\} \right\}, \quad \text{for} \quad 0 < \lambda < m - \mu,$$

and

$$\mathbb{P}(Z \leq \mu - \lambda) \leq \exp\left\{ -\{mH_{1-p}(1 - p + \frac{\lambda}{m})\} \right\}, \quad \text{for} \quad 0 < \lambda < \mu,$$

where $H_p(x) = x \log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right)$ is the relative entropy of x with respect to p.

Note that

$$\begin{aligned}
P(\widehat{\tau}_2 > 2\alpha_n) &\leq P(\widehat{\tau}_4 > \alpha_n) \\
&= P\left[\sum_{g \notin \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\} > c_2 \alpha_n G_n \right] \\
&= P\left[\frac{1}{(G_n - G_{\mathcal{A}_n})} \sum_{g \notin \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\} > c_2 \frac{\alpha_n G_n}{(G_n - G_{\mathcal{A}_n})} \right] \\
&\leq P\left[\frac{1}{(G_n - G_{\mathcal{A}_n})} \sum_{g \notin \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\} \geq \alpha_n \right] \\
&= P\left[\sum_{g \notin \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\} \geq \alpha_n (G_n - G_{\mathcal{A}_n}) \right] \\
&= P\left[S_n - \mathbb{E}(S_n) \geq \alpha_n (G_n - G_{\mathcal{A}_n}) - \mathbb{E}(S_n) \right] \text{ say,}
\end{aligned} \tag{83}$$

where the second inequality holds due to $c_2 G_n > G_n - G_{\mathcal{A}_n}$ and

$$S_n = \sum_{g \notin \mathcal{A}_n} 1\left\{ \frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n \right\}.$$

To find an upper bound for (83), we use Lemma A.6 taking $m = G_n - G_{\mathcal{A}_n}$, $Z_i = 1\left\{\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n\right\}$, $\mathbb{E}(Z_i) = p_i = P\left(\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n\right) \leq m_g G_n^{-\frac{c_1}{2}} (\log G_n)^{\frac{m_g}{2}-1}$, $\mu \leq s G_n^{-\frac{c_1}{2}+1} (\log G_n)^{\frac{s}{2}-1}$, and $p = \frac{\mu}{G_n - G_{\mathcal{A}_n}} \leq s G_n^{-\frac{c_1}{2}} (\log G_n)^{\frac{s}{2}-1}$, where s denotes the maximum of the group size. For $\lambda = \alpha_n(G_n - G_{\mathcal{A}_n}) - \mu$, we have $0 < \lambda < m - \mu$. Also, we have $p + \frac{\lambda}{G_n - G_{\mathcal{A}_n}} = \alpha_n$. Hence,

$$H_p(p + \frac{\lambda}{G_n - G_{\mathcal{A}_n}}) = \alpha_n \log\left(\frac{\alpha_n}{p}\right) + (1-\alpha_n)\log\left(\frac{1-\alpha_n}{1-p}\right). \tag{84}$$

Also, note that, since $c_1 \geq 2$, $\frac{p}{\alpha_n} \to 0$ as $n \to \infty$ under the assumption that $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$.

Recall that, $\frac{\log\left(\frac{1}{1-y}\right)}{y} \to 1$ as $y \to 0$. Hence, with $y = \frac{p-\alpha_n}{1-\alpha_n}$, the second term in the right hand side of (84) is of the form

$$\log\left(\frac{1-\alpha_n}{1-p}\right) = \frac{p-\alpha_n}{1-\alpha_n}(1+o(1)), \tag{85}$$

where $o(1)$ depends only on n such that $\lim_{n\to\infty} o(1) = 0$. Hence, using (84) and (85), an lower bound of $H_p(\alpha_n)$ is given by

$$\begin{aligned} H_p(\alpha_n) &= \alpha_n \log\left(\frac{\alpha_n}{p}\right) + (1-\alpha_n) \cdot \frac{(p-\alpha_n)}{(1-\alpha_n)}(1+o(1)) \\ &= \alpha_n \log\left(\frac{\alpha_n}{p}\right)(1+o(1)) \\ &\gtrsim \alpha_n(1+o(1)). \end{aligned} \tag{86}$$

With the use of Lemma A.6 and (86) with the assumption $G_{\mathcal{A}_n} = o(G_n)$, the upper bound of (83) is obtained as

$$\begin{aligned} P(\widehat{\tau}_2 > 2\alpha_n) &\leq P\left[\frac{1}{(G_n - G_{\mathcal{A}_n})} \sum_{g \notin \mathcal{A}_n} 1\left\{\frac{n\widehat{\boldsymbol{\beta}}_g^{\mathrm{T}}\mathbf{Q}_{n,g}\widehat{\boldsymbol{\beta}}_g}{\sigma^2} > c_1 \log G_n\right\} \geq \alpha_n\right] \\ &\leq \exp\Big(-\alpha_n G_n (1+o(1))\Big) \\ &\leq \exp\left(-\frac{G_{\mathcal{A}_n}}{c_2}(1+o(1))\right), \end{aligned} \tag{87}$$

where inequality in the last line holds due to (81). Choosing $\alpha_n = c_2 \frac{G_{\mathcal{A}_n}}{G_n}$, we immediately see that $\alpha_n \to 0$ and satisfies (76) and (81). Thus, with the above choice of α_n, and combining (78) and (87), we obtain

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2},\ \widehat{\tau}^{\mathrm{EB}} > 2\alpha_n) \leq G_n \exp\left(-\frac{G_{\mathcal{A}_n}}{c_2}(1+o(1))\right).$$

Since, $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$, we can conclude that

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\mathrm{EB}} > 2\alpha_n) = o(1),\ \text{as } n \to \infty. \tag{88}$$

We now proceed to prove that

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) = o(1), \text{ as } n \to \infty. \tag{89}$$

Case-1 $a \geq 1$.
Now using the monotonicity of the shrinkage coefficient and then by Markov's inequality, the term in the left hand side of (89) can be bounded as

$$\begin{aligned}
&\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) \\
&\leq \sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) > \frac{1}{2}) \\
&\leq 2 \sum_{g \notin \mathcal{A}_n} E\left(E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D})\right),
\end{aligned} \tag{90}$$

where the outer expectation is w.r.t. to $W_{n,g} = \frac{n\widehat{\boldsymbol{\beta}}_g^{\text{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2}$. Since, by definition, $E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) \leq 1$, so, we have

$$\begin{aligned}
E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) &= E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) 1_{\{W_{n,g} > 2c_1 \log G_n\}} + E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) 1_{\{W_{n,g} \leq 2c_1 \log G_n\}} \\
&\leq 1_{\{W_{n,g} > 2c_1 \log G_n\}} + E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) 1_{\{W_{n,g} \leq 2c_1 \log G_n\}}.
\end{aligned} \tag{91}$$

Now consider Lemma 1 of Paul and Chakrabarti (2025), which states that:

Lemma A.7. *Suppose* $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\theta}, \mathbf{I}_n)$, *where* $\mathbf{X} = (X_1, \cdots, X_n)$ *and* $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$. *Consider the class of priors given by, for* $i = 1, \ldots, n$, $\theta_i | \lambda_i^2, \tau \overset{ind}{\sim} \mathcal{N}(0, \lambda_i^2 \tau^2)$ *and* $\lambda_i^2 \overset{ind}{\sim} \pi(\lambda_i^2) = K(\lambda_i^2)^{-a-1} L(\lambda_i^2)$ *with* $a \geq 1$. *Define* $\kappa_i = 1/(1+\lambda_i^2\tau^2), i = 1, \ldots, n$. *Then for any* $\tau \in (0,1)$ *and any* $x_i \in \mathbb{R}$,

$$E(1-\kappa_i | x_i, \tau) \leq \left(\tau^2 e^{\frac{x_i^2}{4}} + K\tau^2 \int_1^\infty \frac{1}{(1+t\tau^2)^{\frac{3}{2}}} t^{-a} L(t) e^{\frac{x_i^2}{2} \cdot \frac{t\tau^2}{1+t\tau^2}} dt\right)(1+o(1)), \tag{92}$$

where the term $o(1)$ *depends only on* τ *and* $\lim_{\tau \to 0} o(1) = 0$ *and* $\int_0^\infty t^{-a-1} L(t) dt = K^{-1}$.

Next, using arguments similar to Lemma A.7, we provide an upper bound on $E(1-\kappa_g \mid \tau, \sigma^2, \mathcal{D})$ for any $\tau \in (0,1)$.

$$E(1-\kappa_g \mid \tau, \sigma^2, \mathcal{D}) \leq \left(\tau^2 e^{\frac{W_{n,g}}{4}} + K \int_1^\infty \frac{t\tau^2}{1+t\tau^2} \cdot \frac{1}{(1+t\tau^2)^{\frac{m_g}{2}}} t^{-a-1} L(t) e^{\frac{t\tau^2}{1+t\tau^2} \cdot \frac{W_{n,g}}{2}} dt\right)(1+o(1)), \tag{93}$$

where the term $o(1)$ depends only on τ and $\lim_{\tau \to 0} o(1) = 0$ and $\int_0^\infty t^{-a-1} L(t) dt = K^{-1}$.
Now, using (91) and (93), we obtain, for any $\tau \in (0,1)$

$$\begin{aligned}
&E\left(E(1-\kappa_g \mid \tau, \sigma^2, \mathcal{D})\right) \\
&\leq P(W_{n,g} > 2c_1 \log G_n) + E\left[\left(\tau^2 e^{\frac{W_{n,g}}{4}} + K \int_1^\infty \frac{t^{-a}\tau^2}{(1+t\tau^2)^{\frac{m_g}{2}+1}} L(t) e^{\frac{t\tau^2}{1+t\tau^2} \cdot \frac{W_{n,g}}{2}} dt\right)(1+o(1)) 1_{\{W_{n,g} \leq 2c_1 \log G_n\}}\right]
\end{aligned} \tag{94}$$

Since, for all $g \notin \mathcal{A}_n$, $W_{n,g} \sim \chi^2_{m_g}$, so, by Lemma A.5,

$$P(W_{n,g} > 2c_1 \log G_n) \lesssim G_n^{-c_1} (\log G_n)^{\frac{s}{2}-1}. \tag{95}$$

Next, for the second term in the right-hand side of (94), noting that the term $(1+o(1))$ is independent of any particular g,

$$\begin{aligned} E\Big[\tau^2 e^{\frac{W_{n,g}}{4}}(1+o(1))1_{\{W_{n,g}\le 2c_1\log G_n\}}\Big] &= \tau^2 \frac{1}{2^{\frac{m_g}{2}}\Gamma(\frac{m_g}{2})}\int_0^{2c_1\log G_n} e^{\frac{u}{4}}\exp\Big(-\frac{u}{2}\Big)u^{\frac{m_g}{2}-1}du(1+o(1)) \\ &= \tau^2 \frac{1}{\Gamma(\frac{m_g}{2})}\int_0^{c_1\log G_n} \exp\Big(-\frac{u}{2}\Big)u^{\frac{m_g}{2}-1}du(1+o(1)) \\ &\lesssim \tau^2(1+o(1)). \end{aligned} \tag{96}$$

Finally, for the third term in the right-hand side of (94), note that

$$\begin{aligned} &\int_0^{2c_1\log G_n}\int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot\frac{1}{(1+t\tau^2)^{\frac{m_g}{2}}}t^{-a-1}L(t)e^{\frac{t\tau^2}{1+t\tau^2}\cdot\frac{u}{2}}\exp\Big(-\frac{u}{2}\Big)u^{\frac{m_g}{2}-1}dtdu \\ &= \int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot\frac{1}{(1+t\tau^2)^{\frac{m_g}{2}}}t^{-a-1}L(t)\left(\int_0^{2c_1\log G_n} e^{-\frac{u}{2(1+t\tau^2)}}u^{\frac{m_g}{2}-1}du\right)dt \\ &= \int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-a-1}L(t)\left(\int_0^{\frac{2c_1\log G_n}{1+t\tau^2}} e^{-\frac{z}{2}}z^{\frac{m_g}{2}-1}dz\right)dt. \end{aligned} \tag{97}$$

We now provide an upper bound on (97) separately for $a=1$ and $a>1$.
For $a>1$, using the boundedness of $L(t)$, it is easy to show that,

$$\int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-a-1}L(t)\left(\int_0^{\frac{2c_1\log G_n}{1+t\tau^2}} e^{-\frac{z}{2}}z^{\frac{m_g}{2}-1}dz\right)dt \lesssim \tau^2. \tag{98}$$

For $a=1$, we have the following

$$\begin{aligned} &\int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-a-1}L(t)\left(\int_0^{\frac{2c_1\log G_n}{1+t\tau^2}} e^{-\frac{z}{2}}z^{\frac{m_g}{2}-1}dz\right)dt \\ &\le \int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-a-1}L(t)\left(\int_0^{\frac{2c_1\log G_n}{t\tau^2}} e^{-\frac{z}{2}}z^{\frac{m_g}{2}-1}dz\right)dt \\ &\le C_7\int_1^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-a-1}L(t)dt, \end{aligned}$$

where C_7 is a global constant independent of n. Next, note that, $1+t\tau^2 \ge t\tau^2$ for any $t>0$ and $t\tau^2 \ge \sqrt{t}$ if and only if $t \ge \frac{1}{\tau^4}$. As a result, we have

$$\int_1^{\frac{1}{\tau^4}} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-2}L(t)dt \le M\tau^2\int_1^{\frac{1}{\tau^4}} t^{-1}dt = M\tau^2\log\Big(\frac{1}{\tau^4}\Big), \tag{99}$$

and

$$\int_{\frac{1}{\tau^4}}^{\infty} \frac{t\tau^2}{1+t\tau^2}\cdot t^{-2}L(t)dt \le M\tau^2\int_{\frac{1}{\tau^4}}^{\infty} t^{-\frac{3}{2}}dt = 2M\tau^4. \tag{100}$$

Hence, for $a = 1$, the upper bound on (97) is obtained as

$$\int_1^\infty \frac{t\tau^2}{1+t\tau^2} \cdot t^{-a-1}L(t) \left(\int_0^{\frac{2c_1 \log G_n}{1+t\tau^2}} e^{-\frac{z}{2}} z^{\frac{m_g}{2}-1} dz \right) dt \leq 2M\tau^2 [\log\left(\frac{1}{\tau^4}\right) + \tau^2]. \tag{101}$$

As a consequence, for $a \geq 1$, the upper bound of the third term in the right-hand side of (94) is of the form

$$E\left[K \int_1^\infty \frac{t\tau^2}{1+t\tau^2} \cdot \frac{1}{(1+t\tau^2)^{\frac{m_g}{2}}} t^{-a-1}L(t) e^{\frac{t\tau^2}{1+t\tau^2}\cdot \frac{W_{n,g}}{2}} dt (1+o(1)) 1_{\{W_{n,g} \leq c_1 \log G_n\}}\right] \lesssim \tau^2 [\log\left(\frac{1}{\tau^2}\right) + \tau^2](1+o(1)). \tag{102}$$

Finally, substituting the upper bounds obtained in (95), (96) and (102) in (94) with $\tau = \alpha_n = \frac{c_2 G_{\mathcal{A}_n}}{G_n}$, the upper bound on (89) can be obtained as

$$\begin{aligned}
&\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) \\
&\lesssim G_n^{-c_1+1} (\log G_n)^{\frac{s}{2}-1} + \frac{(G_{\mathcal{A}_n})^2}{G_n}(1+o(1)) + \frac{(G_{\mathcal{A}_n})^2}{G_n} \log G_n (1+o(1)).
\end{aligned}$$

Hence, for $c_1 \geq 2$ and $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$, we have

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) = o(1), \text{ as } n \to \infty. \tag{103}$$

This completes the proof for $a \geq 1$.
Case-II $\frac{1}{2} < a < 1$. Again using the monotonicity of the shrinkage coefficient, the term in the left-hand side of (89) can be bounded as

$$\begin{aligned}
&\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) \\
&\leq \sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid 2\alpha_n, \sigma^2, \mathcal{D}) > \frac{1}{2}) \\
&\lesssim G_n \left(\frac{G_{\mathcal{A}_n}}{G_n}\right)^{2a} \left[\log\left(\frac{G_n}{G_{\mathcal{A}_n}}\right)\right]^{\frac{s}{2}-1} (1+o(1)).
\end{aligned}$$

Here, inequality in the last line follows due to arguments similar to the proof of Theorem 1, where τ is assumed to be a tuning one.
Note that, for $\frac{1}{2} < a < 1$, there exists $\epsilon \in (0, 1-\frac{1}{2a})$ such that $2a(1-\epsilon) > 1$. Hence, when $\frac{1}{2} < a < 1$, for $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some $0 < \epsilon_1 < \epsilon_2 < 1 - \frac{1}{2a}$, we conclude

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}}, \sigma^2, \mathcal{D}) > \frac{1}{2}, \widehat{\tau}^{\text{EB}} \leq 2\alpha_n) = o(1), \text{ as } n \to \infty \tag{104}$$

and the proof is completed for $\frac{1}{2} < a < 1$. □

Proof of Theorem 4:

Proof. To prove Theorem 4, we employ a similar set of arguments used in the proof of Theorem 2 when τ is used as a tuning parameter. Here, T_1 is as defined in Theorem 2 and $T_2 = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{A}_n}^{\frac{1}{2}} (\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n,EB}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n})$.
As in Theorem 2, it follows that for proving $T_2 \xrightarrow{P} 0$, it suffices to show that

$$\sum_{g \in \mathcal{A}_n} U_{ng} \xrightarrow{P} 0 \text{ as } n \to \infty, \tag{105}$$

where

$$W_{n,g} = \frac{n \widehat{\boldsymbol{\beta}}_g^{\mathrm{T}} \mathbf{Q}_{n,g} \widehat{\boldsymbol{\beta}}_g}{\sigma^2}, \text{ and } U_{n,g} = W_{n,g} E(\kappa_g^2 \mid \widehat{\tau}^{\mathrm{EB}}, \sigma^2, \mathcal{D}).$$

Now to establish (105), let us first fix any $\epsilon_0 > 0$, and take $\gamma_n = \frac{1}{G_n}$, for $n \geq 1$. Note that, for fixed $\mathcal{D} = \{\mathbf{y}\}$ and σ^2, $E(\kappa_g \mid \tau, \sigma^2, \mathcal{D})$ is non-increasing in τ, and $\widehat{\tau}^{\mathrm{EB}} \geq \gamma_n$ for all $n \geq 1$. Therefore, for $\tau_n = \gamma_n$ and using the monotonicity of τ, we only need to show that

$$\sum_{g \in \mathcal{A}_n} U_{n,g}^* = \sum_{g \in \mathcal{A}_n} W_{n,g} E(\kappa_g^2 \mid \gamma_n, \sigma^2, \mathcal{D}) \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{106}$$

Next, we proceed following the steps as mentioned in (48)-(62) with $\tau_n = \frac{1}{G_n}$. Now we define C_n as $C_n = \{W_{n,g} > 2a\rho \frac{\sqrt{G_n}}{\log G_n}\}$ and B_n as $B_n = \{U_{n,g}^* > \frac{\epsilon_0}{|\mathcal{A}_n|}\}$. Hence, to obtain the optimal estimation rate, it is enough to show that

$$\lim_{n \to \infty} \sum_{g \in \mathcal{A}_n} P(B_n) = 0. \tag{107}$$

Here, observe that under the assumption that $|\mathcal{A}_n| = O(G_n^{\epsilon}), 0 < \epsilon < \frac{1}{2}$, we have, for $k = 1, 2$,

$$\sup_{W_{n,g} > 2a\rho \frac{\sqrt{G_n}}{\log G_n}} h_k(W_{n,g}, \frac{1}{G_n}) = o(\frac{1}{|\mathcal{A}_n|}).$$

As a consequence of this, we have for all $g \in \mathcal{A}_n$,

$$P(B_n \cap C_n) = 0.$$

This also ensures

$$\lim_{n \to \infty} \sum_{g \in \mathcal{A}_n} P(B_n \cap C_n) = 0. \tag{108}$$

Next note that for $\tau_n = \frac{1}{G_n}$, we have $G_n \sqrt{\tau_n} \log\left(\frac{1}{\tau_n}\right) \to \infty$ as $n \to \infty$, and hence using similar set of arguments used in (19)-(33), we can show that

$$\lim_{n \to \infty} \sum_{g \in \mathcal{A}_n} P(C_n^c) = 0. \tag{109}$$

Finally combining (108) and (109) implies (107) and (106) holds. As a result, (105) is also established, and using Slutsky's Lemma, the proof of Theorem 4 is completed. $\square$

Proof of Theorem 5:

Proof. Here also, we will show that

$$\sum_{g\in\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) < \frac{1}{2}) = o(1), \text{ as } n\to\infty, \tag{110}$$

and

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) > \frac{1}{2}) = o(1), \text{ as } n\to\infty, \tag{111}$$

both when $0.5 \leq a < 1$, and $a \geq 1$, where $\widehat{\tau}^{\text{EB}}$ is defined in (24) in the main document.
Now, using the same technique as used in the proof of Theorem 3 and taking $\gamma_n = (\frac{1}{G_n})^2$, for $n \geq 1$, and then following steps similar to (70) and (71), we obtain

$$\lim_{n\to\infty}\sum_{g\in\mathcal{A}_n} P(E(\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) > \frac{1}{2}) = 0,$$

which completes the proof of (110). Now, we are left to prove (111). Again, following the same steps as given in the proof of Theorem 3, along with the use of (87) with the use of (23), we obtain

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) > \frac{1}{2},\ (\widehat{\tau}^{\text{EB}})^{\frac{1}{2}} > 2\alpha_n) \leq G_n e^{-\frac{G_{\mathcal{A}_n}}{c_2}(1+o(1))}.$$

Since, $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$, we can conclude that

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) > \frac{1}{2}, (\widehat{\tau}^{\text{EB}})^{\frac{1}{2}} > 2\alpha_n) = o(1), \text{ as } n\to\infty. \tag{112}$$

Next using the monotonicity of $E(1-\kappa_g \mid \tau,\sigma^2,\mathcal{D})$, it follows

$$\begin{aligned}\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) > \frac{1}{2},\ (\widehat{\tau}^{\text{EB}})^{\frac{1}{2}} \leq 2\alpha_n) &\leq \sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid (2\alpha_n)^2,\sigma^2,\mathcal{D}) > \frac{1}{2}) \\ &\lesssim G_n\alpha_n^2[\log\left(\frac{1}{\alpha_n}\right)]^{\frac{s}{2}}.\end{aligned}$$

With the choice of $\alpha_n = \frac{c_2 G_{\mathcal{A}_n}}{G_n}$, $c_2 \geq 1$ for all $n \geq 1$, and since, $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$, we conclude that

$$\sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \widehat{\tau}^{\text{EB}},\sigma^2,\mathcal{D}) > \frac{1}{2}, (\widehat{\tau}^{\text{EB}})^{\frac{1}{2}} \leq 2\alpha_n) = o(1), \text{ as } n\to\infty. \tag{113}$$

Combining (112) and (113), we obtain (111) and the proof is completed. □

Proof of Theorem 6:

Proof. To show that the main document's decision rule (16) is an oracle, we first show the selection consistency part. Using similar arguments used before, again, we have

$$P(\widehat{\mathcal{A}}_n^{(\text{FB})} \neq \mathcal{A}_n) \leq \sum_{g\in\mathcal{A}_n} P(E(1-\kappa_g \mid \sigma^2,\mathcal{D}) < \frac{1}{2}) + \sum_{g\notin\mathcal{A}_n} P(E(1-\kappa_g \mid \sigma^2,\mathcal{D}) > \frac{1}{2}). \tag{114}$$

Here also, our target is to show the following:

$$\sum_{g \in \mathcal{A}_n} P(E(1-\kappa_g \mid \sigma^2, \mathcal{D}) < \frac{1}{2}) = o(1), \text{ as } n \to \infty, \tag{115}$$

and

$$\sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \sigma^2, \mathcal{D}) > \frac{1}{2}) = o(1), \text{ as } n \to \infty, \tag{116}$$

both when $0.5 \leq a < 1$, and $a \geq 1$.
In order to show (115), first note that with the use of D1,

$$\begin{aligned} E(\kappa_g \mid \sigma^2, \mathcal{D}) &= \int_{\gamma_{1n}}^{\gamma_{2n}} E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \pi(\tau_n|\mathcal{D}) d\tau \\ &\leq E(\kappa_g \mid \gamma_{1n}, \sigma^2, \mathcal{D}). \end{aligned}$$

Here, first, we use the fact that given τ_n, σ^2 and $\mathcal{D}$, the posterior mean of κ_g depends only on g^{th} group. Inequality in the last line follows since, for any fixed τ_n and σ^2, $E(\kappa_g \mid \tau_n, \sigma^2, \mathcal{D})$ is non-increasing in τ. As a result of this, we have

$$\sum_{g \in \mathcal{A}_n} P(E(1-\kappa_g \mid \sigma^2, \mathcal{D}) < \frac{1}{2}) \leq \sum_{g \in \mathcal{A}_n} P(E(\kappa_g \mid \gamma_{1n}, \sigma^2, \mathcal{D}) > \frac{1}{2}). \tag{117}$$

Next using the same steps as used in (32)-(33) with $\tau_n = \gamma_{1n}$ and noting that $\log\left(\frac{1}{\gamma_{1n}}\right) \asymp \log(G_n)$, along with the use of (117), (115) is established.
In case of proving (116), note that

$$\begin{aligned} E(1-\kappa_g \mid \sigma^2, \mathcal{D}) &= \int_{\gamma_{1n}}^{\gamma_{2n}} E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \pi(\tau_n|\mathcal{D}) d\tau \\ &\leq E(1-\kappa_g \mid \gamma_{2n}, \sigma^2, \mathcal{D}). \end{aligned}$$

Here, first, we use the fact that given τ_n, σ^2 and $\mathcal{D}$, the posterior mean of $1-\kappa_g$ depends only on g^{th} group. Inequality in the last line follows since, for any fixed τ_n and σ^2, $E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D})$ is non-decreasing in τ. As a consequence, it follows that

$$\begin{aligned} \sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \sigma^2, \mathcal{D}) > \frac{1}{2}) &\leq \sum_{g \notin \mathcal{A}_n} P(E(1-\kappa_g \mid \gamma_{2n}, \sigma^2, \mathcal{D}) > \frac{1}{2}) \\ &\lesssim G_n \gamma_{2n} [\log\left(\frac{1}{\gamma_{2n}}\right)]^{\frac{s}{2}-1}(1+o(1)). \end{aligned}$$

Inequality in the last line follows due to arguments similar to those used in (34)-(38). Finally, under the assumption that $G_n \gamma_{2n} [\log\left(\frac{1}{\gamma_{2n}}\right)]^{\frac{s}{2}-1} \to 0$ as $n \to \infty$, (116) also holds and completes the proof related to variable selection consistency.
Next, we move forward to show the optimal estimation. Here we aim to show,

$$\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{A}_n}^{\frac{1}{2}} (\widehat{\boldsymbol{\beta}}_{\mathcal{A}_n, FB}^{\mathrm{HT}} - \widehat{\boldsymbol{\beta}}_{\mathcal{A}_n}) \xrightarrow{P} 0 \text{ as } n \to \infty.$$

Now making use of the same arguments as used in (41)-(47) of Theorem 2, we only need to show that

$$\sum_{g \in \mathcal{A}_n} W_{n,g} E(\kappa_g^2 \mid \gamma_{1n}, \sigma^2, \mathcal{D}) \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{118}$$

Next we follow the same steps as mentioned in (48)-(66) with $\tau = \gamma_{1n}$. Note that since γ_{1n} satisfies both $G_n\sqrt{\gamma_{1n}}\log\left(\frac{1}{\gamma_{1n}}\right) \to \infty$ as $n \to \infty$ and $\log\left(\frac{1}{\gamma_{1n}}\right) \asymp \log(G_n)$, it is immediate that (118) holds and as a result, $T_2 \xrightarrow{P} 0$ as $n \to \infty$. Finally, the use of Slutsky's Theorem completes the proof of Theorem 6.

□

B Property of the Half-Thresholding (HT) estimator

In the main document, we have stated the results related to the oracle property of our proposed decision rules, irrespective of the level of sparsity to be known or unknown. However, along with the variable selection consistency, one may also be interested in studying the accuracy of the half-thresholding(HT) estimator with respect to an appropriate loss function. Theorem 7, stated and proved below, indicates that in an extremely sparse situation with no signals or in the situation where the true group coefficients are allowed to grow only up to a polynomial rate in n, our HT estimator has a clear edge over the traditional Bayes estimator.

Theorem 7. *Consider the hierarchical framework of (5) where $\pi(\lambda_g^2)$ is as in (6) with $a \geq \frac{1}{2}$ and the half-thresholding (HT) rule (9) based on these. Let $\mathcal{A}_n = \{g : \boldsymbol{\beta}_g^0 \neq \mathbf{0}\}$ and $\widehat{\mathcal{A}}_n = \{g : \widehat{\boldsymbol{\beta}}_g^{HT} \neq \mathbf{0}\}$ denote respectively the set of truly active groups, and the set of groups declared active by the half-thresholding rule (9). Let $\mathbf{Q}_{n,g} = \mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g/n$, for $g = 1, \cdots, G$. Then, we have the following.*
(i) In a sparse situation, under the assumption (A3), if all groups are inactive, our proposed HT estimator of $\boldsymbol{\beta}_g$ is more accurate than that of the posterior mean with respect to squared error loss, i.e., for all sufficiently large n,

$$\sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{HT} - \boldsymbol{\beta}_g^0\right\|_2^2 < \sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{PM} - \boldsymbol{\beta}_g^0\right\|_2^2. \tag{119}$$

(ii) Along with (A2)-(A4) and (C1), we further assume that for any active group $g \in \mathcal{A}_n$, $\max_j |\beta_{gj}^0| \leq n^{C_3}$, for some $0 < C_3 < \infty$. Then also our proposed HT estimator of $\boldsymbol{\beta}_g$ is more accurate than that of the posterior mean with respect to squared error loss, i.e., for all sufficiently large n,

$$\sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{HT} - \boldsymbol{\beta}_g^0\right\|_2^2 < \sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{PM} - \boldsymbol{\beta}_g^0\right\|_2^2. \tag{120}$$

Proof of Theorem 7:

Proof. The proof of (119) follows as a byproduct of the proof of (120). Hence, we now focus on proving (120). From the definition of the mean squared error of an estimator, one can write the mean squared error of the half-thresholding estimator $\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}}$ as

$$\begin{aligned} E_{\boldsymbol{\beta}^0}\left\|\widehat{\boldsymbol{\beta}}^{\mathrm{HT}} - \boldsymbol{\beta}^0\right\|_2^2 &= \sum_{g \in \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} - \boldsymbol{\beta}_g^0\right\|_2^2 + \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} - \boldsymbol{\beta}_g^0\right\|_2^2 \\ &= \sum_{g \in \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} - \boldsymbol{\beta}_g^0\right\|_2^2 + \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}}\right\|_2^2, \end{aligned} \tag{121}$$

where the last line follows from the fact $\boldsymbol{\beta}_g^0 = \mathbf{0}$ for each $g \notin \mathcal{A}_n$. Similarly, the mean squared error of the posterior mean $\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}$ can be obtained as

$$E_{\boldsymbol{\beta}^0}\left\|\widehat{\boldsymbol{\beta}}^{\mathrm{PM}} - \boldsymbol{\beta}^0\right\|_2^2 = \sum_{g \in \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \boldsymbol{\beta}_g^0\right\|_2^2 + \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2. \tag{122}$$

Hence, to establish part (ii) of Theorem 7, it would be enough to show that, for all sufficiently large n,

$$\sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} \right\|_2^2 = \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 + e_{1n}, \text{ with } e_{1n} < 0, \tag{123}$$

and

$$\sum_{g \in \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 = \sum_{g \in \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \boldsymbol{\beta}_g^0 \right\|_2^2 + e_{2n}, \text{ where } e_{2n} = o(e_{1n}). \tag{124}$$

Proof of (123): From the definition of the half-thresholding estimate, $\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}}$ given in (11) of the main document, it follows that,

$$\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} \right\|_2^2 = \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 I\big\{ E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le 0.5 \big\}. \tag{125}$$

Hence, for any g, $\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} \right\|_2^2 \le \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2$. As a result, for obtaining (123), it is enough to establish that for any $g \notin \mathcal{A}_n$, $\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} \right\|_2^2 > 0$ with some positive probability. Therefore, we are interested in exploring the behavior of $\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 I\big\{ E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le 0.5 \big\}$ for $g \notin \mathcal{A}_n$.

Observe that, from Proposition 1, it follows that

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \xrightarrow{P} 0 \text{ as } n \to \infty, \text{ uniformly in } g \notin \mathcal{A}_n.$$

This implies that, for any $g \notin \mathcal{A}_n$

$$I\big\{ E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le 0.5 \big\} = 1 \text{ on the set } B_{1n} \text{ and } \lim_{n\to\infty} P_{\boldsymbol{\beta}_g^0 = 0}(B_{1n}) = 1, \tag{126}$$

where $B_{1n} = \big\{ E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le 0.5 \big\}$. Again, for any $g \notin \mathcal{A}_n$, using the definition of $\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}$,

$$\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 > 0 \text{ on the set } B_{2n} \text{ and } P_{\boldsymbol{\beta}_g^0 = 0}(B_{2n}) = 1, \text{ for all } n, \tag{127}$$

where $B_{2n} = \left\{ \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 > 0 \right\}$. Hence, combining (125)- (127) along with the definitions of B_{1n} and B_{2n}, we have

$$\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 I\big\{ E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le 0.5 \big\} > 0 \text{ on the set } B_{1n} \cap B_{2n},$$

with

$$P_{\boldsymbol{\beta}_g^0 = 0}(B_{1n} \cap B_{2n}) \ge P_{\boldsymbol{\beta}_g^0 = 0}(B_{1n}) + P_{\boldsymbol{\beta}_g^0 = 0}(B_{2n}) - 1 \to 1 \text{ as } n \to \infty. \tag{128}$$

Hence, for all sufficiently large n, uniformly in $g \notin \mathcal{A}_n$, $\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} \right\|_2^2 > 0$ on the set $B_{1n} \cap B_{2n}$ with $P_{\boldsymbol{\beta}_g^0 = 0}(B_{1n} \cap B_{2n}) \to 1$. This along with (125) implies,
$\sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0 = 0} \left[\left\| \widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} \right\|_2^2 I\big\{ E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \le 0.5 \big\} \right] > 0$ for sufficiently large n.

Let us define,

$$e_{1n} = \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} \right\|_2^2 - \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \right\|_2^2$$
$$= - \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0 = 0} \left[\left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \right\|_2^2 I\big\{ E(1 - \kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5 \big\} \right].$$

Then, using our previous arguments, it follows that $e_{1n} < 0$ for sufficiently large n.
Therefore, for all sufficiently large n, we have

$$\sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} \right\|_2^2 = \sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \right\|_2^2 + e_{1n}, \text{ with } e_{1n} < 0,$$

which completes the proof of (123).

It is important to observe that, when all groups are inactive, $\mathcal{A}_n = \emptyset$. In this case, the first terms in both (121) and (122) will disappear. Consequently, (119) trivially follows from here. This concludes the proof of part (i) of the present theorem.

Now we prove (124).
Proof of (124): Note that, for any $g \in \mathcal{A}_n$

$$E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 = E_{\boldsymbol{\beta}_g^0} \left[\left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} = \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \} \right] + E_{\boldsymbol{\beta}_g^0} \left[\left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} \neq \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \} \right]$$
$$= E_{\boldsymbol{\beta}_g^0} \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \boldsymbol{\beta}_g^0 \right\|_2^2 + E_{\boldsymbol{\beta}_g^0} \left[\left(\left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \boldsymbol{\beta}_g^0 \right\|_2^2 \right) 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} \neq \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \} \right]. \tag{129}$$

Next, observe that, using the definition of $\widehat{\boldsymbol{\beta}}_g^{\text{HT}}$, we have

$$1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} \neq \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \} = 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} = \mathbf{0} \}.$$

Using this fact, the second term in the right hand side of (129) is of the form

$$E_{\boldsymbol{\beta}_g^0} \left[\left(\left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \boldsymbol{\beta}_g^0 \right\|_2^2 \right) 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} \neq \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \} \right]$$
$$= E_{\boldsymbol{\beta}_g^0} \left[\left(\left\| \widehat{\boldsymbol{\beta}}_g^{\text{HT}} - \boldsymbol{\beta}_g^0 \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \boldsymbol{\beta}_g^0 \right\|_2^2 \right) 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} = \mathbf{0} \} \right]$$
$$= E_{\boldsymbol{\beta}_g^0} \left[\left(\left\| \mathbf{0} - \boldsymbol{\beta}_g^0 \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \boldsymbol{\beta}_g^0 \right\|_2^2 \right) 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} = \mathbf{0} \} \right]$$
$$= E_{\boldsymbol{\beta}_g^0} \left[\left(2(\boldsymbol{\beta}_g^0)^{\text{T}} \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \right\|_2^2 \right) 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} = \mathbf{0} \} \right]. \tag{130}$$

Hence, for proving (124), first we need to obtain an upper bound of
$\sum_{g \in \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0} \left[\left(2(\boldsymbol{\beta}_g^0)^{\text{T}} \widehat{\boldsymbol{\beta}}_g^{\text{PM}} - \left\| \widehat{\boldsymbol{\beta}}_g^{\text{PM}} \right\|_2^2 \right) 1\{ \widehat{\boldsymbol{\beta}}_g^{\text{HT}} = \mathbf{0} \} \right]$.

Note that,

$$
\begin{aligned}
E_{\boldsymbol{\beta}_g^0}\left(2(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right)^2 &\leq 2\left[E_{\boldsymbol{\beta}_g^0}\left(2(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right)^2 + E_{\boldsymbol{\beta}_g^0}\left(\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right)^2\right] \\
&\leq 8(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\boldsymbol{\beta}_g^0 E_{\boldsymbol{\beta}_g^0}\left(\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right) + 2E_{\boldsymbol{\beta}_g^0}\left(\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right)^2 \\
&\leq 8(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\boldsymbol{\beta}_g^0 E_{\boldsymbol{\beta}_g^0}\left(\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2\right) + 2E_{\boldsymbol{\beta}_g^0}\left(\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2\right)^2. \qquad (131)
\end{aligned}
$$

In the above chain of inequalities, the first inequality holds because of the fact $(a-b)^2 \leq 2(a^2+b^2)$, for any two real numbers a and b. The second inequality here follows from the Cauchy-Schwarz inequality, while the last one follows from the definition of $\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}$, and the fact $\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2 \leq \left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2$, for any arbitrary group g.

Since, $\widehat{\boldsymbol{\beta}}_g \sim \mathcal{N}_{m_g}(\boldsymbol{\beta}_g, \sigma^2(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1})$, using standard theory of multivariate normal distributions, it follows that

$$
\begin{aligned}
E\left(\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2\right) &= \|\boldsymbol{\beta}_g^0\|_2^2 + \sigma^2 Tr((\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1}) \\
&= \|\boldsymbol{\beta}_g^0\|_2^2 + \frac{\sigma^2}{n}\sum_{j=1}^{m_g}\eta_j \\
&\leq \|\boldsymbol{\beta}_g^0\|_2^2 + \frac{\sigma^2 C_3 s}{n}, \qquad (132)
\end{aligned}
$$

where η_j is the j^{th} diagonal element of $n(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1}$. Note that, (C1) implies η_j is bounded for all $j = 1, \ldots, m_g$. Also, observe that, $E((\sum_{j=1}^{m_g}\widehat{\beta}_{gj})^2)^2 \leq m_g \sum_{j=1}^{m_g} E((\widehat{\beta}_{gj})^4)$. As a result,

$$
\begin{aligned}
E((\widehat{\beta}_{gj})^4) &= (\beta_{gj}^0)^4 + 6\sigma^2\eta_j\frac{(\beta_{gj}^0)^2}{n} + 3\frac{\sigma^4}{n^2}(\eta_j)^2 \\
&\leq (\beta_{gj}^0)^4 + 6\frac{\sigma^2 C_3}{n}(\beta_{gj}^0)^2 + 3\frac{\sigma^4 C_3}{n^2}. \qquad (133)
\end{aligned}
$$

Combining (131)-(133), we obtain, for all $g \in \mathcal{A}_n$

$$
E_{\boldsymbol{\beta}_g^0}\left(2(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right)^2 \leq C_5 \max_{j=1,\ldots,m_g}(\beta_{gj}^0)^4, \qquad (134)
$$

for some arbitrary constant C_5, independent of both g and j. On the other hand, using the definition of $\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}}$ followed by (18) and (23), we have

$$
\begin{aligned}
E_{\boldsymbol{\beta}_g^0}[1\{\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} = \mathbf{0}\}] &= P_{\boldsymbol{\beta}_g^0}(E(1-\kappa_g|\tau,\sigma^2,\mathcal{D}) \leq \frac{1}{2}) \\
&\leq \frac{C_7}{\sqrt{n}m_n}\exp\big(-C_8 n m_n^2\big)(1+o(1)), \qquad (135)
\end{aligned}
$$

where $o(1)$ is independent of any $g \in \mathcal{A}_n$ and C_7 and C_8 are some constants independent of any $g \in \mathcal{A}_n$.

Hence, using (134), (135) and by Cauchy-Schwarz inequality along with (A2), for all sufficiently large n,

$$\sum_{g\in\mathcal{A}_n} E_{\boldsymbol{\beta}_g^0}\left[\left(2(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right)1\{\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} = \mathbf{0}\}\right] \leq \sum_{g\in\mathcal{A}_n}\sqrt{E_{\boldsymbol{\beta}_g^0}\left(2(\boldsymbol{\beta}_g^0)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2\right)^2}\sqrt{E_{\boldsymbol{\beta}_g^0}[1\{\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} = \mathbf{0}\}]}$$
$$\lesssim \sum_{g\in\mathcal{A}_n}\max_{j=1,\ldots,m_g}(\beta_{gj}^0)^2 n^{\frac{b}{2}-\frac{1}{4}}\exp\big(-0.5C_8 n^{1-2b}\big)(1+o(1))$$
$$\lesssim n^{\tilde{C}_1}\exp\Big(\tilde{C}_2\times n^{\tilde{C}_3}\Big)(1+o(1)), \tag{136}$$

for some $\tilde{C}_i > 0$, $i = 1, 2, 3$. Here the last line holds since for all $g \in \mathcal{A}_n$, $\max_{j=1,\ldots,m_g}\beta_{gj}^0 \leq n^{C_3}$, for some $C_3 > 0$ and $|\mathcal{A}_n| \leq n$. Let us define

$$e_{2n} = E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{HT}} - \boldsymbol{\beta}_g^0\right\|_2^2 - E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}} - \boldsymbol{\beta}_g^0\right\|_2^2.$$

Next using (129)-(136), we have, for all sufficiently large n, $e_{2n} \lesssim n^{\tilde{C}_1}\exp\Big(\tilde{C}_2\times n^{\tilde{C}_3}\Big)(1+o(1))$.
Hence, we are left to prove that, $e_{2n} = o(e_{1n})$, as $n \to \infty$. For this, we need an upper bound on e_{1n}. Towards this, first observe that,

$$\begin{aligned}
&E_{\boldsymbol{\beta}_g^0=0}\left[\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2 I\big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5\big\}\right]\\
&\geq E_{\boldsymbol{\beta}_g^0=0}\left[\left\|\widehat{\boldsymbol{\beta}}_g^{\mathrm{PM}}\right\|_2^2 I\big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5\big\} I\big\{\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2 \geq \frac{1}{n^2}\big\}\right]\\
&\geq \frac{1}{n^2}E_{\boldsymbol{\beta}_g^0=0}\left[(E(1-\kappa_g|\tau_n,\sigma^2,\mathcal{D}))^2 I\big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5 \cap \left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2 \geq \frac{1}{n^2}\big\}\right]\\
&\geq \frac{1}{n^2}\cdot\frac{1}{n^{(4a+2)}}E_{\boldsymbol{\beta}_g^0=0}\left[I\big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5 \cap \left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2 \geq \frac{1}{n^2}\big\} I\big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \geq \frac{1}{n^{(2a+1)}}\big\}\right]\\
&\geq \frac{1}{n^{(4a+4)}}P_{\boldsymbol{\beta}_g^0=0}(B_{1n}\cap B_{3n}\cap B_{4n})\\
&\geq \frac{1}{n^{(4a+4)}}[P_{\boldsymbol{\beta}_g^0=0}(B_{1n}) + P_{\boldsymbol{\beta}_g^0=0}(B_{3n}) + P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) - 2],
\end{aligned} \tag{136}$$

where B_{1n} is the same as defined earlier, $B_{3n} = \left\{\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2 \geq \frac{1}{n^2}\right\}$ and $B_{4n} = \big\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \geq \frac{1}{n^{(2a+1)}}\big\}$.
Here in the chain of inequalities, the second one holds due to the definition of the posterior mean along with the lower bound on the least square estimate. The fourth one follows due to Markov's inequality on the event B_{4n}. The last one follows from Boole's inequality.
Recall that, Proposition 1 implies, $\lim_{n\to\infty} P_{\boldsymbol{\beta}_g^0=0}(B_{1n}) = 1$. Hence, our next target is to show that, $\lim_{n\to\infty} P_{\boldsymbol{\beta}_g^0=0}(B_{3n}) = 1$ and $\lim_{n\to\infty} P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) = 1$. Towards that, note, under (C1), we have

$$\frac{W_{n,g}}{C_2} \leq \frac{n\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2}{\sigma^2} \leq \frac{W_{n,g}}{C_1}.$$

Hence, for $W_{n,g} \sim \chi^2_{m_g}$,

$$\begin{aligned}
P_{\boldsymbol{\beta}_g^0=0}(B_{3n}) &= P_{\boldsymbol{\beta}_g^0=0}\left(\frac{n\left\|\widehat{\boldsymbol{\beta}}_g\right\|_2^2}{\sigma^2} > \frac{1}{n\sigma^2}\right) \\
&\geq P_{\boldsymbol{\beta}_g^0=0}\left(\frac{W_{n,g}}{C_2} > \frac{1}{n\sigma^2}\right) \\
&\to 1 \text{ as } n \to \infty. \qquad (137)
\end{aligned}$$

Hence, it is only left to show that, $\lim_{n\to\infty} P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) = 1$. For this, we need to obtain a lower bound on $E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D})$. We consider two cases $a \in [\frac{1}{2}, 1)$ and $a \geq 1$, separately.

Case 1: $a \in [\frac{1}{2}, 1)$**.** Using the definition of the posterior distribution of κ_g given $\tau, \sigma^2, \mathcal{D}$, we have

$$\begin{aligned}
E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) &= \frac{\int_0^1 \kappa_g^{a+\frac{m_g}{2}-1}(1-\kappa_g)^{-a} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right)\exp\left\{-\kappa_g \cdot \frac{W_{n,g}}{2}\right\}d\kappa_g}{\int_0^1 \kappa_g^{a+\frac{m_g}{2}-1}(1-\kappa_g)^{-a-1} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right)\exp\left\{-\kappa_g \cdot \frac{W_{n,g}}{2}\right\}d\kappa_g} \\
&\geq \exp\left(-\frac{W_{n,g}}{2}\right)\frac{\int_0^1 \kappa_g^{a+\frac{m_g}{2}-1}(1-\kappa_g)^{-a} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right)d\kappa_g}{\int_0^1 \kappa_g^{a+\frac{m_g}{2}-1}(1-\kappa_g)^{-a-1} L\left(\frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)\right)d\kappa_g}, \qquad (138)
\end{aligned}$$

where the inequality holds using the monotonicity of the function $\kappa \mapsto \exp(-\kappa x^2/2), 0 < \kappa < 1$. Next, using the transformation $t = \frac{1}{\tau_n^2}(\frac{1}{\kappa_g}-1)$ in the integrals above, we obtain

$$\begin{aligned}
E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) &\geq \exp\left(-\frac{W_{n,g}}{2}\right)\tau_n^2 \frac{\int_0^\infty (1+t\tau_n^2)^{-\frac{m_g}{2}-1}t^{-a}L(t)dt}{\int_0^\infty (1+t\tau_n^2)^{-\frac{m_g}{2}}t^{-a-1}L(t)dt} \\
&= K\exp\left(-\frac{W_{n,g}}{2}\right)\tau_n^2 \int_0^\infty (1+t\tau_n^2)^{-\frac{m_g}{2}-1}t^{-a}L(t)dt(1+o(1)) \\
&\geq K\exp\left(-\frac{W_{n,g}}{2}\right)\tau_n^2 \int_{t_0}^{\frac{t_0}{\tau_n^2}} (1+t\tau_n^2)^{-\frac{m_g}{2}-1}t^{-a}L(t)dt(1+o(1)) \\
&\geq K\exp\left(-\frac{W_{n,g}}{2}\right)\tau_n^2(1+t_0)^{-\frac{m_g}{2}-1} \int_{t_0}^{\frac{t_0}{\tau_n^2}} t^{-a}L(t)dt(1+o(1)) \\
&\geq \frac{K}{(1-a)}\exp\left(-\frac{W_{n,g}}{2}\right)\tau_n^2(1+t_0)^{-\frac{s}{2}-1}L(\frac{t_0}{\tau_n^2})(\frac{t_0}{\tau_n^2})^{(1-a)}(1+o(1)) \\
&\geq K_1\exp\left(-\frac{W_{n,g}}{2}\right)\tau_n^{2a}(1+o(1)), \qquad (139)
\end{aligned}$$

where K_1 is a constant depending on t_0, a, s. Here, the equality in the second line holds due to using the Dominated Convergence Theorem on the denominator. The inequality in the fourth line holds because of the monotonicity of the function $t \mapsto (1+t)^{-\frac{m_g}{2}-1}$ in the interval $[t_0, \frac{t_0}{\tau_n^2}]$. (A2) followed by (5) of Lemma A.1 ensures the inequality in the fifth line. The final one holds due to Assumption 1 on $L(\cdot)$.

Case 2: $a \geq 1$**.** In this case, too, following arguments same as those of (138) and (139), we have

$$\begin{aligned} E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) &\geq K \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2 \int_0^{\infty} (1+t\tau_n^2)^{-\frac{m_g}{2}-1} t^{-a} L(t) dt (1+o(1)) \\ &\geq K \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2 \int_1^{\frac{1}{\tau_n^2}} (1+t\tau_n^2)^{-\frac{m_g}{2}-1} t^{-a} L(t) dt (1+o(1)) \\ &\geq 2^{-\frac{m_g}{2}-1} K \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2 \int_1^{\frac{1}{\tau_n^2}} t^{-a} L(t) dt (1+o(1)) \\ &\geq 2^{-\frac{s}{2}-1} K c_0 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2 \int_1^{\frac{1}{\tau_n^2}} t^{-a} dt (1+o(1)). \end{aligned} \tag{140}$$

Since, for $a=1$, $\int_1^{\frac{1}{\tau_n^2}} t^{-a} dt = \log\big(\frac{1}{\tau_n^2}\big)$ and for $a>1$, $\int_1^{\frac{1}{\tau_n^2}} t^{-a} dt = \frac{1}{(a-1)}[1-\tau_n^{2(a-1)}]$, using (140), we obtain

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \geq \begin{cases} K_2 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2 \log\big(\frac{1}{\tau_n^2}\big)(1+o(1)), a=1, \\ K_3 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2(1+o(1)), a>1, \end{cases} \tag{141}$$

where K_2 and K_3 are constants depending on t_0, a, s. Combining (139) and (141), for $a \geq \frac{1}{2}$,

$$E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \geq \begin{cases} K_1 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^{2a}(1+o(1)), 0<a<1, \\ K_2 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2 \log\big(\frac{1}{\tau_n^2}\big)(1+o(1)), a=1, \\ K_3 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^2(1+o(1)), a>1. \end{cases} \tag{142}$$

Hence, we need to show $\lim\limits_{n\to\infty} P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) = 1$, for $a \in [\frac{1}{2}, 1)$ and $a \geq 1$. When $\frac{1}{2} \leq a < 1$,

$$\begin{aligned} P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) &= P_{\boldsymbol{\beta}_g^0=0}\bigg(E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \geq \frac{1}{n^{(2a+1)}}\bigg) \\ &\geq P_{\boldsymbol{\beta}_g^0=0}\bigg(K_1 \exp\Big(-\frac{W_{n,g}}{2}\Big)\tau_n^{2a}(1+o(1)) \geq \frac{1}{n^{(2a+1)}}\bigg) \\ &= P_{\boldsymbol{\beta}_g^0=0}\bigg(\log K_1 - \frac{W_{n,g}}{2} - 2a\log\Big(\frac{1}{\tau_n}\Big) + o(1) \geq -(2a+1)\log n\bigg) \\ &= P_{\boldsymbol{\beta}_g^0=0}\bigg(\frac{W_{n,g}}{2} \leq (2a+1)\log n - 2a\log\Big(\frac{1}{\tau_n}\Big) + o(1) + \log K_1\bigg). \end{aligned} \tag{143}$$

Observe that, under (A4), for sufficiently large n,

$$(2a+1)\log n - 2a\log\Big(\frac{1}{\tau_n}\Big) \geq (2a+1)\log n - 2a\log n = \log n. \tag{144}$$

Therefore, combining (143) and (144), for $a \in (0,1)$,

$$P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) \geq P_{\boldsymbol{\beta}_g^0=0}\Big(W_{n,g} \leq 2\log n + 2\log K_1 + o(1)\Big) \to 1, \text{ as } n \to \infty. \tag{145}$$

For $a \geq 1$, using the lower bound obtained in (142) and following similar arguments of (143) and (144), again we can show that,

$$P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) \to 1, \text{ as } n \to \infty. \tag{146}$$

Combining (145) and (146), for $a > 0$, we obtain

$$P_{\boldsymbol{\beta}_g^0=0}(B_{4n}) \to 1, \text{ as } n \to \infty. \tag{147}$$

Combining (136), (137) and (147), we have, for sufficiently large n,

$$e_{1n} = -\sum_{g \notin \mathcal{A}_n} E_{\boldsymbol{\beta}_g^0=0}\left[\left\|\widehat{\boldsymbol{\beta}}_g^{\text{PM}}\right\|_2^2 I\{E(1-\kappa_g \mid \tau_n, \sigma^2, \mathcal{D}) \leq 0.5\}\right] \lesssim -n^{-(4a+3)}(1+o(1)).$$

This ensures $e_{2n} = o(e_{1n})$, as $n \to \infty$. As a result, (124) is established. The proof of Theorem 7 is completed using (123) and (124). □

We end this subsection with the following remark.

Remark 1. *Theorem 7 establishes that the HT estimator is more accurate than the traditional Bayes estimator if the conditions (A2)-(A4) and (C1) hold. Note that (A4) implies that the level of sparsity is known and hence the global parameter τ_n is used as a tuning one. Hence, a follow-up question is whether the same result holds when the level of sparsity is unknown, when either empirical Bayes or a full Bayes treatment on τ is employed. The next two corollaries provide a positive answer to this question. Proofs of these follow from arguments used in Theorem 7 and Theorems 3 and 6 of the main paper, and hence are skipped.*

Corollary 1. *Consider the hierarchical framework of (5) where $\pi(\lambda_g^2)$ is as in (6) with $a \geq \frac{1}{2}$ and the half-thresholding (HT) rule (14) based on these. Let $\mathcal{A}_n = \{g : \boldsymbol{\beta}_g^0 \neq \mathbf{0}\}$ and $\widehat{\mathcal{A}}_n^{(EB)} = \{g : \widehat{\boldsymbol{\beta}}_{g,EB}^{HT} \neq \mathbf{0}\}$ denote respectively the set of truly active groups, and the set of groups declared active by the half-thresholding rule (14). Let $\mathbf{Q}_{n,g} = \mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g/n$, for $g = 1, \cdots, G$. Along with (A2),(A3) and (C1), we further assume that $|\mathcal{A}_n| = G_{\mathcal{A}_n}$ is unknown and $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some unknown $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$, with any active group $g \in \mathcal{A}_n$ satisfies $\max_j |\beta_{gj}^0| \leq n^{C_3}$, for some $0 < C_3 < \infty$. Then also our proposed empirical Bayes HT estimator of $\boldsymbol{\beta}_g$ is more accurate than that of the posterior mean with respect to squared error loss, i.e., for all sufficiently large n,*

$$\sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_{g,EB}^{HT} - \boldsymbol{\beta}_g^0\right\|_2^2 < \sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{PM} - \boldsymbol{\beta}_g^0\right\|_2^2.$$

Corollary 2. *Consider the hierarchical framework of (5) where $\pi(\lambda_g^2)$ is as in (6) with $a \geq \frac{1}{2}$ and the half-thresholding (HT) rule (16) based on these and τ is assumed to have a non-degenerate prior distribution satisfying (D1). Let $\mathcal{A}_n = \{g : \boldsymbol{\beta}_g^0 \neq \mathbf{0}\}$ and $\widehat{\mathcal{A}}_n^{(FB)} = \{g : \widehat{\boldsymbol{\beta}}_{g,FB}^{HT} \neq \mathbf{0}\}$ denote respectively the set of truly active groups, and the set of groups declared active by the half-thresholding rule (16). Let $\mathbf{Q}_{n,g} = \mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g/n$, for $g = 1, \cdots, G$. Along with (A2),(A3) and (C1), we further assume that $|\mathcal{A}_n| = G_{\mathcal{A}_n}$ is unknown and $G_n^{\epsilon_1} \lesssim G_{\mathcal{A}_n} \lesssim G_n^{\epsilon_2}$ for some unknown $0 < \epsilon_1 < \epsilon_2 < \frac{1}{2}$, with any active group $g \in \mathcal{A}_n$, satisfies $\max_j |\beta_{gj}^0| \leq n^{C_3}$, for some $0 < C_3 < \infty$. Then also our proposed full Bayes HT estimator of $\boldsymbol{\beta}_g$ is more accurate than that of the posterior mean with respect to squared error loss, i.e., for all sufficiently large n,*

$$\sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_{g,FB}^{HT} - \boldsymbol{\beta}_g^0\right\|_2^2 < \sum_{g=1}^{G_n} E_{\boldsymbol{\beta}_g^0}\left\|\widehat{\boldsymbol{\beta}}_g^{PM} - \boldsymbol{\beta}_g^0\right\|_2^2.$$

C Gibbs Sampling

Within the hierarchical form (5) of the main document, and using the prior distributions on τ and σ^2 stated before, the Gibbs samples are drawn from the full conditional distributions as follows:

(1) Sampling from the Posterior Distribution of $\boldsymbol{\beta}_g$:

Since, the full posterior distribution of $\boldsymbol{\beta}$ given $(\boldsymbol{\lambda}^2, \sigma^2, \tau^2, \mathcal{D})$ is

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\lambda}^2, \sigma^2, \tau^2, \mathcal{D}) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \sum_{g=1}^{G}\frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{\lambda_g^2\tau^2}\right)\right]$$

we obtain for $g = 1, \ldots, G$,

$$\pi(\boldsymbol{\beta}_g \mid \boldsymbol{\beta}_{-g}, \boldsymbol{\lambda}^2, \sigma^2, \tau^2, \mathcal{D}) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g - 2\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{y} + \frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{\lambda_g^2\tau^2} + \sum_{g'(\neq g)=1}^{G}\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_{-g}\boldsymbol{\beta}_{-g}\right)\right],$$

This is equivalent to saying for $g = 1, \ldots, G$,

$$\boldsymbol{\beta}_g \mid (\boldsymbol{\beta}_{-g}, \boldsymbol{\lambda}^2, \sigma^2, \tau^2, \mathcal{D}) \sim \mathcal{N}_{m_g}(\boldsymbol{\mu}_g, \sigma^2\boldsymbol{\Sigma}_g),$$

with $\boldsymbol{\mu}_g = (1 + \frac{1}{\lambda_g^2\tau^2})^{-1}(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1}(\mathbf{X}_g^{\mathrm{T}}\mathbf{y} - \sum_{g'(\neq g)=1}^{G}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_{-g}\boldsymbol{\beta}_{-g})$ and
$\boldsymbol{\Sigma}_g = (1 + \frac{1}{\lambda_g^2\tau^2})^{-1}(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1} = (1 - \kappa_g)(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1}$.
With the additional assumption on the block-orthogonality of the design matrix $\mathbf{X}$, we have for $g = 1, \ldots, G$,

$$\boldsymbol{\beta}_g \mid (\boldsymbol{\lambda}^2, \sigma^2, \tau^2, \mathcal{D}) \overset{ind}{\sim} \mathcal{N}_{m_g}(\boldsymbol{\mu}_g, \sigma^2\boldsymbol{\Sigma}_g),$$

with $\boldsymbol{\mu}_g = (1 - \kappa_g)\widehat{\boldsymbol{\beta}}_g$ and $\boldsymbol{\Sigma}_g = (1 - \kappa_g)(\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g)^{-1}$.

(2) Sampling from the Posterior Distribution of σ^2:

The full posterior distribution of σ^2 conditioned on $(\boldsymbol{\beta}, \boldsymbol{\lambda}^2, \tau^2, \mathcal{D})$ is given by

$$\pi(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\lambda}^2, \tau^2, \mathcal{D}) \propto (\sigma^2)^{-(\frac{n}{2} + \sum_{g=1}^{G}\frac{m_g}{2} + 1)} \times \exp\left[-\frac{1}{\sigma^2}\left\{\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2} + \sum_{g=1}^{G}\frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{2\lambda_g^2\tau^2}\right\}\right].$$

Hence,

$$\sigma^2 \mid (\boldsymbol{\beta}, \boldsymbol{\lambda}^2, \tau^2, \mathcal{D}) \sim \text{Inverse Gamma}\left(\frac{n}{2} + \sum_{g=1}^{G}\frac{m_g}{2}, \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2} + \sum_{g=1}^{G}\frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{2\lambda_g^2\tau^2}\right).$$

(3) Sampling from the Posterior Distribution of λ_g^2:

Observe that, for each $g = 1, 2, \cdots, G$,

$$\pi(\lambda_g^2 \mid \boldsymbol{\beta}_g, \sigma^2, \tau^2, \mathcal{D}) \propto (\lambda_g^2)^{-\frac{(m_g+1)}{2}}(1 + \lambda_g^2)^{-1} \times \exp\left[-\frac{1}{\lambda_g^2} \cdot \frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{2\tau^2\sigma^2}\right]$$

Using the Slice-sampling approach of Damlen et al. (1999), posterior sampling is done in two steps:

1. Given λ_g^2, sample u_g from the Uniform distribution supported over the interval $(0, 1 + \lambda_g^2)$.
2. For given $(\boldsymbol{\beta}_g, \sigma^2, \tau^2, \mathcal{D})$, sample λ_g^2 from an inverse-gamma distribution with parameters $\frac{(m_g-1)}{2}$ and $\frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{2\tau^2\sigma^2}$, truncated over the interval $(0, \frac{1-u_g}{u_g})$.

(4) Sampling from the Posterior Distribution of τ^2:

$$\pi(\tau^2|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}^2, \mathcal{D}) \propto \frac{1}{1+\tau^2} \times (\tau^2)^{-\frac{p}{2}} \exp\left(-\frac{1}{\tau^2}\sum_{g=1}^{G}\frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{2\lambda_g^2\sigma^2}\right).$$

Again, using the Slice-sampling approach of Damlen et al. (1999), samples are drawn from the above posterior distribution of τ^2 as follows:

1. Given τ^2, sample u from the Uniform distribution supported over the interval $(0, 1+\tau^2)$.

2. Given $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}^2, \mathcal{D})$, sample τ^2 from an inverse-gamma distribution with parameters $\frac{(p-2)}{2}$ and $\sum_{g=1}^{G}\frac{\boldsymbol{\beta}_g^{\mathrm{T}}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\boldsymbol{\beta}_g}{2\lambda_g^2\sigma^2}$, truncated to have zero probability outside the interval $(0, \frac{1-u}{u})$.

D Additional simulations

D.1 Simulation setup

- Example 5. For this example, we consider the case when the group sizes differ. Let us consider the scenario when $n = 200$ and $p = 50$ predictors are grouped in 16 groups with group sizes 4,3,3,2,2,2,2,2,2,4,4,4,4,2,5 and 5 respectively. Let $\boldsymbol{\beta} = ((0.1, 0.2, 0.3, 0.4), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, (0, 0.4, 0, 0), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$ for weak signal strength and $\boldsymbol{\beta} = ((1, 2, 3, 4), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, (1, 1.1, 1.2, 1.3), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$ for strong signal. In both of the cases, the first two $\mathbf{0}$ denote null vectors of length 3, the remaining are of length 2, except the last two, which are of length 5. In this case, the predictors are also generated the same way as in Example 1, except for necessary dimension changes.

- Example 6. Next consider a situation when the sample size is large. Suppose the sample size $n = 500$ and $p = 100$ covariates are grouped into 25 groups containing four covariates each. We assume that only one group is active and the group coefficients are, $\boldsymbol{\beta} = ((0.1, 0.2, 0.3, 0.4), \mathbf{0}, \cdots, \mathbf{0}, \mathbf{0})$ when the signal strength is weak and $\boldsymbol{\beta} = ((1.1, 1.2, 1.3, 1.4), \mathbf{0}, \cdots, \mathbf{0}, \mathbf{0})$ when the signal strength is strong. Here $\mathbf{0}$ is a null vector of length 4. The data generated scheme is similar to Example 1 except for necessary dimension changes.

- Example 7. Now we revisit example 3, where the design matrix now becomes $\mathbf{Z}$ instead of $\mathbf{X}$, whose columns are generated using (25) of the main file. Here, $n = 200$ and $p = 40$ covariates are grouped into 10 groups containing 4 covariates each. We assume that only the first group is active and the coefficients are $\boldsymbol{\beta} = ((1.1, 1.2, 1.3, 1.4), \mathbf{0}, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{0})$ where $\mathbf{0}$ is a null vector of length 4. The data generated scheme is similar to Example 3 except for necessary dimension changes.

- Example 8. Consider an example of a large p small n problem with $n = 50$ and $p = 100$. 100 predictors are grouped into 20 groups of 5 covariates each. For weak signal strength, the group coefficients are assumed to be, $\boldsymbol{\beta} = (\mathbf{0}, \cdots, \mathbf{0}, \mathbf{0.4}, \mathbf{0.5}, (0.65, 0.60, 0.55, 0.50, 0.45), \mathbf{0}, \cdots, \mathbf{0})$. On the other hand, when the signal strength is strong, we assume $\boldsymbol{\beta} = (\mathbf{1}, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{1.2}, \mathbf{1.4}, \mathbf{0}, \cdots, \mathbf{0})$. In each case, $\mathbf{0}$ denotes a null vector of length 5.

- Example 9. This is another example for large p small n problem with $n = 50$ and $p = 100$. Unlike previous situations, for the same combination of (n, p), we consider two situations where the level of sparsity in the second situation is twice the former one. Here 100 predictors are grouped into 25 groups of 4 covariates each. First, we assume 4 groups to be active out of 25 groups. Let $\boldsymbol{\beta} = (\mathbf{0}, \cdots, \mathbf{0}, \mathbf{1.8}, \mathbf{0.5}, (0.65, 0.60, 0.55, 0.50), \mathbf{0}, \cdots, \mathbf{0}, \mathbf{2.5})$ where the $\mathbf{0}$ denote null vector of length 4. Next, we assume eight groups to be active out of 25 groups. Let $\boldsymbol{\beta} = (\mathbf{1.8}, \mathbf{1.5}, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{1.8}, \mathbf{1.5}, (0.65, 0.60, 0.55, 0.50), \mathbf{0}, \cdots, \mathbf{0}, \mathbf{2.5}, (0.4, 0.45, 0.50, 0.55), \mathbf{2.5})$.

- From Theorem 1 of the main paper, the choice of $\tau_n^2 = (\frac{G_{\mathcal{A}_n}}{G_n})^{2+\delta}$, for any $\delta > 0$ ensures model selection consistency. A question that arises here is how to choose that δ to optimize the performance. The same goes for the choice of (c_1, c_2), also, for the empirical Bayes estimates of τ, defined in the main paper. To provide answers to these questions, we have done some simulation studies corresponding to Examples 1 and 2 of the main document with choice of τ_n obtained from $\tau_n^2 = (\frac{G_{\mathcal{A}_n}}{G_n})^{2+\delta}$, with $\delta = 0.1, 0.2, \cdots, 0.5$. The results are stated in Tables 10 and 11, respectively. On the other hand, for the empirical Bayes estimate, in order to obtain a proper choice of (c_1, c_2), we have repeated Examples 1 and 2 of the main document by keeping $c_1 = 2$ and varying c_2 over $\{1, \ldots, 5\}$ and keeping $c_2 = 1$ and varying c_1 over $\{2, 3, 4, 5\}$. The corresponding results are summarized in Tables 12-15.

- In order to study the importance of the choice of τ based on the proportion of active groups, we revisited Examples 8 and 9, where along with the choices of τ previously mentioned, we consider other choices of τ which depend on *only* G_n. Here, we use five choices of τ_n, $\tau_n = \tau_{1n} = \left(\frac{G_{\mathcal{A}_n}}{G_n}\right)^{1.05}$, $\tau_n = \tau_{2n} = \frac{1}{G_n^2}$, and $\tau_n = \tau_{3n} = G_n^{-1.4}$, $\tau_n = \tau_{4n} = G_n^{-1.6}$, and $\tau_n = \tau_{5n} = \widehat{\tau}^{\text{EB(Mod)}}$, where $\widehat{\tau}^{\text{EB(Mod)}}$ is the modified version of the empirical Bayes estimate of τ defined in (23) of the main document. We have also considered the full Bayes procedure based on half-Cauchy prior on a truncated range. The results are presented in Tables 16 and 17.

D.2 Simulation Results

D.3 Interpretation of simulation results

For each of the above examples, we computed the probability of misclassification (MP), the false positive rate (FPR) and the true positive rate (TPR) for each of the methods mentioned in the main document. Our main findings are the following.

- Example 5 is different from all the remaining examples, as the group size is different in this case. Although the group size is not used in the prior distribution of the group regression coefficients in any of these methods, our half-thresholding rule successfully captures the truth and hence produces better results than those of GSD-SSS, BGL-SS, group SCAD, and group MCP.

- Example 6 shows that when the signal strength is weak, other competing procedures can produce results similar to ours, if the sample size is large ($n = 500$). It may be recalled that for small or moderate sample sizes the situation is quite different, as demonstrated in the main document.

- Examples 5-7 clarify that when signal strength is strong, regardless of sample size n, the performances of BGL-SS and GSD-SSS are comparable with those of our decision rule.

- In example 7, we consider a case where the design matrix is block-diagonal. Since the data generation scheme is similar to that of Example 3 of the main document, from the previous result, we expected that our method would provide better results than those of Yang and Narisetty (2020) and Xu and Ghosh (2015) when the signal strength is weak and would produce comparable results when the signal strength is strong. Table 7 confirms this.

- Examples 8 and 9 show that the rule (10) of the main document will also work even if $p > n$. In this case, too, MP and TPR corresponding to our decision rule are much better than those of GSD-SSS and better than BGL-SS for weak signal strength, and produce comparable results for strong signal strength.

- Example 9 deals with situations when there is a mixture of weak and strong signal strengths. In these cases also, our proposed method outperforms GSD-SSS, BGL-SS, and yields better results than those of group LASSO, group MCP, and group SCAD.

Table 5: Mean True/False Positive Rate based on 100 replications(Example 5)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH	0.0092	0.0034	0.95	0.0051	0.0029	0.98
Normal GH	0.0118	0.0041	0.95	0.0082	0.0036	0.96
GSD-SSS	0.0265	0.0046	0.82	0.0188	0.0029	0.87
Modified GH(EB1)	0.0120	0.0038	0.93	0.0117	0.0034	0.93
Modified GH(EB2)	0.0118	0.0035	0.92	0.0115	0.0031	0.93
Normal GH(EB)	0.0127	0.0045	0.93	0.0122	0.0039	0.93
Modified GH(FBtruncated)	0.0119	0.0036	0.93	0.0129	0.0034	0.92
Modified GH(FBfull)	0.0123	0.0041	0.93	0.0159	0.0039	0.90
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0096	0.0038	0.95	0.0063	0.0029	0.97
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0081	0.0035	0.96	0.0065	0.0031	0.97
Normal GH(FB)	0.0125	0.0043	0.93	0.0144	0.0036	0.91
BGL-SS	0.0133	0.0038	0.92	0.0104	0.0034	0.94
Group LASSO	0.1675	0.1812	0.94	0.1574	0.1691	0.94
Group SCAD	0.0313	0.0215	0.91	0.0228	0.0175	0.94
Group MCP	0.0188	0.0143	0.95	0.0135	0.0112	0.97
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH	0.00113	0.0013	1.00	0.00105	0.0012	1.00
Normal GH	0.00131	0.0015	1.00	0.00131	0.0015	1.00
GSD-SSS	0.00113	0.0013	1.00	0.00113	0.0013	1.00
Modified GH(EB1)	0.00123	0.0014	1.00	0.00131	0.0015	1.00
Modified GH(EB2)	0.00113	0.0013	1.00	0.00105	0.0012	1.00
Normal GH(EB)	0.00123	0.0014	1.00	0.00140	0.0016	1.00
Modified GH(FBtruncated)	0.00166	0.0019	1.00	0.00123	0.0014	1.00
Modified GH(FBfull)	0.00192	0.0022	1.00	0.00157	0.0018	1.00
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.00131	0.0015	1.00	0.00123	0.0014	1.00
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.00123	0.0014	1.00	0.00123	0.0014	1.00
Normal GH(FB)	0.00158	0.0018	1.00	0.00114	0.0013	1.00
BGL-SS	0.00625	0.0072	1.00	0.00313	0.0036	1.00
Group LASSO	0.14717	0.1682	1.00	0.13475	0.1540	1.00
Group SCAD	0.05440	0.0622	1.00	0.02761	0.0315	1.00
Group MCP	0.001841	0.0021	1.00	0.00161	0.0018	1.00

- The results in Tables 10 and 11 show that with an increase in the value of δ, the misclassification probability (MP) and the False Positive Rate (FPR) increase and the True Positive Rate (TPR) decreases, irrespective of the signal strength and magnitude of correlation among the covariates within a group. However, this change is more significant when the signal strength is small. Hence, based on simulation studies, we propose to choose δ to be small and positive, that is, we propose the choice $\delta = 0.1$.
- Tables 12-15 show that, keeping $c_1 = 2$ and varying c_2 over $\{1, \ldots, 5\}$ results in an increase in MP and FPR and a decrease in TPR for both small and large signals. The same result is also obtained with $c_2 = 1$ and $c_1 \in \{2, 3, 4, 5\}$. Based on these results, we recommend to choose $(c_1, c_2) = (2, 1)$.

Table 6: Mean True/False Positive Rate based on 100 replications(Example 6)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH	0.0029	0.0031	1.00	0.0023	0.0024	1.00
Normal GH	0.0029	0.0031	1.00	0.0023	0.0024	1.00
GSD-SSS	0.0029	0.0031	1.00	0.0023	0.0024	1.00
Modified GH(EB1)	0.0035	0.0036	1.00	0.0029	0.0031	1.00
Modified GH(EB2)	0.0033	0.0034	1.00	0.0027	0.0028	1.00
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0033	0.0034	1.00	0.0023	0.0024	1.00
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0033	0.0034	1.00	0.0024	0.0025	1.00
Normal GH(EB)	0.0035	0.0036	1.00	0.0029	0.0031	1.00
Modified GH(FBtruncated)	0.0029	0.0031	1.00	0.0024	0.0025	1.00
Modified GH(FBfull)	0.0034	0.0035	1.00	0.0027	0.0028	1.00
Normal GH(FB)	0.0031	0.0032	1.00	0.0024	0.0025	1.00
BGL-SS	0.0032	0.0033	1.00	0.0024	0.0025	1.00
Group LASSO	0.065	0.0681	1.00	0.0522	0.0544	1.00
Group SCAD	0.0028	0.0029	1.00	0.0023	0.0024	1.00
Group MCP	0.0029	0.0031	1.00	0.0027	0.0028	1.00
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH	0.0027	0.0028	1.00	0.0023	0.0024	1.00
Normal GH	0.0027	0.0028	1.00	0.0023	0.0024	1.00
GSD-SSS	0.0027	0.0028	1.00	0.0023	0.0024	1.00
Modified GH(EB1)	0.0025	0.0026	1.00	0.0029	0.0031	1.00
Modified GH(EB2)	0.0025	0.0026	1.00	0.0029	0.0031	1.00
Normal GH(EB)	0.0025	0.0026	1.00	0.0029	0.0031	1.00
Modified GH(FBtruncated)	0.0029	0.0031	1.00	0.0028	0.0029	1.00
Modified GH(FBfull)	0.0032	0.0033	1.00	0.0031	0.0032	1.00
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0029	0.0031	1.00	0.0023	0.0024	1.00
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0028	0.0029	1.00	0.0023	0.0024	1.00
Normal GH(FB)	0.0031	0.0032	1.00	0.0028	0.0029	1.00
BGL-SS	0.0024	0.0025	1.00	0.0020	0.0021	1.00
Group LASSO	0.0570	0.0594	1.00	0.0522	0.0544	1.00
Group SCAD	0.0028	0.0029	1.00	0.0022	0.0023	1.00
Group MCP	0.0026	0.0027	1.00	0.0022	0.0023	1.00

- Tables 16 and 17 demonstrate that when the proportion of active groups is known, MP and FPR are at their least when τ_n is used as a tuning parameter based on this proportion. This implies that when the knowledge on the proportion is available, it is better to use that knowledge. When the level of sparsity is unknown, the tables demonstrate the following. One gets better results in these cases when one either uses the empirical Bayes method (that estimates τ using quantities which are influenced by the level of sparsity) or the full Bayes method using a prior on τ. The ad-hoc choices of τ_n using only G_n are clear under-performers in such cases.

Table 7: Mean True/False Positive Rate based on 100 replications(Example 7)

	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH	0.00351	0.0039	1.00	0.00252	0.0028	1.00
Normal GH	0.00369	0.0041	1.00	0.00261	0.0029	1.00
GSD-SSS	0.00369	0.0041	1.00	0.00279	0.0031	1.00
Modified GH(EB1)	0.00396	0.0044	1.00	0.00261	0.0029	1.00
Modified GH(EB2)	0.00369	0.0041	1.00	0.00261	0.0029	1.00
Normal GH(EB)	0.00405	0.0045	1.00	0.00297	0.0033	1.00
Modified GH(FBtruncated)	0.00441	0.0049	1.00	0.00306	0.0034	1.00
Modified GH(FBfull)	0.00468	0.0052	1.00	0.00342	0.0038	1.00
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.00387	0.0043	1.00	0.00288	0.0032	1.00
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.00378	0.0042	1.00	0.00279	0.0031	1.00
Normal GH(FB)	0.00468	0.0052	1.00	0.00279	0.0031	1.00
BGL-SS	0.00401	0.0044	1.00	0.00203	0.0022	1.00
Group LASSO	0.19981	0.2221	1.00	0.19981	0.2221	1.00
Group SCAD	0.00387	0.0043	1.00	0.00351	0.0039	1.00
Group MCP	0.00351	0.0039	1.00	0.00243	0.0027	1.00

References

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation.* Cambridge University Press, 1987. 2

Paul Damlen, John Wakefield, and Stephen Walker. Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344, 1999. 33, 34

Wolfgang Gabcke. Derivation of the riemann-siegel formula with explicit estimates of its remainders. *http://dx.doi.org/10.53846/goediss-5113*, 2015. 9

Prasenjit Ghosh and Arijit Chakrabarti. Asymptotic optimality of one-group shrinkage priors in sparse high-dimensional problems. *Bayesian Analysis*, 12(4):1133–1161, 2017. 11

Prasenjit Ghosh, Xueying Tang, Malay Ghosh, and Arijit Chakrabarti. Asymptotic properties of bayes risk of a general class of shrinkage priors in multiple hypothesis testing under sparsity. *Bayesian Analysis*, 11(3): 753–796, 2016. 2, 3

Sayantan Paul and Arijit Chakrabarti. Posterior contraction rate and asymptotic bayes optimality for one group global-local shrinkage priors in sparse normal means problem. *Ann Inst Stat Math*, 2025. 19

Xiaofan Xu and Malay Ghosh. Bayesian variable selection and estimation for group lasso. *Bayesian Analysis*, 10(4):909–936, 2015. 35

Xinming Yang and Naveen N Narisetty. Consistent group selection with bayesian high dimensional modeling. *Bayesian Analysis*, 15(3):909–935, 2020. 35

Table 8: Mean True/False Positive Rate based on 100 replications(Example 8)

Small group coefficients			
Prior	MP	FPR	TPR
Modified GH	0.0137	0.0140	0.988
Normal GH	0.0142	0.0139	0.984
GSD-SSS	0.0445	0.0138	0.782
Modified GH(EB1)	0.0149	0.0141	0.980
Modified GH(EB2)	0.0144	0.0141	0.984
Normal GH(EB)	0.0158	0.0141	0.974
Modified GH(FBtruncated)	0.0131	0.0139	0.991
Modified GH(FBfull)	0.0138	0.0145	0.990
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0146	0.0143	0.984
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0145	0.0142	0.986
Normal GH(FB)	0.0137	0.0140	0.988
BGL-SS	0.0134	0.0144	0.992
Group LASSO	0.1866	0.2151	0.975
Group SCAD	0.0151	0.0176	1.000
Group MCP	0.0151	0.0176	1.000
Large group coefficients			
Prior	MP	FPR	TPR
Modified GH	0.0111	0.0131	1.00
Normal GH	0.0127	0.0149	1.00
GSD-SSS	0.0109	0.0129	1.00
Modified GH(EB1)	0.0121	0.0142	1.00
Modified GH(EB2)	0.0115	0.0135	1.00
Normal GH(EB)	0.0129	0.0152	1.00
Modified GH(FBtruncated)	0.0121	0.0142	1.00
Modified GH(FBfull)	0.0126	0.0148	1.00
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0112	0.0132	1.00
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0113	0.0133	1.00
Normal GH(FB)	0.0118	0.0139	1.00
BGL-SS	0.0121	0.0142	1.00
Group LASSO	0.1133	0.1331	1.00
Group SCAD	0.0301	0.0353	1.00
Group MCP	0.0112	0.0133	1.00

Table 9: Mean True/False Positive Rate based on 100 replications(Example 9)

Number of active groups=4			
Prior	MP	FPR	TPR
Modified GH	0.0175	0.0204	0.998
Normal GH	0.0169	0.0194	0.996
GSD-SSS	0.0326	0.0183	0.892
Modified GH(EB1)	0.0205	0.0232	0.994
Modified GH(EB2)	0.0164	0.0191	0.995
Normal GH(EB)	0.0203	0.0231	0.994
Modified GH(FBtruncated)	0.0158	0.0184	0.998
Modified GH(FBfull)	0.0167	0.0191	0.996
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0188	0.0212	0.994
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0180	0.0209	0.997
Normal GH(FB)	0.0139	0.0181	0.995
BGL-SS	0.0179	0.0199	0.992
Group LASSO	0.1506	0.1764	0.985
Group SCAD	0.0199	0.0214	0.988
Group MCP	0.0288	0.0272	0.963
Number of active groups=8			
Prior	MP	FPR	TPR
Modified GH	0.0102	0.0121	0.994
Normal GH	0.0109	0.0119	0.991
GSD-SSS	0.0558	0.0118	0.851
Modified GH(EB1)	0.0118	0.0121	0.989
Modified GH(EB2)	0.0116	0.0119	0.989
Normal GH(EB)	0.0131	0.0122	0.985
Modified GH(FBtruncated)	0.0115	0.0118	0.989
Modified GH(FBfull)	0.0128	0.0122	0.986
Modified TPBN($\alpha_1 = 0.5, \alpha_2 = 1$)	0.0113	0.0124	0.991
Modified TPBN($\alpha_1 = 1, \alpha_2 = 0.5$)	0.0101	0.0120	0.994
Normal GH(FB)	0.0118	0.0117	0.988
BGL-SS	0.0127	0.0121	0.986
Group LASSO	0.1406	0.1978	0.981
Group SCAD	0.0200	0.0177	0.975
Group MCP	0.0148	0.0159	0.987

Table 10: Mean True/False Positive Rate using different choices of δ based on 100 replications(Example 1)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH($\delta = 0.1$)	0.0162	0.0125	0.95	0.0123	0.0114	0.98
Modified GH($\delta = 0.2$)	0.0163	0.0126	0.95	0.0126	0.0118	0.98
Modified GH($\delta = 0.3$)	0.0164	0.0127	0.95	0.0138	0.0120	0.97
Modified GH($\delta = 0.4$)	0.0174	0.0126	0.94	0.0139	0.0121	0.97
Modified GH($\delta = 0.5$)	0.0174	0.0127	0.94	0.0141	0.0123	0.97
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH($\delta = 0.1$)	0.0109	0.0122	1.00	0.0094	0.0105	1.00
Modified GH($\delta = 0.2$)	0.0111	0.0123	1.00	0.0095	0.0106	1.00
Modified GH($\delta = 0.3$)	0.0112	0.0125	1.00	0.0095	0.0106	1.00
Modified GH($\delta = 0.4$)	0.0114	0.0127	1.00	0.0096	0.0107	1.00
Modified GH($\delta = 0.5$)	0.0114	0.0127	1.00	0.0097	0.0108	1.00

Table 11: Mean True/False Positive Rate using different choices of δ based on 100 replications(Example 2)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH ($\delta = 0.1$)	0.00659	0.0051	0.98	0.00306	0.0034	1.000
Modified GH($\delta = 0.2$)	0.00686	0.0054	0.98	0.00324	0.0036	1.000
Modified GH($\delta = 0.3$)	0.00731	0.0059	0.98	0.00333	0.0037	1.000
Modified GH($\delta = 0.4$)	0.00749	0.0061	0.98	0.00342	0.0038	1.000
Modified GH($\delta = 0.5$)	0.00758	0.0062	0.98	0.00351	0.0039	1.000
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH($\delta = 0.1$)	0.00405	0.0045	1.00	0.00261	0.0029	1.00
Modified GH($\delta = 0.2$)	0.00405	0.0045	1.00	0.00261	0.0029	1.00
Modified GH($\delta = 0.3$)	0.00414	0.0046	1.00	0.00261	0.0029	1.00
Modified GH($\delta = 0.4$)	0.00414	0.0046	1.00	0.00261	0.0029	1.00
Modified GH($\delta = 0.5$)	0.00423	0.0047	1.00	0.00270	0.0030	1.00

Table 12: Mean True/False Positive Rate using $c_2 = 1$ based on 100 replications(Example 1)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_1 = 2$)	0.0172	0.0128	0.94	0.0149	0.0122	0.96
Modified GH(EB2,$c_1 = 3$)	0.0178	0.0131	0.94	0.0152	0.0125	0.96
Modified GH(EB2,$c_1 = 4$)	0.0260	0.0167	0.89	0.0258	0.0165	0.89
Modified GH(EB2,$c_1 = 5$)	0.0295	0.0172	0.86	0.0274	0.0171	0.88
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_1 = 2$)	0.0112	0.0124	1.00	0.0107	0.0119	1.00
Modified GH(EB2,$c_1 = 3$)	0.0114	0.0127	1.00	0.0110	0.0122	1.00
Modified GH(EB2,$c_1 = 4$)	0.0116	0.0129	1.00	0.0113	0.0126	1.00
Modified GH(EB2,$c_1 = 5$)	0.0119	0.0132	1.00	0.0115	0.0128	1.00

Table 13: Mean True/False Positive Rate using $c_2 = 1$ based on 100 replications(Example 2)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_1 = 2$)	0.0084	0.006	0.97	0.00306	0.0034	1.00
Modified GH(EB2,$c_1 = 3$)	0.0123	0.007	0.94	0.00315	0.0035	1.00
Modified GH(EB2,$c_1 = 4$)	0.0160	0.010	0.93	0.00342	0.0038	1.00
Modified GH(EB2,$c_1 = 5$)	0.0187	0.013	0.93	0.00369	0.0041	1.00
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_1 = 2$)	0.00459	0.0051	1.00	0.00252	0.0028	1.00
Modified GH(EB2,$c_1 = 3$)	0.00468	0.0052	1.00	0.00261	0.0029	1.00
Modified GH(EB2,$c_1 = 4$)	0.00504	0.0056	1.00	0.00288	0.0032	1.00
Modified GH(EB2,$c_1 = 5$)	0.00522	0.0058	1.00	0.00306	0.0034	1.00

Table 14: Mean True/False Positive Rate using $c_1 = 2$ based on 100 replications(Example 1)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_2 = 1$)	0.0172	0.0128	0.94	0.0149	0.0122	0.96
Modified GH(EB2,$c_2 = 2$)	0.0176	0.0129	0.94	0.0161	0.0123	0.95
Modified GH(EB2,$c_2 = 3$)	0.0188	0.0131	0.93	0.0163	0.0126	0.95
Modified GH(EB2,$c_2 = 4$)	0.0189	0.0132	0.93	0.0175	0.0128	0.94
Modified GH(EB2,$c_2 = 5$)	0.0201	0.0134	0.92	0.0178	0.0131	0.94
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_2 = 1$)	0.0112	0.0124	1.00	0.0107	0.0119	1.00
Modified GH(EB2,$c_2 = 2$)	0.0113	0.0126	1.00	0.0110	0.0122	1.00
Modified GH(EB2,$c_2 = 3$)	0.0115	0.0128	1.00	0.0113	0.0125	1.00
Modified GH(EB2,$c_2 = 4$)	0.0118	0.0131	1.00	0.0116	0.0129	1.00
Modified GH(EB2,$c_2 = 5$)	0.0119	0.0132	1.00	0.0117	0.0130	1.00

Table 15: Mean True/False Positive Rate using $c_1 = 2$ based on 100 replications(Example 2)

Small group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_2 = 1$)	0.0084	0.006	0.97	0.00306	0.0034	1.00
Modified GH(EB2,$c_2 = 2$)	0.0103	0.007	0.96	0.00315	0.0035	1.00
Modified GH(EB2,$c_2 = 3$)	0.0121	0.009	0.96	0.00324	0.0036	1.00
Modified GH(EB2,$c_2 = 4$)	0.0140	0.010	0.95	0.00342	0.0038	1.00
Modified GH(EB2,$c_2 = 5$)	0.0140	0.010	0.95	0.00351	0.0039	1.00
Large group coefficients						
	$\rho = 0$			$\rho = 0.5$		
Prior	MP	FPR	TPR	MP	FPR	TPR
Modified GH(EB2,$c_2 = 1$)	0.00459	0.0051	1.00	0.00252	0.0028	1.00
Modified GH(EB2,$c_2 = 2$)	0.00468	0.0052	1.00	0.00261	0.0029	1.00
Modified GH(EB2,$c_2 = 3$)	0.00486	0.0054	1.00	0.00279	0.0031	1.00
Modified GH(EB2,$c_2 = 4$)	0.00495	0.0055	1.00	0.00306	0.0034	1.00
Modified GH(EB2,$c_2 = 5$)	0.00495	0.0055	1.00	0.00306	0.0034	1.00

Table 16: Mean True/False Positive Rate using different choices of τ based on 100 replications(Example 8)

Small group coefficients			
Prior	MP	FPR	TPR
Modified GH(τ_{1n})	0.0137	0.0140	0.988
Modified GH(τ_{2n})	0.0151	0.0148	0.983
Modified GH(τ_{3n})	0.0149	0.0144	0.982
Modified GH(τ_{4n})	0.0151	0.0146	0.982
Modified GH(τ_{5n})	0.0144	0.0141	0.984
Modified GH(FBtruncated)	0.0131	0.0139	0.991
Large group coefficients			
Prior	MP	FPR	TPR
Modified GH(τ_{1n})	0.0111	0.0131	1.00
Modified GH(τ_{2n})	0.0126	0.0148	1.00
Modified GH(τ_{3n})	0.0123	0.0145	1.00
Modified GH(τ_{4n})	0.0126	0.0148	1.00
Modified GH(τ_{5n})	0.0115	0.0135	1.00
Modified GH(FBtruncated)	0.0121	0.0142	1.00

Table 17: Mean True/False Positive Rate using different choices of τ based on 100 replications(Example 9)

Number of active groups=4			
Prior	MP	FPR	TPR
Modified GH(τ_{1n})	0.0175	0.0204	0.998
Modified GH(τ_{2n})	0.0186	0.0214	0.996
Modified GH(τ_{3n})	0.0185	0.0211	0.998
Modified GH(τ_{4n})	0.0184	0.0209	0.995
Modified GH(τ_{5n})	0.0164	0.0191	0.995
Modified GH(FBtruncated)	0.0158	0.0184	0.998
Number of active groups=8			
Prior	MP	FPR	TPR
Modified GH (τ_{1n})	0.0102	0.0121	0.994
Modified GH(τ_{2n})	0.0133	0.0139	0.988
Modified GH(τ_{3n})	0.0132	0.0138	0.988
Modified GH(τ_{4n})	0.0127	0.0136	0.989
Modified GH(τ_{5n})	0.0116	0.0119	0.989
Modified GH(FBtruncated)	0.0115	0.0118	0.989