Supplementary Materials for "High-Dimensional Multivariate Posterior Consistency Under Global-Local Shrinkage Priors"

Ray Bai[a,*], Malay Ghosh[a]

[a]*Department of Statistics, University of Florida, Gainesville, FL 32611, United States*

---

1. Proofs for Section 3.3

1.1. Proof of Theorem 1

The proof of Theorem 1 is based on a lemma. This lemma is similar to Lemma 1.1 in Goh et al. [3], with suitable modifications so that we utilize Conditions (A1)-(A3) explicitly. Furthermore, Goh et al. [3] gave a sufficient condition for posterior consistency in the Frobenius norm when $p_n = o(n)$ in Theorem 1 of their paper. However, we are not clear about a particular step in their proof. They assert that

$$\left\{(\mathbf{A},\mathbf{B}) : n^{-1}\left(||\mathbf{Y}_n - \mathbf{X}\mathbf{C})\mathbf{\Sigma}^{-1/2}||_F^2 - ||(\mathbf{Y}_n - \mathbf{X}\mathbf{C}^*)\mathbf{\Sigma}^{-1/2}||_F^2\right) < 2\nu,\right.$$
$$\left.\mathbf{C} = \mathbf{A}\mathbf{B}^\top\right\}$$
$$\supseteq \left\{(\mathbf{A},\mathbf{B}) : n^{-1}\left|||\mathbf{Y}_n - \mathbf{X}\mathbf{C}||_F^2 - ||(\mathbf{Y}_n - \mathbf{X}\mathbf{C}^*)||_F^2\right| < 2\tau_{\min}\nu, \mathbf{C} = \mathbf{A}\mathbf{B}^\top\right\},$$

where $\tau_{\min}$ is the minimum eigenvalue for $\mathbf{\Sigma}$. This does not seem to be true in general, unless the matrix $(\mathbf{Y}_n - \mathbf{X}\mathbf{C})(\mathbf{Y}_n - \mathbf{X}\mathbf{C})^\top - (\mathbf{Y}_n - \mathbf{X}\mathbf{C}^*)(\mathbf{Y}_n - \mathbf{X}\mathbf{C}^*)^\top$ is positive definite, which cannot be assumed. Our proof for Theorem 1 thus gives a different sufficient condition for posterior consistency in this low-dimensional setting. Moreover, the proof of Theorem 2 in the ultrahigh-dimensional case requires a suitable modification of Theorem 1. Thus, we deem it beneficial to write out all the details for Lemma 1 and Theorem 1.

Lemma 1. *Define* $\mathcal{B}_\varepsilon = \{\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F > \varepsilon\}$*, where* $\varepsilon > 0$*. To test* $H_0 : \mathbf{B}_n = \mathbf{B}_0$ *vs.* $H_1 : \mathbf{B}_n \in \mathcal{B}_\varepsilon$*, define a test function* $\Phi_n = 1(\mathbf{Y}_n \in \mathcal{C}_n)$*, where the critical region is* $C_n := \left\{\mathbf{Y}_n : ||\widehat{\mathbf{B}}_n - \mathbf{B}_0||_F > \varepsilon/2\right\}$ *and* $\widehat{\mathbf{B}}_n = (\mathbf{X}_n^\top\mathbf{X}_n)^{-1}\mathbf{X}_n^\top\mathbf{Y}_n$*. Then, under model (10) and assumptions (A1)-(A3), we have that as* $n \to \infty$*,*

*Corresponding author.
Email address: raybai07@ufl.edu (Ray Bai)

1. $\mathbb{E}_{\mathbf{B}_0}(\Phi_n) \leq \exp(-\varepsilon^2 n c_1/16 d_2)$,
2. $\sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1-\Phi_n) \leq \exp(-\varepsilon^2 n c_1/16 d_2)$.

Proof of Lemma 1. Since $\widehat{\mathbf{B}}_n \sim \mathcal{MN}_{p_n \times q}(\mathbf{B}_0, (\mathbf{X}_n^\top \mathbf{X}_n)^{-1}, \mathbf{\Sigma})$ w.r.t. $\mathbb{P}_0$-measure,

$$\mathbf{Z}_n = (\mathbf{X}_n^\top \mathbf{X}_n)^{1/2}(\widehat{\mathbf{B}}_n - \mathbf{B}_0)\mathbf{\Sigma}^{-1/2} \sim \mathcal{MN}_{p_n \times q}(\mathbf{O}, \mathbf{I}_{p_n}, \mathbf{I}_q). \tag{1.1}$$

Using the fact that for square conformal positive definite matrices $\mathbf{A}, \mathbf{B}$, $\lambda_{\min}(\mathbf{A})\text{tr}(\mathbf{B}) \leq \text{tr}(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\text{tr}(\mathbf{B})$, we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{B}_0}(\Phi_n) &= \mathbb{P}_{\mathbf{B}_0}(\mathbf{Y}_n : ||\widehat{\mathbf{B}}_n - \mathbf{B}_0||_F > \varepsilon/2) \\
&= \mathbb{P}_{\mathbf{B}_0}(||(\mathbf{X}_n^\top \mathbf{X}_n)^{-1/2}\mathbf{Z}_n \mathbf{\Sigma}^{1/2}||_F^2 > \varepsilon^2/4) \qquad \text{(by (4))} \\
&= \mathbb{P}_{\mathbf{B}_0}(\text{tr}(\mathbf{\Sigma}^{1/2}\mathbf{Z}_n(\mathbf{X}_n^\top \mathbf{X}_n)^{-1}\mathbf{Z}_n \mathbf{\Sigma}^{1/2}) > \varepsilon^2/4) \\
&\leq \mathbb{P}_{\mathbf{B}_0}(n^{-1}c_1^{-1}\text{tr}(\mathbf{\Sigma}^{1/2}\mathbf{Z}_n^\top \mathbf{Z}_n \mathbf{\Sigma}^{1/2}) > \varepsilon^2/4) \\
&\leq \mathbb{P}_{\mathbf{B}_0}(n^{-1}c_1^{-1}d_2\text{tr}(\mathbf{Z}_n^\top \mathbf{Z}_n) > \varepsilon^2/4) \\
&= \mathbb{P}_{\mathbf{B}_0}\left(||\mathbf{Z}_n||_F^2 > \frac{\varepsilon^2 c_1 n}{4d_2}\right) \\
&= \Pr\left(\chi^2_{p_n q} > \frac{\varepsilon^2 c_1 n}{4d_2}\right) \qquad \text{(by (1.1))},
\end{aligned} \tag{1.2}$$

where the two inequalities follow from Assumptions (A2) and (A3) respectively. By Armagan et al. [1], for all $m > 0$, $\Pr(\chi^2_m \geq x) \leq \exp(-x/4)$ whenever $x \geq 8m$. Using Assumption (A1) and noting that q is fixed, we have by (1.1) that as $n \to \infty$,

$$\mathbb{E}_{\mathbf{B}_0}(\Phi_n) \leq \Pr\left(\chi^2_{p_n q} > \frac{\varepsilon^2 c_1 n}{4d_2}\right) \leq \exp\left(-\frac{\varepsilon^2 c_1 n}{16 d_2}\right),$$

thus establishing the first part of the lemma.

We next show the second part of the lemma. We have

$$\begin{aligned}
\sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1-\Phi_n) &= \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}(\mathbf{Y}_n : ||\widehat{\mathbf{B}}_n - \mathbf{B}_0||_F \leq \varepsilon/2) \\
&\leq \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : \left| ||\widehat{\mathbf{B}}_n - \mathbf{B}_n||_F - ||\mathbf{B}_n - \mathbf{B}_0||_F \right| \leq \varepsilon/2\right) \\
&\leq \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : -\varepsilon/2 + ||\mathbf{B}_n - \mathbf{B}_0||_F \leq ||\widehat{\mathbf{B}}_n - \mathbf{B}_n||_F\right) \\
&= \mathbb{P}_{\mathbf{B}_n}(\mathbf{Y}_n : ||\widehat{\mathbf{B}}_n - \mathbf{B}_n||_F > \varepsilon/2) \\
&\leq \exp\left(-\frac{\varepsilon^2 c_1 n}{16 d_2}\right),
\end{aligned}$$

The last inequality follows from the fact that $\widehat{\mathbf{B}}_n \sim \mathcal{MN}_{p_n \times q}(\mathbf{B}_n, (\mathbf{X}_n^\top \mathbf{X}_n)^{-1}, \mathbf{\Sigma})$. Thus, we may use the same steps that were used to prove the first part of the lemma. Therefore, we have also established the second part of the lemma. □

Proof of Theorem 1. We utilize the proof technique of Theorem 1 in Armagan et al. [1] and modify it suitably for the multivariate case subject to conditions (A1)-(A3). The posterior probability of $\mathcal{B}_n$ is given by

$$\begin{aligned}
\Pi_n(\mathcal{B}_n|\mathbf{Y}_n) &= \frac{\int_{\mathcal{B}_n} \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \Pi_n(d\mathbf{B}_n)}{\int \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \Pi_n(d\mathbf{B}_n)} \\
&\leq \Phi_n + \frac{(1-\Phi_n)J_{\mathcal{B}_\varepsilon}}{J_n} \\
&= I_1 + \frac{I_2}{J_n}, \qquad (1.3)
\end{aligned}$$

where $J_{\mathcal{B}_\varepsilon} = \int_{\mathcal{B}_\varepsilon} \left\{ \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \right\} \Pi_n(d\mathbf{B}_n)$ and $J_n = \int \left\{ \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \right\} \Pi_n(d\mathbf{B}_n)$.

Let $b = \frac{\varepsilon^2 c_1}{16 d_2}$. For sufficiently large n, using Markov's Inequality and the first part of Lemma 1, we have

$$\mathbb{P}_{\mathbf{B}_0}\left(I_1 \geq \exp\left(-\frac{bn}{2}\right)\right) \leq \exp\left(\frac{bn}{2}\right) \mathbb{E}_{\mathbf{B}_0}(I_1) \leq \exp\left(-\frac{bn}{2}\right).$$

This implies that $\sum_{n=1}^{\infty} \mathbb{P}_{\mathbf{B}_0}\left(I_1 \geq \exp(-bn/2)\right) < \infty$. Thus, by the Borel-Cantelli Lemma, $I_1 \to 0$ a.s. $\mathbb{P}_0$ as $n \to \infty$.

We next look at the behavior of I_2. We have

$$\begin{aligned}
\mathbb{E}_{\mathbf{B}_0} I_2 = \mathbb{E}_{\mathbf{B}_0}\left\{(1-\Phi_n)J_{\mathcal{B}_\varepsilon}\right\} &= \mathbb{E}_{\mathbb{B}_0}\left\{(1-\Phi_n)\int_{\mathcal{B}_\varepsilon} \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \Pi_n(d\mathbf{B}_n)\right\} \\
&= \int_{\mathcal{B}_\varepsilon} \int (1-\Phi_n) f(\mathbf{Y}_n|\mathbf{B}_n)\, d\mathbf{Y}_n\, \Pi_n(d\mathbf{B}_n) \\
&\leq \Pi_n(\mathcal{B}_\varepsilon) \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1-\Phi_n) \\
&\leq \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1-\Phi_n) \\
&\leq \exp(-bn),
\end{aligned}$$

where the last inequality follows from the second part of Lemma 1.

Thus, for sufficiently large n, $\mathbb{P}_{\mathbf{B}_0}(I_2 \geq \exp(-bn/2)) \leq \exp(-bn/2)$, which implies that $\sum_{n=1}^{\infty} \mathbb{P}_{\mathbf{B}_0}\left(I_2 \geq \exp(-bn/2)\right) < \infty$. Thus, by the Borel-Cantelli Lemma, $I_2 \to 0$ a.s. $\mathbb{P}_0$ as $n \to \infty$.

We have now shown that both I_1 and I_2 in (1.3) tend towards zero exponentially fast. We now analyze the behavior of J_n. To complete the proof, we need to show that

$$\exp(bn/2) J_n \to \infty \;\; \mathbb{P}_0 \text{ a.s. as } n \to \infty. \qquad (1.4)$$

Note that

$$\exp(bn/2)J_n = \exp(bn/2)\int \exp\left\{-n\frac{1}{n}\ln\frac{f(\mathbf{Y}_n|\mathbf{B}_0)}{f(\mathbf{Y}_n|\mathbf{B}_n)}\right\}\Pi_n(d\mathbf{B}_n)$$
$$\geq \exp\{(b/2-\nu)n\}\Pi_n(\mathcal{D}_{n,\nu}), \tag{1.5}$$

where $\mathcal{D}_{n,\nu} = \left\{\mathbf{B}_n : n^{-1}\ln\left(\frac{f(\mathbf{Y}_n|\mathbf{B}_0)}{f(\mathbf{Y}_n|\mathbf{B}_n)}\right) < \nu\right\}$ for $0 < \nu < b/2$. Therefore, we have

$$\begin{aligned}\mathcal{D}_{n,\nu} &= \Big\{\mathbf{B}_n : n^{-1}\Big(\frac{1}{2}\mathrm{tr}\left[[(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_n)^\top(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_n)]\boldsymbol{\Sigma}^{-1}\right] - \frac{1}{2}\mathrm{tr}\,[[(\mathbf{Y}_n- \\ &\quad \mathbf{X}_n\mathbf{B}_0)^\top(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)]\boldsymbol{\Sigma}^{-1}]\Big) < 2\nu\Big\} \\ &\equiv \Big\{\mathbf{B}_n : n^{-1}\Big(\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_n)^\top(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_n)\boldsymbol{\Sigma}^{-1/2}\right] - \mathrm{tr}\Big[\boldsymbol{\Sigma}^{-1/2} \\ &\quad (\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)^\top(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}\Big]\Big) < 2\nu\Big\} \\ &\equiv \left\{\mathbf{B}_n : n^{-1}\left(||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_n)\boldsymbol{\Sigma}^{-1/2}||_F^2 - ||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F^2\right) < 2\nu\right\}.\end{aligned}$$

Noting that

$$\begin{aligned}||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_n)\boldsymbol{\Sigma}^{-1/2}||_F^2 &\leq ||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F^2 + ||\mathbf{X}_n(\mathbf{B}_n-\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F^2 \\ &\quad + 2||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F||\mathbf{X}_n(\mathbf{B}_n-\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F,\end{aligned}$$

we have

$$\begin{aligned}\Pi_n(\mathcal{D}_{n,\nu}) &\geq \Pi\Big\{\mathbf{B}_n : n^{-1}\Big(2||\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F||\mathbf{X}_n(\mathbf{B}_n-\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F \\ &\quad +||\mathbf{X}_n(\mathbf{B}_n-\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F^2\Big) < 2\nu\Big\} \\ &\geq \Pi\Big\{\mathbf{B}_n : n^{-1}||\mathbf{X}_n(\mathbf{B}_n-\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F < \frac{2\nu}{3\kappa_n}, \\ &\quad ||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F < \kappa_n\Big\},\end{aligned} \tag{1.6}$$

for some positive increasing sequence κ_n such that $\kappa_n \to \infty$ as $n \to \infty$.

Set $\kappa_n = n^{(1+\rho)/2}$ for $\rho > 0$. Since $\mathbf{E}_n = \mathbf{Y}_n - \mathbf{X}_n\mathbf{B}_0$, we have $\mathbf{Z}_n = (\mathbf{Y}_n - \mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2} \sim \mathcal{MN}_{n\times q}(\mathbf{O}, \mathbf{I}_n, \mathbf{I}_q)$. Therefore, as $n \to \infty$,

$$\begin{aligned}\mathbb{P}_{\mathbf{B}_0}(||(\mathbf{Y}_n-\mathbf{X}_n\mathbf{B}_0)\boldsymbol{\Sigma}^{-1/2}||_F > \kappa_n) &= \mathbb{P}_{\mathbf{B}_0}(||\mathbf{Z}_n||_F^2 > \kappa_n^2) \\ &= \Pr\left(\chi^2_{nq} > n^{1+\rho}\right) \\ &\leq \exp\left(-\frac{n^{1+\rho}}{4}\right),\end{aligned}$$

where the last inequality follows from the fact that for all $m > 0$, $\Pr(\chi^2_m \geq x) \leq \exp(-x/4)$ when $x \geq 8m$ and the assumptions that q is fixed and $\rho > 0$. Since

$\sum_{n=1}^{\infty} \mathbb{P}_{\mathbf{B}_0}(||(\mathbf{Y}_n - \mathbf{X}_n\mathbf{B}_0)\mathbf{\Sigma}^{-1/2}||_F > \kappa_n) \leq \sum_{n=1}^{\infty} \exp\left(-\frac{n^{1+\rho}}{4}\right) < \infty$, we have by the Borel-Cantelli Lemma that

$$\mathbb{P}_{\mathbf{B}_0}\left\{||\mathbf{Y}_n - \mathbf{X}_n\mathbf{B}_0||_F > \kappa_n \text{ infinity often } \right\} = 0.$$

For sufficiently large n, we have from (1.6) that

$$\begin{aligned}
\Pi_n(\mathcal{D}_{n,\nu}) &\geq \Pi_n\left\{\mathbf{B}_n : n^{-1}||\mathbf{X}_n(\mathbf{B}_n - \mathbf{B}_0)\mathbf{\Sigma}^{-1/2}||_F < \frac{2\nu}{3\kappa_n}\right\} \\
&\geq \Pi_n\left\{n^{-1}n^{1/2}c_2^{1/2}d_1^{-1/2}||\mathbf{B}_n - \mathbf{B}_0||_F < \frac{2\nu}{3\kappa_n}\right\} \\
&= \Pi_n\left\{\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F < \left(\frac{2d_1^{1/2}\nu}{3c_2^{1/2}}\right) n^{-(1+\rho)/2}n^{1/2}\right\} \\
&= \Pi_n\left\{\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F < \frac{\Delta}{n^{\rho/2}}\right\}, \qquad (1.7)
\end{aligned}$$

where $\Delta = \frac{2d_1^{1/2}\nu}{3c_2^{1/2}}$. The second inequality in (1.7) follows from Assumptions (A2) and (A3) and the fact that

$$\begin{aligned}
||\mathbf{X}_n(\mathbf{B}_n - \mathbf{B}_0)\mathbf{\Sigma}^{-1/2}||_F &= \sqrt{\text{tr}[\mathbf{\Sigma}^{-1/2}(\mathbf{B}_n - \mathbf{B}_0)^\top\mathbf{X}_n^T\mathbf{X}_n(\mathbf{B}_n - \mathbf{B}_0)\mathbf{\Sigma}^{-1/2}]} \\
&\leq \sqrt{\lambda_{\max}(\mathbf{X}_n^\top\mathbf{X}_n)\lambda_{\max}(\mathbf{\Sigma}^{-1})||\mathbf{B}_n - \mathbf{B}_0)||_F^2} \\
&< n^{1/2}c_2^{1/2}d_1^{-1/2}||\mathbf{B}_n - \mathbf{B}_0||_F.
\end{aligned}$$

Therefore, from (1.7), if $\Pi_n\left\{\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F < \frac{\Delta}{n^{\rho/2}}\right\} > \exp(-kn)$ for all $0 < k < b/2 - \nu$, then (1.5) will hold.

Substitute $b = \frac{\varepsilon^2 c_1}{16d_2}, \Delta = \frac{2d_1^{1/2}\nu}{3c_2^{1/2}} \Rightarrow \nu = \frac{3\Delta c_2^{1/2}}{2d_1^{1/2}}$ to obtain that $0 < k < \frac{\varepsilon^2 c_1}{32d_2} - \frac{3\Delta c_2^{1/2}}{2d_1^{1/2}}$. To ensure that $k > 0$, we must have $0 < \Delta < \frac{\varepsilon^2 c_1 d_1^{1/2}}{48c_2^{1/2}d_2}$.

Therefore, if the conditions on Δ and k in Theorem 1 are satisfied, then (1.4) holds. This ensures that the expected value of (1.3) w.r.t. $\mathbb{P}_0$ measure approaches 0 as $n \to \infty$, which ultimately establishes that posterior consistency holds if (11) is satisfied. □

1.2. Proof of Theorem 2

Proof of Theorem 2 also requires the creation of an appropriate test function. In this case, the test must be very carefully constructed since $\mathbf{X}_n$ is no longer nonsingular. We first define some constants and prove a lemma.

For arbitrary $\varepsilon > 0$ and $\widetilde{c}_1$ and d_2 specified in (B3) and (B5), let

$$\widetilde{c}_3 = \frac{\varepsilon^2\widetilde{c}_1}{16d_2}, \qquad (1.8)$$

and

$$m_n = \left\lfloor \frac{n\widetilde{c}_3}{6\ln p_n} \right\rfloor. \qquad (1.9)$$

Lemma 2. *Define the set* $\mathcal{B}_\varepsilon = \{\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F > \varepsilon\}$. *Suppose that Conditions (B1)-(B6) hold under model (10). In order to test* $H_0 : \mathbf{B}_n = \mathbf{B}_0$ *vs.* $H_1 : \mathbf{B}_n \in \mathcal{B}_\varepsilon$, *there exists a test function* $\widetilde{\Phi}_n$ *such that as* $n \to \infty$,

1. $\mathbb{E}_{\mathbf{B}_0}(\widetilde{\Phi}_n) \leq \exp(-n\widetilde{c}_3/2)$,
2. $\sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1 - \widetilde{\Phi}_n) \leq \exp(-n\widetilde{c}_3/4)$,

where $\widetilde{c}_3$ *is defined in (1.8).*

Proof of Lemma 2. By Condition (B1), we must have that $\frac{n}{\ln p_n} \to \infty$. Moreover, by (1.9), $m_n = o(n)$, since $\ln p_n \to \infty$ as $n \to \infty$. Combining this with assumption (B6), we must have that for sufficiently large n, there exists a positive integer $\widetilde{m}_n$, determined by Equation (1.9), such that $0 < s^* < \widetilde{m}_n < n$.

Fix $\varepsilon > 0$. For sufficiently large n so that $s^* < \widetilde{m}_n < n$, define the set $\mathcal{M}$ as the set of models S which properly contain the true model $S^* \subset \{1, \dots, p_n\}$ so that

$$\mathcal{M} = \left\{ S : S \supset S^*, S \neq S^*, |S| \leq \widetilde{m}_n, ||\mathbf{B}_n^{S^c}||_F \leq \frac{\varepsilon}{4(1 + \widetilde{c}_1^{-1}\widetilde{c}_2)} \right\}, \tag{1.10}$$

where S^c denotes $\{1, \dots, p_n\} \setminus S$ and $\widetilde{c}_1$ and $\widetilde{c}_2$ are from Assumptions (B3) and (B4).

Let $\mathbf{X}_S$ denote the submatrix of $\mathbf{X}$ with columns indexed by model S, and let $\mathbf{B}_0^S$ denote the submatrix of $\mathbf{B}_0$ that contains rows of $\mathbf{B}_0$ indexed by S. Define the following set $\mathcal{C}_n$:

$$\mathcal{C}_n = \bigvee_{S \in \mathcal{M}} \left\{ ||(\mathbf{X}_S^\top \mathbf{X}_S)^{-1} \mathbf{X}_S^\top \mathbf{Y}_n - \mathbf{B}_0^S||_F > \varepsilon/2 \right\}, \tag{1.11}$$

where $\bigvee$ indicates the union of all models S contained in $\mathcal{M}$. Essentially, the set $\mathcal{C}_n$ contains the union of all models S that contain the true model S^*, $S \neq S*$, such that the submatrix $\mathbf{X}_S$ has at least s^* columns and at most $\widetilde{m}_n (< n)$ columns, and $||(\mathbf{X}_S^\top \mathbf{X}_S)^{-1} \mathbf{X}_S^\top \mathbf{Y}_S - \mathbf{B}_0^S||_F > \varepsilon/2$, while $||\mathbf{B}_n^{S^c}||_F \leq \frac{\varepsilon}{4(1+\widetilde{c}_1^{-1}\widetilde{c}_2)}$. Given our choice of $\widetilde{m}_n$, $\mathbf{X}_S^T \mathbf{X}_S$ is nonsingular for all models S contained in our sets.

We are now ready to define our test function $\widetilde{\Phi}_n$. To test $H_0 : \mathbf{B}_n = \mathbf{B}_0$ vs. $H_1 : \mathbf{B}_n \in \mathcal{B}_\varepsilon$, define $\widetilde{\Phi}_n = 1(\mathbf{Y}_n \in \mathcal{C}_n)$, where the critical region is defined as in (1.11). We now show that Lemma 2 holds with this choice of $\widetilde{\Phi}$.

Let s be the size of an arbitrary model S. Noting also that there are $\binom{p_n}{s}$ ways to select a model of size s, we therefore have for sufficiently large n,

$$\begin{aligned}
\mathbb{E}_{\mathbf{B}_0}(\widetilde{\Phi}_n) &\leq \sum_{S \in \mathcal{M}} \mathbb{P}_{\mathbf{B}_0}\left(\mathbf{Y}_n : ||(\mathbf{X}_S^\top \mathbf{X}_S)^{-1} \mathbf{X}_S^\top \mathbf{Y}_n - \mathbf{B}_0^S||_F > \varepsilon/2\right) \\
&\leq \sum_{s=s^*+1}^{\widetilde{m}_n} \binom{p_n}{s} \mathbb{P}_{\mathbf{B}_0}\left(\mathbf{Y}_n : ||(\mathbf{X}_S^\top \mathbf{X}_S)^{-1} \mathbf{X}_S^\top \mathbf{Y}_n - \mathbf{B}_0^S||_F > \varepsilon/2\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{s=s^*+1}^{\widetilde{m}_n} \binom{p_n}{s} \mathbb{P}\left(\chi^2_{sq} > \frac{\varepsilon^2 \widetilde{c}_1 n}{4 d_2}\right) \\
&\leq \sum_{s=s^*+1}^{\widetilde{m}_n} \binom{p_n}{s} \exp(-n\widetilde{c}_3) \\
&\leq (\widetilde{m}_n - s^*) \binom{p_n}{\widetilde{m}_n} \exp(-n\widetilde{c}_3) \\
&\leq (\widetilde{m}_n - s^*) \left(\frac{e p_n}{\widetilde{m}_n}\right)^{\widetilde{m}_n} \exp(-n\widetilde{c}_3), \qquad (1.12)
\end{aligned}
$$

where we use Part 1 of Lemma 1 for the second inequality, the fact that $\mathbb{P}(\chi^2_m > x) \leq \exp(-x/4)$ when $x > 8m$ and $\widetilde{m}_n = o(n)$ for the third inequality, and the fact that $\sum_{i=k}^{m} \binom{n}{i} \leq (m-k+1)\binom{n}{m}$ for the fourth inequality in (1.12).

Since $\ln n = o(n)$, we must have for sufficiently large n that $\ln n < \frac{\widetilde{c}_3 n}{6}$. Then from the definition of $\widetilde{m}_n$, we have

$$
\begin{aligned}
\ln(\widetilde{m}_n - s^*) + \widetilde{m}_n \left(1 + \ln\left(\frac{p_n}{\widetilde{m}_n}\right)\right) &\leq \ln(\widetilde{m}_n) + \widetilde{m}_n \left(1 + \ln(p_n)\right) \\
\leq \ln(n) + \frac{\widetilde{c}_3 n}{6 \ln p_n} + \left(\frac{\widetilde{c}_3 n}{6 \ln p_n}\right) \ln(p_n) & \\
\leq \frac{\widetilde{c}_3 n}{6} + \frac{\widetilde{c}_3 n}{6} + \frac{\widetilde{c}_3 n}{6} = \frac{\widetilde{c}_3 n}{2}. & \qquad (1.13)
\end{aligned}
$$

Therefore, from (1.12) and (1.13), we must have that $\mathbb{E}_{\mathbf{B}_0}(\widetilde{\Phi}_n) \leq \exp(-n\widetilde{c}_3/2)$ as $n \to \infty$. This proves the first part of the lemma.

Let $\widetilde{S}$ be an arbitrary set in $\mathcal{M}$, as defined in (1.10) and $\widetilde{S}^c := \{1, \ldots, p_n\} \backslash \widetilde{S}$ be its complement. We observe that as $n \to \infty$,

$$
\begin{aligned}
&\sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1 - \widetilde{\Phi}_n) = \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}(\mathbf{Y}_n \notin \mathcal{C}_n) \\
&= \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\bigcap_{S \in \mathcal{M}} \left\{\mathbf{Y}_n : ||(\mathbf{X}_S^\top \mathbf{X}_S)^{-1} \mathbf{X}_S^\top \mathbf{Y}_n - \mathbf{B}_0^S||_F \leq \varepsilon/2\right\}\right) \\
&\leq \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{Y}_n - \mathbf{B}_0^{\widetilde{S}}||_F \leq \varepsilon/2\right) \\
&\leq \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{Y}_n - \mathbf{B}_0^{\widetilde{S}}||_F \leq \varepsilon/2\right) \\
&= \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{E}_n + \mathbf{B}_n^{\widetilde{S}} + (\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}^c} \mathbf{B}_n^{\widetilde{S}^c} - \mathbf{B}_0^{\widetilde{S}}||_F \leq \varepsilon/2\right) \\
&\leq \sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{E}_n||_F \geq ||\mathbf{B}_n^{\widetilde{S}} - \mathbf{B}_0^{\widetilde{S}}||_F - ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}^c} \mathbf{B}_n^{\widetilde{S}^c}||_F - \varepsilon/2\right). \qquad (1.14)
\end{aligned}
$$

We have $||\mathbf{B}_n^{\widetilde{S}} - \mathbf{B}_0^{\widetilde{S}}||_F = ||\mathbf{B}_n - \mathbf{B}_0||_F - ||\mathbf{B}_n^{\widetilde{S}^c}||_F$. Additionally, by Assumptions (B3) and (B4), $||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}^c} \mathbf{B}_n^{\widetilde{S}^c}||_F \leq c_1^{-1} c_2 ||\mathbf{B}_n^{\widetilde{S}^c}||_F$. Combining these with (1.14), we have that for $\mathcal{B}_\varepsilon = \{||\mathbf{B}_n - \mathbf{B}_0||_F > \epsilon\}$,

$$\begin{aligned}
\sup_{\mathbf{B}_n \in \mathcal{B}_\varepsilon} \mathbb{E}_{\mathbf{B}_n}(1 - \widetilde{\Phi}_n) &\leq \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{E}_n||_F \geq \varepsilon/2 - (1 + c_1^{-1} c_2)||\mathbf{B}_n^{\widetilde{S}^c}||_F\right) \\
&\leq \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{E}_n||_F \geq \varepsilon/4\right) \\
&= \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top (\mathbf{Y}_n - \mathbf{X}_{\widetilde{S}} \mathbf{B}_0^{\widetilde{S}})||_F \geq \varepsilon/4\right) \\
&= \mathbb{P}_{\mathbf{B}_n}\left(\mathbf{Y}_n : ||(\mathbf{X}_{\widetilde{S}}^\top \mathbf{X}_{\widetilde{S}})^{-1} \mathbf{X}_{\widetilde{S}}^\top \mathbf{Y}_n - \mathbf{B}_0^{\widetilde{S}}||_F \geq \varepsilon/4\right) \\
&\leq \mathbb{P}\left(\chi^2_{|\widetilde{S}|q} > \frac{\varepsilon^2 \widetilde{c}_1 n}{16 d_2}\right) \\
&\leq \exp(-n\widetilde{c}_3/4),
\end{aligned}$$

where $\widetilde{c}_3$ is from (1.8). Note that since $\widetilde{S} \in \mathcal{M}$, $|\widetilde{S}| \leq \widetilde{m_n} = o(n)$, and thus, we may invoke the fact that $\mathbb{P}(\chi^2_m > x) \leq \exp(-x/4)$ when $x > 8m$ in the final inequality. □

Proof of Theorem 2. In light of Lemma 2, we suitably modify Theorem 1 for the ultrahigh-dimensional case. Let $\widetilde{\Phi}_n$ be the test function defined in Lemma 2 for sufficiently large n. The posterior probability of $\mathcal{B}_n$ is given by

$$\begin{aligned}
\Pi_n(\mathcal{B}_n|\mathbf{Y}_n) &= \frac{\displaystyle\int_{\mathcal{B}_n} \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \Pi(d\mathbf{B}_n)}{\displaystyle\int \frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)} \Pi(d\mathbf{B}_n)} \\
&\leq \widetilde{\Phi}_n + \frac{(1 - \widetilde{\Phi}_n)\widetilde{J}_{\mathcal{B}_\varepsilon}}{\widetilde{J}_n} \\
&= \widetilde{I}_1 + \frac{\widetilde{I}_2}{\widetilde{J}_n}, \qquad (1.15)
\end{aligned}$$

where $\widetilde{J}_{\mathcal{B}_\varepsilon} = \displaystyle\int_{\mathcal{B}_\varepsilon} \left\{\frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)}\right\} \Pi(d\mathbf{B}_n)$ and $\widetilde{J}_n = \displaystyle\int \left\{\frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)}\right\} \Pi(d\mathbf{B}_n)$.

For sufficiently large n, using Markov's Inequality and the first part of Lemma 2, and taking $\widetilde{c}_3$ as defined in (1.8), we have

$$\mathbb{P}_{\mathbf{B}_0}\left(\widetilde{I}_1 \geq \exp\left(-\frac{n\widetilde{c}_3}{4}\right)\right) \leq \exp\left(\frac{n\widetilde{c}_3}{4}\right) \mathbb{E}_{\mathbf{B}_0}(\widetilde{I}_1) \leq \exp\left(-\frac{n\widetilde{c}_3}{4}\right).$$

This implies that $\sum_{n=1}^\infty \mathbb{P}_{\mathbf{B}_0}\left(\widetilde{I}_1 \geq \exp(-n\widetilde{c}_3/4)\right) < \infty$. Thus, by the Borel-Cantelli Lemma, we have $\mathbf{P}_{\mathbf{B}_0}(\widetilde{I}_1 \geq \exp(-n\widetilde{c}_3/4)$ infinitely often$) = 0$, i.e. $\widetilde{I}_1 \to 0$ a.s. $\mathbb{P}_0$ as $n \to \infty$.

We next look at the behavior of $\widetilde{I}_2$. We have

$$\mathbb{E}_{\mathbf{B}_0}\widetilde{I}_2 = \mathbb{E}_{\mathbf{B}_0}\left\{(1-\widetilde{\Phi}_n)\widetilde{J}_{\mathcal{B}_\varepsilon}\right\} = \mathbb{E}_{\mathbb{B}_0}\left\{(1-\widetilde{\Phi}_n)\int_{\mathcal{B}_\varepsilon}\frac{f(\mathbf{Y}_n|\mathbf{B}_n)}{f(\mathbf{Y}_n|\mathbf{B}_0)}\Pi_n(d\mathbf{B}_n)\right\}$$
$$= \int_{\mathcal{B}_\varepsilon}\int(1-\widetilde{\Phi}_n)f(\mathbf{Y}_n|\mathbf{B}_n)\, d\mathbf{Y}_n\, \Pi_n(d\mathbf{B}_n)$$
$$\leq \pi_n(\mathcal{B}_\varepsilon)\sup_{\mathbf{B}_n\in\mathcal{B}_\varepsilon}\mathbb{E}_{\mathbf{B}_n}(1-\widetilde{\Phi}_n)$$
$$\leq \sup_{\mathbf{B}_n\in\mathcal{B}_\varepsilon}\mathbb{E}_{\mathbf{B}_n}(1-\widetilde{\Phi}_n)$$
$$\leq \exp(-n\widetilde{c}_3/4),$$

where the last inequality follows from the second part of Lemma 2, and $\widetilde{c}_3$ is again from (1.8).

Thus, for sufficiently large n, $\mathbb{P}_{\mathbf{B}_0}(\widetilde{I}_2 \geq \exp(-n\widetilde{c}_3/8)) \leq \exp(-n\widetilde{c}_3/8)$, which implies that $\sum_{n=1}^{\infty}\mathbb{P}_{\mathbf{B}_0}\left(\widetilde{I}_2 \geq \exp(-n\widetilde{c}_3/8)\right) < \infty$. Thus, by the Borel-Cantelli Lemma, $\widetilde{I}_2 \to 0$ a.s. $\mathbb{P}_0$ as $n \to \infty$.

We have now shown that both $\widetilde{I}_1$ and $\widetilde{I}_2$ in (1.15) tend towards zero exponentially fast. We now analyze the behavior of $\widetilde{J}_n$. To complete the proof, we need to show that

$$\exp(n\widetilde{c}_3/8)J_n \to \infty \;\; \mathbb{P}_0 \text{ a.s. as } n \to \infty. \tag{1.16}$$

Note that

$$\exp(n\widetilde{c}_3/8)\widetilde{J}_n = \exp(n\widetilde{c}_3/8)\int \exp\left\{-n\frac{1}{n}\ln\frac{f(\mathbf{Y}_n|\mathbf{B}_0)}{f(\mathbf{Y}_n|\mathbf{B}_n)}\right\}\Pi_n(d\mathbf{B}_n)$$
$$\geq \exp\left\{(\widetilde{c}_3/8-\nu)n\right\}\Pi_n(\widetilde{\mathcal{D}}_{n,\nu}), \tag{1.17}$$

where $\widetilde{\mathcal{D}}_{n,\nu} = \left\{\mathbf{B}_n : n^{-1}\ln\left(\frac{f(\mathbf{Y}_n|\mathbf{B}_0)}{f(\mathbf{Y}_n|\mathbf{B}_n)}\right) < \nu\right\}$ for $0 < \nu < \widetilde{c}_3/8$.

Because of Assumption (B4) which bounds the maximum singular value of $\mathbf{X}_n$ from above by $n\widetilde{c}_2$, the rest of the proof is essentially identical to the remainder of the proof from Theorem 1, with suitable modifications (i.e. replacing c_1 with $\widetilde{c}_1$ and c_2 with $\widetilde{c}_2$ and substituting in the expression in (1.8) for $\widetilde{c}_3$).

Therefore, if the conditions on $\widetilde{\Delta}$ and k in Theorem 2 are satisfied, then (1.17) is satisfied, i.e. $\exp(n\widetilde{c}_3/8)J_n \to \infty$ as $n \to \infty$. This ensures that the expected value of (1.16) w.r.t. $\mathbb{P}_0$ measure approaches 0 as $n \to \infty$, which ultimately establishes that posterior consistency holds if (12) is satisfied. □

2. Proofs for Section 3.4

2.1. Preliminary Lemmas

Before proving Theorems 3 and 4, we first prove two lemmas which characterize the marginal prior density for the rows of **B**. Throughout this section, we denote $\mathbf{b}_i, 1 \leq i \leq p$ as the ith row of $\mathbf{B}$ under (5), with polynomial-tailed hyperpriors of the form (3). Lemma 4 in particular plays a central role in proving our theoretical results in Section 3 of the main article.

Lemma 3. *Under model (5) with polynomial-tailed hyperpriors (3), the marginal density* $\pi(\mathbf{b}_i|\mathbf{\Sigma})$ *is equal to*

$$\pi(\mathbf{b}_i|\mathbf{\Sigma}) = D \int_0^\infty \xi_i^{-q/2-a-1} \exp\left\{-\frac{1}{2\xi_i \tau}||\mathbf{b}_i \mathbf{\Sigma}^{-1/2}||_2^2\right\} L(\xi_i) d\xi_i,$$

where $D > 0$ *is an appropriate constant.*

Proof of Lemma 3. Let $\mathbf{D} = \text{diag}(\xi_1, ..., \xi_p)$. Using Definition 2, the joint prior for (5) with polynomial-tailed priors is

$$\begin{aligned}
\pi(\mathbf{B}, \xi_1, ..., \xi_p|\mathbf{\Sigma}) &\propto |\mathbf{D}|^{-q/2} |\mathbf{\Sigma}|^{-p/2} \exp\left\{-\frac{1}{2}\text{tr}[\mathbf{\Sigma}^{-1}\mathbf{B}^T \tau^{-1} \mathbf{D}^{-1} \mathbf{B}]\right\} \times \prod_{i=1}^p \pi(\xi_i) \\
&\propto \left[\prod_{i=1}^p (\xi_i)^{-q/2}\right] \exp\left\{-\frac{1}{2\tau} \sum_{i=1}^p ||\xi_i^{-1/2} \mathbf{b}_i \mathbf{\Sigma}^{-1/2}||_2^2\right\} \times \prod_{i=1}^p \pi(\xi_i) \\
&\propto \prod_{i=1}^p \left[\xi_i^{-q/2} \exp\left\{-\frac{1}{2\xi_i \tau}||\mathbf{b}_i \mathbf{\Sigma}^{-1/2}||_2^2\right\} \pi(\xi_i)\right] \\
&\propto \prod_{i=1}^p \left[\xi_i^{-q/2-a-1} \exp\left\{-\frac{1}{2\xi_i \tau}||\mathbf{b}_i \mathbf{\Sigma}^{-1/2}||_2^2\right\} L\xi_i)\right]. \qquad (2.1)
\end{aligned}$$

Since the rows, $\mathbf{b}_i$ and the ξ_i's, $1 \leq i \leq p$ are independent, we have from (2.1) that

$$\pi(\mathbf{b}_i, \xi_i|\mathbf{\Sigma}) \propto \xi_i^{-q/2-a-1} \exp\left\{-\frac{1}{2\xi_i \tau}||\mathbf{b}_i \mathbf{\Sigma}^{-1/2}||_2^2\right\}.$$

Integrating out ξ_i gives the desired marginal prior for $\pi(\mathbf{b}_i|\mathbf{\Sigma})$. □

Though we are not able to obtain a closed form solution for $\pi(\mathbf{b}_i|\mathbf{\Sigma})$, we *are* able to obtain a lower bound on it that can be written in closed form, as we illustrate in the next lemma.

Lemma 4. *Suppose Condition (A3) on the eigenvalues of* $\mathbf{\Sigma}$ *and Condition (C1) on the slowly varying function* $L\cdot)$ *in (3) hold . Under model (5) with polynomial-tailed hyperpriors (3) and known* $\mathbf{\Sigma}$*, the marginal density for* $\mathbf{b}_i$*, the* i*th row of* $\mathbf{B}$*, can be bounded below by*

$$\widetilde{C} \exp\left(-\frac{||\mathbf{b}_i||_2^2}{2\tau d_1 t_0}\right), \qquad (2.2)$$

where $\widetilde{C} = D c_0 t_0^{-q/2-a} \left(\frac{q}{2} + a\right)^{-1}$.

Proof of Lemma 4. Following from Lemma 3, we have

$$
\begin{aligned}
\pi(\mathbf{b}_i) &= D \int_0^\infty \xi_i^{-q/2-a-1} \exp\left\{ -\frac{1}{2\xi_i \tau} ||\mathbf{b}_i \boldsymbol{\Sigma}^{-1/2}||_2^2 \right\} L(\xi_i) d\xi_i \\
&\geq D \int_0^\infty \xi_i^{-q/2-a-1} \exp\left\{ -\frac{||\mathbf{b}_i||_2^2}{2\xi_i \tau d_1} \right\} L(\xi_i) d\xi_i \qquad (2.3) \\
&\geq D \int_{t_0}^\infty \xi_i^{-q/2-a-1} \exp\left\{ -\frac{||\mathbf{b}_i||_2^2}{2\xi_i \tau d_1} \right\} L(\xi_i) d\xi_i \\
&\geq D c_0 \int_{t_0}^\infty \xi_i^{-q/2-a-1} \exp\left\{ -\frac{||\mathbf{b}_i||_2^2}{2\xi_i \tau d_1} \right\} d\xi_i \qquad (2.4) \\
&= D c_0 \left(\frac{2\tau d_1}{||\mathbf{b}_i||_2^2} \right)^{q/2+a} \int_0^{||\mathbf{b}_i||_2^2/2\tau d_1 t_0} u^{q/2+a-1} e^{-u} du \qquad (2.5) \\
&\geq D c_0 \left(\frac{2\tau d_1}{||\mathbf{b}_i||_2^2} \right)^{q/2+a} \exp\left(-\frac{||\mathbf{b}_i||_2^2}{2\tau d_1 t_0} \right) \int_0^{||\mathbf{b}_i||_2^2/2\tau d_1 t_0} u^{q/2+a-1} du \\
&= D c_0 t_0^{-q/2-a} \left(\frac{q}{2} + a \right)^{-1} \exp\left(-\frac{||\mathbf{b}_i||_2^2}{2\tau d_1 t_0} \right) \\
&= \widetilde{C} \exp\left(-\frac{||\mathbf{b}_i||_2^2}{2\tau d_1 t_0} \right).
\end{aligned}
$$

where (2.3) follows from Condition (A3), while (2.4) follows from Condition (C1). (2.5) follows from a change of variables $u = \frac{||\mathbf{b}_i||_2^2}{2\xi_i \tau d_1}$. We have thus established the lower bound (2.2) for the marginal density of $\mathbf{b}_i$. □

2.2. Proofs for Theorem 3 and Theorem 4

Before we prove Theorems 3 and 4 for the MBSP model (5) with hyperpriors (3), we first introduce some notation. Because we are operating under the assumption of sparsity, most of the rows of $\mathbf{B}_0$ should contain only entries of zero.

Our proofs depend on partitioning $\mathbf{B}_0$ into sets of active and inactive predictors. To this end, let $\mathbf{b}_j^0$ denote the jth row for the true coefficient matrix $\mathbf{B}_0$ and $\mathbf{b}_j^n$ denote the jth row of $\mathbf{B}_n$, where both $\mathbf{B}_0$ and $\mathbf{B}_n$ depend on n. We also let $\mathcal{A}_n := \left\{ j : \mathbf{b}_j^n \neq \mathbf{0}, 1 \leq j \leq p_n \right\}$ denote the set of indices of the nonzero rows of $\mathbf{B}_0$. This indicates active predictors. Equivalently, $\mathcal{A}_n^c$ is the set of indices of the zero rows (or the inactive predictors).

Proof of Theorem 3. For the low-dimensional setting, let $s = |S|$ denote the size of the true model. Since (A1)-(A3) hold, it is enough to show (by Theorem 1) that, for sufficiently large n and *any* $k > 0$,

$$\Pi_n \left(\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F < \frac{\Delta}{n^{\rho/2}} \right) > \exp(-kn),$$

where $0 < \Delta < \frac{\varepsilon^2 c_1 d_1^{1/2}}{48 c_2^{1/2} d_2}$. We have

$$\begin{aligned}
\Pi_n\left(\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F < \frac{\Delta}{n^{\rho/2}}\right) &= \Pi_n\left(\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F^2 < \frac{\Delta^2}{n^{\rho}}\right) \\
&= \Pi_n\left(\mathbf{B}_n : \sum_{j\in\mathcal{A}_n} ||\mathbf{b}_j^n - \mathbf{b}_j^0||_2^2 + \sum_{j\in\mathcal{A}_n^c} ||\mathbf{b}_j^n||_2^2 < \frac{\Delta^2}{n^{\rho}}\right) \\
&\geq \Pi_n\left(\mathbf{B}_n : \sum_{j\in\mathcal{A}_n} ||\mathbf{b}_j^n - \mathbf{b}_j^0||_2^2 < \frac{s\Delta^2}{p_n n^{\rho}}, \sum_{j\in\mathcal{A}_n^c} ||\mathbf{b}_j||_2^2 < \frac{(p_n - s)\Delta^2}{p_n n^{\rho}}\right) \\
&\geq \left\{\prod_{j\in\mathcal{A}_n} \Pi_n\left(\mathbf{b}_j^n : ||\mathbf{b}_j^n - \mathbf{b}_j^0||_2^2 < \frac{\Delta^2}{p_n n^{\rho}}\right)\right\} \\
&\quad \times \left\{\Pi_n\left(\sum_{j\in\mathcal{A}_n^c} ||\mathbf{b}_j||_2^2 < \frac{(p_n - s)\Delta^2}{p_n n^{\rho}}\right)\right\}. \qquad (2.6)
\end{aligned}$$

Define the density

$$\widetilde{\pi}(\mathbf{b}_j) \propto \exp\left(-\frac{||\mathbf{b}_j||_2^2}{2\tau_n d_1 t_0}\right), \qquad (2.7)$$

Since (C1) holds for the slowly varying component of (3), we have by the lower bound in Lemma 4, (2.6), and (2.7) that it is sufficient to show that

$$\begin{aligned}
&\left\{\widetilde{\Pi}_n\left(\mathbf{b}_j^n : ||\mathbf{b}_j^n - \mathbf{b}_j^0||_2^2 < \frac{\Delta^2}{p_n n^{\rho}}\right)\right\}^s \\
&\quad \times \left\{\widetilde{\Pi}_n\left(\sum_{j\in\mathcal{A}_n^c} ||\mathbf{b}_j||_2^2 < \frac{(p_n - s)\Delta^2}{p_n n^{\rho}}\right)\right\} > \exp(-kn) \qquad (2.8)
\end{aligned}$$

for sufficiently large n and *any* $k > 0$ in order to obtain posterior consistency (again by Theorem 1). We consider the last two terms in the product on the left-hand side of (2.8) separately. Note that

$$\begin{aligned}
\widetilde{\Pi}_n\left(b_{jk}^n : \sum_{k=1}^{q} (b_{jk}^n - b_{jk}^0)^2 < \frac{\Delta^2}{p_n n^{\rho}}\right) &\geq \widetilde{\Pi}_n\left(b_{jk}^n : \sum_{k=1}^{q} |b_{jk}^n - b_{jk}^0| < \frac{\Delta}{\sqrt{p_n n^{\rho}}}\right) \\
&\geq \prod_{k=1}^{q}\left\{\widetilde{\Pi}_n\left(b_{jk}^n : |b_{jk}^n - b_{jk}^0| < \frac{\Delta}{q\sqrt{p_n n^{\rho}}}\right)\right\} \qquad (2.9)
\end{aligned}$$

By (2.7), $\widetilde{\pi}(b_{jk}^n) = \frac{1}{\sqrt{2\pi d_1 t_0}} \exp\left(-\frac{(b_{jk}^n)^2}{2 d_1 t_0}\right)$, i.e. $b_{jk}^n \sim N(0, \tau_n d_1 t_0)$. Therefore, we have

$$\begin{aligned}
&\widetilde{\Pi}_n \left(b_{jk}^n : |b_{jk}^n - b_{jk}^0| < \frac{\Delta}{q\sqrt{p_n n^\rho}} \right) \\
&= \frac{1}{\sqrt{2\pi\tau_n d_1 t_0}} \int_{b_{jk}^0 - \frac{\Delta}{q\sqrt{p_n n^\rho}}}^{b_{jk}^0 + \frac{\Delta}{q\sqrt{p_n n^\rho}}} \exp\left(- \frac{\left(b_{jk}^n\right)^2}{2\tau_n d_1 t_0} \right) \mathrm{d}b_{jk} \\
&= \Pr\left(- \frac{\Delta}{q\sqrt{p_n n^\rho}} \leq X - b_{jk}^0 \leq \frac{\Delta}{q\sqrt{p_n n^\rho}} \right) \\
&= \Pr\left(|X - b_{jk}^0| \leq \frac{\Delta}{q\sqrt{p_n n^\rho}} \right), \qquad (2.10)
\end{aligned}$$

where $X \sim N(b_{jk}^0, \tau_n d_1 t_0)$. By Assumption (C2), b_{jk}^0 is finite for every n. Furthermore, for any random variable $X \sim N(\mu, \sigma^2)$, we have the concentration inequality $\Pr|X - \mu| > t) \leq 2e^{-\frac{t^2}{2\sigma^2}}$ for any $t \in \mathbb{R}$. Setting $X \sim N(b_{jk}^0, \tau_n d_1 t_0)$ and $t = \frac{\Delta}{q\sqrt{p_n n^\rho}}$, we have

$$\begin{aligned}
\Pr\left(|X - b_{jk}^0| \leq \frac{\Delta}{q\sqrt{p_n n^\rho}} \right) &= 1 - \Pr\left(|X - b_{jk}^0| \geq \frac{\Delta}{q\sqrt{p_n n^\rho}} \right) \\
&\geq 1 - 2\exp\left(- \frac{\Delta^2}{2q^2 p_n n^\rho \tau_n d_1 t_0} \right). \qquad (2.11)
\end{aligned}$$

We now consider the second term on the left in (2.8). Since $\mathbb{E}(b_{jk}^2) = \tau_n d_1 t_0$, an application of the Markov inequality gives

$$\begin{aligned}
\widetilde{\Pi}_n \left(b_{jk}^n : \sum_{j \in \mathcal{A}_n^c} \sum_{k=1}^q (b_{jk}^n)^2 < \frac{(p_n - s)\Delta^2}{p_n n^\rho} \right) &\geq 1 - \frac{p_n n^\rho \mathbb{E}\left(\sum_{j \in \mathcal{A}_n^c} \sum_{k=1}^q \left(b_{jk}^n\right)^2 \right)}{(p_n - s)\Delta^2} \\
&= 1 - \frac{p_n q n^\rho \tau_n d_1 t_0}{\Delta^2}. \qquad (2.12)
\end{aligned}$$

Combining (2.8)-(2.12), we obtain as a lower bound for the left-hand side of (2.8),

$$\left\{ 1 - 2\exp\left(- \frac{\Delta^2}{2q^2 p_n n^\rho \tau_n d_1 t_0} \right) \right\}^{qs} \left(1 - \frac{p_n q n^\rho \tau_n d_1 t_0}{\Delta^2} \right). \qquad (2.13)$$

By Assumption (C3), it is clear that (2.13) tends to 1 as $n \rightarrow \infty$, so obviously this quantity is greater than e^{-kn} for any $k > 0$ for sufficiently large n. Since the lower bound in (2.8) holds for all sufficiently large n, we have under the given conditions that the MBSP model (5) achieves posterior consistency in the Frobenius norm. □

Proof of Theorem 4. For the ultra high-dimensional setting, we first let $S^* \subset \{1, 2, ..., p_n\}$ denote the indices of the nonzero rows, and denote the true size of S^* as $s^* = |S^*|$. Since (B1)-(B6) hold, it is enough to show by Theorem 2 that, for sufficiently large n and *any* $k > 0$,

$$\Pi_n \left(\mathbf{B}_n : ||\mathbf{B}_n - \mathbf{B}_0||_F < \frac{\widetilde{\Delta}}{n^{\rho/2}} \right) > \exp(-kn),$$

where $0 < \widetilde{\Delta} < \frac{\varepsilon^2 \widetilde{c}_1 d_1^{1/2}}{48 \widetilde{c}_2^{1/2} d_2}$ and $\rho > 0$.

By Assumption (C1) for the slowly varying component in (3), Lemma 4, and Theorem 2, it is thus sufficient to show that

$$\begin{aligned} &\left\{ \widetilde{\Pi}_n \left(\mathbf{b}_j^n : ||\mathbf{b}_j^n - \mathbf{b}_j^0||_2^2 < \frac{\widetilde{\Delta}^2}{p_n n^\rho} \right) \right\}^{s^*} \\ &\times \left\{ \widetilde{\Pi}_n \left(\sum_{j \in \mathcal{A}_n^c} ||\mathbf{b}_j||_2^2 < \frac{(p_n - s^*)\widetilde{\Delta}^2}{p_n n^\rho} \right) \right\} > \exp(-kn) \end{aligned} \tag{2.14}$$

for sufficiently large n and any $k > 0$, where the density $\widetilde{\pi}_n$ is defined in (2.7). Mimicking the proof for Theorem 3 and given regularity conditions (C1) and (C2), we obtain as a lower bound for the left-hand side of (2.14),

$$\left\{ 1 - 2 \exp \left(- \frac{\widetilde{\Delta}^2}{2q^2 p_n n^\rho \tau_n d_1 t_0} \right) \right\}^{qs^*} \left(1 - \frac{p_n q n^\rho \tau_n d_1 t_0}{\widetilde{\Delta}^2} \right). \tag{2.15}$$

Under Assumption (C3), (2.15) is clearly greater than e^{-kn} for any $k > 0$ and sufficiently large n, since (2.15) tends to 1 as $n \to \infty$. We have thus proven posterior consistency in the Frobenius norm for the ultra high-dimensional case as well. □

3. Gibbs Sampler for the MBSP-TPBN Model

Here, we provide the technical details of the Gibbs sampler for the MBSP-TPBN model (16) and present an efficient method for sampling from the full conditional of $\mathbf{B}$. These algorithms are implemented in the `R` package `MBSP`.

3.1. Full Conditional Densities for the Gibbs Sampler

The full conditional densities are all available in closed form as follows. Letting $\mathbf{T} = \text{diag}(\psi_1, \ldots, \psi_p)$, $\mathbf{b}_i$ denote the ith row of $\mathbf{B}$, and $\mathcal{GIG}(a, b, p)$ denote a generalized inverse Gaussian density with $f(x; a, b, p) \propto x^{(p-1)} e^{-(a/x + bx)/2}$, we have

$$
\begin{aligned}
\mathbf{B}|\text{rest} &\sim \mathcal{MN}_{p\times q}\left((\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1})^{-1}\mathbf{X}^\top\mathbf{Y}, \left(\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1}\right)^{-1}, \mathbf{\Sigma}\right), \\
\mathbf{\Sigma}|\text{rest} &\sim \mathcal{IW}(n+p+d, (\mathbf{Y}-\mathbf{XB})^\top(\mathbf{Y}-\mathbf{XB})+\mathbf{B}^\top\mathbf{T}^{-1}\mathbf{B}+k\mathbf{I}_q), \\
\psi_i|\text{rest} &\overset{\text{ind}}{\sim} \mathcal{GIG}\left(||\mathbf{b}_i\mathbf{\Sigma}^{-1/2}||_2^2, 2\zeta_i, u-\frac{q}{2}\right), i=1,\ldots,p, \\
\zeta_i|\text{rest} &\overset{\text{ind}}{\sim} \mathcal{G}\left(a, \tau+\psi_i\right), i=1,\ldots,p.
\end{aligned}
\tag{3.1}
$$

Because the full conditional densities are available in closed form, we can implement the MBSP-TPBN model (16) straightforwardly using Gibbs sampling.

3.2. *Fast Sampling of the Full Conditional Density for* $\mathbf{B}$

In (3.1), the most computationally intensive operation is sampling from the density, $\pi(\mathbf{B}|\text{rest})$. Much of the computational cost comes from computing the inverse $(\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1})^{-1}$, which requires $O(p^3)$ time complexity if we use Cholesky factorization methods. In the case where $p<n$, this is not a problem. However, when $p \gg n$, then this operation can be prohibitively costly.

In this section, we provide an alternative algorithm for sampling from the density $\mathcal{MN}_{p\times q}\left((\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1})^{-1}\mathbf{X}^\top\mathbf{Y}, \left(\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1}\right)^{-1}, \mathbf{\Sigma}\right)$ in $O(n^2p)$ time. Bhattacharya et al. [2] originally devised an algorithm to efficiently sample from a class of structured multivariate Gaussian distributions. Our algorithm below is a matrix-variate extension of the algorithm given by Bhattacharya et al. [2].

Algorithm 1

Step 1. Sample $\mathbf{U} \sim \mathcal{MN}_{p\times q}(\mathbf{O},\mathbf{T},\mathbf{\Sigma})$ and $\mathbf{M} \sim \mathcal{MN}_{n\times q}(\mathbf{O},\mathbf{I}_n,\mathbf{\Sigma})$.

Step 2. Set $\mathbf{V} = \mathbf{XU}+\mathbf{M}$.

Step 3. Solve for $\mathbf{W}$ in the below system of equations:

$$(\mathbf{XTX}^\top+\mathbf{I}_n)\mathbf{W} = \mathbf{Y}-\mathbf{V}.$$

Step 4. Set $\mathbf{\Theta} = \mathbf{U}+\mathbf{TX}^\top\mathbf{W}$.

With the above algorithm, we have the following proposition.

Proposition 1. *Suppose* $\mathbf{\Theta}$ *is obtained by following Algorithm 1. Then* $\mathbf{\Theta} \sim \mathcal{MN}_{p\times q}\left((\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1})^{-1}\mathbf{X}^\top\mathbf{Y}, \left(\mathbf{X}^\top\mathbf{X}+\mathbf{T}^{-1}\right)^{-1}, \mathbf{\Sigma}\right)$.

Proof. This follows from a trivial modification of Proposition 1 in [2]. □

From Algorithm 1, it is clear that the most computationally intensive step is solving the system of equations in Step 3. However, since $\mathbf{T}$ is a diagonal matrix, it follows from the arguments in [2] that computing the inverse of $(\mathbf{XTX}^\top+\mathbf{I}_n)$ can be done in $O(n^2p)$ time. Once this inverse is obtained, solving the system of equations can be done in $O(n^2q)$ time, and in general, $q \ll p$. It is thus clear that Algorithm 1 is $O(n^2p)$ when $p > n$. Since our algorithm scales linearly with p, it provides a significant reduction in computing time from typical methods based on Cholesky factorization when $p \gg n$.

On the other hand, if $p < n$, then Algorithm 1 provides no time saving, so we simply utilize Cholesky factorization methods to sample from the full conditional density, $\pi(\mathbf{B}|\text{rest})$ in $O(p^3)$ time if $p < n$.

3.3. Convergence of the Gibbs Sampler

In order to ensure quick convergence, we need good initial guesses for $\mathbf{B}$ and $\mathbf{\Sigma}$, $\mathbf{B}^{(\text{init})}$ and $\mathbf{\Sigma}^{(\text{init})}$, respectively. We take as our initial guess for $\mathbf{B}$, $\mathbf{B}^{(\text{init})} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{Y}$, where $\lambda = \delta + \lambda_{\min^+}(\mathbf{X})$, $\lambda_{\min^+}(\mathbf{X})$ is the smallest positive singular value of $\mathbf{X}$, and $\delta = 0.01$. This forces the term $\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_p$ to be positive definite. For $\mathbf{\Sigma}$, we take as our initial guess $\mathbf{\Sigma}^{(\text{init})} = \frac{1}{n} \left(\mathbf{Y} - \mathbf{X}\mathbf{B}^{(\text{init})} \right)^\top \left(\mathbf{Y} - \mathbf{X}\mathbf{B}^{(\text{init})} \right)$.

Figure 1 shows the historical plots of the first 10,000 draws from from the Gibbs sampler for the MBSP-TPBN model described in Section 3.1 for four randomly drawn coefficients b_{ij} in $\mathbf{B}$ from experiments 5 and 6 in Section 5.1. The top two plots correspond to a true nonzero coefficient ($b_{ij}^0 = -3.8103$) and a true zero coefficient ($b_{ij}^0 = 0$) from experiment 5 in 5.1 ($n = 100, p = 500, q = 3$). The bottom two plots correspond to a true nonzero coefficient ($b_{ij}^0 = 3.1436$) and a true zero coefficient ($b_{ij}^0 = 0$) from experiment 6 in ($n = 150, p = 1000, q = 4$).

We consider two different Markov chains with different starting values for $\mathbf{B}^{(\text{init})}$: 1) the ridge estimator described above, and 2) the regularized MLASSO estimator described in Section 5.1. We see from the plots that although both chains start with different initial values of $b_{ij}^{(\text{init})}$, they mix well and seem to rapidly converge to a stationary distribution that captures the true coefficients b_{ij}^0's with high probability.

References

[1] A. Armagan, D.B. Dunson, J. Lee, W.U. Bajwa, and N. Strawn. Posterior consistency in linear models under shrinkage priors. *Biometrika*, 100(4): 1011–1018, 07 2013. URL `https://doi.org/10.1093/biomet/ast028`.

[2] Anirban Bhattacharya, Antik Chakraborty, and Bani K. Mallick. Fast sampling with gaussian scale mixture priors in high-dimensional regression. *Biometrika*, 103(4):985–991, 2016. doi: 10.1093/biomet/asw042. URL `http://dx.doi.org/10.1093/biomet/asw042`.

[3] Gyuhyeong Goh, Dipak K. Dey, and Kun Chen. Bayesian sparse reduced rank multivariate regression. *J. Multivar. Anal.*, 157:14 – 28, 2017. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2017.02.007. URL `http://www.sciencedirect.com/science/article/pii/S0047259X17301112`.

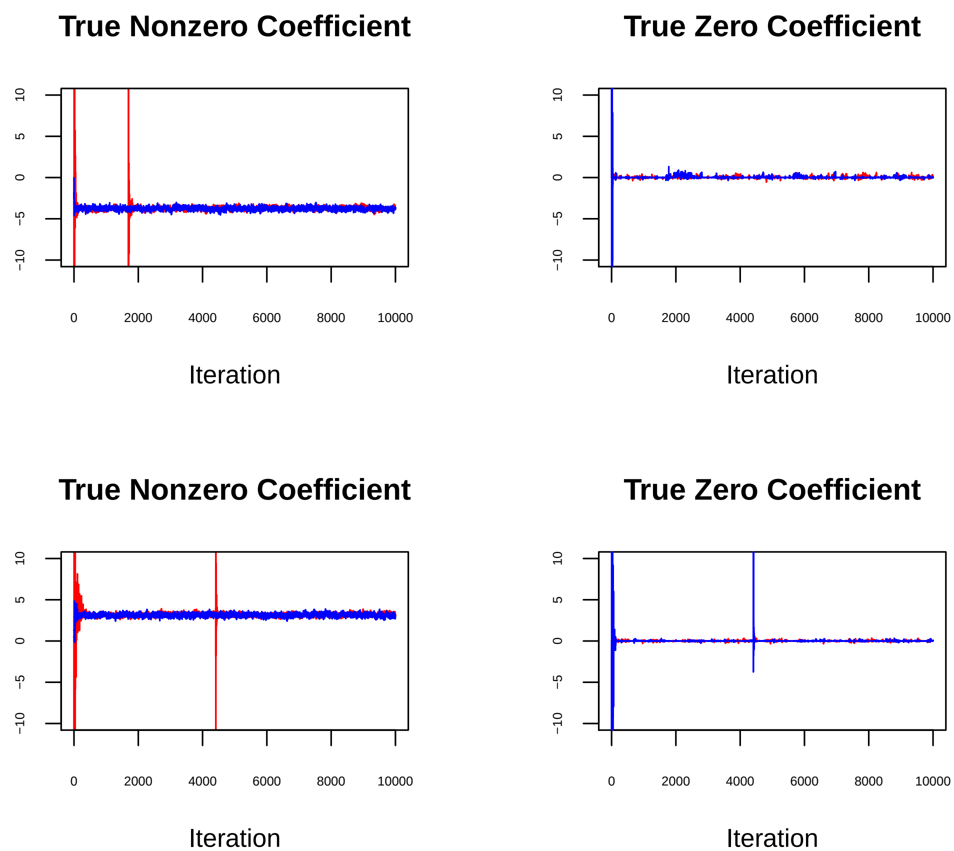

Figure 1: History plots of the first 10,000 draws from the Gibbs sampler for the MBSP-TPBN model described in Section 3.1 for randomly drawn coefficients b_{ij} in $\mathbf{B}_0$ from experiments 5 and 6 in Section 5.1. The top two plots are taken from experiment 5 ($n = 100, p = 500, q = 3$), and the bottom two plots are taken from Experiment 6 ($n = 150, p = 1000, q = 4$).