Supplementary material to "A mass-shifting phenomenon of truncated multivariate normal priors"

Shuang Zhou*[1], Pallavi Ray†[2], Debdeep Pati‡[3], and Anirban Bhattacharya§[3]

[1]School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona, 85287, USA

[2]Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana, 46285, USA

[3]Department of Statistics, Texas A&M University, College Station, Texas, 77843, USA

S1 Summary

In this supplementary document, we collect additional plots, proofs of main results in the manuscript, remaining technical results and additional details on numerical studies. Equations in the main documents are cited as (1), (2) etc., while new equations defined in this document are numbered (S1), (S2) etc.

- § S2 provides supportive materials to the main document, including empirical illustrations of mass-shifting behavior of the tMVN marginal density and an additional mass-shifting theorem for the tMVN families with unequal variances, and additional graphical illustrations in the main document.

- § S3 summarizes useful intermediate results for proving main theorems and corollaries in the manuscript.

*szhou98@asu.edu

†ray_pallavi@lilly.com (This work was done by Pallavi Ray as a former Ph.D. student at Department of Statistics, Texas A&M University.)

‡debdeep@stat.tamu.edu

§anirbanb@stat.tamu.edu

- § S4 collects the proofs of Theorems 2,3 in § 2, Theorem 5 in § 3, Theorem 7 and Theorem 8 in § 4, and Corollary S1 in § S2.2.

- § S5 contains the proofs of all Proposition results which includes Proposition 1 in § 2, Proposition 4 and Proposition 6 in § 3, Proposition S1 in § S4.2, and Proposition S2 in § S4.5.

- § S6 contains the proofs of technical results in Appendix A and the intermediate results in § S3.

- The rest of auxiliary results used in the proofs are listed in § S7. Finally, § S8 provides additional empirical details on prior illustration, posterior computation, remaining numerical results, hyperprior choices and additional sensitivity studies, and diagnostics on MCMC algorithms.

S2 Supporting materials to the main document

S2.1 Empirical illustration of mass-shifting phenomenon of truncated multivariate Normal

We now empirically illustrate the conclusion of Theorem 2 by presenting the univariate marginal density $\widetilde{p}_{1,N}$ for different values of the dimension N and the bandwidth K. The density calculations were performed using the **R** package **tmvtnorm**, which is based on the numerical approximation algorithm proposed in [3] and subsequent refinements in [6, 7, 8]. We consider an N-dimensional

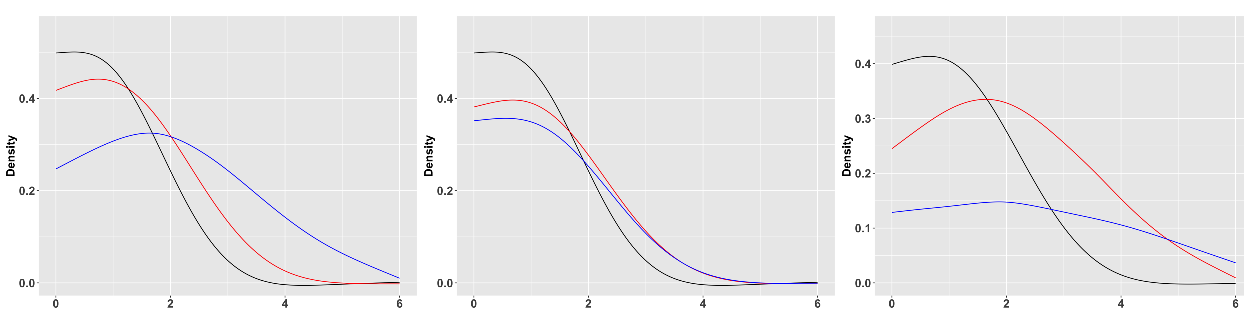

Figure S1: *Left panel shows marginal density functions $\widetilde{p}_{1,N}$ for $K = 2$ (black), $K = 5$ (red) and $K = 20$ (blue) with $N = 100$. Middle panel shows $\widetilde{p}_{1,N}$ for $N = 10$ (black), $N = 50$ (red) and $N = 100$ (blue) with $K = 5$. Right panel shows $\widetilde{p}_{1,N}$ for $(K, N) = (5, 25)$ (black), $(20, 100)$ (red) and $(50, 250)$ (blue).*

correlation matrix $\Sigma_N = (\sigma_{ij})$ which is K-banded with $\sigma_{ii} = 1 + \sigma_0^2$ with $\sigma_0^2 = 10^{-6}$ for $1 \leq i \leq N$

and $\sigma_{ij} = \rho^{|i-j|\wedge 10}$ for $\rho = 0.9$ and for all i, j such that $1 \le |i-j| < K$; and $\sigma_{ij} = 0$ otherwise.[1] The left panel of Figure S1 shows that for N fixed at a moderately large value, the probability assigned to a small neighborhood of the origin decreases with increasing K. Also, the mode of the marginal density increasingly shifts away from zero. A similar effect is seen for a fixed K and increasing N in the middle panel and also for an increasing pair (K, N) in the right panel, where we see the mass-shifting effect is accentuated as both N, K increase. This behavior perfectly aligns with the main message of the theorem that the interplay between the truncation and the dependence brings forth the mass-shifting phenomenon.

S2.2 Mass-shifting phenomenon for the unequal-variance case

In this section we discuss a generalization of Theorem 3 to the case where the scale matrix Σ_N contains unequal variances. Recall $\theta \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma_N)$. We continue to consider an "approximately" banded scale matrix Σ_N such that for some integer $2 \le K \le N-1$ there exists a K-banded symmetric and positive definite matrix $\Sigma'_N = (\sigma'_{ij})$ satisfying $\|\Sigma_N - \Sigma'_N\| \lesssim (N \log K)^{-1} \|\Sigma_N\|$. Compared to the equal-variance scenario considered in Theorem 3, the unequal variances $\{\sigma'_{ii}\}$ will be taken into account for the assumption on the correlation structure of Σ'_N accordingly. We now provide assumptions and notations required in this case, which are very similar to those of Theorem 5. We let $\sigma'^2_{(1)}$ and $\sigma'^2_{(N)}$ denote the smallest and largest variances of Σ'_N separately. Without loss of generality, we assume $\sigma'^2_{11} = \sigma'^2_{(1)}$. As one can always scale the matrix Σ_N such that $\sigma'^2_{(1)} = 1$, we assume $\sigma'^2_{(1)} = 1$ and let $\sigma'^2_{(N)} = \kappa$ for some constant $\kappa \ge 1$. Here κ can be interpreted as the ratio of the largest and smallest variances. In addition, we denote by $\sigma'_{\min}, \sigma'_{\max}$ the smallest and largest off-diagonal entries of Σ'_N within the K-band, respectively. Again, denote $\mu^* = \|\boldsymbol{\mu}_N\|_\infty$ and we assume $\sigma'_{\min}, \sigma'_{\max} \in (0, 1)$.

Corollary S1. (Unequal variance). *Fix $\beta \in [0, 1)$. For the mode $\boldsymbol{\mu}_N$ satisfying $\mu^* \le C_{\rho_{\min}, \rho_{\max}} \beta \cdot G_\alpha(\rho_{\min}, \rho_{\max}) (\log K)^{1/2}$, and if $(\sigma'_{\min}, \sigma'_{\max}, \kappa) \in \mathcal{Q}_\kappa$, where $\mathcal{Q}_\kappa$ takes the same form of $\mathcal{Q}_s$ in Theorem 5 by substituting $s = \kappa$. Then there exists some large enough integer K_0 such that $K > K_0$*

[1] A small nugget term σ_0^2 is added such that Σ_N is positive definite. Details on $(\rho_{\min}, \rho_{\max})$ for all combinations (N, K) are deferred to §S7.2 of the supplement. Here we denote $a \wedge b = \min(a, b)$ for any $a, b \in \mathbb{R}$.

and for any fixed $\delta > 0$ *we have*

$$\alpha_{N,\delta} \leq C'_{\sigma'_{\min},\sigma'_{\max},\kappa}\, \delta\, (\log K)^{1/2} K^{-(1-\beta) G_\alpha\left(\sigma'_{\min}/\kappa,\, \sigma'_{\max}\right)},$$

where the function G_α *is same as defined in Theorem 2 for some* $\alpha \in (0,1)$*, and the constants* $C_{\rho_{\min},\rho_{\max}}, C'_{\sigma'_{\min},\sigma'_{\max},\kappa} > 0$ *do not depend on* K, N*.*

S2.3 Additional plots in the main document

Additional graphical illustrations in the manuscript are summarized in this section.

S2.3.1 Additional plot in § 1

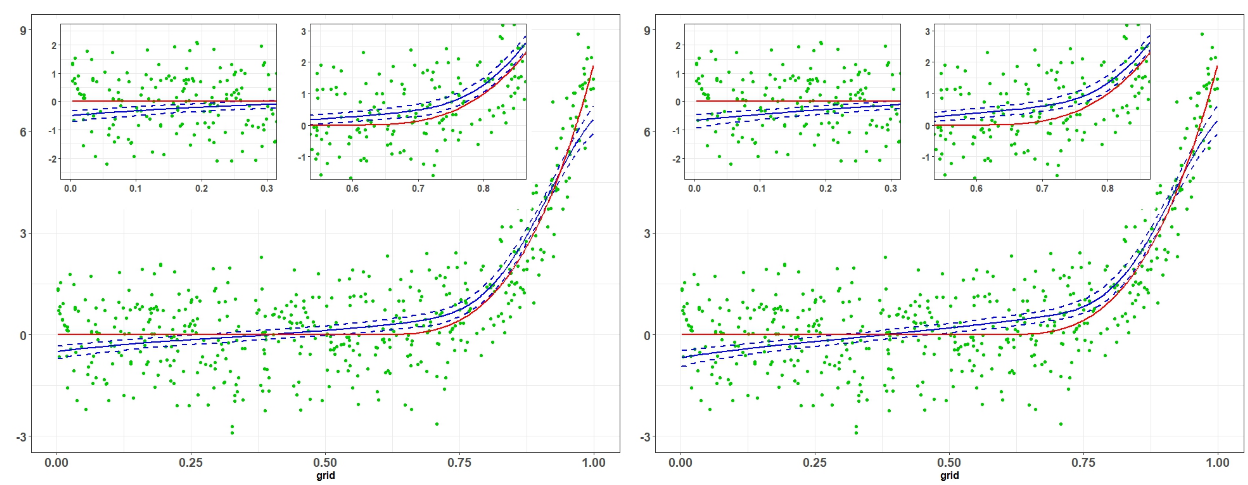

Figure S2: *Monotone function estimation using the basis of [11] and a joint truncated normal prior on the coefficients. Red solid curve corresponds to the true function, blue solid curve is the posterior mean, the region within two dotted blue curves represent a pointwise 95% credible interval, and the green dots are observed data points corresponding to* $N = 50$ *(left panel) and* $N = 250$ *(right panel).*

S2.3.2 Additional plot in § 2

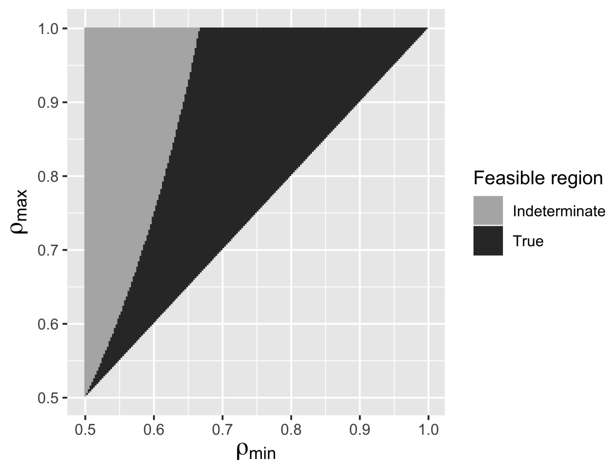

Figure S3: *The region shaded in black depicts $\mathcal{Q}$ from the statement of Theorem 2 in § 2 of the main document.*

S2.3.3 Additional plot in § 3

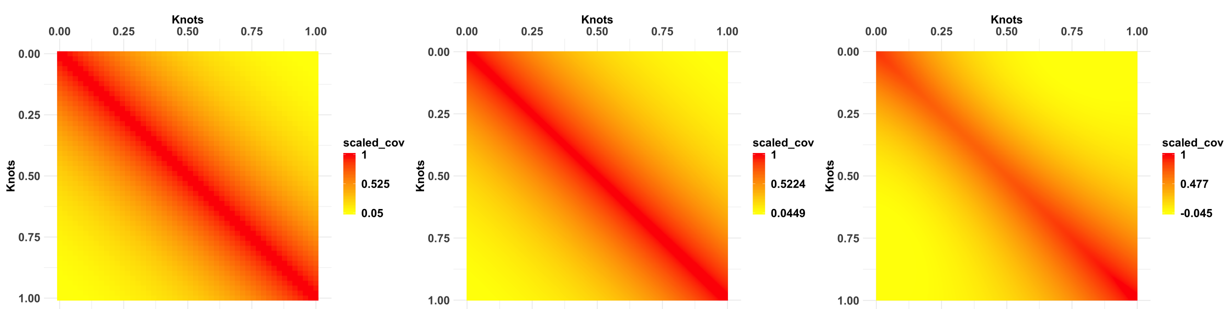

Figure S4: *Scaled posterior scale matrix $\Sigma_N = (\Omega_N^{-1} + \Phi^{\mathrm{T}}\Phi)^{-1}$ defined in § 3 of the manuscript of dimension $N = 50$ (left), $N = 250$ (middle) and $N = 500$ (right).*

S2.3.4 Additional plots in § 4

Model fit of function f_2:

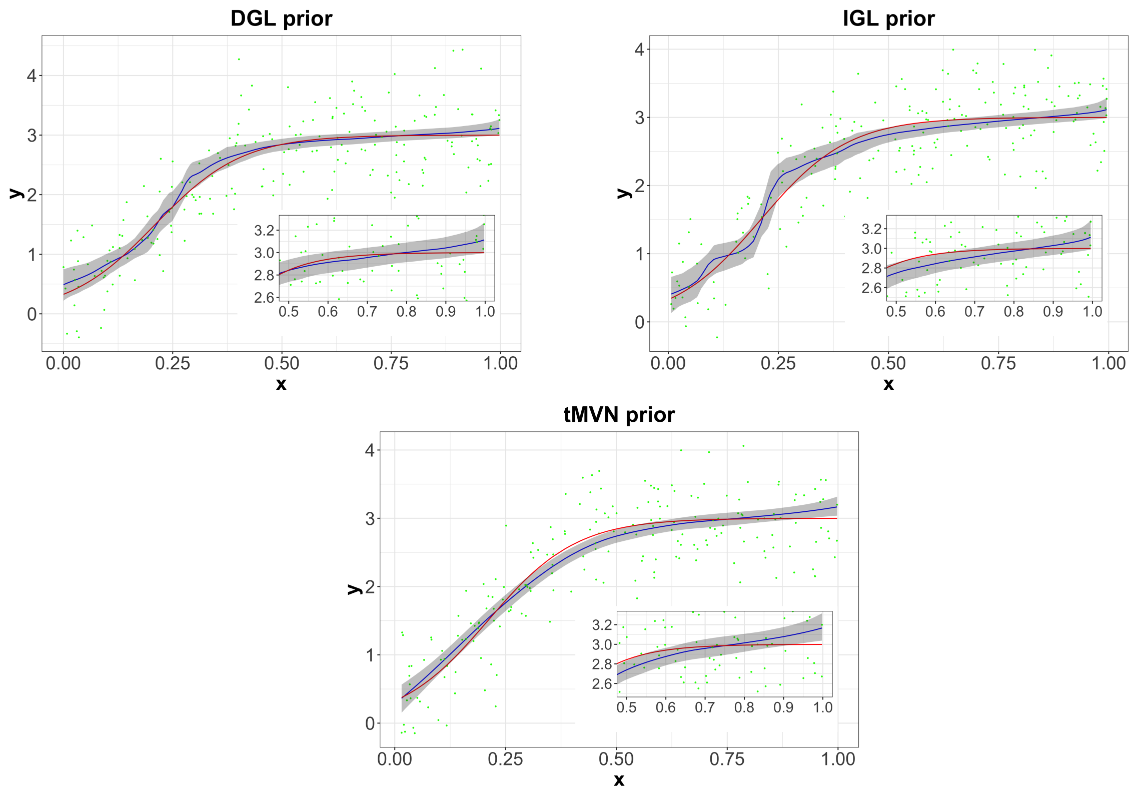

Figure S5: *Same as Figure 2 in the main document for f_2 with zoomed-in inset plots over $x \in [0.5, 1]$.*

Model fit of function f_4:

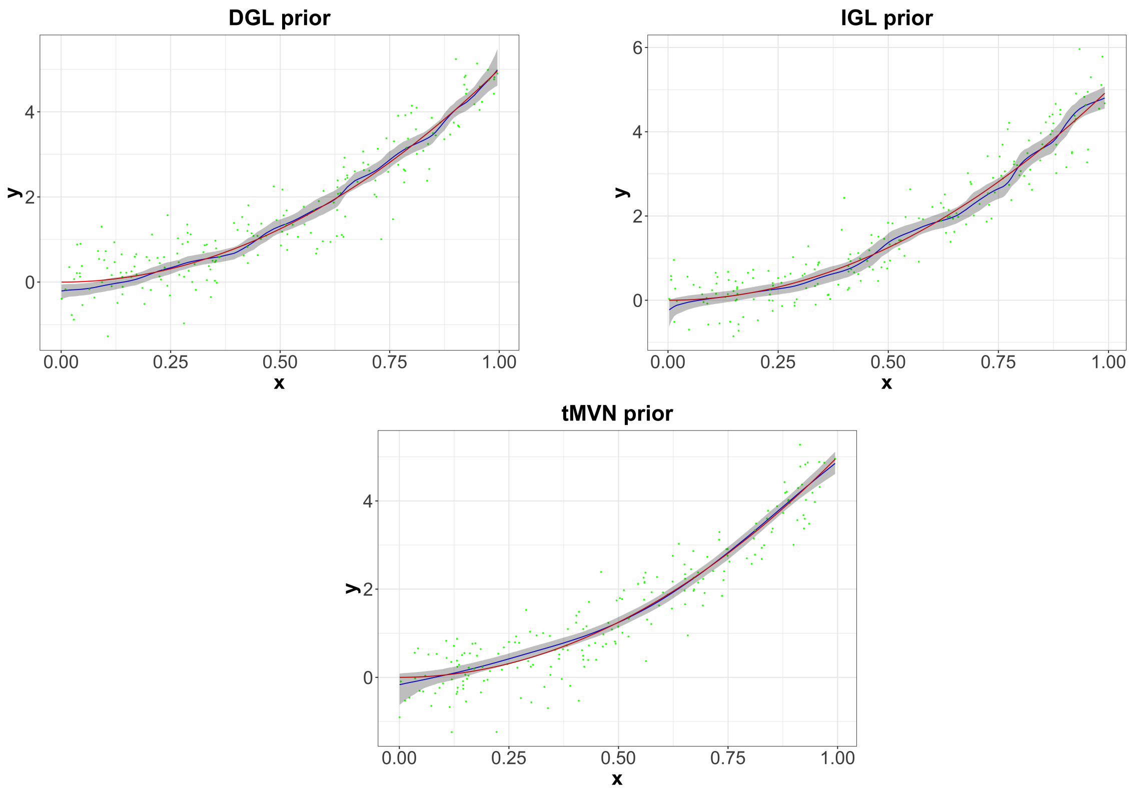

Figure S6: *Same as Figure 2 for* f_4.

S3 Intermediate results

In this section, we provide several intermediate results that play important roles in the proofs of main results in § 2 and § 3. Their proofs are deferred to § S6. We first introduce some notations which will be used throughout the rest of the paper. For any d-dimensional vector $a = [a_1, \ldots, a_d]^{\mathrm{T}}$ we denote its sub-vector by $a_{[i_1\,:\,i_2]} = [a_{i_1}, \ldots, a_{i_2}]^{\mathrm{T}}$ for any $1 \leq i_1 < i_2 \leq d$. For two vectors a and b of the same length, let $a \geq b$ ($a \leq b$) denote the event $a_i \geq b_i$ ($a_i \leq b_i$) for all i. For a fixed integer $d > 0$ and value $\rho \in (0,1)$, we recall that $\Sigma_d(\rho)$ denotes a symmetry-compound correlation matrix with all off-diagonal elements taking on ρ. A key aspect of the compound-symmetry structure that we exploit is for $X \sim \mathcal{N}(\mathbf{0}, \Sigma_d(\rho))$ with $\rho \in (0,1)$, we can represent $X_i \overset{d}{=} \rho^{1/2}\, w + (1-\rho)^{1/2}\, W_i$, where w, W_i's are independent $\mathcal{N}(0,1)$ variables. In addition, we define the matrix $\Sigma_d(\sigma^2, \rho) = (\sigma_{ij})$ with $\sigma_{ii} > 0$ for $1 \leq i \leq d$ and $\sigma_{ij} = \rho$ for $1 \leq i \neq j \leq d$ and for some $0 < \rho < \min_{1 \leq i \leq d} \sigma_{ii}$. Let $\sigma^2_{(1)} = \min_{1 \leq i \leq d}\{\sigma_{ii}\}$ and $\sigma^2_{(d)} = \max_{1 \leq i \leq d}\{\sigma_{ii}\}$. At last, we denote $\bar{\rho}_{(1)} = (\sigma^2_{(1)} - \rho)/\rho$ and $\bar{\rho}_{(d)} = (\sigma^2_{(d)} - \rho)/\rho$. Similarly, for $X \sim \mathcal{N}(\mathbf{0}, \Sigma_d(\sigma^2, \rho))$, one obtains an equivalent expression as $X_i \overset{d}{=} \rho^{1/2}\, w + (\sigma_{ii} - \rho)^{1/2}\, W_i$, for $i = 1, \ldots, N$.

The following Lemma S1 provides a novel comparison result for two-sided Gaussian rectangular probabilities in moderate or high dimensions, which can be considered as an extension of Slepian's inequality summarized in Lemma S4. For a truncated multivariate normal random vector $\theta \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma_N)$ with a mode $\boldsymbol{\mu}_N \geq \mathbf{0}_N$, a scale matrix Σ_N and $\mathcal{C} = [0, \infty)^N$, fix an arbitrary $\delta > 0$. A key ingredient in the mass-shifting theory is to estimate

$$\alpha_{N,\delta} = \mathbb{P}(\theta_1 \leq \delta) = \frac{\mathbb{P}(0 \leq Z_1 \leq \delta, Z_2 \geq 0, \ldots, Z_N \geq 0)}{\mathbb{P}(Z_1 \geq 0, Z_2 \geq 0, \ldots, Z_N \geq 0)},$$

where $Z \sim \mathcal{N}(\boldsymbol{\mu}_N, \Sigma_N)$.

Lemma S2 provides a sandwich bound for the numerator of the ratio in the preceding and Lemma S3 provides a lower bound for the denominator of the ratio in the preceding. All Lemmas will be repeated applied in the proofs of theorems in § 2 and § 3.

Lemma S1. *(Generalized Slepian's inequality.) Let X, Y be d-dimensional Gaussian vectors with finite $\mathbb{E}X_i = \mathbb{E}Y_i$ and $\mathbb{E}X_i^2 = \mathbb{E}Y_i^2$ for all i and $\mathbb{E}(X_iX_j) \leq \mathbb{E}(Y_iY_j)$ for all $i \neq j$. Then for any*

$\ell_1, u_1 \in \mathbb{R}$ such that $\ell_1 < u_1$ and $u_2, \ldots u_d \in \mathbb{R}$, we have

$$\mathbb{P}\big(\ell_1 \le X_1 \le u_1, X_2 \ge u_2, \ldots, X_d \ge u_d\big) \le \mathbb{P}\big(\ell_1 \le Y_1 \le u_1, Y_2 \ge u_2, \ldots, Y_d \ge u_d\big).$$

Lemma S2. *Let $X \sim \mathcal{N}(\boldsymbol{\mu}_d, \Sigma_d(\sigma^2, \rho))$, where the mean vector $\boldsymbol{\mu}_d$ is fixed with $\mu_i \ge 0$ for $i = 1, \ldots, d$, and $\Sigma_d(\sigma^2, \rho)$ is the variance-correlation matrix defined as above and we assume $0 < \rho < \sigma^2_{(1)}$. Fix any $\delta > 0$.*

(Upper bound). *For some $\alpha \in (0, 1)$ and for sufficiently large d, we have the upper bound*

$$\begin{aligned}
&\mathbb{P}(0 \le X_1 < \delta, X_2 \ge 0, \ldots, X_d \ge 0) \\
&\quad \le C\delta \Big\{ \log(d-1) - \rho^{-1/2} \min_{2\le i\le d} \mu_i \Big\}^{-1/2} \\
&\qquad \cdot \exp\Big(- \Big[\{2(1-\alpha)\log(d-1)\}^{1/2} - (\sigma_{11} - \rho)^{-1/2}\big(\delta + \min_{2\le i\le d}\mu_i - \mu_1\big)\Big]^2/2 \\
&\qquad\qquad - \Big[2\,\bar{\rho}_{(1)}\,\{(1-\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2}\min_{2\le i\le d}\mu_i\Big]^2/2\Big) + \exp(-(d-1)^{\alpha}),
\end{aligned}$$

where $C = \{4\pi(\sigma_{11} - \rho)\bar{\rho}_{(1)}\,(1-\alpha)\}^{-1/2}$.

(Lower bound). *In addition, we have the lower bound*

$$\begin{aligned}
&\mathbb{P}(0 \le X_1 < \delta, X_2 \ge 0, \ldots, X_d \ge 0) \\
&\quad \ge C'\delta\,\{\log(d-1)\}^{-1/2}\,\exp\Big\{ - \Big[\{2(1+\alpha)\log(d-1)\}^{1/2} - (\sigma_{11}-\rho)^{-1/2}\big(\min_{2\le i\le d}\mu_i - \mu_1\big)\Big]^2/2 \\
&\qquad\qquad - \Big[2\,\bar{\rho}_{(1)}\,\{(1+\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2}\min_{2\le i\le d}\mu_i\Big]^2/2\Big\},
\end{aligned}$$

where $C' = \{8\pi(\sigma_{11} - \rho)\bar{\rho}_{(1)}\,(1+\alpha)\}^{-1/2}$.

Lemma S3. *Assume the conditions in Lemma S2 are satisfied. Define $\mu^* = \max_{1\le i\le d}\{\mu_j\}$, then*

$$\begin{aligned}
&\mathbb{P}(X_1 \ge 0, X_2 \ge 0, \ldots, X_d \ge 0) \\
&\qquad \ge \frac{\mu^*\rho^{-1/2} + (2\bar{\rho}_{(d)}\log d)^{1/2}}{\big\{\mu^*\rho^{-1/2} + (2\bar{\rho}_{(d)}\log d)^{1/2}\big\}^2 + 1}\,\exp\Big[-\big\{\mu^*\rho^{-1/2} + (2\bar{\rho}_{(d)}\log d)^{1/2}\big\}^2/2\Big].
\end{aligned}$$

S4 Proofs of main results

In this section, we provide proofs of Theorems 2,3 in § 2, Theorem 5 in § 3, Theorems 7 and 8 in § 4, and Corollary S1 in § S2.2. We first introduce some notations which will be used for the rest of the paper. For a constant $a \in \mathbb{R}$, we denote by $[a]$ the largest integer that is no greater than a. For two quantities a, b, we write $a \lesssim (\gtrsim)\, b$ if a/b can be bounded from above (below) by some finite constant, and we write $a \asymp b$ when a/b can be bounded from below and above by two finite constants. For a N-dimensional vector $a = [a_1, \ldots, a_d]^{\mathrm{T}}$, for any subset of indexes $A \subset \{1, \ldots, N\}$ we denote a partition of a based on A by $a = [a_A, a_{A^c}]^{\mathrm{T}}$, where $a_A = \{a_j, j \in A\}$ and $a_{A^c} = \{a_j, j \in A^c\}$ with $A^c = \{1, \ldots, N\} \backslash A$. For a square matrix B, we denote by $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ the smallest and largest eigenvalues of B, respectively. For two square matrices A, B of the same dimension, we say $A \geq B$ if $A - B$ is positive semi-definite. For two random variables X and Y, We write $X \overset{\mathrm{d}}{=} Y$ if X and Y are identical in distribution.

S4.1 Proof of Theorem 2

There are two main steps that consist of the proof of Theorem 2. The first step is to obtain a proper upper bound of the marginal probability $\alpha_{N,\delta}$. The second step is to determine the allowable magnitude of the supremum norm of the mode $\boldsymbol{\mu}_N$ of the truncated multivariate normal distribution in order for the obtained upper bound of $\alpha_{N,\delta}$ to decrease to 0 as $K, N \to \infty$.

Step 1. By definition,

$$\alpha_{N,\delta} = \mathbb{P}(\theta_1 \leq \delta) = \frac{\mathbb{P}(0 \leq Z_1 \leq \delta, Z_2 \geq 0, \ldots, Z_N \geq 0)}{\mathbb{P}(Z_1 \geq 0, Z_2 \geq 0, \ldots, Z_N \geq 0)}, \tag{S4.1}$$

where $Z \sim \mathcal{N}(\boldsymbol{\mu}_N, \Sigma_N)$ where $\Sigma_N \in \mathcal{B}_{N,K}$ and $\boldsymbol{\mu}_N$ is positive component-wise. We now proceed to separately bound the numerator and denominator on the right hand side of equation (S4.1). Denote $\mu_* = \min_{1 \leq j \leq N}(\mu_j)$ and $\mu^* = \max_{1 \leq j \leq N}(\mu_j)$.

We first consider the denominator in equation (S4.1), and use Slepian's lemma to bound it from below. It follows from Slepian's inequality, see comment after Lemma S4 in § S7.1, that if X, Y are d-dimensional Gaussian random variables with $\mathbb{E}(X_i) = \mathbb{E}(Y_i)$ and $\mathbb{E}(X_i^2) = \mathbb{E}(Y_i^2)$ for all i, and

$\mathbb{E}(X_i X_j) \leq \mathbb{E}(Y_i Y_j)$ for all $i \neq j$, then

$$\mathbb{P}(X_1 \geq 0, \ldots, X_d \geq 0) \leq \mathbb{P}(Y_1 \geq 0, \ldots, Y_d \geq 0). \tag{S4.2}$$

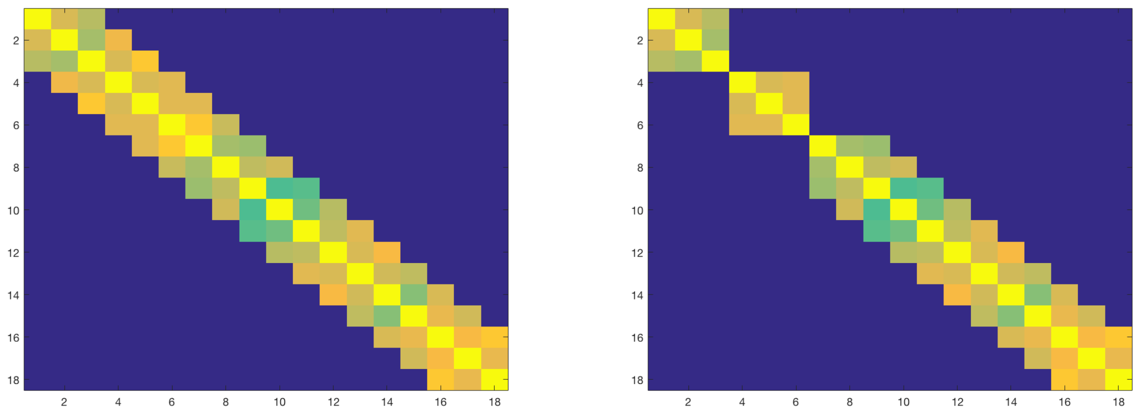

Figure S7: *Left panel: example of* Σ_N *with* $N = 18, K = 3$. *Right panel: the corresponding block approximation* $\widetilde{\Sigma}_N$.

The Slepian's inequality is a prominent example of a Gaussian comparison inequality originally developed to bound the supremum of Gaussian processes. To apply Slepian's inequality to the present context, we construct another N-dimensional centered Gaussian random vector $S \sim \mathcal{N}_N(\boldsymbol{\mu}_N, \widetilde{\Sigma}_N)$ such that (i) $S_{[1:K]} \overset{d}{=} Z_{[1:K]}$, $S_{[(K+1):2K]} \overset{d}{=} Z_{[(K+1):2K]}$ and $S_{[(2K+1):N]} \overset{d}{=} Z_{[(2K+1):N]}$, and (ii) the sub-vectors $S_{[1:K]}$, $S_{[(K+1):2K]}$ and $S_{[(2K+1):N]}$ are mutually independent. The correlation matrix $\widetilde{\Sigma}_N$ of S clearly satisfies $(\Sigma_N)_{ij} \geq (\widetilde{\Sigma}_N)_{ij}$ for all $i \neq j$ by construction. Figure S7 pictorially depicts this *block approximation* in an example with $N = 18$ and $K = 3$. Applying Slepian's inequality, we then have,

$$\begin{aligned}
&\mathbb{P}(Z_1 \geq 0, \ldots, Z_N \geq 0) \geq \mathbb{P}(S_1 \geq 0, \ldots, S_N \geq 0) \\
&\quad = \mathbb{P}(S_{[1:K]} \geq \mathbf{0})\, \mathbb{P}(S_{[(K+1):2K]} \geq \mathbf{0})\, \mathbb{P}(S_{[(2K+1):N]} \geq \mathbf{0}) \\
&\quad = \mathbb{P}(Z_{[1:K]} \geq \mathbf{0})\, \mathbb{P}(Z_{[(K+1):2K]} \geq \mathbf{0})\, \mathbb{P}(Z_{[(2K+1):N]} \geq \mathbf{0}).
\end{aligned} \tag{S4.3}$$

Next, we consider the numerator in equation (S4.1). We have,

$$
\begin{aligned}
&\mathbb{P}(0 \le Z_1 \le \delta, Z_2 \ge 0, \ldots, Z_N \ge 0) \\
&\quad \le \mathbb{P}\big(0 \le Z_1 \le \delta, Z_{[2:K]} \ge \mathbf{0},\, Z_{[(K+1):2K]} \in \mathbb{R}^K,\, Z_{[(2K+1):N]} \ge \mathbf{0}\big) \\
&\quad = \mathbb{P}\big(0 \le Z_1 \le \delta, Z_{[2:K]} \ge \mathbf{0}\big)\mathbb{P}\big(Z_{[(2K+1):N]} \ge \mathbf{0}\big).
\end{aligned} \tag{S4.4}
$$

The last equality crucially uses $Z_{[1:K]}$ and $Z_{[(2K+1):N]}$ are independent, which is a consequence of Σ_N being K-banded. Taking the ratio of equations (S4.3) and (S4.4), the term $\mathbb{P}(Z_{[(2K+1):N]} \ge \mathbf{0})$ cancels so that

$$
\alpha_{N,\delta} \le \frac{\mathbb{P}(0 \le Z_1 \le \delta, Z_{[2:K]} \ge \mathbf{0})}{\mathbb{P}(Z_{[1:K]} \ge \mathbf{0})\,\mathbb{P}(Z_{[K+1:2K]} \ge \mathbf{0})} = R. \tag{S4.5}
$$

To bound the terms $\mathbb{P}(Z_{[1:K]} \ge \mathbf{0})$ and $\mathbb{P}(Z_{[K+1:2K]} \ge \mathbf{0})$ in the denominator of R, we resort to another round of Slepian's inequality. Recall that $\rho_{\min}, \rho_{\max}$ denote the minimum and maximum non-zero correlations in Σ_N. Let $Z'' \sim \mathcal{N}(\boldsymbol{\mu}_K, \Sigma_K(\rho_{\min}))$. Also, recall from equation (2.1) that $\Sigma_K(\rho_{\min})$ denotes the $K \times K$ compound-symmetry correlation matrix with all correlations equal to $\rho_{\min}$. By construction, for any $1 \le i \ne j \le K$, $\mathbb{E}(Z_iZ_j), \mathbb{E}(Z_{K+i}Z_{K+j}) \ge \rho_{\min} = \mathbb{E}(Z''_iZ''_j)$. Thus, applying Slepian's inequality as in equation (S4.2),

$$
\mathbb{P}(Z_{[1:K]} \ge \mathbf{0})\,\mathbb{P}(Z_{[K+1:2K]} \ge \mathbf{0}) \ge \{\mathbb{P}(Z'' \ge \mathbf{0})\}^2.
$$

The numerator of equation (S4.5) cannot be directly tackled by Slepian's inequality, we instead apply Lemma S1. Define a random variable $Z' \sim \mathcal{N}(\boldsymbol{\mu}_K, \Sigma_K(\rho_{\max}))$ and use Lemma S1 to conclude that $\mathbb{P}(0 \le Z_1 \le \delta, Z_{[2:K]} \ge \mathbf{0}) \le \mathbb{P}(0 \le Z'_1 \le \delta, Z'_{[2:K]} \ge \mathbf{0})$.

Substituting these bounds in equation (S4.5), we obtain

$$
R \le R' = \frac{\mathbb{P}(0 \le Z'_1 \le \delta, Z'_2 \ge 0, \ldots, Z'_K \ge 0)}{\{\mathbb{P}(Z''_1 \ge 0, \ldots, Z''_K \ge 0)\}^2}. \tag{S4.6}
$$

The primary reduction achieved by bounding R by R'' is that we only need to estimate Gaussian probabilities under a compound-symmetry covariance structure.

Using the upper bound result of Lemma S2 by letting $\mu = \boldsymbol{\mu}_K$, $\sigma_{ii} = 1$ for $1 \le i \le K$, $\rho = \rho_{\max}$

and $\bar{\rho}_{(1)} = \bar{\rho}_{\max} = (1-\rho_{\max})/\rho_{\max}$, for sufficiently large K such that $\mu^* \leq \sqrt{(1-\rho_{\max})(1-\alpha)\log(K-1)}$ and for any $\alpha \in (0,1)$, we have

$$\begin{aligned}
&\mathbb{P}(0 < Z_1' \leq \delta, Z_2' \geq 0, \ldots, Z_K' \geq 0) \\
&\leq \delta\,[\{2\bar{\rho}_{\max}\,(1-\alpha)\log(K-1)\}^{1/2} - \mu_*/\rho_{\max}^{1/2}]^{-1} \exp\Bigg(- \left[\{2\bar{\rho}_{\max}(1-\alpha)\log(K-1)\}^{1/2} - \mu_*/\rho_{\max}^{1/2}\right]^2/2 \\
&- (1-\alpha)\log(K-1)\Bigg) + \exp\{-(K-1)^{\alpha}\}. \qquad \text{(S4.7)}
\end{aligned}$$

Next, applying Lemma S3 with $a = \mu^*$, $\sigma_{ii} = 1$ for all i, $\rho = \rho_{\min}$ and $\bar{\rho}_{(N)} = \bar{\rho}_{\min} = (1-\rho_{\min})/\rho_{\min}$, one can lower bound the denominator of equation (S4.6),

$$\mathbb{P}(Z''_{[1:K]} \geq \mathbf{0}) \geq \frac{\mu^*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\min}\log K)^{1/2}}{\left\{\mu^*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\min}\log K)^{1/2}\right\}^2 + 1} \exp\left[- \left\{\mu^*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\min}\log K)^{1/2}\right\}^2/2\right]. \qquad \text{(S4.8)}$$

Combining equation (S4.7) and equation (S4.8) leads to the upper bound on R' in equation (S4.6),

$$\begin{aligned}
R' \lesssim\ & \delta(\log K)^{1/2} \exp\Bigg\{ - \left[(2\bar{\rho}_{\max}(1-\alpha)\log(K-1))^{1/2} - \mu_*/\rho_{\max}^{1/2}\right]^2/2 - (1-\alpha)\log(K-1) \\
& \qquad + \left[\mu_*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\min}\log K)^{1/2}\right]^2\Bigg\} \\
& + 4\,\bar{\rho}_{\min}\log K \exp\left\{ - (K-1)^{\alpha} + \left[\mu^*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\min}\log K)^{1/2}\right]^2\right\}. \qquad \text{(S4.9)}
\end{aligned}$$

Since $(\rho_{\min}, \rho_{\max}) \in \mathcal{Q}$, we have $\rho_{\min}/\{2(1-\rho_{\min})\} \geq \rho_{\max}$, or equivalently, $2\bar{\rho}_{\min} < 1/\rho_{\max}$. Thus, we can always find $\alpha > 0$ such that $(1-\alpha)/\rho_{\max} - 2\,\bar{\rho}_{\min} > 0$. Fix such an α, and substitute in equation (S4.9). By choosing K_0 large enough so that for any $K > K_0$, the second term on the right hand side of equation (S4.9) is smaller than the first term; this is possible since the second term decreases exponentially while the first does so polynomially in K.

Step 2. To complete the proof, it remains to determine the condition on $\boldsymbol{\mu}_N$ such that the obtained upper bound of $\alpha_{N,\delta}$ in equation (S4.9) decreases to zero along with K under the assumptions. Since for sufficient large K the first term on the right hand side of equation (S4.9) dominates, it suffices to determine the feasible values of $\boldsymbol{\mu}_N$ such that the exponent of the first term of the upper bound in equation (S4.9) is negative under the assumption $(\rho_{\min}, \rho_{\max}) \in \mathcal{Q}$, which is equivalent to

$\boldsymbol{\mu}_N \in \mathcal{S}$, where

$$\mathcal{S} := \Big\{ \boldsymbol{\mu}_N \in (0,\infty)^N \ : \big\{[2\bar{\rho}_{\max}(1-\alpha)\log(K-1)]^{1/2} - \mu_*/\rho_{\max}^{1/2}\big\}^2/2 \\ + (1-\alpha)\log(K-1) - 2\big[\mu^*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\min}\log K)^{1/2}\big]^2 > 0, \text{ for } (\rho_{\min},\rho_{\max}) \in \mathcal{Q} \Big\}.$$

With some simplifications one can find a condtion subset $\widetilde{\mathcal{S}} \subset \mathcal{S}$ for μ^*, where

$$\widetilde{\mathcal{S}} := \Big\{ \mu^* \in (0,\infty) \ : a\ \log(K-1) - b\ \mu^* \sqrt{\log K} > 0, \\ \mu^* \le \min\big\{\sqrt{(1-\rho_{\max})(1-\alpha)}, \rho_{\min}\big\}\sqrt{\log(K-1)}, \text{ for } (\rho_{\min},\rho_{\max}) \in \mathcal{Q} \Big\}, \tag{S4.10}$$

with

$$a = \frac{1-\alpha}{\rho_{\max}} - 2\,\bar{\rho}_{\min}, \quad \text{and} \quad b = \frac{\sqrt{2(1-\alpha)(1-\rho_{\max})}}{\rho_{\max}} + 2\,\frac{\sqrt{2(1-\rho_{\min})}}{\rho_{\min}}. \tag{S4.11}$$

We now determine feasible values of μ^* in $\widetilde{\mathcal{S}}$. It is straightforward that $b > 0$. As discussed in *Step 1*, given $(\rho_{\min},\rho_{\max}) \in \mathcal{Q}$ there exists some $\alpha \in (0,1)$ such that $a = G_\alpha(\rho_{\min},\rho_{\max}) > 0$. Fix such an α, for any $\beta \in [0,1)$ and any $(\rho_{\min},\rho_{\max}) \in \mathcal{Q}$, as long as choosing $\mu^* \le \beta\min\{1/b, \sqrt{(1-\alpha)(1-\rho_{\max})}/a, \rho_{\min}/a\}a\sqrt{\log(K-1)}$ for sufficiently large K, we have $a\ \log(K-1) - b\,\mu^*\sqrt{\log K} \ge a(1-\beta)\log(K-1)$. Substitute the lower bound in the first term of the right hand side of equation (S4.9), we then obtain the desirable upper bound $\alpha_{N,\delta} \lesssim \delta(\log K)^{1/2}\,(K-1)^{-(1-\beta)G(\rho_{\min},\rho_{\max})}$, for sufficiently large K and some $\alpha \in (0,1)$. Finally, taking $C'_{\rho_{\min},\rho_{\max}} = \min\{1/b, \sqrt{(1-\alpha)(1-\rho_{\max})}/a, \rho_{\min}/a\}$ completes the proof of Theorem 2.

S4.2 Proof of Theorem 3

We now prove Theorem 3 based on Theorem 2. The key observation is that if the scale matrix can be approximated by a banded matrix well enough in operator norm, then the marginal probability $\alpha_{N,\delta}$ of the truncated normal changes only up to a constant when the associated scale matrix is replaced by its banded approximating matrix. We formulate this result as follows. Define random vectors $Z \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma_N)$ and $Z' \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma'_N)$, where $\mathcal{C} = [0,\infty)^N$ and the mode $\boldsymbol{\mu}_N = \{\mu_j\}$

with $\mu_j \geq 0$ for $j = 1, \ldots, N$. Also both Σ_N, Σ'_N are non-negative and Σ'_N is K-banded for some integer $2 \leq K \leq N-1$.

Proposition S1. *For $Z \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma_N)$ and $Z' \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma'_N)$ where $\mathcal{C} = [0, \infty)^N$, if $\boldsymbol{\mu}_N$ is a non-negative fixed vector, and for sufficiently large N, K if there exists some $\varepsilon \lesssim (N \log K)^{-1}$ such that $\|\Sigma_N - \Sigma'_N\| \lesssim \varepsilon \|\Sigma_N\|$, then for any fixed $\delta > 0$, we have*

$$\mathbb{P}(0 \leq Z_1 \leq \delta) \asymp \mathbb{P}(0 \leq Z'_1 \leq \delta).$$

The proof of Proposition S1 is deferred to § S5.4. An application of Proposition S1 and Theorem 2 immediately yields the result of Theorem 3.

S4.3 Proof of Theorem 5

Proof of Part (a). Part (a) of Theorem 5 is an immediate application of Proposition S1 and Corollary S1 by taking $\mu^* = 0, \beta = 0$. To see that, recall $\theta_c \sim \mathcal{N}(\mathbf{0}, \Sigma_N)$ and define $\tilde{\theta}_c \sim \mathcal{N}(\mathbf{0}, \widetilde{\Sigma}_N)$. Given $\|\Sigma_N - \widetilde{\Sigma}_N\| \lesssim (N \log K)^{-1} \|\Sigma_N\|$ and by applying Proposition S1, then for any fixed $\delta > 0$ one obtains $\Pi(0 \leq \theta_{c1} \leq \delta \,|\, Y) \asymp \Pi(0 \leq \tilde{\theta}_{c1} \leq \delta \,|\, Y)$, by applying Proposition S1. Following similar notations and assumptions in Corollary S1, denote $\tilde{\sigma}^2_{(1)} = \min_{1 \leq i \leq d} \tilde{\sigma}_{ii}$, and define $\tilde{\theta}'_c = \tilde{\theta}_c / \tilde{\sigma}_{(1)}$ and one have $\tilde{\theta}'_c \sim \mathcal{N}(\mathbf{0}, \widetilde{\Sigma}'_N)$ where $\widetilde{\Sigma}'_N = \widetilde{\Sigma}_N / \tilde{\sigma}^2_{(1)}$. Define $\tilde{\delta} = \delta / \tilde{\sigma}_{(1)}$, then we have $\Pi(0 \leq \tilde{\theta}_{c1} \leq \delta) = \Pi(0 \leq \tilde{\theta}'_{c1} \leq \tilde{\delta})$. Then applying Corollary S1, we obtain the desired upper bound of $\Pi(0 \leq \theta_{c1} \leq \delta)$, which yields the result in Part (a).

Proof of Part (b). We now provide a detailed proof of Part (b) of Theorem 5. We still follow the main line of arguments in the proof of Theorem 2, that we first obtain a proper upper bound of the marginal posterior probability, then analyze the relation between the posterior mode and the obtained upper bound. However, the posterior mode is a random vector under the true data-generating process, thus bounding the posterior mode leads to a high probability result, which is different from *Step 2* of the proof of Theorem 2.

Let $X \sim \mathcal{N}(\boldsymbol{\mu}_N, \Sigma_N)$ with $\boldsymbol{\mu}_N = \Sigma_N \Phi^{\mathrm{T}} Y$ and $\Sigma_N = (\Phi^{\mathrm{T}} \Phi + \Omega_N^{-1})^{-1}$, and let $\widetilde{X} \sim \mathcal{N}(\boldsymbol{\mu}_N, \widetilde{\Sigma}_N)$ where $\widetilde{\Sigma}_N$ is K-banded and satisfies $\|\Sigma_N - \widetilde{\Sigma}_N\| \lesssim (N \log K)^{-1} \|\Sigma_N\|$ by Proposition 4. Recall $\tilde{\sigma}^2_{(1)}, \tilde{\sigma}^2_{(N)}$ denote the smallest and largest diagonal elements of $\widetilde{\Sigma}_N$ separately. Define $\widetilde{X}' \sim$

$\mathcal{N}(\tilde{\boldsymbol{\mu}}_N, \widetilde{\Sigma}'_N)$, where $\tilde{\boldsymbol{\mu}}_N = \boldsymbol{\mu}_N/\tilde{\sigma}_{(1)} = (\tilde{\rho}_{ij})$ and $\widetilde{\Sigma}'_N = \widetilde{\Sigma}_N/\tilde{\sigma}^2_{(1)}$. Therefore $\widetilde{\Sigma}'_N = (\tilde{\rho}_{ij})$ is also K-banded. Under the assumptions, we have $\tilde{\rho}_{11} = 1$ denoting smallest diagonal element of $\widetilde{\Sigma}'_N$ and $\tilde{\kappa} = \tilde{\sigma}^2_{(N)}/\tilde{\sigma}^2_{(1)}$ is the largest diagonal element of $\widetilde{\Sigma}'_N$. Also recall that $\tilde{\rho}_{\min}$ and $\tilde{\rho}_{\max}$ denote the smallest and largest positive off-diagonal entries of $\widetilde{\Sigma}'_N$, respectively. For any fixed $\delta > 0$, define $\tilde{\delta} = \delta/\tilde{\sigma}_{(1)}$.

Now we are ready to bound the marginal posterior probability $\Pi(0 < \theta_1 < \delta \mid Y)$ for any fixed $\delta > 0$. Define $\widetilde{X}^c = \widetilde{X}' - \tilde{\boldsymbol{\mu}}_N \sim \mathcal{N}(\mathbf{0}_N, \widetilde{\Sigma}'_N)$ with $\tilde{\boldsymbol{\mu}}_N = \widetilde{\Sigma}'_N \Phi^{\mathrm{T}} Y$. By definition,

$$
\begin{aligned}
\Pi(0 < \theta_1 < \delta \mid Y) &\lesssim \frac{\mathbb{P}(0 \le \widetilde{X}_1 \le \delta, \widetilde{X}_2 \ge 0, \dots, \widetilde{X}_N \ge 0 \mid Y)}{\mathbb{P}(\widetilde{X}_1 \ge 0, \widetilde{X}_2 \ge 0, \dots, \widetilde{X}_N \ge 0 \mid Y)} \\
&\le \frac{\mathbb{P}(0 \le \widetilde{X}'_1 \le \tilde{\delta}, \widetilde{X}'_2 \ge 0, \dots, \widetilde{X}'_K \ge 0 \mid Y)}{\mathbb{P}\big(\widetilde{X}'_{[1:K]} \ge 0\big)\mathbb{P}\big(\widetilde{X}'_{[(K+1):2K]} \ge 0 \mid Y\big)} \\
&= \frac{\mathbb{P}(0 \le \widetilde{X}^c_1 + \tilde{\mu}_1 \le \tilde{\delta}, \widetilde{X}^c_2 + \tilde{\mu}_2 \ge 0, \dots, \widetilde{X}^c_K + \tilde{\mu}_K \ge 0 \mid Y)}{\mathbb{P}\big(\widetilde{X}^c_{[1:K]} + \tilde{\boldsymbol{\mu}}_{[1:K]} \ge 0\big)\mathbb{P}\big(\widetilde{X}^c_{[(K+1):2K]} + \tilde{\boldsymbol{\mu}}_{[(K+1):2K]} \ge 0 \mid Y\big)}.
\end{aligned} \tag{S4.12}
$$

The first inequality is an immediate application of Proposition S1. To arrive at the second inequality we first apply the change of variables $\widetilde{X}' = \widetilde{X}/\tilde{\sigma}_{(1)}$ and then apply the same technique used in obtaining the upper bound of R_1 in equation (S4.53) in the proof of Corollary S1, with an application of Lemma S1 and Lemma S4.

Bounding the posterior mode. Note that $\boldsymbol{\mu}_N$ is a Gaussian random vector under the true data-generating distribution denoted by $\mathbb{P}_0$. Under the true model, one obtains the marginal distribution of the mode $\boldsymbol{\mu}_N \sim \mathcal{N}(\mathbf{0}, \Sigma_N \Phi^{\mathrm{T}} \Phi \Sigma_N)$ with $\Sigma_N = (\Phi^{\mathrm{T}}\Phi + \Omega^{-1})^{-1}$. Assuming $N = o(n)$ and under Assumptions 1 and 2, one obtains $C_2^{-1}(N/n)\,\mathrm{I_N} \le \Sigma_N \le C_1^{-1}(N/n)\,\mathrm{I_N}$, where C_1, C_2 are constants defined in Assumption 1. Simple calculations lead to facts $\Sigma_N \Phi^{\mathrm{T}}\Phi\Sigma_N \le \Sigma_N$ and $\|\Sigma_N \Phi^{\mathrm{T}}\Phi\Sigma_N - \Sigma_N\| \le \|\Sigma_N \Omega^{-1}\Sigma_N\| \le (C_2\lambda_0)^{-1}(N/n)^2$, where λ_0 is same as in Assumption 2. Combining these results yields a sharp sandwiched bound $\Sigma_N - (C_2\lambda_0)^{-1}(N/n)^2\mathrm{I_N} \le \Sigma_N\Phi^{\mathrm{T}}\Phi\Sigma_N \le \Sigma_N$. Then, for sufficiently large N, n we have $\Sigma_N\Phi^{\mathrm{T}}\Phi\Sigma_N = \Sigma_N + o(N/n)\mathrm{I_N}$. Then we have for sufficiently large N, n, $\tilde{\boldsymbol{\mu}}_N = \boldsymbol{\mu}_N/\tilde{\sigma}_{(1)} \sim \mathcal{N}(\mathbf{0}_N, \Sigma_N/\tilde{\sigma}^2_{(1)})$.

Since the ratio in the last line of equation (S4.12) depends on the sub-vector $\tilde{\boldsymbol{\mu}}_{[1:2K]}$, as follows we define a high probability set of $\tilde{\boldsymbol{\mu}}_{[1:2K]}$, using the concentration property of Lipschitz functions of dependent Gaussian random variables. We remark that the result can be easily applied to sub-vectors of $\boldsymbol{\mu}_N$ over different dimensions. For any integer $2 \le K < [N/2]$, it is well known that

$\max_{1\leq i\leq 2K}\{\tilde{\mu}_i\}$ is a Lipschitz function of $\{\tilde{\mu}_i, i=1,\dots,2K\}$ with the Lipschitz constant $\sigma^2_{\mu,\max} = \max_{1\leq i\leq 2K}\{\mathrm{var}(\tilde{\mu}_i)\} \leq \sigma^2_{(N)}/\tilde{\sigma}^2_{(1)}$. Now applying the concentration inequality of a Lipschitz function of Gaussian random variables, for any $\epsilon_K > 0$, we have

$$\mathbb{P}\Big\{\Big|\max_{1\leq i\leq 2K}\tilde{\mu}_i - \mathbb{E}\Big(\max_{1\leq i\leq 2K}\tilde{\mu}_i\Big)\Big| > \epsilon_K/2\Big\} \leq 2\exp\left\{-\epsilon_K^2/(8\,\sigma^2_{\mu,\max})\right\}. \tag{S4.13}$$

It is easy to show that $\mathbb{E}(\max_{1\leq i\leq 2K}\{\tilde{\mu}_i\}) = O(\sqrt{\log(2K)})$ by using Sudakov's minoration and Slepian's lemma (see, e.g., Theorem 3.14 in [12]) for sufficiently large K. To simplify the computation, we proceed with a slightly different representation by introducing a positive constant $M = M(K)$ that may depend on K such that $\mathbb{E}(\max_{1\leq i\leq 2K}\{\tilde{\mu}_i\}) = \sqrt{2M\log K}$ for sufficiently large K. Also note that $-\min_{1\leq i\leq 2K}\{\tilde{\mu}_i\} = \max_{1\leq i\leq 2K}\{-\tilde{\mu}_i\} \overset{d}{=} \max_{1\leq i\leq 2K}\{\tilde{\mu}_i\}$, by the symmetry of $\tilde{\mu}_N$ about the origin. Fix such an M, take $\epsilon_K = (2\gamma\log K)^{1/2}$ for some small constant $0<\gamma<M$ to be chosen later, and choose another sufficiently small constant $\beta\in(0,1)$, we define the following event

$$\mathcal{A} = \left\{\tilde{\boldsymbol{\mu}}\in\mathbb{R}^{2K} : \sqrt{2(M-\gamma)\log K} \leq -\min_{1\leq i\leq 2K}\tilde{\mu}_i \leq \sqrt{2(M+\gamma)\log K},\ |\tilde{\mu}_1| \leq (2\log K)^{(1-\beta)/2}\right\}.$$

Note that based on the event $\mathcal{A}$, there exists $\gamma'\in(\gamma, M)$ such that for the set defined as

$$\mathcal{A}' = \left\{\tilde{\boldsymbol{\mu}}\in\mathbb{R}^{K} : \sqrt{2(M-\gamma')\log K} \leq -\min_{2\leq i\leq K}\tilde{\mu}_i + \tilde{\mu}_1 \leq \sqrt{2(M+\gamma')\log K}\right\}. \tag{S4.14}$$

We have $\mathcal{A}\subset\mathcal{A}'$ and

$$\begin{aligned}\mathbb{P}(\mathcal{A}^c) &\leq 8K^{-\gamma/(2\sigma^2_{\mu,\max})} + 2\mathbb{P}(|\tilde{\mu}_1| \geq (2\log K)^{(1-\beta)/2}) \\ &\leq 3(2\log K)^{(1-\beta)/2}\exp\{-c_0(\log K)^{1-\beta}\},\end{aligned} \tag{S4.15}$$

where $c_0 = \tilde{\sigma}^2_{(1)}/\sigma^2_{11}$. The second inequality holds by applying (S4.13) and applying Lemma S7 for sufficiently large K.

Then combining results in equations (S4.12), (S4.14) and (S4.15), we have with at least $\mathbb{P}_0$-

probability of $1-3(2\log K)^{(1-\beta)/2}\exp\{-c_0(\log K)^{1-\beta}\}$,

$$\Pi(0<\theta_1<\delta \mid Y) \leq \frac{\sup_{\tilde{\mu}_{[1:2K]}\in\mathcal{A}'}\mathbb{P}(0\leq \widetilde{X}_1^c+\tilde{\mu}_1\leq\tilde{\delta}, \widetilde{X}_2^c+\tilde{\mu}_2\geq 0,\ldots,\widetilde{X}_K^c+\tilde{\mu}_K\geq 0)}{\inf_{\tilde{\mu}_{[1:2K]}\in\mathcal{A}'}\left\{\mathbb{P}\big(\widetilde{X}^c_{[1:K]}+\tilde{\mu}_{[1:K]}\geq \mathbf{0}_{[1:K]}\big)\right\}\inf_{\tilde{\boldsymbol{\mu}}_{[1:2K]}\in\mathcal{A}'}\left\{\mathbb{P}\big(\widetilde{X}^c_{[(1+K):2K]}+\tilde{\boldsymbol{\mu}}_{[(1+K):2K]}\geq \mathbf{0}_{[(1+K):2K]}\big)\right\}}. \tag{S4.16}$$

Bounding $\Pi(0<\theta_1<\delta \mid Y)$. The rest of the proof follows a similar line of arguments in proving Corollary S1 in § S2.2. To bound the denominator of the obtained upper bound in equation (S4.16), we first bound the term $\inf_{\tilde{\mu}_{[1:2K]}\in\mathcal{A}'}\mathbb{P}(\widetilde{X}^c_{[1:K]}+\tilde{\boldsymbol{\mu}}_{[1:K]}\geq\mathbf{0}_K)$. Recall $\widetilde{X}^c\sim\mathcal{N}(\mathbf{0}_K,\widetilde{\Sigma}'_N)$ with $\widetilde{\Sigma}'_N)=(\tilde{\rho}_{ij})$. Define $\widetilde{Y}\sim\mathcal{N}(\mathbf{0}_K,\Sigma'_K)$, where the covariance matrix $\Sigma'_K=(\rho'_{ij})$ such that $\rho'_{ii}=\tilde{\rho}_{ii}$ for $i=1,\ldots,K$ and $\rho'_{ij}=\tilde{\rho}_{\min}$ for $1\leq i\neq j\leq K$. Then we have $\mathbb{E}(\widetilde{Y}_i\widetilde{Y}_j)\leq\mathbb{E}(\widetilde{X}_i^c\widetilde{X}_j^c)$ for $1\leq i\neq j\leq K$ and $\mathbb{E}(\widetilde{Y}_i^2)=\mathbb{E}[(\widetilde{X}_i^c)^2]$ for $1\leq i\leq K$. Then by Lemma S4 we obtain the lower bound

$$\mathbb{P}(\widetilde{X}^c_{[1:K]}+\tilde{\boldsymbol{\mu}}_{[1:K]}\geq\mathbf{0}_K)\geq\mathbb{P}(\widetilde{Y}_{[1:K]}+\tilde{\boldsymbol{\mu}}_{[1:K]}\geq\mathbf{0}_K).$$

Then applying Lemma S3 by taking $d=K$, $\rho=\tilde{\rho}_{\min}$, $\bar{\rho}_{(K)}=(\tilde{\kappa}-\tilde{\rho}_{\min})/\tilde{\rho}_{\min}$ and $a=\max_{\tilde{\boldsymbol{\mu}}_{[1:2K]}\in\mathcal{A}'}\{\tilde{\mu}_i\}$, for sufficiently large K, we have

$$\begin{aligned}\inf_{\tilde{\boldsymbol{\mu}}_{[1:2K]}\in\mathcal{A}'}\mathbb{P}\left(\widetilde{Y}_{[1:K]}+\tilde{\boldsymbol{\mu}}_{[1:K]}\geq\mathbf{0}_K\right)&\geq\inf_{\tilde{\boldsymbol{\mu}}_{[1:2K]}\in\mathcal{A}'}\mathbb{P}\left(\widetilde{Y}_{[1:K]}\geq\max_{\tilde{\mu}_{[1:2K]}\in\mathcal{A}'}\{\tilde{\mu}_i\}\mathbf{1}_K\right)\\&\geq\frac{\big(\bar{\rho}_{(K)}^{-1/2}+\tilde{\rho}_{\min}^{-1/2}(M+\gamma')^{1/2}\big)\sqrt{\log K}}{\big(\bar{\rho}_{(K)}^{-1/2}+\tilde{\rho}_{\min}^{-1/2}(M+\gamma')^{1/2}\big)^2\log K+1}K^{-\left(\bar{\rho}_{(K)}^{1/2}+\tilde{\rho}_{\min}^{-1/2}(M+\gamma')^{1/2}\right)^2}.\end{aligned} \tag{S4.17}$$

The second inequality is attained by taking $\max_{\tilde{\boldsymbol{\mu}}_{[1:2K]}\in\mathcal{A}'}\{\tilde{\mu}_i\}=\sqrt{2(M+\gamma')\log K}$ under the set $\mathcal{A}'$. We obtain a same lower bound for the term $\inf_{\tilde{\boldsymbol{\mu}}_{[1:2K]}\in\mathcal{A}'}\mathbb{P}(\widetilde{X}^c_{[(K+1):2K]}+\tilde{\boldsymbol{\mu}}_{[(K+1):2K]}\geq\mathbf{0}_K)$ as the leading constants M,γ' are universal for $\{\tilde{\mu}_i:1\leq i\leq 2K\}$.

Now we upper bound the numerator of the ratio in equation (S4.16). Again, define $Z\sim\mathcal{N}(\mathbf{0}_K,\Sigma''_K)$ where $\Sigma''_K=(\rho''_{ij})$ satisfies $\rho''_{ii}=\tilde{\rho}_{ii}$ for $1\leq i\leq K$ and $\rho''_{ij}=\tilde{\rho}_{\max}$ for $1\leq i\neq j\leq K$. Then $\mathbb{E}(Z_iZ_j)\geq\mathbb{E}(\widetilde{X}_i^c\widetilde{X}_j^c)$ for $1\leq i\neq j\leq K$ and $\mathbb{E}(Z_i^2)=\mathbb{E}[(\widetilde{X}_i^c)^2]$ for $1\leq i\leq K$. Applying

Lemma S1 leads to, for any $\tilde{\mu} \in \mathcal{A}'$,

$$\mathbb{P}(0 \leq \widetilde{X}_1^c + \tilde{\mu}_1 \leq \tilde{\delta}, \widetilde{X}_2^c + \tilde{\mu}_2 \geq 0, \ldots, \widetilde{X}_K^c + \tilde{\mu}_K \geq 0) \\ \leq \mathbb{P}(0 \leq Z_1 + \tilde{\mu}_1 \leq \tilde{\delta}, Z_2 + \tilde{\mu}_2 \geq 0, \ldots, Z_K + \tilde{\mu}_K \geq 0). \tag{S4.18}$$

Next step is to obtain an upper bound for the right hand side of the preceding. Let $d = K$, $\sigma_{(1)}^2 = \rho_{11}'' = 1$, $\rho = \tilde{\rho}_{\max}$ and $\bar{\rho}_{(1)} = (1 - \tilde{\rho}_{\max})/\tilde{\rho}_{\max}$, for sufficient large N, K and for some $\alpha \in (0, 1)$, applying Lemma S2 implies

$$\sup_{\tilde{\boldsymbol{\mu}}_{[1:2K]} \in \mathcal{A}'} \mathbb{P}(0 \leq Z_1 + \tilde{\mu}_1 \leq \tilde{\delta}, Z_2 + \tilde{\mu}_2 \geq 0, \ldots, Z_K + \tilde{\mu}_K \geq 0) \\ \leq \tilde{\delta}\left[\{\bar{\rho}_{\max}(1-\alpha) + \tilde{\rho}_{\max}^{-1}(M - \gamma')\} \log(K-1)\right]^{-1/2} \cdot K^{-\left\{(1-\alpha)+\sqrt{\frac{M-\gamma'}{1-\tilde{\rho}_{\max}}}\right\}^2 - \left\{(1-\alpha)\bar{\rho}_{\max}+\sqrt{\frac{M-\gamma'}{\tilde{\rho}_{\max}}}\right\}^2} \\ + e^{-(K-1)^{\alpha}}. \tag{S4.19}$$

The supremum of the probability attains when $-\min_{2\leq i\leq K} \tilde{\mu}_i + \tilde{\mu}_1 = \sqrt{2(M-\gamma')\log K}$ and therefore the inequality holds. The above bound in equation (S4.19) is simplified as $\tilde{\delta}$ is fixed and is dominated by $\sqrt{2(M-\gamma')\log K}$ for sufficiently large K.

Now applying results in equations (S4.17),(S4.18) and (S4.19) to equation (S4.16), we obtain the upper bound of the marginal posterior probability $\Pi(0 \leq \theta_1 \leq \tilde{\delta} \mid Y)$. With $\mathbb{P}_0$-probability at least $1 - 3(2\log K)^{(1-\beta)/2} \exp\{-c_0(\log K)^{1-\beta}\}$, we have

$$\Pi(0 \leq \theta_1 \leq \tilde{\delta} \mid Y) \leq C_{\tilde{\kappa},\tilde{\rho}_{\min},\tilde{\rho}_{\max}}\, \tilde{\delta} \log(K)^{1/2} K^{-\left\{(1-\alpha)+\sqrt{\frac{M-\gamma'}{1-\tilde{\rho}_{\max}}}\right\}^2 - \left\{(1-\alpha)\bar{\rho}_{\max}+\sqrt{\frac{M-\gamma'}{\tilde{\rho}_{\max}}}\right\}^2 + 2\left(\bar{\rho}_{\min,\kappa}^{1/2}+\sqrt{\frac{M+\gamma'}{\tilde{\rho}_{\min}}}\right)^2} \\ + C'_{\bar{\rho}_{\min,\kappa},\tilde{\rho}_{\min}} \log K \cdot e^{-(K-1)^{\alpha}} K^{2\left(\bar{\rho}_{\min,\kappa}^{1/2}+\sqrt{\frac{M+\gamma'}{\tilde{\rho}_{\min}}}\right)^2}, \tag{S4.20}$$

where $C_{\tilde{\kappa},\tilde{\rho}_{\min},\tilde{\rho}_{\max}} = \{\bar{\rho}_{\max}(1-\alpha) + \tilde{\rho}_{\max}^{-1}(M-\gamma')\}^{1/2}/C'_{\bar{\rho}_{\min,\kappa},\tilde{\rho}_{\min}}$ and $C'_{\bar{\rho}_{\min,\kappa},\tilde{\rho}_{\min}} = \{\bar{\rho}_{\min,\kappa}^{-1/2} + \tilde{\rho}_{\min}^{-1/2}(M+\gamma')^{1/2}\}^2$.

To complete the proof it remains to show that under the condition $(\tilde{\rho}_{\min}, \tilde{\rho}_{\max}, \tilde{\kappa}) \in \mathcal{Q}_{\tilde{\kappa}}$ and under set $\mathcal{A}'$, there exist constants $\gamma, \gamma' > 0$ such that the desired bound is attained given $\beta, M > 0$ for sufficiently large $K > 0$. Again, for sufficiently large K, the term $e^{-(K-1)\alpha}$ decreases exponentially fast so the second term of the right hand side of equation (S4.20) is negligible compared to the first terms. Similar to the proof of Theorem 2, we first find sufficient conditions of (γ, γ') such that the first term on the right hand side of equation (S4.20) decreases with K under the assumptions. To that end, for sufficiently large K, and for the defined M and some $\alpha \in (0,1)$, we define the set of γ, γ' as

$$\mathcal{T} = \Bigg\{(\gamma,\gamma') \in (0,\infty) \otimes (\gamma, M) : \Bigg\{\sqrt{1-\alpha} + \sqrt{\frac{M-\gamma'}{1-\tilde{\rho}_{\max}}}\Bigg\}^2 + \Bigg\{\sqrt{\frac{(1-\tilde{\rho}_{\max})(1-\alpha)}{\tilde{\rho}_{\max}}} + \sqrt{\frac{M-\gamma'}{\tilde{\rho}_{\max}}}\Bigg\}^2 - 2\Bigg\{\sqrt{\frac{\tilde{\kappa}-\tilde{\rho}_{\min}}{\tilde{\rho}_{\min}}} + \sqrt{\frac{M+\gamma'}{\tilde{\rho}_{\min}}}\Bigg\}^2 > 0, \text{ for } (\tilde{\rho}_{\min}, \tilde{\rho}_{\max}, \tilde{\kappa}) \in \mathcal{Q}_{\tilde{\kappa}}\Bigg\}.$$

It is easy to see that $(\gamma, \gamma') \in \mathcal{T}$ ensures the first term of the upper bound in equation (S4.20) decreases with K. With some calculations one can obtain a subset $\mathcal{T}' \subset \mathcal{T}$, where

$$\mathcal{T}' = \Bigg\{(\gamma,\gamma') \in (0,\infty) \otimes (\gamma, M) : aM + b\sqrt{M} + c - (a'\gamma' + b'\sqrt{\gamma'}) > 0, \text{ for } (\tilde{\rho}_{\min}, \tilde{\rho}_{\max}, \tilde{\kappa}) \in \mathcal{Q}_{\tilde{\kappa}}\Bigg\},$$

with

$$\begin{aligned}
a &= \frac{1-\alpha}{\tilde{\rho}_{\max}} + \frac{1-\alpha}{1-\tilde{\rho}_{\max}} - \frac{2}{\tilde{\kappa}_{\min}}, \qquad a' = \frac{1-\alpha}{\tilde{\rho}_{\max}} + \frac{1-\alpha}{1-\tilde{\rho}_{\max}} + \frac{2}{\tilde{\rho}_{\min}}, \\
b &= 2\left(\sqrt{\frac{1-\alpha}{1-\tilde{\rho}_{\max}}} + \frac{\sqrt{(1-\alpha)(1-\tilde{\rho}_{\max})}}{\tilde{\rho}_{\max}} - 2\frac{\sqrt{\tilde{\kappa}-\tilde{\rho}_{\min}}}{\tilde{\rho}_{\min}}\right), \\
b' &= 2\left(\sqrt{\frac{1-\alpha}{1-\tilde{\rho}_{\max}}} + \frac{\sqrt{(1-\alpha)(1-\tilde{\rho}_{\max})}}{\tilde{\rho}_{\max}} + 2\frac{\sqrt{\tilde{\kappa}-\tilde{\rho}_{\min}}}{\tilde{\rho}_{\min}}\right), \\
c &= \frac{1-\alpha}{\tilde{\rho}_{\max}} - \frac{2(\tilde{\kappa}-\tilde{\rho}_{\min})}{\tilde{\rho}_{\min}}.
\end{aligned} \tag{S4.21}$$

Then under $\mathcal{T}'$, it can be shown that the first term of the upper bound in equation (S4.20) is bounded above by a multiple of $(\log K)^{1/2} K^{-\{aM+b\sqrt{M}+c-(a'\gamma'+b'\sqrt{\gamma'})\}}$. Thus the rest of the proof is to show that the set $\mathcal{T}'$ is not empty. It suffices to show there exist γ, γ' such that $\gamma' \in (\gamma, M)$ and $aM + b\sqrt{M} > a'\gamma' + b'\sqrt{\gamma'}$ for large enough K and for the defined M. First, note that the assumption $(\tilde{\rho}_{\min}, \tilde{\rho}_{\max}, \tilde{\kappa}) \in \mathcal{Q}_{\tilde{\kappa}}$ leads to $c > 0$ for some $\alpha \in (0,1)$ and this implies $\tilde{\rho}_{\max} > 1/2$. Fix such an α, then it implies $a > 0$ based on $\tilde{\rho}_{\max} > 1/2$. In addition, applying the inequality $a + b \geq 2\sqrt{ab}$ for $a, b \geq 0$ to the first two terms of b implies $b/2 \geq 2(\sqrt{(1-\alpha)/\sqrt{\tilde{\rho}_{\max}}} - \sqrt{(\tilde{\kappa}-\tilde{\rho}_{\min})/\tilde{\rho}_{\min}})$. Further we can show that $\sqrt{(1-\alpha)/\sqrt{\tilde{\rho}_{\max}}} > \sqrt{(\tilde{\kappa}-\tilde{\rho}_{\min})/\tilde{\rho}_{\min}}$ given $c > 0$ and $\tilde{\rho}_{\max} > 1/2$. Therefore, under the assumptions in Theorem 5 we have $a, a', b, b', c > 0$. Then by choosing $\gamma' > 0$ such that $\max\{\gamma', \sqrt{\gamma'}\} \leq \min\{M, (aM + b\sqrt{M})/\{2(a'+b')\}\}$ and choosing any $\gamma < \gamma'$ yields the result

$$\Pi(0 \leq \theta_1 \leq \delta \mid Y) \lesssim \tilde{\delta}(\log K)^{1/2} K^{-c}$$

with at least $\mathbb{P}_0$-probability $1 - 3(2\log K)^{(1-\beta)/2}\exp\{-c_0(\log K)^{1-\beta}\}$ for $K > K_0$, where c is defined in (S4.21), K_0 is a sufficiently large integer and $c_0 = \tilde{\sigma}^2_{(1)}/\sigma^2_{11}$. We then complete the proof of Part (b).

S4.4 Proof of Theorem 7

Write the joint prior $\xi \mid \tau, \Lambda \sim \mathcal{N}_{\mathcal{C}}(\mathbf{0}_N, \tau^2\Lambda\Sigma\Lambda)$, where $\Lambda = \mathrm{Diag}(\{\lambda_1, \ldots, \lambda_N\})$ with $\lambda_j \overset{i.i.d.}{\sim} C_+(0,1)$ and $\tau > 0$ is some fixed constant which may depend on n to be chosen later. The posterior distribution is expressed as $\xi|\tau, \Lambda, Y \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Omega^{-1})$ with $\boldsymbol{\mu}_N = \Omega^{-1}\Phi^{\mathrm{T}}Y$ and $\Omega^{-1} = (\Phi^{\mathrm{T}}\Phi + \Lambda^{-1}\Sigma^{-1}\Lambda^{-1}/\tau^2)$. For simplicity, we adopt the same notation introduced in Section 3, and abuse the notation $\alpha_{N,\delta}$ to denote the marginal posterior probability over $(0,\delta)$ for any fixed $\delta > 0$, i.e., let $\alpha_{N,\delta} := \Pi(0 < \xi_1 < \delta|\tau, Y)$. It suffices to show that for any fixed $\delta > 0$ and for some sequence $\{\tau_n\}$, there exists a lower bound on $\mathbb{E}_0(\alpha_{N,\delta})$ which goes to 1 as $n \to \infty, \tau_n \to 0$ almost surely, where $\mathbb{E}_0(\cdot)$ denotes taking expectation with respect to the true data generating function.

We first obtain a lower bound of $\alpha_{N,\delta}$. Denote $\tilde{\xi} = \xi - \boldsymbol{\mu}_N \sim \mathcal{N}(\mathbf{0}_N, \Omega^{-1})$. It is easy to see that

$$\alpha_{N,\delta} \geq \frac{\mathbb{P}(\|\boldsymbol{\mu}_N\|_\infty < \tilde{\xi}_1 < \delta + \|\boldsymbol{\mu}_N\|_\infty,\ \tilde{\xi}_2 > \|\boldsymbol{\mu}_N\|_\infty,\ \cdots,\ \tilde{\xi}_N > \|\boldsymbol{\mu}_N\|_\infty \mid \tau_n, Y)}{\mathbb{P}(\tilde{\xi}_1 > -\|\boldsymbol{\mu}_N\|_\infty,\ \tilde{\xi}_2 > -\|\boldsymbol{\mu}_N\|_\infty,\ \cdots,\ \tilde{\xi}_N > -\|\boldsymbol{\mu}_N\|_\infty \mid \tau_n, Y)} =: P. \quad \text{(S4.22)}$$

Then $\mathbb{E}_0(\alpha_{N,\delta}) \geq \mathbb{E}_0(P)$. To proceed, we shall first state two high probability results for the posterior scale matrix Ω^{-1} and posterior mode $\boldsymbol{\mu}_N$ separately.

Bound Ω^{-1}. The idea is to show that the posterior scale matrix is dominated by the prior scale matrix with high probability in the presence of a sufficiently small global shrinkage prior. We first bound the operator norm of the matrix $\Lambda\Sigma\Lambda$ with high probability. For some small constant $\beta \in (0,1)$, let $\tau_n = O(n^{-1/(1-\beta)})$, then denote $\mathcal{A}_\Omega = \{\|\Lambda\Sigma\Lambda\| \leq N/(n\tau_n^{2-2\beta})\}$, we shall show $\mathbb{P}(\mathcal{A}_\Omega) \geq 1 - \sqrt{nN\tau_n^{2-2\beta}}$. Under Assumption 1, we have $C_1(n/N)\mathrm{I_N} \leq \Phi^{\mathrm{T}}\Phi \leq C_2(n/N)\mathrm{I_N}$ for constants $C_1, C_2 > 0$. And under Assumption 2, there exists $\lambda_0 > 0$ such that

$$\lambda_0 \tau_n^{-2} \min_{1 \le j \le N} (\lambda_j^{-2}) \le \lambda_{\min}(\Lambda^{-1}\Sigma^{-1}\Lambda^{-1}/\tau_n^2) \le \lambda_{\max}(\Lambda^{-1}\Sigma^{-1}\Lambda^{-1}/\tau_n^2) \le (1/\lambda_0)\tau^{-2} \max_{1 \le j \le N} (\lambda_j^{-2}).$$

Given the above result, one can show

$$\begin{aligned}
\mathbb{P}\Big(\lambda_0 \tau_n^{-2} \min_{1 \le j \le N} (\lambda_j^{-2}) \ge \tau_n^{-2\beta} \|\Phi^{\mathrm{T}}\Phi\|\Big) &\ge \mathbb{P}\Big(\lambda_0 \tau_n^{-2} \min_{1 \le j \le N} (\lambda_j^{-2}) \ge C_2 n/(N\tau_n^{2\beta})\Big) \\
&= \Pi_{j=1}^n \Big\{\mathbb{P}\Big(\lambda_j \le \sqrt{C_2' N/(n\tau_n^{2-2\beta})}\Big)\Big\} \\
&\approx \Big(1 - \sqrt{n\tau_n^{2-2\beta}/(C_2' N)}\Big)^N \\
&\ge 1 - \sqrt{{C'}_2^{-1} n N \tau_n^{2-2\beta}},
\end{aligned} \tag{S4.23}$$

for some constant $C_2' > 0$. The third line of the preceding uses the fact that for $\lambda_j \sim C_+(0,1)$, for sufficiently large $a > 0$ one has $\Pi_\lambda(\lambda_j > a) \approx a^{-1}$ for $j = 1, \ldots, N$. The last line uses the inequality $(1-x)^n \ge 1 - nx$ for any $x \in [0, n^{-1})$ and the fact that $\sqrt{nN}\tau_n^{1-\beta} = O(N/n)$ by choosing $\tau_n = O(n^{-1/(1-\beta)})$. Under the set $\mathcal{A}_\Omega$, one can easily see that $\Omega^{-1} \approx \tau_n^2 \Lambda\Sigma\Lambda$ for sufficiently large n, N. This result indicates that the prior matrix employed with the global-local shrinkage parameters dominates the posterior scale matrix with high probability, and essentially shrink the posterior scale matrix to a zero matrix.

Bound $\|\boldsymbol{\mu}_N\|_\infty$. We now state the concentration property of $\|\boldsymbol{\mu}_N\|_\infty$ under the true data generating function. First note that $\boldsymbol{\mu}_N \sim \mathcal{N}(\mathbf{0}_N, \tau_n^4 \Lambda\Sigma\Lambda\Phi^{\mathrm{T}}\Phi\Lambda\Sigma\Lambda)$. Under Assumption 1 and under the set $\mathcal{A}_\Omega$, it is easy to show that $\|\tau_n^4 \Lambda\Sigma\Lambda\Phi^{\mathrm{T}}\Phi\Lambda\Sigma\Lambda\| \lesssim \tau_n^{2+2\beta}\|\Lambda\Sigma\Lambda\|$. It is well know that $\|\boldsymbol{\mu}_N\|_\infty = \max_{1 \le i \le N}(|\mu_i|)$ is a Lipschitz function with the Lipschitz constant denoted by $\sigma_{\max}^2 = \tau_n^{2+2\beta}\|\Lambda\Sigma\Lambda\|$, under the set $\mathcal{A}_\Omega$. Further, one can show that $\mathbb{E}_0\|\boldsymbol{\mu}_N\|_\infty \le M_0 \tau_n^{1+\beta}\sqrt{\|\Lambda\Sigma\Lambda\| \log N}$ for some constant $M_0 > 0$. Then apply the concentration inequality of the Lipschitz function of Gaussian variables, choosing $t_N = 2M_0\tau_n^{1+\beta}\sqrt{\|\Lambda\Sigma\Lambda\| \log N}$ one can obtain $\mathbb{P}_0(\|\boldsymbol{\mu}_N\|_\infty > t_N) \le 2N^{-2}$. For convenience, we denote $\mathcal{A}_\mu = \{\|\boldsymbol{\mu}_N\|_\infty \le t_N\}$ with $\mathbb{P}(\mathcal{A}_\mu^c) \le 2N^{-2}$.

Now we are ready to bound $\mathbb{E}_0(P)$ from below. Denote $\mathcal{A} = \mathcal{A}_\Omega \cap \mathcal{A}_\mu$. Then one has $\mathbb{P}_0(\mathcal{A}) \ge \mathbb{P}(\mathcal{A}_\Omega) - \mathbb{P}_0(\mathcal{A}_\mu^c) \gtrsim 1 - \sqrt{nN\tau_n^{2-2\beta}} - 2N^{-2}$, based on the above result and equation (S4.23). Then under set $\mathcal{A}$, one can obtain

$$P \geq \frac{\mathbb{P}(t_n < \tilde{\xi}_1 < \delta + t_n,\ \tilde{\xi}_2 > t_n,\ \cdots,\ \tilde{\xi}_N > t_n)}{\mathbb{P}(\tilde{\xi}_1 > -t_n,\ \tilde{\xi}_2 > -t_n,\ \cdots,\ \tilde{\xi}_N > -t_n)} := P'.$$

Then

$$\mathbb{E}_0(P) \geq \mathbb{E}_0(P'\mathbb{1}_{\mathcal{A}}) \geq \mathbb{P}_0(\mathcal{A}) \inf_{\mathcal{A}} P'.$$

It suffices to bound $\inf_{\mathcal{A}} P'$. By change of variables, let $\xi' = \tilde{\xi}/(\tau_n \|\Lambda\Sigma\Lambda\|^{1/2})$ and $\xi' \sim \mathcal{N}(\mathbf{0}_N, \Omega'^{-1})$ with $\|\Omega'^{-1}\| \leq 1$. Denote $t'_n = 2M_0\tau_n^{\beta}\sqrt{\log N}$, then it is straightforward to see that

$$\inf_{\mathcal{A}} P' = \frac{\mathbb{P}(t'_n < \xi'_1 < \delta/(\tau_n \|\Lambda\Sigma\Lambda\|^{1/2}) + t'_n,\ \xi'_2 > t'_n,\ \cdots,\ \xi'_N > t'_n)}{\mathbb{P}(\xi'_1 > -t_n,\ \xi'_2 > -t'_n,\ \cdots,\ \xi'_N > -t'_n)}.$$

It is easy to see that by choosing $\tau_n \asymp O(n^{-1/(1-\beta)})$, we have $t'_n \to 0$ and $\delta/(\tau_n\|\Lambda\Sigma\Lambda\|^{1/2}) \geq \delta\tau_n^{-\beta}\sqrt{n/N} \to \infty$ for any fixed $\delta > 0$, as $n, N \to \infty$. Then,

$$\lim_{n,N\to\infty} \inf_{\mathcal{A}} P' = \frac{\mathbb{P}(0 \leq \xi'_1 \leq \infty,\ \xi'_2 \geq 0,\ \cdots,\ \xi'_N \geq 0)}{\mathbb{P}(\xi'_1 \geq 0,\ \xi'_2 \geq 0,\ \cdots,\ \xi'_N \geq 0)} = 1.$$

Combining the above result with the result that $\lim_{n,N\to\infty} \mathbb{P}_0(A) = 1$ completes the proof of Theorem 7. Note that in the theorem, we let $\alpha = (1-\beta)^{-1} - 1$.

S4.5 Proof of Theorem 8

In this section, we prove equations (4.6) and (4.7). We shall first provide a detailed proof of equation (4.6). As the proof of equation (4.7) is similar to that of (4.6), we omit some details and only highlight the difference. We begin by introducing some new notations used in the proof. For two densities p, q that are absolutely continuous with respect to the Lebesgue measure μ, the Kullback–Leibler divergence between p and q is defined as $KL(p,q) = \int p \log(p/q) d\mu$ and the V-divergence is defined as $V(p,q) = \int p \log^2(p/q) d\mu$. Denote by $f_{C+}(x) = 2/(\pi(1+x^2))\mathbb{1}(x \geq 0)$ the density of the default half-Cauchy distribution $C_+(0,1)$. Denote $P_0 = \mathcal{N}(f_0, \sigma_0^2 \mathrm{I_n})$ and $P_N = \mathcal{N}(\Psi\theta, \sigma^2\mathrm{I_n})$. We denote by $\Pi(\cdot)$ the joint prior distribution for $(\theta, \lambda, \sigma^2)$, and we use $\Pi_{\theta,\lambda}(\cdot)$, $\Pi_{\theta|\lambda}(\cdot), \Pi_{\lambda}(\cdot), \Pi_{\sigma}(\cdot)$ to denote priors for $(\theta, \lambda), \theta|\lambda, \lambda, \sigma^2$ separately. Recall that the prior for θ conditioning on λ is restricted to the set $\mathcal{C}$, for computational convenience, we also define the unconstrained parameter denoted by θ' where $\theta'|\lambda \sim \mathcal{N}(\mathbf{0}_N, \tau^2\Lambda\Omega_N\Lambda)$ and we denote by $\Pi_{\theta'|\lambda}(\cdot)$ the corresponding unconstrained prior. Then we have $\Pi_{\theta|\lambda}(\cdot) = \Pi_{\theta'|\lambda}(\cdot)\mathbb{1}_{\mathcal{C}}(\cdot)/\Pi_{\theta'|\lambda}(\mathcal{C})$ given any $\lambda > \mathbf{0}_N$. For any $1 \leq j \leq N$, when $\lambda_j = 0$ the prior $\Pi_{\theta_j|\lambda_j}(\cdot) = \delta_0(\cdot)$ is degenerated to a Dirac measure at 0. In addition, we define the marginal prior of θ by integrating $\lambda > \mathbf{0}_N$ out,

$$\Pi_{\mathcal{C}}(\theta) = \frac{\int_{\lambda > \mathbf{0}} \Pi_{\theta'|\lambda}(\theta)\mathbb{1}_{\mathcal{C}}(\theta)\Pi(\lambda)d\lambda}{\int_{\lambda > \mathbf{0}} \int_{\mathcal{C}} \Pi_{\theta'|\lambda}(\theta)\Pi(\lambda)d\theta d\lambda}, \tag{S4.24}$$

the denominator is the normalizing constant denoted by $\mathcal{M}_{\mathcal{C}}$.

Bounding $\mathcal{M}_{\mathcal{C}}$. It is easy to check that the scale matrix $\Omega_N = (\omega_{ij})$ induced from a Gaussian process with a Mátern kernel in § 4.1 satisfies Assumption 2 with some $\lambda_0, \alpha_0 > 0$, which indicates that the scale matrix Ω_N is approximately banded. Thus to bound $\mathcal{M}_{\mathcal{C}}$, we will first construct a strictly banded symmetric and positive definite matrix Ω'_N that approximates Ω_N well. Then we apply similar techniques to Ω'_N used in deriving the equation (S4.3) in the proof of Theorem 2 to bound $\mathcal{M}_{\mathcal{C}}$.

Based on Proposition 4, one can show that there exists a K-banded symmetric positive definite matrix $\Omega'_N = (\omega'_{ij})$ such that $\|\Omega_N - \Omega'_N\| \lesssim (N \log K)^{-1}$, as long as the band width $K \gtrsim (\log N)^t$ for some $t > 0$. To simplify the analysis, we assume that Ω'_N has equal variances and it is non-negative. We remark that the analysis can be extended to unequal-variance case, with an applications of techniques used in proving Corollary S1. Denote $\omega'_{ii} = \sigma'^2$ for $1 \leq i \leq N$ and assume $\omega'_{\min} = \min_{1 \leq |i-j| < K}\{\omega'_{ij}\} \in (0, \sigma'^2)$. Then the following Proposition provides a lower bound of $\mathcal{M}_{\mathcal{C}}$.

Proposition S2. *Assume Ω_N satisfies Assumption 2, and for some integer $K = O(N)$ there exists a K-banded symmetric and positive definite matrix Ω'_N such that $\|\Omega_N - \Omega'_N\| \lesssim (N \log N)^{-1}$. Also assume Ω'_N has equal variances and is non-negative. Then there exists some constant $t_0 > 2$ such that*

$$\mathcal{M}_{\mathcal{C}} \gtrsim (\log N)^{-t_0/2} N^{-t_0 \bar{\omega}'_{\min}},$$

where $\bar{\omega}'_{\min} = (\sigma'^2 - \omega'_{\min})/\omega'_{\min}$.

The proof of Proposition S2 is deferred to § S5.5. We remark that as we assume $\omega'_{\min}$ does not change along with N, Proposition S2 posits the lower bound of $\mathcal{M}_{\mathcal{C}}$ decreases at a polynomial rate as N goes to infinity. And choosing $K = O(N)$ serves a technical purpose of controlling the lower bound of $\mathcal{M}_{\mathcal{C}}$. For convenience, we denote the obtained lower bound by $\mathcal{M}'_{\mathcal{C}} = (\log N)^{-t_0/2} N^{-t_0 \bar{\omega}'_{\min}}$ in Proposition S2.

We are now ready to prove equations (4.6) and (4.7).

Proof of equation (4.6). Following the seminal work of [9], it suffices to show that under the conditions in Theorem 8, for the defined sequence of $\{\epsilon_n\}$, there exist a sequence of sieves $\{\mathcal{F}_n\}$ over the parameter space of (θ, σ^2) and a sequence of test functions $\{\phi_n\}$ satisfying the following conditions:

$$\Pi\{(\theta,\sigma^2) : KL(P_0, P_N) \le n\epsilon_n^2; V(P_0, P_N) \le n\epsilon_n^2\} \gtrsim e^{-c_1 n\epsilon_n^2}, \tag{S4.25}$$

$$\Pi(\mathcal{F}_n^c) \lesssim e^{-c_2 n\epsilon_n^2}, \tag{S4.26}$$

$$\mathbb{E}_0(\phi_n) \lesssim e^{-c_3 n\epsilon_n^2}, \qquad \sup_{\substack{(\theta,\sigma^2)\in\mathcal{F}_n : \|\theta-\theta_0\|\ge M_1\sigma_0\sqrt{N}\epsilon_n, \\ \text{or } |\sigma^2-\sigma_0^2|\ge\sigma_0^2\epsilon_n}} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n) \lesssim e^{-c_3 n\epsilon_n^2}, \tag{S4.27}$$

for some constants $c_1, c_2, c_3, M_1 > 0$.

Part I. We first verify condition (S4.25) by following a similar line of arguments in [1]. We have

$$\begin{aligned} KL(P_0, P_N) &= \frac{n}{2}\left[\frac{\sigma_0^2}{\sigma^2} - 1 - \log\frac{\sigma_0^2}{\sigma^2}\right] + \frac{\|f_0 - \Psi\theta\|^2}{2\sigma^2}, \\ V(P_0, P_N) &= \frac{n}{2}\left[\left(\frac{\sigma_0^2}{\sigma^2}\right)^2 + 1 - 2\frac{\sigma_0^2}{\sigma^2}\right] + \frac{\sigma_0^2}{\sigma^2}\frac{\|f_0 - \Psi\theta\|^2}{\sigma^2}. \end{aligned}$$

Similar to [1], define

$$\mathcal{B}_1 = \left\{\sigma^2 : \frac{\sigma_0^2}{\sigma^2} - 1 - \log\frac{\sigma_0^2}{\sigma^2} \le \epsilon_n^2\right\},$$

and

$$\mathcal{B}_2 = \left\{(\theta,\sigma^2) : \frac{\|f_0 - \Psi\theta\|^2}{\sigma^2} \le n\epsilon_n^2; \quad \frac{\sigma_0^2}{\sigma^2}\frac{\|f_0 - \Psi\theta\|^2}{\sigma^2} \le n\epsilon_n^2\right\}.$$

It is easy to see that $\{KL(P_0, P_N) \le n\epsilon_n^2; V(P_0, P_N) \le n\epsilon_n^2\} \supset \mathcal{B}_1 \cap \mathcal{B}_2$. Further we define $\widetilde{\mathcal{B}}_1 = \{\sigma^2 : \sigma_0^2/(1+\epsilon_n) \le \sigma^2 < \sigma_0^2\}$ and it is also easy to see that $\widetilde{\mathcal{B}}_1 \subset \mathcal{B}_1$. Since $\sigma^2 \sim \mathrm{IG}(a_0, b_0)$, we have

$$\Pi_\sigma(\mathcal{B}_1) \ge \Pi_\sigma(\widetilde{\mathcal{B}}_1) \asymp \int_{\sigma_0^2/(1+\epsilon_n)}^{\sigma_0^2} (\sigma^2)^{-a_0-1} e^{-b_0/\sigma^2} \ge (\sigma_0^2)^{-a_0-1} e^{-b_0(1+\epsilon_n)/\sigma_0^2} \gtrsim e^{-C_1 n\epsilon_n^2},$$

for some constant $C_1 > 0$ and for sufficiently large n. Conditioning the set $\mathcal{B}_2$ on $\mathcal{B}_1$, we have

$$\frac{\|\Psi\theta - f_0\|^2}{\sigma^2}(1+\epsilon_n) \leq \frac{2}{\sigma_0^2}\|\Psi\theta - f_0\|^2.$$

Thus $\Pi(\mathcal{B}_2 \mid \mathcal{B}_1) \geq \Pi_{\mathcal{C}}(\|\Psi\theta - f_0\|^2 \leq n\epsilon_n^2)$. Applying the triangular inequality and Lemma 1, we have

$$\begin{aligned}\|\Psi\theta - f_0\|^2 &\leq 2\|\Psi(\theta-\theta_0)\|^2 + 2\|\Psi\theta_0 - f_0\|^2 \\ &\lesssim \|\Psi(\theta-\theta_0)\|^2 + n\|\Psi\theta_0 - f_0\|_\infty^2 \\ &\leq \|\Psi(\theta-\theta_0)\|^2 + n\epsilon_n^2.\end{aligned}$$

Then it suffices to bound $\Pi_{\mathcal{C}}(\|\Psi(\theta-\theta_0)\|^2 \leq n\epsilon_n^2)$. Next, we can show

$$\|\Psi(\theta-\theta_0)\|^2 \leq \lambda_{\max}(\Psi^{\mathrm{T}}\Psi)\|\theta-\theta_0\|^2 \lesssim n\|\theta-\theta_0\|_1^2.$$

The second inequality in the preceding uses Assumption (A2) that $\lambda_{\max}(\Psi^{\mathrm{T}}\Psi) \lesssim n/N$ and Cauchy-Schwarz inequality $\|\theta-\theta_0\| \leq \sqrt{N}\|\theta-\theta_0\|_1$. Then $\Pi_{\mathcal{C}}(\|\Psi(\theta-\theta_0)\|^2 \leq n\epsilon_n^2/2) \geq \Pi_{\mathcal{C}}(\|\theta-\theta_0\|_1 \leq \epsilon_n/2)$. To proceed, we consider two cases separately: (i) $0 < s_0 < N$ and (ii) $s_0 = N$.

Case (i). Recall $\theta = [\theta_{S_0}, \theta_{S_0^c}]$ and $\theta_0 = [\theta_{0S_0}, \theta_{0S_0^c}]$. To simplify the notation we let $\boldsymbol{\theta}_1 = \theta_{S_0}, \boldsymbol{\theta}_2 = \theta_{S_0^c}$ and let $\boldsymbol{\theta}_{01} = \theta_{0S_0}, \boldsymbol{\theta}_{02} = \theta_{0S_0^c}$. Similarly, for the unconstrained $\theta' = [\theta'_{S_0}, \theta'_{S_0^c}]$ we denote $\boldsymbol{\theta}'_1 = \theta'_{S_0}, \boldsymbol{\theta}'_2 = \theta'_{S_0^c}$. Accordingly, we have $\lambda = [\lambda_{S_0}, \lambda_{S_0^c}]$ with $\lambda_{S_0} = \{\lambda_j, j \in S_0\}$ and $\lambda_{S_0^c} = \{\lambda_j, j \in S_0^c\}$ and we partition the diagonal matrix Λ with $\Lambda_1 = \mathrm{diag}(\{\lambda_j, j \in S_0\})$ and $\Lambda_2 = \mathrm{diag}(\{\lambda_j, j \in S_0^c\})$. Now we partition the prior scale matrix $\Omega_N = [\Omega_{11}\,\Omega_{12}; \Omega_{21}\,\Omega_{22}]$, where $\Omega_{11} = (\omega_{ij})$ for $i,j \in S_0$, $\Omega_{12} = (\omega_{ij})$ for $i \in S_0, j \in S_0^c$ with $\Omega_{21} = \Omega_{12}^{\mathrm{T}}$ and $\Omega_{22} = (\omega_{ij})$ for $i,j \in S_0^c$. Denote the unconstrained prior $\theta' \sim \mathcal{N}(\mathbf{0}_N, \tau^2\Lambda\Omega\Lambda)$ conditioning on λ, we write it in the form of

$$[\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2] \mid \lambda \sim \mathcal{N}\left(\mathbf{0}_N, \tau^2\begin{bmatrix}\Lambda_1\Omega_{11}\Lambda_1 & \Lambda_1\Omega_{12}\Lambda_2 \\ \Lambda_2\Omega_{21}\Lambda_1 & \Lambda_2\Omega_{22}\Lambda_2\end{bmatrix}\right). \tag{S4.28}$$

Then for some sufficiently small constant $a_n > 0$ that may depend on (N, n) and will be chosen later, for some $C_2 > 0$, we have

$$
\begin{aligned}
&\Pi_{\mathcal{C}}(\|\theta - \theta_0\|_1 \le C_2\,\epsilon_n) \\
&\quad \ge \Pi_{\theta',\lambda}(\|\boldsymbol{\theta}_1' - \boldsymbol{\theta}_{01}\|_1 \le (C_2 - 1)\,\epsilon_n, \|\boldsymbol{\theta}_2'\|_\infty \le \epsilon_n/(N - s_0), \theta' \ge \mathbf{0}_N)/\Pi_{\theta',\lambda}(\boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}) \\
&\quad \ge \Pi_{\theta',\lambda}\Big(\|\boldsymbol{\theta}_1' - \boldsymbol{\theta}_{01}\|_1 \le (C_2 - 1)\,\epsilon_n, \boldsymbol{\theta}_1' \ge \mathbf{0}_{S_0}, \lambda_{S_0} \ge \mathbf{0} \mid \|\boldsymbol{\theta}_2'\|_1 \le \epsilon_n, \boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}, \min_{j \in S_0^c}\{\lambda_j\} \ge a_n\Big) \\
&\quad \cdot \Big\{\Pi_{\theta',\lambda}\Big(\|\boldsymbol{\theta}_2'\|_\infty \le \epsilon_n/(N - s_0), \boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}, \min_{j \in S_0^c}\{\lambda_j\} \ge a_n\Big)/\Pi(\boldsymbol{\theta}_2' \ge \mathbf{0}_{N-s_0})\Big\} \\
&\quad =: P_1 \cdot P_2
\end{aligned}
\tag{S4.29}
$$

The first inequality holds based on the fact $\Pi_{\theta',\lambda}(\theta' \in \mathcal{C}) \le \Pi_{\theta',\lambda}(\boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c})$. Thus, to lower bound $\Pi_{\mathcal{C}}(\|\theta - \theta_0\|_1 \le C_2\,\epsilon_n)$ it suffices to lower bound P_1 and P_2 in the preceding, separately.

We now bound P_1 in equation (S4.29). First, define sets $\mathcal{A} = \{(\boldsymbol{\theta}_2', \lambda_{S_0^c}) : \|\boldsymbol{\theta}_2'\|_\infty \le \epsilon_n/(N - s_0), \boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}, \min_{j \in S_0^c}\{\lambda_j\} \ge a_n\}$ and $\tilde{\mathcal{A}} = \{\lambda_{S_0} : \max_{j \in S_0}\{\lambda_j\} \le C_2 a_n/(2K_0 s_0)\}$ for some constant $K_0 > 0$ to be determined later. Based on equation (S4.28), we have $\mathbb{E}(\boldsymbol{\theta}_1'|\boldsymbol{\theta}_2') = \Lambda_1\Omega_{12}\Omega_{22}^{-1}\Lambda_2^{-1}\boldsymbol{\theta}_2'$. Under Assumption 2, there exists some constant K_0 such that $\|\Omega_{12}\Omega_{22}^{-1}\|_\infty \le K_0$. Then under the set $\mathcal{A} \cap \tilde{\mathcal{A}}$, one can obtain

$$
\|\mathbb{E}(\boldsymbol{\theta}_1'|\boldsymbol{\theta}_2')\|_1 \le s_0 K_0 \max_{j \in S_0}\{\lambda_j\}/\min_{j \in S_0^c}\{\lambda_j\}(N - s_0)\|\boldsymbol{\theta}_2'\|_\infty \le K_0\, s_0\, a_n^{-1} \max_{j \in S_0}\{\lambda_j\}\epsilon_n \le (C_2 - 1)\epsilon_n/2.
\tag{S4.30}
$$

Then, we lower bound P_1 as

$$
\begin{aligned}
P_1 &\ge \Pi_{\theta',\lambda}(\|\boldsymbol{\theta}_1' - \boldsymbol{\theta}_{01}\|_1 \le (C_2 - 1)\,\epsilon_n, \boldsymbol{\theta}_1' \ge \mathbf{0}_{S_0}, \lambda_{S_0} \in \tilde{\mathcal{A}} \mid \mathcal{A}) \\
&= \Pi_{\theta',\lambda}\Big(\|\boldsymbol{\theta}_1' - \mathbb{E}(\boldsymbol{\theta}_1'|\boldsymbol{\theta}_2') + \mathbb{E}(\boldsymbol{\theta}_1'|\boldsymbol{\theta}_2') - \boldsymbol{\theta}_{01}\|_1 \le (C_2 - 1)\,\epsilon_n, \boldsymbol{\theta}_1' \ge \mathbf{0}_{S_0}, \\
&\qquad\qquad \max_{j \in S_0}\{\lambda_j\} \le C_2 a_n/(2K_0 s_0) \mid \mathcal{A}\Big) \\
&\ge \Pi_{j \in S_0}\Big\{\Pi_{\tilde{\theta}_j,\lambda_j}\Big(|\tilde{\theta}_j - \theta_{0_j}| \le (C_2 - 1)\epsilon_n/(2s_0), \tilde{\theta}_j \ge (C_2 - 1)\epsilon_n/\{2(N - s_0)\}, \\
&\qquad\qquad\qquad \lambda_j \le C_2 a_n/(2K_0 s_0)\Big)\Big\},
\end{aligned}
\tag{S4.31}
$$

where $\tilde{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_1' - \mathbb{E}(\boldsymbol{\theta}_1'|\boldsymbol{\theta}_2') \sim \mathcal{N}(\mathbf{0}_{s_0}, \tau^2\Lambda_1\widetilde{\Omega}^{(1)}\Lambda_1)$ and $\widetilde{\Omega}^{(1)} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Sigma_{21}$. To derive the second inequality, first note that $\{\tilde{\theta}_j \ge \|\mathbb{E}(\boldsymbol{\theta}_1' \mid \boldsymbol{\theta}_2')\|_\infty\} \subset \{\theta_j' \ge 0\}$ for $j \in S_0$. Then following a similar argument in deriving equation (S4.30), one can show $\|\mathbb{E}(\boldsymbol{\theta}_1'|\boldsymbol{\theta}_2')\|_\infty \le (C_2-1)\epsilon_n/\{2(N-s_0)\}$. Then $\{\tilde{\theta}_j' \ge (C_2-1)\epsilon_n/\{2(N-s_0)\}, j \in S_0\} \subset \{\theta_j' \ge 0, j \in S_0\}$ and thus the last inequality in equation (S4.31) is attained.

To bound the last preceding in equation (S4.31), we use a similar proof of Theorem 1 in [4]. For any $a > 0$, we have

$$\int_0^a \frac{1}{\sqrt{2\pi}\lambda} e^{-\frac{x^2}{2\lambda^2}} \frac{2}{\pi(1+\lambda^2)} d\lambda = K' e^{\frac{x^2}{2}} E_1\{(1+a^{-2})x^2/2\} \ge (K'/2)\log[1+2\{(1+a^{-2})x^2/2\}^{-1}],$$

for some constant $K' > 0$, and $E_1(\cdot)$ is the exponential integral function. Then for any $j \in S_0$, under the Assumption (A3), applying the lower bound in the above display with $a = Ca_n/(2K_0 s_0)$, for any $j \in S_0$ we have

$$\begin{aligned}
&\Pi_{\tilde{\theta}_j,\lambda_j}\{|\tilde{\theta}_j - \theta_{0_j}| \le (C_2-1)\epsilon_n/s_0, \tilde{\theta}_j \ge (C_2-1)\epsilon_n/\{2(N-s_0), \lambda_j \le C_2 a_n/(2Ks_0)\} \\
&\quad \ge \int_{\lambda=0}^{C_2 a_n/(2Ks_0)} \int_{\theta_j \in \left[\max\left(\theta_{0_j} - \frac{(C_2-1)\epsilon_n}{2s_0}, \frac{(C_2-1)\epsilon_n}{2(N-s_0)}\right), \theta_{0_j} + \frac{(C_2-1)\epsilon_n}{2s_0}\right]} (2\pi\tau'^2\lambda_j^2)^{-1/2} e^{-\frac{\theta_j^2}{2\tau'^2\lambda_j^2}} d\theta_j f_{C+}(\lambda_j) d\lambda_j \\
&\quad \gtrsim \frac{(C_2-1)\epsilon_n}{4s_0} \frac{\tau'^2}{[1+\{C_2 a_n/(2K_0 s_0)\}^{-2}]E_n} \gtrsim n^{-(1+\mu)},
\end{aligned} \tag{S4.32}$$

where $\tau' = \tau(\widetilde{\Omega}_{jj}^{(1)})^{-1/2}$ and μ is some positive constant. The second inequality in the above display holds based on a similar argument in the proof of Theorem 3.3 in [17]. The last inequality holds by choosing $a_n \lesssim N^{-(2+t_0\bar{\omega}'_{\min})}$, where $\bar{\omega}'_{\min}$ is defined in Proposition S2 and by choosing $\tau \asymp n^{-(1+\alpha)}$ for some constant $0 < \alpha < \mu$ and $E_n \lesssim n^{c'}$ for some $0 < c' < \alpha$. Then with bounds in equations (S4.31) and (S4.32), we have shown $P_1 \gtrsim e^{-c_3 n\epsilon_n^2}$ for some constant $c_3 > 0$.

Next, we shall obtain a lower bound of P_2. Note that it is equivalent to upper bound the term

$$1 - P_2 = \frac{\Pi\Big(\{\boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}, \lambda_{S_0^c} \ge \mathbf{0}_{S_0^c}\}\backslash\{\|\boldsymbol{\theta}_2'\|_\infty \le \epsilon_n/(N-s_0), \boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}, \min_{j\in S_0^c}\{\lambda_j\} \ge a_n\}\Big)}{\Pi\Big(\boldsymbol{\theta}_2' \ge \mathbf{0}_{S_0^c}, \lambda_{S_0^c} \ge \mathbf{0}_{S_0^c}\Big)}. \tag{S4.33}$$

We first bound the denominator of equation (S4.33). A direct application of Proposition S2 for

$(\boldsymbol{\theta}'_2, \lambda_{S_0^c})$ leads to

$$\Pi(\boldsymbol{\theta}'_2 \geq \mathbf{0}_{S_0^c}, \lambda_{S_0^c} \geq \mathbf{0}_{S_0^c}) \gtrsim \{\log(N-s_0)\}^{-t_0/2}(N-s_0)^{-t_0\bar{\omega}'_{\min}}, \tag{S4.34}$$

for positive constants $t_0, \bar{\omega}'_{\min}$ defined in Proposition S2.

Now we bound the numerator of equation (S4.33) by a union bound of the probability. First we have

$$\Pi_{\theta',\lambda}\Big(\Big\{\boldsymbol{\theta}'_2 \geq \mathbf{0}_{S_0^c}, \lambda_{S_0^c} \geq \mathbf{0}_{S_0^c}\Big\} \Big\backslash \Big\{\|\boldsymbol{\theta}'_2\|_\infty \leq \epsilon_n/(N-s_0), \boldsymbol{\theta}'_2 \geq \mathbf{0}_{S_0^c}, \min_{j\in S_0^c}\{\lambda_j\} \geq a_n\Big\}\Big)$$
$$\leq 2\Big\{\Pi_{\theta',\lambda}(\|\boldsymbol{\theta}'_2\|_\infty \geq \epsilon_n/(N-s_0)) + \Pi_\lambda\Big(\min_{j\in S_0^c}\{\lambda_j\} \leq a_n\Big)\Big\}. \tag{S4.35}$$

We first bound the first term on the right hand side of equation (S4.35). For some $b_n > 0$ that may depend on n, N, define the set $\mathcal{A}_\lambda = \{\lambda_{S_0^c} : \max_{j\in S_0^c}\{\lambda_j\} \leq b_n\}$. Then we have

$$\Pi_{\theta',\lambda}(\|\boldsymbol{\theta}'_2\|_\infty \geq \epsilon_n/(N-s_0)) \leq \Pi_{\theta'|\lambda}(\|\boldsymbol{\theta}'_2\|_\infty \geq \epsilon_n/(N-s_0) \mid \mathcal{A}_\lambda) + \Pi_\lambda(\mathcal{A}_\lambda^c). \tag{S4.36}$$

Recall $\boldsymbol{\theta}'_2|\lambda_{S_0^c} \sim \mathcal{N}(\mathbf{0}_{N-s_0}, \tau^2\Lambda_2\Omega_{22}\Lambda_2)$. Then under the set $\mathcal{A}_\lambda$ we have $\|\Lambda_2\Omega_{22}\Lambda_2\| \leq \lambda_0^{-1} b_n^2$, where λ is defined in Assumption 2 and we have $\lambda_0^{-1} \geq \lambda_{\max}(\Omega_N)$. It is easy to show that conditioning on $\lambda_{S_0^c}$, $\boldsymbol{\theta}'_2$ is a sub-Gaussian random vector with parameter $\lambda_0^{-1}\tau^2 b_n^2$. Then by the concentration property of maximum of sub-Gaussian random variables, we have

$$\Pi_{\theta'|\lambda}(\|\boldsymbol{\theta}'_2\|_\infty \geq \epsilon_n/(N-s_0) \mid \mathcal{A}_\lambda) \leq 2\exp\left\{-\frac{\epsilon_n^2}{8(N-s_0)\tau^2 b_n^2}\right\} \lesssim e^{-c'_3 n\epsilon_n^2}, \tag{S4.37}$$

for some constant $c'_3 > 0$. The last inequality holds by the fact that

$$\mathbb{E}(\|\boldsymbol{\theta}'_2\|_\infty \mid \mathcal{A}_\lambda) \leq 2\sqrt{\lambda_0^{-1}\log(N-s_0)}\, b_n\tau = o(\epsilon_n/(N-s_0)),$$

by choosing $\tau \asymp n^{-(1+\alpha)}$ for some constant $\alpha > 0$ and choosing $b_n \asymp \tau^{-1}n^{-1/2}$, and by $N = o(n)$ under Assumption (A1).

Now to bound the second term on the right hand side of equation (S4.36), note that for large enough b_n, we have $\Pi_\lambda(\lambda_j \geq b_n) \asymp b_n^{-1}$ for any $1 \leq j \leq N$. Then we have

$$\Pi_\lambda(\mathcal{A}_\lambda{}^c) = \Pi_\lambda\Big(\max_{j \in S_0^c}\{\lambda_j\} \geq b_n\Big) = \sum_{k=1}^{N-s_0} \binom{N-s_0}{k} (b_n^{-1})^k (1-b_n^{-1})^{N-s_0-k}$$
$$\leq \sum_{k=1}^{N-s_0} \left(\frac{e(N-s_0)}{b_n}\right)^k \lesssim n^{-(1/2+\alpha)}, \tag{S4.38}$$

by choosing $b_n \asymp \tau^{-1} n^{-1/2}$ and $N = o(n)$. Then combining results in equations (S4.36),(S4.37) and (S4.38) leads to

$$\Pi_{\theta',\lambda}(\|\boldsymbol{\theta}_2'\|_\infty \geq \epsilon_n/(N-s_0)) \lesssim n^{-\mu''}, \tag{S4.39}$$

for some constant $0 < \mu'' \leq \min\{1/2+\alpha, c_3' n\epsilon_n^2\}$. Now to bound the second term on the right hand side of equation (S4.35), note that $\Pi(\lambda_j \leq a_n) \asymp a_n$ for $a_n \lesssim N^{-(2+t_0\bar{\omega}'_{\min})}$ and for any $1 \leq j \leq N$. Then

$$\Pi_\lambda\Big(\min_{j \in S_0^c}\{\lambda_j\} \leq a_n\Big) = \sum_{k=1}^{N-s_0} \binom{N-s_0}{k} a_n^k (1-a_n)^{N-s_0-k} \leq \sum_{k=1}^{N-s_0} (e(N-s_0)a_n)^k \lesssim N^{-(1+t_0\bar{\omega}'_{\min})}. \tag{S4.40}$$

where $t_0, \bar{\omega}'_{\min}$ are defined in Proposition S2. Combining results in equations (S4.34), (S4.39) and (S4.40) and the assumption $N = o(n)$, the numerator of the ratio on the right hand side of equation (S4.33) is upper bounded by a multiple of $N^{-(1+t_0\bar{\omega}'_{\min})}$. Then, we obtain $P_2 \geq 1 - 4(\log N)^{-t_0/2} N^{-1}$ for sufficiently large N. Further, combining this result with the lower bound of the term P_1 yields

$$\Pi(\|\theta - \theta_0\|_1 \leq C_2\epsilon_n) \geq P_1 \cdot P_2 \gtrsim \exp(-c_3 n\epsilon_n^2)\{1 - 4(\log N)^{-t_0/2} N^{-1}\},$$

for some constant $c_3 > 0$. We then complete verifying the condition (S4.25).

Case (ii). Conditioning on λ, we write the prior of θ' in the matrix notation as $\boldsymbol{\theta}' \mid \lambda \sim \mathcal{N}(\mathbf{0}_N, \tau^2 \Lambda\Omega\Lambda)$.

Then

$$\begin{aligned}\Pi_{\mathcal{C}}(\|\theta-\theta_0\|_1 \le C_2\,\epsilon_n) &\ge \Pi_{\theta',\lambda}(\|\boldsymbol{\theta}'-\boldsymbol{\theta}_0\|_1 \le C_2\,\epsilon_n', \theta' \ge \mathbf{0}_N)/\Pi_{\theta',\lambda}(\boldsymbol{\theta}' \ge \mathbf{0}_N)\\ &\ge \Pi_{\theta',\lambda}\Big(\cap_{j=1}^N \{|\theta_j'-\boldsymbol{\theta}_{0_j}| \le C_2\,\epsilon_n/N, \boldsymbol{\theta}_j' \ge 0\}\Big)\\ &\ge \left(\frac{\lambda_m}{\lambda_M}\right)^{N/2}\Pi_{j=1}^N \Pi_{\tilde{\theta}_j,\lambda_j}(|\tilde{\theta}_j-\boldsymbol{\theta}_{0_j}| \le C_2\,\epsilon_n/N, \tilde{\theta}_j \ge 0), \end{aligned} \tag{S4.41}$$

where $\tilde{\theta}_j|\lambda_j \sim \mathcal{N}(0, \tau'^2\lambda_j^2)$ with $\tau' = \lambda_m\tau$. Then applying the result used in proving Theorem 3.3 in [18], for $j = 1, \ldots, N$,

$$\Pi_{\tilde{\theta}_j,\lambda_j}(|\tilde{\theta}_j-\boldsymbol{\theta}_{0_j}| \le C_2\,\epsilon_n'/N, \tilde{\theta}_j \ge 0) \gtrsim \frac{\epsilon_n'}{N}\frac{\tau'^2}{E_n^2} \asymp n^{-(1+\alpha')},$$

for some constant $\alpha' > 0$. The last result is arrived by choosing $\tau \asymp n^{-(1+\alpha)}$ and $E_n \lesssim n^{\beta}$ for some constant $\beta \in (0, \alpha)$.

We then complete the verification of condition (S4.25) by combining cases (i) and (ii).

Part II. Next we verify the condition (S4.26). We first construct a sequence of sieves $\{\mathcal{F}_n\}$ that satisfies condition (S4.26). Define

$$\mathcal{F}_n = \left\{(\theta,\sigma^2) : |S| \le T_{s_0}, 0 < \sigma^2 < e^{n\epsilon_n^2/a_0}\right\}, \quad \text{with} \quad T_{s_0} = (s_0+p_0)\mathbb{1}_{s_0<[N/2]}(s_0) + N\mathbb{1}_{s_0\ge[N/2]}(s_0), \tag{S4.42}$$

where the integer $0 < p_0 < [N/2] - s_0$ for $s_0 < [N/2]$, and recall a_0 is the shape parameter of inverse-Gamma prior on σ^2. Then, applying the union bound of probability yields

$$\Pi(\mathcal{F}_n^c) \le 2\Pi_{\theta,\lambda}\big(|S| \ge (s_0+p_0)+1\big) + 2\Pi_{\sigma^2}\left(\sigma^2 \ge e^{n\epsilon_n^2/a_0}\right).$$

We now bound $\Pi_{\theta,\lambda}(|S| \ge (s_0+p_0)+1)$. We define the threshold $\eta_n = \epsilon_n/\sqrt{N}$. For any set S with $|S| = s$ for $1 \le s \le N$,

$$\begin{aligned}\Pi_{\mathcal{C}}(\theta_j \ge \eta_n, j \in S; \theta_j < \eta_n, j \in S^c) &\le \Pi_{\theta',\lambda}(\theta_j' \ge \eta_n, j \in S; \theta_j' < \eta_n, j \in S^c)/\mathcal{M}_{\mathcal{C}}'\\ &\le \Pi_{\theta',\lambda}(\theta_j' \ge \eta_n, j \in S)/\mathcal{M}_{\mathcal{C}}',\end{aligned}$$

where $\mathcal{M}'_{\mathcal{C}}$ denotes the lower bound of normalizing constant $\mathcal{M}_{\mathcal{C}}$ obtained by Proposition S2. Let $\tau'' = \tau\sqrt{\lambda_{\max}(\Omega_S)}$, where $\Omega_S = (\omega_{ij})$ for all $i, j \in S$ is a $|S| \times |S|$ sub-matrix of Ω_N. For any $j \in S$, we have

$$\begin{aligned}\Pi_{\theta'}(\theta'_j \geq \eta_n) &\leq \sqrt{\frac{\lambda_{\max}(\Omega_S)}{\lambda_{\min}(\Omega_S)}} \int_{\lambda_j=0}^{\infty} (\lambda_j \tau'')^{-1} \int_{\eta_n}^{\infty} e^{-\frac{\theta_j^2}{2\lambda_j^2 \tau''^2}} d\theta_j f_{C+}(\lambda_j) d\lambda_j \\ &\lesssim \frac{\sqrt{2/\pi} e^{-\eta_n^2/(2b_n^2\tau''^2)}}{\eta_n/(b_n\tau'')} + b_n^{-1} \lesssim e^{-c_4 n\epsilon_n^2/(s_0+p_0+1)},\end{aligned}$$

for some constant $c_4 > 0$. Under Assumption 2, the term $\sqrt{\lambda_{\max}(\Omega_S)/\lambda_{\min}(\Omega_S)} \leq \lambda_0^{-1}$ for λ_0 defined in Assumption 2. The last inequality in the above display is obtained by choosing τ such that $\tau \asymp n^{-(1+\alpha)}$ and choosing $b_n = \tau^{-1}/\sqrt{n}$, we then have $\exp\{-\eta_n^2/(2b_n^2\tau''^2)\}/[\eta_n/(b_n\tau'')] \lesssim e^{-c_4 n\epsilon_n^2/(s_0+p_0+1)}$ for some constant $c_4 > 0$. Then combining the preceding, we have

$$\begin{aligned}\Pi_{\theta,\lambda}(|S| \geq (s_0+p_0)+1) &= \sum_{k=(s_0+p_0)+1}^{N} \binom{N}{k} \Pi_{\mathcal{C}}(|\theta_j| \geq \eta_n, j \in S; |\theta_j| \leq \eta_n, j \notin S \mid |S| = k) \\ &\leq \sum_{k=(s_0+p_0)+1}^{N} \binom{N}{k} \Pi_{\theta'}(\theta'_j \geq \eta_n, j \in S)/\mathcal{M}'_{\mathcal{C}} \\ &\lesssim (\log N)^{t_0/2} N^{t_0\bar{\omega}'_{\min}} \sum_{k=(s_0+p_0)+1}^{N} \left(\frac{e}{(s_0+p_0)+1} e^{-c_4 n\epsilon_n^2/(s_0+p_0+1)}\right)^k \\ &\lesssim e^{-c'_4 n\epsilon_n^2}, \end{aligned} \tag{S4.43}$$

for some constant $c'_4 > 0$. Recall $t_0, \bar{\omega}'_{\min}$ are positive constants defined in Proposition S2.

Now we bound $\Pi(\sigma^2 \geq e^{n\epsilon_n^2/a_0})$. Since $\sigma^2 \sim \mathrm{IG}(a_0, b_0)$, we have

$$\begin{aligned}\Pi_{\sigma^2}(\sigma^2 > e^{n\epsilon_n^2/a_0}) &= \int_{e^{n\epsilon_n^2/a_0}}^{\infty} = \frac{a_0^{b_0}}{\Gamma(a_0)} \int_{e^{n\epsilon_n^2/a_0}}^{\infty} (\sigma^2)^{-(a_0+1)} e^{-b_0/\sigma^2} d\sigma^2 \\ &\leq \frac{a_0^{b_0}}{\Gamma(a_0)} \int_{e^{n\epsilon_n^2/a_0}}^{\infty} (\sigma^2)^{-(a_0+1)} d\sigma^2 \asymp e^{-n\epsilon_n^2}.\end{aligned} \tag{S4.44}$$

Combing results in equations (S4.43) and (S4.44), we arrive at $\Pi(\mathcal{F}_n^c) \lesssim e^{-c_2 n\epsilon_n^2}$ with $c_2 = \min\{c'_4, 1\}$, then we have verified condition (S4.26).

Part III. Now we show the existence of test functions $\{\phi_n\}$ that satisfy condition (S4.27). To that end, we consider a similar construction of test functions in the proof of Theorem 1 of [15].

For any nonempty subset $S \subset \{1, \ldots, N\}$, define the (restricted) ordinary least squares estimator $\hat{\theta}_S = \hat{\theta} := (\Psi_S^{\mathrm{T}} \Psi_S)^{-1} \Psi_S^{\mathrm{T}} Y$, and $\hat{\sigma}_S^2 = Y^{\mathrm{T}}(\mathrm{I_n} - H_S)Y/(n - |S|)$ where $H_S = \Psi_S(\Psi_S^{\mathrm{T}} \Psi_S)^{-1} \Psi_S^{\mathrm{T}}$. When $S = \{1, \ldots, N\}$, then $\hat{\theta}_S = (\Psi^{\mathrm{T}} \Psi)^{-1} \Psi^{\mathrm{T}} Y$ is the ordinary least squares estimator. Recall $T_{s_0} = (s_0 + p_0) \mathbb{1}_{s_0 < [N/2]}(s_0) + N \mathbb{1}_{s_0 \geq [N/2]}(s_0)$ defined in equation (S4.42). Similar to the construction of test functions in [15], define $\phi_n = \max\{\phi_n', \tilde{\phi}_n\}$, where

$$\phi_n' = \max_{\{S \supset S_0, |S| \leq T_{s_0}\}} \mathbb{1}\Big\{ ||\hat{\theta}_S - \theta_0|| \geq c_5 \, \sigma_0 \sqrt{N} \epsilon_n \Big\},$$
$$\tilde{\phi}_n = \max_{\{S \supset S_0, |S| \leq T_{s_0}\}} \mathbb{1}\Big\{ |\hat{\sigma}_S^2 - {\sigma_0}^2| \geq c_5' \, \sigma_0^2 \, \epsilon_n \Big\}.$$

for some constants $c_5, c_5' > 0$, and recall that the set S_0 contains indexes of pseudo-true nonzero coordinates.

Recall θ_0 denotes the pseudo-true coefficient vector, and denote the bias term by $\delta := \Psi\theta_0 - f_0$. Under the true distribution, we have $Y_i = f_0(x_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma_0^2)$ i.i.d., for $i = 1, \ldots, n$. For brevity, we use ε to denote the random error term $\{\epsilon_i\}$. Given the definition of $\hat{\sigma}_S^2$, we have

$$\begin{aligned} &\mathbb{E}_{f_0, \sigma_0^2} \mathbb{1}\Big\{ |\hat{\sigma}_S^2 - {\sigma_0}^2| \geq c_5' \, \sigma_0^2 \epsilon_n \Big\} \\ &\quad = \mathbb{P}_{f_0, \sigma_0^2}\big(|\varepsilon^{\mathrm{T}}(\mathrm{I_n} - H_S)\varepsilon + 2\varepsilon^{\mathrm{T}}(\mathrm{I_n} - H_S)\delta + \delta^{\mathrm{T}}(\mathrm{I_n} - H_S)\delta| \geq c_5'(n - |S|)\epsilon_n \big). \end{aligned}$$

Applying Lemma 1, we obtain

$$\delta^{\mathrm{T}}(\mathrm{I_n} - H_S)\delta \leq \|\delta\|^2 \leq (n - |S|)\|\delta\|_\infty^2 \lesssim (n - |S|)\epsilon_n^2 \leq (n - |S|)\epsilon_n,$$
$$\varepsilon^{\mathrm{T}}(\mathrm{I_n} - H_S)\delta \leq \|\varepsilon\|\|\delta\| \lesssim \sqrt{(n - |S|)}\epsilon_n \, \|\varepsilon\|.$$

According to [18], one can show $\mathbb{P}(|2\varepsilon^{\mathrm{T}}(\mathrm{I_n} - H_S)\delta| \geq (n - |S|)\epsilon_n) \leq e^{-c_6 n \epsilon_n^2}$ for some constant $c_6 > 0$. Also with a similar discussion in Theorem A.1 of [15] we arrive at

$$\mathbb{E}_{f_0, \sigma_0^2} \mathbb{1}\Big\{ |\hat{\sigma}_S^2 - {\sigma_0}^2| \geq c_5' \, \sigma_0^2 \epsilon_n \Big\} \leq e^{-c_6' n \epsilon_n^2}, \tag{S4.45}$$

for some constants $c_6' > 0$. Under the true data generating distribution,

$$\|\hat{\theta}_S - \theta_{0S}\| \le \|(\Psi_S^{\mathrm{T}}\Psi_S)^{-1}\Psi_S^{\mathrm{T}}\delta\| + \|(\Psi_S^{\mathrm{T}}\Psi_S)^{-1}\Psi_S^{\mathrm{T}}\varepsilon\| \le c_0'\sqrt{N}\epsilon_n + \|(\Psi_S^{\mathrm{T}}\Psi_S)^{-1}\Psi_S^{\mathrm{T}}\varepsilon\|,$$

for some constant $c_0' > 0$. The last inequality holds based on $\|(\Psi_S^{\mathrm{T}}\Psi_S)^{-1}\Psi_S^{\mathrm{T}}\delta\| \le (nk_1/N)^{-1}\sqrt{nk_2/N}\|\delta\| \le c_0'\sqrt{N}\epsilon_n$ by Assumption (A2) and Lemma 1 in the manuscript. Again, similar to the proof of Theorem A.1 in [15], one can show

$$\begin{aligned}\mathbb{P}_{f_0,\sigma_0^2}\left(\|\hat{\theta}_S - \theta_{S0}\| \ge c_5\sqrt{N}\sigma_0\,\epsilon_n\right) &\le \mathbb{P}_{f_0,\sigma_0^2}\left(\|(\Psi_S^{\mathrm{T}}\Psi_S)^{-1}\Psi_S^{\mathrm{T}}\varepsilon\| \ge c_5''\sigma_0\sqrt{N}\epsilon_n\right) \\ &\le \mathbb{P}\left(\chi^2_{|S|} \ge \tilde{c}_5\, n\epsilon_n^2\right) \le e^{-c_6''\, n\epsilon_n^2}.\end{aligned} \tag{S4.46}$$

for some constants $c_5'', \tilde{c}_5, c_6'' > 0$. The random variable $\chi^2_{|S|}$ follows a chi-square distribution with the degree of freedom $|S|$. The last inequality in equation (S4.46) holds since $|S| \lesssim n\epsilon_n^2$ and $\lambda_{\max}((\Psi_S^{\mathrm{T}}\Psi_S)^{-1}) = (k_1 n/N)^{-1}$ under Assumption (A2). Combining the bound results in equations (S4.45) and (S4.46), we have

$$\begin{aligned}\mathbb{E}_{f_0,\sigma_0^2}(\phi_n) &\le \mathbb{E}_{f_0,\sigma_0^2}\sum_{\{S\supset S_0,|S|\le T_{s_0}\}}(\phi_n' + \tilde{\phi}_n) \\ &\le \mathbb{E}_{f_0,\sigma_0^2}\left(\sum_{\{S\supset S_0,|S|\le(s_0+p_0),s_0<[N/2]\}} + \sum_{\{S\supset S_0,[N/2]\le|S|\le N,s_0\ge[N/2]\}}\right)(\phi_n' + \tilde{\phi}_n) \\ &\le \left\{(s_0+p_0)\binom{N}{s_0+p_0} + [N/2]\binom{N}{[N/2]}\right\}\left(e^{-c_6'\,n\epsilon_n^2} + e^{-c_6''\,n\epsilon_n^2}\right) \\ &\lesssim \left(e^{(s_0+p_0)\log N} + e^{\{e[N/2]\}\log N}\right)\left(e^{-c_6'\,n\epsilon_n^2} + e^{-c_6''\,n\epsilon_n^2}\right) \lesssim e^{-\tilde{c}_6\,n\epsilon_n^2},\end{aligned} \tag{S4.47}$$

for some positive constant $\tilde{c}_6 \le \min\{c_6', c_6''\}$. The third inequality arrives based on the fact that the total number of models S with $|S| \ge [N/2]$ is same as the total number of the corresponding complement models S^c of size that is no greater than $[N/2]$. The first inequality in the fourth line of the preceding holds due to the fact that $(s_0+p_0)\log N < n\epsilon_n^2$, since $p_0 < [N/2]-s_0$ when $s_0 < [N/2]$, and the fact when $s_0 \ge [N/2]$, one obtains $e[N/2]\log N < n\epsilon_n^2$. Then, the final inequality result in the preceding is easily obtained by choosing $\tilde{c}_6 \le \min\{c_6', c_6''\}$.

Now we verify the second part of the condition (S4.27). Define the set

$$\mathcal{C}_n = \Big\{\|\theta-\theta_0\| \geq M_1\sigma_0\sqrt{N}\epsilon_n, \text{ or } \sigma^2/\sigma_0^2 \leq (1-\epsilon_n)/(1+\epsilon_n), \text{ or } \sigma^2/\sigma_0^2 \geq (1+\epsilon_n)/(1-\epsilon_n)\Big\}.$$

Then, according to Lemma D.2 in [1], we have

$$\sup_{\substack{(\theta,\sigma^2)\in\mathcal{F}_n:\|\theta-\theta_0\|\geq M_1\sigma_0\sqrt{N}\epsilon_n,\\ \text{or } |\sigma^2-\sigma_0^2|\geq\sigma_0^2\epsilon_n}} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n) \leq \sup_{(\theta,\sigma^2)\in\mathcal{C}_n} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n).$$

Again, following a similar discussion in the proof of Theorem A.1 in [15], define sets

$$\begin{aligned}\tilde{\mathcal{C}}_n &= \Big\{\sigma^2/\sigma_0^2 \leq (1-\epsilon_n)/(1+\epsilon_n) \text{ or } \sigma^2/\sigma_0^2 \geq (1+\epsilon_n)/(1-\epsilon_n)\Big\},\\ \mathcal{C}'_n &= \Big\{\ \|\theta-\theta_0\| \geq M_1\sigma_0\sqrt{N}\epsilon_n, \text{ and } \sigma^2=\sigma_0^2\Big\}.\end{aligned} \tag{S4.48}$$

And we have $\mathcal{C}_n \subset \tilde{\mathcal{C}}_n \cup \mathcal{C}'_n$. Similar to [15], one can show

$$\begin{aligned}\sup_{(\theta,\sigma^2)\in C_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n) &\leq \sup_{(\theta,\sigma^2)\in(C'_n\cup\tilde{C}_n)\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n)\\ &\leq \max\left\{\sup_{(\theta,\sigma^2)\in\tilde{C}_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\tilde{\phi}_n), \sup_{(\theta,\sigma^2)\in C'_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\phi'_n)\right\}.\end{aligned}$$

To proceed we consider two cases: (i) $s_0 < N$ and (ii) $s_0 = N$, separately.

Case (i). Now let $\tilde{S} = \{\theta : |\theta_j/\sigma| \geq a_n\} \cup S_0$ satisfying $|\tilde{S}| < N$, and denote $\tilde{S}^c = \{1,\ldots,N\}\backslash\tilde{S}$. Then, with the same argument in the proof of Theorem A.1 in [15] we have

$$\sup_{(\theta,\sigma^2)\in\tilde{C}_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\tilde{\phi}_n) \leq \sup_{(\theta,\sigma^2)\in\tilde{C}_n\cap\mathcal{F}_n} \mathbb{P}\Big\{|\chi^2_{n-|\tilde{S}|}(m) - (n-|\tilde{S}|)| \geq (n-|\tilde{S}|)\epsilon_n\Big\} \leq e^{-\tilde{c}_7 n\epsilon_n^2}, \tag{S4.49}$$

for some constant $\tilde{c}_7 > 0$. Here $\chi^2_{n-|\tilde{S}|}(m)$ denotes a non-central chi-square random variable with the non-central parameter $m = \theta^{\mathrm{T}}_{\tilde{S}^c}\Psi^{\mathrm{T}}_{\tilde{S}^c}(I-H_{\tilde{S}})\Psi_{\tilde{S}^c}\theta_{\tilde{S}^c}/\sigma^2$. It is easy to show there exists some constant $k_0 > 0$ such that $m \leq nk_0\epsilon_n^2$, based on the fact that $\|\Psi_{\tilde{S}^c}\theta_{\tilde{S}^c}\| \leq \sqrt{\lambda_{\max}(\Psi^{\mathrm{T}}_{\tilde{S}^c}\Psi_{\tilde{S}^c})}\sqrt{N}\|\theta_{\tilde{S}^c}\|_\infty \leq \sqrt{nk_2}\,\epsilon_n$, by Assumption (A2) with choosing $S = \tilde{S}^c$ and $k_2 > 0$ is the constant defined therein.

Next, we have

$$
\begin{aligned}
\sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n') &\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S})^{-1}\Psi_{\tilde S}^{\mathrm{T}} Y-\theta_{0\tilde S}\| \le \sigma_0\sqrt{N}\epsilon_n\right)\\
&\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S})^{-1}\Psi_{\tilde S}^{\mathrm{T}}\varepsilon\| \ge \|\theta_{\tilde S}-\theta_{0\tilde S}\| - (\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S})^{-1}\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S^c}\theta_{\tilde S^c}-\sqrt{N}\epsilon_n\right)\\
&\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S})^{-1}\Psi_{\tilde S}^{\mathrm{T}}\varepsilon\| \ge \sqrt{N}\epsilon_n\right)\\
&\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\chi^2_{|\tilde S|}\ge nk_1\epsilon_n^2\right) \le e^{-c_7' n\epsilon_n^2},
\end{aligned}
$$

for some constant $c_7'>0$, and $\chi^2_{|\tilde S|}$ denotes a centered chi-square random variable of the degree of freedom $|\tilde S|<n\epsilon_n^2$. The second inequality holds based on the following facts that $\|\theta_{\tilde S}-\theta_{0\tilde S}\|\ge \|\theta-\theta_0\|-N(\epsilon_n\sigma/N)$, and $\|(\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S})^{-1}\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S^c}\theta_{\tilde S^c}\| \le \sqrt{\lambda_{\max}((\Psi_{\tilde S}^{\mathrm{T}}\Psi_{\tilde S})^{-1})\lambda_{\max}(\Psi_{\tilde S^c}^{\mathrm{T}}\Psi_{\tilde S^c})}\sqrt{N}\|\theta_{\tilde S^c}\|_1 \le \sqrt{\{N/(k_1 n)\}nk_2}\epsilon_n \lesssim \sqrt{N}\epsilon_n$, by Assumption (A2). Then we complete verifying condition (S4.27) for Case (i).

Case (ii). The arguments are very similar to Case (i), except $S^c=\emptyset$. Then

$$
\sup_{(\theta,\sigma^2)\in\tilde C_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\tilde\phi_n) \le \sup_{(\theta,\sigma^2)\in\tilde C_n\cap\mathcal{F}_n} \mathbb{P}\{|\chi^2_{n-N}-(n-N)|\ge (n-N)\epsilon_n\} \le e^{-\tilde c_7 n\epsilon_n^2}, \quad \text{(S4.50)}
$$

where χ^2_{n-N} is a centered chi-square random variable with the degree of freedom $n-N$ for $N=o(n)$. And, following a similar argument in Case (i) yields

$$
\begin{aligned}
\sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\phi_n') &\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi^{\mathrm{T}}\Psi)^{-1}\Psi^{\mathrm{T}}\varepsilon\| \ge \|\theta-\theta_0\|-\sqrt{N}\epsilon_n\right)\\
&\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi^{\mathrm{T}}\Psi)^{-1}\Psi^{\mathrm{T}}\varepsilon\| \ge \sqrt{N}\epsilon_n\right)\\
&\le \sup_{(\theta,\sigma^2)\in C_n'\cap\mathcal{F}_n} \mathbb{P}\left(\chi^2_N\ge nk_1\epsilon_n^2\right) \le e^{-\tilde c_7' n\epsilon_n^2},
\end{aligned}
$$

for some constant $\tilde c_7'>0$.

Combining Cases (i) and (ii) we have verified the second part of condition (S4.27). Therefore we have shown results in equation (4.6).

Proof of equation (4.7). We now verify equation (4.7) to complete the proof of Theorem 8. The proof of equation (4.7) is very similar to that of equation (4.6). In particular, the proof of condition (S4.25) can be applied directly here, and since we construct the same sieves $\mathcal{F}_n$ thus condition (S4.26) is satisfied. We shall verify the condition (S4.27) by showing the existence of test functions $\{\zeta_n\}$ such that

$$\mathbb{E}_{f_0,\sigma_0^2}(\zeta_n) \lesssim e^{-c_3 n\epsilon_n^2}, \tag{S4.51}$$

$$\sup_{\substack{(\theta,\sigma^2)\in\mathcal{F}_n:\|\Psi\theta-f_0\|\geq M_1\sigma_0\sqrt{n}\epsilon_n,\\ \text{or } |\sigma^2-\sigma_0^2|\geq\sigma_0^2\epsilon_n}} \mathbb{E}_{\theta,\sigma^2}(1-\zeta_n) \lesssim e^{-c_3 n\epsilon_n^2}. \tag{S4.52}$$

Similar to [15], consider the test function $\zeta_n = \max\{\zeta_n', \tilde{\zeta}_n\}$, where

$$\zeta_n' = \max_{\{S\supset S_0, |S|\leq T_{s_0}\}} \mathbb{1}\left\{||\Psi_S\hat{\theta}_S - f_0|| \geq c_5\,\sigma_0\sqrt{n}\epsilon_n\right\},$$

$$\tilde{\zeta}_n = \max_{\{S\supset S_0, |S|\leq T_{s_0}\}} \mathbb{1}\left\{|\hat{\sigma}_S^2 - {\sigma_0}^2| \geq c_5'\,\sigma_0^2\epsilon_n\right\},$$

for some constants $c_5, c_5' > 0$. The argument for deriving exponential upper bounds for type I errors of $\tilde{\zeta}_n$ remains the same as in (S4.45), therefore we omit it here.

We now derive the exponential error bounds for ζ_n' under the true data generating function. For any $S \in \{S : S \supset S_0, |S| \leq T_{s_0}\}$, then we have

$$\begin{aligned}
\|\Psi_S\hat{\theta}_S - f_0\|^2 &\leq \|\Psi_S(\hat{\theta}_S - \theta_{0S})\|^2 + n\|\Psi\theta_0 - f_0\|_\infty^2\\
&\lesssim \|\Psi_S(\hat{\theta}_S - \theta_{0S})\|^2 + n\epsilon_n^2\\
&\leq (nk_2/N)\|\hat{\theta}_S - \theta_{0S}\|^2 + n\epsilon_n^2,
\end{aligned}$$

where the constant $k_2 > 0$ is defined in Assumption (A2). The first inequality holds as we assume $S \supset S_0$, $\theta_{0S^c} = \mathbf{0}_{N-|S|}$ and $\Psi_{S^c}\theta_{0S^c} = \mathbf{0}_{N-|S|}$. The second and third inequality use Lemma 1 and Assumption (A2), respectively. Thus,

$$\mathbb{P}_{f_0,\sigma_0^2}\left(\|\Psi_S\hat{\theta}_S - f_0\| \geq c_5\,\sigma_0\sqrt{n}\epsilon_n\right) \leq \mathbb{P}_{f_0,\sigma_0^2}\left(\|\hat{\theta}_S - \theta_{S0}\| \geq k_2^{-1}c_5''\,\sigma_0\sqrt{N}\epsilon_n\right),$$

for some constant $0 < c_5'' < c_5$. The inequality is obtained based on a similar argument used to derive equation (S4.46). Combining the above result with the result in equation (S4.46) and results in equation (S4.45), and following a similar argument of deriving the result in equation (S4.47) yields the desired bound in equation (S4.51).

Now we show equation (S4.52). Similarly, Recall the set $\tilde{C}_n$ defined in equation (S4.48) and define the set,

$$\mathcal{D}'_n = \left\{\|\Psi\theta - f_0\| \geq M_1\sigma_0\sqrt{n}\epsilon_n, \text{ and } \sigma^2 = \sigma_0^2\right\}.$$

With a same argument, one can show

$$\begin{aligned}
&\sup_{\substack{(\theta,\sigma^2)\in\mathcal{F}_n:\|\Psi\theta-f_0\|\geq M_1\sigma_0\sqrt{n}\epsilon_n,\\ \text{or } |\sigma^2-\sigma_0^2|\geq\sigma_0^2\epsilon_n}} \mathbb{E}_{\theta,\sigma^2}(1-\zeta_n)\\
&\qquad\qquad\leq \max\left\{\sup_{(\theta,\sigma^2)\in\tilde{C}_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\tilde{\zeta}_n), \sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\zeta'_n)\right\}.
\end{aligned}$$

The argument of bounding $\sup_{(\theta,\sigma^2)\in\tilde{C}_n\cap\mathcal{F}_n}\mathbb{E}_{\theta,\sigma^2}(1-\tilde{\zeta}_n)$ is same as the the proof of the result in equation (S4.49), therefore we omit it here. It remains to bound $\sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n}\mathbb{E}_{\theta,\sigma^2}(1-\zeta'_n)$. We proceed to consider two cases: (i) $s_0 < N$ and (ii) $s_0 = N$, respectively.

Case (i). Again, let $\tilde{S} = \{\theta : |\theta_j/\sigma| \geq a_n\} \cup S_0$ satisfying $|\tilde{S}| < N$, and denote $\tilde{S}^c = \{1,\ldots,N\}\backslash\tilde{S}$. Applying Lemma 1, we have

$$\begin{aligned}
&\sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\zeta'_n)\\
&\leq \sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi_{\tilde{S}}(\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}})^{-1}\Psi_{\tilde{S}}^{\mathrm{T}}\varepsilon\| \geq \|\Psi_{\tilde{S}}\theta_{\tilde{S}} - \Psi_{\tilde{S}}\theta_{0\tilde{S}}\| - \Psi_{\tilde{S}}(\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}})^{-1}\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}^c}\theta_{\tilde{S}^c} - \sqrt{n}\epsilon_n\right)\\
&\leq \sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi_{\tilde{S}}(\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}})^{-1}\Psi_{\tilde{S}}^{\mathrm{T}}\,\varepsilon\| \geq \sqrt{n}\epsilon_n\right) \leq e^{-c_3 n\epsilon_n^2},
\end{aligned}$$

for some constant $c_3 > 0$. The second inequality holds by the facts that $\|\Psi_{\tilde{S}}\theta_{\tilde{S}} - \Psi_{\tilde{S}}\theta_{0\tilde{S}}\| \geq \sqrt{\lambda_{\min}(\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}}^{\mathrm{T}})}\|\theta_{\tilde{S}} - \theta_{0\tilde{S}}\| \gtrsim \sqrt{n}\epsilon_n$ under Assumption (A2). And $\|\Psi_{\tilde{S}}(\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}})^{-1}\Psi_{\tilde{S}}^{\mathrm{T}}\Psi_{\tilde{S}^c}\theta_{\tilde{S}^c}\| \lesssim \|\Psi_{\tilde{S}^c}\theta_{\tilde{S}^c}\| \leq \sqrt{nk_2}\epsilon_n$.

Case (ii). The argument follows a similar line of the one in Case (i), except $\tilde{S} = \{1,\ldots,N\}$ and

$\tilde{S}^c = \emptyset$. It is easy to show

$$\sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\zeta'_n) \leq \sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{P}\left(\|(\Psi(\Psi^{\mathrm{T}}\Psi)^{-1}\Psi^{\mathrm{T}}\,\varepsilon\| \geq \|\Psi\theta - \Psi\theta_0\| - \sqrt{n}\epsilon_n\right).$$

With same calculations, one can show $\sup_{(\theta,\sigma^2)\in\mathcal{D}'_n\cap\mathcal{F}_n} \mathbb{E}_{\theta,\sigma^2}(1-\zeta'_n) \leq e^{-c_3 n\epsilon_n^2}$ for some constant $c_3 > 0$.

We then complete the proof of equation (4.7) by combining Case (i) and Case (2), and therefore we complete the proof of Theorem 8.

S4.6 Proof of Corollary S1 in § S2.2

The proof of Corollary S1 follows a similar line of arguments in the proof of Theorem 2. We shall first adopt similar techniques used in the *Step 1* of the proof of Theorem 2 to obtain an upper bound of $\alpha_{N,\delta}$ and then directly show such upper bound decreases to 0 along with K. Under the assumption, there exists a K-banded approximating matrix $\Sigma'_N = (\sigma'_{ij}) \in \mathcal{B}_{N,K}$ to the scale matrix $\Sigma_N = (\sigma_{ij})$ such that $\|\Sigma_N - \Sigma'_N\| \lesssim (N\log K)^{-1}\|\Sigma_N\|$, for some integer $2 \leq K \leq N-1$ and sufficiently large N. Also, recall we assume $\sigma'^2_{(1)} = 1$ and we denote $\sigma'^2_{(N)} = \kappa \geq 1$. Denote $\bar{\rho}_{\max} = (1-\sigma'_{\max})/\sigma'_{\max}$ and $\bar{\rho}_{\kappa,\min} = (\kappa - \sigma'_{\min})/\sigma'_{\min}$. Recall the matrix $\Sigma_N(\sigma^2,\rho) = (\sigma_{ij})$ with $\sigma_{ii} > 0$ for $i = 1,\ldots,N$ and $\sigma_{ij} = \rho$ for some $\rho < \min_{1\leq i\leq N}\{\sigma_{ii}\}$ for any $1 \leq i \neq j \leq N$.

Fix an arbitrary $\delta > 0$, recall $\alpha_{N,\delta} = \mathbb{P}(0 < \theta_1 < \delta)$, where $\theta \sim \mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}_N, \Sigma_N)$. Define independent N-dimensional random vectors $Z \sim \mathcal{N}(\boldsymbol{\mu}_N, \Sigma_N)$ and $Z' \sim \mathcal{N}(\boldsymbol{\mu}_N, \Sigma'_N)$. Again, by definition,

$$\begin{aligned}
\alpha_{N,\delta} &\overset{(i)}{\asymp} \frac{\mathbb{P}(0 \leq Z'_1 \leq \delta, Z'_2 \geq 0, \ldots, Z'_N \geq 0)}{\mathbb{P}(Z'_1 \geq 0, Z'_2 \geq 0, \ldots, Z'_N \geq 0)} \\
&\overset{(ii)}{\leq} \frac{\mathbb{P}\left(0 \leq Z'_1 \leq \delta, Z'_{[2:K]} \geq \mathbf{0}_K\right)}{\mathbb{P}\left(Z'_{[1:K]} \geq \mathbf{0}_K\right)\mathbb{P}\left(Z'_{[K+1:2K]} \geq \mathbf{0}_K\right)} = R_1.
\end{aligned} \tag{S4.53}$$

The inequality (i) holds by using the result $\mathbb{P}(0 < \theta_1 < \delta) \asymp \mathbb{P}(0 < \theta'_1 < \delta)$ in Proposition S1. The inequality (ii) follows by applying Lemma S1 to the numerator and applying Slepian's lemma to the denominator of the ratio in the first line of equation (S4.53). The details are same as in the derivation of inequalities (S4.3) and (S4.4) in the proof of Theorem 2.

Following the line of argument in the proof of Theorem 2, we apply another round of Lemma

S1 to the numerator of R_1 and Lemma S4 to the denominator of R_1 and obtain an upper bound of R_1 which can be expressed as a ratio of probabilities associated with equicorrelated normal random vectors. Then we can proceed by applying Lemma S2 and Lemma S3 to arrive at our final upper bound. To that end, define $\widetilde{Z}' \sim \mathcal{N}(\boldsymbol{\mu}_K, \Sigma_K(\sigma^2, \rho_{\max}))$, then we have $\mathbb{E}(Z_i'^2) = \mathbb{E}(\widetilde{Z}_i'^2)$ for $1 \le i \le K$ and $\mathbb{E}(Z_i'Z_j') \le \mathbb{E}(\widetilde{Z}_i'\widetilde{Z}_j')$ for all $1 \le i,j \le K$. Similarly, define $\widetilde{Z}'' \sim \mathcal{N}(\boldsymbol{\mu}_K, \Sigma_K(\sigma^2, \rho_{\min}))$, then we have $\mathbb{E}(Z_i'^2) = \mathbb{E}(\widetilde{Z}_i''^2)$ for $1 \le i \le 2K$ and $\mathbb{E}(\widetilde{Z}_i''\widetilde{Z}_j'') \le \mathbb{E}(Z_i'Z_j')$ for all $1 \le i,j \le 2K$. Applying Lemma S1 to the numerator of R_1 and applying Slepian's lemma to the denominator of R_1, we have

$$R_1 \le R_1' = \frac{\mathbb{P}\big(0 \le \widetilde{Z}_1' \le \delta, \widetilde{Z}_2' \ge 0, \ldots, \widetilde{Z}_K' \ge 0\big)}{\mathbb{P}\big(\widetilde{Z}_1'' \ge 0, \ldots, \widetilde{Z}_K'' \ge 0\big)\mathbb{P}\big(\widetilde{Z}_{K+1}'' \ge 0, \ldots, \widetilde{Z}_{2K}'' \ge 0\big)}.$$

Now apply the upper bound result of Lemma S2 with $d = K, \mu = \boldsymbol{\mu}_K, \bar{\rho}_{(1)} = \bar{\rho}_{\max}$ to the numerator of R_1' and apply Lemma S3 with $d = K, \mu = \boldsymbol{\mu}_K, \bar{\rho}_{(K)} = \bar{\rho}_{\kappa,\min}$ to both probability terms in the denominator of the ratio R_1', separately. Then, we obtain a similar result as in equation (S4.9) with replacing $\bar{\rho}_{\min}$ by $\bar{\rho}_{\kappa,\min}$.

$$\begin{aligned} R_1' \lesssim \delta(\log K)^{1/2} \exp\Big\{ -\Big[(2\bar{\rho}_{\max}(1-\alpha)\log(K-1))^{1/2} - \mu_*/\rho_{\max}^{1/2}\Big]^2/2 - (1-\alpha)\log(K-1) \\ + \Big[\mu_*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\kappa,\min}\log K)^{1/2}\Big]^2\Big\} \\ + 4\,\bar{\rho}_{\kappa,\min}\log K \exp\Big\{ -(K-1)^{\alpha} + \Big[\mu^*\rho_{\min}^{-1/2} + (2\bar{\rho}_{\kappa,\min}\log K)^{1/2}\Big]^2\Big\}, \end{aligned} \tag{S4.54}$$

where $\mu_* = \min_{2\le j\le N}\{\mu_j\}$ and $\mu^* = \|\mu\|_\infty$. The rest of the proof is same as *Step 2* of the proof of Theorem 2. One can find a set $\widetilde{\mathcal{S}}'$ which takes the same form of $\widetilde{\mathcal{S}}$ defined in equations (S4.10) and (S4.11) by taking

$$b = b_\kappa = \frac{\sqrt{2(1-\alpha)(1-\rho_{\max})}}{\rho_{\max}} + 2\,\frac{\sqrt{2(\kappa-\rho_{\min})}}{\rho_{\min}}.$$

Then fixing an $\alpha \in (0,1)$ such that $a = G_\alpha(\rho_{\min}, \rho_{\max}) > 0$, choosing

$$\mu^* \le \beta \min\Big\{1/b_\kappa, \sqrt{(1-\alpha)(1-\rho_{\max})}/a, \rho_{\min}/a\Big\}\; a\sqrt{\log(K-1)}$$

for sufficiently large K, and letting $C'_{\rho_{\min},\rho_{\max}} = \min\big\{1/b_\kappa, \sqrt{(1-\alpha)(1-\rho_{\max})}/a, \rho_{\min}/a\big\}$ leads to the desired bound of Corollary S1.

S5 Proofs of Propositions

We first introduce some notations that are used in the proof. Recall that for a $N \times N$ matrix A, we denote $\lambda_j(A)$ as its jth eigenvalue, and denote $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ as the minimum and maximum of eigenvalues, respectively. For a matrix A, we define its operator norm as $\|A\| = \{\lambda_{\max}(A^{\mathrm{T}}A)\}^{1/2}$.

S5.1 Proof of the Proposition 1

Now we derive the k-dimensional marginal density function. We denote $\theta^{(k)} = (\theta_1, \ldots, \theta_k)^{\mathrm{T}}$ and $\theta^{(N-k)} = (\theta_{k+1}, \ldots, \theta_N)^{\mathrm{T}}$. We partition Σ_N into appropriate blocks as

$$\Sigma_N = \begin{bmatrix} \Sigma_{k,k} & \Sigma_{k,N-k} \\ \Sigma_{N-k,k} & \Sigma_{N-k,N-k} \end{bmatrix}.$$

We also partition its inverse matrix $\widetilde{\Sigma}_N$,

$$\widetilde{\Sigma}_N = \begin{bmatrix} \widetilde{\Sigma}_{k,k} & \widetilde{\Sigma}_{k,N-k} \\ \widetilde{\Sigma}_{k,N-k} & \widetilde{\Sigma}_{N-k,N-k} \end{bmatrix}.$$

Then the k-dimensional marginal $\widetilde{p}_{k,N}(\theta_1, \ldots, \theta_k)$ is

$$\begin{aligned}
&\left(\frac{1}{2\pi}\right)^{N/2} |\Sigma|^{-1/2} \int_0^\infty \cdots \int_0^\infty \exp\big\{ - \big(\theta^{(k)\mathrm{T}} \widetilde{\Sigma}_{k,k}\, \theta^{(k)} \\
&\quad - 2\theta^{(k)\mathrm{T}}\, \widetilde{\Sigma}_{k,N-k}\, \theta^{(N-k)} + \theta^{(N-k)\mathrm{T}}\, \widetilde{\Sigma}_{N-k,N-k}\, \theta^{(N-k)}\big)/2\big\}\, d\theta^{(N-k)} \\
&= \left(\frac{1}{2\pi}\right)^{k/2} \exp\big\{ - \theta^{(k)\mathrm{T}} \widetilde{\Sigma}_{k,k} \theta^{(k)}/2\big\} \cdot \Pi_{i=1}^k \mathbb{1}_{[0,\infty)}(\theta_i) \left(\frac{1}{2\pi}\right)^{(N-k)/2} \{|\widetilde{\Sigma}_{N-k,N-k}|\}^{-1/2} \\
&\quad \cdot \int_0^\infty \cdots \int_0^\infty \exp\big\{ - \|\widetilde{\Sigma}^{\frac{1}{2}}_{N-k,N-k}\, \big(\theta^{(N-k)} - \Sigma_{N-k,k}\, \Sigma_{k,k}^{-1}\, \theta^{(k)}\big)\|^2/2\big\} d\theta^{(N-k)} \\
&= \left(\frac{1}{2\pi}\right)^{k/2} \exp\{-\theta^{(k)\mathrm{T}} \widetilde{\Sigma}_{k,k}\, \theta^{(k)}/2\}\, \mathbb{P}(\widetilde{X}_{N-k} \le \Sigma_{N-k,k}\, \Sigma_{k,k}^{-1}\, \theta^{(k)}) \cdot \Pi_{i=1}^k \mathbb{1}_{[0,\infty)}(\theta_i).
\end{aligned}$$

where

$$
\begin{aligned}
\widetilde{\Sigma}_{k,k} &= \Sigma_{k,k}^{-1} + \Sigma_{k,k}^{-1}\,\Sigma_{k,N-k}\,\widetilde{\Sigma}_{N-k,N-k}\,\Sigma_{N-k,k}\,\Sigma_{k,k}^{-1},\\
\widetilde{\Sigma}_{k,N-k} &= \Sigma_{k,k}^{-1}\,\Sigma_{k,N-k}\,\widetilde{\Sigma}_{N-k,N-k},\\
\widetilde{\Sigma}_{N-k,N-k} &= (\Sigma_{N-k,N-k} - \Sigma_{N-k,k}\,\Sigma_{k,k}^{-1}\,\Sigma_{k,N-k})^{-1},
\end{aligned}
$$

and $\widetilde{X}_{N-k} \sim \mathcal{N}_{N-k}(\mathbf{0}_{N-k}, \widetilde{\Sigma}_{N-k,N-k}^{-1})$.

S5.2 Proof of Proposition 4

We repeatedly apply Newmann series and Lemma S8 in § S7 to construct the approximation matrix to the posterior scale matrix Σ_N. Under Assumption 2, we have the prior covariance matrix $\Omega_N \in \mathcal{M}(\lambda_0, \alpha, k)$ for some universal constants $\lambda_0, \alpha, k > 0$. Then for any $\epsilon \in (0, \lambda_0/2)$, by choosing $r \geq \log(C/\epsilon)/\alpha$, one can find a r-banded symmetric and positive definite matrix $\Omega_{N,r}$ such that

$$
\|\Omega_N - \Omega_{N,r}\| \leq \epsilon. \tag{S5.1}
$$

Now we let $M = \lambda_{\max}(\Omega_{N,r})$ and $m = \lambda_{\min}(\Omega_{N,r})$. Given (S5.1), we have

$$
\mathrm{I_N}\lambda_0 - \epsilon \leq m \leq M \leq 1/\lambda_0 + \epsilon. \tag{S5.2}
$$

By choosing $\xi = 2/(M+m)$, simple calculation gives $\|\mathrm{I_N} - \xi\,\Omega_{N,r}\| < 1$. We now apply Newmann series to construct a polynomial of $\Omega_{N,r}$ of degree n_1, defined as $\widetilde{\Omega}^{-1} = \xi \sum_{j=0}^{n_1}(I - \xi\,\Omega_{N,r})^j$, for some integer $n_1 > 0$ to be chosen later. Applying Lemma S8 in § S7, we have

$$
\|\Omega_{N,r}^{-1} - \widetilde{\Omega}^{-1}\| \leq \kappa_0^{n_1+1}/(\lambda_0 - \epsilon), \tag{S5.3}
$$

where $\kappa_0 = (M-m)/(M+m)$. Applying Lemma S8 we guarantee $\widetilde{\Omega}^{-1}$ is $(n_1\,r)$-banded and positive definite. Combining results in (S5.2) and (S5.3), we have

$$
\lambda_0/(1+\lambda_0\epsilon) - \kappa_0^{n_1+1}/(\lambda_0-\epsilon) \leq \lambda_{\min}(\widetilde{\Omega}^{-1}) \leq \lambda_{\max}(\widetilde{\Omega}^{-1}) \leq 1/(\lambda_0-\epsilon) + \kappa_0^{n_1+1}/(\lambda_0-\epsilon). \tag{S5.4}
$$

Now we let $\widetilde{\Sigma}^{-1} = \widetilde{\Omega}^{-1} + \Phi^{\mathrm{T}}\Phi$. Under Assumption 1 we have $\widetilde{\Sigma}^{-1}$ is k-banded with $k = \max\{n_1\, r, q\}$. We then define $\tilde{\lambda}_1 = \lambda_{\max}(\widetilde{\Sigma}^{-1})$ and $\tilde{\lambda}_N = \lambda_{\min}(\widetilde{\Sigma}^{-1})$. Thus, given (S5.4), we have

$$C_1\,(n/N) + \lambda_0/(1+\lambda_0\epsilon) - \kappa_0^{n_1+1}/(\lambda_0-\epsilon) \le \tilde{\lambda}_N \le \tilde{\lambda}_1 \le C_2\,(n/N) + 1/(\lambda_0-\epsilon) + \kappa_0^{n_1+1}/(\lambda_0-\epsilon),$$

for constants $0 < C_1 < C_2 < \infty$ in Assumption 2.

We first consider the case where $N/n \to a$ for some constant $a \in (0,1)$, as $n, N \to \infty$. For sufficiently large n, N, we obtain

$$C_1' a + \lambda_0/(1+\lambda_0\epsilon) \le \tilde{\lambda}_N \le \tilde{\lambda}_1 \le C_2'\, a + 1/(\lambda_0-\epsilon), \tag{S5.5}$$

for constants C_1', C_2' satisfying $C_1' < C_1$ and $C_2 < C_2'$.

Secondly, we consider the case where $N/n \to 0$ as $n, N \to \infty$. In this case, n/N dominates in the eigenvalues of $\widetilde{\Sigma}^{-1}$. Thus, for sufficiently large n, N, we have

$$C_1\,(n/N) \le \tilde{\lambda}_N \le \tilde{\lambda}_1 \le C_2\,(n/N). \tag{S5.6}$$

Now we apply Lemma S8 one more time to construct the approximation matrix to the inverse of $\widetilde{\Sigma}^{-1}$. Again, by taking $\gamma = 2/(\tilde{\lambda}_1 + \tilde{\lambda}_N)$, we have $\|\mathrm{I_N} - \gamma\,\widetilde{\Sigma}^{-1}\| < 1$. Now we define $\Sigma' = \gamma\sum_{j=0}^{m_1}(\mathrm{I_N} - \gamma\,\widetilde{\Sigma}^{-1})^j$ for some positive integer m_1. Also, it follows

$$\|\widetilde{\Sigma} - \Sigma'\| \le \tilde{\kappa}^{m_1+1}/\tilde{\lambda}_N, \tag{S5.7}$$

where $\tilde{\kappa} = (\tilde{\lambda}_1 - \tilde{\lambda}_N)/(\tilde{\lambda}_1 + \tilde{\lambda}_N)$. By construction Σ' is $(m_1\, k)$-banded.

Now we estimate $\tilde{\kappa}$. For large enough N, n in the first case, we can upper bound

$$\tilde{\kappa} \le \kappa_1 = \frac{(C_2' - C_1')\, a + 1/(\lambda_0-\epsilon) - \lambda_0/(1+\lambda_0\epsilon)}{(C_2' + C_1')\, a + 1/(\lambda_0-\epsilon) + \lambda_0/(1+\lambda_0\epsilon)}.$$

The inequality holds since the map $x \mapsto (1-x)/(1+x)$ is non-increasing in $x \in (0,1)$. Combing this with the result in (S5.5) and taking $x = \tilde{\lambda}_N/\tilde{\lambda}_1$ leads to the expression of κ_1. Based on (S5.7), we have $\|\widetilde{\Sigma} - \Sigma'\| \le \kappa_1^{m_1+1}/\{C_1' a + \lambda_0/(1+\lambda_0\epsilon)\}$. For N, n in the second case, following a similar line of argument, we have $\|\widetilde{\Sigma} - \Sigma'\| \leq \tilde{\kappa}^{m_1+1} N/(C_1 n)$ with $\tilde{\kappa} = (C_2 - C_1)/(C_2 + C_1)$.

We recall the posterior scale matrix $\Sigma_N = (\Omega_N^{-1} + \Phi^{\mathrm{T}}\Phi)^{-1}$. Then we have

$$\begin{aligned}
\|\Sigma_N - \Sigma'\| &\leq \|\Sigma_N - \widetilde{\Sigma}\| + \|\widetilde{\Sigma} - \Sigma'\| \\
&\leq \|\Sigma_N\|(\|\Omega_N^{-1} - \Omega_{N,r}^{-1}\| + \|\Omega_{N,r}^{-1} - \widetilde{\Omega}^{-1}\|)\|\widetilde{\Sigma}\| + \|\widetilde{\Sigma} - \Sigma'\| \\
&\leq \|\Sigma_N\|\|\widetilde{\Sigma}\|(c_1\,\epsilon + c_2\,\kappa_0^{n_1+1}) + \|\widetilde{\Sigma} - \Sigma'\|
\end{aligned}$$

where $c_1 = \|\Omega^{-1}\|\|\Omega_{N,r}^{-1}\|$ and $c_2 = 1/(\lambda_0 - \epsilon)$. The first inequality follows from the triangular inequality and the second inequality follows from the identity $\|A^{-1} - B^{-1}\| = \|A^{-1}\|\|A - B\|\|B^{-1}\|$ for invertible matrices A, B. The last inequality follows from results in (S5.1) and (S5.3).

For N, n in the first case, $\|\Sigma_N\|$ and $\|\widetilde{\Sigma}\|$ are upper bounded by some constants that are free of n, N given (S5.5). Then we obtain

$$\|\Sigma_N - \Sigma'\| \leq C'(\epsilon + \kappa_0^{n_1+1} + \kappa_1^{m_1+1}),$$

where $C' = \max\{c_1, c_2, C_1'\,a + \lambda_0/(1+\lambda_0\epsilon)\}/\{C_1'\,a + \lambda_0/(1+\lambda_0\epsilon)\}^2$.

For N, n in the second case, for sufficiently large N, n we have $\|\Sigma_N\| \asymp (N/n)$ given (S5.6). Then we have

$$\|\Sigma_N - \Sigma'\| \leq C''\,\{(N/n)^2(\epsilon + \kappa_0^{n_1+1}) + (N/n)\tilde{\kappa}^{m_1+1}\},$$

where $C'' = C_1^{-2}\max\{c_1, c_2, C_1\}$. Letting $\kappa = \max\{\kappa_0, \kappa_1, \tilde{\kappa}\}$, $n_0 = \min\{n_1, m_1\}$, and $\delta_{\epsilon,\kappa} = (\epsilon + \kappa^{n_0+1})\{(N/n)$ for $N \leq n$ yields the result in Proposition 4.

S5.3 Proof of Proposition 6

Given the prior $\theta \sim \mathcal{N}_{\mathcal{C}}(\mathbf{0}_N, \Sigma_N)$, denote by $\Pi_{\mathcal{C}}(\cdot)$ the prior distribution measure. Denote the set by $B_n = \{\|\Phi\theta - f_0\| \leq \eta\sqrt{N}\}$, for some constant $\eta > 0$. Denote by $P_N = \mathcal{N}(\Phi\theta, \mathbf{I_n})$ and $P_0 = \mathcal{N}(f_0, \mathbf{I_n})$, then define the Kullback–Leibler neighborhood of f_0 of radius α_n as $B_{KL}(f_0, \alpha_n) = \{\theta \in \mathbb{R}^N : KL(P_0, P_N) \lesssim n\alpha_n^2, V(P_0, P_N) \lesssim n\alpha_n^2\}$, where the definitions of $KL(P, Q), V(P, Q)$ can be found in the proof of Theorem 8 in §S4.5. Applying Lemma 1 in [5], it suffices to show there exists a sequence $\{\alpha_n\}$ satisfying $\alpha_n \to 0$ and $n\alpha_n^2 \to \infty$ as $n \to \infty$, then

$$\Pi_{\mathcal{C}}(B_n)/\Pi_{\mathcal{C}}(B_{KL}(f_0, \alpha_n)) \leq e^{-2n\alpha_n^2}. \tag{S5.8}$$

Under the same context in § 3 of the manuscript, for the true function $f_0(x) \equiv 0$ for $x \in [0, 1]$, we assume the pseudo-true parameter $\theta_0 = \mathbf{0}_N$ and $[f_0(x_1), \ldots, f_0(x_n)]^{\mathrm{T}} = \Phi\theta_0$, based on a proper basis expansion such as defined in equation (A.1) in Appendix A of the manuscript. Then, for Φ satisfying Assumption 1, one can show that

$$\|\Phi\theta - f_0\| \geq \|\Phi\theta - \Phi\theta_0\| \geq \sqrt{\lambda_{\min}(\Phi^{\mathrm{T}}\Phi)}\|\theta - \theta_0\| \geq \sqrt{C_1 n/N}\|\theta\|,$$

where $C_1 > 0$ is the constant defined in Assumption 1. The preceding implies $\Pi_{\mathcal{C}}(B_n) \leq \Pi_{\mathcal{C}}(\|\theta - \theta_0\| \leq \eta C_1^{-1/2}\sqrt{N^2/n})$, for sufficiently large N.

Let $\alpha_n \asymp (N/n)^{1/2}$ and $n\alpha_n^2 \asymp N$. Then, following a similar argument in the proof of Theorem 8, we have

$$B_{KL}(f_0, \alpha_n) \supset \left\{\theta \in \mathbb{R}^N : \|\theta\| \leq C_2^{-1}\sqrt{N}\alpha_n\right\},$$

where C_2 is the constant defined in Assumption 1.

Then, to show equation (S5.8), it suffices to show that there exists a constant $\eta > 0$ such that

$$\frac{\Pi_{\mathcal{C}}(\|\theta\| < t_n)}{\Pi_{\mathcal{C}}(\|\theta\| < r_n)} = o(e^{-2n\alpha_n^2}), \tag{S5.9}$$

where $r_n = C_2^{-1}Nn^{-1/2}, t_n = \eta C_1^{-1/2}Nn^{-1/2}$. Define $\tilde{\theta} \sim \mathcal{N}(\mathbf{0}_N, \Sigma_N)$ and denote by $\Pi(\cdot)$ its distribution measure, then for any measurable set A one has $\Pi_{\mathcal{C}}(A) = \Pi(A \cap \mathcal{C})/\Pi(\mathcal{C})$. It is easy to see that the above quotients are equivalent to

$$\frac{\Pi(\|\theta\| < t_n, \theta \geq \mathbf{0}_N)}{\Pi(\|\theta\| < r_n, \theta \geq \mathbf{0}_N)}. \tag{S5.10}$$

We first bound the numerator of equation (S5.10) from above. Denote $\lambda_M = \lambda_{\max}(\Sigma)$ and $\lambda_m = \lambda_{\min}(\Sigma)$, by Assumption 2, λ_m, λ_M are bounded from zero and infinity, and do not change along with n, N. Then it is bounded from above by

$$\begin{aligned}\Pi(\|\theta\| < t_n) &\leq \left(\frac{\lambda_M}{\lambda_m}\right)^{N/2} \int_{\|\theta\|<t_n} \frac{1}{(2\pi\lambda_M)^{N/2}} e^{-\frac{\|\theta\|^2}{2\lambda_M}} d\theta \\ &= \left(\frac{\lambda_M}{\lambda_m}\right)^{N/2} \int_0^{t_n^2/\lambda_M} \frac{x^{\frac{N}{2}-1} e^{-\frac{x}{2}}}{2^{N/2}\Gamma(N/2)} dx \\ &\leq \frac{\{t_n^2/(2\lambda_m)\}^{N/2}}{\Gamma(N/2+1)}. \end{aligned} \tag{S5.11}$$

The second line arrives by using the definition of a chi-square random variable with the degree of freedom N.

We now bound the denominator of equation (S5.10) from below. Note that

$$\begin{aligned}\Pi(\|\theta\| < r_n, \theta \geq \mathbf{0}_N) &\geq \Pi(0 < \theta_j < r_n/\sqrt{N}, j = 1, \ldots, N) \\ &\geq \left(\frac{\lambda_m}{\lambda_M}\right)^{N/2} \Pi_{j=1}^N \int_0^{r_n/\sqrt{N}} \frac{1}{\sqrt{2\pi\lambda_m}} e^{-\frac{x_j^2}{2\lambda_m}} dx_j \\ &\geq e^{-r_n^2/(2\lambda_m)} \left(\frac{r_n^2}{2\pi\lambda_M N}\right)^{N/2}. \end{aligned} \tag{S5.12}$$

With results in equations (S5.11) and (S5.12), we now bound the term in equation (S5.10) from above by

$$e^{r_n^2/(2\lambda_m)} \left(\frac{\pi\lambda_M t_n^2}{\lambda_m r_n^2}\right)^{N/2} \frac{N^{N/2}}{\Gamma(N/2+1)} \leq \exp\{r_n^2/(2\lambda_m) - N\log(r_n/t_n)/2\} = o(e^{-2n\alpha_n^2}),$$

the first inequality uses the Stirling's approximation as $\Gamma(N/2+1) \geq \sqrt{eN}\{N/(2e)\}^{N/2}$. Then the final result of the preceding is attained by choosing sufficiently small $\eta > 0$. We then verified equation (S5.9) and complete the proof of Proposition 6.

S5.4 Proof of Proposition S1

($\Rightarrow$). We first show $\mathbb{P}(0 \leq Z_1 < \delta) \lesssim \mathbb{P}(0 \leq Z_1' < \delta)$ for any fixed $\delta > 0$. Define the set $\mathcal{C}_\mu = \{x \in \mathbb{R}^N : x_j \geq -\mu_j, j = 1, \ldots, N\}$ for any fixed vector $\boldsymbol{\mu}_N \in [0, \infty)^N$. For any $\delta > 0$, define $\mathcal{C}_\delta = C_\mu \cap \{0 < x_1 < \delta\}$. Define two N-dimensional random vectors $X \sim \mathcal{N}(\mathbf{0}, \Sigma_N)$ and

$X' \sim \mathcal{N}(\mathbf{0}, \Sigma'_N)$. By definition, we have

$$\mathbb{P}(0 \le Z_1 < \delta) = \frac{\mathbb{P}(X \in \mathcal{C}_\delta)}{\mathbb{P}(X \in \mathcal{C}_\mu)}. \tag{S5.13}$$

First we lower-bound the denominator of the ratio in equation (S5.13). Given $\|\Sigma_N - \Sigma'_N\| \le \varepsilon \|\Sigma_N\|$, for some ε to be chosen later, simple calculation leads to $\Sigma_N^{-1} \le \Sigma_N'^{-1} + \varepsilon \|\Sigma_N'^{-1}\| \mathrm{I_N}$. Then we have

$$\begin{aligned}
\mathbb{P}(X \in \mathcal{C}_\mu) &= \int_{\mathcal{C}_\mu} (2\pi)^{-N/2} |\Sigma_N|^{-1/2} e^{-\frac{1}{2} x^{\mathrm{T}} \Sigma_N^{-1} x} dx \\
&\ge \int_{\mathcal{C}_\mu} (2\pi)^{-N/2} |\Sigma_N|^{-1/2} e^{-\frac{1}{2} x^{\mathrm{T}} (\Sigma_N'^{-1} + \varepsilon \|\Sigma_N'^{-1}\| \mathrm{I_N}) x} dx \\
&\ge e^{-\varepsilon P_N/2} (|\Sigma'_N|/|\Sigma_N|)^{1/2} \mathbb{P}(X' \in \mathcal{C}_\mu \cap \mathcal{A}),
\end{aligned} \tag{S5.14}$$

where $\mathcal{A} = \{X' : \|X'\|^2 \le P_N / \|\Sigma_N'^{-1}\|\}$ for some $P_N \gtrsim N$ to be chosen later. Next, to bound $\mathbb{P}(X' \in \mathcal{C}_\mu \cap \mathcal{A})$, we shall show that there exist $P_N, \varepsilon > 0$ such that $\mathbb{P}(X' \in \mathcal{A}^c) < \mathbb{P}(X' \in \mathcal{C}_\mu)$. Then there exist some $\eta \in (0,1)$ such that $\mathbb{P}(X' \in \mathcal{C}_\mu \cap \mathcal{A}) \ge \mathbb{P}(X' \in \mathcal{C}_\mu) - \mathbb{P}(X' \in \mathcal{A}^c) \ge (1-\eta)\mathbb{P}(X' \in \mathcal{C}_\mu)$. To that end, we bound $\mathbb{P}(X' \in \mathcal{A}^c)$ from above and bound $\mathbb{P}(X' \in \mathcal{C}_\mu)$ from below separately.

We first bound $\mathbb{P}(X' \in \mathcal{A}^c)$. Note that

$$\begin{aligned}
\mathbb{P}(X' \in \mathcal{A}^c) &= \mathbb{P}(\|X'\|^2 > P_N / \|\Sigma_N'^{-1}\|) \le \mathbb{P}(\|\Sigma'_N\| \|Z_0\|^2 > P_N / \|\Sigma_N'^{-1}\|) \\
&\le \mathbb{P}(\|Z_0\|^2 > c\, P_N\}) \lesssim e^{-c_1 P_N},
\end{aligned} \tag{S5.15}$$

where $Z_0 \sim \mathcal{N}(\mathbf{0}_N, \mathrm{I_N})$ and $c, c_1 > 0$ are some constants. Note that $\|Z_0\|^2$ is a chi-square random variable with the degree of freedom N, the last inequality is obtained by applying the concentration result of a N-dimensional chi-square random vector.

Next we bound $\mathbb{P}(X' \in \mathcal{C}_\mu)$ from below. Since $\Sigma'_N = (\rho'_{ij})$ is a K-banded matrix and under the assumption that the entries of Σ'_N are positive within the band, we adopt the same *block approximation* technique used in the proof of Theorem 2. Define the random vector $X'' \sim \mathcal{N}(\mathbf{0}_N, \Sigma'_N(\rho'_{\min}))$ where $\Sigma'_N(\rho'_{\min}) = (\rho''_{ij})$ satisfies $\rho''_{ii} = \rho'_{ii}$ for $1 \le i \le N$ and $\rho''_{ij} = \rho'_{\min}$ for all i, j satisfying $1 \le |i-j| \le K-1$. Denote $\sigma'^2_{(N)} = \max_{1 \le i \le N}\{\rho'_{ii}\}$ and $\sigma'^2_{(1)} = \min_{1 \le i \le N}\{\rho'_{ii}\}$. Then by applying Slepian's inequality, we have $\mathbb{P}(X' \in \mathcal{C}_\mu) \ge \mathbb{P}(X'' \in \mathcal{C}_\mu)$. Now let $m = [N/K] > 1$. Define the random vector Z'' such that (i), $Z''_{[(1+(i-1)K):iK]} \overset{d}{=} X''_{[(1+(i-1)K):iK]}$ for $i = 1, \ldots, m$

and $Z''_{[(1+sK):N]} \overset{d}{=} X''_{[(1+sK):N]}$. And (ii), the sub-vectors $\{Z''_{[(1+(i-1)K):iK]}, i = 1, \ldots, m\}$ and $Z''_{[(1+mK):N]}$ are mutually independent. Then we have $\mathbb{E}\{(X''_i)^2\} = \mathbb{E}\{(Z''_i)^2\}$ for $1 \leq i \leq N$ and $\mathbb{E}(X''_i X''_j) \geq \mathbb{E}(Z''_i Z''_j)$ for all $1 \leq i \neq j \leq N$. For sufficiently large N, K, we have

$$\begin{aligned}\mathbb{P}(X' \in \mathcal{C}_\mu) &\geq \mathbb{P}(X'' \in \mathcal{C}_\mu) \geq \mathbb{P}(Z'' \geq \mu^* \mathbf{1}_N) \\ &= \Pi_{i=1}^m \{\mathbb{P}(Z''_{[(1+(i-1)K):iK]} \geq \mu^* \mathbf{1}_K)\} \mathbb{P}(Z''_{[(1+mK):N]} \geq \mu^* \mathbf{1}_{N-mK}) \\ &\gtrsim (\log K)^{-(m+1)/2} K^{-(m+1)\bar{\rho}'_{(N)}} > \exp(-c_2 N \log K), \end{aligned} \tag{S5.16}$$

for sufficiently large K, N, where $\bar{\rho}'_{(N)} = (\sigma'^2_{(N)} - \rho'_{\min})/\rho'_{\min}$, $\mu^* = \max_{1 \leq i \leq N} |\mu_i|$ is a finite and positive constant under the assumption, and $c_2 > 0$ is some constant. The first and second inequalities hold by applying Slepian's inequality, the third inequality is obtained by applying Lemma S3 by taking $d = K$, $a = \mu^*$, $\rho = \rho'_{\min}$ and $\bar{\rho}_{(K)} = \bar{\rho}'_{(N)}$, which leads to $\mathbb{P}(Z''_{[(1+(i-1)K):iK]} \geq \mu^* \mathbf{1}_K) \gtrsim (\log K)^{-1/2} K^{-\bar{\rho}'_{(N)}}$, for $i = 1, \ldots, s$. A same lower bound is obtained for $\mathbb{P}(Z''_{[(1+mK):N]} \geq \mu^* \mathbf{1}_{N-mK})$. Comparing the result in equation (S5.16) with result in equation (S5.15), it is obvious that choosing $P_N > c_2 N \log K$ leads to $\mathbb{P}(X' \in \mathcal{A}^c) < \eta \mathbb{P}(X' \in \mathcal{C}_\mu)$ for some constant $\eta \in (0, 1)$ and for any positive integer N. Combining this result with equations (S5.14) and (S5.16), we obtain

$$\mathbb{P}(X \in \mathcal{C}_\mu) \geq e^{-\varepsilon P_N/2} (|\Sigma'_N|/|\Sigma_N|)^{1/2} (1-\eta) \mathbb{P}(X' \in \mathcal{C}_\mu), \tag{S5.17}$$

for some constant $\eta \in (0, 1)$.

Now we bound the numerator of the ratio in equation (S5.13). Similarly, we have

$$\begin{aligned}\mathbb{P}(X \in \mathcal{C}_\delta) &\leq \int_{\mathcal{C}_\delta} (2\pi)^{-N/2} |\Sigma_N|^{-1/2} e^{-\frac{1}{2} x^{\mathrm{T}} \Sigma'^{-1}_N x} e^{\varepsilon \|\Sigma'^{-1}_N\| \|x\|^2/2} dx \\ &\leq e^{\varepsilon P_N/2} (|\Sigma'_N|/|\Sigma_N|)^{1/2} \{\mathbb{P}(X' \in \mathcal{C}_\delta) + \mathbb{P}(X' \in \mathcal{A}^c)\}. \end{aligned} \tag{S5.18}$$

Next, we shall show $\mathbb{P}(X' \in \mathcal{A}^c) < \mathbb{P}(X' \in \mathcal{C}_\delta)$. To proceed, we consider the defined random vector

Z'' above, since $(\Sigma''_N(\rho'_{\min}))_{ij} \le (\Sigma'_N)_{ij}$ for all $i \neq j$, then applying Lemma S1 leads to

$$
\begin{aligned}
\mathbb{P}(X' \in \mathcal{C}_\delta) \ge \mathbb{P}(Z'' \in \mathcal{C}_\delta) &\ge \mathbb{P}(-\mu_1 \le Z''_1 \le \delta - \mu_1, Z''_2 \ge -\mu_2, \ldots, Z''_N \ge -\mu_N) \\
&\ge \mathbb{P}(-\mu_1 \le Z''_1 \le \delta - \mu_1, Z''_2 \ge \mu^*, \ldots, Z''_K \ge \mu^*) \\
&\quad \cdot \Pi_{i=2}^m \{\mathbb{P}(Z''_{[(1+(i-1)K):iK]} \ge \mu^* \mathbf{1}_K)\} \mathbb{P}(Z''_{[(1+mK):N]} \ge \mu^* \mathbf{1}_{N-mK}) \\
&\gtrsim (\log K)^{-(m+1)/2} (K-1)^{-(1+\alpha)\sigma'^2_{(1)}/\rho'_{\min}} K^{-m\bar{\rho}'_{(N)}} \ge \exp(-c_3 N \log K),
\end{aligned}
$$

where $\bar{\rho}'_{(1)} = (\sigma'^2_{(1)} - \rho'_{\min})/\rho'_{\min}$, $\alpha \in (0,1)$ and $c_3 > 0$ are some constants. The third inequality holds based on an application of the lower bound of Lemma S2 with $d = K$, $\rho = \rho'_{\min}$ and $\bar{\rho}_{(1)} = \bar{\rho}'_{(1)}$ to bound $\mathbb{P}(-\mu_1 \le Z''_1 \le \delta - \mu_1, Z''_2 \ge \mu^*, \ldots, Z''_K \ge \mu^*)$ from below and a same lower bound for $\mathbb{P}(Z''_{[(K+1):N]} \ge \mu^* \mathbf{1}_{N-K})$ as in deriving equation (S5.16). Again, it is easy to see that by choosing $P_N \gtrsim N \log K$ we have $\mathbb{P}(X' \in \mathcal{A}^c) < \mathbb{P}(X' \in \mathcal{C}_\delta)$. With this result and equation (S5.18), we obtain the upper bound for the numerator of the ratio in equation (S5.14) as

$$
\mathbb{P}(X \in \mathcal{C}_\delta) \le 2e^{\varepsilon P_N/2} (|\Sigma'_N|/|\Sigma_N|)^{1/2} \mathbb{P}(X' \in \mathcal{C}_\delta). \tag{S5.19}
$$

Now combining results in equations (S5.17) and (S5.19), by choosing $\varepsilon \le c P_N^{-1}$ for some constant $c > 0$, we obtain

$$
\frac{\mathbb{P}(X \in \mathcal{C}_\delta)}{\mathbb{P}(X \in \mathcal{C}_\mu)} \le 2(1-\eta) e^c \frac{\mathbb{P}(X' \in \mathcal{C}_\delta)}{\mathbb{P}(X' \in \mathcal{C}_\mu)}, \tag{S5.20}
$$

for some $\eta \in (0,1)$.

($\Leftarrow$). We now follow a similar line of arguments to show $\mathbb{P}(0 \le Z_1 < \delta) \gtrsim \mathbb{P}(0 \le Z'_1 < \delta)$. It suffices to lower bound the ratio in equation (S5.13). We first bound the denominator of the ratio in equation (S5.13) from above. With a similar calculation as in equation (S5.18), one can show that

$$
\begin{aligned}
\mathbb{P}(X \in \mathcal{C}_\mu) &\le e^{\varepsilon P_N/2} (|\Sigma'_N|/|\Sigma_N|)^{1/2} \{\mathbb{P}(X' \in \mathcal{C}_\mu) + \mathbb{P}(X' \in \mathcal{A}^c)\} \\
&\le 2e^{\varepsilon P_N/2} (|\Sigma'_N|/|\Sigma_N|)^{1/2} \mathbb{P}(X' \in \mathcal{C}_\mu).
\end{aligned} \tag{S5.21}
$$

The second inequality adopts the fact that $\mathbb{P}(X' \in \mathcal{A}^c) < \mathbb{P}(X' \in \mathcal{C}_\mu)$. Now to lower bound the numerator, with a similar calculation in equation (S5.14), we have

$$
\begin{aligned}
\mathbb{P}(X \in \mathcal{C}_\delta) &\geq e^{-\varepsilon P_N/2}(|\Sigma_N'|/|\Sigma_N|)^{1/2}\{\mathbb{P}(X' \in \mathcal{C}_\delta) - \mathbb{P}(X' \in \mathcal{A}^c)\} \\
&\geq (1-\eta')e^{-\varepsilon P_N/2}(|\Sigma_N'|/|\Sigma_N|)^{1/2}\mathbb{P}(X' \in \mathcal{C}_\delta),
\end{aligned} \tag{S5.22}
$$

for some constant $\eta' \in (0,1)$. Again, the second inequality uses the result that $\mathbb{P}(X' \in \mathcal{A}^c) < \mathbb{P}(X' \in \mathcal{C}_\delta)$. Then there exist some $\eta' \in (0,1)$ such that $\mathbb{P}(X' \in \mathcal{C}_\delta) - \mathbb{P}(X' \in \mathcal{A}^c) \geq (1-\eta')\mathbb{P}(X' \in \mathcal{C}_\delta)$. Then combining results of equations (S5.21) and (S5.22) and by choosing $\varepsilon \leq cP_N^{-1}$ for some constant $c > 0$,

$$
\frac{\mathbb{P}(X \in \mathcal{C}_\delta)}{\mathbb{P}(X \in \mathcal{C}_\mu)} \geq \{2(1-\eta')e^c\}^{-1}\frac{\mathbb{P}(X' \in \mathcal{C}_\delta)}{\mathbb{P}(X' \in \mathcal{C}_\mu)}. \tag{S5.23}
$$

Combining results in equations (S5.20) and (S5.23), we complete the proof of Proposition S1.

S5.5 Proof of Proposition S2

First note that the scale matrix Ω_N defined in § 4.1 is a correlation matrix. Also it is easy to verify that Ω_N satisfies Assumption 2 that $\Omega_N \in \mathcal{M}(\alpha_0, \lambda_0, k)$ for some constants $\alpha_0, \lambda_0 > 0$. Given any fixed τ, λ, recall the unconstrained parameter $\theta' \mid \lambda \sim \mathcal{N}(\mathbf{0}_N, \tau^2\Lambda\Omega_N\Lambda)$. Next, by applying Proposition 4, one can construct a K-banded approximating matrix Ω_N' such that $\|\Omega_N - \Omega_N'\| \leq \varepsilon\|\Omega_N\|$ by choosing $K \gtrsim \log(1/\varepsilon)$. Here we choose $K = O(N)$ and then the condition $\varepsilon \lesssim (N\log K)^{-1}$ is satisfied. Since $\|\Omega_N\| = O(1)$ under Assumption 2, we have $\|\Omega_N - \Omega_N'\| \leq c'(N\log N)^{-1}$ for some constant $c' > 0$. Define $\theta'' \mid \lambda \sim \mathcal{N}(\mathbf{0}_N, \tau^2\Lambda\Omega_N'\Lambda)$, then by Proposition S1, one can show that for any fixed $\lambda > \mathbf{0}_N$, there exists some constant $C' > 0$ such that

$$
\mathbb{P}(\theta' \geq \mathbf{0}_N \mid \lambda) \geq C'\,\mathbb{P}(\theta'' \geq \mathbf{0}_N \mid \lambda).
$$

We now lower bound $\mathbb{P}(\theta'' \geq \mathbf{0}_N \mid \lambda)$ for any fixed $\lambda > \mathbf{0}_N$. To that end, we adopt the *block approximation* technique in the proof of Theorem 2. First recall that we assume $\omega_{ii}' = \sigma'^2$ for $1 \leq i \leq N$ and $\omega_{\min}'$ denotes the smallest off-diagonal positive elements of Ω_N'. Then define $\widetilde{\Omega}_N' = (\tilde{\omega}_{ij}')$ with $\tilde{\omega}_{ii}' = \sigma'^2$ for $1 \leq i \leq N$ and $\tilde{\omega}_{ij}' = \omega_{\min}'$ for all i, j satisfying $1 < |i-j| < K$. And we assume $\omega_{\min}' < \sigma'^2$. Then define the random vector $\tilde{\theta} \sim \mathcal{N}(\mathbf{0}_N, \tau^2\Lambda\widetilde{\Omega}_N'\Lambda)$ given the same fixed λ, we have $\mathbb{E}(\tilde{\theta}_i\tilde{\theta}_j) \le \mathbb{E}(\theta_i''\theta_j'')$ for all $1 \le i \ne j \le N$. We now construct a new random vector using the *block approximation* technique. Recall that $K = O(N)$ and we denote $m_0 = [N/K] > 1$. Now define the random vector $\tilde{\theta}'$ such that $\tilde{\theta}'_{[(1+iK):(1+i)K]} \overset{d}{=} \tilde{\theta}_{[(1+iK):(1+i)K]}$ for $i = 0, \ldots, m_0 - 1$ and $\tilde{\theta}'_{[(1+m_0K):N]} \overset{d}{=} \tilde{\theta}_{[(1+m_0K):N]}$. And also the sub-vectors $\{\tilde{\theta}'_{[(1+iK):(1+i)K]}, i = 0, \ldots, m_0 - 1\}$ and $\tilde{\theta}'_{[(1+m_0K):N]}$ are mutually independent. Then we have $\mathbb{E}(\tilde{\theta}_i'^2) = \mathbb{E}(\tilde{\theta}_i^2)$ for $1 \le i \le N$ and $\mathbb{E}(\tilde{\theta}_i'\tilde{\theta}_j') \le \mathbb{E}(\tilde{\theta}_i\tilde{\theta}_j)$ for $1 \le i \ne j \le N$. Then applying Lemma S4,

$$\begin{aligned}\mathcal{M}_{\mathcal{C}} &= \mathbb{E}_\lambda\{\Pi(\theta \ge \mathbf{0}_N \mid \lambda)\} \ge \mathbb{E}_\lambda\{\Pi(\tilde{\theta} \ge \mathbf{0}_N \mid \lambda)\} \ge \mathbb{E}_\lambda\{\Pi(\tilde{\theta}' \ge \mathbf{0}_N \mid \lambda)\} \\ &= \mathbb{E}_\lambda\big\{\Pi_{i=0}^{m_0-1}\Pi\big(\tilde{\theta}'_{[(1+iK):(1+i)K]} \ge \mathbf{0}_K \mid \lambda\big)\Pi\big(\tilde{\theta}'_{[(1+m_0K):N]} \ge \mathbf{0}_{N-m_0K} \mid \lambda\big)\big\}. \end{aligned} \tag{S5.24}$$

The second equality holds by the definition of $\tilde{\theta}'$. To proceed, we first bound $\mathbb{P}\big(\tilde{\theta}'_{[1:K]} \ge \mathbf{0}_K \mid \lambda\big)$ from below, similar arguments can be applied to bounding the rest of probability terms of (S5.24). Note that $\mathrm{var}(\tilde{\theta}_j') = \tau^2\lambda_j^2\sigma'^2$ and $\mathrm{cov}(\tilde{\theta}_i', \tilde{\theta}_j') = \tau^2\lambda_i\lambda_j\omega'_{\min}$ for all $1 \le i, j \le K$. Then we use the equivalent expression of $\tilde{\theta}'$ as

$$\tilde{\theta}_i' \overset{d}{=} \tau\lambda_i\left(\sqrt{1-\omega'_{\min}}\,w_i + \sqrt{\omega'_{\min}}\,W\right), \quad i = 1, \ldots, K,$$

where $\{w_i\}, W$ are independently and identically distributed standard normal random variables. For any fixed $\lambda, \tau > \mathbf{0}$, we have

$$\begin{aligned}\Pi(\tilde{\theta}_1' \ge 0, \ldots, \tilde{\theta}_K' \ge 0 \mid \lambda) &= \Pi\Big\{\tau\lambda_i\Big(\sqrt{1-\omega'_{\min}}w_i + \sqrt{\omega'_{\min}}W\Big) \ge 0, i = 1, \ldots, K \mid \lambda\Big\} \\ &= \Pi\Big(\sqrt{1-\omega'_{\min}}w_i + \sqrt{\omega'_{\min}}W \ge 0, i = 1, \ldots, K\Big) \\ &\gtrsim (\log K)^{-1/2}K^{-\bar{\omega}'_{\min}},\end{aligned}$$

where $\bar{\omega}'_{\min} = (\sigma'^2 - \omega'_{\min})/\omega'_{\min}$. The last inequality holds by applying Lemma S3. Following similar arguments, we obtain $\mathbb{P}(\tilde{\theta}_{[(1+iK):(1+i)K]} \ge \mathbf{0}_K \mid \lambda) \gtrsim (\log K)^{-1/2}K^{-\bar{\omega}'_{\min}}$ for $i = 1, \ldots, m_0 - 1$ and $\mathbb{P}(\tilde{\theta}_{[(1+m_0K):N]} \ge \mathbf{0}_{N-m_0K} \mid \lambda) \ge (\log K)^{-1/2}K^{-\bar{\omega}'_{\min}}$. Combining these results with (S5.24), then we have

$$\mathcal{M}_{\mathcal{C}} \gtrsim (\log K)^{-(m_0+1)/2}K^{-(m_0+1)\bar{\omega}'_{\min}}.$$

Since the above lower bound is decreasing in K and m_0 is a constant, substitute $K = N/m_0$ in the above lower bound leads to the desired result $\mathcal{M}_{\mathcal{C}} \gtrsim (\log N)^{-(m_0+1)/2} N^{-(m_0+1)\bar{\omega}'_{\min}}$. By taking $t_0 = m_0 + 1$ completes the proof.

S6 Proofs of technical results

S6.1 Proofs in Appendix A

Proof of Lemma 1. For any function $f \in C[0,1]$ and f' is Lipschitz, then there exists some finite constant $L > 0$ such that $|f'(x) - f'(y)| \le L|x - y|$ for any $x, y \in [0,1]$ and $x \neq y$. Now denote its expansion with respect to basis expansion (M) under a grid $\{u_j\}$ by $f_N(x)$. Given the definition of f_N and applying the fundamental theorem of calculus to f, we have for any $x \in [0,1]$,

$$
\begin{aligned}
|f_N(x) - f(x)| &= \left| \{f(0) + \int_0^x \sum_{j=0}^{N-1} f'(u_j) h_j(s)\, ds\} - \{f(0) + \int_0^x f'(s)\, ds\} \right| \\
&\le \int_0^x \left| \sum_{j=0}^{N-1} f'(u_j) h_j(s) - f'(s) \sum_{j=0}^{N-1} h_j(s) \right| ds \\
&\le \int_0^x \sum_{j=0}^{N-1} |f'(u_j) - f'(s)| h_j(s)\, ds \le \delta_N L.
\end{aligned}
$$

The second inequality holds for the facts $|f'(u_j) - f'(s)| \le L|s - u_j|$ and $h_j(s) \neq 0$ for $|s - u_j| \le \delta_N$ for any $1 \le j \le N$. The inequality holds uniformly for any $x \in [0,1]$, therefore we complete the proof.

Proof of Lemma 2. We first show the upper bound. For any nonempty set $S \in \{1, \ldots, N\}$, recall Ψ_S is a $n \times |S|$ sub-matrix of Ψ with columns $\Psi_j, j \in S$ and denote an arbitrary $|S|$-dimensional vector by $\theta_S = \{\theta_j \in \mathbb{R}, j \in S\}$. Note that $\Psi_S \theta_S$ is a $n \times 1$ vector, we have

$$
\|\Psi_S \theta_S\|^2 = \sum_{i=1}^n \left(\sum_{j=1}^{|S|} \Psi_{ij} \theta_j \right)^2 \le \sum_{i=1}^n \left(\sum_{j=1}^{|S|} \Psi_{ij}^2 \sum_{j=1}^{|S|} \theta_j^2 \right) \le n \delta_N \|\theta_S\|^2 \approx (n/N) \|\theta_S\|^2,
$$

for sufficiently large N. The second inequality is obtained by applying Cauchy-Schwarz inequality, and the third inequality holds by the fact that $\|\Psi\|_\infty \le \delta_N$ by construction, and $\delta_N = 1/(N-1)$.

Now we prove the lower bound. Note that under the assumption $\min_{ij}|x_i - u_j| \geq c\delta_N^{3/2}$, we have $\Psi_{ij} \geq 2c\,\delta_N^{1/2}$ for all $1 \leq i, j \leq N$. Then

$$\|\Psi_S\theta_S\|^2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{|S|}\Psi_{ij}\theta_j\right)^2 \geq \sum_{i=1}^{n}(2c\,\delta_N^{1/2})^2\|\theta_S\|^2 \approx 4c^2(n/N)\|\theta_S\|^2,$$

for a fixed constant $c > 0$ and for sufficiently large N. Then we complete the proof.

Proof of Lemma 3. For any $0 \leq a < b \leq 1$, determine the largest integer $0 \leq j_1' \leq N-1$ such that $u_{j_1'} \leq a$ and the smallest integer $0 \leq j_S' \leq N-1$ such that $b \leq u_{j_S'}$. Then the shortest interval contains $[a, b]$ is $[u_{j_1'}, u_{j_S'}]$. By restricting the function to be non-decreasing, one has $\theta_j \geq 0$ for $j = 1, \ldots, N$. Given the construction in (M), the flatness of f over the interval $[a, b]$ is equivalent to

$$f'(x) = \sum_{l=j_1'}^{j_S'}\theta_{l+1}h_l(x) = 0,$$

for $x \in [a, b]$. It implies $\theta_{l+1} = 0$ for all $l = j_1', \ldots, j_S'$. Then we complete the proof.

S6.2 Proof of Lemma S1

We first prove the result for centered multivariate normal vectors. For random vectors $X \sim \mathcal{N}(\mathbf{0}, \Sigma_X)$ and $Y \sim \mathcal{N}(\mathbf{0}, \Sigma_Y)$, to show $\mathbb{P}(\ell_1 \leq X_1 \leq u_1, X_2 \geq u_2, \ldots, X_d \geq u_d) \leq \mathbb{P}(\ell_1 \leq Y_1 \leq u_1, Y_2 \geq u_2, \ldots, Y_d \geq u_d)$, it suffices to show

$$\begin{aligned}&\mathbb{P}(Y_1 \geq u_1, Y_2 \geq u_2, \ldots, Y_d \geq u_d) - \mathbb{P}(X_1 \geq u_1, X_2 \geq u_2, \ldots, X_d \geq u_d)\\ \leq\, &\mathbb{P}(Y_1 \geq \ell_1, Y_2 \geq u_2, \ldots, Y_d \geq u_d) - \mathbb{P}(X_1 \geq \ell_1, X_2 \geq u_2, \ldots, X_d \geq u_d).\end{aligned} \tag{S6.1}$$

We define d-dimensional indicator functions $G(x) = \mathbb{1}_{[u_1,\infty)}(x_1)\prod_{j=2}^{d}\mathbb{1}_{(u_j,\infty)}(x_j)$ and $F(x) = \mathbb{1}_{[\ell_1,\infty)}(x_1)\prod_{j=2}^{d}\mathbb{1}_{(u_j,\infty)}(x_j)$, then it is equivalent to show

$$\mathbb{E}\{G(Y)\} - \mathbb{E}\{G(X)\} \leq \mathbb{E}\{F(Y)\} - \mathbb{E}\{F(X)\}. \tag{S6.2}$$

We now construct non-decreasing approximating functions of G, F with continuous second order derivatives respectively. Let $\nu \in C^2(\mathbb{R})$ be a non-decreasing twice differentiable function with

$\nu(t) = 0$ for $t \le 0$, $\nu(t) \in [0,1]$ for $t \in [0,1]$, and $\nu(t) = 1$ for $t \ge 1$. Also, choose ν so that $\|\nu'\|_\infty < C$ for some universal constant $C > 0$. For $\eta > 0$, we define $m_\eta(x) = \nu(\eta x)$. It is clear that $m_\eta(x)$ approximates $\mathbb{1}_{[0,\infty)}(x)$ for large η. In fact, for any $x \neq 0$, $\lim_{\eta\to\infty} m_\eta(x) = \mathbb{1}_{[0,\infty)}(x)$.

Given the above, let $g_j^\eta(x_j) = \nu\{\eta(x_j - u_j)\}$ for $j = 1,\dots,d$, and $f_1^\eta = \nu\{\eta(x - \ell_1)\}$, $f_j^\eta = \nu\{\eta(x_j - u_j)\}$ for $j = 2,\dots,d$. Define

$$g^\eta(x) = \Pi_{j=1}^d g_j^\eta(x_j) \quad \text{and} \quad f^\eta(x) = \Pi_{j=1}^d f_j^\eta(x_j).$$

It then follows that g^η and f^η provide increasingly better approximations of G and F as $\eta \to \infty$. It thus suffices to show

$$\mathbb{E}\{g^\eta(Y)\} - \mathbb{E}\{g^\eta(X)\} \le \mathbb{E}\{f^\eta(Y)\} - \mathbb{E}\{f^\eta(X)\}, \tag{S6.3}$$

for sufficiently large $\eta > 0$ to be chosen later. We henceforth drop the superscript η from g and f for notation brevity.

We proceed to utilize an interpolation technique commonly used to prove comparison inequalities (see Chapter 7 of [16]). We construct a sequence of interpolating random variables based on the independent random variables X, Y:

$$S_t = (1-t^2)^{1/2}X + tY, \quad t \in [0,1].$$

Specifically, we have $S_0 = X$, $S_1 = Y$, and for any $t \in [0,1]$, $S_t \sim \mathcal{N}(\mathbf{0}, \widetilde{\Sigma}_t)$ where $\widetilde{\Sigma}_t = (1-t^2)\Sigma_X + t^2\Sigma_Y$. For any twice differentiable function h, we have the following identity

$$\mathbb{E}\{h(Y)\} - \mathbb{E}\{h(X)\} = \int_0^1 \frac{d}{dt}\mathbb{E}\{h(S_t)\}\, dt. \tag{S6.4}$$

Applying a multivariate version of Stein's lemma (Lemma 7.2.7 in [16]) to the integrand in equation (S6.4), one obtains

$$\frac{d}{dt}\mathbb{E}\{h(S_t)\} = t\sum_{i,j=1}^d \mathbb{E}\left[\{\mathbb{E}(Y_iY_j) - \mathbb{E}(X_iX_j)\}\frac{\partial^2 h}{\partial x_i \partial x_j}(S_t)\right]. \tag{S6.5}$$

To show equation (S6.3), we define the difference $\Delta = [\mathbb{E}\{f(Y)\} - \mathbb{E}\{f(X)\}] - [\mathbb{E}\{g(Y)\} - \mathbb{E}\{g(X)\}]$. We further decompose Δ as

$$\begin{aligned}
\Delta &= [\mathbb{E}\{f(Y)\} - \mathbb{E}\{f(X)\}] - [\mathbb{E}\{g(Y)\} - \mathbb{E}\{g(X)\}] \\
&= \int_0^1 dt \left\{ \frac{d}{dt}\mathbb{E}\{f(S_t)\} - \frac{d}{dt}\mathbb{E}\{g(S_t)\} \right\} \\
&= \int_0^1 dt \left\{ t \sum_{i,j=1}^{d} \mathbb{E}\bigg[\{\mathbb{E}(Y_iY_j) - \mathbb{E}(X_iX_j)\} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(S_t) - \frac{\partial^2 g}{\partial x_i \partial x_j}(S_t) \right) \bigg] \right\} \\
&= 2\int_0^1 dt \left\{ t \sum_{j=2}^{d} \mathbb{E}\bigg[\{\mathbb{E}(Y_1Y_j) - \mathbb{E}(X_1X_j)\} \left(\frac{\partial^2 f}{\partial x_1 \partial x_j}(S_t) - \frac{\partial^2 g}{\partial x_1 \partial x_j}(S_t) \right) \bigg] \right\} \\
&\quad + \int_0^1 dt \left\{ t \sum_{i,j=2}^{d} \mathbb{E}\bigg[\{\mathbb{E}(Y_iY_j) - \mathbb{E}(X_iX_j)\} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(S_t) - \frac{\partial^2 g}{\partial x_i \partial x_j}(S_t) \right) \bigg] \right\} \\
&= \Delta_1 + \Delta_2.
\end{aligned}$$

The second equation follows from (S6.4) and the third equation follows from (S6.5). First we show $\Delta_1 \geq 0$. Since $\mathbb{E}(Y_1Y_j) \geq \mathbb{E}(X_1X_j)$ for all $j > 1$, it suffices to show that for any fixed $t \in [0,1]$ and for any $j = 2, \ldots, d$,

$$D_1 = \mathbb{E}\left(\frac{\partial^2 f}{\partial x_1 \partial x_j}(S_t) - \frac{\partial^2 g}{\partial x_1 \partial x_j}(S_t) \right) \geq 0.$$

We consider a generic interpolating random variable $S \sim \mathcal{N}\big(\mathbf{0}, \widetilde{\Sigma}\big)$ by dropping the t-subscript; let $\phi(s_1, \ldots, s_d)$ denote its probability density function. Then we have

$$\begin{aligned}
D_1 &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{f_1'(s_1)f_j'(s_j) - g_1'(s_1)g_j'(s_j)\}\, \Pi_{l\neq 1,j} f_l(s_l)\, \phi(s_1, \ldots, s_d)\, ds_1 \ldots ds_d \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \{f_1'(s_1) - g_1'(s_1)\}\, \phi(s_1, \ldots, s_N)\, ds_1 \right] f_j'(s_j)\, \Pi_{l\neq 1,j} f_l(s_l)\, ds_2 \ldots ds_d.
\end{aligned}$$

To guarantee D_1 is non-negative we need the integral over s_1 to be non-negative. Based on the definition of f_1 and g_1, the integral over s_1 can be simplified to

$$\begin{aligned}
&\int_{-\infty}^{\infty} \{f_1'(s_1) - g_1'(s_1)\}\, \phi(s_1, \ldots, s_N)\, ds_1 \\
&= \int_{\ell_1}^{\ell_1+1/\eta} \big\{\eta\, \nu'\big(\eta(s_1 - \ell_1)\big)\big\}\, \phi(s_1, \ldots, s_N)\, ds_1 - \int_{u_1}^{u_1+1/\eta} \big\{\eta\, \nu'\big(\eta(s_1 - u_1)\big)\big\}\, \phi(s_1, \ldots, s_N)\, ds_1 \\
&= \int_0^{1/\eta} \eta\, \nu'(\eta s_1)\{\phi(s_1 + \ell_1, s_2, \ldots, s_N) - \phi(s_1 + u_1, s_2, \ldots, s_N)\} ds_1. \qquad \text{(S6.6)}
\end{aligned}$$

Let us denote the inverse of the covariance matrix $\widetilde{\Sigma}$ as

$$\widetilde{\Sigma}^{-1} = \begin{bmatrix} \widetilde{\Sigma}_{11}^{-1} & \widetilde{\Sigma}_{12}^{-1} \\ \widetilde{\Sigma}_{21}^{-1} & \widetilde{\Sigma}_{22}^{-1} \end{bmatrix},$$

where $\widetilde{\Sigma}_{11}^{-1}$ is a scalar. To check the non-negativity of the last line in equation (S6.6), we now estimate the term

$$\frac{\phi(s_1 + \ell_1, s_2, \ldots, s_d)}{\phi(s_1 + u_1, s_2, \ldots, s_d)} = e^{\{(u_1^2 - \ell_1^2) + 2 s_1\, (u_1 - \ell_1)\}\, \widetilde{\Sigma}_{11}^{-1}/2 + (u_1 - \ell_1)\, \widetilde{\Sigma}_{12}^{-1}\, \tilde{s}_2},$$

where $\tilde{s}_2 = (s_2, \ldots, s_d)^{\mathrm{T}}$. Since $s_j \in [0, 1/\eta]$, we have $s_1\, (u_1 - \ell_1)\}\, \widetilde{\Sigma}_{11}^{-1} > 0$. We denote $\tilde{\rho} = \max\{\widetilde{\Sigma}_{12}^{-1}\}$ as the largest element of $\widetilde{\Sigma}_{12}^{-1}$. Then, one can choose η large enough such that

$$(u_1 + \ell_1)\, \widetilde{\Sigma}_{11}^{-1} - 2(d-1)\tilde{\rho}/\eta \geq 0,$$

to guarantee $D_1 \geq 0$. For example $\eta = 4(d-1)\, \tilde{\rho}\, \widetilde{\Sigma}_{11}/(u_1 + \ell_1)$ satisfies the above inequality.

Now we show $\Delta_2 \geq 0$. We have $\mathbb{E}(Y_i Y_j) \geq \mathbb{E}(X_i X_j)$ for all $i, j = 2, \ldots, d$. For any $i, j \geq 2$, for any fixed $t \in [0, 1]$, we define

$$D_2 = \mathbb{E}\bigg(\frac{\partial^2 f}{\partial x_i \partial x_j}(S_t) - \frac{\partial^2 g}{\partial x_i \partial x_j}(S_t)\bigg) = \mathbb{E}\{(f_1 - g_1) f_i'\, g_j'\, \Pi_{k \neq 1, i, j} f_k\}.$$

Since $f_1 - g_1 \geq 0$, and $f_j' \geq 0$ for all $j > 1$, it follows that $D_2 \geq 0$ and thus $\Delta_2 \geq 0$. Combining with the non-negativity of Δ_1 completes the proof for centered case.

Now we consider the non-centered multivariate normal vectors. Consider $X \sim \mathcal{N}(\mu, \Sigma_X)$ and $Y \sim \mathcal{N}(\mu, \Sigma_Y)$. Define $\widetilde{X} := X - \mu$ and $\widetilde{Y} := Y - \mu$. Let $\tilde{\ell_1} = \ell_1 - \mu_1$, and $u_j = u_j - \mu_j$ for

$j = 1, \ldots, d$. Then it is equivalent to show $\mathbb{P}(\tilde{\ell}_1 \le \widetilde{X}_1 \le \tilde{u}_1, \widetilde{X}_2 \ge \tilde{u}_2, \ldots, \widetilde{X}_d \ge \tilde{u}_d) \le \mathbb{P}(\tilde{\ell}_1 \le \widetilde{Y}_1 \le \tilde{u}_1, \widetilde{Y}_2 \ge \tilde{u}_2, \ldots, \widetilde{Y}_d \ge \tilde{u}_d)$, which has been proved in the centered case. Thus we complete the proof of Lemma S1.

S6.3 Proof of Lemma S2

Recall $X \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma_d(\sigma^2, \rho))$, where μ is a non-negative fixed mean vector and $\Sigma_d(\sigma^2, \rho) = (\sigma_{ij})$ denotes a variance-correlation matrix with $\sigma_{ii} > 0$ for $1 \le i \le N$ and $\sigma_{ij} = \rho$ for $1 \le i \ne j \le d$. Recall $\sigma^2_{(1)} = \min_{1 \le i \le d} \rho_{ii}$, and under the assumption we have $\rho < \sigma^2_{(1)}$.

Define $X' := X - \boldsymbol{\mu}$, we will repeatedly use its equivalent expression

$$X'_i = \rho^{1/2}\, w + (\sigma_{ii} - \rho)^{1/2}\, W_i, \quad i = 1, \ldots, N, \tag{S6.7}$$

where w, W_i's are independent standard normal variables.

Proof of the upper bound. Recall $\bar{\rho}_{(1)} = (\sigma^2_{(1)} - \rho)/\rho$. For any fixed $\delta > 0$, and for $\alpha \in (0,1)$ we have

$$\begin{aligned}
&\mathbb{P}(0 \le X_1 < \delta, X_2 \ge 0, \ldots, X_d \ge 0) \\
&= \mathbb{P}(0 \le X'_1 + \mu_1 < \delta, X'_2 + \mu_2 \ge 0, \ldots, X'_d + \mu_d \ge 0) \\
&= \mathbb{P}\Big(-\mu_1 \le \rho^{1/2}\, w + (\sigma_{11} - \rho)^{1/2}\, W_1 \le \delta - \mu_1, \rho^{1/2}\, w \ge \max_{2 \le i \le d} (\sigma_{ii} - \rho)^{1/2}\, W_i - \min_{2 \le i \le d} \mu_i \Big) \\
&= \mathbb{P}\Big(\Big\{ -\mu_1 \le \rho^{1/2}\, w + (\sigma_{11} - \rho)^{1/2}\, W_1 \le \delta - \mu_1, w \ge \bar{\rho}_{(1)}^{1/2} \max_{2 \le i \le d} W_i - \rho^{-1/2} \min_{2 \le i \le d} \mu_i \Big\} \\
&\qquad \cup \Big[\max_{2 \le i \le d} W_i \ge \{2(1-\alpha)\log(d-1)\}^{1/2} \Big] \cup \Big[\max_{2 \le i \le d} W_i \le \{2(1-\alpha)\log(d-1)\}^{1/2} \Big] \Big) \\
&\le \mathbb{P}\Big[-\mu_1 \le \rho^{1/2}\, w + (\sigma_{11} - \rho)^{1/2}\, W_1 \le \delta - \mu_1, w \ge \big\{ 2\, \bar{\rho}_{(1)}\, (1-\alpha) \log(d-1) \big\}^{1/2} - \rho^{-1/2} \min_{2 \le i \le d} \mu_i \Big] \\
&\qquad + \mathbb{P}\Big[\max_{2 \le i \le d} W_i \le \{2(1-\alpha)\log(d-1)\}^{1/2} \Big] \\
&= P_1 + P_2.
\end{aligned} \tag{S6.8}$$

First, we estimate P_1 in (S6.8). For a sufficiently large d and fixed μ_i's, it is easy to see that

$\mu_i < \{2(1-\alpha)\log(d-1)\}^{1/2}$ for $i = 1, \ldots, d$. Then applying the expression of X in (S6.7), we have

$$\begin{aligned}
P_1 = \mathbb{P}\Big[& W_1 \in \Big\{ -\mu_1/(\sigma_{11}-\rho)^{1/2} - \Big(\frac{\rho}{\sigma_{11}-\rho}\Big)^{1/2} w, (\delta-\mu_1)/(\sigma_{11}-\rho)^{1/2} - \Big(\frac{\rho}{\sigma_{11}-\rho}\Big)^{1/2} w \Big\} \,\Big| \\
& \qquad w \geq \big\{2\,\bar{\rho}_{(1)}\,(1-\alpha)\log(d-1)\big\}^{1/2} - \rho^{-1/2} \min_{2\leq i\leq d} \mu_i \Big] \\
& \cdot \mathbb{P}\Big[w \geq \big\{2\,\bar{\rho}_{(1)}\,(1-\alpha)\log(d-1)\big\}^{1/2} - \rho^{-1/2} \min_{2\leq i\leq d} \mu_i \Big] \\
\leq \mathbb{P}\Big[& W_1 \in \Big(-\mu_1(\sigma_{11}-\rho)^{-1/2} - \Big[\{2(1-\alpha)\log(d-1)\}^{1/2} - (\sigma_{11}-\rho)^{-1/2} \min_{2\leq i\leq d} \mu_i\Big], \\
& \qquad (\delta-\mu_1)(\sigma_{11}-\rho)^{-1/2} - \Big[\{2(1-\alpha)\log(d-1)\}^{1/2} - (\sigma_{11}-\rho)^{-1/2} \min_{2\leq i\leq d} \mu_i\Big] \Big) \Big] \\
& \cdot \mathbb{P}\Big(w \geq \big\{2\,\bar{\rho}_{(1)}\,(1-\alpha)\log(d-1)\big\}^{1/2} - \rho^{-1/2} \min_{2\leq i\leq d} \mu_i \Big) \\
\leq \delta & \{2\pi(\sigma_{11}-\rho)\}^{-1/2} \exp\Big(- \Big[(\delta + \min_{2\leq i\leq d} \mu_i - \mu_1)(\sigma_{11}-\rho)^{-1/2} - \{2(1-\alpha)\log(d-1)\}^{1/2} \Big]^2 / 2 \Big) \\
& \cdot \Big[\{2\,\bar{\rho}_{(1)}\,(1-\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2} \min_{2\leq i\leq d} \mu_i \Big]^{-1} \\
& \cdot \exp\Big(- \big[2\,\bar{\rho}_{(1)}\,\{(1-\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2} \min_{2\leq i\leq d} \mu_i \big]^2 / 2 \Big).
\end{aligned} \tag{S6.9}$$

The last inequality follows from Lemma S7 in § S7.

Now we move to estimate the term P_2 in (S6.8). We have,

$$\begin{aligned}
\mathbb{P}\Big[\max_{2\leq i\leq d} W_i \leq \{2\,(1-\alpha)\log(d-1)\}^{1/2} \Big] &= \big(1 - \mathbb{P}\big[Z \geq \{2\,(1-\alpha)\,\log(d-1)\}^{1/2} \big] \big)^{d-1} \\
&\leq \exp\big(-(d-1)\,\mathbb{P}\big[Z \geq \{2\,(1-\alpha)\,\log(d-1)\}^{1/2} \big] \big) \\
&\leq \exp(-(d-1)^{\,\alpha}),
\end{aligned} \tag{S6.10}$$

where $Z \sim \mathcal{N}(0,1)$. Then combining bound results in (S6.9) and (S6.10), for sufficiently large d we attain the desired upper bound

$$
\begin{aligned}
&\mathbb{P}(0 \le X_1 < \delta, X_2 \ge 0, \dots, X_d \ge 0) \\
&\quad \le C\delta \Big\{ \log(d-1) - \rho^{-1/2} \min_{2\le i\le d} \mu_i \Big\}^{-1/2} \\
&\qquad \cdot \exp\Big(- \Big[\{2(1-\alpha)\log(d-1)\}^{1/2} - (\sigma_{11}-\rho)^{-1/2}\big(\delta + \min_{2\le i\le d}\mu_i - \mu_1\big)\Big]^2/2 \\
&\qquad\qquad - \Big[2\,\bar\rho_{(1)}\,\{(1-\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2} \min_{2\le i\le d}\mu_i \Big]^2/2\Big) + \exp(-(d-1)^{\,\alpha}), \qquad \text{(S6.11)}
\end{aligned}
$$

where $C = \{4\pi(\sigma_{11}-\rho)\bar\rho_{(1)}\,(1-\alpha)\}^{-1/2}$ for some $\alpha \in (0,1)$.

Proof of the lower bound. The lower bound is derived in a similar manner, thus we omit details and only state the different steps. Using the expression (S6.7), we arrive at

$$
\begin{aligned}
&\mathbb{P}(0 \le X_1 < \delta, X_2 \ge 0, \dots, X_d \ge 0) \\
&\ge \mathbb{P}\Big(\Big\{ -\mu_1 \le \rho^{1/2}\, w + (\sigma_{11}-\rho)^{1/2}\, W_1 \le \delta - \mu_1, w \ge \bar\rho_{(1)}^{-1/2} \max_{2\le i\le d} W_i - \rho^{-1/2}\min_{2\le i\le d}\mu_i \Big\} \\
&\qquad \cap \Big[\max_{2\le i\le d} W_i \le \{2(1+\alpha)\log(d-1)\}^{1/2} \Big]\Big) \\
&\ge \mathbb{P}\Big[-\mu_1 \le \rho^{1/2}\, w + (\sigma_{11}-\rho)^{1/2}\, W_1 \le \delta - \mu_1, w \ge \big\{2\,\bar\rho_{(1)}\,(1+\alpha)\log(d-1)\big\}^{1/2} - \rho^{-1/2}\min_{2\le i\le d}\mu_i \Big] \\
&\quad \cdot \mathbb{P}\Big[\max_{2\le i\le d} W_i \le \{2(1+\alpha)\log(d-1)\}^{1/2}\Big] =: P_1' \cdot P_2'.
\end{aligned}
$$

Following a similar argument, we can bound P_1' by

$$
\begin{aligned}
P_1' &\ge \delta\{2\pi(\sigma_{11}-\rho)\}^{-1/2} \exp\Big\{ - \Big[\{2(1+\alpha)\log(d-1)\}^{1/2} - (\sigma_{11}-\rho)^{-1/2}\big(\min_{2\le i\le d}\mu_i - \mu_1\big)\Big]^2/2 \Big\} \\
&\quad \cdot \Big[\{2\,\bar\rho_{(1)}\,(1+\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2}\min_{2\le i\le d}\mu_i\Big]^{-1} \\
&\quad \cdot \exp\Big(- \Big[2\,\bar\rho_{(1)}\,\{(1+\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2}\min_{2\le i\le d}\mu_i\Big]^2/2\Big), \qquad \text{(S6.12)}
\end{aligned}
$$

for some $\alpha \in (0,1)$. By using the lower bound in Lemma S7,

$$
\begin{aligned}
P_2' &= \big(1 - \mathbb{P}\big[Z \ge \{2\,(1+\alpha)\,\log(d-1)\}^{1/2}\big]\big)^{d-1} \ge \exp\big(-(d-1)\,\mathbb{P}\big[Z \ge \{2\,(1+\alpha)\,\log(d-1)\}^{1/2}\big]\big) \\
&\ge \exp(-(d-1)^{\,-\alpha}) \ge 1/2, \qquad \text{(S6.13)}
\end{aligned}
$$

for sufficiently large d and for some $\alpha \in (0,1)$. Combining (S6.12) and (S6.13) leads to

$$\begin{aligned}
&\mathbb{P}(0 \le X_1 < \delta, X_2 \ge 0, \ldots, X_d \ge 0) \\
&\quad \ge C'\delta \{\log(d-1)\}^{-1/2} \exp\Big\{ - \Big[\{2(1+\alpha)\log(d-1)\}^{1/2} - (\sigma_{11}-\rho)^{-1/2}\big(\min_{2\le i\le d}\mu_i - \mu_1\big)\Big]^2/2 \\
&\qquad\qquad - \Big[2\,\bar{\rho}_{(1)}\,\{(1+\alpha)\log(d-1)\}^{1/2} - \rho^{-1/2}\min_{2\le i\le d}\mu_i\Big]^2/2\Big\}, \quad \text{(S6.14)}
\end{aligned}$$

where $C' = \{8\pi(\sigma_{11}-\rho)\bar{\rho}_{(1)}\,(1+\alpha)\}^{-1/2}$. By combining (S6.11) and (S6.14) yields the sandwich bound in Lemma S2.

S6.4 Proof of Lemma S3

Recall $X' = X - \boldsymbol{\mu}$, then it suffices to lower-bound $\mathbb{P}(X' \ge \mu^* \mathbf{1}_d)$. Recall that $\bar{\rho}_{(d)} = (\sigma^2_{(d)} - \rho)/\rho$. We now show that for any scalar $a \ge 0$, we have

$$\mathbb{P}(X' \ge a\mathbf{1}_d) \ge \frac{a\rho^{-1/2} + (2\,\bar{\rho}_{(d)}\,\log N)^{1/2}}{\big\{a\rho^{-1/2} + (2\,\bar{\rho}_{(d)}\,\log d)^{1/2}\big\}^2 + 1} \exp\Big[-\frac{1}{2}\Big\{a\rho^{-1/2} + (2\,\bar{\rho}_{(d)}\,\log d)^{1/2}\Big\}^2\Big], \quad \text{(S6.15)}$$

where recall that $\mathbf{1}_d$ denotes a N-dimensional vector of ones. By taking $a = \max_{1\le i\le d}\{\mu_i\}$ leads to the desired lower bound.

Now we prove the lower bound in (S6.15). First,

$$\begin{aligned}
\mathbb{P}(X \ge a\mathbf{1}_d) &= \mathbb{P}\big(\rho^{1/2}\,w + (\sigma_{ii}-\rho)^{1/2}\,W_i \ge a, \text{ for } i = 1,\ldots,d\big) \qquad \text{(S6.16)} \\
&= \mathbb{E}\big(\,\mathbb{P}\big[w \ge \rho^{-1/2}\{a - (\sigma_{ii}-\rho)^{1/2}\,W_i\},\, i = 1,\ldots,d \mid W_1,\ldots,W_d\big]\big) \\
&\overset{\text{(i)}}{=} \mathbb{E}\big\{\,\mathbb{P}\big(w \ge \rho^{-1/2}\big[a + \max_i\{(\sigma_{ii}-\rho)^{1/2}W_i\}\big] \mid W_1,\ldots,W_N\big)\big\} \\
&= \mathbb{E}\Big\{1 - \Phi\Big(a\rho^{-1/2} + \bar{\rho}_{(d)}^{1/2}\max_i W_i\Big)\Big\},
\end{aligned}$$

where $W = [W_1,\ldots,W_d]^{\mathrm{T}}$. Here, (i) holds since $-W_i \overset{\mathrm{d}}{=} W_i$ for $i = 1,\ldots,d$ and $\max_{1\le i\le d}(-W_i) \overset{\mathrm{d}}{=} \max_{1\le i\le d}(W_i)$.

We now proceed to lower bound the right hand side of the last equation in (S6.16). To that end, we define $g(a,b) = 1 - \Phi(a\rho^{-1/2} + \bar{\rho}_{(d)}^{1/2}\,b)$, where $g : \mathbb{R}_+ \times \mathbb{R} \to [0,1]$. Importantly, g is non-increasing function of a, b for $a, b \in \mathbb{R}$, and g is a convex function of (a,b) for $a, b > 0$. For any fixed $a > 0$, since $g(a, \max_i W_i)$ is non-increasing in $\max_i W_i$, we have $g(a, \max_i W_i) \ge g(a, \max_i |W_i|)$. We then apply Jensen's inequality,

$$\mathbb{E}\Big\{g\Big(a, \max_{1\le i\le d}|W_i|\Big)\Big\} \ge g\Big\{a, \mathbb{E}\Big(\max_{1\le i\le d}|W_i|\Big)\Big\} \ge g\big\{a, (2\log d)^{1/2}\big\}.$$

The last inequality holds by applying Lemma S6 in § S7. To lower bound $g\big\{a, (2\log d)^{1/2}\big\}$ we apply Lemma S7 in § S7. Eventually, we obtain

$$\mathbb{E}\left\{g\Big(a, \max_{1\le i\le d}|W_i|\Big)\right\} \ge \frac{a\rho^{-1/2} + (2\bar{\rho}_{(d)}\log d)^{1/2}}{\big\{a\rho^{-1/2} + (2\bar{\rho}_{(d)}\log d)^{1/2}\big\}^2 + 1}\exp\big[-\big\{a\rho^{-1/2} + (2\bar{\rho}_{(d)}\log d)^{1/2}\big\}^2/2\big]. \tag{S6.17}$$

Taking $a = \mu^*$ completes the proof.

S7 Auxiliary results

S7.1 Technical results

Lemma S4. *(Slepian's lemma) Let X, Y be centered Gaussian vectors on $\mathbb{R}^d$. Suppose $\mathbb{E}X_i^2 = \mathbb{E}Y_i^2$ for all i, and $\mathbb{E}(X_iX_j) \le \mathbb{E}(Z_iZ_j)$ for all $i \ne j$. Then, for any $x \in \mathbb{R}$,*

$$\mathbb{P}\Big(\max_{1\le i\le d}X_i \le x\Big) \le \mathbb{P}\Big(\max_{1\le i\le d}Y_i \le x\Big).$$

We use the Slepian's lemma in the following way in the main document. We have,

$$\mathbb{P}(X_1 \ge 0, \ldots, X_d \ge 0) = \mathbb{P}\Big(\min_{1\le i\le d}X_i \ge 0\Big) = \mathbb{P}\Big(\max_{1\le i\le d}X_i \le 0\Big),$$

where the second equality uses $X \overset{d}{=} -X$. We use Slepian's inequality to arrive at equation (S4.2) in the main document.

Lemma S5. *(Lemma 3.1 in [14].) Let $X \sim \mathcal{N}(0, \Sigma_1)$ and $Y \sim \mathcal{N}(0, \Sigma_2)$. For any set $\mathcal{D} \in \mathbb{R}^N$, if $\Sigma_1 - \Sigma_2$ is positive semi-definite, then*

$$P(Y \in \mathcal{D}) \le (|\Sigma_1|/|\Sigma_2|)^{1/2}P(X \in \mathcal{D}).$$

Lemma S6. *Let $Z_1, \dots, Z_N$ be iid $\mathcal{N}(0,1)$ random variables. Then we have*

$$C_1\sqrt{2\log N} \le \mathbb{E} \max_{i=1,\dots,N} Z_i \le \mathbb{E} \max_{i=1,\dots,N} |Z_i| \le \sqrt{2\log N}. \tag{S7.1}$$

for some constant $0 < C_1 < 1$.

Lemma S7. *(Mill's ratio bound) Let $X \sim \mathcal{N}(0,1)$. We have, for $x > 0$, that*

$$\frac{x}{x^2+1}e^{-x^2/2} \le 1 - \Phi(x) \le \frac{1}{x}e^{-x^2/2},$$

where $\Phi(\cdot)$ is cumulative distribution function of X.

Lemma S8. *(Lemma 2.1 in [2]) Let matrix A be k-banded, symmetric, and positive definite. We denote $M = \|A\|$ and $m = 1/\|A^{-1}\|$, and for $n \in \mathbb{N}_0$, we define*

$$B_n = \gamma \sum_{j=0}^{n} (I - \gamma A)^j, \tag{S7.2}$$

where $\gamma = 2/(M+m)$. Then B_n is a symmetric positive definite (nk)-banded matrix, also, $\|A^{-1} - B_n\| \le \kappa^{n+1}/m$, $\kappa = (M-m)/(M+m) < 1$.

S7.2 Correlation $(\rho_{\min}, \rho_{\max})$ in Figure S1

Table S1: Values of $(\rho_{\min}, \rho_{\max})$ for (N, K) considered in Figure S1

	N	K	$\rho_{\min}$	$\rho_{\max}$
Case I	100	2	0.43	0.48
	100	5	0.295	0.45
	100	20	0.154	0.23
Case II	10	5	0.447	0.68
	50	5	0.30	0.46
	100	5	0.295	0.45
Case III	25	5	0.327	0.50
	100	20	0.154	0.23
	250	50	0.057	0.08

S8 Additional details on the numerical studies

S8.1 Prior draws

We consider equation (4.1) and the prior specified in § 4. Prior samples on both θ and ξ of dimension $N = 100$ were drawn. Figure S8 shows prior draws for the first and third components of both θ and ξ.

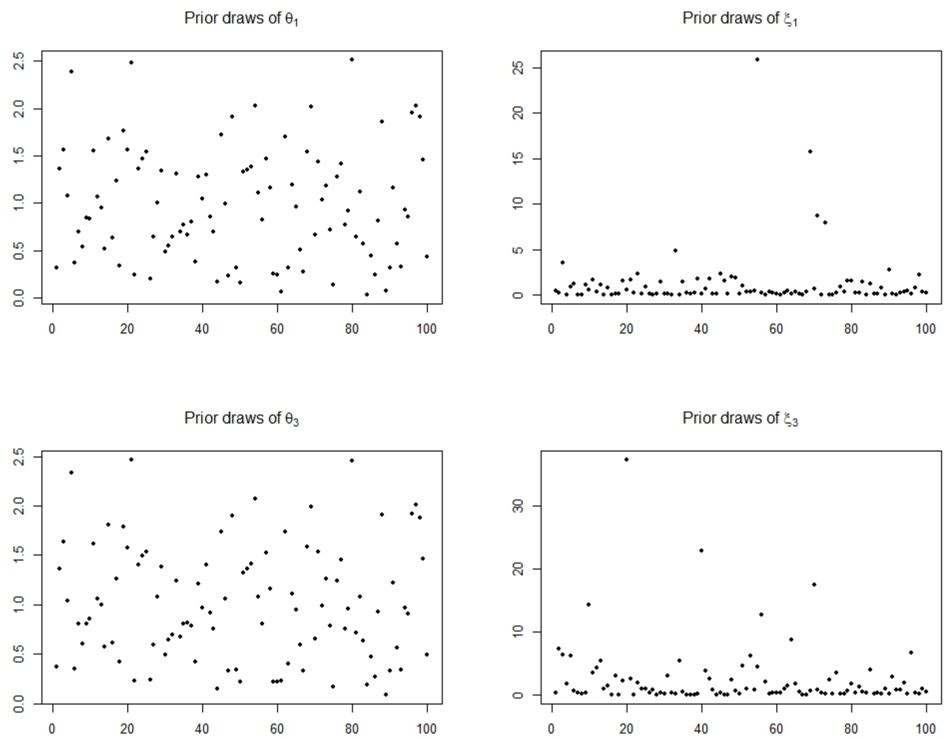

Figure S8: *Showing prior draws from distribution of θ (left panel) and ξ (right panel). Top and bottom panels correspond to first and third components respectively, for both θ and ξ.*

S8.2 Posterior Computations

We now consider model (4.2) and the prior specified in § 4. Then the full conditional distribution of θ

$$\pi(\theta \mid Y, \zeta, \lambda, \tau, \sigma) \propto \exp\left\{ -\frac{1}{2\sigma^2}\|\widetilde{Y} - \Psi\Lambda\theta\|^2 \right\} \exp\left\{ -\frac{1}{2\tau^2}\theta^{\mathrm{T}} K^{-1}\theta \right\} \mathbb{1}_{\mathcal{C}_\theta}(\theta)$$

can be approximated by

$$\pi(\theta \mid Y, \zeta, \lambda, \tau, \sigma) \propto \exp\left\{-\frac{1}{2\sigma^2}\|\widetilde{Y} - \Psi_\lambda\,\theta\|^2\right\} \exp\left\{-\frac{1}{2\tau^2}\theta^{\mathrm{T}} K^{-1}\theta\right\} \left\{\prod_{j=1}^{N+1} \frac{e^{\eta_\theta\,\theta_j}}{1+e^{\eta_\theta\,\theta_j}}\right\}$$
$$= \left[\exp\left\{-\frac{1}{2\sigma^2}\|\widetilde{Y} - \Psi_\lambda\,\theta\|^2\right\} \left\{\prod_{j=1}^{N+1} \frac{e^{\eta_\theta\,\theta_j}}{1+e^{\eta_\theta\,\theta_j}}\right\}\right] \exp\left\{-\frac{1}{2\tau^2}\theta^{\mathrm{T}} K^{-1}\theta\right\}$$

where η_θ is a large valued constant, $\widetilde{Y} = Y - \zeta 1_n$ and $\Psi_\lambda = \Psi\Lambda$. The above is same as equation (5) of [13] and thus falls under the framework of their sampling scheme. For more details on the sampling scheme and the approximation, one can refer to [13].

Note that $\lambda_j \sim \mathcal{C}_+(0,1),\ j = 1,\ldots,N$, can be equivalently given by $\lambda_j \mid w_j \sim \mathcal{N}(0, w_j^{-1})\mathbb{1}(\lambda_j > 0)\,,\ w_j \sim \mathcal{G}(0.5, 0.5)\,,\ j = 1,\ldots,N$. Thus the full conditional distribution of λ can be approximated by:

$$\pi(\lambda \mid Y, \zeta, \theta, w, \tau, \sigma) \propto \left[\exp\left\{-\frac{1}{2\sigma^2}\|\widetilde{Y} - \Psi_\theta\,\lambda\|^2\right\} \left\{\prod_{j=1}^{N+1} \frac{e^{\eta_\lambda\,\lambda_j}}{1+e^{\eta_\lambda\,\lambda_j}}\right\}\right] \exp\left\{-\frac{1}{2}\lambda^{\mathrm{T}} W\lambda\right\}$$

where η_λ plays the same role as η_θ, $w = (w_1,\ldots,w_N)^{\mathrm{T}}$, $W = \mathrm{diag}(w_1,\ldots,w_N)$, $\Psi_\theta = \Psi\Theta$ and $\Theta = \mathrm{diag}(\theta_1,\ldots,\theta_N)$. Thus, λ can be sampled efficiently using algorithm proposed in [13].

S8.3 Additional numerical studies and plots of model fits

In this section, we provide details on model comparisons and additional plots discussed in § 4.3 of the main documents. First, we discuss the improvement due to the shrinkage. Specifically, the variants of tMVN model considered are:

- **tMVN *with fixed hyperparameters:*** We set $\Lambda = \mathrm{I}_N$ and $\tau = 1$ in (4.2), and also fix ν and ℓ, so that we have a truncated normal prior on the coefficients. This was implemented as a part of the motivating examples in the introduction. We fix $\nu = 0.75$ and ℓ so that the correlation $k(1)$ between the maximum separated points in the covariate domain equals 0.05.

- **tMVN *with hyperparameter updates:*** The only difference from the previous case is that ν and ℓ are both assigned priors described previously and updated within the MCMC algorithm.

- **tMVN *with global shrinkage:*** We continue with $\Lambda = \mathrm{I_N}$ and place a half-Cauchy prior on the global shrinkage parameter τ. The hyperparameters ν and ℓ are updated.

- **DGL-tMVN:** This is the proposed procedure where the λ_js are also assigned half-Cauchy priors and the hyperparameters are updated.

Figures S9, S10, S11 and S12 display the model fits of functions f_1, f_2, f_3 and f_4 respectively based the on four variants of tMVN priors discussed above. The figures suggest that the tMVN prior with fixed hyperparameters leads to a large bias in the flat region. Adding some global structure to it, for instance, by updating the GP hyperparamaters and adding a global shrinkage term improves prediction around the flat region. However it still lacked the flexibility to transition from the flat region to the strictly increasing region. By including component-wise local parameters additionally (i.e. the DGL-tMVN) improves the overall prediction and performs the best, both visually and also in terms of MSPE.

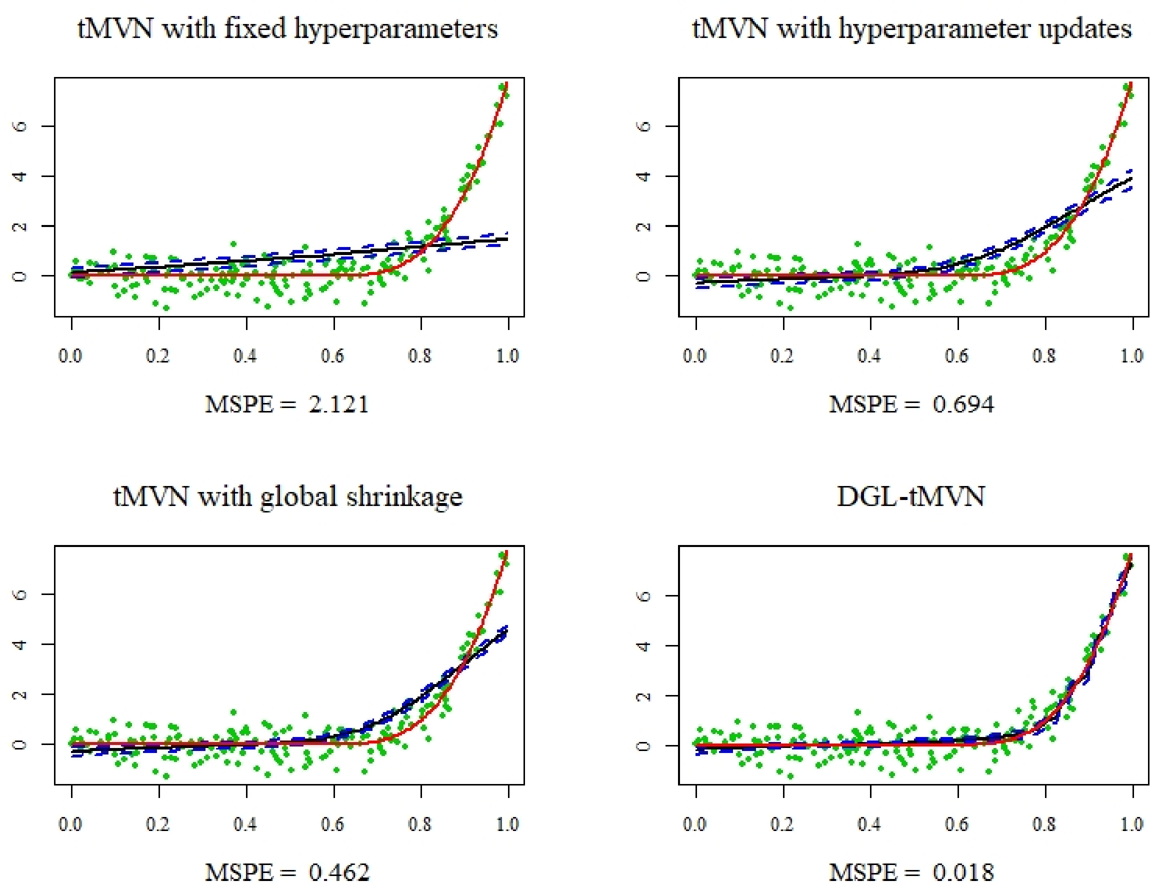

Figure S9: *Out-of-sample prediction accuracy for f_1 using the four variants. Red solid curve corresponds to the true function, black solid curve is the mean prediction, the region within two dotted blue curves represent 95% pointwise prediction interval and the green dots are* 200 *test data points.* MSPE *values corresponding to each of the method are also shown in the plots.*

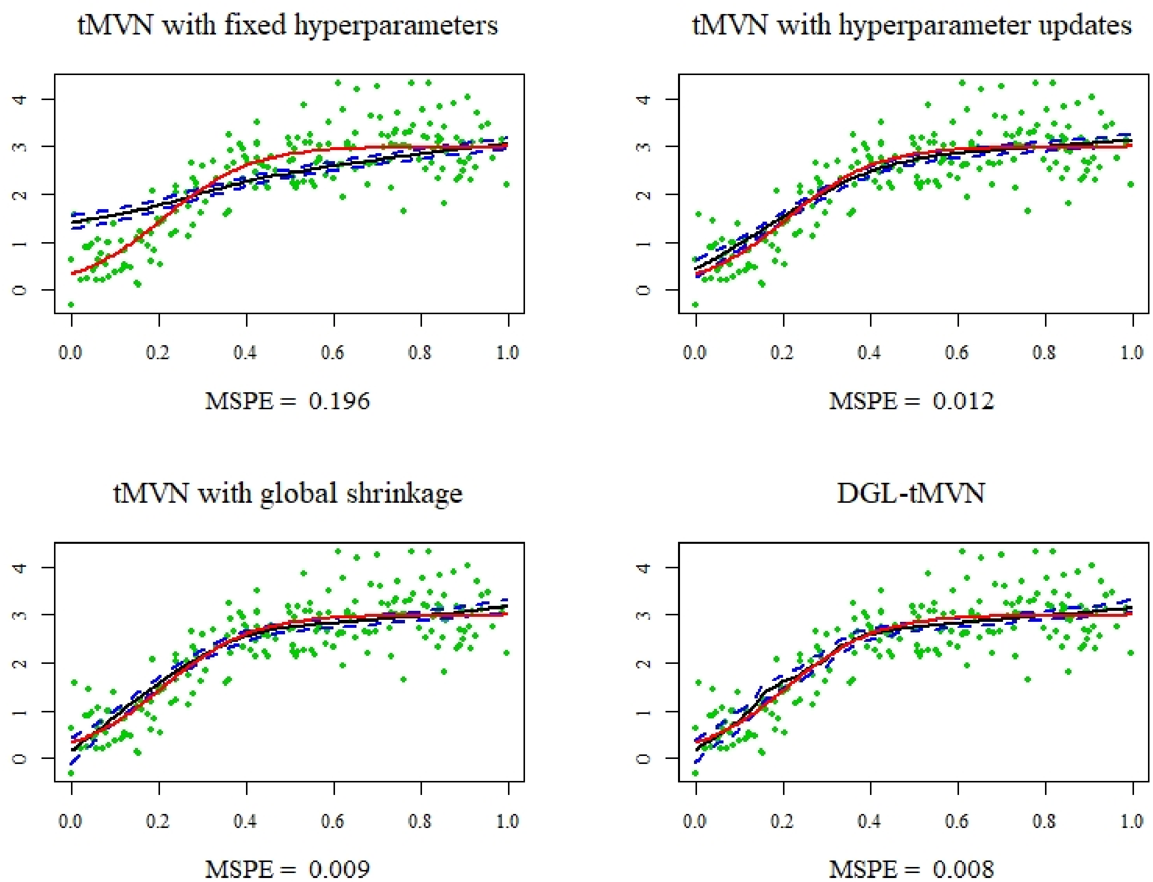

Figure S10: *Same as Figure S9, now for the function* f_2.

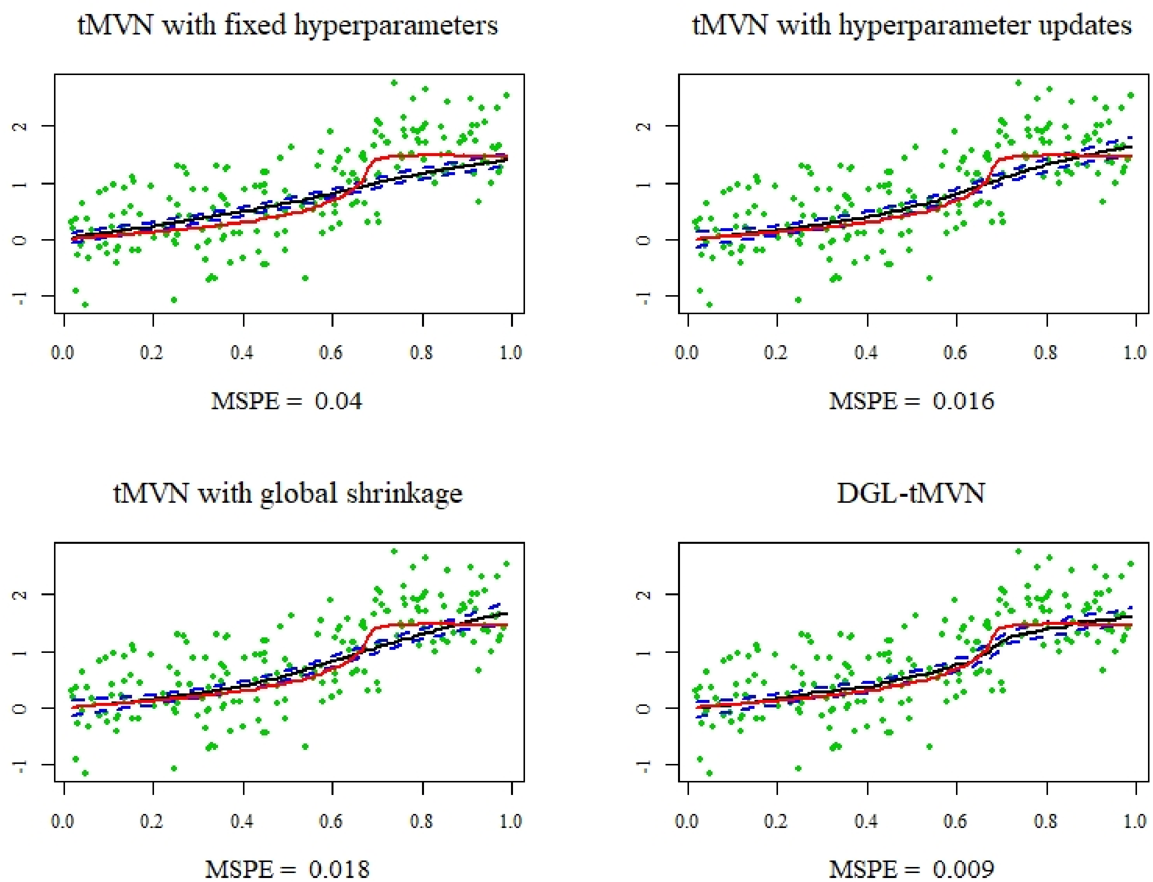

Figure S11: *Same as Figure S9, now for the function* f_3.

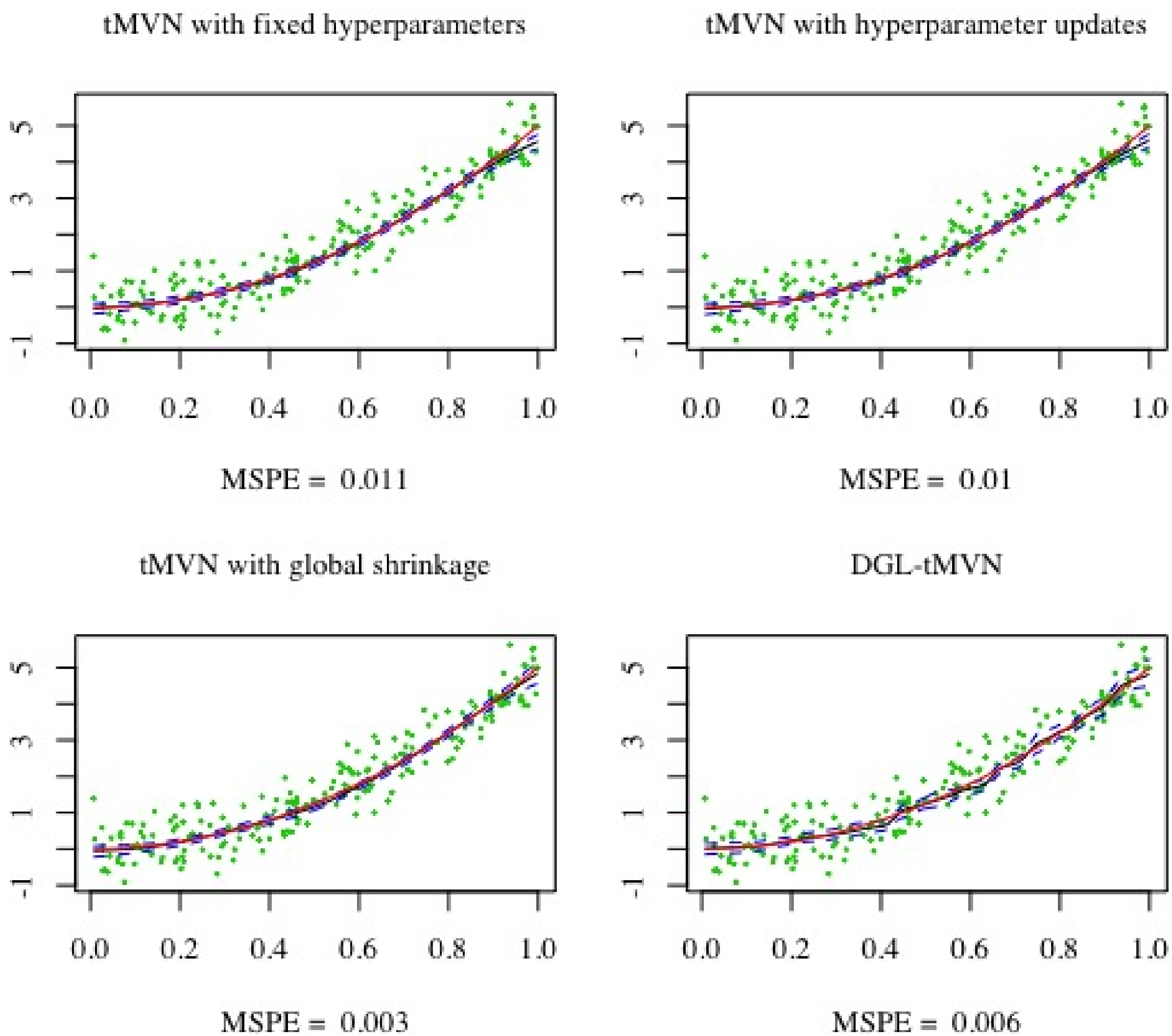

Figure S12: *Same as Figure S9, now for the function f_4.*

S8.4 Performance of bsar

Consider the simulation set-up specified corresponding to the four variants of tMVN priors in § 4.3. Figure S13 shows the out-of-sample prediction performance of bsar, developed by [10], and implemented by the **R** package **bsamGP**.

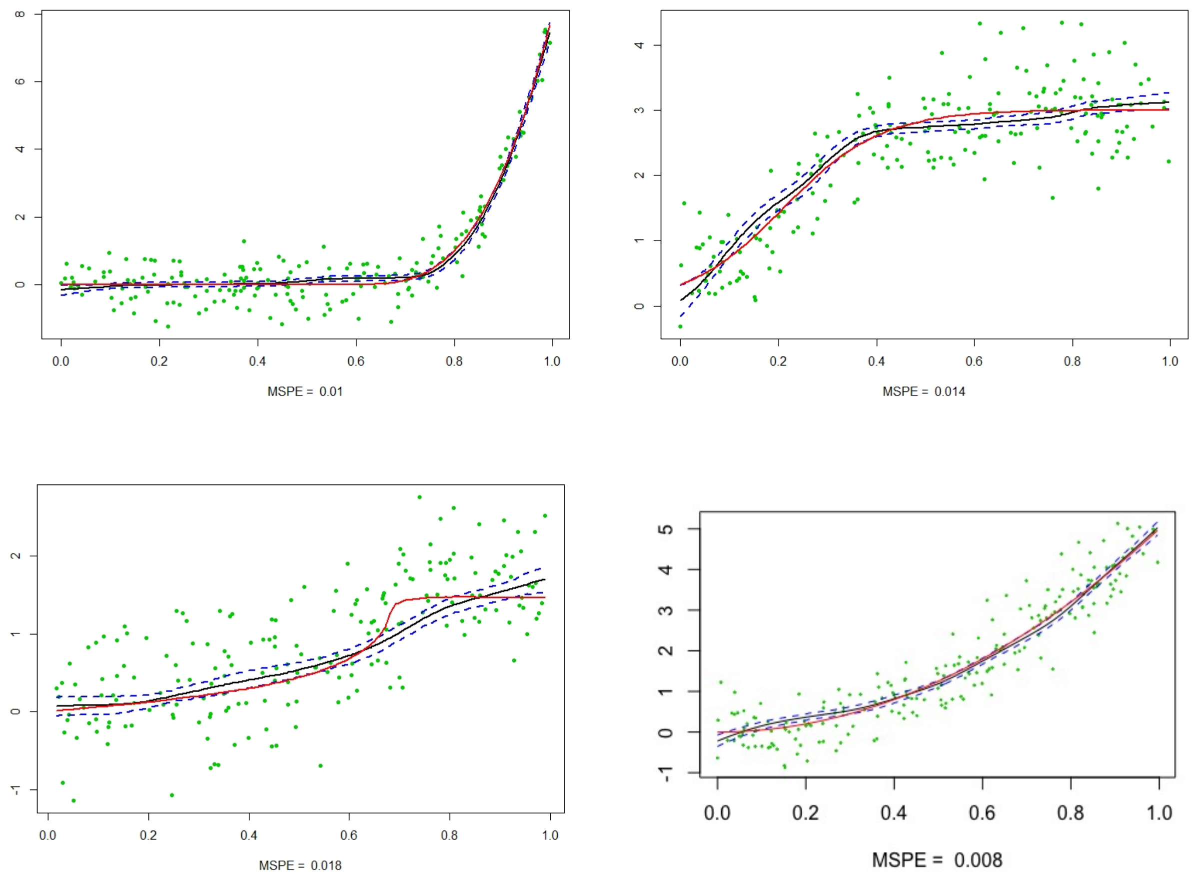

Figure S13: *Figure portraying out-of-sample prediction accuracy using bsar for f_1 (top left), f_2 (top right), f_3 (bottom left) and f_4 (bottom right). Red solid curve corresponds to the true function, black solid curve is the mean prediction, the region within two dotted blue curves represent 95% pointwise prediction interval and the green dots are* 200 *test data points.* MSPE *values corresponding to each of the method are also shown in the plots.*

S8.5 Sensitivity studies

In this section, we provide the results of the sensitivity studies mentioned in the manuscript.

Choice of covariance kernel: Consider the simulation set-up specified corresponding to the four variants of tMVN priors in § S8.3. For squared-exponential kernel with length-scale parameter $\ell > 0$, we place a compactly supported prior $\ell \sim \mathcal{U}(0.2, 1)$ to ensure that the correlation between the furthest two points in the covariate domain ranges from 10^{-6} to approximately 0.5. Figures S14, S15, S16 and S17 show the out-of-sample prediction performance based on squared-exponential covariance kernel.

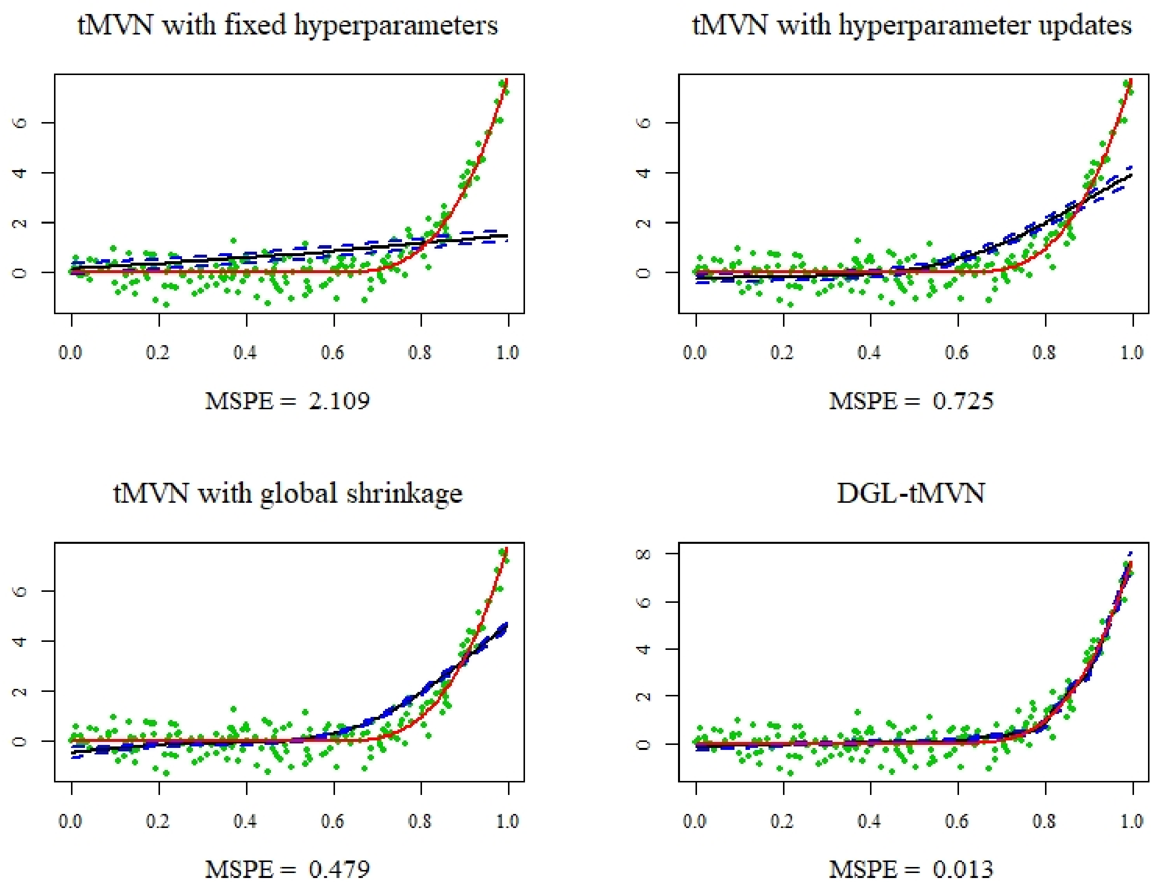

Figure S14: *Out-of-sample prediction accuracy for f_1 using the four variants and squared-exponential kernel. Red solid curve corresponds to the true function, black solid curve is the mean prediction, the region within two dotted blue curves represent 95% pointwise prediction interval and the green dots are* 200 *test data points.* MSPE *values corresponding to each of the method are also reported.*

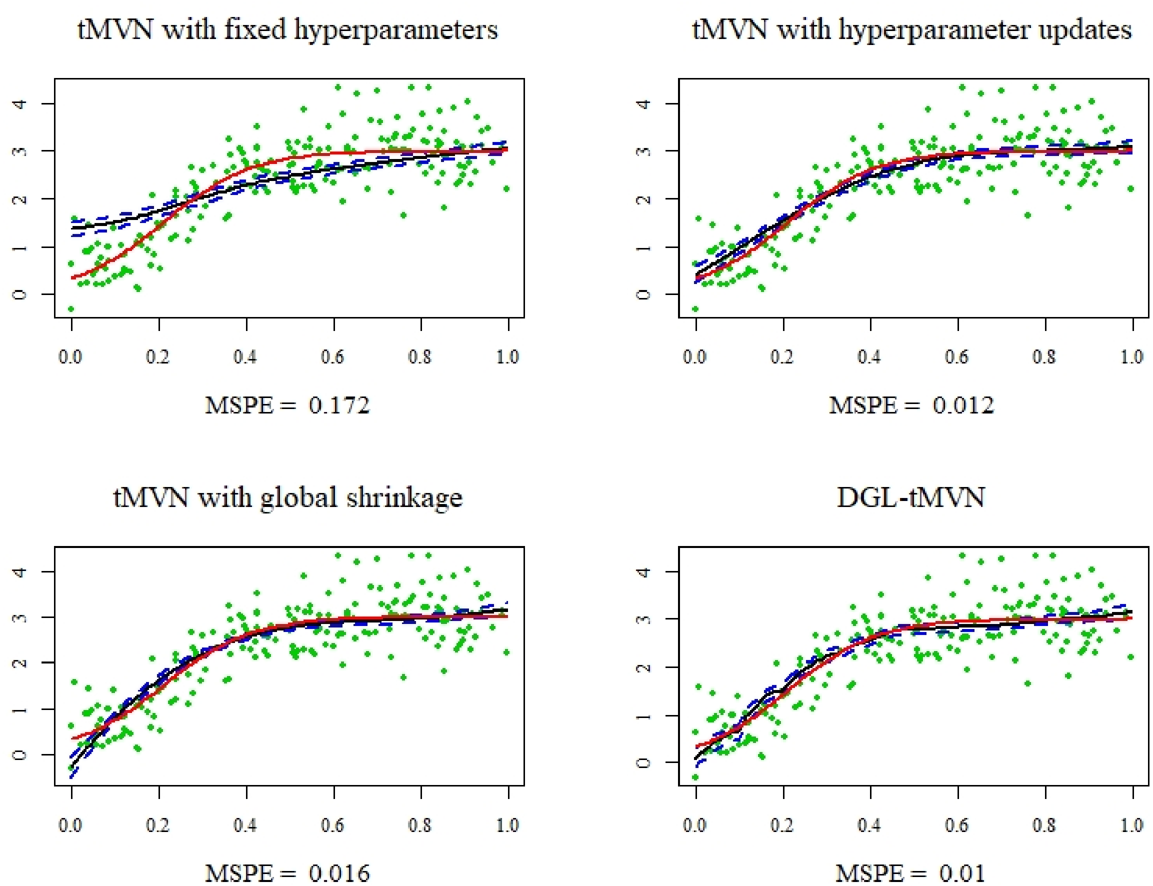

Figure S15: *Same as Figure S14 for the function* f_2.

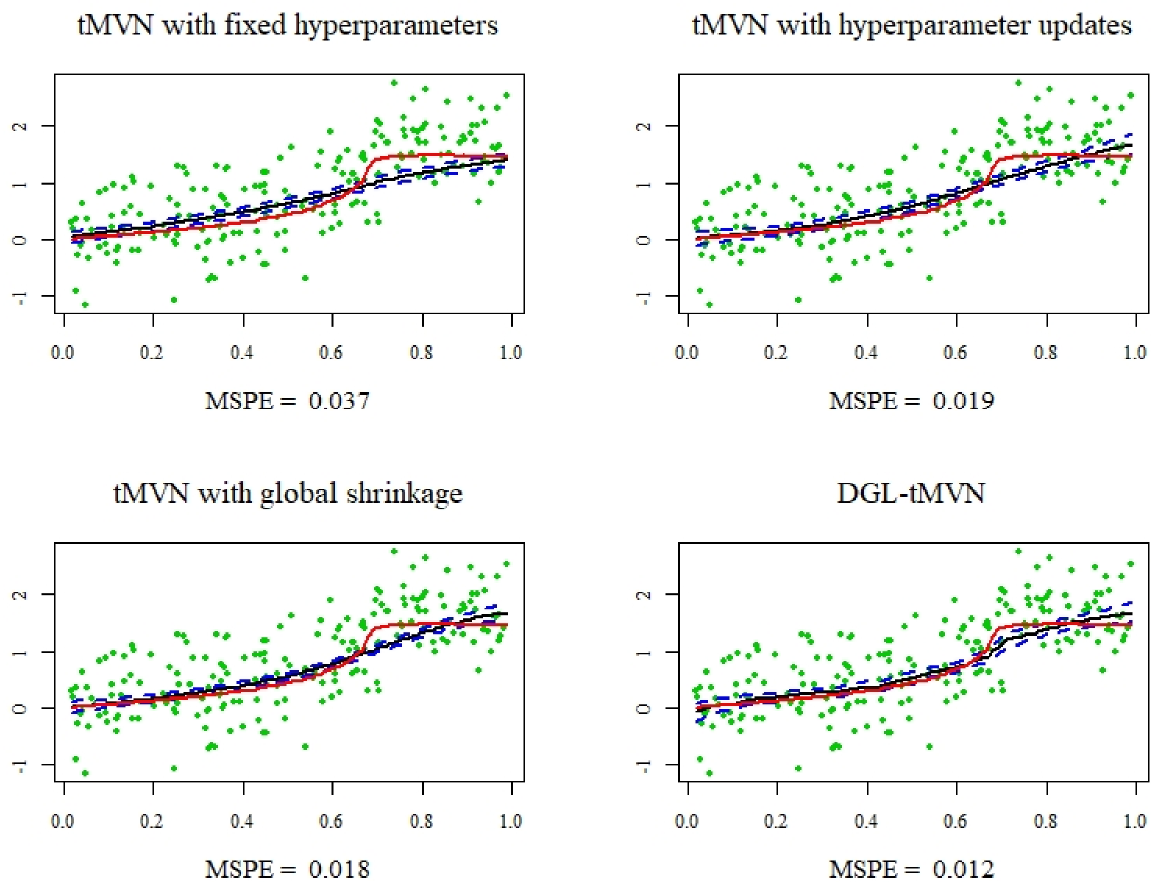

Figure S16: *Same as Figure S14 for the function* f_3.

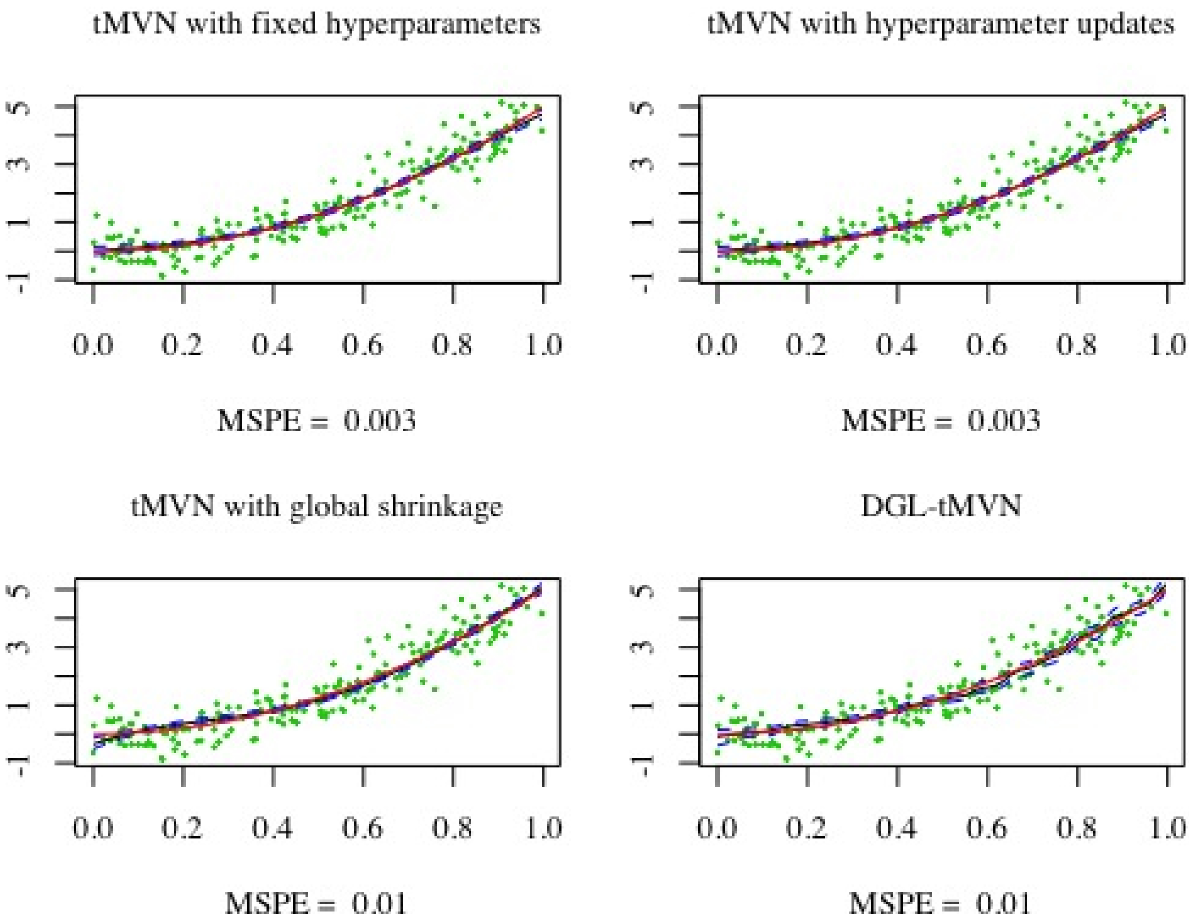

Figure S17: *Same as Figure S14 for the function* f_4.

Choice of priors on ν and ℓ: Consider the simulation settings specified corresponding to the four variants of tMVN priors in § 4.3. We placed compactly supported priors on ν and ℓ of a Matérn kernel as: $\nu \sim \mathcal{U}(0.05, 2)$ and $\ell \sim \mathcal{U}(0.05, 2)$. Figure S18 shows the out-of-sample prediction performance based on the choices of priors on ν and ℓ.

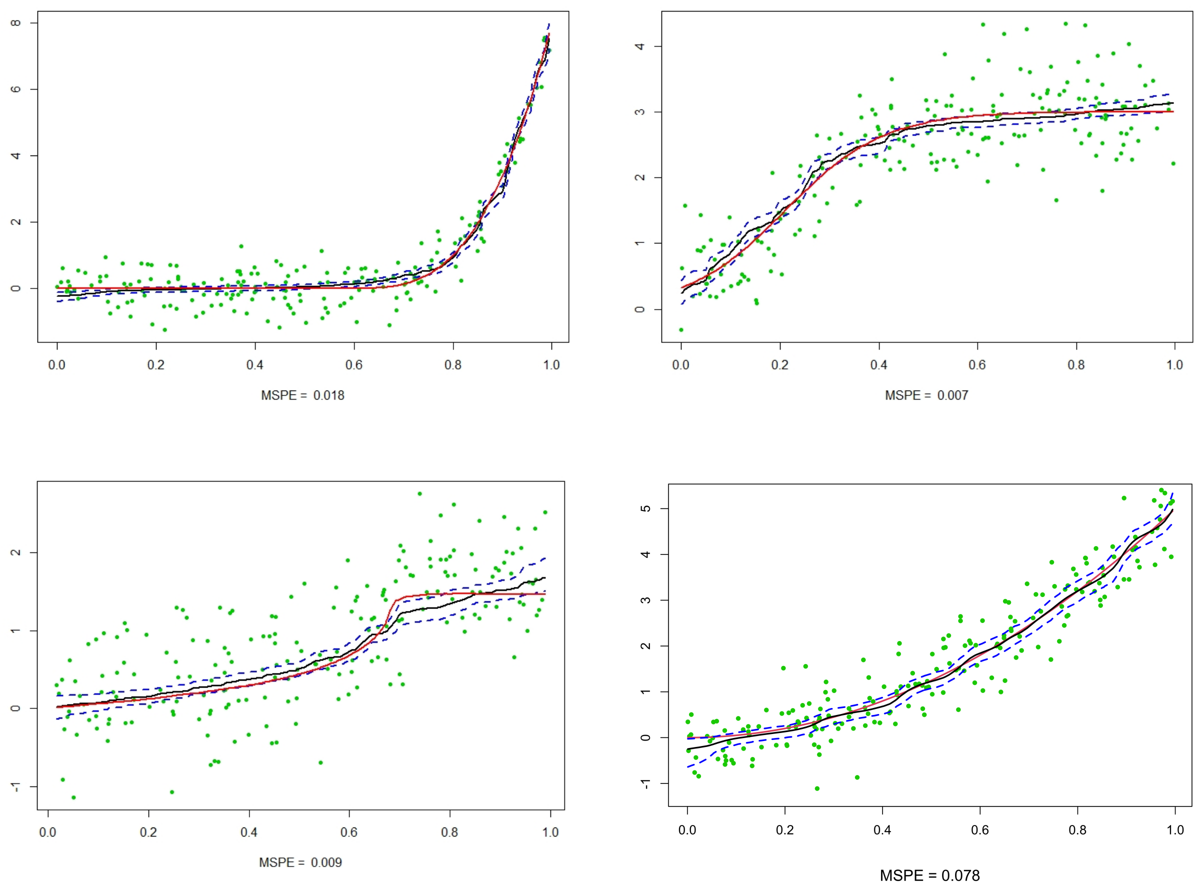

Figure S18: *Out-of-sample prediction accuracy using* DGL-*t*MVN *prior for f_1 (top left), f_2 (top right), f_3 (bottom left) and f_4 (bottom right) using Matérn kernel with $\nu \sim \mathcal{U}(0.05, 2)$ and $\ell \sim \mathcal{U}(0.05, 2)$. Red solid curve corresponds to the true function, black solid curve is the mean prediction, the region within two dotted blue curves represent 95% pointwise prediction interval and the green dots are* 200 *test data points.* MSPE *values corresponding to each of the method are also reported in the plots.*

Choice of δ_τ for simulation studies: Consider the variant of DGL-tMVN prior described in § 4.4. We used $\delta = 1.2$ and $\delta = 1.8$ for the sensitivity study and applied on simulation example of function f_4 in § 4.3. Figure S19 shows the estimation accuracy for estimating function f_4.

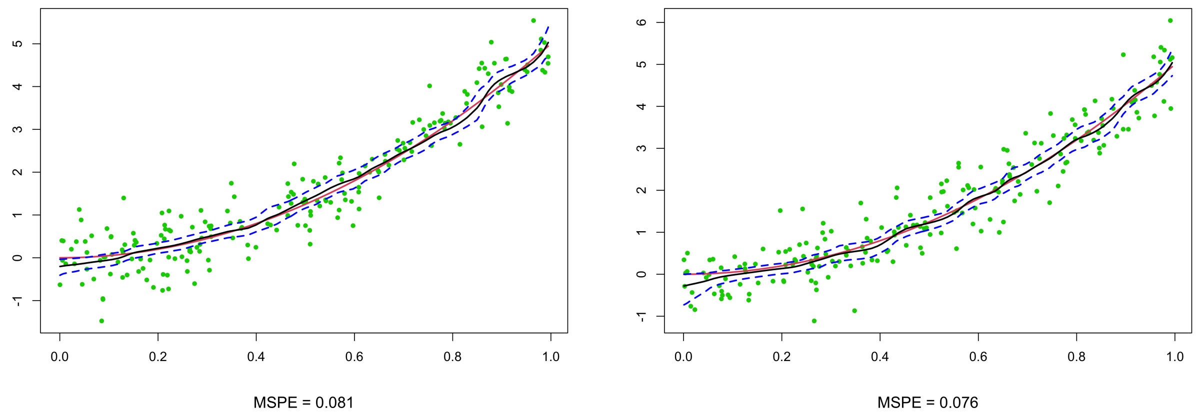

Figure S19: *Estimation accuracy based on $\delta_\tau = 1.2$ (left panel) and $\delta_\tau = 1.8$ (right panel) for fitting function f_4 in § 4.3. The black solid curve is the posterior mean, the region within two dotted blue curves represent 95% pointwise credible interval and the green dots are the observed data points. The* MSPE *values are shown in the sub-plots.*

Choice of δ_τ for real data analyses: Consider the variant of DGL-tMVN prior described in § 4.4. We used $\delta = 0.3$, $\delta = 0.7$, and $\delta = 1$ for the sensitivity study and applied on the real data sets described in § 5. Figures S20 and S21 show the estimation accuracy for these two data sets.

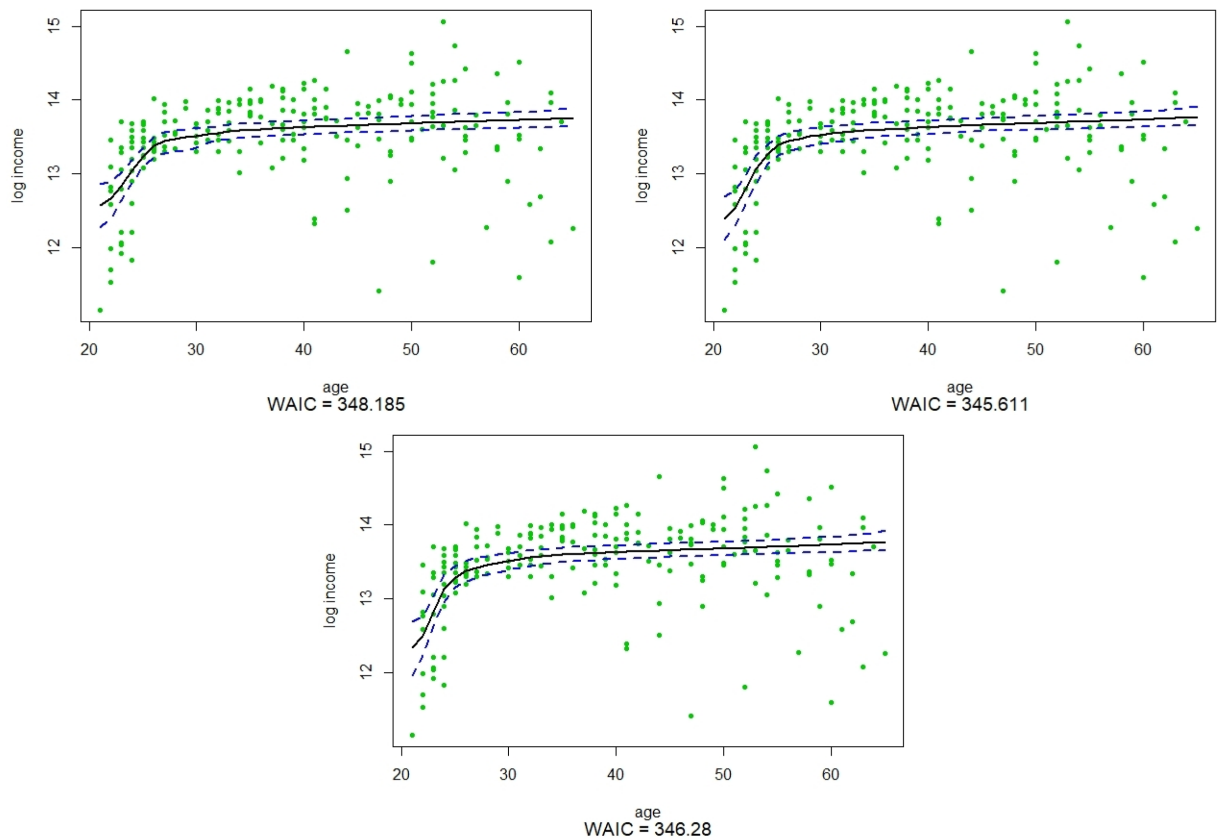

Figure S20: *Estimation accuracy based on $\delta_\tau = 0.3$ (top left panel), $\delta_\tau = 0.7$ (top right panel), and $\delta_\tau = 1$ (bottom panel) for the age and income data set used in § 5. The black solid curve is the posterior mean, the region within two dotted blue curves represent 95% pointwise credible interval and the green dots are the observed data points.*

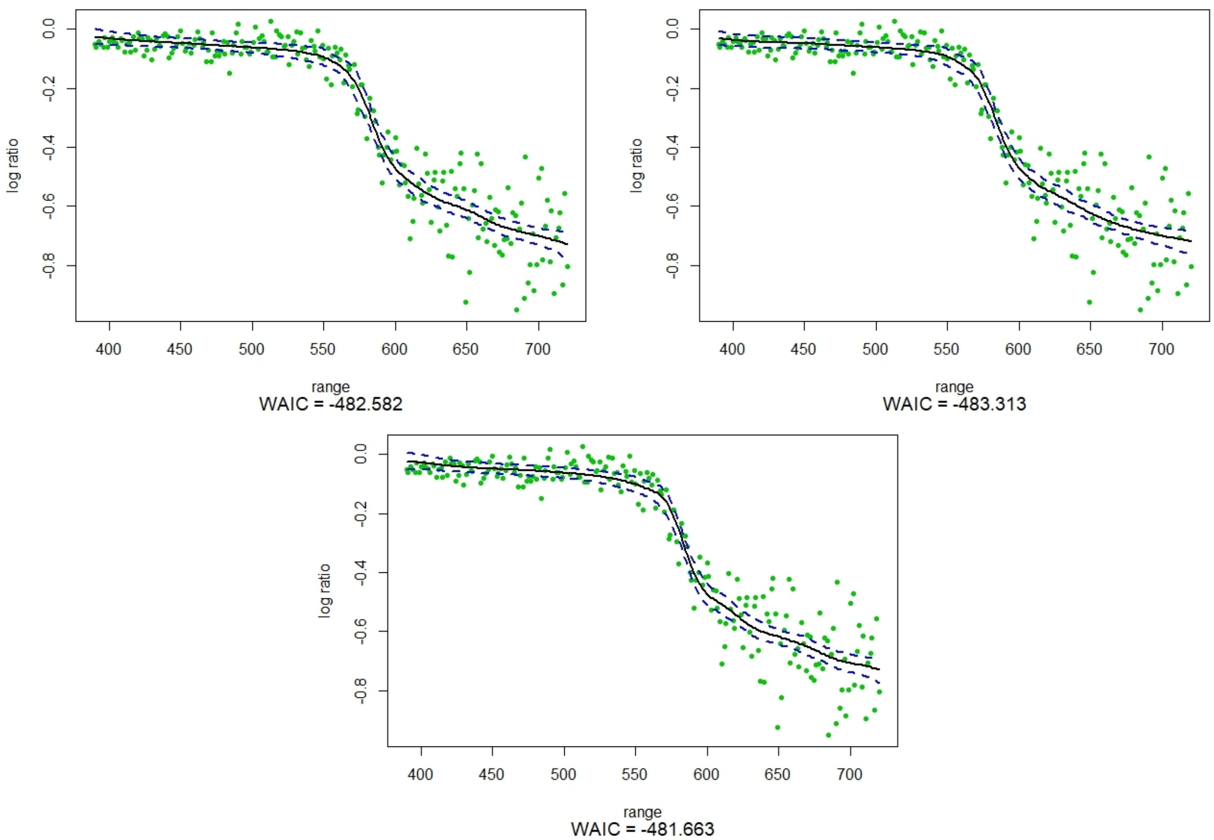

Figure S21: *Same as figure S20 for the LiDAR data set.*

S8.6 Diagnostics of MCMC samples

Here we provide results on the mixing behavior and computational efficiency of the Gibbs samplers based on the DGL, IGL, and tMVN priors respectively, as discussed in § 4.3. Figure S22 shows the boxplots of the effective sample sizes (ESS) and Table S2 reports the Monte Carlo standard errors (MCSE) of the MCMC samples of predicted function values based on 200 test points averaged over 25 replicates. Figure S23 and Table S3 provide the same for the real data sets discussed in § 5.

Table S2: The averaged standard deviations of MCMC samples of estimated function values over 200 test points and 25 replicates based on three different priors compared to the averaged standard deviations of the replicated response points for functions f_1, f_2, f_3 and f_4

	DGL	IGL	tMVN	SD(response)
f_1	0.00955	0.003389	0.009092	1.7520
f_2	0.00673	0.001664	0.004581	0.7610
f_3	0.00974	0.002348	0.005817	1.0117
f_4	0.01444	0.002180	0.006978	1.5171

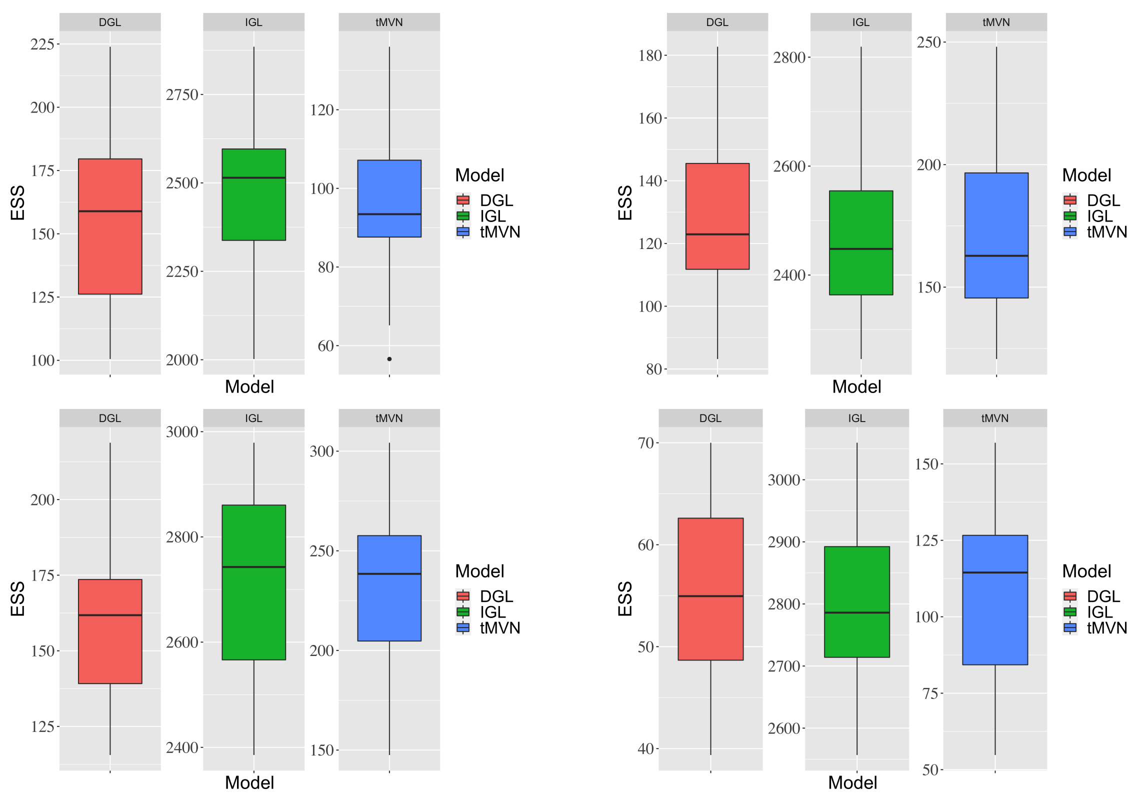

Figure S22: Boxplots of averaged effective sample sizes of estimated function values over 200 test samples based on DGL, IGL and tMVN over 25 replicated data sets for functions f_1 (top left panel), f_2 (top right panel), f_3 (bottom left panel) and f_4 (bottom right panel).

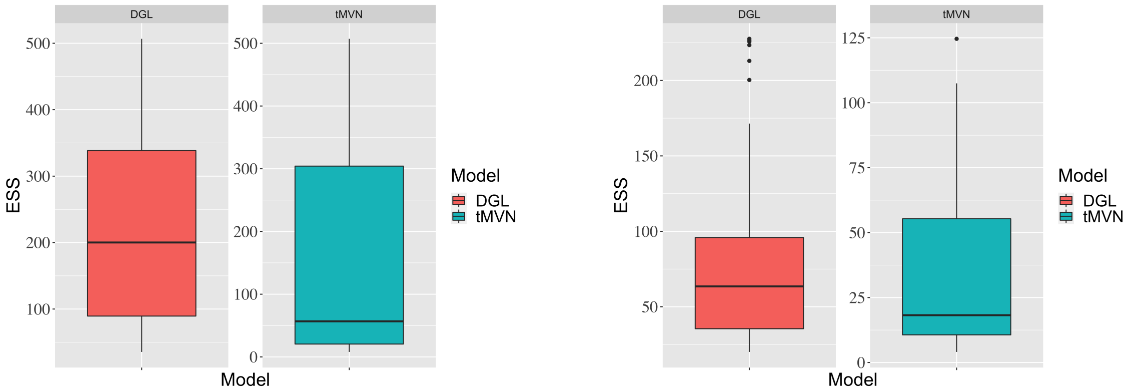

Figure S23: Boxplots of effective sample sizes of estimated function values of models with DGL and tMVN for the Age and income data (top left panel) and LiDAR (top right panel).

Table S3: Monte Carlo standard errors in estimating function values of models with DGL and tMVN against the standard deviation of the observed values for different data sets

	DGL	tMVN	SD(response)
Age-income	0.0016	0.0016	0.6363
LiDAR	8×10^{-4}	0.0014	0.2825

References

[1] Ray Bai, Gemma E Moran, Joseph L Antonelli, Yong Chen, and Mary R Boland. Spike-and-Slab group LASSOs for grouped regression and sparse generalized additive models. *Journal of the American Statistical Association*, pages 1–14, 2020. 25, 35

[2] Peter Bickel and Marko Lindner. Approximating the inverse of banded matrices by banded matrices with applications to probability and statistics. *Theory of Probability & Its Applications*, 56(1):1–20, 2012. 62

[3] Jack Cartinhour. One-dimensional marginal density functions of a truncated multivariate Normal density function. *Communications in Statistics-Theory and Methods*, 19(1):197–203, 1990. 2

[4] Carlos M Carvalho, Nicholas G Polson, and James G Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. 28

[5] Ismaël Castillo. Lower bounds for posterior rates with Gaussian process priors. *Electronic Journal of Statistics*, 2:1281–1299, 2008. 44

[6] Alan Genz. Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1(2):141–149, 1992. 2

[7] Alan Genz. Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics*, pages 400–400, 1993. 2

[8] Alan Genz and Frank Bretz. *Computation of Multivariate Normal and t Probabilities*, volume 195. Springer Science & Business Media, 2009. 2

[9] Subhashis Ghosal, Jayanta K. Ghosh, and Aad van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, 2000. 25

[10] Peter J Lenk and Taeryon Choi. Bayesian analysis of shape-restricted functions using Gaussian process priors. *Statistica Sinica*, 27:43–69, 2017. 69

[11] Hassan Maatouk and Xavier Bay. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582, 2017. 4

[12] Pascal Massart. Concentration inequalities and model selection. 2007. 17

[13] Pallavi Ray, Debdeep Pati, and Anirban Bhattacharya. Efficient Bayesian shape-restricted function estimation with constrained Gaussian process priors. *Statistics and Computing*, 30(4): 839–853, 2020. 64

[14] Qi-Man Shao. A Gaussian correlation inequality and its applications to the existence of small ball constant. *Stochastic processes and their applications*, 107(2):269–287, 2003. 61

[15] Qifan Song and Faming Liang. Nearly optimal Bayesian shrinkage for high dimensional regression. *arXiv preprint arXiv:1712.08964*, 2017. 32, 33, 34, 35, 37

[16] Roman Vershynin. *High–Dimensional Probability : An Introduction with Applications in Data Science*. Cambridge University Press, Cambridge, United Kingdom New York, NY, 2018. ISBN 978-1108415194. 54

[17] Ran Wei and Subhashis Ghosal. Contraction properties of shrinkage priors in logistic regression. *Journal of Statistical Planning and Inference*, 207:215–229, 2020. 28

[18] Ran Wei, Brian J Reich, Jane A Hoppin, and Subhashis Ghosal. Sparse Bayesian additive nonparametric regression with application to health effects of pesticides mixtures. *Statistica Sinica*, 30:55–79, 2020. 31, 33