# VARIATIONAL BAYES FOR HIGH-DIMENSIONAL STRUCTURED MIXTURE MODEL

*Fudan University*

## Supplementary Material

Section S1 contains the proofs of the results under the case of a known noise variance in Section 3.2. Section S2 extends theoretical results to the unknown variance case with the proofs. Section S3 provides CAVI updates and their derivation. Section S4 reports additional simulation results. Section S5 presents additional information and results on the two real applications.

## S1 Proof of Theoretical Results Under Known Variance Case

In this section, we provide the proofs of the theoretical results in Section 3.2, where a known $\sigma_y^2$ is assumed. Before entering the proof of the main results, we first state several important lemmas whose proofs are deferred in Section S1.2. We denote the log-likelihood as $l_n(\boldsymbol{\theta})$ and, for any model $S$, $l_n(\boldsymbol{\theta}_S)$ represents the log-likelihood with $\boldsymbol{\theta}_{S^c} = \mathbf{0}$. We define

$Z_n(\boldsymbol{\theta}_S) := l_n(\boldsymbol{\theta}_S) - l_n(\boldsymbol{\theta}_{0S})$ and

$$V_n = \sup_{S \in \mathcal{S}} \sup_{\boldsymbol{\theta} \in \Theta(M)} \frac{1}{n} \frac{|Z_n(\boldsymbol{\theta}_S) - \mathbb{E}Z_n(\boldsymbol{\theta}_S)|}{\|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_1 \vee \lambda_0},$$

where $\mathcal{S}$ is the considered model space and $\lambda_0 = \sqrt{\log p/n}$.

**Lemma S1.1.** Under Condition 1, for $\boldsymbol{\theta} \in \Theta(M)$ there exists some constant $c_R > 0$, such that for any constant $R \geq c_R$, we have as $n \to \infty$,

$$P(V_n \leq R\lambda_0) \to 1.$$

Further under Conditions 1 and 2, for some constants $c_1, c_2, c_3 > 0$, it holds that on $\{V_n \leq R\lambda_0\}$,

$$Z_n(\boldsymbol{\theta}_S) \geq -c_1 n \|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 - 2c_2(|S| + 2)\log p,$$

$$Z_n(\boldsymbol{\theta}_S) \leq -c_3 n \|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 + c_2(|S| + 2)\log p.$$

Lemma S1.1 constructs an event with probability going to 1 and states that if this event holds, the divergence between the log-likelihoods of $\boldsymbol{\theta}_S$ and $\boldsymbol{\theta}_{0S}$, i.e., $Z_n(\boldsymbol{\theta}_S)$, is bounded both above and below in terms of the $\ell_2$ error $\|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2$ and model size $|S|$.

**Lemma S1.2.** Let $\tilde{\Theta}$ be a subset of the parameter space $\Theta$ and $\boldsymbol{\theta}_0$ be the underlying true parameter. For any event $\mathcal{E}$ and any distribution $Q$ for $\boldsymbol{\theta}$, if there exists some $C > 0$ and $\xi_n > 0$ such that

$$\mathbb{E}_{\boldsymbol{\theta}_0}[\Pi(\boldsymbol{\theta} \in \tilde{\Theta} \mid \boldsymbol{Y}) 1_{\mathcal{E}}] \leq C e^{-\xi_n}, \tag{S1.1}$$

then

$$\mathbb{E}_{\boldsymbol{\theta}_0}[Q(\boldsymbol{\theta} \in \tilde{\Theta})1_{\mathcal{E}}] \leq \frac{2}{\xi_n}\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\mathrm{KL}[Q(\boldsymbol{\theta})\|\Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})]1_{\mathcal{E}}] + Ce^{-\xi_n/2}\right). \quad \text{(S1.2)}$$

Lemma S1.2 builds a connection between the variational distribution and the exact posterior distribution. To leverage Lemma S1.2, we define the following events for $R$ defined in Lemma S1.1 and $L, M_1, M_2 > 0$:

$$\mathcal{E}_{n,1}(R) = \{V_n \leq R\lambda_0\},$$

$$\mathcal{E}_{n,2}(L) = \{\Pi(\boldsymbol{\theta} \in \Theta(M) : |S| \leq Ls_0 \mid \boldsymbol{Y}) \geq 3/4\},$$

$$\mathcal{E}_{n,3}(M_1, M_2) = \left\{\Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) \leq e^{-M_2 s_0 \log p}\right\}, \quad \text{(S1.3)}$$

where $\tilde{\boldsymbol{Z}} = (\boldsymbol{Z}, \boldsymbol{T})$. We further define

$$\mathcal{E}_n = \mathcal{E}_{n,1}(R) \cap \mathcal{E}_{n,2}(L) \cap \mathcal{E}_{n,3}(M_1, M_2),$$

and our ultimate goals are to establish the upper bounds in (S1.1) and (S1.2) on $\mathcal{E}_n$ with respect to some predefined subspaces $\tilde{\Theta}$. We first state the following lemma to bound the KL divergence between the variational posterior and the exact posterior.

**Lemma S1.3.** Under Conditions 1-3, for sufficiently large $p$ and some constant $C_K > 0$, we have

$$\mathrm{KL}\left[Q^*(\boldsymbol{\theta})\|\Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})\right] 1_{\mathcal{E}_n} \leq C_K s_0 \log p. \quad \text{(S1.4)}$$

Based on the results of Lemma S1.3, we are left to prove the posterior contraction results on the event $\mathcal{E}_n$.

**Lemma S1.4.** Under Conditions 1-3, on $\mathcal{E}_{n,1}(R)$, we have

$$\int_{\Theta(M)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) \geq q^{s_0}(1-q)^{p-s_0} \exp\{-2c_2(s_0+2)\log p\}$$
$$\times \left(\frac{1}{2c_1 n\tau^2+1}\right)^{\frac{s_0}{2}} \left(\frac{1}{2c_1 n\sigma_\alpha^2+1}\right)$$
$$\times \exp\left\{-\frac{c_1 n}{2c_1 n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2\right\} \exp\left\{-\frac{c_1 n}{2c_1 n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2\right\},$$

where $\boldsymbol{\theta}_{0,-} = (\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T)^T$.

**Lemma S1.5.** Under Conditions 1-3, for $L \geq 2 + 5c_2(s_0+2)/s_0$,

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq Ls_0 \mid Y\right) 1_{\mathcal{E}_{n,1}(R)}\right] \preceq \exp\left\{-\frac{1}{5}(L-2)s_0 \log p\right\}.$$

**Lemma S1.6.** Under Conditions 1-3, for $M_1 \geq \sqrt{2\lambda_2(5c_2+2+8c_2/s_0)/c_3}$ and $L$ satisfying the condition in Lemma S1.5,

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right]$$
$$\preceq \exp\left\{-\frac{c_3 M_1^2 s_0 \log p}{2\lambda_2}\right\} + \exp\left\{-\frac{(L-2)s_0 \log p}{5}\right\}.$$

Lemma S1.4 provides a lower bound on the normalizing constant for obtaining the posterior probability of any subspace. Lemma S1.5 establishes a bound on the posterior probability of selecting models with sizes larger than a multiple of $s_0$, forming the basis for proving Lemma 1. Lemma S1.6 bounds the posterior probability of choosing $\boldsymbol{\theta}$ with large $\ell_2$ distance from $\boldsymbol{\theta}_0$, preparing for the proof of Theorem 1.

**Lemma S1.7.** Under Conditions 1-3, for $R$ defined in Lemma S1.1, $L$ defined in Lemma S1.5, $M_1$ defined in Lemma S1.6, and $M_2 \leq (cM_1^2/2\lambda_2) \wedge ((L-2)/5)$, as $p \to \infty$, we have

$$P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_n^c\right] \to 0.$$

Lemma S1.7 guarantees $\mathcal{E}_n$ holds with probability going to 1 under $\boldsymbol{\theta}_0$. Combining Lemma S1.5, S1.6, and S1.7, we are able to prove Lemma 1 and Theorem 1. To further obtain model selection consistency of the VB posterior, we need the following result on the exact posterior.

**Lemma S1.8.** Under Conditions 1-4, for $\kappa_n \geq (2c_2 s_0/c_3) \vee (2((3c_2+2)s_0 + 8c_2)/c_2 s_0)$ and $L$ defined in Lemma S1.5, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : S \neq S_0 \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \\
&\preceq 3\exp\left\{-\frac{c_2\kappa_n s_0 \log p}{2}\right\} + \exp\left\{-\frac{(L-2)s_0\log p}{5}\right\}.
\end{aligned}$$

## S1.1 Proof of Main Results

**Proof of Lemma 1**

By Lemma S1.5, for $L_n \geq L_0 := 2 + 5c_2(s_0+2)/s_0$,

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L_n s_0 \mid Y\right) 1_{\mathcal{E}_n}\right] \preceq \exp\left\{-\frac{1}{5}(L_n - 2)s_0 \log p\right\}.$$

By assigning $\xi_n = \frac{1}{5}(L_n - 2)s_0 \log p$ in Lemma S1.2, we have for some constant $C_L \geq 10C_K$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L_n s_0\right) 1_{\mathcal{E}_n}\right] \\
&\preceq \frac{10}{(L_n - 2)s_0 \log p}\left[C_K s_0 \log p + \exp\left\{-\frac{1}{5}(L_n - 2)s_0 \log p\right\}\right] \\
&\leq \frac{10C_K}{L_n - 2}(1 + o(1)) \leq \mathcal{O}\left(\frac{C_L}{L_n}\right).
\end{aligned}$$

Thus the targeted expectation can be bounded by

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L_n s_0)\right] \\
&\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L_n s_0\right) 1_{\mathcal{E}_n}\right] + P_{\boldsymbol{\theta}_0}[\mathcal{E}_n^c] \leq \mathcal{O}\left(\frac{C_L}{L_n}\right) + o(1).
\end{aligned}$$

□

**Proof of Theorem 1**

By Lemma S1.6, for $M_n \geq M_0 := 2\lambda_2(5c_2 + 2 + 8c_2/s_0)/c_3$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \geq \frac{\sqrt{M_n s_0 \log p}}{\|\boldsymbol{X}\| \vee \|\tilde{\boldsymbol{Z}}\|} \mid Y\right) 1_{\mathcal{E}_n}\right] \\
&\preceq \exp\left\{-\frac{c_3 M_n s_0 \log p}{2\lambda_2}\right\} + \exp\left\{-\frac{(L_n - 2)s_0 \log p}{5}\right\}.
\end{aligned}$$

Since $M_n$ grows more slowly than $L_n$, by assigning $\xi_n = \frac{c_3}{2\lambda_2} M_n s_0 \log p$ in Lemma S1.2, we have for some constant $C_M \geq 4\lambda_2 C_K / c_3$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \geq \frac{\sqrt{M_n s_0 \log p}}{\|\boldsymbol{X}\| \vee \|\tilde{\boldsymbol{Z}}\|}\right) 1_{\mathcal{E}_n}\right] \\
&\preceq \frac{4\lambda_2}{c_3 M_n s_0 \log p}\left[C_K s_0 \log p + 2\exp\left\{-\frac{c_3 M_n s_0 \log p}{2\lambda_2}\right\}\right] \\
&\leq \frac{4\lambda_2 C_K}{c_3 M_n}(1 + o(1)) \leq \mathcal{O}\left(\frac{C_M}{M_n}\right).
\end{aligned}$$

Thus the targeted expectation can be bounded by

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta}\in\Theta(M):\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\geq\frac{\sqrt{M_n s_0\log p}}{\|\boldsymbol{X}\|\vee\|\tilde{\boldsymbol{Z}}\|}\right)\right]\\
&\leq\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta}\in\Theta(M):\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\geq\frac{\sqrt{M_n s_0\log p}}{\|\boldsymbol{X}\|\vee\|\tilde{\boldsymbol{Z}}\|}\right)1_{\mathcal{E}_n}\right]+P_{\boldsymbol{\theta}_0}[\mathcal{E}_n^c]\\
&\leq\mathcal{O}\left(\frac{C_M}{M_n}\right)+o(1).
\end{aligned}$$

□

**Proof of Theorem 2**

By Lemma S1.8, for $\kappa_n\geq\kappa_0:=(2c_2s_0/c_3)\vee(2((3c_2+2)s_0+8c_2)/c_2s_0)$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M):S\neq S_0\mid Y\right)1_{\mathcal{E}_n}\right]\\
&\preceq 3\exp\left\{-\frac{c_2\kappa_n s_0\log p}{2}\right\}+\exp\left\{-\frac{(L_n-2)s_0\log p}{5}\right\}.
\end{aligned}$$

Since $\kappa_n$ grows more slowly than $L_n$, by assigning $\xi_n=\frac{c_2}{2}\kappa_n s_0\log p$ in Lemma S1.2, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta}\in\Theta(M):S\neq S_0\right)1_{\mathcal{E}_n}\right]\\
&\preceq\frac{4}{c_2\kappa_n s_0\log p}\left[C_K s_0\log p+4\exp\left\{-\frac{c_2\kappa_n s_0\log p}{2}\right\}\right]\\
&\leq\frac{4C_K}{c_2\kappa_n}(1+o(1))\leq\mathcal{O}\left(\frac{C_\kappa}{\kappa_n}\right),
\end{aligned}$$

where $C_M\geq 4C_K/c_2$. Thus the targeted expectation can be bounded by

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta}\in\Theta(M):S\neq S_0\right)\right]\\
&\leq\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta}\in\Theta(M):S\neq S_0\right)1_{\mathcal{E}_n}\right]+P_{\boldsymbol{\theta}_0}[\mathcal{E}_n^c]\leq\mathcal{O}\left(\frac{C_\kappa}{\kappa_n}\right)+o(1).
\end{aligned}$$

□

## S1.2 Proof of Technical Lemmas

**Proof of Lemma S1.1**

This lemma is modified from Lemma 3.1 and Lemma 3.2 in Zhang et al. (2025), where we substitute the value of $\lambda_0$ into the inequality and refine the constant terms.

**Proof of Lemma S1.2**

This lemma is modified from Theorem 5 in Ray and Szabó (2022).

**Proof of Lemma S1.3**

The exact posterior can be written as

$$\Pi(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \sum_{S \in \mathcal{S}} \hat{w}_S \Pi_S(\boldsymbol{\theta}_S \mid \boldsymbol{Y}) \otimes \delta_0(\boldsymbol{\theta}_{S^c}),$$

where $\hat{w}_S$ denotes the posterior model weights satisfying $0 \leq \hat{w}_S \leq 1$ and $\sum_{S \in \mathcal{S}} \hat{w}_S = 1$ and $\Pi_S$ denotes the distribution with respect to model $S$. Since the VB posterior $Q^*$ minimizes the KL divergence from the exact posterior, we have

$$\mathrm{KL}[Q^*(\boldsymbol{\theta}) \| \Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})] \leq \mathrm{KL}[Q(\boldsymbol{\theta}) \| \Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})],$$

for any $Q \in \mathcal{Q}$. To establish an upper bound on the KL divergence, we carefully design a $\check{Q} \in \mathcal{Q}$, where for any model $S$, the VB posterior can be expressed as

$$\check{Q}(\boldsymbol{\theta}) = N_S(\boldsymbol{\theta}_S; \boldsymbol{\mu}_S, \boldsymbol{D}_S) \otimes \delta_0(\boldsymbol{\theta}_{S^c}) = \prod_{j \in S} N(\theta_j; \mu_{Sj}, \sigma_{Sj}^2) \otimes \prod_{j \in S^c} \delta_0(\theta_j),$$

where $N_S$ denotes an $|S|+2$ dimensional normal distribution corresponding to $(\boldsymbol{\beta}_S, \boldsymbol{\gamma}_S, \boldsymbol{\alpha})$. We set

$$\boldsymbol{\mu}_S = \boldsymbol{\theta}_{0S} - (1+\tau^{-2})\Sigma_S \boldsymbol{\theta}_{0S},$$
$$\Sigma_S^{-1} = (2c_3 n + \tau^{-2}) I_S + \begin{pmatrix} \tilde{\boldsymbol{Z}}_S^T \tilde{\boldsymbol{Z}}_S & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & n\lambda_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{X}_S^T \boldsymbol{X}_S \end{pmatrix},$$

with $\boldsymbol{D}_S = \operatorname{diag}(\Sigma_S)$ being the diagonal matrix. Since the chosen $\check{Q}$ is only absolutely continuous with respect to $\hat{w}_S \Pi_S(\boldsymbol{\theta}_S \mid \boldsymbol{Y}) \otimes \delta_0(\boldsymbol{\theta}_{S^c})$ in $\Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})$, we have

$$\begin{aligned} \mathrm{KL}[\check{Q}(\boldsymbol{\theta}) \| \Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})] &= \mathbb{E}_{\check{Q}}\left[\log \frac{dN_S(\boldsymbol{\theta}_S; \boldsymbol{\mu}_S, \boldsymbol{D}_S) \otimes \delta_0(\boldsymbol{\theta}_{S^c})}{\hat{w}_S d\Pi_S(\boldsymbol{\theta}_S \mid \boldsymbol{Y}) \otimes \delta_0(\boldsymbol{\theta}_{S^c})}\right] \\ &= \log \frac{1}{\hat{w}_S} + \mathrm{KL}[N_S(\boldsymbol{\theta}_S; \boldsymbol{\mu}_S, \boldsymbol{D}_S) \| \Pi_S(\boldsymbol{\theta}_S \mid \boldsymbol{Y})]. \end{aligned} \tag{S1.5}$$

We first claim that on the event $\mathcal{E}_n$, there exists a model $\check{S}$ satisfying the following properties:

$$|\check{S}| \le L s_0, \quad \|\boldsymbol{\theta}_{0\check{S}^c}\|_2 \le \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|}, \quad \hat{w}_{\check{S}} \ge (2e)^{-1} p^{-L s_0}.$$

On $\mathcal{E}_n$, we have

$$\begin{aligned} &\Pi\left(\boldsymbol{\theta} : \|\boldsymbol{\theta}_{0S^c}\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) \\ &\le \Pi\left(\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) \le \exp\{-M_2 s_0 \log p\} \to 0. \end{aligned}$$

Thus the posterior weights satisfy

$$\sum_{\substack{S:|S|\leq Ls_0,\\ \|\boldsymbol{\theta}_{0S^c}\|_2\leq \frac{M_1\sqrt{s_0\log p}}{\|\boldsymbol{X}\|\wedge\|\tilde{\boldsymbol{Z}}\|}}} \hat{w}_S \geq \frac{3}{4} - \exp\{-M_2 s_0 \log p\} \geq \frac{1}{2},$$

with the number of elements in the sum bounded by

$$\sum_{d=0}^{Ls_0} \binom{p}{d} \leq \sum_{d=0}^{Ls_0} \frac{p^d}{d!} \leq e p^{Ls_0},$$

which implies that there exists at least a model $\check{S}$ of size $|\check{S}| \leq Ls_0$ with posterior weight $\hat{w}_{\check{S}} \geq (2e)^{-1} p^{-Ls_0}$ satisfying $\|\boldsymbol{\theta}_{0\check{S}^c}\|_2 \leq \frac{M_1\sqrt{s_0\log p}}{\|\boldsymbol{X}\|\wedge\|\tilde{\boldsymbol{Z}}\|}$. Thus in (S1.5), the first term is bounded by $2 + Ls_0 \log p$.

Denote $\check{Q}_{\check{S}} = N_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \boldsymbol{D}_{\check{S}})$ and the second KL term in (S1.5) can be rewritten as

$$\mathrm{KL}[\check{Q}_{\check{S}} \| N_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})] + \mathbb{E}_{\check{Q}_{\check{S}}}\left[\log \frac{dN_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})}{d\Pi_{\check{S}}(\boldsymbol{\theta}_{\check{S}} \mid \boldsymbol{Y})}\right]. \tag{S1.6}$$

The KL divergence between the two multivariate-normal distributions $\check{Q}_{\check{S}}$ and $N_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})$ can be expressed as

$$\mathrm{KL}[\check{Q}_{\check{S}} \| N_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})] = \frac{1}{2}\left[\log \frac{|\Sigma_{\check{S}}|}{|\boldsymbol{D}_{\check{S}}|} - |\check{S}| - 2 + \mathrm{tr}(\Sigma_{\check{S}}^{-1} \boldsymbol{D}_{\check{S}})\right].$$

By definitions, $\mathrm{tr}(\Sigma_{\check{S}}^{-1} \boldsymbol{D}_{\check{S}}) = |\check{S}| + 2$ while their determinants are

$$\begin{aligned} |\Sigma_{\check{S}}| &\leq \left(\lambda_{\min}(\tilde{\boldsymbol{Z}}_{\check{S}}^T \tilde{\boldsymbol{Z}}_{\check{S}}) + 2c_3 n + \tau^{-2}\right)^{-|\check{S}_{\boldsymbol{\beta}}|-1} (n\lambda_2)^{-1} \left(\lambda_{\min}(\boldsymbol{X}_{\check{S}}^T \boldsymbol{X}_{\check{S}}) + 2c_3 n + \tau^{-2}\right)^{-|\check{S}_{\boldsymbol{\gamma}}|} \\ &\leq n^{-|\check{S}|-2} \left(\lambda_1 + 2c_3 + \frac{1}{n\tau^2}\right)^{-|\check{S}|-1} \lambda_2^{-1}, \end{aligned}$$

and

$$|\boldsymbol{D}_{\check{S}}^{-1}| = \prod_{j=1}^{|\check{S}_{\boldsymbol{\beta}}|+1} \left[(\tilde{\boldsymbol{Z}}_{\check{S}}^T \tilde{\boldsymbol{Z}}_{\check{S}})_{jj} + 2c_3 n + \frac{1}{\tau^2}\right] n\lambda_2 \prod_{\ell=1}^{|\check{S}_{\boldsymbol{\gamma}}|} \left[(\boldsymbol{X}_{\check{S}}^T \boldsymbol{X}_{\check{S}})_{\ell\ell} + 2c_3 n + \frac{1}{\tau^2}\right]$$
$$\leq \left(\|\tilde{\boldsymbol{Z}}\|^2 + 2c_3 n + \tau^{-2}\right)^{|\check{S}_{\beta}|+1} n\lambda_2 \left(\|\boldsymbol{X}\|^2 + 2c_3 n + \tau^{-2}\right)^{|\check{S}_{\gamma}|}$$
$$\leq n^{|\check{S}|+2} \left(\lambda_2 + 2c_3 + \frac{1}{n\tau^2}\right)^{|\check{S}|+1} \lambda_2.$$

Combining the above bounds, we have

$$\mathrm{KL}[\check{Q}_{\check{S}} \| N_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})] \leq \frac{1}{2} \log \left(\frac{\lambda_2 + 2c_3 + \frac{1}{n\tau^2}}{\lambda_1 + 2c_3 + \frac{1}{n\tau^2}}\right)^{|\check{S}|+1} \leq \frac{Ls_0 + 1}{2} \log \left(\frac{\lambda_2 + 2c_3 + \frac{1}{n\tau^2}}{\lambda_1 + 2c_3 + \frac{1}{n\tau^2}}\right),$$

and thus the first term in (S1.6) is bounded by $\mathcal{O}(Ls_0)$. We are only left with the second non-diagonal term in (S1.6).

The probability density of the exact posterior $\Pi_{\check{S}}(\boldsymbol{\theta}_{\check{S}} \mid \boldsymbol{Y})$ is proportional to

$$\exp\left\{ l_n(\boldsymbol{\theta}_{\check{S}}) - l_n(\boldsymbol{\theta}_{0\check{S}}) - \frac{1}{2\tau^2} \|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 \right.$$
$$\left. - \frac{1}{\tau^2} \boldsymbol{\theta}_{0\check{S},-}^T (\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) - \frac{1}{2\sigma_\alpha^2} \|\boldsymbol{\alpha}\|_2^2 \right\},$$

where $\boldsymbol{\theta}_{\check{S},-} = (\boldsymbol{\beta}_{\check{S}}^T, \boldsymbol{\gamma}_{\check{S}}^T)^T$. Thus the non-diagonal term can be rewritten as

$$\mathbb{E}_{\check{Q}_{\check{S}}} \left[\log \frac{dN_{\check{S}}(\boldsymbol{\theta}_{\check{S}}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})}{d\Pi_{\check{S}}(\boldsymbol{\theta}_{\check{S}} \mid \boldsymbol{Y})}\right]$$
$$= \mathbb{E}_{\check{Q}_{\check{S}}} \left[\log \frac{\frac{1}{D_N} \exp\{-\frac{1}{2}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})^T \Sigma_{\check{S}}^{-1} (\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}) - (\boldsymbol{\theta}_{0\check{S}} - \boldsymbol{\mu}_{\check{S}})^T \Sigma_{\check{S}}^{-1} (\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})\}}{\frac{1}{D_\Pi} \exp\{Z_n(\boldsymbol{\theta}_{\check{S}}) - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 - \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\}}\right]$$
$$= \log \frac{D_\Pi}{D_N} + \mathbb{E}_{\check{Q}_{\check{S}}} \left[-\frac{1}{2}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})^T \Sigma_{\check{S}}^{-1} (\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}) - (\boldsymbol{\theta}_{0\check{S}} - \boldsymbol{\mu}_{\check{S}})^T \Sigma_{\check{S}}^{-1} (\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}) \right.$$
$$\left. - Z_n(\boldsymbol{\theta}_{\check{S}}) + \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 + \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) + \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\right], \tag{S1.7}$$

where

$$D_N = \int_{\Theta_{\check{S}}(M)} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})^T \Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}) - (\boldsymbol{\theta}_{0\check{S}} - \boldsymbol{\mu}_{\check{S}})^T \Sigma_{S}^{-1}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})\right\} d\boldsymbol{\theta}_{\check{S}}$$
$$D_\Pi = \int_{\Theta_{\check{S}}(M)} \exp\left\{Z_n(\boldsymbol{\theta}_{\check{S}}) - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 - \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\right\} d\boldsymbol{\theta}_{\check{S}}.$$

We first bound the ratio of the normalizing constants. Define the subspace

$$B_{\check{S}} = \left\{\boldsymbol{\theta}_{\check{S}} \in \Theta_{\check{S}}(M) : \|\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}\|_2 \le \frac{2M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|}\right\}.$$

If we define $\bar{\boldsymbol{\theta}}_{\check{S}} \in \mathbb{R}^{p_n}$ with $\boldsymbol{\theta}_{\check{S}}$ for $\check{S}$ and $\mathbf{0}$ for $\check{S}^c$, then on $\mathcal{E}_n$,

$$\begin{aligned}
\Pi_{\check{S}}(B_{\check{S}}^c \mid \boldsymbol{Y}) &\le \Pi_{\check{S}}\left(\boldsymbol{\theta}_{\check{S}} \in \Theta_{\check{S}}(M) : \|\bar{\boldsymbol{\theta}}_{\check{S}} - \boldsymbol{\theta}_0\|_2 + \|\boldsymbol{\theta}_{0\check{S}^c}\|_2 > \frac{2M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) \\
&\le \Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) \\
&\le \exp\{-M_2 s_0 \log p\} \le \frac{1}{2}.
\end{aligned}$$

Using Bayes' formula, we have on $\mathcal{E}_n$,

$$\begin{aligned}
&\Pi_{\check{S}}(B_{\check{S}} \mid \boldsymbol{Y}) \\
&= \frac{\int_{B_{\check{S}}} \exp\left\{Z_n(\boldsymbol{\theta}_{\check{S}}) - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 - \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\right\} d\boldsymbol{\theta}_{\check{S}}}{\int_{\Theta_{\check{S}}(M)} \exp\left\{Z_n(\boldsymbol{\theta}_{\check{S}}) - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 - \frac{1}{\tau^2}\boldsymbol{\theta}_{0S,-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\right\} d\boldsymbol{\theta}_{\check{S}}} \ge \frac{1}{2}.
\end{aligned}$$

Thus on $\mathcal{E}_n$, the log ratio $\log(D_\Pi/D_N)$ is bounded by

$$
\begin{aligned}
&\log \frac{2\int_{B_{\check{S}}} \exp\left\{Z_n(\boldsymbol{\theta}_{\check{S}}) - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 - \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\right\} d\boldsymbol{\theta}_{\check{S}}}{\int_{\Theta_{\check{S}}(M)} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}) - (\boldsymbol{\theta}_{0\check{S}} - \boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})\right\} d\boldsymbol{\theta}_{\check{S}}} \\
&\leq \log \sup_{\boldsymbol{\theta}_{\check{S}} \in B_{\check{S}}} \exp\Big\{-c_3 n\|\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}\|_2^2 + c_2(|\check{S}|+2)\log p \\
&- \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}\|_2^2 - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2 - \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-} - \boldsymbol{\theta}_{0\check{S},-}) \\
&+\frac{1}{2}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0\check{S}}) + (\boldsymbol{\theta}_{0\check{S}} - \boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})\Big\} + \log 2 \\
&\leq c_2(Ls_0+2)\log p - \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 + \log 2 \\
&+ \sup_{\boldsymbol{\theta}_{\check{S}} \in B_{\check{S}}} \Big\{-\frac{1}{2}(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})^T\left[(2c_3 n + \tau^{-2})I_{\check{S}} - \Sigma_{\check{S}}^{-1}\right](\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}) \\
&+ \left[(\boldsymbol{\theta}_{0\check{S}} - \boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1} - \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S}}^T\right](\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})\Big\}.
\end{aligned}
$$

By the definitions of $\boldsymbol{\mu}_{\check{S}}$ and $\Sigma_{\check{S}}$, if we define $\tilde{\boldsymbol{\beta}}_{\check{S}} = (\boldsymbol{\beta}_{\check{S}}^T, \alpha_1)^T$, the sup term is bounded by

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}_{\check{S}} \in B_{\check{S}}} \left\{\frac{1}{2}\|\tilde{\boldsymbol{Z}}_{\check{S}}(\tilde{\boldsymbol{\beta}}_{\check{S}} - \tilde{\boldsymbol{\beta}}_{0\check{S}})\|_2^2 + \frac{1}{2}n\lambda_2(\alpha_2 - \alpha_{20})^2 + \frac{1}{2}\|\boldsymbol{X}_{\check{S}}(\boldsymbol{\gamma}_{\check{S}} - \boldsymbol{\gamma}_{0\check{S}})\|_2^2 + \boldsymbol{\theta}_{0\check{S}}^T(\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}})\right\} \\
&\leq \sup_{\boldsymbol{\theta}_{\check{S}} \in B_{\check{S}}} \left\{\frac{1}{2}n\lambda_2\|\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}\|_2^2 + \|\boldsymbol{\theta}_{0\check{S}}\|_2\|\boldsymbol{\theta}_{\check{S}} - \boldsymbol{\theta}_{0\check{S}}\|_2\right\} \\
&\leq \frac{2\lambda_2}{\lambda_1}M_1^2 s_0 \log p + M\|\boldsymbol{\theta}_{0\check{S}}\|_2.
\end{aligned}
$$

Thus the log ratio of the normalizing constants is controlled by $\mathcal{O}(s_0 \log p)$.

We now deal with the expectation term in (S1.7). By Lemma S1.1, on

$\mathcal{E}_n$, it is bounded by

$$
\begin{aligned}
&2c_2(|\check{S}|+2)\log p + \mathbb{E}_{\check{Q}_{\check{S}}}\Bigg[-\frac{1}{2}(\boldsymbol{\theta}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}) \\
&-(\boldsymbol{\theta}_{0\check{S}}-\boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}) + c_1 n\|\boldsymbol{\theta}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}\|_2^2 \\
&+\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-}\|_2^2 + \frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-\alpha}-\boldsymbol{\theta}_{0\check{S},-}) + \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2\Bigg] \\
&= 2c_2(|\check{S}|+2)\log p - \frac{1}{2}(\boldsymbol{\mu}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\mu}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}) - \frac{1}{2}\operatorname{tr}(\Sigma_{\check{S}}^{-1}\boldsymbol{D}_{\check{S}}) \\
&-(\boldsymbol{\theta}_{0\check{S}}-\boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\mu}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}) + \left(c_1 n + \frac{1}{2\tau^2}\right)\left[\|\boldsymbol{\mu}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}\|_2^2 + \operatorname{tr}(\boldsymbol{D}_{\check{S}})\right] \\
&+\frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S}}^T(\boldsymbol{\mu}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}) + \mathbb{E}_{\check{Q}_{\check{S}}}\left[\frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2 - \frac{1}{2\tau^2}(\|\boldsymbol{\alpha}\|_2^2 - \|\boldsymbol{\alpha}_0\|_2^2)\right] \\
&\le 2c_2(|\check{S}|+2)\log p - \frac{|\check{S}|+2}{2} + \frac{1}{2}(1+\tau^{-2})^2\boldsymbol{\theta}_{0\check{S}}^T\Sigma_{\check{S}}\boldsymbol{\theta}_{0\check{S}} \\
&+\left(c_1 n + \frac{1}{2\tau^2}\right)\left[(1+\tau^{-2})^2\boldsymbol{\theta}_{0\check{S}}^T\Sigma_{\check{S}}^2\boldsymbol{\theta}_{0\check{S}} + \operatorname{tr}(\boldsymbol{D}_{\check{S}})\right] + \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 \\
&-\left(\frac{1}{\tau^2}+\frac{1}{\tau^4}\right)\boldsymbol{\theta}_{0\check{S}}^T\Sigma_{\check{S}}\boldsymbol{\theta}_{0\check{S}} + \frac{1}{2\sigma_\alpha^2}\left[\boldsymbol{\theta}_{0\check{S}}^T\left(I_{\check{S}}-(1+\tau^{-2})\Sigma_{\check{S}}\right)^2\boldsymbol{\theta}_{0\check{S}} + \operatorname{tr}(\boldsymbol{D}_{\check{S}})\right].
\end{aligned}
$$

By the definition of $\boldsymbol{D}_{\check{S}}$ and $\Sigma_{\check{S}}$, the trace is bounded by

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{D}_{\check{S}}) = &\sum_{j=1}^{|\check{S}_{\boldsymbol{\beta}}|+1}\frac{1}{(\tilde{\boldsymbol{Z}}_{\check{S}}^T\tilde{\boldsymbol{Z}}_{\check{S}})_{jj}+2c_3 n+\tau^{-2}} + \frac{1}{n\lambda_2+2c_3 n+\tau^{-2}} \\
&+\sum_{\ell=1}^{|\check{S}_{\boldsymbol{\gamma}}|}\frac{1}{(\boldsymbol{X}_{\check{S}}^T\boldsymbol{X}_{\check{S}})_{\ell\ell}+2c_3 n+\tau^{-2}} \le \frac{|\check{S}|+2}{2c_3 n},
\end{aligned}
$$

and the quadratic forms are bounded by

$$
\frac{1}{n\lambda_2+2c_3 n+\tau^{-2}}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 \le \boldsymbol{\theta}_{0\check{S}}^T\Sigma_{\check{S}}\boldsymbol{\theta}_{0\check{S}} \le \frac{1}{n\lambda_1+2c_3 n+\tau^{-2}}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2,
$$

and

$$
\frac{1}{(n\lambda_2+2c_3 n+\tau^{-2})^2}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 \le \boldsymbol{\theta}_{0\check{S}}^T\Sigma_{\check{S}}^2\boldsymbol{\theta}_{0\check{S}} \le \frac{1}{(n\lambda_1+2c_3 n+\tau^{-2})^2}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2.
$$

On $\mathcal{E}_n$, the expectation term is bounded by

$$
\begin{aligned}
&2c_2(|\check{S}|+2)\log p - \frac{|\check{S}|+2}{2} + \frac{1}{2}(1+\tau^{-2})^2 \frac{1}{n\lambda_1 + 2c_3 n + \tau^{-2}} \|\boldsymbol{\theta}_{0\check{S}}\|_2^2 \\
&+ \frac{|\check{S}|+2}{4c_3 n\sigma_\alpha^2} + \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 - \frac{1}{\tau^2}\left(1+\frac{1}{\tau^2}\right)\frac{1}{n\lambda_2 + 2c_3 n + \tau^{-2}}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 \\
&+ \left(c_1 n + \frac{1}{2\tau^2}\right)\left[(1+\tau^{-2})^2 \frac{1}{(n\lambda_1 + 2c_3 n + \tau^{-2})^2}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 + \frac{|\check{S}|+2}{2c_3 n}\right] \\
&+ \frac{1}{2\sigma_\alpha^2}\left[1 - 2\frac{1+\tau^{-2}}{n\lambda_2 + 2c_3 n + \tau^{-2}} + \frac{(1+\tau^{-2})^2}{(n\lambda_2 + 2c_3 n + \tau^{-2})^2}\right]\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 \\
&\preceq 2c_2 L s_0 \log p + \frac{c_1}{2c_3} L s_0 + \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 \\
&+ \frac{1}{n(\lambda_1 + 2c_3)}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 + \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2 + \frac{c_1}{n(\lambda_1+2c_3)^2}\|\boldsymbol{\theta}_{0\check{S}}\|_2^2,
\end{aligned}
$$

which is of the order $\mathcal{O}(s_0 \log p)$.

Combining all of the above bounds gives the result. □

**Proof of Lemma S1.4**

By the definition of the prior $\Pi$, the desired integral is a weighted sum over all models. Since each summand is non-negative, the integral is bounded below by

$$
q^{s_0}(1-q)^{p-s_0}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{s_0}\left(\frac{1}{2\pi\sigma_\alpha^2}\right)\int_{\Theta_{S_0}(M)} \frac{L_n(\boldsymbol{\theta}_{S_0})}{L_n(\boldsymbol{\theta}_{0S_0})}\exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S_0,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_{S_0}.
$$

By Lemma S1.1, on $\mathcal{E}_{n,1}(R)$, the integral term in the above display is bounded below by

$$\int_{\Theta_{S_0}(M)} \exp\{Z_n(\boldsymbol{\theta}_{S_0})\} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S_0,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_{S_0}$$
$$\geq \exp\{-2c_2(s_0+2)\log p\}$$
$$\times \int_{\Theta_{S_0}(M)} \exp\left\{-c_1 n\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_2^2 - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S_0,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_{S_0}.$$

Since $\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_1 \leq \sqrt{s_0+2}\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_2$ and

$$\left\{\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_2^2 \leq \frac{M^2}{s_0+2}\right\} \supseteq \bigcap_{j=1}^{s_0+2} \left\{(\theta_j - \theta_{0j})^2 \leq \frac{M^2}{(s_0+2)^2}\right\},$$

we can shrink the integral space and bound the integral from below by

$$\int_{\left\{\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_2 \leq \frac{M}{\sqrt{s_0+2}}\right\}} \exp\left\{-c_1 n\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_2^2 - \frac{\|\boldsymbol{\theta}_{S_0,-}\|_2^2}{2\tau^2} - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_{S_0}$$
$$\geq \int_{\bigcap_j \left\{(\theta_j - \theta_{0j})^2 \leq \frac{M^2}{(s_0+2)^2}\right\}} \exp\left\{-c_1 n\|\boldsymbol{\theta}_{S_0} - \boldsymbol{\theta}_{0S_0}\|_2^2 - \frac{\|\boldsymbol{\theta}_{S_0,-}\|_2^2}{2\tau^2} - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_{S_0}$$
$$= \prod_{j \in S_0} \int_{\theta_{0j} - \frac{M}{s_0+2}}^{\theta_{0j} + \frac{M}{s_0+2}} \exp\left\{-c_1 n(\theta_j - \theta_{0j})^2 - \frac{1}{2\tau^2}\theta_j^2\right\} d\theta_j$$
$$\times \prod_{j=1}^{2} \int_{\alpha_{0j} - \frac{M}{s_0+2}}^{\alpha_{0j} + \frac{M}{s_0+2}} \exp\left\{-c_1 n(\alpha_j - \alpha_{0j})^2 - \frac{1}{2\sigma_\alpha^2}\alpha_j^2\right\} d\alpha_j.$$

By defining $u_j = (\theta_j - \frac{2c_1 n}{2c_1 n + \frac{1}{\tau^2}}\theta_{0j}) \Big/ \sqrt{\frac{1}{2c_1 n + \frac{1}{\tau^2}}}$, each integral for $j \in S_0$ can be written as

$$
\begin{aligned}
&\int_{\theta_{0j}-\frac{M}{s_0+2}}^{\theta_{0j}+\frac{M}{s_0+2}} \exp\left\{-c_1 n(\theta_j - \theta_{0j})^2 - \frac{1}{2\tau^2}\theta_j^2\right\} d\theta_j \\
&= \int_{\theta_{0j}-\frac{M}{s_0+2}}^{\theta_{0j}+\frac{M}{s_0+2}} \exp\left\{-\frac{2c_1 n + \frac{1}{\tau^2}}{2}\left(\theta_j - \frac{2c_1 n}{2c_1 n + \frac{1}{\tau^2}}\theta_{0j}\right)^2 - \frac{c_1 n}{2c_1 n\tau^2 + 1}\theta_{0j}^2\right\} d\theta_j \\
&= \exp\left\{-\frac{c_1 n}{2c_1 n\tau^2 + 1}\theta_{0j}^2\right\} \sqrt{\frac{1}{2c_1 n + \frac{1}{\tau^2}}} \\
&\times \int_{\sqrt{\frac{1}{\tau^2}}\sqrt{\frac{1}{2c_1 n\tau^2+1}}\theta_{0j} - \frac{M}{s_0+2}\sqrt{2c_1 n + \frac{1}{\tau^2}}}^{\sqrt{\frac{1}{\tau^2}}\sqrt{\frac{1}{2c_1 n\tau^2+1}}\theta_{0j} + \frac{M}{s_0+2}\sqrt{2c_1 n + \frac{1}{\tau^2}}} \exp\left\{-\frac{1}{2}u_j^2\right\} du_j.
\end{aligned}
$$

As $n \to \infty$, the integral interval goes to $\mathbb{R}$, and thus

$$
\int_{\theta_{0j}-\frac{M}{s_0+2}}^{\theta_{0j}+\frac{M}{s_0+2}} \exp\left\{-c_1 n(\theta_j - \theta_{0j})^2 - \frac{1}{2\tau^2}\theta_j^2\right\} d\theta_j = \exp\left\{-\frac{c_1 n}{2c_1 n\tau^2 + 1}\theta_{0j}^2\right\} \sqrt{\frac{2\pi}{2c_1 n + \frac{1}{\tau^2}}}.
$$

Similar results hold for the integral of $\alpha_j$, and we have

$$
\begin{aligned}
&\int_{\Theta(M)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) \geq q^{s_0}(1-q)^{p-s_0} \exp\{-2c_2(s_0+2)\log p\} \\
&\times \left(\frac{1}{2c_1 n\tau^2 + 1}\right)^{\frac{s_0}{2}} \left(\frac{1}{2c_1 n\sigma_\alpha^2 + 1}\right) \\
&\times \exp\left\{-\frac{c_1 n}{2c_1 n\tau^2 + 1}\|\boldsymbol{\theta}_{0,-}\|_2^2\right\} \exp\left\{-\frac{c_1 n}{2c_1 n\sigma_\alpha^2 + 1}\|\boldsymbol{\alpha}_0\|_2^2\right\}.
\end{aligned}
$$

□

**Proof of Lemma S1.5**

By defining the event $\mathcal{A}(L) = \{\boldsymbol{\theta} \in \Theta(M) : |S| \geq Ls_0\}$, the left-hand side can be rewritten as

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi(\mathcal{A}(L) \mid \boldsymbol{Y}) 1_{\mathcal{E}_{n,1}(R)}\right] = \mathbb{E}_{\boldsymbol{\theta}_0}\left[\frac{\int_{\mathcal{A}(L)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta})}{\int_{\Theta(M)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta})} 1_{\mathcal{E}_{n,1}(R)}\right].$$

Using the results in Lemma S1.4, the right-hand side is bounded above by

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{A}(L)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) 1_{\mathcal{E}_{n,1}(R)}\right] q^{-s_0}(1-q)^{s_0-p}(2c_1 n\tau^2+1)^{\frac{s_0}{2}}(2c_1 n\sigma_\alpha^2+1)$$
$$\times \exp\left\{\frac{c_1 n}{2c_1 n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2 + \frac{c_1 n}{2c_1 n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2 + 2c_2(s_0+2)\log p\right\}.$$

Given the definition of the prior $\Pi$, the expectation term is

$$\sum_{S:|S|\geq Ls_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|}\frac{1}{2\pi\sigma_\alpha^2}$$
$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\Theta_S(M)} \frac{L_n(\boldsymbol{\theta}_S)}{L_n(\boldsymbol{\theta}_0)} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S 1_{\mathcal{E}_{n,1}(R)}\right].$$

We define $\boldsymbol{\theta}_{\tilde{S}} \in \mathbb{R}^{|S\cup S_0|+2}$ to contain $\boldsymbol{\theta}_S$ for $S$ and $\mathbf{0}$ for $S^c \cap S_0$ and the corresponding true parameter is given by $\boldsymbol{\theta}_{0\tilde{S}}$ with $\boldsymbol{\theta}_0$ for $S_0$ and $\mathbf{0}$ for $S_0^c \cap S$.

Then we can rewrite $L_n(\boldsymbol{\theta}_S)/L_n(\boldsymbol{\theta}_0) = \exp\{Z_n(\boldsymbol{\theta}_{\tilde{S}})\}$, and by Lemma S1.1,

$$
\begin{aligned}
&\int_{\Theta_S(M)} \exp\{Z_n(\boldsymbol{\theta}_{\tilde{S}})\} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S \\
&\leq \exp\{c_2(|\tilde{S}|+2)\log p\} \\
&\times \int_{\Theta_S(M)} \exp\{-c_3 n\|\boldsymbol{\theta}_{\tilde{S}} - \boldsymbol{\theta}_{0\tilde{S}}\|_2^2\} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S \\
&\leq \exp\{c_2(|S|+s_0+2)\log p\} \\
&\times \int_{\Theta_S} \exp\{-c_3 n(\|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 + \|\boldsymbol{\theta}_{0S^c}\|_2^2)\} \exp\left\{-\frac{\|\boldsymbol{\theta}_{S,-}\|_2^2}{2\tau^2} - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S \\
&= \exp\{c_2(|S|+s_0+2)\log p\} \exp\{-c_3 n\|\boldsymbol{\theta}_{0S^c}\|_2^2\} \frac{2\pi}{2c_3 n + \frac{1}{\sigma_\alpha^2}} \left(\frac{2\pi}{2c_3 n + \frac{1}{\tau^2}}\right)^{\frac{|S|}{2}} \\
&\times \exp\left\{-\frac{c_3 n}{2c_3 n\sigma_\alpha^2 + 1}\|\boldsymbol{\alpha}_0\|_2^2\right\} \exp\left\{-\frac{c_3 n}{2c_3 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S,-}\|_2^2\right\}.
\end{aligned}
$$

Since the above display does not include $\boldsymbol{Y}$, the expectation only involves the indicator variable $1_{\mathcal{E}_{n,1}(R)}$ with $\mathbb{E}_{\boldsymbol{\theta}_0}[1_{\mathcal{E}_{n,1}(R)}] \leq 1$. Thus we have

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{A}(L)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) 1_{\mathcal{E}_{n,1}(R)}\right] \\
&\leq \sum_{S:|S|\geq Ls_0} \exp\left\{-\frac{c_3 n}{2c_3 n\sigma_\alpha^2 + 1}\|\boldsymbol{\alpha}_0\|_2^2 - \frac{c_3 n}{2c_3 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S,-}\|_2^2 - c_3 n\|\boldsymbol{\theta}_{0S^c}\|_2^2\right\} \\
&\times \exp\{c_2(|S|+s_0+2)\log p\} q^{|S|}(1-q)^{p-|S|} \left(\frac{1}{2c_3 n\tau^2 + 1}\right)^{\frac{|S|}{2}} \frac{1}{2c_3 n\sigma_\alpha^2 + 1}.
\end{aligned}
$$

Then the posterior probability of our desired event is bounded by

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi(\mathcal{A}(L) \mid \boldsymbol{Y}) 1_{\mathcal{E}_{n,1}(R)}\right] \leq q^{-s_0}(1-q)^{s_0-p}(2c_1 n\tau^2+1)^{\frac{s_0}{2}}(2c_1 n\sigma_\alpha^2+1)\\
&\times \exp\left\{\frac{c_1 n}{2c_1 n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2+\frac{c_1 n}{2c_1 n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2+2c_2(s_0+2)\log p\right\}\\
&\times \sum_{S:|S|\geq Ls_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{2c_3 n\tau^2+1}\right)^{\frac{|S|}{2}}\frac{1}{2c_3 n\sigma_\alpha^2+1}\\
&\times \exp\left\{c_2(|S|+s_0+2)\log p-\frac{c_3 n}{2c_3 n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2-\frac{c_3 n}{2c_3 n\tau^2+1}\|\boldsymbol{\theta}_{0S,-}\|_2^2-c_3 n\|\boldsymbol{\theta}_{0S^c}\|_2^2\right\}\\
&\preceq q^{-s_0}(1-q)^{s_0}(2c_1 n\tau^2+1)^{\frac{s_0}{2}}\exp\left\{\frac{c_1 n}{2c_1 n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2+3c_2(s_0+2)\log p\right\}\\
&\times \sum_{S:|S|\geq Ls_0} q^{|S|}(1-q)^{-|S|}\left(\frac{1}{2c_3 n\tau^2+1}\right)^{\frac{|S|}{2}}\\
&\times \exp\left\{-\frac{c_3 n}{2c_3 n\tau^2+1}\|\boldsymbol{\theta}_{0S,-}\|_2^2-c_3 n\|\boldsymbol{\theta}_{0S^c}\|_2^2+c_2|S|\log p\right\}
\end{aligned}
$$

Since the model size is a non-negative integer, we define $k=\lfloor (L-1)s_0\rfloor$ where $(L-1)s_0-1<k\leq (L-1)s_0$. The probability is bounded above by

$$
\begin{aligned}
&q^{-s_0}(1-q)^{s_0}(2c_1 n\tau^2+1)^{\frac{s_0}{2}}\exp\left\{\frac{c_1 n}{2c_1 n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2+3c_2(s_0+2)\log p\right\}\\
&\times \sum_{S:|S|\geq s_0+k} q^{|S|}(1-q)^{-|S|}\left(\frac{1}{2c_3 n\tau^2+1}\right)^{\frac{|S|}{2}}\\
&\times \exp\left\{-\frac{c_3 n}{2c_3 n\tau^2+1}\|\boldsymbol{\theta}_{0S,-}\|_2^2-c_3 n\|\boldsymbol{\theta}_{0S^c}\|_2^2+c_2|S|\log p\right\},
\end{aligned}
$$

and, by listing all candidate models, the above is bounded by

$$
\begin{aligned}
& q^{-s_0}(1-q)^{s_0}(2c_1 n\tau^2+1)^{\frac{s_0}{2}} \exp\left\{3c_2(s_0+2)\log p\right\} \\
& \times \sum_{d=s_0+k} \binom{p}{d} q^d (1-q)^{-d} \left(\frac{1}{2c_3 n\tau^2+1}\right)^{\frac{d}{2}} \exp\left\{c_2 d \log p\right\} \\
& \preceq \exp\left\{3c_2(s_0+2)\log p + (c_2+1)s_0 \log p\right\} \\
& \sum_{d=s_0+k} p^{d-s_0} \left(\frac{q}{1-q}\right)^{d-s_0} \left(\frac{1}{2c_3 n\tau^2+1}\right)^{\frac{d-s_0}{2}} \exp\{c_2 \log p\}^{d-s_0}.
\end{aligned}
$$

By Condition 3,

$$
p\left(\frac{q}{1-q}\right)\sqrt{\frac{1}{2c_3 n\tau^2+1}}\exp\{c_2\log p\} \preceq p^{-1}.
$$

Thus the desired probability is bounded by

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi(\mathcal{A}(L)\mid \boldsymbol{Y})1_{\mathcal{E}_{n,1}(R)}\right] &\preceq \exp\left\{3c_2(s_0+2)\log p + (c_2+1)s_0\log p\right\} p^{-k} \\
&\leq \exp\left\{[4c_2(s_0+2)-(L-2)s_0]\log p\right\}.
\end{aligned}
$$

Then for $L \geq 5c_2(s_0+2)/s_0+2$, we have

$$
\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi(\mathcal{A}(L)\mid \boldsymbol{Y})1_{\mathcal{E}_{n,1}(R)}\right] \preceq \exp\left\{-\frac{1}{5}(L-2)s_0\log p\right\}.
$$

□

**Proof of Lemma S1.6**

We denote the desired event as $\mathcal{B}(\eta) = \{\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \eta\}$. By defining $\mathcal{B}(L, \eta) = \{\boldsymbol{\theta} \in \Theta(M) : |S| < Ls_0, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \eta\}$, we have

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \\
&\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq Ls_0 \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] + \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(L, \eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right].
\end{aligned}
$$

By Lemma S1.5, for some constant $L$, the first term on the right-hand side is bounded. For the second term, it can be written as

$$
\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(L, \eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] = \mathbb{E}_{\boldsymbol{\theta}_0}\left[\frac{\int_{\mathcal{B}(L,\eta)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta})}{\int_{\Theta(M)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta})} 1_{\mathcal{E}_{n,1}(R)}\right].
$$

Then by Lemma S1.4, the last display is bounded above by

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{B}(L,\eta)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) 1_{\mathcal{E}_{n,1}(R)}\right] q^{-s_0}(1-q)^{s_0-p}(2c_1 n\tau^2 + 1)^{\frac{s_0}{2}}(2c_1 n\sigma_\alpha^2 + 1) \\
&\times \exp\left\{\frac{c_1 n}{2c_1 n\tau^2 + 1}\|\boldsymbol{\theta}_{0,-}\|_2^2 + \frac{c_1 n}{2c_1 n\sigma_\alpha^2 + 1}\|\boldsymbol{\alpha}_0\|_2^2 + 2c_2(s_0 + 2)\log p\right\}.
\end{aligned}
$$

The expectation term can be calculated as

$$
\begin{aligned}
&\sum_{S:|S| \leq Ls_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|}\frac{1}{2\pi\sigma_\alpha^2} \\
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{B}_S(\eta)} \frac{L_n(\boldsymbol{\theta}_S)}{L_n(\boldsymbol{\theta}_0)} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S 1_{\mathcal{E}_{n,1}(R)}\right].
\end{aligned}
$$

Using a similar idea in the proof of Lemma S1.5, the integral term is bounded above by

$$
\begin{aligned}
&\exp\{c_2(|S| + s_0 + 2)\log p\} \\
&\times \int_{\mathcal{B}_S(\eta)} \exp\left\{-c_3 n\left(\|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 + \|\boldsymbol{\theta}_{0S^c}\|_2^2\right) - \frac{\|\boldsymbol{\theta}_{S,-}\|_2^2}{2\tau^2} - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S.
\end{aligned}
$$

On the set $\mathcal{B}_S(\eta)$, we have $\|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 + \|\boldsymbol{\theta}_{0S^c}\|_2^2 > \eta^2$. Thus

$$\begin{aligned}\|\boldsymbol{\theta}_S\|_2^2 &= \|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 + \|\boldsymbol{\theta}_{0S}\|_2^2 - 2(\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S})^T \boldsymbol{\theta}_{0S} \\ &\geq \eta^2 - \|\boldsymbol{\theta}_{0S^c}\|_2^2 + \|\boldsymbol{\theta}_{0S}\|_2^2 - 2M\|\boldsymbol{\theta}_{0S}\|_2 \geq \eta^2 - \|\boldsymbol{\theta}_{0,-}\|_2^2 - 2M\|\boldsymbol{\theta}_0\|_2.\end{aligned}$$

Then we have

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{B}(L,\eta)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) 1_{\mathcal{E}_{n,1}(R)}\right] \\ &\leq \sum_{S:|S|\leq Ls_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|}\frac{1}{2\pi\sigma_\alpha^2}\exp\{c_2(|S|+s_0+2)\log p\} \\ &\times \mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{B}_S(\eta)} \exp\left\{-c_3 n\eta^2 - \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} d\boldsymbol{\theta}_S 1_{\mathcal{E}_{n,1}(R)}\right] \\ &\preceq \sum_{S:|S|\leq Ls_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|}\frac{1}{2\pi\sigma_\alpha^2}\exp\{c_2(|S|+s_0+2)\log p\} \\ &\times \exp\left\{-\left(c_3 n + \frac{1}{2\tau^2}\right)\eta^2 + \frac{1}{2\tau^2}\left(\|\boldsymbol{\theta}_{0,-}\|_2^2 + 2M\|\boldsymbol{\theta}_0\|_2\right)\right\}.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(L,\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \preceq q^{-s_0}(1-q)^{s_0-p}(2c_1 n\tau^2+1)^{\frac{s_0}{2}}(2c_1 n\sigma_\alpha^2+1) \\ &\times \exp\left\{\frac{c_1 n}{2c_1 n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2 + \frac{c_1 n\|\boldsymbol{\alpha}_0\|_2^2}{2c_1 n\sigma_\alpha^2+1} + 2c_2(s_0+2)\log p\right\} \\ &\times \sum_{S:|S|\leq Ls_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|}\frac{1}{2\pi\sigma_\alpha^2}\exp\{c_2(|S|+s_0+2)\log p\} \\ &\times \exp\left\{-\left(c_3 n + \frac{1}{2\tau^2}\right)\eta^2 + \frac{1}{2\tau^2}\left(\|\boldsymbol{\theta}_{0,-}\|_2^2 + 2M\|\boldsymbol{\theta}_0\|_2\right)\right\} \\ &\preceq \left(\frac{2c_1 n\tau^2+1}{2\pi\tau^2}\right)^{\frac{s_0}{2}}\frac{2c_1 n\sigma_\alpha^2+1}{2\pi\sigma_\alpha^2}\exp\left\{[3c_2(s_0+2)+c_2 s_0]\log p\right\} \\ &\times \exp\left\{-\left(c_3 n + \frac{1}{2\tau^2}\right)\eta^2 + \frac{1}{\tau^2}\left(\|\boldsymbol{\theta}_0\|_2^2 + 2M\|\boldsymbol{\theta}_0\|_2\right)\right\} \\ &\times \sum_{S:|S|\leq Ls_0}\left(\frac{q}{1-q}\right)^{|S|-s_0}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|-s_0}\exp\{c_2(|S|-s_0)\log p\}.\end{aligned}$$

Denote $k = \lfloor Ls_0 \rfloor$. We can rewrite the sum as

$$\begin{aligned}
&\sum_{d=0}^{k} \binom{p}{d} \left(\frac{q}{1-q}\right)^{d-s_0} \left(\frac{1}{\sqrt{2\pi}\tau}\right)^{d-s_0} \exp\{c_2 \log p\}^{d-s_0} \\
&\leq p^{s_0} \sum_{d=0}^{k} p^{d-s_0} \left(\frac{q}{1-q}\right)^{d-s_0} \left(\frac{1}{\sqrt{2\pi}\tau}\right)^{d-s_0} \exp\{c_2 \log p\}^{d-s_0} \\
&= p^{s_0} \sum_{d=0}^{s_0} \left[p^{1+c_2} \left(\frac{q}{1-q}\right) \left(\frac{1}{\sqrt{2\pi}\tau}\right)\right]^{d-s_0} \\
&+ p^{s_0} \sum_{d=s_0+1}^{k} \left[p^{1+c_2} \left(\frac{q}{1-q}\right) \left(\frac{1}{\sqrt{2\pi}\tau}\right)\right]^{d-s_0}.
\end{aligned}$$

When $d \leq s_0$, by Condition 3,

$$\sum_{d=0}^{s_0} \left[p^{1+c_2} \left(\frac{q}{1-q}\right) \left(\frac{1}{\sqrt{2\pi}\tau}\right)\right]^{d-s_0} \preceq \sum_{d=0}^{s_0} \left(\sqrt{\frac{n}{(n \vee p^2)^{1+c_2}}} p^{c_2}\right)^{d-s_0} \preceq p^{s_0}.$$

When $d > s_0$, similarly, we have

$$\sum_{d=s_0+1}^{k} \left[p^{1+c_2} \left(\frac{q}{1-q}\right) \left(\frac{1}{\sqrt{2\pi}\tau}\right)\right]^{d-s_0} \preceq \sum_{d=0}^{s_0} \left(\sqrt{\frac{n}{(n \vee p^2)^{1+c_2}}} p^{c_2}\right)^{d-s_0} \preceq 1.$$

Combining the results,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0} \left[\Pi\left(\mathcal{B}(L,\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \preceq \left(\frac{c_1 n}{\pi}\right)^{\frac{s_0+2}{2}} \\
&\times \exp\left\{[3c_2(s_0+2) + (c_2+2)s_0] \log p\right\} \exp\left\{-\left(c_3 n + \frac{1}{2\tau^2}\right)\eta^2\right\}.
\end{aligned}$$

Let $\eta = M_1\sqrt{s_0 \log p_n}/\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|$. Since $\|\boldsymbol{X}\|^2 \leq \lambda_{\max}(\boldsymbol{X}^T\boldsymbol{X}) \leq n\lambda_2$ and, similarly, $\|\tilde{\boldsymbol{Z}}\|^2 \leq n\lambda_2$, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0} \left[\Pi\left(\mathcal{B}(L,\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \\
&\preceq \exp\left\{[4c_2(s_0+2) + (c_2+2)s_0] \log p\right\} \exp\left\{-\frac{c_3 M_1^2 s_0 \log p}{\lambda_2}\right\}.
\end{aligned}$$

For $M_1 \geq \sqrt{2\lambda_2(5c_2 + 2 + 8c_2/s_0)/c_3}$,

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(L,\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \preceq \exp\left\{-\frac{c_3 M_1^2 s_0 \log p}{2\lambda_2}\right\}.$$

□

**Proof of Lemma S1.7**

By the union bound, we have

$$P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_n^c\right] \leq P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_{n,1}(R)^c\right] + P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_{n,1}(R) \cap \mathcal{E}_{n,2}(L)^c\right] + P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_{n,1}(R) \cap \mathcal{E}_{n,3}(M_1, M_2)^c\right].$$

By Lemma S1.1, the first term is bounded as $P_{\boldsymbol{\theta}_0}[\mathcal{E}_{n,1}(R)^c] \to 0$. For the second term, by Markov's inequality and Lemma S1.5, we have

$$\begin{aligned}
P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_{n,1}(R) \cap \mathcal{E}_{n,2}(L)^c\right] &= P_{\boldsymbol{\theta}_0}\left[\Pi(\boldsymbol{\theta} \in \Theta(M) : |S| \leq Ls_0 \mid \boldsymbol{Y}) 1_{\mathcal{E}_{n,1}(R)} \leq 3/4\right] \\
&= P_{\boldsymbol{\theta}_0}\left[\Pi(\boldsymbol{\theta} \in \Theta(M) : |S| > Ls_0 \mid \boldsymbol{Y}) 1_{\mathcal{E}_{n,1}(R)} > 1/4\right] \\
&\leq 4\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi(\boldsymbol{\theta} \in \Theta(M) : |S| > Ls_0 \mid \boldsymbol{Y}) 1_{\mathcal{E}_{n,1}(R)}\right] \to 0.
\end{aligned}$$

Similarly, for the third term, for $M_2 \leq (cM_1^2/2\lambda_2) \wedge ((L-2)/5)$, by Lemma S1.6,

$$\begin{aligned}
&P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_{n,1}(R) \cap \mathcal{E}_{n,3}(M_1, M_2)^c\right] \\
&= P_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)} > \exp\{-M_2 s_0 \log p\}\right] \\
&\leq \exp\{M_2 s_0 \log p\} \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \frac{M_1\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \\
&\preceq \exp\{M_2 s_0 \log p\}\left[\exp\left\{-\frac{c_3 M_1^2 s_0 \log p}{2\lambda_2}\right\} + \exp\left\{-\frac{(L-2)s_0 \log p}{5}\right\}\right] \to 0.
\end{aligned}$$

□

**Proof of Lemma S1.8**

Similar to the proof of Lemma S1.6, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M): S\neq S_0 \mid \boldsymbol{Y}\right)1_{\mathcal{E}_{n,1}(R)}\right]\\
&\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M): |S|\geq Ls_0 \mid \boldsymbol{Y}\right)1_{\mathcal{E}_{n,1}(R)}\right]\\
&+ \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M): S\neq S_0, |S|< Ls_0 \mid \boldsymbol{Y}\right)1_{\mathcal{E}_{n,1}(R)}\right],
\end{aligned}$$

where the first term is bounded for some constant $L$. The posterior probability of a specific model $S'$ can be calculated as

$$\begin{aligned}
&\Pi\left(\boldsymbol{\theta}\in\Theta(M): S=S' \mid \boldsymbol{Y}\right) = C_n q^{|S'|}(1-q)^{|S'|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S'|}\left(\frac{1}{2\pi\sigma_\alpha^2}\right)\\
&\times\int_{\Theta_{S'}(M)} L_n(\boldsymbol{\theta}_{S'})\exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S',-}\|_2^2\right\}\exp\left\{-\frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\}d\boldsymbol{\theta}_{S'}\\
&= C_n q^{|S'|}(1-q)^{|S'|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S'|}\left(\frac{1}{2\pi\sigma_\alpha^2}\right)L_n(\boldsymbol{\theta}_{0S'})\\
&\times\int_{\Theta_{S'}(M)}\exp\{Z_n(\boldsymbol{\theta}_{S'})\}\exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S',-}\|_2^2\right\}\exp\left\{-\frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\}d\boldsymbol{\theta}_{S'},
\end{aligned}$$

where $C_n$ is the normalizing constant. We are going to construct both upper and lower bounds for $\Pi(\boldsymbol{\theta}\in\Theta(M): S=S' \mid \boldsymbol{Y})$ on $\mathcal{E}_{n,1}(R)$.

For the upper bound, by the calculation in the proof of Lemma S1.5, the integral term is bounded above by

$$\begin{aligned}
&\exp\{c_2(|S'|+2)\log p\}\frac{2\pi}{2c_3n+\frac{1}{\sigma_\alpha^2}}\left(\frac{2\pi}{2c_3n+\frac{1}{\tau^2}}\right)^{\frac{|S'|}{2}}\\
&\times\exp\left\{-\frac{c_3n}{2c_3n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2-\frac{c_3n}{2c_3n\tau^2+1}\|\boldsymbol{\theta}_{0S,-}\|_2^2\right\}.
\end{aligned}$$

Thus the posterior probability is bounded above by

$$
\begin{aligned}
&\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y}) \\
&\leq C_n q^{|S'|}(1-q)^{|S'|} L_n(\boldsymbol{\theta}_{0S'}) \exp\{c_2(|S'|+2)\log p\} \frac{1}{2c_3 n\sigma_\alpha^2 + 1} \\
&\times \left(\frac{1}{2c_3 n\tau^2 + 1}\right)^{\frac{|S'|}{2}} \exp\left\{-\frac{c_3 n}{2c_3 n\sigma_\alpha^2 + 1}\|\boldsymbol{\alpha}_0\|_2^2 - \frac{c_3 n}{2c_3 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S',-}\|_2^2\right\}.
\end{aligned}
$$

Similarly, for the lower bound, by the calculation in the proof of Lemma S1.4, the posterior probability is bounded below by

$$
\begin{aligned}
&\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y}) \\
&\geq C_n q^{|S'|}(1-q)^{|S'|} L_n(\boldsymbol{\theta}_{0S'}) \exp\{-2c_2(|S'|+2)\log p\} \frac{1}{2c_1 n\sigma_\alpha^2 + 1} \\
&\times \left(\frac{1}{2c_1 n\tau^2 + 1}\right)^{\frac{|S'|}{2}} \exp\left\{-\frac{c_1 n}{2c_1 n\sigma_\alpha^2 + 1}\|\boldsymbol{\alpha}_0\|_2^2 - \frac{c_1 n}{2c_1 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S',-}\|_2^2\right\}.
\end{aligned}
$$

Then, by Condition 4, the ratio can be calculated as

$$
\begin{aligned}
&\frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \\
&\preceq \left(\frac{q}{1-q}\right)^{|S'|-s_0} \left(\frac{1}{2c_3 n\tau^2 + 1}\right)^{\frac{|S'|}{2}} \left(\frac{1}{2c_1 n\tau^2 + 1}\right)^{-\frac{s_0}{2}} \\
&\times \frac{L_n(\boldsymbol{\theta}_{0S'})}{L_n(\boldsymbol{\theta}_{0S_0})} \exp\{c_2(|S'| + 2s_0 + 6)\log p\} \\
&\times \exp\left\{-\frac{c_3 n}{2c_3 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S',-}\|_2^2 + \frac{c_1 n}{2c_1 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S_0,-}\|_2^2\right\} \\
&\preceq \exp\{[3c_2 s_0 + (2 - c_2 + c_2\kappa_n s_0)(s_0 - |S'|) + 6c_2]\log p\} \\
&\times \frac{L_n(\boldsymbol{\theta}_{0S'})}{L_n(\boldsymbol{\theta}_{0S_0})} \exp\left\{-\frac{c_3 n}{2c_3 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S',-}\|_2^2 + \frac{c_1 n}{2c_1 n\tau^2 + 1}\|\boldsymbol{\theta}_{0S_0,-}\|_2^2\right\}.
\end{aligned}
$$

We split the model space into the following possibly overlapping sets:

Over-fitted models: $\mathcal{S}_1 = \{S : S \supset S_0, S \neq S_0, |S| < Ls_0\}$,

Large models: $\mathcal{S}_2 = \{S : s_0 < |S| < Ls_0\}$,

Under-fitted models: $\mathcal{S}_3 = \{S : S \not\supset S_0, |S| \leq s_0\}$.

**Over-fitted models:** if $S' \in \mathcal{S}_1$, then $L_n(\boldsymbol{\theta}_{0S'}) = L_n(\boldsymbol{\theta}_{0S_0})$. Let $k = \lfloor Ls_0 \rfloor$. Then

$$\frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})}$$
$$\preceq \exp\{[3c_2 s_0 + (2 - c_2 + c_2\kappa_n s_0)(s_0 - |S'|) + 6c_2] \log p\}.$$

The sum of probability ratios in $\mathcal{S}_1$ is

$$\sum_{S' \in \mathcal{S}_1} \frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})}$$
$$\preceq \sum_{d=s_0+1}^{k} \binom{p - s_0}{d - s_0} \exp\{[3c_2 s_0 + (2 - c_2 + c_2\kappa_n s_0)(s_0 - d) + 6c_2] \log p\}$$
$$\preceq p^{3c_2 s_0 + 6c_2} \sum_{d=s_0+1}^{k} \left(p^{1 - c_2 + c_2\kappa_n s_0}\right)^{(s_0 - d)}$$
$$\preceq p^{3c_2 s_0 + 7c_2 - 1 - c_2\kappa_n s_0}.$$

If $\kappa_n > 6 + 2(7 - 1/c_2)/s_0$, we have

$$\sum_{S' \in \mathcal{S}_1} \frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \preceq \exp\left\{-\frac{c_2}{2}\kappa_n s_0 \log p\right\}.$$

**Large models:** for $S' \in \mathcal{S}_2$, we define $\tilde{S} = S' \cup S_0$ and define $\boldsymbol{\theta}_{\tilde{S}}$ to include $\boldsymbol{\theta}_{0S'}$ for $S'$ and $\mathbf{0}$ for $(S')^c \cap S_0$ and $\boldsymbol{\theta}_{0\tilde{S}}$ to include $\boldsymbol{\theta}_{0S_0}$ for $S_0$ and $\mathbf{0}$ for

$S' \cap S_0^c$. Then

$$
\begin{aligned}
&\frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \\
&\preceq \exp\{[3c_2 s_0 + (2 - c_2 + c_2 \kappa_n s_0)(s_0 - |S'|) + 6c_2] \log p\} \\
&\times \frac{L_n(\boldsymbol{\theta}_{\tilde{S}})}{L_n(\boldsymbol{\theta}_{0\tilde{S}})} \exp\left\{ -\frac{c_3 n}{2c_3 n\tau^2 + 1} \|\boldsymbol{\theta}_{0S',-}\|_2^2 + \frac{c_1 n}{2c_1 n\tau^2 + 1} \|\boldsymbol{\theta}_{0S_0,-}\|_2^2 \right\} \\
&\preceq \exp\{[5c_2 s_0 + (2 - 2c_2 + c_2 \kappa_n s_0)(s_0 - |S'|) + 8c_2] \log p\} \exp\left\{ -c_3 n \|\boldsymbol{\theta}_{0(S')^c}\|_2^2 \right\} \\
&\preceq \exp\{[5c_2 s_0 + (2 - 2c_2 + c_2 \kappa_n s_0)(s_0 - |S'|) + 8c_2] \log p\} \exp\left\{ -c_3 \kappa_n^2 s_0 \log p \right\}.
\end{aligned}
$$

The sum of probability ratios in $\mathcal{S}_2$ is

$$
\begin{aligned}
&\sum_{S' \in \mathcal{S}_2} \frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \\
&\preceq \exp\left\{ -c_3 \kappa_n^2 s_0 \log p \right\} \exp\{(5c_2 s_0 + 8c_2) \log p\} \\
&\times \sum_{d=s_0+1}^{k} \sum_{h=0}^{s_0-1} \binom{p - s_0}{d - h} s_0 \binom{s_0 - 1}{h} \exp\{(2 - 2c_2 + c_2 \kappa_n s_0)(s_0 - d) \log p\} \\
&\preceq p^{-c_3 \kappa_n^2 s_0} p^{5c_2 s_0 + 8c_2 + s_0} \sum_{d=s_0+1}^{k} \left( p^{1 - 2c_2 + c_2 \kappa_n s_0} \right)^{(s_0 - d)} \\
&\preceq p^{-c_3 \kappa_n^2 s_0 + 5c_2 s_0 + 10c_2 + s_0 - 1 - c_2 \kappa_n s_0}.
\end{aligned}
$$

If $\kappa_n \geq 2c_2/c_3$, then $c_3 \kappa_n^2 s_0 + c_2 \kappa_n s_0 \geq 3c_2 \kappa_n s_0$. Further if $\kappa_n \geq 2((5c_2 + 1)s_0 + 10c_2 - 1)/3c_2 s_0$, then

$$
\sum_{S' \in \mathcal{S}_2} \frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \preceq \exp\left\{ -\frac{3c_2}{2} \kappa_n s_0 \log p \right\}.
$$

**Under-fitted models:** for $S' \in \mathcal{S}_3$, similarly, we have

$$\frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \preceq \exp\left\{-c_3 \kappa_n^2 s_0 \log p\right\}$$

$$\times \exp\{[5c_2 s_0 + (2 - 2c_2 + c_2 \kappa_n s_0)(s_0 - |S'|) + 8c_2] \log p\},$$

and the sum of probability ratios in $\mathcal{S}_3$ is

$$\sum_{S' \in \mathcal{S}_3} \frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})}$$

$$\preceq \exp\left\{-c_3 \kappa_n^2 s_0 \log p\right\} \exp\{(5c_2 s_0 + 8c_2) \log p\}$$

$$\times \sum_{d=0}^{s_0} \sum_{h=0}^{d} \binom{p - s_0}{d - h} s_0 \binom{s_0 - 1}{h} \exp\{(2 - 2c_2 + c_2 \kappa_n s_0)(s_0 - d) \log p\}$$

$$\preceq p^{-c_3 \kappa_n^2 s_0} p^{(3c_2 + c_2 \kappa_n s_0 + 2)s_0 + 8c_2}.$$

For $\kappa_n \geq 2c_2 s_0 / c_3$, we have $c_3 \kappa_n^2 s_0 - c_2 \kappa_n s_0^2 \geq c_2 \kappa_n s_0^2$. If $\kappa_n \geq 2((3c_2 + 2)s_0 + 8c_2)/c_2 s_0^2$, then

$$\sum_{S' \in \mathcal{S}_3} \frac{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y})}{\Pi(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y})} \preceq \exp\left\{-\frac{c_2}{2} \kappa_n s_0^2 \log p\right\}.$$

Combining the results leads to, on $\mathcal{E}_{n,1}(R)$,

$$\sum_{S' \neq S_0, |S'| < L s_0} \frac{\Pi\left(\boldsymbol{\theta} \in \Theta(M) : S = S' \mid \boldsymbol{Y}\right)}{\Pi\left(\boldsymbol{\theta} \in \Theta(M) : S = S_0 \mid \boldsymbol{Y}\right)} \preceq 3 \exp\left\{-\frac{c_2}{2} \kappa_n s_0 \log p\right\},$$

which implies that

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : S \neq S_0, |S'| < L s_0 \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R)}\right] \preceq 3 \exp\left\{-\frac{c_2}{2} \kappa_n s_0 \log p\right\}.$$

$\square$

## S2 Theoretical Results Under Unknown Variance Case

In this section, we elaborate on the discussion in Section 3.3, extending the theoretical results in Section 3.2 to the case of an unknown noise variance. The parameter studied here is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \sigma_y) \in \mathbb{R}^{p+3}$ with the true value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0, \sigma_{y0})$.The Gaussian components in the priors of $\beta_j$ and $\alpha_k$ are adjusted as $N(0, \tau_{\boldsymbol{\beta}}^2)$ and $N(0, \sigma_\alpha^2)$ to avoid coupling, respectively, while the other priors remain unchanged.

Compared with the known-variance setting in Section 3.2, we now include an inverse gamma prior $\pi(\sigma_y^2)$ in the joint prior and its variational counterpart $q(\sigma_y^2)$ in the variational family. Similar to the case of a known variance, we assume bounded covariate spaces $\mathcal{Z}$ and $\mathcal{X}$. The parameter space is defined as

$$\Theta(M) := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_1 \le M, |\log \sigma_y| \le M\},$$

following a common practice in the analysis of mixture of regressions with an unknown $\sigma_y^2$ (Städler et al., 2010; Zhang et al., 2025). The condition $|\log \sigma_y| < M$ ensures that $0 < \underline{\sigma} \le \sigma_{y0} \le \bar{\sigma} < \infty$ for some constants $\underline{\sigma}$ and $\bar{\sigma}$, and thus no modification of the beta-min condition is required. Under the same regularity conditions specified in Section 3.2, the variational posterior $Q^*$ satisfies the following posterior contraction properties on model selection and parameter estimation.

**Lemma S2.1.** Under Conditions 1-3, there exists a constant $L_0' > 2$ such that, for any sequence $L_n \geq L_0'$, as $n \to \infty$, the VB posterior $Q^*$ satisfies

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L_n s_0)\right] \leq \mathcal{O}\left(\frac{C_L'}{L_n}\right) + o(1),$$

with some constant $C_L' > 0$.

**Theorem S2.1.** Under the conditions in Lemma S2.1, there exists some constant $M_0' > 0$ such that, for any sequence $M_n \geq M_0'$ growing more slowly than $L_n$ in Lemma S2.1, as $n \to \infty$, the VB posterior $Q^*$ satisfies

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \geq \frac{\sqrt{M_n s_0 \log p}}{\|\boldsymbol{X}\| \vee \|\tilde{\boldsymbol{Z}}\|}\right)\right] \leq \mathcal{O}\left(\frac{C_M'}{M_n}\right) + o(1),$$

with some constant $C_M' > 0$.

**Theorem S2.2.** Under Conditions 1-4, for any $\kappa_n$ growing more slowly than $L_n$ defined in Lemma S2.1, as $n \to \infty$, the VB posterior $Q^*$ satisfies

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[Q^*\left(\boldsymbol{\theta} \in \Theta(M) : S \neq S_0\right)\right] \leq \mathcal{O}\left(\frac{C_\kappa'}{\kappa_n}\right) + o(1),$$

with some constant $C_\kappa' > 0$.

These results confirm the robustness of our theoretical framework in the presence of an unknown noise variance $\sigma_y^2$.

## S2.1 Proof of Main Results

Similar to the proof in Section S1, we first derive the upper and lower bounds on the log-likelihood. However, the inclusion of $\sigma_y$ poses a challenge in the analysis of high-dimensional mixture of regressions, since existing results are typically established on a transformed parameter vector $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}/\sigma_y, \boldsymbol{\alpha}/\sigma_y, 1/\sigma_y, \boldsymbol{\gamma})$ (Städler et al., 2010; Zhang et al., 2025). To eliminate the discrepancy between the parameters studied in the exact posterior and those defined in the variational posterior, we need to refine the results to focus directly on the original parameter $\boldsymbol{\theta}$.

**Lemma S2.2.** Under Condition 1, for $\boldsymbol{\theta} \in \Theta(M)$ there exists some constant $c'_R > 0$, such that for any constant $R \geq c'_R$, we have as $n \to \infty$,

$$P(V_n \leq R\lambda_0) \to 1.$$

Further under Conditions 1 and 2, for some constants $c'_1, c'_2, c'_3 > 0$, it holds that on $\{V_n \leq R\lambda_0\}$,

$$Z_n(\boldsymbol{\theta}_S) \geq -c'_1 n \|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 - 2c'_2(|S| + 3)\log p,$$

$$Z_n(\boldsymbol{\theta}_S) \leq -c'_3 n \|\boldsymbol{\theta}_S - \boldsymbol{\theta}_{0S}\|_2^2 + c'_2(|S| + 3)\log p.$$

The variational posterior contraction properties are established by leveraging Lemma S1.2. The events $\mathcal{E}_{n,1}$, $\mathcal{E}_{n,2}$, and $\mathcal{E}_{n,3}$ are defined in a similar way as in Eq.(S1.3) with some positive constants $R'$, $L'$, $M'_1$, and $M'_2$. We further define $\mathcal{E}_n = \mathcal{E}_{n,1}(R') \cap \mathcal{E}_{n,2}(L') \cap \mathcal{E}_{n,3}(M_1', M_2')$. We first derive the bound on the KL divergence between the variational and exact posteriors.

**Lemma S2.3.** Under Conditions 1-3, for sufficiently large $p$ and some constant $C_K' > 0$, we have

$$\mathrm{KL}\left[Q^*(\boldsymbol{\theta}) \| \Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})\right] 1_{\mathcal{E}_n} \leq C_K' s_0 \log p. \tag{S2.8}$$

Then we are left to derive the exact posterior contraction rates and show that the event $\mathcal{E}_n$ holds with probability going to 1.

**Lemma S2.4.** Under Conditions 1-3, on $\mathcal{E}_{n,1}(R')$, we have

$$\begin{aligned}
&\int_{\Theta(M)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) \succeq q^{s_0}(1-q)^{p-s_0} \exp\{-2c_2'(s_0+3)\log p\} \\
&\times \left(\frac{1}{2c_1' n\tau^2+1}\right)^{\frac{s_0}{2}} \left(\frac{1}{2c_1' n\sigma_\alpha^2+1}\right) \sqrt{\frac{1}{c_1' n}} \\
&\times \exp\left\{-\frac{c_1' n}{2c_1' n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2\right\} \exp\left\{-\frac{c_1' n}{2c_1' n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2\right\}.
\end{aligned}$$

**Lemma S2.5.** Under Conditions 1-3, for $L' \geq 2 + 5c_2'(s_0+3)/s_0$,

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L's_0 \mid Y\right) 1_{\mathcal{E}_{n,1}(R)}\right] \preceq \exp\left\{-\frac{1}{5}(L'-2)s_0 \log p\right\}.$$

**Lemma S2.6.** Under Conditions 1-3, for $M_1' \geq \sqrt{2\lambda_2(5c_2' + 2 + 12c_2'/s_0)/c_3'}$ and $L'$ satisfying the condition in Lemma S1.5,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \frac{M_1'\sqrt{s_0 \log p}}{\|\boldsymbol{X}\| \wedge \|\tilde{\boldsymbol{Z}}\|} \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R')}\right] \\
&\preceq \exp\left\{-\frac{c_3' M_1'^2 s_0 \log p}{2\lambda_2}\right\} + \exp\left\{-\frac{(L'-2)s_0 \log p}{5}\right\}.
\end{aligned}$$

**Lemma S2.7.** Under Conditions 1-3, for $R'$ defined in Lemma S1.1, $L'$ defined in Lemma S1.5, $M_1'$ defined in Lemma S1.6, and $M_2' \leq (cM_1'^2/2\lambda_2) \wedge ((L'-2)/5)$, as $p \to \infty$, we have

$$P_{\boldsymbol{\theta}_0}\left[\mathcal{E}_n^c\right] \to 0.$$

**Lemma S2.8.** Under Conditions 1-4, for $\kappa_n \geq (2c_2's_0/c_3') \vee (2((3c_2'+2)s_0 + 11c_2')/c_2's_0)$ and $L'$ defined in Lemma S1.5, we have

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : S \neq S_0 \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R')}\right] \\
&\preceq 3\exp\left\{-\frac{c_2'\kappa_n s_0 \log p}{2}\right\} + \exp\left\{-\frac{(L'-2)s_0 \log p}{5}\right\}.
\end{aligned}$$

With the preliminary lemmas, we are ready to prove the main theoretical results under the unknown $\sigma_y^2$ setting. The proofs of the technical lemmas are deferred to Section S2.2.

**Proof of Lemma S2.1**

By choosing $L_n \geq L_0' := 2 + 5c_2'(s_0+3)/s_0$ in Lemma S2.5, we follow the proof of Lemma 1 to obtain the result.

**Proof of Theorem S2.1**

By choosing $M_n \geq M_0' := 2\lambda_2(5c_2' + 2 + 12c_2'/s_0)/c_3'$ in Lemma S2.6, we follow the proof of Lemma 1 to obtain the result.

**Proof of Theorem S2.2**

By choosing $\kappa_n \geq \kappa_0 := (2c_2' s_0/c_3')\vee(2((3c_2'+2)s_0+11c_2')/c_2' s_0)$ in Lemma S2.8, we follow the proof of Lemma 2 to obtain the result.

## S2.2 Proof of Technical Lemmas

**Proof of Lemma S2.2**

This lemma is modified from Lemma 3.1 and Lemma 3.2 in Zhang et al. (2025), but our proof avoids the auxiliary reparameterization of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\sigma_y$, leading to a more direct argument under the original model formulation.

Following the proof in Zhang et al. (2025), for a model $S$, we define $\boldsymbol{\theta}_{\bar{S}} \in \mathbb{R}^{p+3}$ with $\boldsymbol{\theta}_S$ for $S$ and $\mathbf{0}$ for $S^c$. Then the parameters in the structured mixture model can be represented as

$$\begin{aligned}\boldsymbol{\psi}(\boldsymbol{x}, \boldsymbol{z}, t, S) &= \left(\frac{\exp\{\boldsymbol{x}^T\boldsymbol{\gamma}_{\bar{S}}\}}{1+\exp\{\boldsymbol{x}^T\boldsymbol{\gamma}_{\bar{S}}\}}, \boldsymbol{z}^T\boldsymbol{\beta}_{\bar{S}} + t\alpha_1, \boldsymbol{z}^T\boldsymbol{\beta}_{\bar{S}} + t\alpha_2, \sigma_y\right)\\ &:= (\psi_1, \psi_2, \psi_3, \psi_4),\end{aligned}$$

with a fixed dimension of 4 independent of $p$. For conciseness, we omit $(\boldsymbol{x}, \boldsymbol{z}, t)$ in the notation $\boldsymbol{\psi}(\boldsymbol{x}, \boldsymbol{z}, t, S)$ in the following. We denote the density of $Y$ as $f_{\boldsymbol{\psi}(S)}(Y)$, the log-likelihood as $\ell_{\boldsymbol{\psi}(S)}(Y) = \log f_{\boldsymbol{\psi}(S)}(Y)$, and the score function as $\boldsymbol{s}_{\boldsymbol{\psi}(S)}(Y) = \partial\ell_{\boldsymbol{\psi}(S)}(Y)/\partial\boldsymbol{\psi}(S)$. By direct calculation, we can verify that, under the case of the original parameterization, there also exists a function $G(\cdot)$ for any $S$ such that

$$\sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{z}\in\mathcal{Z},\boldsymbol{\theta}_S\in\Theta(M)} \|\boldsymbol{s}_{\boldsymbol{\psi}(S)}(Y)\|_\infty \leq G(Y) := C(Y^2 + |Y| + 1),$$

where the finite constant $C$ only depends on $\mathcal{X}$, $\mathcal{Z}$, and $M$. We can then adopt the same peeling device as Eq.(8) in Zhang et al. (2025) to prove the first part in Lemma S2.2, i.e., for any constant $R \geq c_R'$, as $n \to \infty$,

$$P(V_n \leq R\lambda_0) \to 1.$$

The second part follows similar calculations in the proof of Lemma 3.2 in Zhang et al. (2025). □

**Proof of Lemma S2.3**

The proof follows a similar strategy to that of Lemma S1.3, except for the non-Gaussian factors induced from the inverse gamma distributions $\pi(\sigma_y^2)$ and $q(\sigma_y^2)$. The surrogate variational distribution $\check{Q} \in \mathcal{Q}$ is defined as,

$$\begin{aligned}
&\check{Q}(\boldsymbol{\theta}) = N_S(\boldsymbol{\theta}_{S,-\sigma_y}; \boldsymbol{\mu}_S, \boldsymbol{D}_S) \otimes \delta_0(\boldsymbol{\theta}_{S^c}) \otimes \mathrm{IG}(\sigma_y^2; a_y, b_y) \\
&= \prod_{j=1}^{2} N(\alpha_j; \mu_{\alpha_j}, \sigma_{\alpha_j}^2) \otimes \prod_{j\in S} N(\theta_j; \mu_{Sj}, \sigma_{Sj}^2) \otimes \prod_{j\in S^c} \delta_0(\theta_j) \otimes \mathrm{IG}(\sigma_y^2; a_y, b_y),
\end{aligned}$$

where $\boldsymbol{\theta}_{S,-\sigma_y} = (\boldsymbol{\beta}_S, \boldsymbol{\gamma}_S, \boldsymbol{\alpha})$. We set

$$\boldsymbol{\mu}_S = \boldsymbol{\theta}_{0S,-\sigma_y} - (1+\tau^{-2})\Sigma_S \boldsymbol{\theta}_{0S,-\sigma_y},$$

$$\Sigma_S^{-1} = (2c_3' n + \tau^{-2}) I_S + \begin{pmatrix} \tilde{\boldsymbol{Z}}_S^T \tilde{\boldsymbol{Z}}_S & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & n\lambda_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{X}_S^T \boldsymbol{X}_S \end{pmatrix},$$

with $\boldsymbol{D}_S = \mathrm{diag}(\Sigma_S)$ being the diagonal matrix, and

$$a_y = a_0, \quad b_y^{\frac{1}{2}} = (a_y - 1)\left[\sigma_{y0} r(a_y) + \sigma_{y0}\sqrt{r(a_y)^2 - 1/(a_y-1)}\right],$$

where $r(a_y) = \Gamma(a_y - 1/2)/\Gamma(a_y)$. Then we have

$$\begin{aligned} &\mathrm{KL}[\check{Q}(\boldsymbol{\theta}) \| \Pi(\boldsymbol{\theta} \mid \boldsymbol{Y})] \\ &= \mathbb{E}_{\check{Q}}\left[\log \frac{dN_S(\boldsymbol{\theta}_{S,-\sigma_y}; \boldsymbol{\mu}_S, \boldsymbol{D}_S) \otimes \delta_0(\boldsymbol{\theta}_{S^c}) \otimes d\,\mathrm{IG}(\sigma_y^2; a_y, b_y)}{\hat{w}_S d\Pi_S(\boldsymbol{\theta}_S \mid \boldsymbol{Y}) \otimes \delta_0(\boldsymbol{\theta}_{S^c})}\right] \\ &= \log \frac{1}{\hat{w}_S} + \mathrm{KL}[N_S(\boldsymbol{\theta}_{S,-\sigma_y}; \boldsymbol{\mu}_S, \boldsymbol{D}_S) \otimes \mathrm{IG}(\sigma_y^2; a_y, b_y) \| \Pi_S(\boldsymbol{\theta}_S \mid \boldsymbol{Y})]. \end{aligned} \tag{S2.9}$$

Denote $\check{Q}_{\check{S}} = N_{\check{S}}(\boldsymbol{\theta}_{\check{S},-\sigma_y}; \boldsymbol{\mu}_{\check{S}}, \boldsymbol{D}_{\check{S}})$ and the second KL term in (S2.9) can be rewritten as

$$\mathrm{KL}[\check{Q}_{\check{S}} \| N_{\check{S}}(\boldsymbol{\theta}_{\check{S},-\sigma_y}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}})] + \mathbb{E}_{\check{Q}_{\check{S}}}\left[\log \frac{dN_{\check{S}}(\boldsymbol{\theta}_{\check{S},-\sigma_y}; \boldsymbol{\mu}_{\check{S}}, \Sigma_{\check{S}}) \otimes d\,\mathrm{IG}(\sigma_y^2; a_y, b_y)}{d\Pi_{\check{S}}(\boldsymbol{\theta}_{\check{S}} \mid \boldsymbol{Y})}\right]. \tag{S2.10}$$

The first term in (S2.10) remains the same as the proof of Lemma S1.6 and is bounded by $\mathcal{O}(L's_0)$. For the second term, we have

$$
\begin{aligned}
&\mathbb{E}_{\check{Q}_{\check{S}}}\left[\log\frac{dN_{\check{S}}(\boldsymbol{\theta}_{\check{S},-\sigma_y};\boldsymbol{\mu}_{\check{S}},\Sigma_{\check{S}})\otimes d\,\mathrm{IG}(\sigma_y^2;a_y,b_y)}{d\Pi_{\check{S}}(\boldsymbol{\theta}_{\check{S}}\mid \boldsymbol{Y})}\right]\\
&=\log\frac{D_\Pi}{D_N}+\mathbb{E}_{\check{Q}_{\check{S}}}\left[-\frac{1}{2}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})\right.\\
&-(\boldsymbol{\theta}_{0\check{S},-\sigma_y}-\boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})-Z_n(\boldsymbol{\theta}_{\check{S}})+\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-}\|_2^2\\
&\left.+\frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-})+\frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2-2(a_y-a_0)\log\sigma_y-\frac{b_y-b_0}{\sigma_y^2}\right],
\end{aligned}
\tag{S2.11}
$$

where

$$
\begin{aligned}
D_N&=\int_{\Theta_{\check{S}}(M)}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})\right.\\
&\qquad\left.-(\boldsymbol{\theta}_{0\check{S},-\sigma_y}-\boldsymbol{\mu}_{\check{S}})^T\Sigma_S^{-1}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})-2(a_y+1)\log\sigma_y-\frac{b_y}{\sigma_y^2}\right\}d\boldsymbol{\theta}_{\check{S}},\\
D_\Pi&=\int_{\Theta_{\check{S}}(M)}\exp\left\{Z_n(\boldsymbol{\theta}_{\check{S}})-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-}\|_2^2-\frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-})\right.\\
&\qquad\left.-\frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2-2(a_0+1)\log\sigma_y-\frac{b_0}{\sigma_y^2}\right\}d\boldsymbol{\theta}_{\check{S}}.
\end{aligned}
$$

To bound the ratio of the normalizing constants, define the subspace

$$
B_{\check{S}}=\left\{\boldsymbol{\theta}_{\check{S}}\in\Theta_{\check{S}}(M):\|\boldsymbol{\theta}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}\|_2\le\frac{2M_1'\sqrt{s_0\log p}}{\|\boldsymbol{X}\|\wedge\|\tilde{\boldsymbol{Z}}\|}\right\}.
$$

Following a similar practice, on $\mathcal{E}_n$, $\log(D_\Pi/D_N)$ is bounded by

$$\log 2\int_{B_{\check S}} \exp\left\{Z_n(\boldsymbol\theta_{\check S}) - \frac{1}{2\tau^2}\|\boldsymbol\theta_{\check S,-} - \boldsymbol\theta_{0\check S,-}\|_2^2 - \frac{1}{\tau^2}\boldsymbol\theta_{0\check S,-}^T(\boldsymbol\theta_{\check S,-} - \boldsymbol\theta_{0\check S,-}) - \frac{1}{2\sigma_\alpha^2}\|\boldsymbol\alpha\|_2^2\right.$$
$$\left. -2(a_0+1)\log\sigma_y - \frac{b_0}{\sigma_y^2}\right\} d\boldsymbol\theta_{\check S} - \log\int_{\Theta_{\check S}(M)} \exp\left\{-\frac{1}{2}(\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y})^T\Sigma_{\check S}^{-1}(\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y})\right.$$
$$\left. -(\boldsymbol\theta_{0\check S,-\sigma_y} - \boldsymbol\mu_{\check S})^T\Sigma_{\check S}^{-1}(\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y}) - 2(a_y+1)\log\sigma_y - \frac{b_y}{\sigma_y^2}\right\} d\boldsymbol\theta_{\check S}$$
$$\le c_2(Ls_0+3)\log p - \frac{1}{2\tau^2}\|\boldsymbol\alpha_0\|_2^2 + \log 2$$
$$+ \sup_{\boldsymbol\theta_{\check S}\in B_{\check S}}\left\{-\frac{1}{2}(\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y})^T\left[(2c_3n+\tau^{-2})I_{\check S} - \Sigma_{\check S}^{-1}\right](\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y}) - c_3n(\sigma_y-\sigma_{y0})^2\right.$$
$$\left. + \left[(\boldsymbol\theta_{0\check S,-\sigma_y} - \boldsymbol\mu_{\check S})^T\Sigma_{\check S}^{-1} - \frac{1}{\tau^2}\boldsymbol\theta_{0\check S,-\sigma_y}^T\right](\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y}) - 2(a_0-a_y)\log\sigma_y - \frac{b_0-b_y}{\sigma_y^2}\right\}.$$

By the definitions of $\boldsymbol\mu_{\check S}$, $\Sigma_{\check S}$, $a_y$, and $b_y$, if we define $\tilde{\boldsymbol\beta}_{\check S} = (\boldsymbol\beta_{\check S}^T, \alpha_1)^T$, the sup term is bounded by

$$\sup_{\boldsymbol\theta_{\check S}\in B_{\check S}}\left\{\frac{1}{2}\|\tilde{\boldsymbol Z}_{\check S}(\tilde{\boldsymbol\beta}_{\check S} - \tilde{\boldsymbol\beta}_{0\check S})\|_2^2 + \frac{1}{2}n\lambda_2(\alpha_2-\alpha_{20})^2\right.$$
$$\left. + \frac{1}{2}\|\boldsymbol X_{\check S}(\boldsymbol\gamma_{\check S} - \boldsymbol\gamma_{0\check S})\|_2^2 + \boldsymbol\theta_{0\check S,-\sigma_y}^T(\boldsymbol\theta_{\check S,-\sigma_y} - \boldsymbol\theta_{0\check S,-\sigma_y}) - \frac{b_0-b_y}{\sigma_y^2}\right\}$$
$$\preceq \frac{2\lambda_2}{\lambda_1}M_1'^2 s_0\log p.$$

Thus the log ratio of the normalizing constants is controlled by $\mathcal{O}(s_0\log p)$.

Similarly, for the expectation term in (S2.11), by Lemma S2.2, on $\mathcal{E}_n$, it is bounded by

$$
\begin{aligned}
&2c_2(|\check{S}|+3)\log p + \mathbb{E}_{\check{Q}_{\check{S}}}\Bigg[-\frac{1}{2}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y}) \\
&-(\boldsymbol{\theta}_{0\check{S},-\sigma_y}-\boldsymbol{\mu}_{\check{S}})^T\Sigma_{\check{S}}^{-1}(\boldsymbol{\theta}_{\check{S},-\sigma_y}-\boldsymbol{\theta}_{0\check{S},-\sigma_y}) + c_1 n\|\boldsymbol{\theta}_{\check{S}}-\boldsymbol{\theta}_{0\check{S}}\|_2^2 + \frac{1}{2\tau^2}\|\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-}\|_2^2 \\
&+\frac{1}{\tau^2}\boldsymbol{\theta}_{0\check{S},-}^T(\boldsymbol{\theta}_{\check{S},-}-\boldsymbol{\theta}_{0\check{S},-}) + \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\alpha}\|_2^2 - 2(a_y-a_0)\log\sigma_y - \frac{b_y-b_0}{\sigma_y^2}\Bigg] \\
&\leq 2c_2(|\check{S}|+3)\log p - \frac{|\check{S}|+2}{2} + \frac{1}{2}(1+\tau^{-2})^2\boldsymbol{\theta}_{0\check{S},-\sigma_y}^T\Sigma_{\check{S}}\boldsymbol{\theta}_{0\check{S},-\sigma_y} \\
&+\left(c_1 n + \frac{1}{2\tau^2}\right)\left[(1+\tau^{-2})^2\boldsymbol{\theta}_{0\check{S},-\sigma_y}^T\Sigma_{\check{S}}^2\boldsymbol{\theta}_{0\check{S},-\sigma_y} + \mathrm{tr}(\boldsymbol{D}_{\check{S}})\right] + \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 \\
&-\left(\frac{1}{\tau^2}+\frac{1}{\tau^4}\right)\boldsymbol{\theta}_{0\check{S},-\sigma_y}^T\Sigma_{\check{S}}\boldsymbol{\theta}_{0\check{S},-\sigma_y} + \frac{1}{2\sigma_\alpha^2}\left[\boldsymbol{\theta}_{0\check{S},-\sigma_y}^T\left(I_{\check{S}}-(1+\tau^{-2})\Sigma_{\check{S}}\right)^2\boldsymbol{\theta}_{0\check{S},-\sigma_y} + \mathrm{tr}(\boldsymbol{D}_{\check{S}})\right] \\
&+c_1 n\left(\frac{b_y}{a_y-1} - 2\sigma_{y0}b_y^{1/2}r(a_y) + \sigma_{y0}^2\right) - 2(a_y-a_0)\log\sigma_y - (b_y-b_0)\frac{a_y}{b_y}.
\end{aligned}
$$

On $\mathcal{E}_n$, by the definitions of $a_y$ and $b_y$, the expectation term is bounded by

$$
\begin{aligned}
&2c_2(|\check{S}|+3)\log p - \frac{|\check{S}|+2}{2} + \frac{1}{2}(1+\tau^{-2})^2\frac{1}{n\lambda_1+2c_3'n+\tau^{-2}}\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2 \\
&+\frac{|\check{S}|+2}{4c_3'n\sigma_\alpha^2} + \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 - \frac{1}{\tau^2}\left(1+\frac{1}{\tau^2}\right)\frac{1}{n\lambda_2+2c_3'n+\tau^{-2}}\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2 \\
&+\left(c_1'n+\frac{1}{2\tau^2}\right)\left[(1+\tau^{-2})^2\frac{1}{(n\lambda_1+2c_3'n+\tau^{-2})^2}\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2 + \frac{|\check{S}|+2}{2c_3'n}\right] \\
&+\frac{1}{2\sigma_\alpha^2}\left[1-2\frac{1+\tau^{-2}}{n\lambda_2+2c_3'n+\tau^{-2}} + \frac{(1+\tau^{-2})^2}{(n\lambda_2+2c_3'n+\tau^{-2})^2}\right]\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2 - (b_y-b_0)\frac{a_0}{b_y} \\
&\preceq 2c_2'L's_0\log p + \frac{c_1'}{2c_3'}Ls_0 + \frac{1}{2\tau^2}\|\boldsymbol{\alpha}_0\|_2^2 \\
&+\frac{1}{n(\lambda_1+2c_3')}\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2 + \frac{1}{2\sigma_\alpha^2}\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2 + \frac{c_1'}{n(\lambda_1+2c_3')^2}\|\boldsymbol{\theta}_{0\check{S},-\sigma_y}\|_2^2,
\end{aligned}
$$

which is of the order $\mathcal{O}(s_0\log p)$.

Combining all of the above bounds gives the result. □

**Proof of Lemma S2.4**

By the definition of the prior $\Pi$, the desired integral is bounded below by

$$q^{s_0}(1-q)^{p-s_0}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{s_0}\left(\frac{1}{2\pi\sigma_\alpha^2}\right)$$
$$\times\int_{\Theta_{S_0}(M)}\frac{L_n(\boldsymbol{\theta}_{S_0})}{L_n(\boldsymbol{\theta}_{0S_0})}\exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S_0,-}\|_2^2-\frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\}p_\sigma(\sigma_y)d\boldsymbol{\theta}_{S_0},$$

where $p_\sigma(\sigma_y)$ defines a distribution derived from the inverse gamma distribution on $\sigma_y^2$. By Lemma S2.2, on $\mathcal{E}_{n,1}(R')$, the integral term in the above display is bounded below by

$$\exp\{-2c_2'(s_0+3)\log p\}$$
$$\int_{\bigcap_j\left\{(\theta_j-\theta_{0j})^2\le\frac{M^2}{(s_0+3)^2}\right\}}\exp\left\{-c_1'n\|\boldsymbol{\theta}_{S_0}-\boldsymbol{\theta}_{0S_0}\|_2^2-\frac{\|\boldsymbol{\theta}_{S_0,-}\|_2^2}{2\tau^2}-\frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\}p_\sigma(\sigma_y)d\boldsymbol{\theta}_{S_0}$$
$$=\exp\{-2c_2'(s_0+3)\log p\}\prod_{j\in S_0}\int_{\theta_{0j}-\frac{M}{s_0+3}}^{\theta_{0j}+\frac{M}{s_0+3}}\exp\left\{-c_1'n(\theta_j-\theta_{0j})^2-\frac{1}{2\tau^2}\theta_j^2\right\}d\theta_j$$
$$\times\prod_{j=1}^{2}\int_{\alpha_{0j}-\frac{M}{s_0+3}}^{\alpha_{0j}+\frac{M}{s_0+3}}\exp\left\{-c_1'n(\alpha_j-\alpha_{0j})^2-\frac{1}{2\sigma_\alpha^2}\alpha_j^2\right\}d\alpha_j$$
$$\times\int_{\sigma_{y0}-\frac{M}{s_0+3}}^{\sigma_{y0}+\frac{M}{s_0+3}}\exp\left\{-c_1'n(\sigma_y-\sigma_{y0})^2\right\}p_\sigma(\sigma_y)d\sigma_y.$$

For the integral of $\sigma_y$, we define $u_\sigma=\sqrt{2c_1'n}(\sigma_y-\sigma_{y0})$ and obtain

$$\int_{\sigma_{y0}-\frac{M}{s_0+3}}^{\sigma_{y0}+\frac{M}{s_0+3}}\exp\left\{-c_1n(\sigma_y-\sigma_{y0})^2\right\}p_\sigma(\sigma_y)d\sigma_y$$
$$=\sqrt{\frac{1}{2c_1'n}}\int_{-\frac{M}{s_0+3}\sqrt{2c_1'n}}^{\frac{M}{s_0+3}\sqrt{2c_1'n}}\exp\left\{-\frac{u_\sigma^2}{2}\right\}p_\sigma\left(\sqrt{\frac{1}{2c_1'n}}u_\sigma+\sigma_{y0}\right)du_\sigma,$$

which, as $n\to\infty$, converges to $\sqrt{\frac{2\pi}{2c_1'n}}\mathbb{E}_{u_\sigma}p_\sigma(\sigma_{y0})\asymp\sqrt{\frac{1}{c_1'n}}$ with $\mathbb{E}_{u_\sigma}$ denoting the expectation from a Gaussian distribution on $u_\sigma$. Combining the results on the integral of $\theta_j$ for $j \in S_0$ and $\alpha_k$, we have

$$\int_{\Theta(M)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) \succeq q^{s_0}(1-q)^{p-s_0} \exp\{-2c_2'(s_0+3)\log p\}$$
$$\times \left(\frac{1}{2c_1' n\tau^2+1}\right)^{\frac{s_0}{2}} \left(\frac{1}{2c_1' n\sigma_\alpha^2+1}\right) \sqrt{\frac{1}{c_1' n}}$$
$$\times \exp\left\{-\frac{c_1' n}{2c_1' n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2\right\} \exp\left\{-\frac{c_1' n}{2c_1' n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2\right\}.$$

□

**Proof of Lemma S2.5**

Define $\mathcal{A}(L') = \{\boldsymbol{\theta} \in \Theta(M) : |S| \geq L's_0\}$. By Lemma S2.4, the left-hand side is bounded above by

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{A}'(L)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) 1_{\mathcal{E}_{n,1}(R')}\right] q^{-s_0}(1-q)^{s_0-p}(2c_1' n\tau^2+1)^{\frac{s_0}{2}}(2c_1' n\sigma_\alpha^2+1)$$
$$\times \sqrt{c_1' n} \exp\left\{\frac{c_1' n}{2c_1' n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2 + \frac{c_1' n}{2c_1 n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2 + 2c_2'(s_0+3)\log p\right\}.$$

Given the definition of the prior $\Pi$, the expectation term is

$$\sum_{S:|S|\geq L's_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|} \frac{1}{2\pi\sigma_\alpha^2}$$
$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\Theta_S(M)} \frac{L_n(\boldsymbol{\theta}_S)}{L_n(\boldsymbol{\theta}_0)} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} p_\sigma(\sigma_y) d\boldsymbol{\theta}_S 1_{\mathcal{E}_{n,1}(R)}\right].$$

Similarly, we define $\boldsymbol{\theta}_{\tilde{S}} \in \mathbb{R}^{|S\cup S_0|+3}$ to contain $\boldsymbol{\theta}_S$ for $S$ and $\mathbf{0}$ for $S^c \cap S_0$.

By Lemma S2.2, we have

$$\int_{\Theta_S(M)} \exp\{Z_n(\boldsymbol{\theta}_{\tilde{S}})\} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} p_\sigma(\sigma_y) d\boldsymbol{\theta}_S$$

$$\leq \exp\{c_2'(|\tilde{S}|+3)\log p\}$$

$$\times \int_{\Theta_S(M)} \exp\{-c_3' n\|\boldsymbol{\theta}_{\tilde{S}} - \boldsymbol{\theta}_{0\tilde{S}}\|_2^2\} \exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2 - \frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\} p_\sigma(\sigma_y) d\boldsymbol{\theta}_S$$

$$\preceq \exp\{c_2'(|S|+s_0+3)\log p - c_3' n\|\boldsymbol{\theta}_{0S^c}\|_2^2\} \frac{2\pi}{2c_3'n + \frac{1}{\sigma_\alpha^2}} \left(\frac{2\pi}{2c_3'n+\frac{1}{\tau^2}}\right)^{\frac{|S|}{2}} \sqrt{\frac{1}{c_3'n}}$$

$$\times \exp\left\{-\frac{c_3'n}{2c_3'n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2\right\} \exp\left\{-\frac{c_3'n}{2c_3'n\tau^2+1}\|\boldsymbol{\theta}_{0S,-}\|_2^2\right\},$$

since the value of $\sigma_y$ is bounded and $p_\sigma(\cdot)$ is a density function. The remaining proof then follows the proof of Lemma S1.5. □

**Proof of Lemma S2.6**

Similarly, we define $\mathcal{B}(\eta) = \{\boldsymbol{\theta} \in \Theta(M) : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \eta\}$ and $\mathcal{B}(L', \eta) = \{\boldsymbol{\theta} \in \Theta(M) : |S| < L's_0, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \eta\}$. We have

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R')}\right]$$

$$\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta} \in \Theta(M) : |S| \geq L's_0 \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R')}\right] + \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\mathcal{B}(L',\eta) \mid \boldsymbol{Y}\right) 1_{\mathcal{E}_{n,1}(R')}\right].$$

By Lemma S2.5, for some constant $L'$, the first term on the right-hand side is bounded. By Lemma S2.4, the second term is bounded above by

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{B}(L',\eta)} \frac{L_n(\boldsymbol{\theta})}{L_n(\boldsymbol{\theta}_0)} d\Pi(\boldsymbol{\theta}) 1_{\mathcal{E}_{n,1}(R')}\right] q^{-s_0}(1-q)^{s_0-p}(2c_1'n\tau^2+1)^{\frac{s_0}{2}}(2c_1'n\sigma_\alpha^2+1)$$

$$\times \sqrt{c_1'n} \exp\left\{\frac{c_1'n}{2c_1'n\tau^2+1}\|\boldsymbol{\theta}_{0,-}\|_2^2 + \frac{c_1'n}{2c_1'n\sigma_\alpha^2+1}\|\boldsymbol{\alpha}_0\|_2^2 + 2c_2'(s_0+3)\log p\right\}.$$

The expectation term can be written as

$$\sum_{S:|S|\leq L's_0} q^{|S|}(1-q)^{p-|S|}\left(\frac{1}{\sqrt{2\pi}\tau}\right)^{|S|}\frac{1}{2\pi\sigma_\alpha^2}$$

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\int_{\mathcal{B}_S(\eta)}\frac{L_n(\boldsymbol{\theta}_S)}{L_n(\boldsymbol{\theta}_0)}\exp\left\{-\frac{1}{2\tau^2}\|\boldsymbol{\theta}_{S,-}\|_2^2-\frac{\|\boldsymbol{\alpha}\|_2^2}{2\sigma_\alpha^2}\right\}p_\sigma(\sigma_y)d\boldsymbol{\theta}_S 1_{\mathcal{E}_{n,1}(R')}\right].$$

The remaining proof follows the ideas in the proof of Lemma S1.6. □

**Proof of Lemma S2.7**

The proof directly follows that of Lemma S1.7. □

**Proof of Lemma S2.8**

Similar to the proof of Lemma S2.6, we have

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M):S\neq S_0\mid \boldsymbol{Y}\right)1_{\mathcal{E}_{n,1}(R')}\right]\leq \mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M):|S|\geq L's_0\mid \boldsymbol{Y}\right)1_{\mathcal{E}_{n,1}(R')}\right]$$

$$+\,\mathbb{E}_{\boldsymbol{\theta}_0}\left[\Pi\left(\boldsymbol{\theta}\in\Theta(M):S\neq S_0,|S|<L's_0\mid \boldsymbol{Y}\right)1_{\mathcal{E}_{n,1}(R')}\right],$$

where the first term is bounded for some constant $L'$. For the second term, we follow the same idea in the proof of Lemma S1.8. □

# S3 Details of the Variational Algorithm

In this section, we provide the updates for the non-hierarchical factors, including $\pi_i$ and $c_i$ for $i=1,\ldots,n$, and $(a_1,b_1)$ for $\sigma_y^2$, as well as the derivations of the CAVI updates of our variational algorithm in Section 4.

## S3.1 CAVI Updates For Non-Hierarchical Factors

For the non-hierarchical factors, the updates for $q(\delta_i)$ for $i = 1, \ldots, n$ are Bernoulli with log odds ratio equal to

$$\begin{aligned}\log\frac{\pi_i}{1-\pi_i} &= x_i^T(\boldsymbol{\eta}^\gamma \odot \boldsymbol{\mu}^\gamma) \\ &- \frac{a_1}{2b_1}\left\{\mu_1^2 + \sigma_1^2 - \mu_2^2 - \sigma_2^2 - 2\left[y_i - z_i^T(\boldsymbol{\eta}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}^{\boldsymbol{\beta}})\right](\mu_1 - \mu_2)\right\} t_i,\end{aligned} \tag{S3.1}$$

where $\odot$ denotes the element-by-element product. The updates for $q(\omega_i)$ for $i = 1, \ldots, n$ are $\mathrm{PG}(1, c_i)$ with

$$c_i = \sqrt{x_i^T\left\{\mathrm{D}\left(\boldsymbol{\eta}^\gamma \odot \sigma^{\gamma 2}\right) + \left[\boldsymbol{\eta}^\gamma(\boldsymbol{\eta}^\gamma)^T + \mathrm{D}\left(\boldsymbol{\eta}^\gamma \odot (1-\boldsymbol{\eta}^\gamma)\right)\right] \odot \boldsymbol{\mu}^\gamma(\boldsymbol{\mu}^\gamma)^T\right\} x_i}. \tag{S3.2}$$

The update for $q(\sigma_y^2)$ is given by $a_1 = n/2 + 1 + \sum_{j=1}^{p_Z}\eta_j^{\boldsymbol{\beta}}/2 + a_0$ and

$$\begin{aligned} b_1 = \frac{1}{2}&\left[\boldsymbol{Y}^T\boldsymbol{Y} + (m_1^2 + \sigma_1^2)\boldsymbol{T}^T\mathbb{E}\boldsymbol{\Delta}\boldsymbol{T} + (m_2^2 + \sigma_2^2)\boldsymbol{T}^T(\mathbf{I} - \mathbb{E}\boldsymbol{\Delta})\boldsymbol{T}\right. \\ &\left. -2\boldsymbol{Y}^T\left(\mu_1\mathbb{E}\boldsymbol{\Delta}\boldsymbol{T} + \mu_2(\mathbf{I} - \mathbb{E}\boldsymbol{\Delta})\boldsymbol{T}\right)\right] \\ &- \left[\boldsymbol{Y} - \mu_1\mathbb{E}\boldsymbol{\Delta}\boldsymbol{T} - \mu_2(\mathbf{I} - \mathbb{E}\boldsymbol{\Delta})\boldsymbol{T}\right]^T\boldsymbol{Z}\left(\boldsymbol{\eta}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}^{\boldsymbol{\beta}}\right) \\ &+ \frac{1}{2}\left(\boldsymbol{\eta}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}^{\boldsymbol{\beta}}\right)^T\boldsymbol{Z}^T\boldsymbol{Z}\left(\boldsymbol{\eta}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}^{\boldsymbol{\beta}}\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{Z}^T\boldsymbol{Z}\,\mathrm{D}\left(\boldsymbol{\eta}^{\boldsymbol{\beta}} \odot \sigma^{\boldsymbol{\beta}2}\right)\right) \\ &+ \frac{1}{2}\left(\boldsymbol{\eta}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}^{\boldsymbol{\beta}}\right)^T\mathrm{D}\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)\left(\left(1 - \boldsymbol{\eta}^{\boldsymbol{\beta}}\right) \odot \boldsymbol{\mu}^{\boldsymbol{\beta}}\right) \\ &+ \frac{1}{2\sigma_\alpha^2}(\mu_1^2 + \sigma_1^2 + \mu_2^2 + \sigma_2^2) + \frac{1}{2\tau_{\boldsymbol{\beta}}^2}\sum_{j=1}^{p_Z}\eta_j^{\boldsymbol{\beta}}\left(\mu_j^{\boldsymbol{\beta}2} + \sigma_j^{\boldsymbol{\beta}2}\right) + b_0.\end{aligned} \tag{S3.3}$$

### S3.2 Proofs for the CAVI Updates

Since the updates of the non-hierarchical factors can be directly obtained based on (4.8), we focus on the derivation of the hierarchical ones.

The optimal variational distribution minimizes the KL divergence, or equivalently, maximizes the evidence lower bound (ELBO)

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) &= \int q(\boldsymbol{\theta}, \boldsymbol{\phi}) \log \frac{\pi(\boldsymbol{\theta}) L_n(\boldsymbol{\theta}, \boldsymbol{\phi})}{q(\boldsymbol{\theta}, \boldsymbol{\phi})} d\boldsymbol{\theta} d\boldsymbol{\phi} \\ &= \mathbb{E}_q \left[ \log(\pi(\boldsymbol{\theta}) L_n(\boldsymbol{\theta}, \boldsymbol{\phi})) - \log q(\boldsymbol{\theta}, \boldsymbol{\phi}) \right].\end{aligned}$$

Thus, we need to calculate the expectation with respect to the variational distribution. Before diving into the details, we first present some preliminary results to facilitate the calculation. The marginal expectation of $\omega_i$ from $\mathrm{PG}(1, c_1)$ is given by $\mathbb{E}_q[\omega_i] = \tanh(c_i/2)/2c_i$. The expectation of $1/\sigma_y^2$ is given by $a_1/b_1$. The marginal expectation of $\boldsymbol{\gamma}$ can be calculated conditionally as

$$\mathbb{E}_q[\boldsymbol{\gamma}] = \mathbb{E}_{I^\gamma} \mathbb{E}_{\boldsymbol{\gamma}|I^\gamma}[\boldsymbol{\gamma}] = \mathbb{E}_{I^\gamma} \left[ I^\gamma \odot \boldsymbol{\mu}^\gamma \right] = \boldsymbol{\eta}^\gamma \odot \boldsymbol{\mu}^\gamma.$$

We are ready to derive the updates. For simplicity, we use $C$ to represent a constant that does not affect the optimization process, with its exact value potentially varying across different lines. For $\beta_j$ for $j = 1, \ldots, p_Z$, if conditional on $I_j^{\boldsymbol{\beta}} = 0$, the variational posterior is $\delta_0$, and if conditional on

$I_j^{\boldsymbol{\beta}} = 1$, we can express the ELBO as a function of $\mu_j^{\boldsymbol{\beta}}$ and $\sigma_j^{\boldsymbol{\beta}2}$,

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) &= \mathbb{E}_{q(I_j^{\boldsymbol{\beta}}=1)} \left[ -\frac{1}{2\sigma_y^2} \|\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta} - \alpha_1 \boldsymbol{\Delta} \boldsymbol{T} - \alpha_2 (\mathbf{I} - \boldsymbol{\Delta}) \boldsymbol{T}\|_2^2 - \frac{\beta_j^2}{2\tau_{\boldsymbol{\beta}}^2 \sigma_y^2} \right] \\
&+ \mathbb{E}_{q(I_j^{\boldsymbol{\beta}}=1)} \left[ \frac{1}{2} \log \sigma_j^{\boldsymbol{\beta}2} + \frac{1}{2\sigma_j^{\boldsymbol{\beta}2}} \left( \beta_j - \mu_j^{\boldsymbol{\beta}} \right)^2 \right] + C \\
&= \frac{a_1}{2b_1} \Big[ 2 \left[ \boldsymbol{Y} - \mu_1 \mathbb{E} \boldsymbol{\Delta} \boldsymbol{T} - \mu_2 (\mathbf{I} - \mathbb{E} \boldsymbol{\Delta}) \boldsymbol{T} \right]^T z_j \mu_j^{\boldsymbol{\beta}} - 2 \mu_j^{\boldsymbol{\beta}} z_j^T \boldsymbol{Z}_{-j} \left( \boldsymbol{\eta}_{-j}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}_{-j}^{\boldsymbol{\beta}} \right) \\
&- \left( \mu_j^{\boldsymbol{\beta}2} + \sigma_j^{\boldsymbol{\beta}2} \right) z_j^T z_j \Big] - \frac{a_1}{2\tau_{\boldsymbol{\beta}}^2 b_1} \left( \mu_j^{\boldsymbol{\beta}2} + \sigma_j^{\boldsymbol{\beta}2} \right) + \frac{1}{2} \log \sigma_j^{\boldsymbol{\beta}2} + C,
\end{aligned}$$

where the subscript $q(I_j^{\boldsymbol{\beta}} = 1)$ denotes that the expectation is taken with respect to the variational distribution conditional on $I_j^{\boldsymbol{\beta}} = 1$. We then take derivatives with respect to $\mu_j^{\boldsymbol{\beta}}$ and $\sigma_j^{\boldsymbol{\beta}2}$ to get

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi})}{\partial \mu_j^{\boldsymbol{\beta}}} &= \frac{a_1}{b_1} \Big[ \left[ \boldsymbol{Y} - \mu_1 \mathbb{E} \boldsymbol{\Delta} \boldsymbol{T} - \mu_2 (\mathbf{I} - \mathbb{E} \boldsymbol{\Delta}) \boldsymbol{T} \right]^T z_j \\
&\quad - z_j^T \boldsymbol{Z}_{-j} \left( \boldsymbol{\eta}_{-j}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}_{-j}^{\boldsymbol{\beta}} \right) - (z_j^T z_j + \tau_{\boldsymbol{\beta}}^{-2}) \mu_j^{\boldsymbol{\beta}} \Big], \\
\frac{\partial \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi})}{\partial \sigma_j^{\boldsymbol{\beta}2}} &= -\frac{a_1}{2b_1} \left( z_j^T z_j + \tau_{\boldsymbol{\beta}}^{-2} \right) + \frac{1}{2\sigma_j^{\beta 2}}.
\end{aligned}$$

Then the updates conditional on $I_j^{\boldsymbol{\beta}} = 1$ is given by the maximizers

$$\begin{aligned}
\mu_j^{\boldsymbol{\beta}} &= \frac{\left[ \boldsymbol{Y} - \mu_1 \mathbb{E} \boldsymbol{\Delta} \boldsymbol{T} - \mu_2 (\mathbf{I} - \mathbb{E} \boldsymbol{\Delta}) \boldsymbol{T} \right]^T z_j - (\boldsymbol{\eta}_{-j}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}_{-j}^{\boldsymbol{\beta}}) \boldsymbol{Z}_{-j}^T z_j}{z_j^T z_j + \tau_{\boldsymbol{\beta}}^{-2}}, \\
\sigma_j^{\boldsymbol{\beta}} &= \frac{1}{a_1 (z_j^T z_j + \tau_{\boldsymbol{\beta}}^{-2}) / b_1}.
\end{aligned}$$

For $I_j^{\boldsymbol{\beta}}$ for $j = 1, \ldots, p_Z$, we maximize the ELBO with respect to $\eta_j^{\boldsymbol{\beta}}$:

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = C + \mathbb{E}_q \Bigg[ -\frac{a_1}{2b_1} \left\| \boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\beta} - \alpha_1 \boldsymbol{\Delta T} - \alpha_2 (\mathbf{I} - \boldsymbol{\Delta}) \boldsymbol{T} \right\|_2^2 \\
&+ I_j^{\boldsymbol{\beta}} \left( \log \frac{q_{\boldsymbol{\beta}}}{\sqrt{2\pi \tau_{\boldsymbol{\beta}}^2}} - \frac{1}{2} \log \sigma_y^2 - \frac{a_1 \beta_j^2}{2\tau_{\boldsymbol{\beta}}^2 b_1} \right) + \left(1 - I_j^{\boldsymbol{\beta}}\right) \log \left(1 - q_{\boldsymbol{\beta}}\right) \Bigg] \\
&- \mathbb{E}_q \left[ I_j^{\boldsymbol{\beta}} \left( \log \frac{\eta_j^{\boldsymbol{\beta}}}{\sqrt{2\pi \sigma_j^{\boldsymbol{\beta} 2}}} - \frac{\left(\beta_j - \mu_j^{\boldsymbol{\beta}}\right)^2}{2\sigma_j^{\boldsymbol{\beta} 2}} \right) + \left(1 - I_j^{\boldsymbol{\beta}}\right) \log \left(1 - \eta_j^{\boldsymbol{\beta}}\right) \right] \\
&= C + \frac{a_1}{b_1} \Bigg[ \left[ \boldsymbol{Y} - \mu_1 \mathbb{E}\boldsymbol{\Delta T} - \mu_2 (\mathbf{I} - \mathbb{E}\boldsymbol{\Delta}) \boldsymbol{T} \right]^T z_j \eta_j^{\boldsymbol{\beta}} \mu_j^{\boldsymbol{\beta}} \\
&- (\boldsymbol{\eta}_{-j}^{\boldsymbol{\beta}} \odot \boldsymbol{\mu}_{-j}^{\boldsymbol{\beta}})^T \boldsymbol{Z}_{-j}^T z_j \eta_j^{\boldsymbol{\beta}} \mu_j^{\boldsymbol{\beta}} - \frac{1}{2} z_j^T z_j \eta_j^{\boldsymbol{\beta}} (\mu_j^{\boldsymbol{\beta} 2} + \sigma_j^{\boldsymbol{\beta} 2}) \Bigg] + (1 - \eta_j^{\boldsymbol{\beta}}) \log \frac{1 - q_{\boldsymbol{\beta}}}{1 - \eta_j^{\boldsymbol{\beta}}} \\
&+ \eta_j^{\boldsymbol{\beta}} \left[ \log \frac{q_{\boldsymbol{\beta}} \sqrt{\sigma_j^{\boldsymbol{\beta} 2}}}{\eta_j^{\boldsymbol{\beta}} \sqrt{\tau_{\boldsymbol{\beta}}^2}} - \frac{1}{2} (\log b_1 - \psi(a_1)) - \frac{a_1}{2\tau_{\boldsymbol{\beta}}^2 b_1} \left( \mu_j^{\boldsymbol{\beta} 2} + \sigma_j^{\boldsymbol{\beta} 2} \right) + \frac{1}{2} \right].
\end{aligned}
$$

After taking derivative with respect to $\eta_j^{\boldsymbol{\beta}}$, the update for $q(I_j^{\boldsymbol{\beta}})$ is given by the optimizer solving

$$
\log \frac{\eta_j^{\boldsymbol{\beta}}}{1 - \eta_j^{\boldsymbol{\beta}}} = \frac{\mu_j^{\boldsymbol{\beta} 2}}{2\sigma_j^{\boldsymbol{\beta} 2}} - \frac{1}{2} (\log b_1 - \psi(a_1)) + \log \frac{q_{\boldsymbol{\beta}} \sigma_j^{\boldsymbol{\beta}}}{(1 - q_{\boldsymbol{\beta}}) \tau_{\boldsymbol{\beta}}}.
$$

Similarly, conditional on $I_\ell^\gamma = 0$, the update for $\gamma_\ell$ is $\delta_0$ for $\ell = 1, \ldots, p_X$.

Conditional on $I_\ell^\gamma = 1$.

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbb{E}_{q(I_\ell^\gamma=1)}\left[\sum_{i=1}^{n}\left[\left(\delta_i - \frac{1}{2}\right)x_i^T\boldsymbol{\gamma} - \frac{1}{2}\omega_i\left(x_i^T\boldsymbol{\gamma}\right)^2\right] - \frac{\gamma_\ell^2}{2\tau_\gamma^2}\right]$$
$$+ \mathbb{E}_{q(I_\ell^\gamma=1)}\left[\frac{1}{2}\log\sigma_\ell^{\gamma 2} + \frac{1}{2\sigma_\ell^{\gamma 2}}\left(\gamma_\ell - \mu_\ell^\gamma\right)^2\right] + C$$
$$= \mathbf{1}^T(\mathbb{E}\boldsymbol{\Delta} - 1/2)x_\ell\mu_\ell^\gamma - (\boldsymbol{\eta}_{-\ell}^\gamma \odot \boldsymbol{\mu}_{-\ell}^\gamma)^T\boldsymbol{X}_{-\ell}^T\mathbb{E}\boldsymbol{\Omega}x_\ell\mu_\ell^\gamma$$
$$- \frac{1}{2}x_\ell^T\mathbb{E}\boldsymbol{\Omega}x_\ell\left(\mu_\ell^{\gamma 2} + \sigma_\ell^{\gamma 2}\right) - \frac{1}{2\tau_\gamma^2}(\mu_\ell^{\gamma 2} + \sigma_\ell^{\gamma 2}) + \frac{1}{2}\log\sigma_\ell^{\gamma 2} + C.$$

We take first derivatives with respect to $\mu_j^\gamma$ and $\sigma_j^{\gamma 2}$ to obtain the updates

$$\mu_\ell^\gamma = \frac{\mathbf{1}^T(\mathbb{E}\boldsymbol{\Delta} - 1/2)x_\ell - (\boldsymbol{\eta}_{-\ell}^\gamma \odot \boldsymbol{\mu}_{-\ell}^\gamma)^T\boldsymbol{X}_{-\ell}^T\mathbb{E}\boldsymbol{\Omega}x_\ell}{x_\ell^T\mathbb{E}\boldsymbol{\Omega}x_\ell + \tau_\gamma^{-2}},$$
$$\sigma_\ell^{\gamma 2} = \frac{1}{x_\ell^T\mathbb{E}\boldsymbol{\Omega}x_\ell + \tau_\gamma^{-2}},$$

For $I_\ell^\gamma$ for $\ell = 1, \ldots, p_X$, the ELBO is calculated as

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = C + \mathbb{E}_q\left[\sum_{i=1}^{n}\left[\left(\delta_i - \frac{1}{2}\right)x_i^T\boldsymbol{\gamma} - \frac{1}{2}\omega_i\left(x_i^T\boldsymbol{\gamma}\right)^2\right]\right.$$
$$\left. + I_\ell^\gamma\left(\log\frac{q_\gamma}{\sqrt{2\pi\tau_\gamma^2}} - \frac{\gamma_\ell^2}{2\tau_\gamma^2}\right) + (1 - I_\ell^\gamma)\log(1 - q_\gamma)\right]$$
$$- \mathbb{E}_q\left[I_\ell^\gamma\left(\log\frac{\eta_\ell^\gamma}{\sqrt{2\pi\sigma_\ell^{\gamma 2}}} - \frac{(\gamma_\ell - \mu_\ell^\gamma)^2}{2\sigma_\ell^{\gamma 2}}\right) + (1 - I_\ell^\gamma)\log(1 - \eta_\ell^\gamma)\right],$$

and the updates are derived by solving

$$\log\frac{\eta_\ell^\gamma}{1 - \eta_\ell^\gamma} = \frac{\mu_\ell^{\gamma 2}}{2\sigma_\ell^{\gamma 2}} + \log\frac{q_\gamma\sigma_\ell^\gamma}{(1 - q_\gamma)\tau_\gamma}.$$

# S4 Additional Simulation Results

In this section, we present additional numerical results on the comparison of the proposed VB method with the scalable MCMC-based BVSA (Zhang et al., 2025) and other subgroup identification approaches.

## S4.1 Variable Selection with Correlated Covariates

We adopt the same settings as in Section 5.1 except that we now consider the scenario where the prognostic covariates $\boldsymbol{Z}$, as well as the predictive covariates $\boldsymbol{X}$, can be correlated. To account for this, we generate the covariates $\boldsymbol{z}_i$ independently from a normal distribution with a mean vector $\mathbf{0}$ and pairwise covariate correlations of $\rho = 0.25$. The same generation process is applied to $\boldsymbol{x}_i$.

We conduct VSM and BVSA following the same procedure as in Section 5.1 and run 100 independent trials. The results are summarized in Table 1. Similar conclusions hold under the scenario of correlated covariates. Although the performance of VSM deteriorates due to the presence of correlation, it remains comparable to BVSA across all settings when $n = 200$. As $n$ increases to 300, VSM shows a more substantial improvement compared to BVSA, particularly when $p = 2000$, further supporting our theoretical conclusions.

Table 1: Finite sample results on variable selection performance under structured mixture model settings when $\rho = 0.25$. VSM: our proposed method; BVSA: scalable MCMC-based method (Zhang et al., 2025).

| $p$ | $n$ | Method | $\boldsymbol{\beta}$ | | | | $\boldsymbol{\gamma}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | TPR | FDR | F1 | Ext | TPR | FDR | F1 | Ext |
| 100 | 200 | VSM | 1 | 0 | 1 | 100% | 0.930 | 0.053 | 0.933 | 77% |
| | | BVSA | 1 | 0 | 1 | 100% | 0.955 | 0.075 | 0.934 | 76% |
| | 300 | VSM | 1 | 0 | 1 | 100% | 0.988 | 0.024 | 0.980 | 98% |
| | | BVSA | 1 | 0 | 1 | 100% | 0.993 | 0.075 | 0.953 | 97% |
| 500 | 200 | VSM | 1 | 0 | 1 | 100% | 0.893 | 0.130 | 0.874 | 63% |
| | | BVSA | 1 | 0 | 1 | 100% | 0.900 | 0.043 | 0.922 | 68% |
| | 300 | VSM | 1 | 0 | 1 | 100% | 0.990 | 0.088 | 0.945 | 93% |
| | | BVSA | 1 | 0 | 1 | 100% | 0.948 | 0.035 | 0.951 | 91% |
| 2000 | 200 | VSM | 1 | 0 | 1 | 100% | 0.668 | 0.122 | 0.737 | 29% |
| | | BVSA | 1 | 0 | 1 | 100% | 0.625 | 0.121 | 0.710 | 20% |
| | 300 | VSM | 1 | 0 | 1 | 100% | 0.915 | 0.058 | 0.919 | 79% |
| | | BVSA | 1 | 0 | 1 | 100% | 0.795 | 0.047 | 0.848 | 56% |

## S4.2 Estimation Comparisons with the MCMC Method

To evaluate parameter estimation accuracy, we analyze how the $\ell_2$ errors of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$ change with increasing sample size. We consider both $p = 100$ and $p = 500$ settings and use the same true values of $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$, and $\boldsymbol{\alpha}_0$ as specified in Section 5.1. The averaged $\ell_2$ errors are obtained from 100 independent trials and presented in Figure 1 for $p = 100$ and Figure 2 for $p = 500$.

As $n$ increases, the $\ell_2$ errors and their standard errors decrease to-

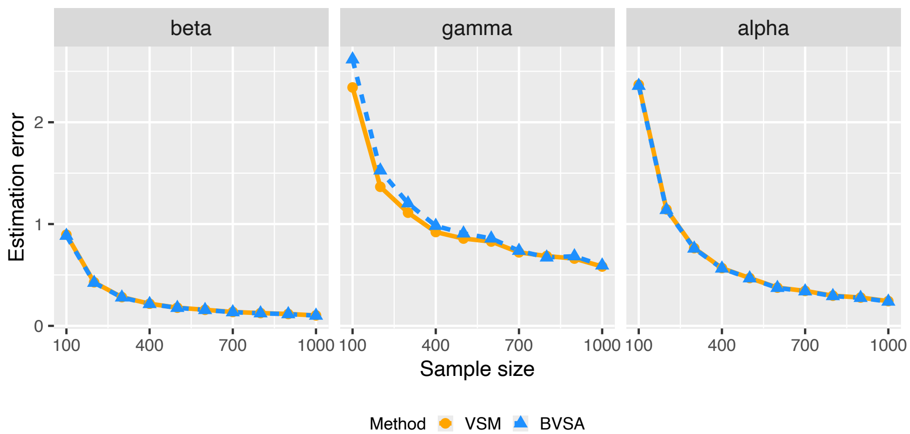


Figure 1: The $\ell_2$ errors of parameter estimation with growing sample sizes when $p = 100$.

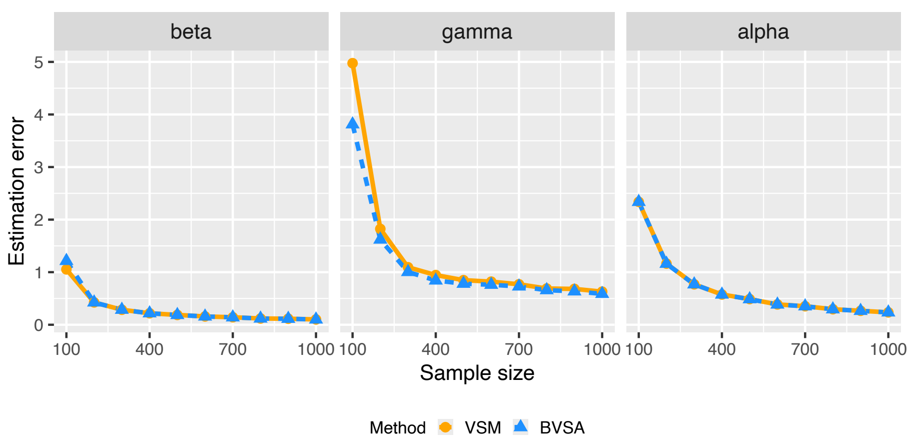


Figure 2: The $\ell_2$ errors of parameter estimation with growing sample sizes when $p = 500$.

wards 0 for all parameters. The estimation errors of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are nearly identical between VSM and BVSA across different model dimensions and sample sizes, indicating that the variational approximation induces negligible bias in estimating the linear coefficients. For the logistic coefficient $\boldsymbol{\gamma}$, the gaps between the $\ell_2$ errors of VSM and BVSA are slightly larger. Nevertheless, their overall performance remains comparable. Note that when $p = 100$, VSM achieves smaller $\ell_2$ errors, suggesting that the proposed variational method can even attain higher estimation accuracy than the scalable MCMC-based counterpart.

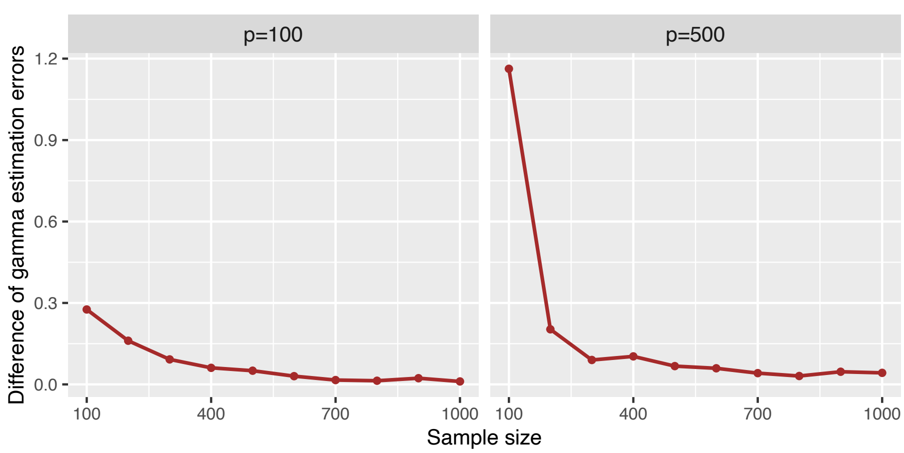


Figure 3: The difference of $\ell_2$ errors of $\boldsymbol{\gamma}$ estimation between VSM and BVSA with growing sample sizes when $p = 100$ and $p = 500$.

To empirically illustrate the asymptotic behavior of variational approximation error, we examine how the gap between the $\boldsymbol{\gamma}$ estimation errors from

VSM and BVSA changes with $n$. As shown in Figure 3, the gap exhibits an overall shrinking trend as $n$ grows, suggesting that the approximation error introduced by VB converges to 0 with larger samples. This observation is consistent with our theoretical findings, where the approximation error is of order $\mathcal{O}(C_M/M_n)$ in Theorem 1 and vanishes as $M_n \to \infty$.

### S4.3 Selection Frequencies Under Traditional Subgroup Settings

In this subsection, we provide additional analyses under the traditional subgroup settings. The comparison methods are the same as those adopted in Section 5.2, which are implemented using R codes (BVSA) and R packages `partykit` for MOB, `MrSGUIDE` for GUIDE, `SubgrpID` for PRIM and SeqBT, and `FindIt` for FindIt. BVSA is implemented as described in Section 5.1, and parameters for subgroup identification methods are set to the recommended values.

We consider the two settings evaluated in Section 5.2 along with an additional setting S0:

$$\mathbf{S0}: Y = 1 + Z_1 + Z_2 + 40t + \varepsilon,$$
$$\mathbf{S1}: Y = 1 + Z_2 + 40tI_{(X_1>0,X_4<1,X_6=2)} + \varepsilon,$$
$$\mathbf{S2}: Y = 1 + Z_1 + Z_2 + Z_4 + I_{(Z_6=2)} + Z_7 + 40tI_{(X_1>0,X_4<1,X_6=2)} + \varepsilon,$$

where $\varepsilon \sim N(0,1)$. Setting S0 contains no meaningful subgroups and is designed to examine the robustness of VSM against falsely identifying spurious subgroup structures.

We set $p = 20$ and $n = 200$, and generate all prognostic and predictive covariates in the same way as described in Section 5.2. The selection frequencies of predictive covariates averaged over 100 random replications are reported in Table 2. We highlight two observations. First, under setting S0, VSM tends not to select covariates when no subgroup exists and achieves selection frequencies comparable to BVSA. In contrast, other subgroup identification methods, particularly those based on splitting rules, are more likely to mistakenly identify subgroups and select inactive covariates. Second, under settings S1 and S2, VSM and BVSA successfully identify all important predictive covariates, while other methods struggle to differentiate between predictive and prognostic covariates, frequently selecting inactive predictive covariates.

### S4.4 Sensitivity Analysis of Initialization

In this subsection, we investigate the sensitivity of our proposed VSM to different initialization approaches. We focus on the logistic component that governs the predictive variable selection, while keeping the linear component randomly initialized as implemented in our numerical studies. We compare the proposed GUIDE-based initialization (**GUIDE**, Loh 2002) with the

Table 2: Selection frequencies of predictive variables when $p = 20$.

(a) **S0** : $Y = 1 + Z_1 + Z_2 + 40t + \varepsilon$

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| VSM | 0.02 | 0.04 | 0.04 | 0 | 0.01 | 0 | 0.02 | 0.03 | 0.01 | 0.02 |
| BVSA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GUIDE | 0.27 | 0.28 | 0.22 | 0.07 | 0.05 | 0.10 | 0.11 | 0.10 | 0.08 | 0.11 |
| FindIt | 0.08 | 0.01 | 0 | 0.02 | 0.03 | 0.05 | 0 | 0 | 0 | 0.02 |
| PRIM | 0.53 | 0.55 | 0.07 | 0.08 | 0 | 0 | 0.06 | 0.03 | 0.04 | 0.02 |
| MOB | 1 | 1 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0 |
| SeqBT | 0.18 | 0.21 | 0.19 | 0.12 | 0.06 | 0.08 | 0.14 | 0.09 | 0.09 | 0.13 |

(b) **S1** : $Y = 1 + Z_2 + 40tI_{(X_1>0, X_4<1, X_6=2)} + \varepsilon$

| | $\mathbf{X_1}$ | $X_2$ | $X_3$ | $\mathbf{X_4}$ | $X_5$ | $\mathbf{X_6}$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| VSM | 0.90 | 0.05 | 0.01 | 0.78 | 0.01 | 0.97 | 0.01 | 0.01 | 0.01 | 0.01 |
| BVSA | 0.92 | 0.01 | 0 | 0.85 | 0.02 | 1 | 0 | 0 | 0 | 0 |
| GUIDE | 0.86 | 0.18 | 0.07 | 0.13 | 0.05 | 0.98 | 0.03 | 0.06 | 0.04 | 0.07 |
| FindIt | 0.99 | 0.84 | 0.81 | 1 | 0.86 | 1 | 0.93 | 0.86 | 0.80 | 0.89 |
| PRIM | 0.53 | 0.16 | 0.15 | 0.50 | 0 | 0.12 | 0.12 | 0.16 | 0.16 | 0.12 |
| MOB | 0.19 | 1 | 0.02 | 0.05 | 0 | 0.51 | 0 | 0.01 | 0 | 0 |
| SeqBT | 0.08 | 0 | 0.01 | 0.02 | 0 | 0.93 | 0 | 0 | 0 | 0 |

(c) **S2** : $Y = 1 + Z_1 + Z_2 + Z_4 + I_{(Z_6=2)} + Z_7 + 40tI_{(X_1>0, X_4<1, X_6=2)} + \varepsilon$

| | $\mathbf{X_1}$ | $X_2$ | $X_3$ | $\mathbf{X_4}$ | $X_5$ | $\mathbf{X_6}$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| VSM | 0.90 | 0.03 | 0 | 0.75 | 0.04 | 0.97 | 0 | 0 | 0.01 | 0 |
| BVSA | 0.90 | 0.01 | 0 | 0.82 | 0.02 | 1 | 0 | 0.01 | 0 | 0 |
| GUIDE | 0.87 | 0.22 | 0.14 | 0.26 | 0.04 | 0.98 | 0.21 | 0.08 | 0.13 | 0.09 |
| FindIt | 1 | 0.84 | 0.84 | 0.99 | 0.87 | 1 | 0.92 | 0.82 | 0.80 | 0.86 |
| PRIM | 0.51 | 0.19 | 0.23 | 0.43 | 0 | 0.12 | 0.19 | 0.22 | 0.14 | 0.19 |
| MOB | 0.90 | 0.74 | 0.01 | 0.52 | 0 | 0.86 | 0.75 | 0.03 | 0 | 0.03 |
| SeqBT | 0.09 | 0 | 0 | 0.01 | 0 | 0.94 | 0 | 0 | 0 | 0 |

following alternative approaches:

- **MOB**: initialize via the subgroup method MOB (Seibold et al., 2016);
- **Random**: initialize by randomly selecting active predictive covariates with a predetermined size of $\min(5, 0.2p_X)$;
- **Zero**: initialize without any active predictive covariate.

We begin with the mixture model setting described in Section 5.1 and conduct simulations with $n = 200$ and varying $p \in \{100, 500\}$. The results averaged from 100 random replications are summarized in Table 3. All initialization strategies yield similar performance across all metrics, indicating that the proposed VSM is highly robust to initialization when the model is correctly specified.

To comprehensively examine initialization sensitivity, we consider the traditional subgroup scenario in Section 5.2. We adopt setting S1 with $n = 200$ and $p \in \{20, 200\}$. Table 4 reports the results averaged from 100 random trials. GUIDE initialization achieves the highest F1 and exact recovery (Ext) scores among all methods, demonstrating its advantage in identifying active predictive covariates under model misspecification. Its advantage becomes more evident in the high-dimensional case when $p = 200$. In contrast, the Zero initialization exhibits conservative behavior,

Table 3: Predictive variable selection results under structured mixture model settings in Section 5.1 for $n = 200$ and $p \in \{100, 500\}$. "GUIDE": initialization via GUIDE (Loh, 2002); "MOB": initialization via MOB (Seibold et al., 2016); "Random": initialization with random predictive variable selection; "Zero": initialization without any predictive variable.

| Mixture model | $p = 100$ | | | | $p = 500$ | | | |
|---|---|---|---|---|---|---|---|---|
| | TPR | FDR | F1 | Ext | TPR | FDR | F1 | Ext |
| **GUIDE** | 0.955 | 0.036 | 0.956 | 90% | 0.903 | 0.126 | 0.881 | 64% |
| MOB | 0.945 | 0.027 | 0.955 | 89% | 0.903 | 0.126 | 0.881 | 64% |
| Random | 0.953 | 0.032 | 0.956 | 91% | 0.900 | 0.141 | 0.871 | 64% |
| Zero | 0.948 | 0.031 | 0.954 | 90% | 0.905 | 0.121 | 0.884 | 64% |

with notably lower TPR and FDR values, underscoring the importance of providing an informative initialization to guide the variational optimization.

It can be concluded that VSM is strongly robust to initialization when the model is correctly specified. Furthermore, under model misspecification, the proposed GUIDE-based initialization can lead to more accurate variable selection.

## S4.5 Sensitivity Analysis of Hyperparameters

To examine the sensitivity of VSM to the spike-and-slab prior hyperparameters, we conduct analyses under the structured mixture model setting specified in Section 5.1 with $n = 200$ and $p \in \{100, 500\}$. We vary each hyperpa-

Table 4: Predictive variable selection results for S1: $Y = 1+Z_2+40tI_{(X_1>0,X_4<1,X_6=2)}+\varepsilon$ for $n = 200$ and $p \in \{20, 200\}$ in Section 5.2. "GUIDE": initialization via GUIDE (Loh, 2002); "MOB": initialization via MOB (Seibold et al., 2016); "Random": initialization with random predictive variable selection; "Zero": initialization without any predictive variable.

| Tree model | $p = 20$ | | | | $p = 200$ | | | |
|---|---|---|---|---|---|---|---|---|
| | TPR | FDR | F1 | Ext | TPR | FDR | F1 | Ext |
| **GUIDE** | 0.873 | 0.031 | 0.891 | 77% | 0.703 | 0.268 | 0.662 | 38% |
| MOB | 0.727 | 0.024 | 0.786 | 69% | 0.533 | 0.161 | 0.586 | 25% |
| Random | 0.724 | 0.047 | 0.793 | 68% | 0.507 | 0.196 | 0.591 | 20% |
| Zero | 0.580 | 0.008 | 0.693 | 63% | 0.477 | 0.067 | 0.585 | 19% |

rameter independently while keeping the others fixed at their recommended values in Section 4.3. Specifically, we consider $q_{\boldsymbol{\gamma}}, q_{\boldsymbol{\beta}} \in \{0.1, 0.2, 0.3, 0.5\}$ and $\tau_{\boldsymbol{\gamma}}, \tau_{\boldsymbol{\beta}} \in \{0.8, 1, 1.3, 1.5\}$. The variable selection performance summarized from 100 random trials is provided in Table 5.

We can draw the following conclusions from Table 5. First, for $q_{\boldsymbol{\beta}}$ and $\tau_{\boldsymbol{\beta}}$ in the prior of the linear coefficients, the performance on both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ remains nearly identical across different values, indicating that VSM is highly robust to their choices. Second, varying $\tau_{\boldsymbol{\gamma}}$ in the prior of the logistic coefficients has only a minor influence on the predictive variable selection, and the prognostic variable selection remains perfect in all settings.

Third, the choice of $q_{\boldsymbol{\gamma}}$ plays a more substantial role in identifying the

Table 5: Performance on variable selection under structured mixture model settings in Section 5.1 with $n = 200$ and $p \in \{100, 500\}$. When one prior hyperparameter varies, other hyperparameters are set as recommended.

| Hyper | $p$ | Value | $\boldsymbol{\beta}$ TPR | $\boldsymbol{\beta}$ FDR | $\boldsymbol{\beta}$ F1 | $\boldsymbol{\beta}$ Ext | $\boldsymbol{\gamma}$ TPR | $\boldsymbol{\gamma}$ FDR | $\boldsymbol{\gamma}$ F1 | $\boldsymbol{\gamma}$ Ext |
|---|---|---|---|---|---|---|---|---|---|---|
| $q_{\boldsymbol{\gamma}}$ | 100 | 0.1 | 1 | 0 | 1 | 100% | 0.728 | 0.001 | 0.827 | 83% |
| | | 0.2 | 1 | 0 | 1 | 100% | 0.828 | 0.010 | 0.893 | 86% |
| | | 0.3 | 1 | 0 | 1 | 100% | 0.895 | 0.010 | 0.934 | 88% |
| | | 0.5 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | 500 | 0.1 | 1 | 0 | 1 | 100% | 0.693 | 0.008 | 0.799 | 56% |
| | | 0.2 | 1 | 0 | 1 | 100% | 0.790 | 0.025 | 0.859 | 59% |
| | | 0.3 | 1 | 0 | 1 | 100% | 0.863 | 0.033 | 0.901 | 65% |
| | | 0.5 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| $q_{\boldsymbol{\beta}}$ | 100 | 0.1 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 0.2 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 0.3 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 0.5 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | 500 | 0.1 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| | | 0.2 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| | | 0.3 | 1 | 0 | 1 | 100% | 0.903 | 0.128 | 0.880 | 64% |
| | | 0.5 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| $\tau_{\boldsymbol{\gamma}}$ | 100 | 0.8 | 1 | 0 | 1 | 100% | 0.958 | 0.062 | 0.943 | 87% |
| | | 1 | 1 | 0 | 1 | 100% | 0.953 | 0.051 | 0.947 | 88% |
| | | 1.3 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 1.5 | 1 | 0 | 1 | 100% | 0.955 | 0.033 | 0.957 | 90% |
| | 500 | 0.8 | 1 | 0 | 1 | 100% | 0.905 | 0.159 | 0.864 | 62% |
| | | 1 | 1 | 0 | 1 | 100% | 0.908 | 0.150 | 0.871 | 64% |
| | | 1.3 | 1 | 0 | 1 | 100% | 0.903 | 0.143 | 0.872 | 63% |
| | | 1.5 | 1 | 0 | 1 | 100% | 0.905 | 0.133 | 0.879 | 63% |
| $\tau_{\boldsymbol{\beta}}$ | 100 | 0.8 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 1 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 1.3 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | | 1.5 | 1 | 0 | 1 | 100% | 0.955 | 0.036 | 0.956 | 90% |
| | 500 | 0.8 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| | | 1 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| | | 1.3 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |
| | | 1.5 | 1 | 0 | 1 | 100% | 0.903 | 0.126 | 0.881 | 64% |

true predictive model. Large values of $q_{\boldsymbol{\gamma}}$, e.g., 0.5, tend to include more predictive covariates in the model, whereas smaller values of $q_{\boldsymbol{\gamma}}$, e.g., 0.1, induce more sparse results. Overall, the variations in the performance on $\boldsymbol{\gamma}$ are mild. The choice of $q_{\boldsymbol{\gamma}}$ exhibits a clear trade-off between TPR and FDR, and the recommended value of 0.5 achieves a satisfactory balance with high F1 and Ext scores across different $p$.

# S5 Additional Information on the Real Application

## S5.1 Posterior Inclusion Probabilities in the IWPC Dataset

The posterior inclusion probabilities of all covariates are shown in Figure 4.

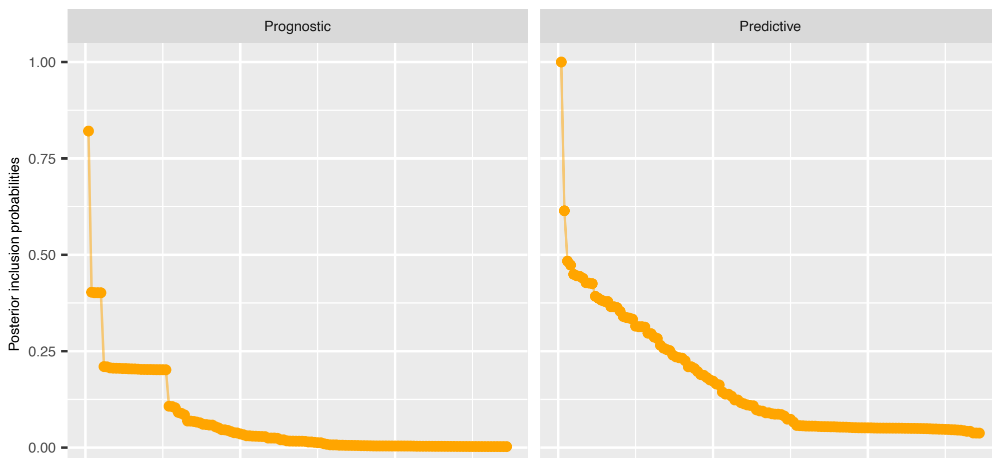


Figure 4: Posterior inclusion probabilities for the IWPC dataset with prognostic on the left and predictive on the right.

## S5.2 Additional Results on the ACTG 320 Study

The ACTG 320 dataset comprises 852 observations, with 423 patients receiving the three-drug regimen and 429 patients in the control group. The pre-treatment covariates include sex, injection-drug use ($dr$), hemophilia ($hemo$), weight ($wt$), and Karnofsky score ($Ks$), months of prior zidovudine ($zido$), age, log baseline CD4 counts ($L_c$), log baseline HIV-1 RNA concentration with base 10 ($L_r$), and indicators for African ($Afri$) and Hispanic ($Hisp$) ethnicity.

The variational posterior inclusion probabilities of all covariates estimated from VSM, including noise variables, are shown in Figure 5. Based on the results that $L_r$ and $L_c$ are selected as the only active prognostic and predictive covariates, the estimated model that includes only the identified active variables is given by

$$\begin{aligned}
Y \sim & \hat{\pi} N(-63.1 + 10.99L_r - 39.9L_c + 139.25t, 8.57^2) \\
& + (1-\hat{\pi})N(-63.1 + 10.99L_r - 39.9L_c - 7.63t, 8.57^2), \\
\log&[\hat{\pi}/(1-\hat{\pi})] = 3.49 + 1.06L_r - 3.38L_c.
\end{aligned}$$

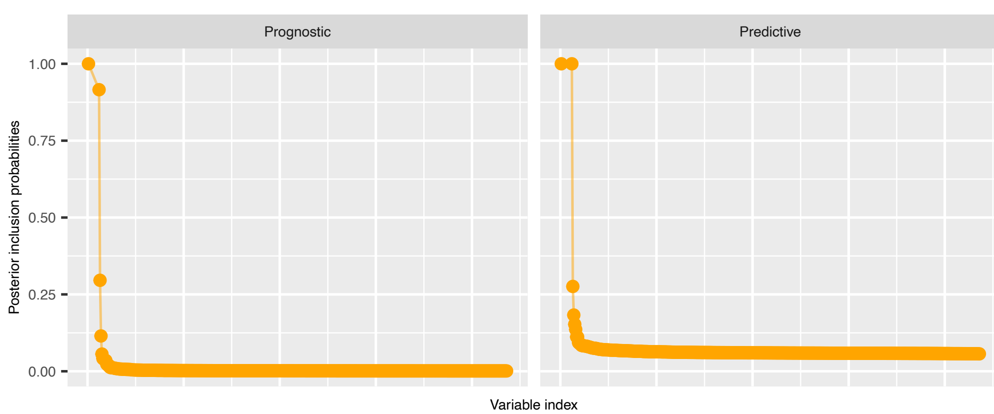


Figure 5: Posterior inclusion probabilities for the ACTG 320 study with prognostic on the left and predictive on the right.

To validate the identified subgroups, we empirically examine the treatment effects across different subgroups. Patients are divided into two subgroups based on the estimated $\hat{\pi}$ using a threshold of 0.5. Figure 6 presents box plots of the response under treatment and control conditions for each subgroup. In Group 1, the two box plots are distinctly separated, indicating a strong treatment effect. In contrast, in Group 2, the box plots exhibit substantial overlap, suggesting a weaker or negligible treatment effect. This difference in treatment effects between the two subgroups underscores the validity and meaningfulness of the identified subgroups.

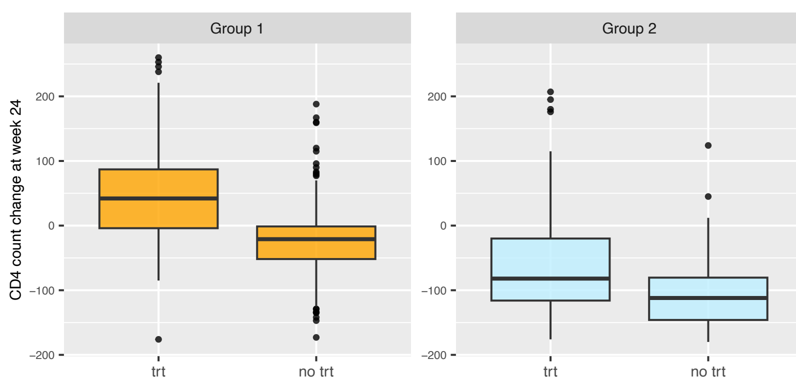


Figure 6: CD4 count change at week 24 under treatment and no treatment in two subgroups, where the subgroup membership is determined by the predicted subgroup proportion with a threshold of 0.5.

## Bibliography

Loh, W.-Y. (2002). Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica 12*(2), 361–386.

Ray, K. and B. Szabó (2022). Variational Bayes for high-dimensional linear regression with sparse priors. *Journal of the American Statistical Association 117*(539), 1270–1281.

Seibold, H., A. Zeileis, and T. Hothorn (2016). Model-based recursive partitioning for subgroup analyses. *The International Journal of Biostatistics 12*(1), 45–63.

Städler, N., P. Bühlmann, and S. Van De Geer (2010). $\ell 1$-penalization for mixture regression models. *Test 19*(2), 209–256.

Zhang, R., N. N. Narisetty, X. He, and J. Shen (2025). Bayesian variable selection on structured logistic-normal mixture models for subgroup analysis. *Electronic Journal of Statistics 19*(1), 2876–2922.