JOINT BAYESIAN VARIABLE AND DAG SELECTION CONSISTENCY FOR HIGH-DIMENSIONAL REGRESSION MODELS WITH NETWORK-STRUCTURED COVARIATES

University of Cincinnati and Inha University

Supplementary Material

S1 Proofs

In this section, we provide proofs for Lemma 1, Theorems 1 to 5, and Corollary 1.

Proof of Lemma 1. It follows from the hierarchical models in (3.1) to (3.6), we have

$$
\begin{aligned}
&\pi\left(\gamma, \mathscr{D} | Y, X\right) \\
=&\int \pi\left(Y | \gamma, \beta_\gamma\right) \prod_{i=1}^{n} \pi\left(X_i | (L, D)\right) \pi_{U,\alpha(\mathscr{D})}^{\Theta_{\mathscr{D}}}((L, D)) \\
&\times \pi\left(\beta_\gamma | \gamma\right) \pi(\gamma) \pi(\mathscr{D}) d\beta_\gamma d(L, D) \\
=&\pi(\gamma)\pi(\mathscr{D}) \int \pi\left(Y | \gamma, \beta_\gamma\right) \pi\left(\beta_\gamma | \gamma\right) d\beta_\gamma
\end{aligned}
$$

$$\times \int \prod_{i=1}^{n} \pi\left(X_i|(L,D)\right)\pi^{\Theta_{\mathscr{D}}}_{U,\alpha(\mathscr{D})}((L,D))d(L,D). \tag{S1.1}$$

First, note that by the conjugacy of the DAG–Wishart distribution, we have

$$\begin{aligned}&\int \prod_{i=1}^{n} \pi\left(X_i|(L,D)\right)\pi^{\Theta_{\mathscr{D}}}_{U,\alpha(\mathscr{D})}((L,D))d(L,D)\\ =&\frac{z_{\mathscr{D}}(U+X^TX, n+\alpha(\mathscr{D}))}{z_{\mathscr{D}}(U,\alpha(\mathscr{D}))},\end{aligned}$$

where $z_{\mathscr{D}}(.,.)$ is the normalized constant for the DAG–Wishart distribution.

Next, note that

$$\begin{aligned}&\int \pi\left(Y|\gamma,\beta_\gamma\right)\pi\left(\beta_\gamma|\gamma\right)d\beta_\gamma\\
\propto&\int (\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}(Y-X_\gamma\beta_\gamma)^T(Y-X_\gamma\beta_\gamma)\right\}\\
&\times(\tau^2\sigma^2)^{-\frac{1}{2}|\gamma|}\exp\left\{-\frac{1}{2\tau^2\sigma^2}\beta_\gamma^T\beta_\gamma\right\}d\beta_\gamma\\
\propto&(\tau^2)^{-\frac{1}{2}|\gamma|}(\sigma^2)^{-\frac{n+|\gamma|}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left(\beta_\gamma^T\left(X_\gamma^TX_\gamma+\frac{1}{\tau^2}I\right)\beta_\gamma-2\beta_\gamma^TX_\gamma^TY\right)\right\}d\beta_\gamma\\
&\times\exp\left\{-\frac{1}{2\sigma^2}Y^TY\right\}\\
\propto&(\tau^2)^{-\frac{1}{2}|\gamma|}(\sigma^2)^{-\frac{n+|\gamma|}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left(Y^TY-Y^TX_\gamma\left(X_\gamma^TX_\gamma+\frac{1}{\tau^2}I\right)^{-1}X_\gamma^TY\right)\right\}\\
&\times\int\exp\left\{-\frac{1}{2\sigma^2}\left(\beta_\gamma-\left(X_\gamma^TX_\gamma+\frac{1}{\tau^2}I\right)^{-1}X_\gamma^TY\right)^T\left(X_\gamma^TX_\gamma+\frac{1}{\tau^2}I\right)\right.\\
&\qquad\left.\times\left(\beta_\gamma-\left(X_\gamma^TX_\gamma+\frac{1}{\tau^2}I\right)^{-1}X_\gamma^TY\right)\right\}d\beta_\gamma\end{aligned}$$

$$\propto(\sigma^2)^{-\frac{n}{2}} \det\left(\tau^2 X_\gamma^T X_\gamma + I\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I_n + \tau^2 X_\gamma X_\gamma^T\right)^{-1} Y\right)\right\}, \tag{S1.2}$$

where the last term follows from the Woodbury matrix identity. Therefore, by (S1.1), under the proposed hierarchical model and known σ^2, we have

$$\begin{aligned}
&\pi\left(\gamma, \mathscr{D}|Y, X\right)\\
&\propto\pi(\gamma|\mathscr{D})\pi(\mathscr{D})\frac{z_{\mathscr{D}}(U + X^T X, n + \alpha(\mathscr{D}))}{z_{\mathscr{D}}(U, \alpha(\mathscr{D}))}\\
&\quad\times \det\left(\tau^2 X_\gamma^T X_\gamma + I_{|\gamma|}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I_n + \tau^2 X_\gamma X_\gamma^T\right)^{-1} Y\right)\right\},
\end{aligned} \tag{S1.3}$$

where $z_{\mathscr{D}}(\cdot,\cdot)$ is the normalized constant in the DAG–Wishart distribution.

□

Proof of Theorem 1. It follows from Assumption 2, Assumption 3 and model (3.6) that, for large enough $n > N$,

$$\begin{aligned}
\frac{\pi(\gamma_0|\mathscr{D})}{\pi(\gamma_0|\mathscr{D}_0)} &= \exp\left(b\gamma_0^T(G - G_0)\gamma_0\right)\\
&\leq \exp\left(b|\gamma_0|^2\right) \leq \exp\left(o\left(\log p/d^4\right)\right).
\end{aligned} \tag{S1.4}$$

Let $S = \frac{1}{n}X^T X$ denote the sample covariance matrix of X. It follows from (3.7), (S1.4), and Lemma 5.1 in Cao et al. (2019b) that

$$\frac{\pi(\gamma_0, \mathscr{D}|Y, X)}{\pi(\gamma_0, \mathscr{D}_0|Y, X)}$$

$$\leq \prod_{i=1}^{p} M \exp\left(o\left(\log p/d^4\right)\right) \left(\frac{\delta_2}{\delta_1}\right)^{\frac{d}{2}} n^{2c} \left(\sqrt{\frac{\delta_2}{n}}\frac{q}{1-q}\right)^{\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0)} \frac{|\tilde{S}^{\geq i}_{\mathscr{D}_0}|^{\frac{1}{2}}}{|\tilde{S}^{\geq i}_{\mathscr{D}}|^{\frac{1}{2}}} \frac{\left(\tilde{S}_{i|pa_i(\mathscr{D}_0)}\right)^{\frac{n+c_i(\mathscr{D}_0)-3}{2}}}{\left(\tilde{S}_{i|pa_i(\mathscr{D})}\right)^{\frac{n+c_i(\mathscr{D})-3}{2}}}$$

$$\triangleq \prod_{i=1}^{q} PR_i(\mathscr{D}, \mathscr{D}_0), \tag{S1.5}$$

where $c_i(\mathscr{D}) = \alpha_i(\mathscr{D}) - \nu_i(\mathscr{D}), c_i(\mathscr{D}_0) = \alpha_i(\mathscr{D}_0) - \nu_i(\mathscr{D}_0)$, $\tilde{S} = S + \frac{U}{n}$, $\tilde{S}_{i|pa_i(\mathscr{D})} = \tilde{S}_{ii} - (\tilde{S}^{>}_{\mathscr{D}\cdot i})^T(\tilde{S}^{>i}_{\mathscr{D}})^{-1}\tilde{S}^{>}_{\mathscr{D}\cdot i}$, and M is some large enough constant. Define the event E_n as

$$E_n = \left\{\|S - \Sigma_0\|_{\max} \geq c'\sqrt{\frac{\log p}{n}}\right\}. \tag{S1.6}$$

It follows from Lemma A.3 of Bickel and Levina (2008), Hanson-Wright inequality from Rudelson and Vershynin (2013) and the union-sum inequality, there exists constants c', m_1, m_2, such that

$$\bar{P}\left(\|\tilde{S} - \Sigma_0\|_{\max} \geq c'\sqrt{\frac{\log p}{n}}\right) \leq m_1 p^{2-m_2(c')^2/4} \to 0. \tag{S1.7}$$

For all the following analyses, we will restrict ourselves to the event E_n^c.

We now analyze the behavior of $PR_i(\mathscr{D}, \mathscr{D}_0)$ under different scenarios in a sequence of three lemmas (Lemmas 1-3). Recall that our goal is to find an upper bound (independent of $\mathscr{D}$ and i) for $PR_i(\mathscr{D}, \mathscr{D}_0)$, such that the upper bound converges to 0 as $n \to \infty$.

Lemma 1. *If $pa_i(\mathscr{D}) \supset pa_i(\mathscr{D}_0)$, then there exists N_1 (not depending on i or $\mathscr{D}$) such that for $n \geq N_1$ we have $PR_i(\mathscr{D}, \mathscr{D}_0) \leq (2p)^{-\frac{\alpha_1}{\kappa}(\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0))}$,* *for any constant* $\kappa > 1$.

Proof of Lemma 1. Since $pa_i(\mathscr{D}) \supset pa_i(\mathscr{D}_0)$, we can write $|\tilde{S}^{\geq i}_{\mathscr{D}}| = |\tilde{S}^{\geq i}_{\mathscr{D}_0}||R_{\tilde{S}^{\geq i}_{\mathscr{D}_0}}|$. Here $R_{\tilde{S}^{\geq i}_{\mathscr{D}_0}}$ is the Schur complement of $\tilde{S}^{\geq i}_{\mathscr{D}_0}$, defined by $R_{\tilde{S}^{\geq i}_{\mathscr{D}_0}} = D - B^T \left(\tilde{S}^{\geq i}_{\mathscr{D}_0}\right)^{-1} B$, for appropriate sub matrices B and D of $\tilde{S}^{\geq i}_{\mathscr{D}}$. Since $\tilde{S}^{\geq i}_{\mathscr{D}} \geq \left(\frac{U}{n}\right)^{\geq i}_{\mathscr{D}}$ [1] and $R^{-1}_{\tilde{S}^{\geq i}_{\mathscr{D}_0}}$ is a principal submatrix of $\left(\tilde{S}^{\geq i}_{\mathscr{D}}\right)^{-1}$, the largest eigenvalue of $R^{-1}_{\tilde{S}^{\geq i}_{\mathscr{D}_0}}$ is bounded above by $\frac{n}{\delta_2}$. Therefore,

$$\left(\frac{|\tilde{S}^{\geq i}_{\mathscr{D}_0}|}{|\tilde{S}^{\geq j}_{\mathscr{D}}|}\right)^{\frac{1}{2}} = |R^{-1}_{\tilde{S}^{\geq j}_{\mathscr{D}_0}}|^{1/2} \leq \left(\sqrt{\frac{n}{\delta_2}}\right)^{\nu_i(\mathscr{D}) - \nu_i(\mathscr{D}_0)}. \tag{S1.8}$$

Denote $S_{j|\mathscr{D}_j} = S_{jj} - (S^{>}_{\mathscr{D}\cdot j})^T (S^{>j}_{\mathscr{D}})^{-1} S^{>}_{\mathscr{D}\cdot j}$. It immediately follows that

$$\tilde{S}_{i|pa_i(\mathscr{D})} \geq S_{i|pa_i(\mathscr{D})}. \tag{S1.9}$$

Since we are restricting ourselves to the event E^c_n, it follows from (S1.6) that

$$||S^{\geq i}_{\mathscr{D}_0} - (\Sigma_0)^{\geq i}_{\mathscr{D}_0}||_{(2,2)} \leq (\nu_i(\mathscr{D}_0) + 1)c' \sqrt{\frac{\log p}{n}}.$$

Therefore,

$$\begin{aligned}
&||(S^{\geq i}_{\mathscr{D}_0})^{-1} - ((\Sigma_0)^{\geq i}_{\mathscr{D}_0})^{-1}||_{(2,2)} \\
=&||(S^{\geq i}_{\mathscr{D}_0})^{-1}||_{(2,2)} ||S^{\geq i}_{\mathscr{D}_0} - (\Sigma_0)^{\geq i}_{\mathscr{D}_0}||_{(2,2)} ||((\Sigma_0)^{\geq i}_{\mathscr{D}_0})^{-1}||_{(2,2)} \\
\leq&(||(S^{\geq i}_{\mathscr{D}_0})^{-1} - ((\Sigma_0)^{\geq i}_{\mathscr{D}_0})^{-1}||_{(2,2)} + \frac{1}{\epsilon_0})(\nu_i(\mathscr{D}_0) + 1)c' \sqrt{\frac{\log p}{n}}. \qquad (S1.10)
\end{aligned}$$

[1] For matrices A and B, we say $A \geq B$ if $A - B$ is positive semi-definite

Recall $d = \max_{1\le i\le p-1} \nu_i(\mathscr{D}_0)$. By the assumption that $d\sqrt{\frac{\log p}{n}} \to 0$ and (S1.10), for large enough n, we have

$$||(S_{\mathscr{D}_0}^{\ge i})^{-1} - ((\Sigma_0)_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)} \le \frac{4c\prime}{\epsilon_0} d\sqrt{\frac{\log p}{n}} = o(1) \tag{S1.11}$$

and

$$\frac{1}{S_{i|pa_i(\mathscr{D}_0)}} = \left[(S_{\mathscr{D}_0}^{\ge i})^{-1}\right]_{ii} \ge \frac{\epsilon_0}{2}. \tag{S1.12}$$

Note that for any $\mathscr{D}$, $||\tilde{S}_{\mathscr{D}}^{\ge i} - S_{\mathscr{D}}^{\ge i}||_{\max} \le \frac{\delta_2}{n}$ gives us $||\tilde{S}_{\mathscr{D}_0}^{\ge i} - S_{\mathscr{D}_0}^{\ge i}||_{(2,2)} \le (\nu_i(\mathscr{D}_0) + 1)\frac{\delta_2}{n}$. Therefore,

$$\begin{aligned}
&||(\tilde{S}_{\mathscr{D}_0}^{\ge i})^{-1} - (S_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)} \\
=&||(\tilde{S}_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)}||\tilde{S}_{\mathscr{D}_0}^{\ge i} - S_{\mathscr{D}_0}^{\ge i}||_{(2,2)}||(S_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)} \\
\le&(||(\tilde{S}_{\mathscr{D}_0}^{\ge i})^{-1} - (S_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)} + ||(S_{\mathscr{D}_0}^{\ge i})^{-1} - ((\Sigma_0)_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)} + \frac{1}{\epsilon_0}) \times (\frac{1}{\epsilon_0} + o(1)) \\
&\times (pa_i(\mathscr{D}_0) + 1)\frac{\delta_2}{n}.
\end{aligned} \tag{S1.13}$$

Following from (S1.11), (S1.12), and $\frac{d}{n} \to 0$, for large enough n, (S1.13) yields

$$||(\tilde{S}_{\mathscr{D}_0}^{\ge i})^{-1} - (S_{\mathscr{D}_0}^{\ge i})^{-1}||_{(2,2)} \le \frac{8\delta_2}{\epsilon_0^2}\frac{d}{n} \text{ and } \frac{1}{\tilde{S}_{i|pa_i(\mathscr{D})}} = \left[(\tilde{S}_{\mathscr{D}}^{\ge i})^{-1}\right]_{ii} \ge \frac{\epsilon_0}{4}. \tag{S1.14}$$

Hence, it follow from (S1.14) and (S1.12) that,

$$\left|\frac{1}{S_{i|pa_i(\mathscr{D}_0)}} - \frac{1}{\tilde{S}_{i|pa_i(\mathscr{D}_0)}}\right| \le \frac{8\delta_2}{\epsilon_0^2}\frac{d}{n} \tag{S1.15}$$

and

$$|S_{i|pa_i(\mathscr{D}_0)} - \tilde{S}_{i|pa_i(\mathscr{D}_0)}| \le c_1\frac{d}{n}, \tag{S1.16}$$

where $c_1 = 64\delta_2/\epsilon_0^4$ is a constant.

Further note that when $pa_i(\mathscr{D}_0) \subset pa_i(\mathscr{D})$, $n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D})} \sim \chi^2_{n-\nu_i(\mathscr{D})}$ and $n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D}_0)} \overset{d}{=} n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D})} \oplus \chi^2_{\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0)}$ under the true model. By Lemma 4.1 in (Cao et al., 2019a), we get

$$P\left[\left|n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D})} - (n - \nu_i(\mathscr{D}))\right| > \sqrt{(n-\nu_i(\mathscr{D}))\log p}\right] \le 2p^{-\frac{1}{8}} \to 0, \tag{S1.17}$$

and

$$\begin{aligned} P\Bigg[& \left|n(D_0)_{ii}^{-1} S_{j|pa_i(\mathscr{D}_0)} - n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D})} - (\nu_i(\mathscr{D}) - \nu_i(\mathscr{D}_0))\right| \\ & > \sqrt{(\nu_i(\mathscr{D}) - \nu_i(\mathscr{D}_0))\log p}\Bigg] \\ \le & 2p^{-\frac{1}{8}} \to 0, \end{aligned} \tag{S1.18}$$

Following from (S1.8), (S1.9), (S1.14), (S1.16), (S1.17), (S1.18), and Assumption 4, for larger enough $n > N_1$ and some constant M', we have

$$\begin{aligned} & PR_i(\mathscr{D}, \mathscr{D}_0) \\ \le & M' \exp\left(o\left(\log p/d^4\right)\right)\left(\frac{\delta_2}{\delta_1}\right)^{\frac{d}{2}} n^{2c}(2p^{-\alpha_1})^{\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0)} \\ & \times \left(1 + \frac{n(D_0)_{ii}^{-1} S_{j|pa_i(\mathscr{D})} - n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D}_0)} + c_1 \frac{d}{(D_0)_{ii}}}{n(D_0)_{ii}^{-1} S_{i|pa_i(\mathscr{D})}}\right)^{\frac{n+c-3}{2}} \\ \le & M' \exp\left(o\left(\log p/d^4\right)\right)\left(\frac{\delta_2}{\delta_1}\right)^{\frac{d}{2}} n^{2c}(2p^{-\alpha_1})^{\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0)} \\ & \times \exp\left\{\frac{\nu_i(\mathscr{D}) - \nu_i(\mathscr{D}_0) + \sqrt{\nu_i(\mathscr{D}) - \nu_i(\mathscr{D}_0)\log p} + c_1 \frac{d}{(D_0)_{ii}}}{n - \nu_i(\mathscr{D}) - \sqrt{(n-\nu_i(\mathscr{D}))\log p}} \frac{n+c-3}{2}\right\} \end{aligned}$$

$$\leq (2p)^{-\frac{\alpha_1}{\kappa}(\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0))}, \text{ for any constant } \kappa > 1. \tag{S1.19}$$

The second inequality follows from $\frac{d}{\log p} \to 0$ and $\frac{\nu_i(\mathscr{D})}{n} \to 0$, as $n \to \infty$ and $\frac{\epsilon_0}{2} \leq (D_0)_{ii} \leq \frac{2}{\epsilon_0}$. □

Lemma 2. *If $pa_i(\mathscr{D}) \subset pa_i(\mathscr{D}_0)$, then there exists N_2 (not depending on i or $\mathscr{D}$) such that for $n \geq N_1$ we have $PR_i(\mathscr{D}, \mathscr{D}_0) \leq p^{-\frac{2\alpha_1}{\kappa}d}$.*

Proof of Lemma 2. Now we move to discuss the scenario when $pa_i(\mathscr{D})$ is a subset of $pa_i(\mathscr{D}_0)$, i.e., $pa_i(\mathscr{D}) \subset pa_i(\mathscr{D}_0)$. Since $pa_i(\mathscr{D}_0) \supset pa_i(\mathscr{D})$, we can write $|\tilde{S}^{\geq i}_{\mathscr{D}_0}| = |\tilde{S}^{\geq i}_{\mathscr{D}}||R_{\tilde{S}^{\geq i}_{\mathscr{D}}}|$, where $R_{\tilde{S}^{\geq i}_{\mathscr{D}}}$ denotes the Schur complement of $\tilde{S}^{\geq i}_{\mathscr{D}}$, defined by $R_{\tilde{S}^{\geq i}_{\mathscr{D}}} = \tilde{D} - \tilde{B}^T(\tilde{S}^{\geq i}_{\mathscr{D}})^{-1}\tilde{B}$ for appropriate sub matrices $\tilde{B}$ and $\tilde{D}$ of $\tilde{S}^{\geq i}_{\mathscr{D}_0}$.

It follows by (S1.10) that if restrict to E_n^c, we have $||(\tilde{S}^{\geq i}_{\mathscr{D}_0})^{-1} - ((\Sigma_0)^{\geq i}_{\mathscr{D}_0})^{-1}||_{(2,2)} \leq \frac{4c'}{\epsilon_0}d\sqrt{\frac{\log p}{n}}$ and $||R^{-1}_{\tilde{S}^{\geq i}_{\mathscr{D}}} - R^{-1}_{(\Sigma_0)^{\geq i}_{\mathscr{D}}}||_{(2,2)} \leq \frac{4c'}{\epsilon_0}d\sqrt{\frac{\log p}{n}}$, for $n > N_2'$, where $R_{(\Sigma_0)^{\geq i}_{\mathscr{D}}}$ represents the Schur complement of $(\Sigma_0)^{\geq i}_{\mathscr{D}}$ defined by $R_{(\Sigma_0)^{\geq i}_{\mathscr{D}}} = \bar{D} - \bar{B}^T((\Sigma_0)^{\geq i}_{\mathscr{D}})^{-1}\bar{B}$ for appropriate sub matrices $\bar{B}$ and $\bar{D}$ of $(\Sigma_0)^{\geq i}_{\mathscr{D}_0}$. Hence, there exists N_2'' such that

$$\left(\frac{|\tilde{S}^{\geq i}_{\mathscr{D}_0}|}{|\tilde{S}^{\geq i}_{\mathscr{D}}|}\right)^{\frac{1}{2}} = \frac{1}{|R^{-1}_{\tilde{S}^{\geq i}_{\mathscr{D}}}|^{1/2}} \leq \frac{1}{\left(\lambda_{\min}\left(R^{-1}_{(\Sigma_0)^{\geq i}_{\mathscr{D}}}\right) - K\frac{d}{\epsilon_0^3}\sqrt{\frac{\log p}{n}}\right)^{\frac{\nu_i(\mathscr{D}_0)-\nu_i(\mathscr{D})}{2}}}$$

$$\leq \left(\frac{1}{\epsilon_0/2}\right)^{\frac{\nu_i(\mathscr{D}_0)-\nu_i(\mathscr{D})}{2}} \quad \text{for } n > N_2'. \tag{S1.20}$$

Since $pa_i(\mathscr{D}) \subset pa_i(\mathscr{D}_0)$, we get $\tilde{S}_{i|pa_i(\mathscr{D}_0)} \leq \tilde{S}_{i|pa_i(\mathscr{D})}$.

Let $K_1 = 4c'/\epsilon_0^3$. By (S1.5) and $2 < c_i(\mathscr{D}), c_i(\mathscr{D}_0) < c$, it follows that there exists N_2''' such that for $n \geq N_2'''$, we get

$$\begin{aligned}
& PR_i(\mathscr{D}, \mathscr{D}_0) \\
\leq & M' \exp\left(o\left(\log p/d^4\right)\right)\left(\frac{\delta_2}{\delta_1}\right)^{\frac{d}{2}} n^{2c}\left(\sqrt{\frac{2n}{\delta_2\epsilon_0}}q^{-1}\right)^{\nu_i(\mathscr{D}_0)-\nu_i(\mathscr{D})}\left(\frac{\frac{1}{(\Sigma_0)_{i|pa_i(\mathscr{D})}} + K_1 d\sqrt{\frac{\log p}{n}}}{\frac{1}{(\Sigma_0)_{i|pa_i(\mathscr{D}_0)}} - K_1 d\sqrt{\frac{\log p}{n}}}\right)^{\frac{n+2-3}{2}} \\
\leq & \left(\exp\left\{\frac{2\log M' + o\left(\log p/d^4\right) + d\log\left(\frac{2\delta_2}{\epsilon_0\delta_1}\right) + 4c\log n}{n-1} + \frac{2\log(q^{-1}\sqrt{n})\left(\nu_i(\mathscr{D}_0) - \nu_i(\mathscr{D})\right)}{n-1}\right\}\right)^{\frac{n-1}{2}} \\
& \times\left(1 + \frac{\left(\frac{1}{(\Sigma_0)_{i|pa_i(\mathscr{D}_0)}} - \frac{1}{(\Sigma_0)_{i|pa_i(\mathscr{D})}}\right) - 2K_1 d\sqrt{\frac{\log p}{n}}}{\frac{1}{(\Sigma_0)_{i|pa_i(\mathscr{D})}} + K_1 d\sqrt{\frac{\log p}{n}}}\right)^{-\frac{n-1}{2}} \qquad \text{(S1.21)}
\end{aligned}$$

It then follows from Proposition 5.2 in Cao et al. (2019b) that,

$$\begin{aligned}
& PR_i(\mathscr{D}, \mathscr{D}_0) \\
\leq & \left(\exp\left\{\frac{2\log M' + o\left(\log p/d^4\right) + d\log\left(\frac{2\delta_2}{\epsilon_0\delta_1}\right) + 4c\log n}{n-1} + \frac{2\log(p^{\alpha_1}\sqrt{n})\left(\nu_i(\mathscr{D}_0) - \nu_i(\mathscr{D})\right)}{n-1}\right\}\right)^{\frac{n-1}{2}} \\
& \times\left(1 + \frac{\epsilon_0 s_n^2(\nu_i(\mathscr{D}_0) - \nu_i(\mathscr{D})) - 2K_1 d\sqrt{\frac{\log p}{n}}}{2/\epsilon_0}\right)^{-\frac{n-1}{2}}. \qquad \text{(S1.22)}
\end{aligned}$$

Note that $\frac{d\log p + d\log n}{ns_n^2} \to 0$ and $\frac{d\sqrt{\frac{\log p}{n}}}{s_n^2} \to 0$ as $n \to \infty$. Since $e^x \leq 1 + 2x$ for $x < \frac{1}{2}$, there exists N_2'''' such that for $n \geq N_2''''$,

$$\frac{\epsilon_0 s_n^2(\nu_i(\mathscr{D}_0) - \nu_i(\mathscr{D})) - 2K_1 d\sqrt{\frac{\log p}{n}}}{2/\epsilon_0} \geq \frac{\epsilon_0 s_n^2}{2},$$

and

$$\exp\left\{\frac{2\log M' + o\left(\log p/d^4\right) + d\log\left(\frac{2\delta_2}{\epsilon_0\delta_1}\right) + 4c\log n}{n-1} + \frac{2\log(p^{\alpha_1}\sqrt{n})\left(\nu_i(\mathscr{D}_0) - \nu_i(\mathscr{D})\right)}{n-1}\right\}$$

$$\leq 1 + \frac{\epsilon_0^2 s_n^2}{8}.$$

It follows by (S1.21) and the above arguments that

$$PR_i(\mathscr{D}, \mathscr{D}_0) \leq \left(\frac{1 + \frac{\epsilon_0^2}{8}s_n^2}{1 + \frac{\epsilon_0^2}{4}s_n^2}\right)^{\frac{n-1}{2}}$$

for $n \geq \max(N_2', N_2'', N_2''', N_2'''')$. Since there exist a $(L_0)_{ji}$ such that $s_n^2 \leq (L_0)_{ji}^2 \leq \frac{1}{\epsilon_0}\left(\frac{[(L_0)_{ji}]^2}{(D_0)_{ii}}\right) \leq \frac{(\Omega_0)_{jj}}{\epsilon_0} \leq \frac{1}{\epsilon_0^2}$ and $\epsilon_0^2 s_n^2 \leq 1$, it follows that there exists $N_2 = \max(N_2', N_2'', N_2''', N_2'''')$ such that for $n \geq N_2$, such that

$$PR_i(\mathscr{D}, \mathscr{D}_0) \leq \left(1 - \frac{\frac{\epsilon_0^2}{8}s_n^2}{1 + \frac{\epsilon_0^2}{4}s_n^2}\right)^{\frac{n-1}{2}} \leq \exp\left\{-\left(\frac{\frac{\epsilon_0^2}{8}s_n^2}{1 + \frac{\epsilon_0^2}{4}s_n^2}\right)\left(\frac{n-1}{2}\right)\right\}$$

$$\leq e^{-\frac{1}{10}\epsilon_0^2 s_n^2\left(\frac{n-1}{2}\right)} \leq p^{-\frac{2\alpha_1}{\kappa}d}. \qquad \text{(S1.23)}$$

The last inequality follows from $\frac{d\log p}{ns_n^2} \to 0$, as $n \to \infty$. □

Lemma 3. *If $pa_i(\mathscr{D})$ is not necessarily a superset or a subset of $pa_i(\mathscr{D}_0)$, i.e. $pa_i(\mathscr{D}_0) \neq pa_i(\mathscr{D})$, $pa_i(\mathscr{D}_0) \nsubseteq pa_i(\mathscr{D})$, and $pa_i(\mathscr{D}_0) \nsupseteq pa_i(\mathscr{D})$, then there exists N_3 (not depending on i or $\mathscr{D}$) such that for $n \geq N_3$ we have $PR_i(\mathscr{D}, \mathscr{D}_0) \leq (2p)^{-\frac{\alpha_1}{\kappa}\nu_i(\mathscr{D})}$.*

Proof of Lemma 3. Next consider the scenario when $pa_i(\mathscr{D})$ is not necessarily a superset or a subset of $pa_i(\mathscr{D}_0)$, i.e. $pa_i(\mathscr{D}_0) \neq pa_i(\mathscr{D})$, $pa_i(\mathscr{D}_0) \nsubseteq$

$pa_i(\mathscr{D})$, and $pa_i(\mathscr{D}_0) \not\supseteq pa_i(\mathscr{D})$. Let $\mathscr{D}^*$ be an arbitrary DAG with $pa_i(\mathscr{D}^*) = pa_i(\mathscr{D}) \cap pa_i(\mathscr{D}_0)$. Immediately we get $pa_i(\mathscr{D}^*) \subset pa_i(\mathscr{D}_0)$ and $pa_i(\mathscr{D}^*) \subset pa_i(\mathscr{D})$. It follows from (S1.5) that

$$\begin{aligned}
&PR_i(\mathscr{D}, \mathscr{D}_0) \\
\leq &M \exp\left(o\left(\log p/d^4\right)\right) \left(\frac{\delta_2}{\delta_1}\right)^{\frac{d}{2}} n^{2c} \left(\sqrt{\frac{\delta_2}{n}} \frac{q}{1-q}\right)^{\nu_i(\mathscr{D})-\nu_i(\mathscr{D}^*)} \frac{|\tilde{S}^{\geq i}_{\mathscr{D}^*}|^{\frac{1}{2}}}{|\tilde{S}^{\geq i}_{\mathscr{D}}|^{\frac{1}{2}}} \frac{\left(\tilde{S}_{i|pa_i(\mathscr{D}^*)}\right)^{\frac{n+c_i(\mathscr{D}^*)-3}{2}}}{\left(\tilde{S}_{i|pa_i(\mathscr{D})}\right)^{\frac{n+c_i(\mathscr{D})-3}{2}}} \\
&\times \left(\frac{\delta_2}{\delta_1}\right)^{\frac{d}{2}} n^{2c} \left(\sqrt{\frac{\delta_2}{n}} \frac{q}{1-q}\right)^{\nu_i(\mathscr{D}^*)-\nu_i(\mathscr{D}_0)} \frac{|\tilde{S}^{\geq i}_{\mathscr{D}_0}|^{\frac{1}{2}}}{|\tilde{S}^{\geq i}_{\mathscr{D}^*}|^{\frac{1}{2}}} \frac{\left(\tilde{S}_{i|pa_i(\mathscr{D}_0)}\right)^{\frac{n+c_i(\mathscr{D}_0)-3}{2}}}{\left(\tilde{S}_{i|pa_i(\mathscr{D}^*)}\right)^{\frac{n+c_i(\mathscr{D}^*)-3}{2}}} \\
\leq &PR_i(\mathscr{D}, \mathscr{D}^*) \times PR_i(\mathscr{D}^*, \mathscr{D}_0). \qquad \text{(S1.24)}
\end{aligned}$$

Note that $pa_i(\mathscr{D}^*) \subset pa_i(\mathscr{D})$. It follows from (S1.19) that

$$PR_i(\mathscr{D}, \mathscr{D}^*) \leq (2p)^{-\frac{\alpha_1}{\kappa}(\nu_i(\mathscr{D})-\nu_i(\mathscr{D}^*))}, \text{ for any constant } \kappa > 1 \text{ and } n \geq N_1. \qquad \text{(S1.25)}$$

Following from (S1.23) and the fact that $pa_i(\mathscr{D}^*) \subset pa_i(\mathscr{D}_0)$, we have

$$PR_i(\mathscr{D}, \mathscr{D}^*) \leq p^{-\frac{2\alpha_1}{\kappa}d}, \text{ for } n \geq N_2. \qquad \text{(S1.26)}$$

By (S1.24) and $\nu_i(\mathscr{D}^*) < d$, we get

$$\begin{aligned}
PR_i(\mathscr{D}, \mathscr{D}^*) &\leq (2p)^{-\frac{\alpha_1}{\kappa}(\nu_i(\mathscr{D})-\nu_i(\mathscr{D}^*))} p^{-\frac{2\alpha_1}{\kappa}d} \\
&< (2p)^{-\frac{\alpha_1}{\kappa}\nu_i(\mathscr{D})}, \text{for } n > N_3 = \max\{N_1, N_2\}. \qquad \text{(S1.27)}
\end{aligned}$$

□

The proof of Theorem 1 immediately follows after these three lemmas. For any $\mathscr{D} \neq \mathscr{D}_0$, there exists at least one $1 \leq i \leq p-1$, such that $pa_i(\mathscr{D}) \neq pa_i(\mathscr{D}_0)$. It follows from Lemmas 1-3 that, for large enough $n > N_3$, under the true variable indicator γ_0,

$$\max_{\mathscr{D} \neq \mathscr{D}_0} \frac{\pi(\gamma_0, \mathscr{D}|Y, X)}{\pi(\gamma_0, \mathscr{D}_0|Y, X)} \xrightarrow{\bar{P}} 0, \text{ as } n \to \infty. \tag{S1.28}$$

□

Proof of Theorem 2. Now for the fixed $\mathscr{D}$ case, it follows from Lemma 1 and model (3.6) that

$$\begin{aligned} &\frac{\pi(\gamma, \mathscr{D}|Y, X)}{\pi(\gamma_0, \mathscr{D}|Y, X)} \\ =&\frac{\exp\left(-a1^T\gamma + b\gamma^T G\gamma\right)}{\exp\left(-a1^T\gamma_0 + b\gamma_0^T G\gamma_0\right)} \frac{\det\left(\tau^2 X_\gamma^T X_\gamma + I_{|\gamma|}\right)^{-\frac{1}{2}}}{\det\left(\tau^2 X_{\gamma_0}^T X_{\gamma_0} + I_{|\gamma_0|}\right)^{-\frac{1}{2}}} \\ &\times \frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I + \tau^2 X_\gamma X_\gamma^T\right)^{-1} Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I + \tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1} Y\right)\right\}}. \end{aligned} \tag{S1.29}$$

For any model γ presenting the variable space, denote $Q_\gamma = \det\left(\tau^2 X_\gamma^T X_\gamma + I_{|\gamma|}\right)^{-\frac{1}{2}}$, $P_\gamma = X_\gamma(X_\gamma^T X_\gamma)^{-1} X_\gamma^T$,

$$R_\gamma^* = Y^T\left(I_n + \tau^2 X_\gamma X_\gamma^T\right)^{-1} Y \text{ and } R_\gamma = Y^T\left(I_n - P_\gamma\right) Y.$$

Our method of proving variable selection consistency involves utilizing properties of R_γ and approximating R_γ^* and $R_{\gamma_0}^*$ with R_γ and R_{γ_0} respectively.

Using the Woodbury matrix identity, we have

$$R^*_{\gamma_0} = Y^T\left(I_n + \tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1} Y = Y^T\left(I_n - X_{\gamma_0}\left(I_n/\tau^2 + X_{\gamma_0}^T X_{\gamma_0}\right)^{-1} X_{\gamma_0}^T\right) Y.$$

Note that for $1 \le i \le p$,

$$R^*_{\gamma_0} = Y^T\left(I_n + \tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1} Y = Y^T\left(I_n - X_{\gamma_0}\left(I_n/\tau^2 + X_{\gamma_0}^T X_{\gamma_0}\right)^{-1} X_{\gamma_0}^T\right) Y$$

and $R_{\gamma_0} = Y^T\left(I_n - X_\gamma\left(X_\gamma^T X_\gamma\right)^{-1} X_\gamma^T\right) Y$. It follows that

$$R^*_{\gamma_0} - R_{\gamma_0} \ge 0 \tag{S1.30}$$

and

$$\begin{aligned} & R^*_{\gamma_0} - R_{\gamma_0} \\ =& Y^T X_{\gamma_0} (X_{\gamma_0}^T X_{\gamma_0})^{-\frac{1}{2}} \left(I_n - \left(I_n + (X_{\gamma_0}^T X_{\gamma_0})^{-1}/\tau^2\right)^{-1}\right) (X_{\gamma_0}^T X_{\gamma_0})^{-\frac{1}{2}} X_{\gamma_0}^T Y \\ \le& \frac{1}{1 + n\epsilon_0\tau^2/2} Y^T P_{\gamma_0} Y. \end{aligned} \tag{S1.31}$$

Note that $\frac{R_{\gamma_0}}{\sigma^2} \sim \chi^2_{n-|\gamma_0|}$ and $\frac{Y^T P_{\gamma_0} Y}{\sigma^2} \sim \chi^2_{|\gamma_0|}\left(\frac{\beta_0^T X_{\gamma_0}^T X_{\gamma_0}\beta_0}{\sigma^2}\right)$. Here we denote χ^2_m as the centered chi-squared distribution with degrees of freedom $m > 0$ and $\chi^2_m(\lambda)$ as the noncentral chi-squared distribution with noncentral parameter λ. It follows from Lemmas 4.1 and 4.2 in Cao et al. (2019a), and Assumption 2 that

$$P\left[\left|\frac{R_{\gamma_0}}{\sigma^2} - (n - |\gamma_0|)\right| > \sqrt{(n - |\gamma_0|)\log p}\right] \le 2p^{-\frac{1}{8}} \to 0, \quad \text{as } n \to \infty, \tag{S1.32}$$

and

$$
\begin{aligned}
&P\left[\frac{Y^T P_{\gamma_0} Y}{\sigma^2} - \left(|\gamma_0| + \frac{\beta_0^T X_{\gamma_0}^T X_{\gamma_0}\beta_0}{\sigma^2}\right) > n\log p - |\gamma_0| - \frac{\beta_0^T X_{\gamma_0}^T X_{\gamma_0}\beta_0}{\sigma^2}\right] \\
&\leq \exp\left\{-\frac{|\gamma_0|}{2}\left\{\frac{n\log p}{\frac{\beta_0^T X_{\gamma_0}^T X_{\gamma_0}\beta_0}{\sigma^2}} - \log\left(1 + \frac{n\log p}{\frac{\beta_0^T X_{\gamma_0}^T X_{\gamma_0}\beta_0}{\sigma^2}}\right)\right\}\right\} \\
&\leq \exp\left\{-\frac{|\gamma_0|}{4}\left\{\frac{\log p}{1 + \frac{\epsilon_0}{2\sigma^2}\beta_0^T\beta_0}\right\}\right\} \preceq \exp\left\{-c'\sqrt{\log p}\right\} \to 0, \quad \text{as } n \to \infty.
\end{aligned} \tag{S1.33}
$$

Further note that,

$$
\begin{aligned}
R_\gamma^* - R_\gamma =& Y^T X_\gamma (X_\gamma^T X_\gamma)^{-\frac{1}{2}} \left(I_n - \left(I_n + (X_\gamma^T X_\gamma)^{-1}/\tau^2\right)^{-1}\right)(X_\gamma^T X_\gamma)^{-\frac{1}{2}} X_\gamma^T Y \\
\geq& \frac{\epsilon_0}{\epsilon_0 + 2n\tau^2} Y^T P_\gamma Y.
\end{aligned} \tag{S1.34}
$$

In the case when all the active elements of the true model γ_0 are contained in model γ, it follows that $\frac{R_{\gamma_0} - R_\gamma}{\sigma^2} \sim \chi^2_{|\gamma|-|\gamma_0|}$. Again, by Lemma 4.1 in Cao et al. (2019a), it follows that

$$
\begin{aligned}
&P\left[\left|\frac{Y^T(P_\gamma - P_{\gamma_0})Y}{\sigma^2} - (|\gamma| - |\gamma_0|)\right| > \sqrt{(|\gamma| - |\gamma_0|)\log p}\right] \\
&\leq 2p^{-\frac{1}{8}} \to 0,
\end{aligned} \tag{S1.35}
$$

and

$$
\begin{aligned}
&P\left[\left|\frac{R_{\gamma_0} - R_\gamma}{\sigma^2} - (|\gamma| - |\gamma_0|)\right| > \sqrt{(|\gamma| - |\gamma_0|)\log p}\right] \\
&\leq 2p^{-\frac{1}{8}} \to 0,
\end{aligned} \tag{S1.36}
$$

as $n \to \infty$. Hence, by (S1.29), (S1.31), (S1.33), (S1.36) and $R^*_\gamma - R_\gamma \geq 0$, we have

$$\begin{aligned}
&\frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2 X_\gamma X_\gamma^T\right)^{-1}Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1}Y\right)\right\}}\\
=&\exp\left\{\frac{1}{2\sigma^2}(R^*_{\gamma_0} - R^*_\gamma)\right\}\\
\leq&\exp\left\{\frac{1}{2\sigma^2}\left(\left(R_{\gamma_0} + \frac{1}{1+n\epsilon_0\tau^2/2}Y^T P_{\gamma_0} Y\right) - R_\gamma\right)\right\}\\
\leq&\exp\left\{\frac{1}{2\sigma^2}\left(|\gamma| - |\gamma_0| + \sqrt{(|\gamma|-|\gamma_0|)\log p} + \frac{1}{1+n\epsilon_0\tau^2/2}n\log p\right)\right\}.
\end{aligned} \tag{S1.37}$$

Next, note that it follows from $\gamma \supset \gamma_0$, Assumption 2 and the arguments leading up to (S1.8) that for large enough n (not depending on γ, $\mathscr{D}$), $\frac{Q_\gamma}{Q_{\gamma_0}} \leq n^{-\frac{|\gamma|-|\gamma_0|}{2}}\left(n\tau^2\right)^{\frac{|\gamma|-|\gamma_0|}{2}}$.

Therefore, it follows from Assumption 3, $\gamma \supset \gamma_0$, (S1.29) and (S1.37) that, for large enough $n \geq N_4$,

$$\begin{aligned}
&\frac{\pi\left(\gamma, \mathscr{D}|Y,X\right)}{\pi\left(\gamma_0, \mathscr{D}|Y,X\right)}\\
=&\frac{\exp\left(-a1^T\gamma + b\gamma^T G\gamma\right)}{\exp\left(-a1^T\gamma_0 + b\gamma_0^T G\gamma_0\right)}\frac{Q_\gamma}{Q_{\gamma_0}}\frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2 X_\gamma X_\gamma^T\right)^{-1}Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1}Y\right)\right\}}\\
\leq&\exp\left\{-a(|\gamma|-|\gamma_0|) + bR_n^2\right\}\left(\tau^2\right)^{\frac{|\gamma|-|\gamma_0|}{2}}\\
&\times\exp\left\{\frac{1}{2\sigma^2}\left(|\gamma| - |\gamma_0| + \sqrt{(|\gamma|-|\gamma_0|)\log p} + \frac{1}{1+n\epsilon_0\tau^2/2}n\log p\right)\right\}\\
\leq&\exp\left\{-\frac{\alpha_1}{\kappa}(|\gamma|-|\gamma_0|)\log p\right\}.
\end{aligned} \tag{S1.38}$$

Next, when $\gamma \subset \gamma_0$, Let Z be a standard normal distribution. When $\gamma \subset \gamma_0$, it follows from Lemma L.1 in Cao et al. (2019a), Assumption 2 and the relation between noncentral chi-squared and normal distribution that,

$$\begin{aligned} & P\left(\frac{R_\gamma - R_{\gamma_0}}{\sigma^2} < 4|\gamma_0| \log n \log p\right) \\ < & P\left((Z - \sqrt{\lambda})^2 < 4|\gamma_0| \log n \log p\right) \\ < & e^{-\frac{n\epsilon_0|\gamma_0|\rho_1^2}{4\sigma^2}} \to 0, \text{ as } n \to \infty. \end{aligned} \tag{S1.39}$$

where $\rho_1 = \min_{j \in \gamma_0} |\beta_{0j}|$ and $\lambda = \frac{\beta_0{}^T\left(X_{\gamma_0}^T P_{\gamma_0} X_{\gamma_0}\right)\beta_0}{\sigma^2} > \frac{n\epsilon_0|\gamma_0|\rho_1^2}{\sigma^2}$. it again follows from (S1.29), (S1.31), (S1.39) and $R_\gamma^* - R_\gamma \geq 0$, with probability tending to 1, we have

$$\begin{aligned} & \frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I + \tau^2 X_\gamma X_\gamma^T\right)^{-1} Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I + \tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1} Y\right)\right\}} \\ = & \exp\left\{-\frac{1}{2\sigma^2}(R_\gamma^* - R_{\gamma_0}^*)\right\} \\ \leq & \exp\left\{-\frac{1}{\sigma^2}\left(R_\gamma - \left(R_{\gamma_0} + \frac{1}{1 + n\epsilon_0\tau^2/2} Y^T P_{\gamma_0} Y\right)\right)\right\} \\ \leq & \exp\left\{-4|\gamma_0| \log n \log p + \frac{1}{1 + n\epsilon_0\tau^2/2} n \log p\right\}. \end{aligned} \tag{S1.40}$$

Next, note that it follows from $\gamma \subset \gamma_0$, Assumption 2 and the arguments leading up to (S1.20) that for large enough n (not depending on γ, $\mathscr{D}$), $\frac{Q_\gamma}{Q_{\gamma_0}} \leq n^{-\frac{|\gamma| - |\gamma_0|}{2}} \left(\frac{1}{\epsilon_0/2}\right)^{\frac{|\gamma_0| - |\gamma|}{2}}$. Therefore, it follows from $\tau^2 \sim \sqrt{\log p}$, $a \sim \log p$, $\gamma \subset \gamma_0$, (S1.29) and (S1.40) that, for large enough $n \geq N_5$, with probability tending to 1,

$$
\begin{aligned}
&\frac{\pi\left(\gamma,\mathscr{D}|Y,X\right)}{\pi\left(\gamma_0,\mathscr{D}|Y,X\right)}\\
=&\frac{\exp\left(-a1^T\gamma+b\gamma^TG\gamma\right)}{\exp\left(-a1^T\gamma_0+b\gamma_0^TG\gamma_0\right)}\frac{Q_\gamma}{Q_{\gamma_0}}\frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2X_\gamma X_\gamma^T\right)^{-1}Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2X_{\gamma_0}X_{\gamma_0}^T\right)^{-1}Y\right)\right\}}\\
\leq&\exp\left\{a|\gamma_0|\right\}n^{\frac{|\gamma_0|}{2}}\left(\frac{1}{\epsilon_0/2}\right)^{\frac{|\gamma_0|-|\gamma|}{2}}\\
&\times\exp\left\{-4|\gamma_0|\log n\log p+\frac{1}{1+n\epsilon_0\tau^2/2}n\log p\right\}\\
\leq&\exp\left\{-2|\gamma_0|\log p\right\}. \qquad \text{(S1.41)}
\end{aligned}
$$

Next, consider the scenario when $\gamma\not\subseteq\gamma_0$ and $\gamma\not\supseteq\gamma_0$. Denote $\gamma'=\gamma\cap\gamma_0$. It follows from (S1.29) that

$$
\begin{aligned}
&\frac{\pi\left(\gamma,\mathscr{D}|Y,X\right)}{\pi\left(\gamma_0,\mathscr{D}|Y,X\right)}\\
=&\frac{\pi\left(\gamma,\mathscr{D}|Y,X\right)}{\pi\left(\gamma',\mathscr{D}|Y,X\right)}\frac{\pi\left(\gamma',\mathscr{D}|Y,X\right)}{\pi\left(\gamma_0,\mathscr{D}|Y,X\right)}\\
=&\frac{\pi(\gamma|\mathscr{D})}{\pi(\gamma'|\mathscr{D})}\frac{Q_\gamma}{Q_{\gamma'}}\frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2X_\gamma X_\gamma^T\right)^{-1}Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2X_{\gamma'}X_{\gamma'}^T\right)^{-1}Y\right)\right\}}\\
&\times\frac{\pi(\gamma'|\mathscr{D})}{\pi(\gamma_0|\mathscr{D})}\frac{Q_{\gamma'}}{Q_{\gamma_0}}\frac{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2X_{\gamma'}X_{\gamma'}^T\right)^{-1}Y\right)\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\left(Y^T\left(I+\tau^2X_{\gamma_0}X_{\gamma_0}^T\right)^{-1}Y\right)\right\}}. \qquad \text{(S1.42)}
\end{aligned}
$$

Since $\gamma'\subset\gamma$ and $\gamma'\subset\gamma_0$, following the same arguments leading up to (S1.38) and (S1.41), we have for large enough $n>\max\{N_4,N_5\}$, with probability tending to 1,

$$
\frac{\pi\left(\gamma,\mathscr{D}|Y,X\right)}{\pi\left(\gamma_0,\mathscr{D}|Y,X\right)}
$$

$$\le \exp\left\{-\frac{\alpha_1}{\kappa}(|\gamma| - |\gamma'|)\log p\right\}\exp\{-2|\gamma_0|\log p\}$$
$$\le \exp\left\{-\frac{\alpha_1}{\kappa}(|\gamma| - |\gamma'|)\log p - 2|\gamma_0|\log p\right\}. \quad \text{(S1.43)}$$

Theorem 2 immediately follows after (S1.38), (S1.41) and (S1.43). For any $\gamma \neq \gamma_0$, for large enough $n > \max\{N_4, N_5\}$, we have

$$\max_{(\gamma,\mathscr{D})\neq(\gamma_0,\mathscr{D}_0)} \frac{\pi(\gamma,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}|Y,X)} \xrightarrow{\bar{P}} 0, \text{ as } n\to\infty. \quad \text{(S1.44)}$$

□

Proof of Theorem 4. We have

$$\frac{1-\pi(\gamma,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}|Y,X)}$$
$$= \sum_{(\gamma,\mathscr{D})\neq(\gamma_0,\mathscr{D}_0)} \frac{\pi(\gamma,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}$$
$$= \sum_{\gamma\neq\gamma_0} \frac{\pi(\gamma,\mathscr{D}_0|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)} + \sum_{\mathscr{D}\neq\mathscr{D}_0} \frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)} + \sum_{\gamma\neq\gamma_0,\mathscr{D}\neq\mathscr{D}_0} \frac{\pi(\gamma,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}. \quad \text{(S1.45)}$$

Note that it follows from the proof of Theorem 2 that for large enough constant $N > \max\{N_4, N_5\}$,

$$\sum_{\gamma\neq\gamma_0} \frac{\pi(\gamma,\mathscr{D}_0|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}$$
$$\le \sum_{\gamma\subset\gamma_0} \frac{\pi(\gamma,\mathscr{D}_0|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)} + \sum_{\gamma\supset\gamma_0} \frac{\pi(\gamma,\mathscr{D}_0|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)} + \sum_{\gamma\not\subseteq\gamma_0,\gamma\not\supseteq\gamma_0} \frac{\pi(\gamma,\mathscr{D}_0|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}$$
$$\le \sum_{|\gamma|=1}^{|\gamma_0|} \binom{|\gamma_0|}{|\gamma|} \exp\left\{-\frac{\alpha_1}{\kappa}|\gamma_0|\log p\right\} + \sum_{|\gamma|=|\gamma_0|+1}^{R_n} \binom{p-|\gamma_0|}{|\gamma|-|\gamma_0|} \exp\{-2(|\gamma|-|\gamma_0|)\log p\}$$

$$+\sum_{|\gamma|=1}^{R_n}\binom{p}{|\gamma|}\exp\left\{-\frac{\alpha_1}{\kappa}(|\gamma|-|\gamma'|)\log p-2|\gamma_0|\log p\right\}.$$

Further note that the upper bound of the binomial coefficient satisfies $\binom{p}{k}\leq p^k$, for any $1\leq k\leq p$. It follows that when $\alpha_1>2\kappa$ for some $\kappa>1$,

$$\sum_{\gamma\neq\gamma_0}\frac{\pi(\gamma,\mathscr{D}_0|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\to 0,\quad \text{as } n\to\infty. \tag{S1.46}$$

Next, it follows from Lemmas 1-3 that if we restrict to E_n^c, then for large enough constant $N>N_3$, we have

$$\begin{aligned}
&\sum_{\mathscr{D}\neq\mathscr{D}_0}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\\
\leq&\sum_{j=1}^{p-1}\sum_{pa_j(\mathscr{D})\neq pa_j(\mathscr{D}_0)}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\\
\leq&\sum_{j=1}^{p-1}\left(\sum_{pa_j(\mathscr{D})\subset pa_j(\mathscr{D}_0)}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}+\sum_{pa_j(\mathscr{D})\supset pa_j(\mathscr{D}_0)}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\right.\\
&\qquad\left.+\sum_{pa_j(\mathscr{D})\not\subseteq,\not\supseteq pa_j(\mathscr{D}_0)}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\right)\\
\leq&\sum_{j=1}^{p-1}\left(\sum_{\nu_j(\mathscr{D})=1}^{\nu_j(\mathscr{D}_0)-1}\binom{\nu_j(\mathscr{D}_0)}{|Z_j|}p^{-\frac{2\alpha_1}{\kappa}d}+\sum_{\nu_j(\mathscr{D})=\nu_j(\mathscr{D}_0)+1}^{R_n}\binom{p-\nu_j(\mathscr{D}_0)}{\nu_j(\mathscr{D})-\nu_j(\mathscr{D}_0)}(2p)^{-\frac{\alpha_1}{\kappa}(\nu_i(\mathscr{D})-\nu_i(\mathscr{D}_0))}\right.\\
&\qquad\left.+\sum_{\nu_i(\mathscr{D})=1}^{R_n}\binom{p}{\nu_i(\mathscr{D})}(2p)^{-\frac{\alpha_1}{\kappa}\nu_i(\mathscr{D})}\right).
\end{aligned}\tag{S1.47}$$

Again it follows from $\binom{p}{k}\leq p^k$, for any $1\leq k\leq p$ that when $\alpha_1>2\kappa$ for some $\kappa>1$,

$$\sum_{\mathscr{D}\neq\mathscr{D}_0}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\to 0,\quad \text{as } n\to\infty. \tag{S1.48}$$

Finally, by (S1.46) and (S1.48), note that

$$\sum_{\gamma\neq\gamma_0,\mathscr{D}\neq\mathscr{D}_0}\frac{\pi(\gamma,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}\leq\sum_{\gamma\neq\gamma_0}\frac{\pi(\gamma,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}|Y,X)}\sum_{\mathscr{D}\neq\mathscr{D}_0}\frac{\pi(\gamma_0,\mathscr{D}|Y,X)}{\pi(\gamma_0,\mathscr{D}_0|Y,X)}$$
$$\rightarrow 0,\quad \text{as } n\rightarrow\infty. \tag{S1.49}$$

Therefore, following from (S1.45), (S1.46), (S1.48) and (S1.49), we have $\pi(\gamma_0,\mathscr{D}_0|Y)\rightarrow 1$, as $n\rightarrow\infty$, which completes our proof of the strong model selection result in Theorem 4. □

Proof of Corollary 1. Note that with the extra layer of inverse gamma distribution on σ^2, by integrating out σ^2 in the proof of Lemma 1, the (marginal) joint posterior distribution is given by

$$\begin{aligned}
&\pi\left(\gamma,\mathscr{D}|Y,X\right)\\
=&\int\pi\left(Y|\gamma,\beta_\gamma\right)\prod_{i=1}^{n}\pi\left(X_i|(L,D)\right)\pi_{U,\alpha(\mathscr{D})}^{\Theta_{\mathscr{D}}}(L,D)\\
&\times\pi\left(\beta_\gamma|\gamma\right)\pi(\gamma)\pi(\mathscr{D})\pi(\sigma^2)d\beta_\gamma d(L,D)d(\sigma^2)\\
\propto&\pi(\gamma|\mathscr{D})\pi(\mathscr{D})\frac{z_{\mathscr{D}}(U+X^TX,n+\alpha(\mathscr{D}))}{z_{\mathscr{D}}(U,\alpha(\mathscr{D}))}Q_\gamma\\
&\times\left(\frac{1}{2}\left(Y^T\left(I+\tau^2X_\gamma X_\gamma^T\right)^{-1}Y\right)+b_0\right)^{-(\frac{n}{2}+a_0)},
\end{aligned}\tag{S1.50}$$

where $Q_\gamma=\det\left(\tau^2X_\gamma^TX_\gamma+I_{|\gamma|}\right)^{-\frac{1}{2}}$. The proofs for Lemmas 1-3 will go through with the new posterior. For the variable selection consistency, it follows from (S1.50) that,

$$\begin{aligned}
&\frac{\pi\left(\gamma, \mathscr{D}|Y, X\right)}{\pi\left(\gamma_0, \mathscr{D}|Y, X\right)}\\
=&\frac{\exp\left(-a1^T\gamma + b\gamma^T G\gamma\right)}{\exp\left(-a1^T\gamma_0 + b\gamma_0^T G\gamma_0\right)}\frac{Q_\gamma}{Q_{\gamma_0}}\\
&\times\frac{\left(\frac{1}{2}\left(Y^T\left(I+\tau^2 X_\gamma X_\gamma^T\right)^{-1}Y\right)+b_0\right)^{-(\frac{n}{2}+a_0)}}{\left(\frac{1}{2}\left(Y^T\left(I+\tau^2 X_{\gamma_0} X_{\gamma_0}^T\right)^{-1}Y\right)+b_0\right)^{-(\frac{n}{2}+a_0)}}\\
=&\frac{\exp\left(-a1^T\gamma + b\gamma^T G\gamma\right)}{\exp\left(-a1^T\gamma_0 + b\gamma_0^T G\gamma_0\right)}\frac{Q_\gamma}{Q_{\gamma_0}}\left(\frac{R_\gamma^* + 2b_0}{R_{\gamma_0}^* + 2b_0}\right)^{-(\frac{n}{2}+a_0)}. \qquad \text{(S1.51)}
\end{aligned}$$

It follows from the arguments leading up to (S1.41) and $1 + x \leq e^x$ that when $\gamma \supset \gamma_0$, we have

$$\begin{aligned}
&\frac{\pi\left(\gamma, \mathscr{D}|Y, X\right)}{\pi\left(\gamma_0, \mathscr{D}|Y, X\right)}\\
\leq&\exp\left\{-a(|\gamma|-|\gamma_0|)+bR_n^2\right\}\left(n\tau^2\epsilon_0\right)^{\frac{1}{2}|\gamma_0|}\times\left(1+\frac{R_{\gamma_0}^* - R_\gamma^*}{R_\gamma^* + 2a_0}\right)^{\frac{n}{2}+b_0}\\
\leq&\exp\left\{-a(|\gamma|-|\gamma_0|)+bR_n^2\right\}\left(\tau^2\right)^{\frac{|\gamma|-|\gamma_0|}{2}}\\
&\times\exp\left\{\frac{(\frac{n}{2}+a_0)\left(|\gamma|-|\gamma_0|+\sqrt{(|\gamma|-|\gamma_0|)\log p}+\frac{1}{1+n\epsilon_0\tau^2/2}n\log p\right)}{n-|\gamma|-\sqrt{(n-|\gamma|)\log p}+2b_0}\right\}\\
\leq&\exp\left\{-\frac{\alpha_1}{\kappa}(|\gamma|-|\gamma_0|)\log p\right\}. \qquad \text{(S1.52)}
\end{aligned}$$

Next, when $\gamma \subset \gamma_0$, it follows by the arguments leading up to (S1.41) and $1 - x \leq e^{-x}$ that,

$$\begin{aligned}
&\frac{\pi\left(\gamma, \mathscr{D}|Y, X\right)}{\pi\left(\gamma_0, \mathscr{D}|Y, X\right)}\\
=&\frac{\exp\left(-a1^T\gamma + b\gamma^T G\gamma\right)}{\exp\left(-a1^T\gamma_0 + b\gamma_0^T G\gamma_0\right)}\left(n\tau^2\epsilon_0\right)^{\frac{1}{2}|\gamma_0|}\left(1-\frac{R_\gamma^* - R_{\gamma_0}^*}{R_\gamma^* + 2b_0}\right)^{\frac{n}{2}+a_0}
\end{aligned}$$

$$\leq \exp\left\{a|\gamma_0|\right\} n^{\frac{|\gamma_0|}{2}} \left(\frac{1}{\epsilon_0/2}\right)^{\frac{|\gamma_0|-|\gamma|}{2}}$$

$$\times \exp\left\{\frac{\left(\frac{n}{2}+a_0\right)\left(-4|\gamma_0|\log p + \frac{1}{1+n\epsilon_0\tau^2/2} n\log p\right)}{n-|\gamma|+\sqrt{(n-|\gamma|)\log p} + \frac{1}{1+n\epsilon_0\tau^2/2}Y^T P_\gamma Y + 2b_0}\right\}$$

$$\leq \exp\left\{(-2|\gamma_0|\log p)\right\}. \tag{S1.53}$$

When $\gamma \nsubseteq \gamma_0$ and $\gamma \nsupseteq \gamma_0$, the exact same results as the previous case without the inverse gamma prior can be obtained by following the arguments leading up to (S1.43). Similarly, Corollary 1 can be acquired from the same arguments leading up to (S1.49). □

Proof of Theorem 5. We start proving Theorem 5 by considering the ratio between posterior ratios under two settings corresponding to $b > 0$ and $b = 0$ respectively. Specifically, let $\pi_1(\gamma, \mathscr{D}|Y, X)$ represent the posterior probability under $b > 0$ and $\pi_1(\gamma, \mathscr{D}|Y, X)$ represent the posterior probability under $b = 0$. It follows from (S1.29) that

$$\frac{\frac{\pi_1(\gamma,\mathscr{D}_0|Y,X)}{\pi_1(\gamma_0,\mathscr{D}_0|Y,X)}}{\frac{\pi_2(\gamma,\mathscr{D}_0|Y,X)}{\pi_2(\gamma_0,\mathscr{D}_0|Y,X)}} = \exp\left\{\gamma^T G_0 \gamma - \gamma_0^T G_0 \gamma_0\right\}. \tag{S1.54}$$

Note that by Condition 1, for any γ, $\gamma^T G_0 \gamma = \sum_{1\leq i,j\leq p} (G_0)_{ij}\gamma_i\gamma_j$ will be maximized at $\gamma = \gamma_0$. Therefore, for any γ, we have

$$\frac{\pi_1(\gamma, \mathscr{D}_0|Y, X)}{\pi_1(\gamma_0, \mathscr{D}_0|Y, X)} \leq \frac{\pi_2(\gamma, \mathscr{D}_0|Y, X)}{\pi_2(\gamma_0, \mathscr{D}_0|Y, X)}. \tag{S1.55}$$

In addition, over all possible scenarios of γ, there exists at least one $\gamma \neq$

γ_0 such that $\gamma^T G_0 \gamma < \gamma_0^T G_0 \gamma_0$ and $\frac{\pi_1(\gamma, \mathscr{D}_0|Y,X)}{\pi_1(\gamma_0, \mathscr{D}_0|Y,X)}$ is strictly smaller than $\frac{\pi_2(\gamma, \mathscr{D}_0|Y,X)}{\pi_2(\gamma_0, \mathscr{D}_0|Y,X)}$. Hence, it follows from (S1.55) that

$$\sum_{\gamma \neq \gamma_0} \frac{\pi_1(\gamma, \mathscr{D}_0|Y,X)}{\pi_1(\gamma_0, \mathscr{D}_0|Y,X)} < \sum_{\gamma \neq \gamma_0} \frac{\pi_2(\gamma, \mathscr{D}_0|Y,X)}{\pi_2(\gamma_0, \mathscr{D}_0|Y,X)}, \tag{S1.56}$$

which is equivalent to

$$\frac{1 - \pi_1(\gamma_0, \mathscr{D}_0|Y,X)}{\pi_1(\gamma_0, \mathscr{D}_0|Y,X)} < \frac{1 - \pi_2(\gamma_0, \mathscr{D}_0|Y,X)}{\pi_2(\gamma_0, \mathscr{D}_0|Y,X)}. \tag{S1.57}$$

Therefore, we have

$$\pi_1(\gamma_0, \mathscr{D}_0|Y,X) > \pi_2(\gamma_0, \mathscr{D}_0|Y,X).$$

□

References

Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *Ann. Statist. 36*, 199–227.

Cao, X., K. Khare, and M. Ghosh (2019a). High-dimensional posterior consistency for hierarchical non-local priors in regression. *Bayesian Anal., to appear*.

Cao, X., K. Khare, and M. Ghosh (2019b). Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models. *Ann. Statist. 47*(1), 319–348.

Rudelson, M. and R. Vershynin (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Comm. Probab. 18*.