Supplement to "A comprehensive Bayesian Framework for Envelope Models"

Saptarshi Chakraborty
and
Zhihua Su
Department of Statistics, University of Florida

This Supplement provides technical details, additional simulation results, proofs of theorems and derivations of results presented in the original text. The referred equations in this Supplement are labeled as (S1), (S2) etc., whereas labels such as (1), (2), etc. refer to equations from the main text. Supplementary tables and figures presented herein are labeled as Table S1, Figure S2 etc. which contrast labels such as Table 1 and Figure 2 etc. used to refer tables and figures from the main text. Materials are organized by the sections in which they appear in the main text.

A Inverse Wishart distribution and matrix normal distribution

In this section we briefly review the inverse Wishart and the matrix normal distributions used in prior construction for the proposed Bayesian approaches. For $d \geq 1$, a random positive definite matrix $\boldsymbol{Z} \in \mathbb{S}_+^{d\times d}$ is said to follow an inverse Wishart distribution $\mathrm{IW}_d(\boldsymbol{\Phi}, \nu)$ if $\boldsymbol{Z}^{-1}$ follows a Wishart distribution $\mathrm{W}_d(\boldsymbol{\Phi}^{-1}, \nu)$, where $\boldsymbol{\Phi} \in \mathbb{S}_+^{d\times d}$ is a positive definite matrix, and $\nu > d-1$. Here, $\mathbb{S}_+^{d\times d}$ denotes the open cone of positive definite $d\times d$ matrices in $\mathbb{S}^{d\times d}$, where $\mathbb{S}^{d\times d}$ is the set of all symmetric $d\times d$ matrices with real-valued entries. The density function of $\boldsymbol{Z}$ with respect to Lebesgue measure on $\mathbb{S}_+^{d\times d}$ is given by[1]

$$\frac{|\boldsymbol{\Phi}|^{\nu/2}}{2^{d\nu/2}\Gamma_d(\nu/2)}|\boldsymbol{Z}|^{-(d+\nu+1)/2}\exp\left[-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Phi}\boldsymbol{Z}^{-1}\right)\right].$$

Here Γ_d is the multivariate gamma function of dimension d, and it is defined as

$$\Gamma_d(a) = \pi^{d(d-1)/4}\prod_{j=1}^{d}\Gamma(a+(1-j)/2), \quad \text{for } a>0$$

with $\Gamma(\cdot)$ being the usual Gamma function. For more details on the inverse Wishart density and its support, see Andersson and Wojnar (2004).

The matrix normal distribution, on the other hand, is a generalization of the multivariate normal distribution. A random matrix $\boldsymbol{Z} \in \mathbb{R}^{d_1\times d_2}$ is said to follow the matrix normal distribution $\mathrm{MN}_{d_1,d_2}(\boldsymbol{Z}_0, \boldsymbol{U}, \boldsymbol{V})$, if its probability density function with respect to Lebesgue measure on $\mathbb{R}^{d_1\times d_2}$ is given by

$$\frac{\exp\{-\frac{1}{2}\operatorname{trace}[\boldsymbol{V}^{-1}(\boldsymbol{Z}-\boldsymbol{Z}_0)^T\boldsymbol{U}^{-1}(\boldsymbol{Z}-\boldsymbol{Z}_0)]\}}{(2\pi)^{d_1d_2/2}|\boldsymbol{V}|^{d_1/2}|\boldsymbol{U}|^{d_2/2}}$$

where $\boldsymbol{Z}_0 \in \mathbb{R}^{d_1\times d_2}$, $\boldsymbol{U} \in \mathbb{S}_+^{d_1\times d_1}$ and $\boldsymbol{V} \in \mathbb{S}_+^{d_2\times d_2}$.

[1] by Lebesgue measure on $\mathbb{S}_+^{d\times d}$, we mean the restriction of the standard Lebesgue measure on $\mathbb{R}^{d\times d}$ to the open subset $\mathbb{S}_+^{d\times d}$.

B Supplementary Notes on Approximate Bayesian Inference on the Envelope Dimension

In most frequentist envelope approaches, the envelope dimension u is selected using a model selection approach, such as AIC/BIC, likelihood ratio test, cross-validation, etc, and inferences on the model parameters are made conditional on the selected u. One exception is the weighted envelope estimation approach (Eck and Cook, 2017) where estimated envelope regression coefficients $\boldsymbol{\beta}$ are first obtained for all possible u-values and then averaged with certain BIC-based weights (discussed below). However, inference on u itself is typically infeasible in a frequentist framework. By contrast, the Bayesian paradigm, in addition to permitting model selection through similar criteria (e.g., via AIC/BIC, Deviance information criterion (Spiegelhalter et al., 2002), etc.), also permits a coherent approach to inferring u with all other envelope model parameters. To this, one assigns some prior distribution $\pi(u)$ on u, and subsequently obtains the joint posterior of $\pi(u, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \text{data})$. However, the complicated trans-dimensional nature of the model parameters for different u makes implementation extremely challenging. For example, a reversible jump MCMC sampler (Green, 1995) to sample from this joint posterior would require trans-dimensional jumps between $\boldsymbol{\Omega} \in \mathbb{S}_+^{u\times u}$ for different u-values across MCMC iterations, making implementation highly non-trivial. A possible alternative is to evaluate the marginal likelihood for a given u directly: $m(\text{data} \mid u) = \int \text{likelihood}(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid u, \text{data})\ d(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A})$ and subsequently use $m(\text{data} \mid u)$ to obtain the marginal posterior $\Pr(u \mid \text{data})$. However, the integral defining $m(\text{data} \mid u)$ is highly intractable and computationally infeasible in applications with constrained supports such as ours, and accurate computation of this integral for Bayesian envelope models is currently an open problem. Assuming vague prior on model parameters, we instead suggest a simple BIC-based approximation of the above marginal probability following Kass and Raftery (1995) and obtain the marginal u-posterior as:

$$\Pr(u = k \mid \text{data}) \approx \frac{\exp(-\text{BIC}(k)/2)\ \pi(u = k)}{\sum_{k'=0}^{r} \exp(-\text{BIC}(k')/2)\ \pi(u = k')};\ k = 0, 1, \ldots, r \tag{S1}$$

where $\text{BIC}(u) = -2 \log \widetilde{L}(u) + \rho(u) \log n$ with $\widetilde{L}(u)$ denoting the maximized value of the likelihood function, and $\rho(u)$ is the effective number of parameters in the model. For MCMC-based implementation, $\widetilde{L}$ can be approximated by the maximum of log-likelihood observed across MCMC draws (Frühwirth-Schnatter, 2011, Section 10.3.4). For the response envelope, predictor envelope, and the envelope probit models, the effective numbers of parameters ρ are $r(r+1)/2 + r + pu$, $r + r(r+1)/2 + p + p(p+1)/2 + rm$ and $1 + p + p(p+1)/2 + m$ respectively.

Inference on u can be made through the marginal posterior $\Pr(u \mid \text{data})$. For point estimation of u, one may consider the posterior mode $\hat{u} = \arg\max_{k=0,\ldots,r} \Pr(u = k \mid \text{data})$. Subsequently one can estimate the envelope model parameters $\{\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}\}$ conditional on $u = \hat{u}$ as done in a typical frequentist envelope approach. A more coherent *Bayesian model averaging* (BMA) approach acknowledging the estimation uncertainty in u marginalizes u entirely to produce the joint marginal (u-free) BMA posterior:

$$\pi(\mu, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}, \boldsymbol{A} \mid \text{data}) = \sum_{k=0}^{r} \pi(\mu, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}, \boldsymbol{A} \mid u = k, \text{data})\ \Pr(u = k \mid \text{data}). \tag{S2}$$

Leveraging its finite mixture representation, one may readily compute approximate moments of

$$\pi(\mu, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}, \boldsymbol{A} \mid \text{data})$$

given the computed moments of $u = k$ specific posteriors $\{\pi(\mu, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}, \boldsymbol{A} \mid u = k, \text{data})\}$. In particular, the marginal posterior mean for $\boldsymbol{\beta}$ is

$$E(\boldsymbol{\beta} \mid \text{data}) = \sum_{k=0}^{r} E(\boldsymbol{\beta} \mid u = k, \text{data}) \Pr(u = k \mid \text{data}) \tag{S3}$$

which is a weighted average of posterior means of $\boldsymbol{\beta}$ obtained from $u = k$ specific models. To compute posterior quantities beyond moments e.g., quantiles, one may leverage approximate draws from the joint posterior $\pi(\mu, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}, \boldsymbol{A} \mid u, \text{data})$ generated posthoc using MCMC draws from each u-conditioned posterior. There, one would iteratively generate random $\{u\}$ with probabilities $\{\Pr(u = k \mid \text{data})\}$ and then conditional on that u randomly pick a draw from the MCMC samples generated for $\pi(\mu, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}, \boldsymbol{A} \mid u, \text{data})$. (The u-specific posterior MCMC samples are drawn separately and possibly parallely for different u's).

C Various Technical Details and Additional Results on the Reparameterized Bayesian Response Envelope Model and its Implementation

C.1 Justification of model identifiability

First, we note that the identifiability of the envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is proved in Cook et al. (2010). The one-to-one correspondence between $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ and $\boldsymbol{A}$ (Su et al., 2016), ensures that the matrix $\boldsymbol{A}$ is identifiable. Because

$$\boldsymbol{C}_{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{I}_u \\ \boldsymbol{A} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Gamma}(\boldsymbol{A}) = \boldsymbol{C}_{\boldsymbol{A}}(\boldsymbol{C}_{\boldsymbol{A}}^T \boldsymbol{C}_{\boldsymbol{A}})^{-1/2},$$

therefore $\boldsymbol{\Gamma}(\boldsymbol{A})$ is also identifiable. Similarly, since

$$\boldsymbol{D}_{\boldsymbol{A}} = \begin{pmatrix} -\boldsymbol{A}^T \\ \boldsymbol{I}_{r-u} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_0(\boldsymbol{A}) = \boldsymbol{D}_{\boldsymbol{A}}(\boldsymbol{D}_{\boldsymbol{A}}^T \boldsymbol{D}_{\boldsymbol{A}})^{-1/2},$$

therefore $\boldsymbol{\Gamma}_0(\boldsymbol{A})$ is identifiable. Consequently, under the regression model

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{\varepsilon}, \qquad E(\boldsymbol{\varepsilon}) = 0, \qquad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma},$$

we have $\boldsymbol{\eta} = \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{\beta}$, $\boldsymbol{\Omega} = \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{\Sigma}\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\boldsymbol{\Omega}_0 = \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\boldsymbol{\Sigma}\boldsymbol{\Gamma}_0(\boldsymbol{A})$. This implies that $\boldsymbol{\eta}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are identified. Thus all the parameters are identifiable in our reparameterized envelope model

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\boldsymbol{X} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A}).$$

C.2 Difference between the reparameterization of response envelope model (6) and the one used in Khare et al. (2017)

The main difference between the proposed reparameterization and the one used in Khare et al. (2017) is that the proposed reparameterization is free of any manifold structure. The key parameter of interest in an envelope model is the envelope subspace $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. In its most general formulation, $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ belongs to an $r \times u$ Grassmann manifold, which is the set of all u-dimensional subspaces in an r-dimensional Euclidean space. However, direct handling of abstract random manifolds for Bayesian modeling is extremely difficult. The main purpose of the both reparameterizations, the one considered in our paper and the one considered in Khare et al. (2017), is to identify this target Grassmanian manifold with appropriately chosen matrices that are more manageable for probabilistic handling in Bayesian inference, and facilitate incorporation of (induced) prior information on $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ through richer classes of probability distributions on matrix spaces.

Our reparameterization focuses on a specific basis of $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ wherein the (possibly top) u rows form an identity matrix and the remaining $r - u$ rows form an unrestricted Euclidean matrix, called $\boldsymbol{A}$. This parameterization has previously been used in Ma and Zhu (2013) and Su et al. (2016), and

the key advantange of this parameterization lies in the fact that $\boldsymbol{A}$ and $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$, and $\boldsymbol{A}$ and $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})^{\perp}$ have one-to-one correspondences. Since $\boldsymbol{A}$ is unconstrained, this reparameterization completely avoids any manifold structure. Consequently, any distribution for unconstrained Euclidean matrices can be potentially considered here for prior elicitation on $\boldsymbol{A}$ which induces a prior on $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$.

In contrast, the reparameterization in Khare et al. (2017) looks for an orthonormal basis for $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ of a specific form. The orthonormality is required to diagonalize $\boldsymbol{\Sigma}$; in addition, the element with the largest absolute value in each column of the is required to have a positive sign. Let us denote this basis matrix by $\boldsymbol{G}$. It can be shown that $\boldsymbol{G}$ is a unique orthonormal basis for $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$. Similarly, a unique orthonormal basis for $\mathcal{E}_{\boldsymbol{\Sigma}}(\boldsymbol{\beta})^{\perp}$ can be constructed, which we denote by $\boldsymbol{G}_0$. Then $\boldsymbol{O} = (\boldsymbol{G}, \boldsymbol{G}_0)$ forms an $r \times r$ orthogonal matrix, and is a point on a restricted Stiefel manifold. A restricted Stiefel manifold is a Stiefel manifold (the space of all orthogonal matrices) that is restricted to conform to the sign requirements. Any distribution on a Stiefel manifold can be potentially considered in this framework as prior distribution for $\boldsymbol{O}$; Khare et al. (2017) consider the matrix Bingham distribution for $\boldsymbol{O}$. Thus, to summarize, Khare et al. (2017) convert the original Grassmann manifold parameterization to a restricted Stiefel manifold parameterization, while our reparameterization is free of any manifold structure.

C.3 Log likelihood and log posterior density under response envelope model (6)

Given n independent observations $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ from the response envelope model (6)), the log likelihood for the model parameters is given by

$$\log L(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}) = \text{const.} - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\,\text{trace}\left[\left(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^T - \mathbb{X}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\right.$$
$$\left.\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^T - \mathbb{X}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\right].$$

Here "const." stands for an unspecified constant free of the model parameters, and as defined in the main text $\mathbb{Y}^T = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ and $\mathbb{X}^T = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$. Together with the prior distribution as discussed in the main text, the un-normalized log posterior density of the model parameters is given by:

$$\begin{aligned}
&\log \pi(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= \text{const.} - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\,\text{trace}\left[\left(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^T - \mathbb{X}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\left(\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\right. \\
&\left.\left(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^T - \mathbb{X}\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\right] - \frac{\nu + u + 1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\,\text{trace}\left(\boldsymbol{\Psi}\boldsymbol{\Omega}^{-1}\right) - \frac{\nu_0 + r - u + 1}{2}\log|\boldsymbol{\Omega}_0| \\
&- \frac{1}{2}\,\text{trace}\left(\boldsymbol{\Psi}_0\boldsymbol{\Omega}_0^{-1}\right) - \frac{1}{2}\,\text{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right] - \frac{p}{2}\log|\boldsymbol{\Omega}| \\
&- \frac{1}{2}\,\text{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\,\boldsymbol{M}\,\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right]. \qquad \text{(S4)}
\end{aligned}$$

We consider the following parameter transformation: $(\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}) \to (\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A})$ where $\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\overline{\boldsymbol{X}}$, which aids substantial simplification in the forms of the conditional densities required for random generation from the posterior distribution. Clearly, the absolute Jacobian of the above transformation is one, and the un-normalized log posterior density for the transformed parameters is given by:

$$\begin{aligned}
&\log \pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= \text{const.} - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\,\text{trace}\left[\left(\mathbb{Y} - \mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T - \mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\left(\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\right.
\end{aligned}$$

$$\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\Big]-\frac{\nu+u+1}{2}\log|\boldsymbol{\Omega}|-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Psi}\boldsymbol{\Omega}^{-1}\right)-\frac{\nu_0+r-u+1}{2}\log|\boldsymbol{\Omega}_0|$$
$$-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Psi}_0\boldsymbol{\Omega}_0^{-1}\right)-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]-\frac{p}{2}\log|\boldsymbol{\Omega}|$$
$$-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\ \boldsymbol{M}\ \left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right]. \tag{S5}$$

C.4 Proof of Theorem 3.1 (Posterior Propriety)

Suppose we have n independent observations $(\boldsymbol{X}_1,\boldsymbol{Y}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{Y}_n)$ from the response envelope model (6). Define $\mathbb{Y}\in\mathbb{R}^{n\times r}$ and $\mathbb{X}\in\mathbb{R}^{n\times p}$ by $\mathbb{Y}^T=(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n)$ and $\mathbb{X}^T=(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$, and let $\mathbf{1}_n$ denote an n-dimensional vector of 1's. At the outset, we consider the transformation $(\boldsymbol{\mu},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A})\to(\widetilde{\boldsymbol{\mu}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A})$ as discussed in Section C.3. Because the transformation has absolute Jacobian of transformation 1, it is equivalent to show that the posterior density of the transformed parameters $(\widetilde{\boldsymbol{\mu}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A})$, as given in (S5) is proper. To this end, we consider the corresponding unnormalized posterior density $f(\boldsymbol{\mu},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y})$ as given in (S4) . Note that

$$\begin{aligned}
&\log f(\widetilde{\boldsymbol{\mu}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y})\\
=&-\frac{n}{2}\log|\boldsymbol{\Omega}|-\frac{n}{2}\log|\boldsymbol{\Omega}_0|-\frac{1}{2}\operatorname{trace}\Big[\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\left(\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\\
&\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\Big]-\frac{\nu+u+1}{2}\log|\boldsymbol{\Omega}|-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Psi}\boldsymbol{\Omega}^{-1}\right)-\frac{\nu_0+r-u+1}{2}\log|\boldsymbol{\Omega}_0|\\
&-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Psi}_0\boldsymbol{\Omega}_0^{-1}\right)-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]-\frac{p}{2}\log|\boldsymbol{\Omega}|\\
&-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\ \boldsymbol{M}\ \left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right]. \qquad \text{(S6)}
\end{aligned}$$

Our objective is to show that f is integrable with respect to the Lebesgue measure on $\mathbb{R}^r\times\mathbb{R}^{u\times p}\times\mathbb{S}_+^{u\times u}\times\mathbb{S}_+^{(r-u)\times(r-u)}\times\mathbb{R}^{(r-u)\times u}$.
For notational convenience, we denote by $\int f(t)\,dt$ an appropriate Lebesgue integral. The following proof incorporates standard techniques from Bayesian analysis for proving posterior propriety, and uses some tricks similar to the ones used in Khare et al. (2017). Note that

$$\begin{aligned}
&\operatorname{trace}\left[\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\right]\\
=&\operatorname{trace}\Big[\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\\
&\qquad\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\left(\mathbb{P}_{\mathbf{1}_n}+\mathbb{Q}_{\mathbf{1}_n}\right)\left(\mathbb{Y}-\mathbf{1}_n\widetilde{\boldsymbol{\mu}}^T-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\Big]\\
=&\operatorname{trace}\Big[\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\\
&\qquad\left\{n\left(\overline{\boldsymbol{Y}}-\widetilde{\boldsymbol{\mu}}\right)\left(\overline{\boldsymbol{Y}}-\widetilde{\boldsymbol{\mu}}\right)^T+\left(\mathbb{Y}-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\mathbb{Q}_{\mathbf{1}_n}\left(\mathbb{Y}-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\right\}\Big]\\
=&\operatorname{trace}\Big[\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\\
&\qquad\left\{n\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)^T+\left(\mathbb{Y}_c-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\left(\mathbb{Y}_c-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\right\}\Big]\\
=&n\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)\\
&+\operatorname{trace}\left[\left(\mathbb{Y}_c-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\mathbb{Y}_c-\mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\right]
\end{aligned}$$

$$+ \operatorname{trace}\left[\mathbb{Y}_c\, \boldsymbol{\Gamma}_0(\boldsymbol{A})\, \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\, \mathbb{Y}_c^T\right] \tag{S7}$$

Also,

$$\begin{aligned}
&\operatorname{trace}\left[\left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\right] \\
&\qquad + \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\,\boldsymbol{M}\,\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right] \\
&= \operatorname{trace}\left[\boldsymbol{\Omega}^T\left\{\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)^T\left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\right)\boldsymbol{\Gamma}(\boldsymbol{A}) + \left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\,\boldsymbol{M}\,\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right\}\right] \\
&= \operatorname{trace}\left[\boldsymbol{\Omega}^T\left\{\left(\mathbb{Y}_c\boldsymbol{\Gamma}(\boldsymbol{A}) - \mathbb{X}_c\boldsymbol{\eta}^T\right)^T\left(\mathbb{Y}_c\boldsymbol{\Gamma}(\boldsymbol{A}) - \mathbb{X}_c\boldsymbol{\eta}^T\right) + \left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\,\boldsymbol{M}\,\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right\}\right] \\
&= \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left\{\boldsymbol{\eta}\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\boldsymbol{\eta}^T - 2\boldsymbol{\eta}\left(\mathbb{X}_c^T\mathbb{Y}_c + \boldsymbol{M}\boldsymbol{e}^T\right)\boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{\Gamma}(\boldsymbol{A})^T\left(\mathbb{Y}^T\mathbb{Y} + \boldsymbol{e}\boldsymbol{M}\boldsymbol{e}^T\right)\boldsymbol{\Gamma}(\boldsymbol{A})\right\}\right] \\
&= \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}\boldsymbol{M}\boldsymbol{e}^T - \check{\boldsymbol{e}}\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\check{\boldsymbol{e}}^T\right)\boldsymbol{\Gamma}(\boldsymbol{A})\right] \\
&\qquad + \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right] \\
&= \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})\right] + \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right].
\end{aligned} \tag{S8}$$

where

$$\begin{aligned}
\check{\boldsymbol{e}} &= \left(\mathbb{Y}_c^T\mathbb{X}_c + \boldsymbol{e}\boldsymbol{M}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)^{-1} \\
\text{and } \widetilde{\boldsymbol{G}} &= \mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}\boldsymbol{M}\boldsymbol{e}^T - \check{\boldsymbol{e}}\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\check{\boldsymbol{e}}^T
\end{aligned} \tag{S9}$$

and $\widetilde{\boldsymbol{G}}$ is positive semi definite (see, e.g., the proof of Lemma 2 in Khare et al. (2017)). Therefore, from (S6), (S7) and (S8) we get

$$\begin{aligned}
&\log f(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= -\frac{1}{2}\left[n\left(\widetilde{\boldsymbol{\mu}} - \overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}} - \overline{\boldsymbol{Y}}\right)\right] \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right] \\
&\quad - \frac{n + p + \nu + u + 1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})\right] - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\right] \\
&\quad - \frac{n + \nu_0 + (r - u) + 1}{2}\log|\boldsymbol{\Omega}_0| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})\right] - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\right] \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right].
\end{aligned} \tag{S10}$$

Since

$$\begin{aligned}
&\int \exp\left(-\frac{1}{2}\operatorname{trace}\left[n\left(\widetilde{\boldsymbol{\mu}} - \overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}} - \overline{\boldsymbol{Y}}\right)\right]\right) d\widetilde{\boldsymbol{\mu}} \\
&\qquad = (2\pi)^{r/2} n^{r/2} |\boldsymbol{\Omega}|^{1/2} |\boldsymbol{\Omega}_0|^{1/2}, \\
\text{and } &\int \exp\left(-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]\right) d\boldsymbol{\eta} \\
&\qquad = (2\pi)^{up/2} |\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}|^{-u/2} |\boldsymbol{\Omega}|^{p/2},
\end{aligned}$$

therefore

$$
\begin{aligned}
\log I_1 &= \log\left(\int\int f(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) d\widetilde{\boldsymbol{\mu}}\, d\boldsymbol{\eta}\right) \\
&= C_1 + \frac{1}{2}\log|\boldsymbol{\Omega}| + \frac{1}{2}\log|\boldsymbol{\Omega}_0| + \frac{p}{2}\log|\boldsymbol{\Omega}| \\
&\quad - \frac{n+p+\nu+u+1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{\Psi}\right)\right] \\
&\quad - \frac{n+\nu_0+(r-u)+1}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A}) + \boldsymbol{\Psi}_0\right)\right] \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right] \\
&= C_1 - \frac{(n+\nu-1)+u+1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{\Psi}\right)\right] \\
&\quad - \frac{(n+\nu_0-1)+(r-u)+1}{2}\log|\boldsymbol{\Omega}_0| \\
&\qquad\qquad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A}) + \boldsymbol{\Psi}_0\right)\right] \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]
\end{aligned} \tag{S11}
$$

where

$$
C_1 = \left(\frac{r}{2} + \frac{up}{2}\right)\log(2\pi) + \frac{r}{2}\log n + \frac{up}{2} - \frac{u}{2}\log|\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}|.
$$

Since $n+\nu > u$ and $n+\nu_0 > r-u$, therefore, I_1 is integrable with respect to $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$. Now, integrating I_1 from (S11) with respect to $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$, and then by taking logarithm, we get

$$
\begin{aligned}
\log I_2 &= \log\left(\int\int I_1\, d\boldsymbol{\Omega}\, d\boldsymbol{\Omega}_0\right) \\
&= C_1 + \frac{(n+\nu-1)\,u}{2}\log 2 + \log\Gamma_u\left(\frac{n+\nu-1}{2}\right) \\
&\qquad\qquad - \frac{n+\nu-1}{2}\log\left|\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{\Psi}\right| \\
&\quad + \frac{n+\nu_0-1)\,(r-u)}{2}\log 2 + \log\Gamma_{r-u}\left(\frac{n+\nu_0-1}{2}\right) \\
&\qquad\qquad - \frac{n+\nu_0-1}{2}\log\left|\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A}) + \boldsymbol{\Psi}_0\right| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right] \\
&= C_2 - \frac{n+\nu-1}{2}\log\left(|\boldsymbol{\Psi}|\left|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Psi}^{-1/2} + \boldsymbol{I}_u\right|\right) \\
&\quad - \frac{n+\nu_0-1}{2}\log\left(|\boldsymbol{\Psi}_0|\left|\boldsymbol{\Psi}_0^{-1/2}\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Psi}_0^{-1/2} + \boldsymbol{I}_{r-u}\right|\right) \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right] \\
&= C_3 - \frac{n+\nu-1}{2}\log\left|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Psi}^{-1/2} + \boldsymbol{I}_u\right| \\
&\quad - \frac{n+\nu_0-1}{2}\log\left|\boldsymbol{\Psi}_0^{-1/2}\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Psi}_0^{-1/2} + \boldsymbol{I}_{r-u}\right| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]
\end{aligned}
$$

$$\leq C_3 - \frac{1}{2}\,\text{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]. \tag{S12}$$

where

$$C_2 = C_1 + \frac{(n+\nu-1)\,u}{2}\log 2 + \log\Gamma_u\left(\frac{n+\nu-1}{2}\right) + \frac{(n+\nu_0-1)\,(r-u)}{2}\log 2 + \log\Gamma_{r-u}\left(\frac{n+\nu_0-1}{2}\right)$$

and

$$C_3 = C_2 - \frac{(n+\nu-1)\,u}{2}\log|\boldsymbol{\Psi}| - \frac{(n+\nu_0-1)\,(r-u)}{2}\log|\boldsymbol{\Psi}_0|$$

and the last inequality follows from the fact that both

$$\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Psi}^{-1/2} \text{ and } \boldsymbol{\Psi}_0^{-1/2}\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Psi}_0^{-1/2}$$

are positive semi definite, and that both determinant and logarithm are order-preserving functions. From (S12) it follows that

$$\begin{aligned}
&\int\int\int\int\int f(\widetilde{\boldsymbol{\mu}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y})\,d\widetilde{\boldsymbol{\mu}}\,d\boldsymbol{\eta}\,d\boldsymbol{\Omega}\,d\boldsymbol{\Omega}_0\,d\boldsymbol{A} = \int I_2\,d\boldsymbol{A}\\
&\leq \exp(C_3)\int\exp\left(-\frac{1}{2}\,\text{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right)d\boldsymbol{A}\\
&= \exp(C_3)\,(2\pi)^{\frac{(r-u)u}{2}}\,|\boldsymbol{K}|^{u/2}\,|\boldsymbol{L}|^{(r-u)/2} < \infty.
\end{aligned}$$

This completes the proof. □

C.5 Proof of Theorem 3.2 (Harris Ergodicity)

At the outset we consider the following parameter transformation: $(\boldsymbol{\mu},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}) \to (\widetilde{\boldsymbol{\mu}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A})$ where $\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\overline{\boldsymbol{X}}$, as discussed before in Section C.3.

The Metropolis within Gibbs sampler described in Algorithm 3.1, and its generalization to the cases $u = 0$ and $u = r$ block-wise updates (in a deterministic/random/mixed order) the following $4+u$ parameter blocks: $\widetilde{\boldsymbol{\mu}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{a}_1,\cdots,\boldsymbol{a}_u$, where $\boldsymbol{a}_j$ denotes the j-th column of $\boldsymbol{A}$ (if $1 \leq u \leq r-1$, otherwise some of the parameters do not appear in the model; see Remark 3.3 in the main text). We discuss the three components of Harris ergodicity separately.

First, we note that the acceptance probabilities in the Metropolis steps in S.1, if present, i.e., if $1 \leq u \leq r$, are strictly positive everywhere, since the target full conditional density of $\boldsymbol{a}_j$ is positive everywhere. The Gibbs steps S.3–S.5, when viewed as Metropolis steps with same proposal and target (full conditional) densities, of course have everywhere positive acceptance probabilities (the acceptance probabilities are all one). This, together with the fact that the proposal densities (the full conditional densities in steps S.3–S.5 and independent multivariate normal densities in step S.1) are all everywhere positive, ensures that $K(x,B) > 0$ whenever $\lambda(B) > 0$ for any $x \in \mathbb{R}^r \times \mathbb{R}^{u\times p} \times \mathbb{S}_+^{u\times u} \times \mathbb{S}_+^{(r-u)\times(r-u)} \times \mathbb{R}^{(r-u)\times u}$ (see Geyer, 1998, Section 3.1.9), where λ is the Lebesgue measure, B is a Borel set on $\mathbb{R}^r \times \mathbb{R}^{u\times p} \times \mathbb{S}_+^{u\times u} \times \mathbb{S}_+^{(r-u)\times(r-u)} \times \mathbb{R}^{(r-u)\times u}$, and $K(\cdot,\cdot)$ denotes the Markov transition kernel of the chain associated with Algorithm 3.1. This proves ϕ-irreducibility of the full chain, with the Lebesgue measure λ being taken as the underlying measure ϕ.

The fact that $K(x,B) > 0$ whenever $\lambda(B) > 0$, for any $x \in \mathbb{R}^r \times \mathbb{R}^{u\times p} \times \mathbb{S}_+^{u\times u} \times \mathbb{S}_+^{(r-u)\times(r-u)} \times \mathbb{R}^{(r-u)\times u}$ and any Borel set B on $\mathbb{R}^r \times \mathbb{R}^{u\times p} \times \mathbb{S}_+^{u\times u} \times \mathbb{S}_+^{(r-u)\times(r-u)} \times \mathbb{R}^{(r-u)\times u}$, i.e., each measurable set with positive Lebesgue measure can be accessed in a single step from any point, ensures aperiodicity (see, e.g., Dutta, 2012).

Finally, to establish Harris recurrence for the chain, we first note that when $u = 0$ and/or $u = r$, the Algorithm is a ϕ-irreducible Gibbs sampler, which is known to be Harris recurrent (see, e.g.,

Tierney, 1994 and Roberts and Rosenthal, 2006, Corollary 13). When $1 \leq u \leq r-1$, we invoke Corollary 18 and Proposition 15 in Roberts and Rosenthal (2006), and show that the target posterior density $\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$ is integrable with respect to any $1 \leq k \leq 4+u$ elements of the set $\{\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_u\}$, where $\boldsymbol{a}_j$ denotes the jth column of $\boldsymbol{A}$[2]. From (S10), it follows that for an appropriate (finite, due to posterior propriety) constant C_0,

$$
\begin{aligned}
&\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= C_0 \, \exp\left\{-\frac{1}{2}\left[n\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c+\boldsymbol{M}\right)\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(n+p+\nu+u+1)/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Psi}+\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})\right)\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}_0|^{-(n+\nu_0+(r-u)+1)/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Psi}_0+\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})\right)\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\} \\
&\leq C_0 \exp\left\{-\frac{1}{2}\left[n\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c+\boldsymbol{M}\right)\left(\boldsymbol{\eta}-\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(n+p+\nu+u+1)/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\right]\right\} \times |\boldsymbol{\Omega}_0|^{-(n+\nu_0+(r-u)+1)/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\} \\
&=: f_0(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}), \text{ say}
\end{aligned}
$$

where $\check{\boldsymbol{e}} = \left(\mathbb{Y}_c^T\mathbb{X}_c + \boldsymbol{e}\boldsymbol{M}\right)\left(\mathbb{X}_c^T\mathbb{X}_c+\boldsymbol{M}\right)^{-1}$ and $\widetilde{\boldsymbol{G}} = \mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}\boldsymbol{M}\boldsymbol{e}^T - \check{\boldsymbol{e}}\left(\mathbb{X}_c^T\mathbb{X}_c+\boldsymbol{M}\right)\check{\boldsymbol{e}}^T$, and the inequality follows from the fact that $\widetilde{\boldsymbol{G}}$ (see, e.g., Khare et al., 2017, Lemma 2), $\boldsymbol{\Gamma}^T(\boldsymbol{A})\,\widetilde{\boldsymbol{G}}\,\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\,\mathbb{Y}_c^T\mathbb{Y}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})$ are all positive semi-definite matrices. Therefore, the Lebesgue integrability of $\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$, with respect to one or more parameter blocks, is ensured by that of $f_0(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$. Now, for any $\widetilde{\boldsymbol{\mu}} \in \mathbb{R}^r$

$$
\exp\left\{-\frac{1}{2}\left[n\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)\right]\right\} \leq 1
$$

and

$$
\int \exp\left\{-\frac{1}{2}\left[n\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}}-\overline{\boldsymbol{Y}}\right)\right]\right\} d\widetilde{\boldsymbol{\mu}}
$$

[2]It is to be noted that the generic Metropolis within Gibbs samplers considered in Roberts and Rosenthal (2006) have state spaces $\mathscr{X} \subset \mathbb{R}^d$, and the samplers update one of the d (univariate) coordinates at a time. However, their Harris recurrence results can be easily extended to algorithms that update the coordinates block-wise. Moreover, although Theorem 16, Corollary 17 and Corollary 18 in Roberts and Rosenthal (2006) only consider random-scan Metropolis within Gibbs samplers, these can be extended to deterministic or mixed scan samplers, without altering any arguments in the original proof. For example, in a deterministic or mixed scan sampler, the set $C_{n,r}$, as defined in the original proof of Theorem 16 in Roberts and Rosenthal (2006), can be empty for some values of n. This is because for deterministic/mixed scan samplers, I_n, the direction of the proposed move of the full chain at time n, can possibly take only one deterministic value among $\{1, \cdots, d\}$, d being the number of blocks (coordinates) in the chain, so that $\{I_n = d\}$ is either a sure or an impossible event. However, the subsequent statement $\cup_{n,r=1}^{\infty} C_{n,r} = H$ remains intact, as it considers all possible n's. This ensures validity of the results even for deterministic/mixed scan algorithms.

$$= (2n\pi)^{r/2}|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Omega}_0|^{1/2}.$$

Therefore, for any $\widetilde{\boldsymbol{\mu}} \in \mathbb{R}^r$, $f_0(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \leq f_1^{(0)}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$, and $\int f_0(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\widetilde{\boldsymbol{\mu}} = f_1^{(1)}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$, where

$$\begin{aligned}
&f_1^{(\xi)}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= (2n\pi)^{\xi r/2} C_0 \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(n-\xi+p+\nu+u+1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\right]\right\} \times |\boldsymbol{\Omega}_0|^{-(n-\xi+\nu_0+(r-u)+1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]\right\}
\end{aligned}$$

for $\xi = 0, 1$. Again, for any $\boldsymbol{\eta} \in \mathbb{R}^{u\times p}$

$$\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]\right\} \leq 1$$

and

$$\begin{aligned}
&\int \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]\right\} d\boldsymbol{\eta} \\
&= (2\pi)^{up/2}|\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}|^{-u/2}|\boldsymbol{\Omega}|^{p/2}.
\end{aligned}$$

Therefore, for all $\boldsymbol{\eta} \in \mathbb{R}^{u\times p}$, $f_1^{(\xi)}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \leq f_2^{(\xi,0)}(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$ and $\int f_1^{(\xi)}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\eta} = f_2^{(\xi,1)}(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$, where

$$\begin{aligned}
&f_2^{(\xi,\kappa)}(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= n^{\xi r/2}(2\pi)^{(\xi r+\kappa up)/2}|\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}|^{-\xi u/2}\, C_0\, |\boldsymbol{\Omega}|^{-(n-\xi+(1-\kappa)p+\nu+u+1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}_0|^{-(n-\xi+\nu_0+(r-u)+1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]\right\}
\end{aligned}$$

for $\kappa = 0, 1$, and hence it is enough to show that $f_2^{(\xi,\kappa)}(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$ is Lebesgue integrable with respect to any $1 \leq k \leq 2 + u$ elements of $\{\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_u\}$ for $\xi, \kappa = 0, 1$. The proof is completed by first noting that $f_2^{(\xi,\kappa)}$ is proportional to the product of $\mathrm{IW}_u(\boldsymbol{\Psi}, n - \xi + (1-\kappa)p + \kappa + \nu)$ density at $\boldsymbol{\Omega}$, $\mathrm{IW}_u(\boldsymbol{\Psi}_0, n - \xi + \nu_0)$ density $\boldsymbol{\Omega}_0$, and $\mathrm{MN}_{r-u,u}(\boldsymbol{A}_0, \boldsymbol{K}, \boldsymbol{L})$ density at $\boldsymbol{A}$, and that moments of all (nonnegative) orders exist for a (matrix) normal distribution.

□

C.6 Conditional posterior distributions for the parameter blocks in the response envelope model (6)

A Metropolis-within-Gibbs MCMC sampler for the posterior density (S4) is obtained by the full conditional and collapsed marginal distributions of various parameter blocks. We consider the transformed parameters $(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A})$ as given in (S5) for posterior simulation. At each iteration, given $\widetilde{\boldsymbol{\mu}}$, $\boldsymbol{\mu}$ is obtained from the transformation $\boldsymbol{\mu} = \widetilde{\boldsymbol{\mu}} + \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\overline{\boldsymbol{X}}$.

Consider the following factorization of the joint posterior density of $(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A})$ as obtained from (S10):

$$\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) = \pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y})\ \pi(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

A collapsed Metoropolis-within-Gibbs sampler utilizing the above factorization generates Markov chain samples from the joint posterior density by first generating MCMC samples for $(\boldsymbol{A}, (\boldsymbol{\Omega}, \boldsymbol{\Omega}_0))$ from their (joint) marginal posterior density, and subsequently generating exact samples for $(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta})$ from the corresponding joint conditional posterior density given $(\boldsymbol{A}, (\boldsymbol{\Omega}, \boldsymbol{\Omega}_0))$. For generating MCMC samples for $(\boldsymbol{A}, (\boldsymbol{\Omega}, \boldsymbol{\Omega}_0))$ one iteratively updates $(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0)$ and $\boldsymbol{A}$ by drawing exact and MCMC samples from the associated conditional posterior densities respectively. In particular, $(\boldsymbol{\Omega}, \boldsymbol{\Omega}_0)$ is updated by sampling from the density $\pi((\boldsymbol{\Omega}, \boldsymbol{\Omega}_0) \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) = \pi(\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})\pi(\boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})$ (i.e., $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ are independently generated conditional on $\boldsymbol{A}, \mathbb{X}, \mathbb{Y}$). On the other hand, $\boldsymbol{A}$ is updated by Markov chain sampling from the conditional density $\pi(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y})$ (see the following section). Finally, given $(\boldsymbol{A}, (\boldsymbol{\Omega}, \boldsymbol{\Omega}_0))$, $(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta})$ are generated by exact sampling from $\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) = \pi(\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\ \pi(\boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y})$ (i.e., $\widetilde{\boldsymbol{\mu}}$ and $\boldsymbol{\eta}$ are independently generated given $(\boldsymbol{A}, (\boldsymbol{\Omega}, \boldsymbol{\Omega}_0))$. In the following explicit forms of the relevant conditional distributions are provided. Except for $\boldsymbol{A}$, all other relevant conditional distributions are standard distributions from which direct exact random sampling is possible.

1. The conditional density $\pi(\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y})$ is given by

   $$\log \pi(\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) = \text{const.} - \frac{1}{2}\left[n\left(\widetilde{\boldsymbol{\mu}} - \overline{\boldsymbol{Y}}\right)^T \left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\widetilde{\boldsymbol{\mu}} - \overline{\boldsymbol{Y}}\right)\right]$$

   which means

   $$\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y} \sim \mathrm{N}_r\left(\overline{\boldsymbol{Y}}, n^{-1}\boldsymbol{\Sigma}\right).$$

2. The posterior conditional density $\pi(\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y})$ is given by

   $$\log \pi(\boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) = \text{const.} - \frac{1}{2}\,\text{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)\left(\boldsymbol{\eta} - \boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\right]$$

   where $\check{\boldsymbol{e}}$ is as given in (S9). This means

   $$\boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \sim \mathrm{MN}_{u,p}\left(\boldsymbol{\Gamma}(\boldsymbol{A})^T\check{\boldsymbol{e}},\ \boldsymbol{\Omega},\ \left(\mathbb{X}_c^T\mathbb{X}_c + \boldsymbol{M}\right)^{-1}\right).$$

3. The conditional density $\pi(\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})$ is given by

   $$\log \pi(\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) = \text{const.} - \frac{(n-1+\nu)+u+1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\,\text{trace}\left(\widetilde{\boldsymbol{\Psi}}\,\boldsymbol{\Omega}^{-1}\right)$$

   where $\widetilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\ \widetilde{\boldsymbol{G}}\ \boldsymbol{\Gamma}(\boldsymbol{A})$, and $\widetilde{\boldsymbol{G}}$ is as defined in (S9). This means

   $$\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y} \sim \mathrm{IW}_u\left(\widetilde{\boldsymbol{\Psi}}, n-1+\nu\right).$$

4. The posterior conditional density $\pi(\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})$ is given by

   $$\log \pi(\boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) = \text{const.} - \frac{(n-1+\nu_0)+r-u+1}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\,\text{trace}\left(\widetilde{\boldsymbol{\Psi}}_0\,\boldsymbol{\Omega}_0^{-1}\right)$$

   where $\widetilde{\boldsymbol{\Psi}}_0 = \boldsymbol{\Psi}_0 + \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\ \boldsymbol{Y}_c^T\boldsymbol{Y}_c\ \boldsymbol{\Gamma}_0(\boldsymbol{A})$. This implies

   $$\boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y} \sim \mathrm{IW}_{r-u}\left(\widetilde{\boldsymbol{\Psi}}_0, n-1+\nu_0\right).$$

5. The posterior conditional density $\pi(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y})$ is given by

$$\log \pi(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) = \text{const.} - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A})\, \widetilde{\boldsymbol{G}}\, \boldsymbol{\Gamma}(\boldsymbol{A})\right] - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\, \mathbb{Y}_c^T \mathbb{Y}_c\, \boldsymbol{\Gamma}_0(\boldsymbol{A})\right] - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T\right]. \tag{S13}$$

This density does not correspond to any standard distribution, and i.i.d. rejection sampling will be inefficient due to high-dimensionality. Instead a Metropolis sampling scheme is adopted; details are provided in the following section.

C.7 Metropolis-Hastings sampler for sampling from the conditional posterior density of A as given in (S13)

Metropolis-Hastings with random-walk proposal for column-wise updates. A simple random-walk Metropolis scheme for updating $\boldsymbol{A}$ column-wise from the target conditional posterior density $\pi(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y})$ are described as follows. For a specific $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ write the conditional posterior density of $\pi(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y})$ as $\exp[g(\boldsymbol{A})]$. Let $\boldsymbol{a}_j \in \mathbb{R}^{r-u}$ denote the j-th column of $\boldsymbol{A}$, $j = 1, \ldots, u$. Given the tuning parameter $\tau > 0$, for $j = i_1, \ldots, i_u$, where $\{i_1, \ldots, i_u\}$ denotes a random permutation of $\{1, \ldots, u\}$, do the following:

1. Generate $\boldsymbol{a}_j^* \sim \mathrm{N}_{r-u}(\boldsymbol{a}_j, \tau^2 \boldsymbol{I}_{r-u})$. Replace the j-th column of $\boldsymbol{A}$ by $\boldsymbol{a}_j^*$ and call the resulting matrix $\boldsymbol{A}^*$. Calculate $\rho(\boldsymbol{A}^*, \boldsymbol{A}) = \exp\left[g(\boldsymbol{A}^*) - g(\boldsymbol{A})\right]$.

2. Perform a Bernoulli experiment with probability of success $\min[1, \rho(\boldsymbol{A}^*, \boldsymbol{A})]$. If a success is achieved, update $\boldsymbol{a}_j$ to $\boldsymbol{a}_j^*$; otherwise retain $\boldsymbol{a}_j$.

Metropolis-Hastings with likelihood-driven proposal. The random-walk proposal is computationally efficient, but requires tuning and may suffer from poor mixing in high dimensions if not well-tuned. Here we provide a "likelihood-driven" proposal that requires no tuning and achieves good mixing even in high-dimension, particularly when a vague prior distribution is used on $\boldsymbol{A}$. The proposal utilizes the 2-dimensional generalized matrix Bingham sampling scheme proposed in Khare et al. (2017), while properly exploiting the Euclidean identification of the envelope parameter (and in addition uses a transformation on the von Mises distribution for sampling; see below). This collectively reduces the computation load substantially compared to the Khare et al. (2017) implementation, especially when the envelope dimension u is small relative to the number of responses r (see the run time comparisons provided in Section C.13 of this supplement), while producing a potentially substantially better mixing chain than a sub-optimally tuned random walk sampler.

Consider the likelihood-part of the conditional posterior density $\pi(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y})$ of $\boldsymbol{A}$ given $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$:

$$q(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) \propto \exp\left(-\frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A})\, \widetilde{\boldsymbol{G}}\, \boldsymbol{\Gamma}(\boldsymbol{A})\right] - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\, \mathbb{Y}_c^T \mathbb{Y}_c\, \boldsymbol{\Gamma}_0(\boldsymbol{A})\right]\right)$$

For Metropolis update of $\boldsymbol{A}$ we seek to generate proposals by drawing (Gibbs) samples from the density $q(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y})$.

Now consider eigen-decompositions of $\boldsymbol{\Omega} = \boldsymbol{P} \boldsymbol{\Lambda} \boldsymbol{P}^T$ and $\boldsymbol{\Omega}_0 = \boldsymbol{P}_0 \boldsymbol{\Lambda}_0 \boldsymbol{P}_0^T$ where $\boldsymbol{P} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{P}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are orthogonal matrices and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_u)$ and $\boldsymbol{\Lambda}_0 = \operatorname{diag}(\lambda_{u+1}, \ldots, \lambda_r)$ are diagonal matrices carrying eigenvalues of $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ respectively. Then

$$q(\boldsymbol{A} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) \propto \exp\left(-\frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Lambda}^{-1} \widetilde{\boldsymbol{\Gamma}}^T(\boldsymbol{A})\, \widetilde{\boldsymbol{G}}\, \widetilde{\boldsymbol{\Gamma}}(\boldsymbol{A})\right] - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Lambda}_0^{-1} \widetilde{\boldsymbol{\Gamma}}_0^T(\boldsymbol{A})\, \mathbb{Y}_c^T \mathbb{Y}_c\, \widetilde{\boldsymbol{\Gamma}}_0(\boldsymbol{A})\right]\right)$$

$$= \exp\left(-\sum_{l=1}^{r} \widetilde{\boldsymbol{O}}_{\cdot l}^T \boldsymbol{H}_l \widetilde{\boldsymbol{O}}_{\cdot l}^T\right)$$

where $\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{A}) = \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{P}$ and $\widetilde{\boldsymbol{\Gamma}}_0(\boldsymbol{A}) = \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{P}_0$ are both semi-orthogonal matrices, $\widetilde{\boldsymbol{O}} = (\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{A}), \widetilde{\boldsymbol{\Gamma}}_0(\boldsymbol{A}))$, $\widetilde{\boldsymbol{O}}_{\cdot l}$ is the l-th column of $\widetilde{\boldsymbol{O}}$, $l = 1, \ldots, r$, and $\boldsymbol{H}_l = \widetilde{\boldsymbol{G}}/(2\lambda_l)$ for $l = 1, \ldots, u$ and $\boldsymbol{H}_l = \mathbb{Y}_c^T\mathbb{Y}_c/(2\lambda_l)$ for $l = u+1, \ldots, r$. A proposal for $\boldsymbol{A}$ may thus be generated through $\widetilde{\boldsymbol{O}}$ having density $q(\widetilde{\boldsymbol{O}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) \propto \exp\left(-\sum_{l=1}^{r} \widetilde{\boldsymbol{O}}_{\cdot l}^T \boldsymbol{H}_l \widetilde{\boldsymbol{O}}_{\cdot l}^T\right)$. We observe the notable similarity of this density with the density of the orthogonal envelope parameter matrix considered in Khare et al. (2017). Hence for random generation from the above density we adapt the Khare et al. (2017) approach, with modifications that utilize the Euclidean matrix $\boldsymbol{A}$-based identification of the envelope space.

Remark C.1. Compared to the Khare et al. (2017) approach, our approach can be substantially more efficient in applications. In Khare et al. (2017) a Gibbs sampling scheme is proposed that updates $\widetilde{\boldsymbol{O}}$ by iteratively updating all $\binom{r}{2}$ pairs of columns. In our current parameterization, however, both $\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\boldsymbol{\Gamma}_0(\boldsymbol{A})$ uniquely determine $\boldsymbol{A}$ (see (3) in Section 3.1 of the main text), and hence only updating the first u columns of $\widetilde{\boldsymbol{O}}$ (corresponding to $\boldsymbol{\Gamma}(\boldsymbol{A})$), or alternatively the last $r-u$ columns of $\widetilde{\boldsymbol{O}}$ (corresponding to $\boldsymbol{\Gamma}_0(\boldsymbol{A})$) is sufficient. In applications this can reduce the computation burden substantially, particularly when u (or $r-u$) is much smaller than r. For example, when u (or $r-u$) is much smaller than r, one only needs to consider $\binom{u}{2} + u(r-u)$ (or $\binom{r-u}{2} + (r-u)u$) pairs of columns in $\widetilde{\boldsymbol{O}}$ that contain at least one of the first u (or the last $r-u$) columns of $\widetilde{\boldsymbol{O}}$. This saves $\binom{r-u}{2}$ pairs (or $\binom{u}{2}$ pairs if the last $r-u$ columns are updated) as compared to the Khare et al. (2017) approach.

Now to update a specific pair (i, j) of columns of $\widetilde{\boldsymbol{O}}$, consider the conditional density of $(\widetilde{\boldsymbol{O}}_{\cdot i}, \widetilde{\boldsymbol{O}}_{\cdot j})$ conditional on the remaining columns (denoted by $\widetilde{\boldsymbol{O}}_{\cdot,-(i,j)}$):

$$q(\widetilde{\boldsymbol{O}}_{\cdot i}, \widetilde{\boldsymbol{O}}_{\cdot i} \mid \widetilde{\boldsymbol{O}}_{\cdot,-(i,j)}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) \propto \exp\left[-\widetilde{\boldsymbol{O}}_{\cdot i}^T \boldsymbol{H}_i \widetilde{\boldsymbol{O}}_{\cdot i}^T - \widetilde{\boldsymbol{O}}_{\cdot j}^T \boldsymbol{H}_j \widetilde{\boldsymbol{O}}_{\cdot j}^T\right]$$

Following Khare et al. (2017), we sample from the above density by first generating a 2×2 orthogonal matrix $\boldsymbol{Z}$ from the density

$$q(\boldsymbol{Z} \mid \widetilde{\boldsymbol{O}}_{\cdot,-(i,j)}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) \propto \exp\left(-\boldsymbol{Z}_{\cdot,1}^T \boldsymbol{N}_{\cdot,(i,j)}^T \boldsymbol{H}_i \boldsymbol{N}_{\cdot,(i,j)} \boldsymbol{Z}_{\cdot,1} - \boldsymbol{Z}_{\cdot,2}^T \boldsymbol{N}_{\cdot,(i,j)}^T \boldsymbol{H}_j \boldsymbol{N}_{\cdot,(i,j)} \boldsymbol{Z}_{\cdot,2}\right)$$

and then defining $\widetilde{\boldsymbol{O}}_{\cdot i}^{\text{new}} = \boldsymbol{N}_{\cdot,(i,j)} \boldsymbol{Z}_{\cdot,1}$ and $\widetilde{\boldsymbol{O}}_{\cdot j}^{\text{new}} = \boldsymbol{N}_{\cdot,(i,j)} \boldsymbol{Z}_{\cdot,2}$. Here $\boldsymbol{Z}_{\cdot,k}$ denotes the k-th column of $\boldsymbol{Z}$ ($k = 1, 2$) and $\boldsymbol{N}_{\cdot,(i,j)} \in \mathbb{R}^{r \times 2}$ is a semi-orthogonal matrix whose columns are orthogonal to $\widetilde{\boldsymbol{O}}_{\cdot,-(i,j)}$. Note that a convenient specific choice of $\boldsymbol{N}_{\cdot,(i,j)}$ is $\widetilde{\boldsymbol{O}}_{\cdot,(i,j)} = (\widetilde{\boldsymbol{O}}_{\cdot,i}, \widetilde{\boldsymbol{O}}_{\cdot,j})$, the current values of $\widetilde{\boldsymbol{O}}_{\cdot,i}$ and $\widetilde{\boldsymbol{O}}_{\cdot,j}$. The above density is called the 2×2 generalized matrix Bingham density ($\text{GB}_{2,2}(\boldsymbol{A}, \boldsymbol{B})$ with $\boldsymbol{A} = \boldsymbol{N}_{\cdot,(i,j)}^T \boldsymbol{H}_i \boldsymbol{N}_{\cdot,(i,j)}$ and $\boldsymbol{B} = \boldsymbol{N}_{\cdot,(i,j)}^T \boldsymbol{H}_j \boldsymbol{N}_{\cdot,(i,j)}$) in Khare et al. (2017) who propose a rejection-sampling approach to random generation from the above density. In the following section we provide an alternative sampling scheme utilizing a von Mises connection to the density that is computationally more efficient than this rejection sampler.

Once all column-pairs (i, j) under consideration (say, all pairs with at least one column from the first u columns of $\widetilde{\boldsymbol{O}}$) have been updated, this process produces an updated $\widetilde{\boldsymbol{O}}^{\text{new}}$ matrix. The corresponding $\boldsymbol{\Gamma}(\boldsymbol{A})^{\text{new}}$ is obtained from the first u columns $\widetilde{\boldsymbol{O}}_{\cdot,(1:u)}^{\text{new}}$ of $\widetilde{\boldsymbol{O}}^{\text{new}}$ as $\boldsymbol{\Gamma}(\boldsymbol{A})^{\text{new}} = \widetilde{\boldsymbol{O}}_{\cdot,(1:u)}^{\text{new}} \boldsymbol{P}^T$ where $\boldsymbol{P}$ is the same eigenvector matrix as obtained in the eigen decomposition of $\boldsymbol{\Omega}$. From $\boldsymbol{\Gamma}(\boldsymbol{A})^{\text{new}}$, the proposal of $\boldsymbol{A}$, namely, $\boldsymbol{A}^*$ is obtained through the reverse transformation as described in Section 3.1 in the main text.

Once the proposal $\boldsymbol{A}^*$ is obtained, a Metropolis-Hasting step is performed. This entails calculating acceptance ratio

$$\rho(\boldsymbol{A}, \boldsymbol{A}^*) = \exp\left[g(\boldsymbol{A}^*) - g(\boldsymbol{A})\right.$$

$$+\sum_{i,j}\Big(\log\pi(\widetilde{\boldsymbol{O}}_{\cdot i},\widetilde{\boldsymbol{O}}_{\cdot j}\mid\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot i},\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot j})-\log\pi(\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot i},\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot j}\mid\widetilde{\boldsymbol{O}}_{\cdot i},\widetilde{\boldsymbol{O}}_{\cdot j})\Big)$$
$$+\log|J(\widetilde{\boldsymbol{O}}^{\text{new}})|-\log|J(\widetilde{\boldsymbol{O}})|\Big]$$

then performing a Bernoulli experiment with probability of success $\min\{1,\rho(\boldsymbol{A},\boldsymbol{A}^*)\}$, and finally updating $\boldsymbol{A}$ to $\boldsymbol{A}^*$ if a success is achieved in the Bernoulli experiment, or otherwise retaining the current value of $\boldsymbol{A}$. Here $|J(\widetilde{\boldsymbol{O}})|$ denotes the denotes the Jacobian of the transformation $\boldsymbol{A}\mapsto\widetilde{\boldsymbol{O}}=(\boldsymbol{\Gamma}(\boldsymbol{A}),\boldsymbol{\Gamma}_0(\boldsymbol{A}))$, defined as matrix-volume of the Jacobian matrix: $\frac{\partial\,\text{vec}(\widetilde{\boldsymbol{O}})}{\partial^T\,\text{vec}(\boldsymbol{A})}$. Note that the transformation is trans-dimensional, and hence determinant of the Jacobian matrix does not exist; instead we use matrix volume (Ben-Israel, 1999). The matrix volume of a full column rank rectangular matrix $\boldsymbol{B}\in\mathbb{R}^{m\times n}$ is defined as square-root of the determinant of $\boldsymbol{B}^T\boldsymbol{B}$. We provide analytic expression for the Jacobian of the transformation in a following section. Note that ignoring the Jacobian adjustment in the Metropolis-Hastings acceptance ratio effectively amounts to sampling from the conditional posterior of $\boldsymbol{O}=(\boldsymbol{\Gamma}(\boldsymbol{A}),\boldsymbol{\Gamma}_0(\boldsymbol{A}))$ instead of the conditional posterior of $\boldsymbol{A}$. The posterior of $\boldsymbol{O}=(\boldsymbol{\Gamma}(\boldsymbol{A}),\boldsymbol{\Gamma}_0(\boldsymbol{A}))$ is associated with the prior on $\boldsymbol{O}$ induced by the (matrix normal) prior on $\boldsymbol{A}$. More specifically, the induced prior on $\boldsymbol{O}$ is given by: $\pi(\boldsymbol{O})\propto\exp\left(-\frac{1}{2}\,\text{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right)|J(\boldsymbol{O})|$. If the final objective is to make inference on the regression parameter $\boldsymbol{\beta}$, sampling can alternatively be done on the $\boldsymbol{O}$-scale. This aids computational efficiency on the sampler. In our computational experiments performed in this paper, we elected to ignore the Jacobian adjustment to achieve computational efficiency, and to aid sampling in the $\boldsymbol{\Gamma}(\boldsymbol{A})$ parameterization.

The Gibbs transition densities $\pi(\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot i},\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot j}\mid\widetilde{\boldsymbol{O}}_{\cdot i},\widetilde{\boldsymbol{O}}_{\cdot j})$ are obtained from the 2×2 generalized matrix Bingham densities. In particular, $\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot,(i,j)}=\widetilde{\boldsymbol{O}}_{\cdot,(i,j)}\boldsymbol{Z}$ where $\boldsymbol{Z}\sim\text{GB}_{2,2}(\boldsymbol{A}=\widetilde{\boldsymbol{O}}^T_{\cdot,(i,j)}\boldsymbol{H}_i\widetilde{\boldsymbol{O}}_{\cdot,(i,j)},\boldsymbol{B}=\widetilde{\boldsymbol{O}}^T_{\cdot,(i,j)}\boldsymbol{H}_j\widetilde{\boldsymbol{O}}_{\cdot,(i,j)})$ implies that the conditional density of $\widetilde{\boldsymbol{O}}^{\text{new}}_{\cdot,(i,j)}$ given the current $\widetilde{\boldsymbol{O}}_{\cdot,(i,j)}$ is simply the above $\text{GB}_{2,2}$ density evaluated at the observed value of $\boldsymbol{Z}$, conditional on the current value of $\widetilde{\boldsymbol{O}}_{\cdot,(i,j)}$). (The Jacobian of transformation $\boldsymbol{Z}\mapsto\widetilde{\boldsymbol{O}}_{\cdot,(i,j)}$ has matrix volume 1 as $\widetilde{\boldsymbol{O}}_{\cdot,(i,j)}$ is semi-orthogonal). See the following subsection for an explicit form of the $\text{GB}_{2,2}(\boldsymbol{A},\boldsymbol{B})$ density (including the normalizing constant), along with a von-Mises based random sampling scheme from the density.

Remark C.2. Under the aforementioned random-walk Metropolis scheme, the columns of $\boldsymbol{A}$ can be updated in a systematic or in a random order, thereby producing a fixed scan and a mixed / random scan sampler respectively. (Note that a mixed scan means that the columns of $\boldsymbol{A}$ are updated randomly while the other parameters are updated systematically.) Since for any Gibbs-type sampler, a fixed-scan variant and a random-scan variant both converge to the same target distribution (Geyer, 1998), an MCMC chain generated by updating columns of $\boldsymbol{A}$ randomly, and one generated by updating columns of $\boldsymbol{A}$ systematically, are both guaranteed to converge to the same target posterior distribution.

C.8 Explicit closed form expression for the generalized matrix Bingham $\text{GB}_{2,2}(\mathbf{A},\mathbf{B})$ density and exact random sampling scheme using von Mises distribution

A random 2×2 orthogonal matrix $\boldsymbol{Z}=(\boldsymbol{Z}_{\cdot1},\boldsymbol{Z}_{\cdot1})$ is said to follow the 2×2 generalized matrix Bingham $\text{GB}_{2,2}(\boldsymbol{A},\boldsymbol{B})$ distribution if $\boldsymbol{Z}$ has density

$$f(\boldsymbol{Z}\mid\boldsymbol{A},\boldsymbol{B})\propto\exp\left[-\boldsymbol{Z}^T_{\cdot1}\boldsymbol{A}\boldsymbol{Z}^T_{\cdot1}-\boldsymbol{Z}^T_{\cdot2}\boldsymbol{B}\boldsymbol{Z}^T_{\cdot2}\right].$$

Since $\boldsymbol{Z}$ is an arbitrary 2×2 orthogonal matrix, we can write $\boldsymbol{Z}$ as (Khare et al., 2017)

$$\boldsymbol{Z}=\begin{pmatrix}\cos\theta & s\sin\theta\\ \sin\theta & -s\cos\theta\end{pmatrix}$$

where $\theta \in [-\pi, \pi)$ and $s \in \{-1, 1\}$ are arbitrary. Using transformations it is straightforward to show that in the resulting density θ are s are independent with $s \sim$ Discrete-Uniform$(\{-1, 1\})$, and θ having the marginal density

$$f(\theta \mid \boldsymbol{A}, \boldsymbol{B}) \propto \exp[-(\boldsymbol{A}_{1,1} + \boldsymbol{B}_{2,2}) \cos^2\theta - (\boldsymbol{A}_{2,2} + \boldsymbol{B}_{1,1}) \sin^2\theta - (\boldsymbol{A}_{1,2} + \boldsymbol{A}_{2,1} + \boldsymbol{B}_{2,2} + \boldsymbol{B}_{2,2}) \sin\theta\cos\theta].$$

Write $a = -(\boldsymbol{A}_{1,1} + \boldsymbol{B}_{2,2})$, $b = -(\boldsymbol{A}_{2,2} + \boldsymbol{B}_{1,1})$, and $c = -(\boldsymbol{A}_{1,2} + \boldsymbol{A}_{2,1} + \boldsymbol{B}_{2,2} + \boldsymbol{B}_{2,2})$. Then the exponent in the above density is given by

$$\begin{aligned} a\cos^2\theta + b\sin^2\theta + c\sin\theta\cos\theta &= \frac{a}{2}\,(1 + \cos 2\theta) + \frac{b}{2}\,(1 - \cos 2\theta) + \frac{c}{2}\sin 2\theta \\ &= \frac{a+b}{2}\,\cos 2\theta + \frac{c}{2}\,\sin 2\theta + \frac{a+b}{2} = \frac{r}{2}\cos(2\theta - \alpha) + \frac{a+b}{2} \end{aligned}$$

where $r = \sqrt{(a-b)^2 + c^2}$ and $\alpha = \arctan(c/(a-b))$. This implies $f(\theta \mid \boldsymbol{A}, \boldsymbol{B}) \propto \exp[(r/2)\cos(2\theta - \alpha)]$, which in turn implies that 2θ follows a von Mises vM$(\mu = \alpha, \kappa = r/2)$ distribution (Mardia and Jupp, 2009). Hence the complete (including the normalizing constant) density of θ is given by:

$$f(\theta \mid \boldsymbol{A}, \boldsymbol{B}) = \frac{1}{\pi I_0(r/2)} \exp\left[\frac{r}{2}\cos(2\theta - \alpha)\right];\ -\pi \le \theta < \pi.$$

where $I_0(\cdot)$ denotes the modified Bessel function of the first kind of order 0. From this, the complete GB$_{2,2}(\boldsymbol{A}, , \boldsymbol{B})$ density including the normalizing constant can be derived as follows:

$$\begin{aligned} f(\boldsymbol{Z} \mid \boldsymbol{A}, \boldsymbol{B}) &= f(\theta \mid \boldsymbol{A}, \boldsymbol{B})\ f(s \mid \boldsymbol{A}, \boldsymbol{B}) \\ &= \frac{1}{2\pi I_0(r/2)} \exp\left[\frac{r}{2}\cos(2\theta - \alpha)\right] \\ &= \frac{1}{2\pi I_0(r/2)} \exp\left[-\frac{a+b}{2}\right]\ \exp\left[\frac{r}{2}\cos(2\theta - \alpha) + \frac{a+b}{2}\right] \\ &= \frac{1}{2\pi I_0(r/2)} \exp\left[-\frac{a+b}{2}\right]\ \exp\left[-\boldsymbol{Z}_{\cdot 1}^T \boldsymbol{A} \boldsymbol{Z}_{\cdot 1}^T - \boldsymbol{Z}_{\cdot 2}^T \boldsymbol{B} \boldsymbol{Z}_{\cdot 2}^T\right] \\ &= \frac{1}{2\pi I_0(r/2)} \exp\left[\frac{1}{2}\,(\text{trace}\,\boldsymbol{A} + \text{trace}\,\boldsymbol{B})\right]\ \exp\left[-\boldsymbol{Z}_{\cdot 1}^T \boldsymbol{A} \boldsymbol{Z}_{\cdot 1}^T - \boldsymbol{Z}_{\cdot 2}^T \boldsymbol{B} \boldsymbol{Z}_{\cdot 2}^T\right]. \end{aligned}$$

For random generation from the GB$_{2,2}(\boldsymbol{A}, \boldsymbol{B})$ density one can simply use the above von Mises connection. More specifically, one can draw $\phi \sim$ vM$(\mu = \alpha, \kappa = r/2)$ and $s \sim$ Discrete-Uniform$(\{-1, 1\})$ independently, and subsequently define $\theta = \phi/2$ and

$$\boldsymbol{Z} = \begin{pmatrix} \cos\theta & s\sin\theta \\ \sin\theta & -s\cos\theta \end{pmatrix}$$

where α and r are as defined above. The resulting $\boldsymbol{Z}$ will have the target GB$_{2,2}(\boldsymbol{A}, \boldsymbol{B})$ distribution.

C.9 Jacobian of the Transformation $\mathbf{A} \mapsto (\boldsymbol{\Gamma}(\mathbf{A}), \boldsymbol{\Gamma}_\mathbf{0}(\mathbf{A}))$

Consider the transformation $\boldsymbol{A} \mapsto (\boldsymbol{\Gamma}(\boldsymbol{A}), \boldsymbol{\Gamma}_0(\boldsymbol{A}))$, where $\boldsymbol{\Gamma}(\boldsymbol{A}) = \boldsymbol{C}_{\boldsymbol{A}} \left(\boldsymbol{C}_{\boldsymbol{A}}^T \boldsymbol{C}_{\boldsymbol{A}}\right)^{-1/2}$ with $\boldsymbol{C}_{\boldsymbol{A}}^T = (\boldsymbol{I}_u, \boldsymbol{A}^T)$ and $\boldsymbol{\Gamma}_0(\boldsymbol{A}) = \boldsymbol{D}_{\boldsymbol{A}} \left(\boldsymbol{D}_{\boldsymbol{A}}^T \boldsymbol{D}_{\boldsymbol{A}}\right)^{-1/2}$ with $\boldsymbol{D}_{\boldsymbol{A}} = (-\boldsymbol{A}^T, \boldsymbol{I}_{r-u})$, and we are interested in the Jacobian matrix of the transformation $\boldsymbol{A} \mapsto (\boldsymbol{\Gamma}(\boldsymbol{A}), \boldsymbol{\Gamma}_0(\boldsymbol{A}))$ defined as

$$\frac{\partial \text{vec}((\boldsymbol{\Gamma}(\boldsymbol{A}), \boldsymbol{\Gamma}_0(\boldsymbol{A})))}{\partial^T \text{vec}(\boldsymbol{A})} = \left(\frac{\partial \text{vec}(\boldsymbol{\Gamma}(\boldsymbol{A}))}{\partial^T \text{vec}(\boldsymbol{A})}, \frac{\partial \text{vec}(\boldsymbol{\Gamma}_0(\boldsymbol{A}))}{\partial^T \text{vec}(\boldsymbol{A})}\right).$$

Below we provide explicit forms for the two blocks, viz., $\frac{\partial \text{vec}(\boldsymbol{\Gamma}(\boldsymbol{A}))}{\partial^T \text{vec}(\boldsymbol{A})}$ and $\frac{\partial \text{vec}(\boldsymbol{\Gamma}_0(\boldsymbol{A}))}{\partial^T \text{vec}(\boldsymbol{A})}$ are separately. First we have

$$\begin{aligned}\frac{\partial \text{vec}(\boldsymbol{\Gamma}(\boldsymbol{A}))}{\partial^T \text{vec}(\boldsymbol{A})} &= \frac{\partial \text{vec}\left(\boldsymbol{C}_{\boldsymbol{A}}(\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2}\right)}{\partial^T \text{vec}(\boldsymbol{C}_{\boldsymbol{A}})} \frac{\partial \text{vec}(\boldsymbol{C}_{\boldsymbol{A}})}{\partial^T \text{vec}(\boldsymbol{A})} \\ &= \left[(\boldsymbol{I}_u \otimes \boldsymbol{C}_{\boldsymbol{A}}) \frac{\partial \text{vec}\left((\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2}\right)}{\partial^T \text{vec}(\boldsymbol{C}_{\boldsymbol{A}})} + \left((\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2} \otimes \boldsymbol{I}_r\right)\right] \frac{\partial \text{vec}(\boldsymbol{C}_{\boldsymbol{A}})}{\partial^T \text{vec}(\boldsymbol{A})}.\end{aligned}$$

Here

$$\frac{\partial \text{vec}(\boldsymbol{C}_{\boldsymbol{A}})}{\partial^T \text{vec}(\boldsymbol{A})} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} & \dots & \boldsymbol{0} \\ \vdots & & & \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{M} \end{pmatrix} \text{ and } \boldsymbol{M} = \begin{pmatrix} \boldsymbol{0}_{u\times(r-u)} \\ \boldsymbol{I}_{r-u} \end{pmatrix},$$

and

$$\begin{aligned}\frac{\partial \text{vec}\left((\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2}\right)}{\partial^T \text{vec}(\boldsymbol{C}_{\boldsymbol{A}})} = &- \left[(\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2} \otimes \boldsymbol{I}_u + \boldsymbol{I}_u \otimes (\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1/2}\right]^{-1} \times \\ &\left[(\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1} \otimes (\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1}\boldsymbol{C}_{\boldsymbol{A}}^T + (\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1}\boldsymbol{C}_{\boldsymbol{A}}^T \otimes (\boldsymbol{C}_{\boldsymbol{A}}^T\boldsymbol{C}_{\boldsymbol{A}})^{-1}\boldsymbol{K}_{r,u}\right]\end{aligned}$$

where $\boldsymbol{K}_{r,u}$ denotes the commutation matrix of order (r, u). Next, we have

$$\begin{aligned}\frac{\partial \text{vec}(\boldsymbol{\Gamma}_0(\boldsymbol{A}))}{\partial^T \text{vec}(\boldsymbol{A})} &= \frac{\partial \text{vec}\left(\boldsymbol{D}_{\boldsymbol{A}}(\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2}\right)}{\partial^T \text{vec}(\boldsymbol{D}_{\boldsymbol{A}})} \frac{\partial \text{vec}(\boldsymbol{D}_{\boldsymbol{A}})}{\partial^T \text{vec}(\boldsymbol{A})} \\ &= \left[(\boldsymbol{I}_{r-u} \otimes \boldsymbol{D}_{\boldsymbol{A}}) \frac{\partial \text{vec}\left((\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2}\right)}{\partial^T \text{vec}(\boldsymbol{D}_{\boldsymbol{A}})} + \left((\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2} \otimes \boldsymbol{I}_r\right)\right] \frac{\partial \text{vec}(\boldsymbol{D}_{\boldsymbol{A}})}{\partial^T \text{vec}(\boldsymbol{A})}.\end{aligned}$$

Here

$$\frac{\partial \text{vec}(\boldsymbol{D}_{\boldsymbol{A}})}{\partial^T \text{vec}(\boldsymbol{A})} = \frac{\boldsymbol{K}_{(r-u),r}\ \partial \text{vec}(\boldsymbol{D}_{\boldsymbol{A}}^T)}{\partial^T \text{vec}(\boldsymbol{A})} = \boldsymbol{K}_{(r-u),r} \begin{pmatrix} -\boldsymbol{I}_{u(r-u)} \\ \boldsymbol{0}_{(r-u)^2\times u(r-u)} \end{pmatrix}$$

and

$$\begin{aligned}\frac{\partial \text{vec}\left((\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2}\right)}{\partial^T \text{vec}(\boldsymbol{D}_{\boldsymbol{A}})} = &- \left[(\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2} \otimes \boldsymbol{I}_{r-u} + \boldsymbol{I}_{r-u} \otimes (\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1/2}\right]^{-1} \times \\ &\left[(\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1} \otimes (\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1}\boldsymbol{D}_{\boldsymbol{A}}^T + (\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1}\boldsymbol{D}_{\boldsymbol{A}}^T \otimes (\boldsymbol{D}_{\boldsymbol{A}}^T\boldsymbol{D}_{\boldsymbol{A}})^{-1}\boldsymbol{K}_{r,(r-u)}\right].\end{aligned}$$

C.10 Comparison of posterior standard deviation of the Bayesian envelope model and the bootstrap-based standard errors of the frequentist envelope model

We consider numerical experiments with response dimensions $r = 20$. The dimension of the response envelope was fixed at $u = 2$, and the number of predictors was fixed at $p = 5$. The model parameters were generated following the same procedure in Section 7.1 of the main text. Then 100 independent datasets were generated for each sample size $n = 50, 100, 200, 500$ and 1000.

In each simulated dataset, we fitted the proposed Bayesian response envelope model and the frequentist response envelope model (Cook et al., 2016) (as implemented in R package Benvlp (Lee and Su, 2020)). For the Bayesian model, we ran the MCMC samplers for 10000 iterations, after discarding the first 5000 iterations as burn-in. For the frequentist model, we evaluated bootstrap standard errors

using residual bootstrap with 10000 bootstrap samples. The comparison of the posterior standard deviations from the Bayesian model with the bootstrap-based standard errors from the frequentist model for the first element in $\boldsymbol{\beta}$ is included in Table S1. Other elements in $\boldsymbol{\beta}$ follow the same pattern.

n	Posterior SD	Bootstrap SE
50	0.174	0.159
100	0.116	0.113
200	0.079	0.083
500	0.049	0.056
1000	0.035	0.042

Table S1: Comparison of the posterior standard deviation (posterior SD) of the Bayesian envelope model and the bootstrap-based standard errors (bootstrap SE) of the frequentist envelope model.

The posterior standard deviations and the bootstrap standard errors are similar, especially when the sample size is larger. This is not surprising as the Bayesian models considered herein contain vague prior information, which gets dominated by the data (likelihood) information particularly when the sample size is large. Consequently, variability in the corresponding posterior distributions reflect primarily the variability in the likelihood. Such similarity between a Bayesian posteiror standard deviation and a Bootstrap standard error is not specific to the envelope models, and can be observed in practice in various other problems; see e.g., the example and the notes provided on p. 271-272 of Friedman et al. (2001).

C.11 Maximum a posteriori Estimation in response envelope model

A computationally fast point estimator for the Bayesian response envelope model parameter vector is obtained through the mode of the target posterior density (maximum *a posteriori* (MAP) estimation). Algorithm C.1 below provides a strategy for MAP estimation for the model parameters $\{\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Omega}}_0, \widehat{\boldsymbol{A}}\}$. The algorithm utilizes the decomposition in (S10) of the target posterior density:

$$\begin{aligned}\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) = \pi(\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y})\ \pi(\boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) \times \\ \pi(\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})\ \pi(\boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})\ \pi(\boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\end{aligned} \tag{S14}$$

wherein all the (conditional) densities on the right hand side except that for $\boldsymbol{A}$ have standard forms, and closed form expressions for their respective modes are available.

Algorithm C.1. Computation of the MAP estimators $\{\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Omega}}_0, \widehat{\boldsymbol{A}}\}$ for model (6).

S.1 Find $\widehat{\boldsymbol{A}} = \arg\max_{\boldsymbol{A} \in \mathbb{R}^{(r-u)\times u}} l(\boldsymbol{A})$, where

$$\begin{aligned}l(\boldsymbol{A}) = &-\frac{n+\nu-1}{2} \log\left|\boldsymbol{\Gamma}^T(\boldsymbol{A})\, \widetilde{\boldsymbol{G}}\, \boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{\Psi}\right| - \frac{n+\nu_0-1}{2} \log\left|\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\, \mathbb{Y}_c^T \mathbb{Y}_c\, \boldsymbol{\Gamma}_0(\boldsymbol{A}) + \boldsymbol{\Psi}_0\right| \\ &- \frac{1}{2} \operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right].\end{aligned} \tag{S15}$$

Then obtain $\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\boldsymbol{\Gamma}(\boldsymbol{A}_0)$ using transformation as described in (4).

S.2 Compute

$$(i) \quad \widehat{\boldsymbol{\eta}} = \boldsymbol{\Gamma}^T(\widehat{\boldsymbol{A}})\check{\boldsymbol{e}}, \qquad (ii) \quad \widehat{\boldsymbol{\Omega}} = \frac{1}{n+p+\nu+u+1}\left(\boldsymbol{\Gamma}^T(\widehat{\boldsymbol{A}})\widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}(\widehat{\boldsymbol{A}}) + \boldsymbol{\Psi}\right),$$

$$(iii)\quad \widehat{\boldsymbol{\Omega}}_0 = \frac{1}{n+\nu_0+r-u+1}\left(\boldsymbol{\Gamma}_0^T(\widehat{\boldsymbol{A}})\mathbb{Y}_c^T\mathbb{Y}_c\boldsymbol{\Gamma}_0(\widehat{\boldsymbol{A}})+\boldsymbol{\Psi}_0\right),\quad (iv)\quad \widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{Y}} - \boldsymbol{\Gamma}^T(\widehat{\boldsymbol{A}})\widehat{\boldsymbol{\eta}}\overline{\boldsymbol{X}}.$$

where $\widetilde{\boldsymbol{G}}$ and $\check{\boldsymbol{e}}$ are as defined in (S9).

Remark C.3. The objective function $l(\boldsymbol{A})$ in Step S.1 is smooth, and can be solved using the Nelder-Mead algorithm or gradient based optimization. Starting value of $\boldsymbol{A}$ can be set at one of the four $\sqrt{n}$-consistent starting values considered in Cook et al. (2016).

To test the computational cost, we considered a numerical experiment using the settings in Section 7.1 of the main text as a starting point. Since the envelope subspace is on an $r \times u$ Grassmann manifold, we increased both r and u and studied four scenarios (a) $r = 20$, $u = 2$ (setting in Section 6.1), (b) $r = 20$, $u = 10$, (c), $r = 70$, $u = 2$ and (d) $r = 70$, $u = 10$. Other parameters were generated the same way as in Section 6.1 of the main text. For each scenario, we simulated 100 replicates, and obtained the MAP estimates of the model parameters, with the objective function for $\boldsymbol{A}$ in Step S.1 being optimized using the Nelder-Mead algorithm. The median computing time as well as the average mean squared errors (MSE) $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{true}\|^2/(rp)$ were recorded. The computing time is the total time for obtaining the MAP estimators, i.e. for both Step S.1 and Step S.2, while the computing time for Step S.2 is miniscule. The results are summarized in Table S2.

r	u	n	Median time (in min)	MSE
20	2	100	0.0093	0.0107
		200	0.0086	0.0053
		500	0.0085	0.0020
		1000	0.0082	0.0010
20	10	100	0.0430	0.0198
		200	0.0389	0.0093
		500	0.0346	0.0040
		1000	0.0319	0.0021
70	2	100	0.1063	0.0158
		200	0.0890	0.0069
		500	0.0870	0.0026
		1000	0.0830	0.0013
70	10	100	0.8087	0.0540
		200	0.5820	0.0232
		500	0.5188	0.0081
		1000	0.4675	0.0041

Table S2: Computing time and average MSE of the MAP estimator

As expected, the MAP estimators are indeed orders of magnitude faster to compute than MCMC based approximation of the posterior. The median computing times for the MCMC chains for scenario (a) were 5.43, 7.53, 6.02, and 7.94 minutes for $n = 100$, 200, 500 and 1000 respectively, while the computing times for the MAP estimators for scenario (a) are 0.0093, 0.0086, 0.0085 and 0.0082 minutes for $n = 100$, 200, 500 and 1000 respectively. Even with a moderate problem ($r = 70$ and $u = 10$) in Scenario (d), the median computing time is still well under a minute. Of course, the objective of MAP estimation is vastly different from full MCMC sampling. The latter aids approximation of the entire posterior distribution and hence permits a full Bayesian inference. The former on the other hand produces just a single point estimate that by itself does not aid any quantification of modeling and estimation uncertainty and therefore does not permit inference.

C.12 Fast Approximate Bayesian Inference using Blackbox Variational Methods

Recall the decomposition (S14) for the joint posterior density of the response envelope model as the product of the marginal posterior density of $\boldsymbol{A}$ *times* the conditional posterior density of the remaining parameters given $\boldsymbol{A}$ as leveraged in Section C.11 for deriving MAP estimates. This decomposition can also be used for fast approximate Bayesian inference via some blackbox variational method, such as mean-field Automatic Differential Variational Inference (ADVI; Kucukelbir et al. 2017). To this end, first note that the conditional posterior distribution of $\boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0$ given $\boldsymbol{A}$ can be efficiently sampled by leveraging the decomposition

$$\pi(\widetilde{\boldsymbol{\mu}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) = \pi(\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y})\ \pi(\boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y})\ \pi(\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y})\ \pi(\boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y}).$$

Each density on the right-hand side is standard and can be efficiently simulated given (approximate) posterior draws for $\boldsymbol{A}$. Specifically, we have

$$\begin{aligned}
\boldsymbol{\Omega} \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y} &\sim \mathrm{IW}_u(\Psi + \boldsymbol{\Gamma}^T(\boldsymbol{A})\widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}(\boldsymbol{A}),\ n + p + \nu),\\
\boldsymbol{\Omega}_0 \mid \boldsymbol{A}, \mathbb{X}, \mathbb{Y} &\sim \mathrm{IW}_{r-u}(\Psi_0 + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{Y}_{\boldsymbol{\mu_Y}}^T\mathbb{Y}_{\boldsymbol{\mu_Y}}\boldsymbol{\Gamma}(\boldsymbol{A}),\ n + p + \nu),\\
\boldsymbol{\eta} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y} &\sim \mathrm{MN}_{u,p}\left(\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}, \boldsymbol{\Omega}, (\mathbb{X}_{\boldsymbol{\mu_X}}^T\mathbb{X}_{\boldsymbol{\mu_X}} + \boldsymbol{M})^{-1}\right),\ \text{and}\\
\widetilde{\boldsymbol{\mu}} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y} &\sim \mathrm{N}_r\left(\overline{Y}, \frac{1}{n}\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)\right).
\end{aligned}$$

Hence one approximate sampling only from the marginal posterior distribution of $\boldsymbol{A}$ with unnormalized log density $l(\boldsymbol{A})$ as given in (S15) in the context of MAP estimation of $\boldsymbol{A}$. Since the elements of $\boldsymbol{A}$ are unconstrained, a blackbox variational Bayes algorithm, e.g., mean-field ADVI as implemented in probabilistic programming language stan (Gelman et al., 2015; Carpenter et al., 2017), requiring only (log) posterior density (up to a constant multiple) for approximate sampling, can be readily employed.

However, we make a cautionary note here that while a black-box variational approach can indeed produce an immediate fast approximation of the target posterior for the proposed envelope model, care must be taken while using the estimates for anything other than quick initial experimentation with the model. Variational approaches, specifically the commonly used mean-field variational approaches (as considered here) are known to underestimate the spread of the target posterior distribution. This leads to an inadequate quantification of estimation uncertainty, thereby potentially producing highly misleading inferences; see the simulation results and the associated discussion in Section C.14. Moreover, the simulation results presented in Sections C.13 and C.14 show notably poorer performances of the variational Bayes estimators even from a point estimation (via posterior mean) perspective, when compared with the proposed MCMC-based estimators and several other (frequentist or Bayesian) competing methods/estimators.

C.13 Additional Simulation Results I: Comparing estimation accuracy, prediction performance and computational cost of the proposed Bayesian response envelope model with a number of competing approaches

We considered numerical experiments with response dimensions $r = 20$. The envelope dimension was fixed at $u = 2$, and the number of predictors was fixed at $p = 7$. The model parameters were generated following the same procedure in Section 6.1 of the main text. Then 100 independent datasets were generated for each sample size $n = 50, 100, 200, 500$, and 1000.

In each dataset, we considered 11 separate (point) estimators of the regression coefficient $\boldsymbol{\beta}$ and corresponding predictions (on a newly generated dataset from the true model):

(i) "Bayes u_{true}": MCMC posterior mean from the proposed Bayesian envelope model with true u.

(ii) "Bayes $\hat{u}$": MCMC posterior mean from the proposed Bayesian envelope model with $u = \hat{u} =$ BIC-based approximate posterior mode for u.

(iii) "Bayes BMA": MCMC posterior mean from the proposed approximate envelope BMA "u-integrated" model.

(iv) "Bayes manifold u_{true}": MCMC posterior mean from the manifold Bayesian envelope model (Khare et al., 2017) with true u.

(v) "Bayes VB u_{true}": Variational Bayes (ADVI; Kucukelbir et al. 2017) posterior mean for the proposed Bayesian envelope model with true u.

(vi) "Bayes std.": MCMC posterior mean for the standard (full; $u = r$) Bayesian linear regression model.

(vii) "Frequentist u_{true}": Frequentist envelope estimator with true u, through its implementation in R package `Renvlp` (Lee and Su, 2020). Performed with residual bootstrap to facilitate inference.

(viii) "Frequentist wtd": Frequentist weighted envelope estimator of Eck and Cook 2017. Performed with residual bootstrap to facilitate inference.

(ix) "Frequentist RRR": Frequentist reduced rank regression with optimally chosen rank using R package rrr and Renvlp. Only point estimation is performed.

(x) "remMap": remMap model (Peng et al. 2010) with cross-validation selected dimension as implemented in R package remMap. Only point estimation is performed.

(xi) "Bayes RRR u_{true}". Bayesian reduced rank regression with rank = u_{true} based on the stan program of Files et al. (2019). Due to the extremely high computational burden (see Section C.13.3, an optimal rank selection was not performed, and the true envelope dimension u_{true} was used as a proxy for the optimal rank.

Prior to making any comparisons, we make a comparative note between the "exactness" of the frequentist vs. Bayesian inferences. For the frequentist approaches we consider bootstrap-based approximations of the corresponding sampling distributions which usually produce more accurate inferences than asymptotic-normal-based approximations. It is important to note, however, that bootstrap approximations are still theoretically justified only when n is large enough; consider, e.g., the asymptotic error rates of various bootstrap statistics provided in Hall (2013, Chapter 3). By contrast, MCMC-based Bayesian inferences are exact for any n (no approximation is made on the posterior distribution based on the sample size), with the precision of the resulting estimates/inferences depending entirely on the computing power. This is particularly important in situations with small or moderate sample sizes where the bootstrap approximation can be less reliable.

For all MCMC-based implementations except Bayes RRR, we ran the respective samplers for 10,000 iterations after discarding the initial 5,000 iterations as burn-in. We then computed the MCMC posterior means for $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ on the retained MCMC iterations while also noting the median effective sample sizes for all components of $\boldsymbol{\beta}$. For frequentist envelope approaches, viz., "Frequentist u_{true}" and "Frequentist wtd." we considered $B = 10{,}000$ residual bootstrap resamples to approximate the corresponding sampling distributions. For the ADVI-based "Bayes VB u_{true}" we obtained 10,000 random draws from the optimized variational distribution. For the stan-based "Bayes RRR" implementation, 1000 MCMC samples were retained after an initial warmup of 1000 iterations. For the final three non-envelope-based approaches, viz., "Frequentist RRR", "remMap" and "Bayes RRR" we centered the training data (both the predictors and the responses) at their respective sample means prior to

model fit; the same (training data) sample means were used to subsequently center the test data prior to predictions from the fitted methods. Below we compare and discuss the (point) estimation accuracy, prediction accuracy, and computational costs for these approaches as evaluated from the replicated experiments. The computational costs were first measured through CPU running times for each method. Subsequently, the raw running times were rescaled to aid approximate comparability with a view to the (median) effective ($\boldsymbol{\beta}$-) sample sizes from the "Bayes u_{true}" MCMC sampler; see section C.13.3 for more details.

C.13.1 Comparing (Point) Estimation Accuracy

To quantify the (point) estimation accuracy of each model/estimator we consider the component-wise mean squared error (MSE) obtained from replicated values of each point estimator of $\boldsymbol{\beta}$. For the (i, j)th element in $\boldsymbol{\beta}^{\text{true}}$, suppose $\hat{\beta}^k_{i,j}$ denotes the computed value of the estimator in the kth replication. Then the MSE for the (i, j) component is computed as $\sum_{k=1}^{100}(\beta^{\text{true}}_{i,j} - \hat{\beta}^k_{i,j})^2/100$. For exposition, we focus on the averages (across all (i, j)'s) of these component-specific MSEs for different methods, r, and n, which are displayed in Table S3. Some notable observations from the table are discussed below.

First, we observe that the standard Bayesian estimator ($u = r$ model) incurs much larger average MSEs than the proposed Bayes envelope estimators (all three of Bayes u_{true}, Bayes $\hat{u}$, and Bayes BMA) for all n and r. This concords with the findings discussed in Section 7.1 of the main text. Second, the frequentist envelope estimators (Frequentist u_{true} and Frequentist wtd.) appear to produce somewhat larger average MSEs than the corresponding Bayes MCMC-based envelope estimators (Bayes u_{true} and Bayes BMA) across the board (all n and r). However, the differences are modest – less than 10% in most cases. Third, Bayes u_{true} and Bayes manifold u_{true} have comparable average MSEs particularly when the sample size is large. This is unsurprising as both approaches furnish envelope models with vague proper priors for the model parameters and use MCMC for posterior approximation.

Fourth, for small n's, Bayes u_{true} enjoys smaller average MSEs than Bayes $\hat{u}$ and Bayes BMA; with Bayes BMA and Bayes $\hat{u}$ having largely comparable performances. However, the differences diminish as the sample size increases. This is unsurprising with a view to the u-selection table in the main text: for small sample sizes, there is much variability incurred in estimating u by the posterior distribution of u which gets translated into the variability of the final $\boldsymbol{\beta}$ estimator. However, as the sample size grows the posterior distribution of u gets more and more concentrated around u_{true}, and hence the estimators from Bayes $\hat{u}$ and Bayes BMA effectively reduces to the estimator from Bayes u_{true}. Fifth, the variational Bayes u_{true} posterior mean produces a considerably poorer point estimator of $\boldsymbol{\beta}$ compared to the proposed MCMC-based Bayesian envelope point estimator from Bayes u_{true}, with the former having an average MSE that is 2 to 61 times as big as the average MSEs of the latter. In fact, for the smallest sample size $n = 50$, the VB approach produces the *poorest* estimator with the largest MSEs among all the methods/estimators considered. However, the performance of Bayes VB u_{true} gets better with larger sample sizes; however, it still produces substantially noisier estimates than the other Bayesian envelope approaches considered.

Finally, the final three non-envelope-based approaches, viz. Frequentist RRR, remMap, and Bayes RRR u_{true}, incur substantially bigger (about twice as big or even bigger) average MSEs compared to the proposed MCMC-based Bayesian envelope estimators. Of these three, Frequentist RRR consistently produces the smallest average MSEs and the Bayes RRR u_{true} produces the largest. For smaller and moderate n (≤ 200), Frequentist RRR estimators perform as good or better than Bayes VB u_{true}; however, as the sample size grows, the VB approach produces more accurate estimators. The Bayes RRR u_{true} appears to produce poor estimates even with large n's. We note that this may be attributable to the disparity between the optimal rank of a Bayesian reduced rank regression model, and the corresponding envelope dimension u_{true} of an envelope estimator. Due to extremely high computational cost, an optimal rank selection was not performed for Bayes RRR and the true

envelope dimension u_{true} was simply used as a proxy for the optimal rank. However, when performed the consequent *optimal* model may incur less variability. However, with a view to the performance of the frequentist reduced rank regression estimator (for which we did consider an optimal rank selection), it seems unlikely that an optimally tuned Bayes RRR estimator would incur smaller errors than the proposed Bayesian envelope estimators.

Method	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Bayes u_{true}	3.389	1.458	0.737	0.267	0.144
Bayes $\hat{u}$	5.259	1.717	0.758	0.271	0.139
Bayes BMA	5.103	1.681	0.751	0.270	0.139
Bayes manifold u_{true}	3.249	1.402	0.728	0.266	0.145
Bayes VB u_{true}	190.480	4.465	1.433	0.882	0.813
Bayes std.	25.025	11.095	5.284	2.066	1.043
Frequentist u_{true}	3.631	1.499	0.782	0.288	0.161
Frequentist wtd.	7.621	2.128	0.843	0.299	0.161
Frequentist RRR	27.980	3.364	1.515	0.576	0.308
remMap	30.194	13.265	5.851	2.189	1.082
Bayes RRR u_{true}	24.472	20.781	19.553	18.516	18.228

Table S3: Comparison of replication MSEs (total across coordinates of $\boldsymbol{\beta}$) among various methods/ point estimators of $\boldsymbol{\beta}$ for different sample sizes (n).

C.13.2 Comparing (Point) Prediction Accuracy

Method	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Bayes u_{true}	7.55	7.42	7.35	7.30	7.30
Bayes $\hat{u}$	7.65	7.43	7.35	7.30	7.30
Bayes BMA	7.65	7.42	7.35	7.30	7.30
Bayes manifold u_{true}	7.52	7.40	7.37	7.30	7.30
Bayes VB u_{true}	17.12	7.57	7.40	7.33	7.33
Bayes std.	8.67	7.91	7.59	7.39	7.35
Frequentist u_{true}	7.55	7.41	7.37	7.30	7.30
Frequentist wtd.	7.76	7.44	7.37	7.30	7.30
Frequentist RRR	8.85	7.51	7.40	7.31	7.31
remMap	8.98	8.00	7.62	7.39	7.35
Bayes RRR	8.61	8.37	8.30	8.22	8.20

Table S4: Comparison of unit average mean squared prediction errors from several competing methods considered in the simulation. The right five columns display the computed average mean squared prediction error.

Next, we focused on the prediction accuracy for each model/estimating method under consideration. For this, after training the various models/estimators on each replicated training dataset, we generated a new test dataset $\{(\boldsymbol{X}_i^{\text{new}}, \boldsymbol{Y}_i^{\text{new}}) : i = 1, \ldots, n\}$ (the same n as in the training data), and computed the unit average mean squared prediction error defined as $(rn)^{-1} \sum_{i=1}^{n} \|\boldsymbol{Y}_i^{\text{pred}} - \boldsymbol{Y}_i^{\text{new}}\|^2$,

where $\boldsymbol{Y}_i^{\text{pred}} = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\beta}}\boldsymbol{X}_i^{\text{new}}$ and $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ are the estimates from each model/method. For the non-envelope-based approaches, viz., Frequentist RRR, remMap, and Bayes RRR u_{true}, both training and test data were centered (separately for the predictors and the responses) by the respective training sample means prior to model fit and prediction, and the corresponding $\hat{\boldsymbol{\mu}}$'s were set to zero. Note at the outset that with growing sample sizes these unit average prediction errors are expected to stabilize for each method; however, unlike estimation error, prediction errors will not converge to zero with $n \to \infty$ due to the inherent variabilities in $\boldsymbol{Y}$'s (measurement noise). The results are summarized in Table S4. From the table, it follows that the overall prediction performances of the proposed Bayesian envelope approaches (Bayes u_{true}, Bayes $\hat{u}$, and Bayes BMA) are consistently better or as good as the other approaches. The overall prediction performances for the different methods generally resemble the estimation performance patterns noted in Section C.13.1, although there are some differences. The prediction performances for the different methods are discussed in detail below.

First, we note that the standard Bayesian estimator ($u = r$ model) has a uniformly poorer prediction performance compared to the proposed Bayes envelope estimators (all three of Bayes u_{true}, Bayes $\hat{u}$, and Bayes BMA) for all n and r. Second, the frequentist envelope estimators (Frequentist u_{true} and Frequentist wtd.) have comparable prediction errors as their corresponding Bayes MCMC-based envelope estimators (Bayes u_{true} and Bayes BMA) across the board (all n and r). Third, Bayes u_{true} and Bayes manifold u_{true} also have virtually identical unit average prediction errors. Fourth, for small n's, Bayes u_{true} enjoys smaller prediction errors than Bayes $\hat{u}$ and Bayes BMA; with Bayes BMA and Bayes $\hat{u}$ having largely comparable performances, with the differences diminishing as the sample size increases. However, unlike estimation performance, the differences between the prediction performances of these three estimators are not as prominent even for small sample sizes. Fifth, for the small sample size of $n = 50$, Bayes VB u_{true} produces extremely poor prediction as noted through the considerably large prediction errors which are much bigger than their counterparts from the other models/estimators. However, as the sample size increases the average unit prediction errors become less severe and comparable to the other approaches/estimators.

Finally, the final three non-envelope-based approaches, viz. Frequentist RRR, remMap, and Bayes RRR u_{true}, produce substantially bigger average unit prediction errors compared to the proposed MCMC-based Bayesian envelope estimators for the small sample size of $n = 50$. As the sample size grows, the average prediction errors get noticeably smaller for remMap and Frequentist RRR, virtually becoming identical to the envelope-based approaches, but not for Bayes RRR. Again this could be attributable to the disparity between the optimal rank of a Bayesian reduced rank regression model, and the corresponding envelope dimension u_{true} of an envelope estimator, as noted in the discussion of estimation performances, and for large n an optimally tuned Bayes RRR implementation may incur comparable prediction error as the proposed method.

C.13.3 Comparing Computing Costs

Finally, we focus on the computational costs for the various methods/estimators in terms of CPU running times and effective sample sizes (for the proposed MCMC-based Bayesian approach). All computations were done on SLURM HPC parallel computing clusters at the Center for Computational Research at the State University of New York at Buffalo, with computing nodes having 16 – 20 GB of allocated memory each, and clock speeds ranging between 2.10 – 2.40 GHz.

To permit a fair comparison of computational costs for the proposed Bayesian envelope model MCMC sampler relative to the Khare et al. (2017) MCMC sampler, we first compare their respective effective sample sizes (ESS). The ESS is a simple univariate measure of the effectiveness of MCMC draws, estimating the number of independent and identically distributed random samples required to achieve the same level of precision as achieved by MCMC samples. The average ESS ratio across elements in $\boldsymbol{\beta}$ and replications for the proposed method to the manifold method for $u = u_{\text{true}}$ is

summarized in Table S5. The values suggest that the ESS of the two methods are very comparable, with the proposed method having a slightly lower ESS on average than the manifold method. This slight reduction in ESS is more than compensated by the substantial reduction in computing times which we discuss next.

Next, we focus on the running times of the proposed method and compare them with those of the various competing methods. Prior to any comparisons, we make two notes. First, for the "ensemble" models (both Bayes BMA and Frequentist wtd.) the total CPU running times under sequential model fits are recorded. In practice, however, one can of course fit models with different u's in parallel which can substantially reduce the total observed user running time. Second, the cost of running Bayes $\hat{u}$ and Bayes BMA are virtually the same – both require running all individuals u-sepcific models together with (approximate) u-specific probabilities. The two approaches only differ in how the subsequent point estimate is computed – Bayes $\hat{u}$ obtains a single point estimate based on the posterior u probabilities while Bayes BMA uses the posterior u probabilities to obtain a model-averaged point estimate; given MCMC draws from all models, the difference between the running times for these computations (selecting a model vs. averaging models) are insignificant. Hence, we only report the running times of Bayes BMA below.

n	ESS Ratio
50	0.936
100	0.892
200	0.900
500	0.908
1000	0.901

Table S5: Average ratio of ESS of the proposed method to the manifold method.

Since the proposed approach permits MCMC-based full Bayesian inference, to ensure a fair cost-comparison with a common denominator, we rescaled the individual observed running times of Frequentist u_{true}, Frequentist wtd., Frequentist RRR, remMap, and Bayes RRR in each replicate to get the corresponding "expected" running times necessary to produce approximately a median ESS number of posterior samples/bootstrap resamples as obtained for the proposed method with $u = u_{\text{true}}$. This entailed multiplying the observed running times of Frequentist u_{true} and Frequentist wtd. by a factor of $\frac{\text{median ESS}}{B=10000}$, those of Frequentist RRR and remMap by a factor of median ESS (since only a single point estimate was obtained from these models and not a full bootstrap distribution), and that of Bayes RRR by a factor of $\frac{\text{median ESS}}{1000}$ (since Bayes RRR was run for 1000 final iterations). The running times of the mean-field variational Bayes implementation were left unchanged as the bulk of the computation time in VB implementations are spent on optimizing the approximating variational family which does not depend on the size of the subsequent Monte Carlo draws from the optimized approximating family; sampling from the optimized variational family is usually very fast.

The obtained running times are summarized in two tables – Table S6 showing the median (across replicates) *absolute* running times in minutes for the proposed method with $u = u_{\text{true}}$ and the two choices of r and Table S7 displaying the median (across replicates) *relative/proportional* running times of all 11 methods/estimators relative to the running time of the proposed method, with the run-time ratios being computed separately in each replication first and then summarized through medians. We make the following observations from these tables.

From Table S6 it can be seen that the median running times for the different n's are all within 5 minutes. The variability observed in the running times is attributable to the differences in n's (which primarily affect the likelihood computation at each iteration), in r's and also in the computing

Method	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Bayes u_{true}	3.90	2.31	3.42	3.49	3.79

Table S6: Median absolute CPU running times (in minutes) for the proposed Bayesian envelope method with $u = u_{\text{true}}$.

cores/processors present in the computing cluster. The relative running times reported in Table S7 avoid the variability in the computing architecture thus permitting a more principled cost comparison among the different methods as discussed below.

method	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Bayes u_{true}	1.00	1.00	1.00	1.00	1.00
Bayes $\hat{u}$	1.34	1.00	1.00	1.00	1.00
Bayes BMA	36.62	36.87	36.56	36.01	38.59
Bayes manifold u_{true}	6.53	7.09	7.73	9.82	12.04
Bayes VB u_{true}	0.09	0.11	0.10	0.11	0.12
Bayes std.	0.04	0.04	0.04	0.05	0.09
Frequentist u_{true} $\times$ med ESS / 10000	3.54	2.13	1.76	1.45	1.31
Frequentist wtd. $\times$ med ESS / 10000	36.12	35.18	31.72	30.44	27.83
Frequentist RRR $\times$ med ESS	16.94	21.17	22.55	22.36	21.37
remMap $\times$ med ESS	37.55	53.32	71.76	122.75	184.10
Bayes RRR u_{true} $\times$ med ESS / 1000	0.60	2.25	6.51	21.90	60.75

Table S7: Median of CPU running times (ratios) for various competing methods/estimators relative to the running times of the proposed Bayesian envelope method with $u = u_{\text{true}}$.

First, Table S7 shows that the implementation of the standard Bayesian model is extremely fast costing on average only 5-10% of the corresponding running times of the proposed method for all n and r. This is unsurprising as the standard Bayesian model does not require generating draws for the envelope parameters, a step that takes up the bulk of the computation times for the envelope Based approaches. However, this perceived gain in speed comes at the cost of a substantially increased noise incurred in the estimation and prediction from the model (Sections C.13.1 and C.13.2), which is, of course, the main motivation for developing a Bayesian envelope model. Second, we see that the (scaled) running times for Frequentist u_{true} is larger than its Bayesian counterpart, while those for Frequentist wtd. is comparable or somewhat smaller than Bayes BMA. This is because the computational cost for the proposed method grows with increasing $u < r$, and the "ensemble" approaches (Frequentist wtd. and Bayes BMA) require computation for all u, including large $u < r$. In applications, the smaller u's are precisely the situations that permit the biggest gain in estimation and prediction efficiency (Cook et al., 2010, 2013), and the proposed approach is seemingly more computationally efficient than the frequentist approach in such situations. However, unlike MCMC sampling bootstrapping is (embarrassingly) parallelizable which means that the computation times for the frequentist implementations can be reduced in a parallel computing environment. Second, the proposed approach enjoys a significantly reduced computing time compared to the manifold based approach of Khare and Hobert (2012), with the proposed method on average being $650 - 1900\%$ times faster than the manifold method. This is a direct consequence of the Euclidean space-based formulation of the underlying envelope parameter which makes the proposed approach substantially less computationally demanding than the Steifel manifold-based sampler of Khare et al. (2017). Third, Bayes VB u_{true}

appears to enjoy a substantially reduced computational cost compared to the proposed MCMC-based Bayes u_{true}. However, the poorer estimation and prediction performance of the latter particularly for small sample sizes (Sections C.13.1 and C.13.2) curtail the practical usability of the method. Third, the stan-based implementation of Bayes RRR has a somewhat comparable (rescaled) computation cost when the sample size is small, but a substantially bigger cost when the sample size is large. This is a consequence of the Hamiltonian Monte Carlo-based implementation used in stan (Files et al., 2019) which requires multiple evaluations of the likelihood and the gradient in each iteration, which grows substantially with increasing n. Fourth, the expected cost of cross-validation-tuned remMap under a bootstrap-based inferential computation appears to be substantially bigger than that of Bayes BMA particularly when the sample size is large, although the runtime for the former can be reduced in a parallel computing environment leveraging parallelizability of bootstrap resampling. We note however that the appropriateness of bootstrap resampling for inference in remMap has not been discussed to the best of our knowledge. Finally, Frequentist RRR appears to enjoy a substantially smaller computing time, even after accounting for rank selection. This is due to the fact that analytical closed forms exists for the Frequentist RRR estimators, which substantially reduces its overall computing time.

C.14 Additional Simulation Results II: Running the Proposed MCMC sampler under larger dimensions and sample sizes

To note how the proposed MCMC algorithm fare with larger dimensions and sample sizes, we performed analgous, but more sizeable simulation experiments. More specifically, we adopted the same setting as the one described in Section 6.1 of the main text, and looked into three cases: $r = 100, p = 20$, and $r = 200, p = 20$, and $r = 200, p = 100$, which correspond to $2,000$, $4,000$, and $20,000$ regression coefficients (number of elements in $\boldsymbol{\beta}$). For each case we considered two u's, namely $u = 1$ and $u = 6$. We varied the sample sizes from 500 to 5000, and 100 replications were generated for each sample size. The Bayesian envelope model was fit by running the proposed MCMC sampler for 10,000 iterations (after discarding an initial burn-in of 5,000 iterations). To aid comparison, an ADVI-based (Kucukelbir et al., 2017) variational implementation was also considered with 1000 Monte carlo draws generated from the variational family after optimizing. All computations were done on SLURM HPC parallel computing clusters at the Center for Computational Research at the University at Buffalo, with computing nodes having allocated memory ranging between 24 to 64 GB, and clock speeds ranging between 2.10-2.40 GHz.

For the MCMC-based implementation, in each iteration, we not only generated the model parameters but also evaluated the likelihood function useful for potential statistical inference and model selection. Thus two computation times were obtained for the MCMC-based implementation, one for generating the posterior draws together with iteration-wise log-likelihood computation, and one for only generating the posterior draws without log-likelihood computation. In Table S8, we report the median computation times in minutes for these two MCMC implementations along with variational implementations along the columns "MCMC logLik" (MCMC with log-likelihood evaluations), "MCMC" (MCMC without log-likelihood evaluations) and "VB". The average ratio of effective sample sizes to MCMC iteration size across all elements in $\boldsymbol{\beta}$ is included in the column "ESS Ratio" which provides a simple univariate summary of the efficiency of MCMC draws. We make the following observations from the table.

First, In all cases, we notice that the proposed MCMC algorithm provides good effective sample sizes, particularly when the envelope dimension u is small, which is precisely the situation where envelope models achieve the greatest gains in estimation and prediction efficiency. When u is increased, the model becomes less parsimonious and requires a bigger number of effective parameters to sample from, particularly for the envelope parameter $\boldsymbol{A}$. This also explains why a larger u also effectuates a greater running time of the algorithm. Second, we note that the computing time for the MCMC

r	p	u	n	MCMC logLik Time	MCMC Time	%ESS	VB Time
100	20	1	500	16.95	16.85	86.55	6.99
			1000	14.07	13.66	92.55	5.71
			2000	14.78	13.93	95.95	6.47
			5000	29.82	24.38	98.53	15.78
		6	500	75.34	79.74	44.44	5.10
			1000	82.60	83.82	44.19	6.42
			2000	84.59	86.47	44.83	7.81
			5000	81.98	82.19	44.50	10.66
200		1	500	104.41	100.88	86.93	24.42
			1000	107.20	101.40	95.41	23.21
			2000	113.03	101.33	98.55	26.34
			5000	129.54	100.84	99.75	28.41
		6	500	267.67	289.39	40.64	22.77
			1000	241.23	244.98	41.42	21.14
			2000	279.82	281.40	40.87	26.44
			5000	517.20	491.29	38.44	36.12
	100	1	500	51.19	50.34	92.39	18.18
			1000	99.90	94.91	98.10	25.99
			2000	83.54	75.39	99.40	24.36
			5000	94.89	75.37	100.05	23.97
		6	500	236.19	232.98	53.00	23.23
			1000	236.08	233.33	39.68	24.20
			2000	236.41	232.30	30.77	24.83
			5000	286.57	310.03	26.21	28.27

Table S8: Comparison of median (across replications) running times (in minutes) of the MCMC and variational implementations of the proposed Bayesian envelope model for larger n, p, and r, and different u's.

r	p	u	n	MSE MCMC	postvar MCMC	MSE VB	postvar VB
100	20	1	500	0.591	0.620	14.350	1.406
			1000	0.263	0.274	10.036	0.576
			2000	0.122	0.128	3.498	0.027
			5000	0.048	0.049	1.581	0.007
		6	500	5.250	5.339	132.260	13.638
			1000	2.525	2.531	109.002	2.319
			2000	1.224	1.235	57.645	0.206
			5000	0.485	0.486	20.021	0.087
200		1	500	3.217	3.867	222.338	6.567
			1000	1.458	1.629	64.872	3.005
			2000	0.716	0.759	61.439	0.454
			5000	0.286	0.289	20.944	0.088
		6	500	10.163	10.597	2876.678	23.823
			1000	4.564	4.681	2907.111	5.104
			2000	2.226	2.239	2950.035	0.302
			5000	0.865	0.870	2979.612	0.121
	100	1	500	3.296	4.542	537.011	2.895
			1000	1.547	1.801	303.465	0.120
			2000	0.763	0.822	512.522	2.165
			5000	0.305	0.309	121.425	0.185
		6	500	19.845	23.221	13631.671	37.953
			1000	8.699	9.582	12781.418	3.021
			2000	4.240	4.434	12864.314	1.530
			5000	1.664	1.697	12768.959	0.604

Table S9: The total (across $\boldsymbol{\beta}$ components) replication MSEs of posterior means and the average (across replications) total (across $\boldsymbol{\beta}$ components) posterior variances obtained from the MCMC and variational implementations for larger n, p, and r, and different u's.

sampler also steadily increases with the number of responses r; however, it does not vary significantly with the number of predictors p. We note in this context that the median computation times for the $r = 200, p = 20$ configuration are somewhat larger than the median times for the $r = 200, p = 100$ configuration. This is due to the fact that these computations were done on a computing cluster consisting of CPUs of varying hardware specifications. Larger memories were allocated to run the jobs corresponding to the $r = 200, p = 100$ configuration compared to those associated with the $r = 200, p = 20$ configuration; these resulted in the former configuration being run on more powerful (in terms of clock speed) computers. Third, the computation time for the MCMC sampler also increases with n, but this is mainly due to the evaluation of the likelihood; the cost for MCMC draws of model parameters do not increase substantially with the sample size. Finally, the average computational cost for the variational implementation is substantially smaller than the corresponding MCMC implementation with the latter being 2-15 times as expensive as the former on average. However, the variational implementation has a substantially poorer estimation accuracy and estimation uncertainty quantification which collectively curtails its practical utility as we discuss next.

To assess the accuracy of the MCMC and variational implementations we focused on two metrics – (a) the average MSE of the corresponding posterior mean and (b) the average posterior variance, where the average is computed across all replicates and all components of $\boldsymbol{\beta}$. Here (a) produces a frequentist measure of how good the posterior mean is as a point estimator relative to the "truth", while (b) leverages a quantifiable summary of estimation uncertainty produced by each method, which in turn can be used to assess honesty/reliability of the uncertainty quantification (with a view to the computed frequentist MSE). These results are displayed in the columns of Table S9.

The table shows a reasonable accuracy of the MCMC-based point estimator as quantified through average MSEs of the posterior mean estimator. The average MSE for the MCMC posterior mean estimator drops with larger sample sizes for each specific r, p, and u, which illustrates a frequentist consistency (in the L^2 sense) of the estimator. In addition, the MCMC-based (averaged) posterior variances produce honest and reliable quantification of the estimation uncertainty, as seen through their closeness with the average MSEs. By contrast, the variational implementation-based posterior mean point estimator has a much poorer accuracy, with the average MSEs two to four orders of magnitude bigger than the corresponding MCMC-based point estimators. More importantly, the corresponding posterior variances do not appear to provide honest quantification of estimation uncertainty: there are large differences (sometimes between two to four orders of magnitude) between the average MSEs and the average posterior variances. Indeed, the average posterior variance computed from the variational implementation is sometimes much smaller than its MCMC counterpart, thus potentially leading to extremely unreliable statistical inference.

With a view to these simulation results we therefore recommend against using the ADVI variational implementation for the proposed Bayesian envelope model for principled statistical inference and suggest using the MCMC implementation instead, despite the latter demanding substantially heavier computations.

C.15 Additional Simulation Results III: Comparing the estimation performance of the proposed model with competing approaches models under varying levels of signal-to-noise-ratio

This section compares under replication the performances of the proposed Bayesian envelope estimators with several other competing approaches under various signal-to-noise ratios (SNR). To aid precise control over the SNR in the simulated data, we fixed the sample size at $n = 500$, and let $r = 20$, $p = 10$, and $u_{\text{true}} = 2$. Next, we set $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\mu}_{\boldsymbol{X}} = \mathbf{0}$, and filled the elements of $\boldsymbol{A}$ and $\boldsymbol{\eta}$ with Uniform$(0, 1)$ random variates, and subsequently normalized $\boldsymbol{\eta}$ so that $\|\boldsymbol{\eta}\|_F = 1$ effectuating $\|\boldsymbol{\beta}\|_F = 1$. Here $\|\cdot\|_F$ denotes the Frobenius norm. Following Cook et al. (2015) the covariance matrices were created as

$[\boldsymbol{\Omega}]_{ij} = \sigma^2(-0.9)^{|i-j|}$ and $[\boldsymbol{\Omega}_0]_{ij} = \sigma^2(-0.9)^{|i-j|}$ with varying constants $\sigma^2 > 0$. Since $\|\boldsymbol{\beta}\|_F = 1$, the SNR is simply $1/\sigma^2$ which we varied from 0.10 to 1. Our setup is similar to the *reduced-rank envelope* model setup of Cook et al. (2015) with u_{true} = reduced rank dimension (d_{true}) = 2, and the "$\boldsymbol{B}$"-matrix used to define the reduced-rank envelope regression coefficient $\boldsymbol{\beta}^* = \boldsymbol{\Gamma}\boldsymbol{\eta}\boldsymbol{B}$ being set to the (u_{true} = 2-dimensional) identity matrix $\boldsymbol{B} = \boldsymbol{I}_2$. From each "true" model corresponding to one specific SNR level, we simulated $r = 100$ replicated random datasets.

Subsequently, in each simulated dataset, we fitted the proposed Bayesian envelope model, the frequentist envelope model, the remMap model, and the frequentist RRR model and obtained various point estimators of $\boldsymbol{\beta}$ obtained from these models. More specifically, we focused on the MCMC posterior mean associated with "Bayes u_{true}" and "Bayes BMA" (our approaches), and Manifold MCMC Bayes u_{true} (Khare et al., 2017) and the frequentist point estimates associated with "Frequentist u_{true}", "Frequentist wtd.", remMap, and "Frequentist RRR" (see Section C.13 for a description of these models/estimators). The accuracy of each estimator was defined in terms of total (across all components) replication MSEs (see section C.13.1 for a definition). These total MSEs for the different estimators under different SNRs are displayed in Table S10.

SNR	Bayes BMA	Bayes u_{true}	Man. u_{true}	Freq. u_{true}	remMap	Freq. RRR	Bayes Full
0.100	0.120	0.285	0.228	0.318	2.810	3.070	4.050
0.111	0.117	0.268	0.271	0.295	2.500	2.800	3.650
0.125	0.113	0.279	0.266	0.302	2.200	2.510	3.240
0.143	0.109	0.274	0.288	0.298	1.890	2.190	2.840
0.167	0.105	0.224	0.211	0.247	1.620	1.850	2.430
0.200	0.100	0.160	0.155	0.178	1.340	1.460	2.030
0.250	0.096	0.166	0.184	0.183	1.070	1.040	1.620
0.333	0.092	0.112	0.120	0.123	0.806	0.608	1.220
0.500	0.087	0.153	0.142	0.160	0.574	0.306	0.811
0.600	0.086	0.118	0.101	0.124	0.503	0.233	0.675
0.675	0.085	0.115	0.103	0.120	0.463	0.202	0.601
0.750	0.084	0.119	0.082	0.126	0.429	0.181	0.541
0.825	0.084	0.107	0.091	0.115	0.399	0.166	0.491
0.900	0.083	0.089	0.082	0.099	0.375	0.149	0.450
1.000	0.083	0.079	0.065	0.094	0.349	0.138	0.405

Table S10: Comparing estimation performance via total MSEs of the proposed method with various competing methods under varied levels of signal-to-noise ratios (SNRs).

We make the following observations from the table. First, we notice that the MSEs of remMap and RRR are both about 10 times as large MSE as the envelope model when the SNR is small. This is consistent with the observation in Cook et al. (2010) that the envelope model is especially effective when SNR is small, because its rationale is to use dimension reduction tools to remove the immaterial information in the data, which amounts to reduce the noise level after the dimension reduction. In contrast, RRR places rank constraints on the coefficient matrix, which is also performing dimension reduction but not necessarily lead to removal of immaterial information. Second, we observe that the proposed BMA estimator has a smaller MSE than the frequentist envelope model across all SNRs, and especially when SNR level is low, which is consistent with the observation that the Bayes estimator tends to have smaller MSE than the frequentist estimator. When SNR level grows, their difference becomes smaller. We also notice that with lower SNR, BMA estimator also has smaller MSE than the manifold implementation of envelope model in Khare et al. (2017), even though the

manifold implementation uses the true dimension. This is because it is harder to estimate the envelope subspace when the signal is weaker, and the manifold implementation may not be able to provide a stable estimator of the envelope subspace even under the true dimension. BMA is a data-driven method, which gives more weights to dimensions that fits the data better. It is possible that a slight overestimate of the dimension guarantees the inclusion of all relevant information in this case. But as SNR level increases, the manifold implementation starts to show the advantage of knowing the true dimension, and provide a slightly smaller MSE. However, in real applications, the dimension of the envelope subspace is unknown.

C.16 Additional Simulation Results IV: Estimation variance ratios associated with the estimation MSE ratios displayed in Table 2 in the main text

The table below displays the ratios of estimation variances (via posterior mean) of $\boldsymbol{\beta}$ components computed from the Bayesian standard regression model to the envelope-based models as discussed in Section 5.1 of the main text (analogous to the the estimation MSEs presented in the main text).

n	$V_{i,j,\mathcal{M}_{\text{std}}}/V_{i,j,\mathcal{M}_{\hat{u}}}$	$V_{i,j,\mathcal{M}_{\text{std}}}/M_{i,j,\mathcal{M}_{\text{BMA}}}$	$V_{i,j,\mathcal{M}_{\text{std}}}/V_{i,j,\mathcal{M}_{u_{\text{true}}}}$
50	4.91 (2.72, 8.72)	5.05 (2.74, 9.26)	7.77 (3.49, 17.23)
100	6.54 (3.67, 17.34)	6.65 (3.69, 17.76)	7.56 (3.91, 21.93)
200	7.11 (3.47, 16.95)	7.16 (3.49, 16.92)	7.52 (3.58, 18.96)
500	7.62 (3.65, 21.17)	7.67 (3.65, 21.07)	7.86 (3.65, 21.77)
1000	7.67 (3.74, 23.39)	7.67 (3.74, 23.39)	7.67 (3.74, 23.39)

Table S11: Medians (ranges) of the component-wise estimation variance ratios $V_{i,j,\mathcal{M}_{\text{std}}}/V_{i,j,\mathcal{M}_{\hat{u}}}$, $V_{i,j,\mathcal{M}_{\text{std}}}/M_{i,j,\mathcal{M}_{\text{BMA}}}$, and $V_{i,j,\mathcal{M}_{\text{std}}}/V_{i,j,\mathcal{M}_{u_{\text{true}}}}$.

C.17 Additional Results from the Real Data Analysis

In this section we compare the statistical significance results obtained from the Bayesian envelope models discussed on Section 7.2.1 of the main text with a Frequentist RRR model with an optimally chosen rank $d = 12$, and the Frequentist envelope model with the same $u = 1$ as used in the Bayesian model. At the outset, recall from Figure 2C in the main text that these two frequentist approaches have cross-validation prediction performance nearly identical to the proposed Bayesian envelope model. However, our focus here lies in inferring the underlying predictor-response relationship through a statistical significance analysis performed on the regression coefficients. To quantify "significance" for the two frequentist approaches we considered $B = 1000$ residual bootstrap resamples, and computed 95% bootstrap confidence percentile intervals for the coordinates for $\boldsymbol{\beta}$ from each model. Analogously to the Bayesian approaches, the "significance" for a coordinate of $\boldsymbol{\beta}$ in each model was then determined if the corresponding 95% confidence interval excluded zero. Figure S1 display the resulting significance results for the two frequentist approaches, alongside analogous results from the Bayesian approaches as shown in Figure 3 of the main text. We make the following observations from the figure. First, it appears that the Frequentist RRR model produces a very sporadic significance pattern, somewhat similar to the full Bayesian model. As noted before this irregularity makes a scientific understanding of the underlying predictor response relationship very challenging. By contrast, the frequentist envelope approach produces a much more systematic significance pattern, comparable to the Bayesian envelope model results. This is unsurprising since both the frequentist and the Bayesian envelope models share the common likelihood, and the Bayesian approach uses vague proper prior. Second, while the frequentist envelope model does produce a noticeably regular significance pattern, interestingly,

the total number of components detected to be significant is much bigger for the frequentist approach compared to the Bayesian approach. Furthermore, the Frequentist RRR model also appears to identify more significant coefficients than the Bayesian full model.

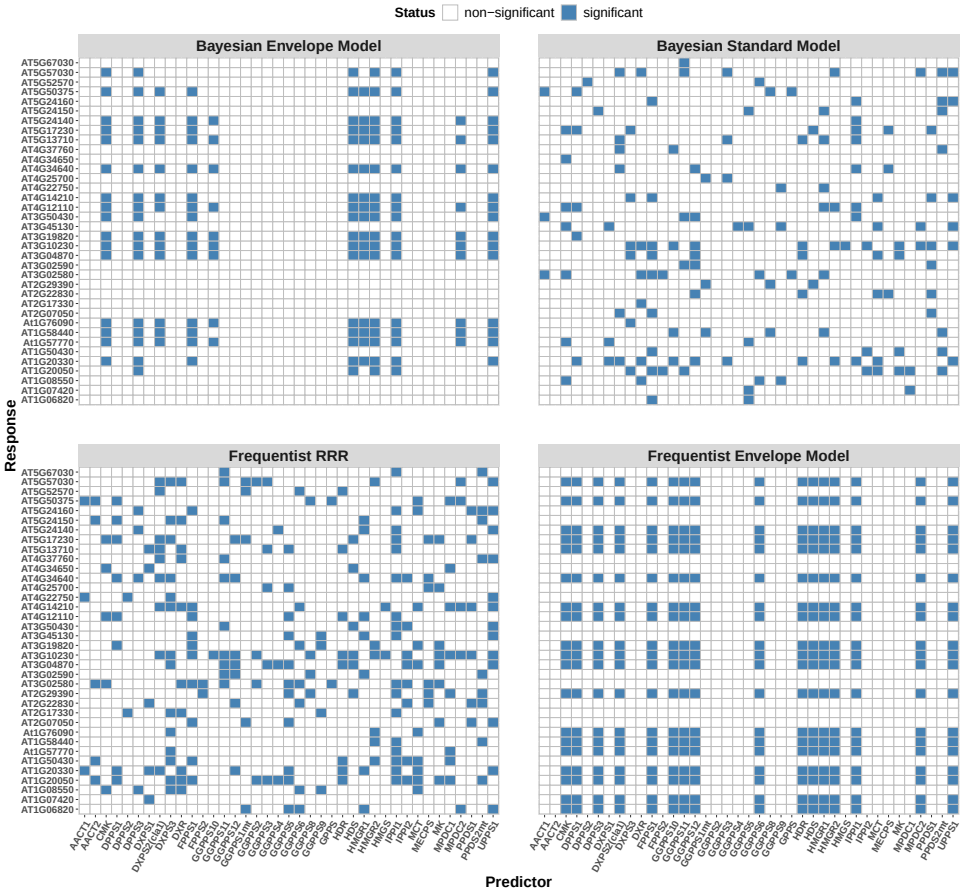


Figure S1: Significance of estimated regression coefficients under the envelope model and the standard model.

D Various Technical Details and Additional Results on the Bayesian Predictor Envelope/PLS Model and its Implementation

D.1 Log likelihood, log posterior density, and MCMC sampler for Bayesian predictor envelope model

Let $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ be n independent observations from the predictor envelope model (10), and let $\mathbb{Y} \in \mathbb{R}^{n\times r}$ and $\mathbb{X} \in \mathbb{R}^{n\times p}$ be data matrices $\mathbb{Y}^T = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ and $\mathbb{X}^T = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$. Then the log likelihood under the model (10) $L(\boldsymbol{\mu_X}, \boldsymbol{\mu_Y}, \boldsymbol{\eta}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A})$ is given as follows.

$$\log L(\boldsymbol{\mu_X}, \boldsymbol{\mu_Y}, \boldsymbol{\eta}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}) = \text{const.} - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\text{trace}\Big\{\Big(\mathbb{X} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\Big)$$
$$\Big[\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\Big]^{-1}\Big(\mathbb{X} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\Big)^T\Big\} - \frac{n}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| - \frac{1}{2}\text{trace}\Big\{\big[\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{Y}}^T$$
$$- \Big(\mathbb{X} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\Big)\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\big]\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\Big[\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{Y}}^T - \Big(\mathbb{X} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\Big)\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\Big]^T\Big\}. \quad \text{(S16)}$$

The associated unnormalized log posterior density is as follows:

$$
\begin{aligned}
&\log \pi(\boldsymbol{\mu_Y}, \boldsymbol{\mu_X}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
=& \text{const.} - \frac{n}{2} \log |\boldsymbol{\Sigma_{Y|X}}| - \frac{1}{2} \operatorname{trace}\left[\left(\mathbb{Y}_{\boldsymbol{\mu_Y}} - \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right) \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\mathbb{Y}_{\boldsymbol{\mu_Y}} - \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T \right] - \frac{n}{2} \log |\boldsymbol{\Omega}| \\
&- \frac{n}{2} \log |\boldsymbol{\Omega}_0| - \frac{1}{2} \operatorname{trace}\left[\mathbb{X}_{\boldsymbol{\mu_X}} \left(\boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A}) \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^T(\boldsymbol{A}) \right) \mathbb{X}_{\boldsymbol{\mu_X}}^T \right] - \frac{\nu_{\boldsymbol{Y}} + r + 1}{2} \log |\boldsymbol{\Sigma_{Y|X}}| \\
&- \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{Y}} \right) - \frac{m}{2} \log |\boldsymbol{\Sigma_{Y|X}}| - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{M} \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} \right) \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} \right)^T \right] \\
&- \frac{\nu_{\boldsymbol{X}} + m + 1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}_{\boldsymbol{X}} \right) - \frac{\nu_{0,\boldsymbol{X}} + (p-m) + 1}{2} \log |\boldsymbol{\Omega}_0| - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Omega}_0^{-1} \boldsymbol{\Psi}_{0,\boldsymbol{X}} \right) \\
&- \frac{1}{2} \operatorname{trace}\left[\boldsymbol{K}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T \right]. \qquad \text{(S17)}
\end{aligned}
$$

Algorithm D.1. One iteration of the Metropolis-within-Gibbs MCMC sampler for updating $\{\boldsymbol{\mu_X}, \boldsymbol{\mu_Y}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}\}$ when $1 \le m \le p-1$.

S.1 Generate

$$
\begin{pmatrix} \boldsymbol{\mu_X} \\ \boldsymbol{\mu_Y} \end{pmatrix} \text{ from } \mathrm{N}_{p+r} \left(\begin{pmatrix} \overline{\boldsymbol{X}} \\ \overline{\boldsymbol{Y}} \end{pmatrix}, \frac{1}{n} \begin{pmatrix} \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{\Omega} \boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A}) \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^T(\boldsymbol{A}) & \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{\Omega} \boldsymbol{\eta} \\ \boldsymbol{\eta}^T \boldsymbol{\Omega} \boldsymbol{\Gamma}^T(\boldsymbol{A}) & \boldsymbol{\Sigma_{Y|X}} + \boldsymbol{\eta}^T \boldsymbol{\Omega} \boldsymbol{\eta} \end{pmatrix} \right).
$$

S.2 Generate $\boldsymbol{\Sigma_{Y|X}}$ from $\mathrm{IW}_r(\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{Y}}, \widetilde{\nu}_{\boldsymbol{Y}})$, where $\widetilde{\nu}_{\boldsymbol{Y}} = \nu_{\boldsymbol{Y}} + n + m$ and $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{Y}}$ is

$$
\boldsymbol{\Psi_Y} + \left(\mathbb{Y}_{\boldsymbol{\mu_Y}} - \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T \left(\mathbb{Y}_{\boldsymbol{\mu_Y}} - \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right) + \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} \right)^T \boldsymbol{M} \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} \right),
$$

where $\mathbb{Y}_{\boldsymbol{\mu_Y}} = \mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T$ and $\mathbb{X}_{\boldsymbol{\mu_X}} = \mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T$.

S.3 Generate $\boldsymbol{\eta}$ from $\mathrm{MN}_{m,r} \left(\widetilde{\boldsymbol{M}}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \widetilde{\boldsymbol{e}}, \widetilde{\boldsymbol{M}}^{-1}, \boldsymbol{\Sigma_{Y|X}} \right)$, where $\widetilde{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_{\boldsymbol{\mu_X}}^T \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}(\boldsymbol{A})$ and $\widetilde{\boldsymbol{e}} = \boldsymbol{e} + \mathbb{X}_{\boldsymbol{\mu_X}}^T \mathbb{Y}_{\boldsymbol{\mu_Y}}$.

S.4 Generate $\boldsymbol{\Omega}$ from $\mathrm{IW}_m \left(\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}}, \widetilde{\nu}_{\boldsymbol{X}} \right)$, where $\widetilde{\nu}_{\boldsymbol{X}} = \nu_{\boldsymbol{X}} + n$, and $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}} = \boldsymbol{\Psi}_{\boldsymbol{X}} + \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_{\boldsymbol{\mu_X}}^T \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}(\boldsymbol{A})$.

S.5 Generate $\boldsymbol{\Omega}_0$ from $\mathrm{IW}_{p-m}(\widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}}, \widetilde{\nu}_{0,\boldsymbol{X}})$, where $\widetilde{\nu}_{0,\boldsymbol{X}} = \nu_{0,\boldsymbol{X}} + n$, and

$$
\widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}} = \boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A}) \mathbb{X}_{\boldsymbol{\mu_X}}^T \mathbb{X}_{\boldsymbol{\mu_X}} \boldsymbol{\Gamma}_0(\boldsymbol{A}).
$$

S.6 Generate a Markov chain realization for $\boldsymbol{A}$ from stationary density proportional to $\exp(h(\boldsymbol{A}))$, where $h(\boldsymbol{A})$ is the full conditional posterior density of $\boldsymbol{A}$ obtained from (S17) (see (S25)).

Simple Metropolis steps for updating $\boldsymbol{A}$ columnwise are described as follows. Let $\boldsymbol{a}_j \in \mathbb{R}^{p-m}$ denote the j-th column of $\boldsymbol{A}$, $j = 1, \ldots, m$. Given the tuning parameter $\tau > 0$, for $j = i_1, \ldots, i_m$, where $\{i_1, \ldots, i_m\}$ denotes a random permutation of $\{1, \ldots, m\}$, do the following:

(a) Generate $\boldsymbol{a}_j^* \sim \mathrm{N}_{p-m}(\boldsymbol{a}_j, \tau^2 \boldsymbol{I}_{p-m})$. Replace the j-th column of $\boldsymbol{A}$ by $\boldsymbol{a}_j^*$ and call the resulting matrix $\boldsymbol{A}^*$. Calculate $\rho(\boldsymbol{A}^*, \boldsymbol{A}) = \exp\left[h(\boldsymbol{A}^*) - h(\boldsymbol{A})\right]$.

(b) Perform a Bernoulli experiment with probability of success $\min[1, \rho(\boldsymbol{A}^*, \boldsymbol{A})]$. If a success is achieved, update $\boldsymbol{a}_j$ to $\boldsymbol{a}_j^*$; otherwise retain $\boldsymbol{a}_j$.

Once $\boldsymbol{A}$ is updated, $\boldsymbol{C_A}$, $\boldsymbol{D_A}$ and $\boldsymbol{\Sigma_X}$ are updated from (9) and (10).

Remark D.1. Similar to the MCMC sampler for the response envelope model, Algorithm D.1 can be easily modified to account for the cases of $m = 0$ and $m = p$. The parameter $\boldsymbol{A}$ is not present in such cases, and hence the Metropolis step S.6 does not arise. In addition, when $m = 0$, then $\boldsymbol{\eta} = 0, \boldsymbol{\Gamma}_0(\boldsymbol{A}) = \boldsymbol{I}_r$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Omega}_0$, and steps S.3 and S.4 are not needed. On the other hand, when $m = p$, the model degenerates to the standard linear regression model with $\boldsymbol{\Gamma}(\boldsymbol{A}) = \boldsymbol{I}_r$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Omega}$, and step S.5 is to be skipped.

D.2 Proof of Theorem 4.1 (Posterior Propriety)

The proof is similar to the proof for Theorem 3.1. Our objective is to show the integrability of the unnormalized posterior density $\pi(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$ with respect to Lebesgue measure on $\mathbb{R}^r \times \mathbb{R}^p \times \mathbb{S}_+^{r\times r} \times \mathbb{R}^{m\times r} \times \mathbb{S}_+^{m\times m} \times \mathbb{S}_+^{(p-m)\times(p-m)} \times \mathbb{R}^{(r-u)\times u}$. From the log posterior distribution (S17),

$$
\begin{aligned}
&\log \pi(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= -\frac{\nu_{\boldsymbol{Y}} + n + m + r + 1}{2} \log |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| - \frac{\nu_{\boldsymbol{X}} + n + m + 1}{2} \log |\boldsymbol{\Omega}| \\
&\quad - \frac{\nu_{0,\boldsymbol{X}} + n + (p-m) + 1}{2} \log |\boldsymbol{\Omega}_0| \\
&\quad - \frac{1}{2} \operatorname{trace}\left[\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T - \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right) \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T - \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T\right] \\
&\quad - \frac{1}{2} \operatorname{trace}\left[\left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \left(\boldsymbol{\Gamma}(\boldsymbol{A})\, \boldsymbol{\Omega}\, \boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\, \boldsymbol{\Omega}_0\, \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1} \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right)^T\right] \\
&\quad - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T \boldsymbol{M} \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right] \\
&\quad - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \boldsymbol{\Psi}_{\boldsymbol{Y}}\right) - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}_{\boldsymbol{X}}\right) - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Omega}_0^{-1} \boldsymbol{\Psi}_{0,\boldsymbol{X}}\right) \\
&\quad - \frac{1}{2} \operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]. \qquad \text{(S18)}
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\operatorname{trace}\left[\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T - \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right) \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T - \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T\right] \\
&= \operatorname{trace}\Big[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T - \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T \\
&\qquad\qquad (\mathbb{P}_{\mathbf{1}_n} + \mathbb{Q}_{\mathbf{1}_n}) \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{Y}}^T - \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)\Big] \\
&= \operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\overline{\boldsymbol{Y}} - \boldsymbol{\mu}_{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A}) \left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}}\right)\right) \left(\overline{\boldsymbol{Y}} - \boldsymbol{\mu}_{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A}) \left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}}\right)\right)^T\right] \\
&\quad + \operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T \left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)\right] \\
&= n \left(\boldsymbol{\mu}_{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A}) \left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)\right)^T \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A}) \left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)\right) \\
&\quad + \operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T \left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)\right] \qquad \text{(S19)}
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\operatorname{trace}\left[\left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right) \left(\boldsymbol{\Gamma}(\boldsymbol{A})\, \boldsymbol{\Omega}\, \boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\, \boldsymbol{\Omega}_0\, \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1} \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right)^T\right] \\
&= \operatorname{trace}\left[\left(\boldsymbol{\Gamma}(\boldsymbol{A})\, \boldsymbol{\Omega}\, \boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\, \boldsymbol{\Omega}_0\, \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1} \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right)^T (\mathbb{P}_{\mathbf{1}_n} + \mathbb{Q}_{\mathbf{1}_n}) \left(\mathbb{X} - \mathbf{1}_n \boldsymbol{\mu}_{\boldsymbol{X}}^T\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \text{trace}\left[\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(n\left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}}\right)\left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}}\right)^T + \mathbb{X}_c^T\mathbb{X}_c\right)\right] \\
&= n\left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}}\right) \\
&\qquad + \text{trace}\left[\left(\boldsymbol{\Gamma}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0\,\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)\mathbb{X}_c^T\mathbb{X}_c\right] \\
&= n\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right) \\
&\qquad + \text{trace}\left[\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\,\boldsymbol{\Gamma}(\boldsymbol{A})\right] + \text{trace}\left[\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\,\boldsymbol{\Gamma}_0(\boldsymbol{A})\right]. \qquad \text{(S20)}
\end{aligned}
$$

Note that,

$$
\begin{aligned}
&n\left(\boldsymbol{\mu}_{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)\right)^T\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)\right) \\
&\quad + n\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)^T\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right) \\
&= n\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T\boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right) \qquad \text{(S21)}
\end{aligned}
$$

where

$$
\overline{\boldsymbol{Z}} = \begin{pmatrix}\overline{\boldsymbol{X}}\\ \overline{\boldsymbol{Y}}\end{pmatrix},\ \boldsymbol{\mu}_{\boldsymbol{Z}} = \begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{X}}\\ \boldsymbol{\mu}_{\boldsymbol{Y}}\end{pmatrix},\ \text{and}\ \boldsymbol{\Delta} = \begin{pmatrix}\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A}) & \boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\eta}\\ \boldsymbol{\eta}^T\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) & \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} + \boldsymbol{\eta}^T\,\boldsymbol{\Omega}\,\boldsymbol{\eta}\end{pmatrix}.
$$

Also,

$$
\begin{aligned}
&\text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T\left(\mathbb{Y}_c - \mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)\right] \\
&\qquad + \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right] \\
&= \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta} - 2\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{Y}_c + \mathbb{Y}_c^T\mathbb{Y}_c\right)\right] \\
&\qquad + \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta}^T\boldsymbol{M}\boldsymbol{\eta} - 2\boldsymbol{\eta}^T\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e} + \boldsymbol{e}^T\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{M}^{-1}\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right] \\
&= \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta}^T\left(\boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}\right)\boldsymbol{\eta} - 2\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})\left(\mathbb{X}_c^T\mathbb{Y}_c + \boldsymbol{e}\right)\right)\right] \\
&\qquad + \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}^T\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right] \\
&= \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T\check{\boldsymbol{M}}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right] \\
&\qquad + \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}^T\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e} - \check{\boldsymbol{e}}^T\boldsymbol{\Gamma}(\boldsymbol{A})\check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right] \\
&= \text{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\widetilde{\boldsymbol{e}}\right)^T\check{\boldsymbol{M}}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\widetilde{\boldsymbol{e}}\right)\right] + \text{trace}\left(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\widetilde{\boldsymbol{H}}\right) \qquad \text{(S22)}
\end{aligned}
$$

where

$$
\begin{aligned}
\check{\boldsymbol{M}} &= \boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A}) \\
\check{\boldsymbol{e}} &= \mathbb{X}_c^T\mathbb{Y}_c + \boldsymbol{e} \\
\text{and}\ \widetilde{\boldsymbol{H}} &= \mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}^T\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e} - \check{\boldsymbol{e}}^T\boldsymbol{\Gamma}(\boldsymbol{A})\check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}, \qquad \text{(S23)}
\end{aligned}
$$

with $\widetilde{\boldsymbol{H}}$ positive semi definite (Proposition D.2.1). Therefore from (S18), (S19), (S20), (S21), and (S22) we get

$$
\log f(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})
$$

$$
\begin{aligned}
&= \text{const.} - \frac{1}{2}\left[n\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)\right] \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \check{\boldsymbol{M}}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right] \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{Y}} + \widetilde{\boldsymbol{H}}\right)\right] - \frac{\nu_{\boldsymbol{Y}} + n + m + r + 1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{X}} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\right)\right] - \frac{\nu_{\boldsymbol{X}} + n + m + 1}{2}\log|\boldsymbol{\Omega}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}_0(\boldsymbol{A})\right)\right] - \frac{\nu_{0,\boldsymbol{X}} + n + (p-m) + 1}{2}\log|\boldsymbol{\Omega}_0| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right] \qquad \text{(S24)}
\end{aligned}
$$

Since

$$
\begin{aligned}
&\int \exp\left[-\frac{n}{2}\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)\right] d\boldsymbol{\mu}_{\boldsymbol{Z}} \\
&= (2\pi)^{\frac{p+r}{2}} n^{\frac{p+r}{2}} |\boldsymbol{\Delta}|^{1/2} = (2n\pi)^{\frac{p+r}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{1/2} |\boldsymbol{\Omega}|^{1/2} |\boldsymbol{\Omega}_0|^{1/2}
\end{aligned}
$$

and

$$
\begin{aligned}
&\int \exp\left[-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \check{\boldsymbol{M}}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right]\right] d\boldsymbol{\eta} \\
&= (2\pi)^{\frac{mr}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{\frac{m}{2}} |\check{\boldsymbol{M}}|^{-\frac{r}{2}} \\
&= (2\pi)^{\frac{mr}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{\frac{m}{2}} |\boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})|^{-\frac{r}{2}} \le (2\pi)^{\frac{mr}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{\frac{m}{2}} |\boldsymbol{M}|^{-\frac{r}{2}},
\end{aligned}
$$

So

$$
\begin{aligned}
&\log \int\int\int f(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\mu}_{\boldsymbol{X}}\, d\boldsymbol{\mu}_{\boldsymbol{Y}} \\
&\le C_1 + \frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| + \frac{1}{2}\log|\boldsymbol{\Omega}| + \frac{1}{2}\log|\boldsymbol{\Omega}_0| + \frac{m}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{Y}} + \widetilde{\boldsymbol{H}}\right)\right] - \frac{\nu_{\boldsymbol{Y}} + n + m + r + 1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{X}} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\right)\right] - \frac{\nu_{\boldsymbol{X}} + n + m + 1}{2}\log|\boldsymbol{\Omega}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}_0(\boldsymbol{A})\right)\right] - \frac{\nu_{0,\boldsymbol{X}} + n + (p-m) + 1}{2}\log|\boldsymbol{\Omega}_0| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right] \\
&= C_1 - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{Y}} + \widetilde{\boldsymbol{H}}\right)\right] - \frac{(\nu_{\boldsymbol{Y}} + n - 1) + r + 1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{X}} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\right)\right] - \frac{\nu_{\boldsymbol{X}} + n + m + 1}{2}\log|\boldsymbol{\Omega}| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}_0(\boldsymbol{A})\right)\right] - \frac{\nu_{0,\boldsymbol{X}} + n + (p-m) + 1}{2}\log|\boldsymbol{\Omega}_0| \\
&\quad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right].
\end{aligned}
$$

where

$$
C_1 = \left(\frac{p+r}{2} + \frac{mr}{2}\right)\log(2\pi) + \frac{p+r}{2}\log n - \frac{r}{2}\log|\boldsymbol{M}|
$$

Therefore,

$$
\begin{aligned}
&\log \int\int\int\int\int\int f(\boldsymbol{\mu_Y}, \boldsymbol{\mu_X}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\mu_X}\, d\boldsymbol{\mu_Y}\, d\boldsymbol{\Sigma_{Y|X}}\, d\boldsymbol{\Omega}\, d\boldsymbol{\Omega}_0 \\
&\leq C_1 + \frac{(\nu_{\boldsymbol{Y}} + n - 1)r}{2} \log 2 + \log \Gamma_r \left(\frac{\nu_{\boldsymbol{Y}} + n - 1}{2} \right) - \frac{\nu_{\boldsymbol{Y}} + n - 1}{2} \log |\boldsymbol{\Psi_Y} + \widetilde{\boldsymbol{H}}| \\
&\quad + \frac{(\nu_{\boldsymbol{X}} + n - 1)m}{2} \log 2 + \log \Gamma_m \left(\frac{\nu_{\boldsymbol{X}} + n - 1}{2} \right) - \frac{\nu_{\boldsymbol{X}} + n - 1}{2} \log |\boldsymbol{\Psi_X} + \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c^T \mathbb{X}_c \boldsymbol{\Gamma}(\boldsymbol{A})| \\
&\quad + \frac{(\nu_{0,\boldsymbol{X}} + n - 1)m}{2} \log 2 + \log \Gamma_{p-m} \left(\frac{\nu_{0,\boldsymbol{X}} + n - 1}{2} \right) \\
&\quad - \frac{\nu_{0,\boldsymbol{X}} + n - 1}{2} \log |\boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A}) \mathbb{X}_c^T \mathbb{X}_c \boldsymbol{\Gamma}_0(\boldsymbol{A})| - \frac{1}{2} \operatorname{trace} \left[\boldsymbol{K}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T \right] \\
&= C_2 - \frac{\nu_{0,\boldsymbol{X}} + n - 1}{2} \log |\boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A}) \mathbb{X}_c^T \mathbb{X}_c \boldsymbol{\Gamma}_0(\boldsymbol{A})| - \frac{1}{2} \operatorname{trace} \left[\boldsymbol{K}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T \right] \\
&\leq C_2 - \frac{\nu_{\boldsymbol{Y}} + n - 1}{2} \log |\boldsymbol{\Psi_Y}| - \frac{\nu_{\boldsymbol{X}} + n - 1}{2} \log |\boldsymbol{\Psi_X}| \\
&\qquad - \frac{\nu_{0,\boldsymbol{X}} + n - 1}{2} \log |\boldsymbol{\Psi}_{0,\boldsymbol{X}}| - \frac{1}{2} \operatorname{trace} \left[\boldsymbol{K}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T \right] \\
&= C_3 - \frac{1}{2} \operatorname{trace} \left[\boldsymbol{K}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T \right].
\end{aligned}
$$

where

$$
C_2 = C_1 + \frac{(\nu_{\boldsymbol{Y}} + n - 1)r}{2} \log 2 + \log \Gamma_r \left(\frac{\nu_{\boldsymbol{Y}} + n - 1}{2} \right) + \frac{(\nu_{\boldsymbol{X}} + n - 1)m}{2} \log 2 + \log \Gamma_m \left(\frac{\nu_{\boldsymbol{X}} + n - 1}{2} \right)
$$

and

$$
C_3 = C_2 - \frac{\nu_{\boldsymbol{Y}} + n - 1}{2} \log |\boldsymbol{\Psi_Y}| - \frac{\nu_{\boldsymbol{X}} + n - 1}{2} \log |\boldsymbol{\Psi_X}|.
$$

Hence, finally

$$
\begin{aligned}
&\int\int\int\int\int\int\int f(\boldsymbol{\mu_Y}, \boldsymbol{\mu_X}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\mu_X}\, d\boldsymbol{\mu_Y}\, d\boldsymbol{\Sigma_{Y|X}}\, d\boldsymbol{\Omega}\, d\boldsymbol{\Omega}_0\, d\boldsymbol{A} \\
&\leq \exp(C_3) \int \exp \left(-\frac{1}{2} \operatorname{trace} \left[\boldsymbol{K}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0) \boldsymbol{L}^{-1} (\boldsymbol{A} - \boldsymbol{A}_0)^T \right] \right) d\boldsymbol{A} \\
&= \exp(C_3)\, (2\pi)^{\frac{(p-m)m}{2}}\, |\boldsymbol{K}|^{\frac{m}{2}} |\boldsymbol{L}|^{\frac{p-m}{2}} < \infty.
\end{aligned}
$$

This completes the proof. □

Proposition D.2.1. $\widetilde{\boldsymbol{H}} = \mathbb{Y}_c^T \mathbb{Y}_c + \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} - \check{\boldsymbol{e}}^T \boldsymbol{\Gamma} \check{\boldsymbol{M}}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \check{\boldsymbol{e}}$ *is (almost surely) positive semidefinite, where* $\check{\boldsymbol{M}}$ *and* $\check{\boldsymbol{e}}$ *are as defined in* (S23) *in the proof for Theorem 4.1.*

Proof. Define

$$
\boldsymbol{H}_1 = \begin{pmatrix} \mathbb{Y}_c^T \mathbb{Y}_c & \mathbb{Y}_c^T \mathbb{X}_c \boldsymbol{\Gamma}(\boldsymbol{A}) \\ \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c^T \mathbb{Y}_c & \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c \mathbb{X}_c \boldsymbol{\Gamma} \end{pmatrix} = \begin{pmatrix} \mathbb{Y}_c^T \\ \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c \end{pmatrix} \begin{pmatrix} \mathbb{Y}_c^T \\ \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c \end{pmatrix}^T
$$

and

$$
\boldsymbol{H}_2 = \begin{pmatrix} \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T \boldsymbol{e} & \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \\ \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} & \boldsymbol{M} \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{M}^{-1/2} \\ \boldsymbol{M}^{1/2} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{M}^{-1/2} \\ \boldsymbol{M}^{1/2} \end{pmatrix}^T.
$$

From the above representation, it is clear that $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are both positive semi-definite. Therefore, so is

$$
\boldsymbol{H}_1 + \boldsymbol{H}_2 = \begin{pmatrix} \mathbb{Y}_c^T \mathbb{Y}_c + \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T \boldsymbol{e} & \mathbb{Y}_c^T \mathbb{X}_c \boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \\ \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c^T \mathbb{Y}_c + \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} & \boldsymbol{\Gamma}^T(\boldsymbol{A}) \mathbb{X}_c \mathbb{X}_c \boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{M} \end{pmatrix}
$$

This, together with the fact that $\check{\boldsymbol{M}} = \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{M}$ is positive definite (since $\boldsymbol{M}$ is positive definite by assumption), implies that the Schur complement (of $\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A}) + \boldsymbol{M}$),

$$\widetilde{\boldsymbol{H}} = \mathbb{Y}_c^T\mathbb{Y}_c + \boldsymbol{e}^T\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T\boldsymbol{e} - \check{\boldsymbol{e}}^T\boldsymbol{\Gamma}\check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}$$

is also positive semi-definite. This completes the proof. □

D.3 Proof of Theorem 4.2 (Harris Ergodicity)

We note that ϕ-irreducibility and aperiodicity of the Markov chain can be established using similar arguments as used in the proof of Theorem 3.2. Furthermore, when $m = 0$ and/or $m = p$, the algorithm becomes a (full) Gibbs sampler, which together with ϕ-irreducibility, ensures Harris recurrence. We therefore only prove Harris recurrence for the case $1 \leq m \leq p-1$ by showing that the joint posterior density $\pi(\boldsymbol{\mu_X}, \boldsymbol{\mu_Y}, \boldsymbol{\eta}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid, \mathbb{X}, \mathbb{Y})$ is Lebesgue integrable with respect to any $1 \leq k \leq 5+m$ elements of the set of parameter blocks $\{\boldsymbol{\mu_Z}, \boldsymbol{\eta}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_m\}$, where $\boldsymbol{\mu}_{\boldsymbol{Z}}^T = (\boldsymbol{\mu}_{\boldsymbol{X}}^T, \boldsymbol{\mu}_{\boldsymbol{Y}}^T)$ and $\boldsymbol{a}_j$ is the j-th column of $\boldsymbol{A}$. The proof is similar to the proof Harris recurrence in Theorem 3.2. From (S24), it follows that for an appropriately chosen (finite, due to posterior propriety) constant C_0, the joint posterior density is given by

$$\begin{aligned}
&\pi(\boldsymbol{\mu_Y}, \boldsymbol{\mu_X}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= C_0 \exp\left\{-\frac{1}{2}\left[n\left(\boldsymbol{\mu_Z} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu_Z} - \overline{\boldsymbol{Z}}\right)\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \check{\boldsymbol{M}}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right]\right\} \\
&\quad \times |\boldsymbol{\Sigma_{Y|X}}|^{-(\nu_{\boldsymbol{Y}}+n+m+r+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\Psi_Y} + \widetilde{\boldsymbol{H}}\right)\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}+n+m+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Psi_X} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\right)\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}+n+(p-m)+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}_0^{-1}\left(\boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}_0(\boldsymbol{A})\right)\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]\right\} \\
&\leq C_0 \exp\left\{-\frac{1}{2}\left[n\left(\boldsymbol{\mu_Z} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu_Z} - \overline{\boldsymbol{Z}}\right)\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \boldsymbol{M}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right]\right\} \\
&\quad \times |\boldsymbol{\Sigma_{Y|X}}|^{-(\nu_{\boldsymbol{Y}}+n+m+r+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\boldsymbol{\Psi_Y}\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}+n+m+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi_X}\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}+n+(p-m)+1)/2} \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]\right\} \\
&=: f_0(\boldsymbol{\mu_Y}, \boldsymbol{\mu_X}, \boldsymbol{\Sigma_{Y|X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}), \text{ say}
\end{aligned}$$

where

$$\check{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})$$

$$\check{\boldsymbol{e}} = \mathbb{X}_c^T \mathbb{Y}_c + \boldsymbol{e}$$
$$\text{and } \widetilde{\boldsymbol{H}} = \mathbb{Y}_c^T \mathbb{Y}_c + \boldsymbol{e}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{M}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \boldsymbol{e} - \check{\boldsymbol{e}}^T \boldsymbol{\Gamma}(\boldsymbol{A}) \check{\boldsymbol{M}}^{-1} \boldsymbol{\Gamma}^T(\boldsymbol{A}) \check{\boldsymbol{e}}.$$

It will, therefore, be enough to show that f_0 is Lebesgue integrable with respect to any one or more parameter blocks. We note that for any $\boldsymbol{\mu}_{\boldsymbol{Z}} \in \mathbb{R}^{p+r}$,

$$\exp\left\{-\frac{1}{2}\left[n\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)\right]\right\} \leq 1$$

and

$$\int \exp\left\{-\frac{1}{2}\left[n\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)\right]\right\} d\boldsymbol{\mu}_{\boldsymbol{Z}} = (2n\pi)^{(p+r)/2} |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{1/2} |\boldsymbol{\Omega}|^{1/2} |\boldsymbol{\Omega}_0|^{1/2},$$

so that for all $\boldsymbol{\mu}_{\boldsymbol{Z}}^T = (\boldsymbol{\mu}_{\boldsymbol{X}}^T, \boldsymbol{\mu}_{\boldsymbol{Y}}^T) \in \mathbb{R}^{p+r}$,

$$f_0(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \leq f_1^{(0)}(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

and

$$\int f_0(\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\mu}_{\boldsymbol{Z}} = f_1^{(1)}(\boldsymbol{\eta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

where

$$\begin{aligned}
&f_1^{(\xi)}(\boldsymbol{\eta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= (2n\pi)^{\xi(p+r)/2} C_0 \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \boldsymbol{M}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right]\right\} \\
&\quad \times |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{-(\nu_{\boldsymbol{Y}} - \xi + n + m + r + 1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{Y}}\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}} - \xi + n + m + 1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right]\right\} \\
&\quad \times |\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}} - \xi + n + (p-m) + 1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]\right\}
\end{aligned}$$

for $\xi = 0, 1$. Again, for all $\boldsymbol{\eta} \in \mathbb{R}^{m \times r}$

$$\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \boldsymbol{M}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right]\right\} \leq 1$$

and

$$\int \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)^T \boldsymbol{M}\left(\boldsymbol{\eta} - \check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}\right)\right]\right\} d\boldsymbol{\eta} = (2\pi)^{mr/2} |\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{m/2} |\boldsymbol{M}|^{-r/2}$$

which means

$$f_1^{(\xi)}(\boldsymbol{\eta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \leq f_2^{(\xi,0)}(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

for all $\boldsymbol{\eta} \in \mathbb{R}^{m \times r}$, and

$$\int f_1^{(\xi)}(\boldsymbol{\eta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\eta} \leq f_2^{(\xi,1)}(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

where

$$f_2^{(\xi,\kappa)}(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

$$= n^{\xi(p+r)/2}(2\pi)^{\xi(p+r)/2+\kappa mr/2}|M|^{-r/2}C_0\,|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}|^{-(\nu_{\boldsymbol{Y}}-\xi+n+(1-\kappa)m+r+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{Y}}\right]\right\}$$
$$\times|\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}-\xi+n+m+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right]\right\}$$
$$\times|\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}-\xi+n+(p-m)+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right]\right\}$$
$$\times\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\}.$$

It is therefore enough to show that, for $\xi,\kappa=0,1$, $f_2^{(\xi,\kappa)}(\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y})$ is Lebesgue integrable with respect to any $1\le k\le 3+m$ elements of the set $\{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m\}$. The proof is completed by noting that for $\xi,\kappa=0,1$ $f_2^{(\xi,\kappa)}$ is proportional to the product of (independent) $\mathrm{IW}_r(\boldsymbol{\Psi}_{\boldsymbol{Y}},\nu_{\boldsymbol{Y}}-\xi+n+(1-\kappa)m)$, $\mathrm{IW}_m(\boldsymbol{\Psi}_{\boldsymbol{X}},\nu_{\boldsymbol{X}}-\xi+n)$, $\mathrm{IW}_{p-m}(\boldsymbol{\Psi}_{0,\boldsymbol{X}},\nu_{\boldsymbol{X}}0-\xi+n)$ and $\mathrm{MN}_{p-m,m}(\boldsymbol{A}_0,\boldsymbol{K},\boldsymbol{L})$ densities for $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}},\boldsymbol{\Omega},\boldsymbol{\Omega}_0$ and $\boldsymbol{A}$ respectively.

□

D.4 Maximum a Posteriori Estimation for Bayesian Predictor Envelope Model

This section provides an algorithm for MAP estimation for the parameters in the predictor envelope model. Derivations are provided in Section D.4 of this supplement.

Algorithm D.2. Computation of MAP estimators $\{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}},\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}},\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}|\boldsymbol{X}},\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\Omega}},\widehat{\boldsymbol{\Omega}}_0,\widehat{\boldsymbol{A}}\}$ of the predictor envelope model (10) parameters $\{\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\mu}_{\boldsymbol{Y}},\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\}$.

Step 1. Obtain $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}=\overline{\boldsymbol{X}}$ and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}=\overline{\boldsymbol{Y}}$.

Step 2. Find $\widehat{\boldsymbol{A}}=\arg\max_{\boldsymbol{A}\in\mathbb{R}^{(p-m)\times m}}\widetilde{l}(\boldsymbol{A})$, where

$$\widetilde{l}(\boldsymbol{A})=-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]-\frac{\nu_{\boldsymbol{X}}+n+m+1}{2}\log\left|\boldsymbol{\Psi}_{\boldsymbol{X}}+\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})\right|$$
$$-\frac{\nu_{\boldsymbol{Y}}+n+m+r+1}{2}\log\left|\boldsymbol{\Psi}_{\boldsymbol{Y}}+\widetilde{\boldsymbol{H}}\right|-\frac{\nu_{0,\boldsymbol{X}}+n+(p-m)+1}{2}\log\left|\boldsymbol{\Psi}_{0,\boldsymbol{X}}+\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}_0(\boldsymbol{A})\right|.$$

Here $\mathbb{Y}_c=\mathbb{Y}-\mathbf{1}_n\overline{\boldsymbol{Y}}^T$, $\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}=\mathbb{X}-\mathbf{1}_n\overline{\boldsymbol{X}}^T$, $\widetilde{\boldsymbol{H}}=\mathbb{Y}_c^T\mathbb{Y}_c+\boldsymbol{e}^T\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}-\check{\boldsymbol{e}}^T\boldsymbol{\Gamma}(\boldsymbol{A})\check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\check{\boldsymbol{e}}$, $\check{\boldsymbol{M}}=\boldsymbol{M}+\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_c^T\mathbb{X}_c\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\check{\boldsymbol{e}}=\mathbb{X}_c^T\mathbb{Y}_c+\boldsymbol{e}$. Then update $\boldsymbol{C}_{\boldsymbol{A}}$, $\boldsymbol{D}_{\boldsymbol{A}}$, $\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\boldsymbol{\Gamma}_0(\boldsymbol{A})$ as follows

$$\widehat{\boldsymbol{C}}_{\boldsymbol{A}}=\begin{pmatrix}\boldsymbol{I}_m\\ \widehat{\boldsymbol{A}}\end{pmatrix},\quad \widehat{\boldsymbol{D}}_{\boldsymbol{A}}=\begin{pmatrix}-\widehat{\boldsymbol{A}}^T\\ \boldsymbol{I}_{p-m}\end{pmatrix},\quad \boldsymbol{\Gamma}(\widehat{\boldsymbol{A}})=\widehat{\boldsymbol{C}}_{\boldsymbol{A}}\left(\widehat{\boldsymbol{C}}_{\boldsymbol{A}}^T\widehat{\boldsymbol{C}}_{\boldsymbol{A}}\right)^{-1/2},\ \text{and}\ \boldsymbol{\Gamma}_0(\widehat{\boldsymbol{A}})=\widehat{\boldsymbol{D}}_{\boldsymbol{A}}\left(\widehat{\boldsymbol{D}}_{\boldsymbol{A}}^T\widehat{\boldsymbol{D}}_{\boldsymbol{A}}\right)^{-1/2}.$$

Step 3. Compute

(i) $\widehat{\boldsymbol{\eta}}=\check{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}(\widehat{\boldsymbol{A}})^T\check{\boldsymbol{e}}$

(ii) $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}|\boldsymbol{X}}=\dfrac{1}{\nu_{\boldsymbol{Y}}+n+m+r+1}\left(\boldsymbol{\Psi}_{\boldsymbol{Y}}+\widetilde{\boldsymbol{H}}\right)$

(iii) $\widehat{\boldsymbol{\Omega}}=\dfrac{1}{\nu_{\boldsymbol{X}}+n+m+1}\left(\boldsymbol{\Psi}_{\boldsymbol{X}}+\boldsymbol{\Gamma}^T(\widehat{\boldsymbol{A}})\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\widehat{\boldsymbol{A}})\right)$

(iv) $\widehat{\boldsymbol{\Omega}}_0=\dfrac{1}{\nu_{0,\boldsymbol{X}}+n+(p-m)+1}\left(\boldsymbol{\Psi}_{0,\boldsymbol{X}}+\boldsymbol{\Gamma}_0^T(\widehat{\boldsymbol{A}})\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}_0(\widehat{\boldsymbol{A}})\right)$.

D.5 Full conditional posterior distributions for the parameter blocks in the predictor envelope model

1. The joint full conditional density of $(\boldsymbol{\mu}_{\boldsymbol{X}}^T, \boldsymbol{\mu}_{\boldsymbol{Y}}^T)^T$, as obtained from (S17) is given by

$$\begin{aligned}
&\log \pi(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\mu}_{\boldsymbol{Y}} \mid \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) \\
&\quad = \text{const.} - \frac{n}{2} \operatorname{trace}\left[\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)^T \left(\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{X}} - \overline{\boldsymbol{X}}\right)^T\right] \\
&\qquad - \frac{n}{2}\left[\left(\boldsymbol{\mu}_{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}})\right)^T \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{Y}} - \overline{\boldsymbol{Y}} - \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}})\right)\right] \\
&\quad = \text{const.} - \frac{n}{2}\left[\left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T \boldsymbol{\Delta}^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{Z}} - \overline{\boldsymbol{Z}}\right)^T\right]
\end{aligned}$$

   where $\overline{\boldsymbol{Y}} = \mathbf{1}_n^T\mathbb{Y}/n$, $\overline{\boldsymbol{X}} = \mathbf{1}_n^T\mathbb{X}/n$, and as defined before,

$$\overline{\boldsymbol{Z}} = \begin{pmatrix}\overline{\boldsymbol{X}} \\ \overline{\boldsymbol{Y}}\end{pmatrix}, \ \boldsymbol{\mu}_{\boldsymbol{Z}} = \begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\mu}_{\boldsymbol{Y}}\end{pmatrix}, \text{ and } \boldsymbol{\Delta} = \begin{pmatrix}\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A}) & \boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}\,\boldsymbol{\eta} \\ \boldsymbol{\eta}^T\,\boldsymbol{\Omega}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) & \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} + \boldsymbol{\eta}^T\,\boldsymbol{\Omega}\,\boldsymbol{\eta}\end{pmatrix}.$$

   This means,

$$\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\mu}_{\boldsymbol{Y}}\end{pmatrix} \Bigg| \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{Y}, \mathbb{X} \sim \mathrm{N}_{p+r}\left(\begin{pmatrix}\overline{\boldsymbol{X}} \\ \overline{\boldsymbol{Y}}\end{pmatrix}, \frac{1}{n}\boldsymbol{\Delta}\right)$$

2. The posterior conditional log density of $\boldsymbol{\eta}$ given the other parameters is given by

$$\begin{aligned}
&\log \pi(\boldsymbol{\eta} \mid \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{X}, \mathbb{Y}) \\
&\quad = \text{const.} - \frac{n}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}| - \frac{1}{2}\operatorname{trace}\left[(\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}} - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})\,\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\,(\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}} - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})^T\right] \\
&\qquad - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right] \\
&\quad = \text{const.} - \frac{1}{2}\operatorname{trace}\left[\left(\boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\right)\boldsymbol{\eta}\,\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\boldsymbol{\eta}^T - 2\,\boldsymbol{\eta}\,\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{e} + \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}}\right)^T\boldsymbol{\Gamma}(\boldsymbol{A})\right].
\end{aligned}$$

   This means

$$\boldsymbol{\eta} \mid \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{Y}, \mathbb{X} \sim \mathrm{N}_{m,r}\left(\widetilde{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\widetilde{\boldsymbol{e}},\ \widetilde{\boldsymbol{M}}^{-1},\ \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}\right).$$

   where

$$\begin{aligned}
\widetilde{\boldsymbol{M}} &= \boldsymbol{M} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A}) \\
\text{and } \widetilde{\boldsymbol{e}} &= \boldsymbol{e} + \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}}.
\end{aligned}$$

3. Straight-forward calculations show that

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} \mid \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{Y}, \mathbb{X} \sim \mathrm{IW}_r(\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{Y}}, \widetilde{\nu}_{\boldsymbol{Y}})$$

   where

$$\begin{aligned}
\widetilde{\nu}_{\boldsymbol{Y}} &= \nu_{\boldsymbol{Y}} + n + m \\
\text{and } \widetilde{\boldsymbol{\Psi}}_{\boldsymbol{Y}} &= \boldsymbol{\Psi}_{\boldsymbol{Y}} + (\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}} - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})^T\,(\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}} - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}) \\
&\qquad + \left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right).
\end{aligned}$$

4. Simple calculations yield

$$\boldsymbol{\Omega} \mid \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{Y}, \mathbb{X} \sim \mathrm{IW}_m(\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}}, \widetilde{\nu}_{\boldsymbol{X}})$$

where

$$\widetilde{\nu}_{\boldsymbol{X}} = \nu_{\boldsymbol{X}} + n$$
$$\text{and } \widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}} = \boldsymbol{\Psi}_{\boldsymbol{X}} + \boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A}).$$

5. Similarly,

$$\boldsymbol{\Omega}_0 \mid \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\eta}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{A}, \mathbb{Y}, \mathbb{X} \sim \mathrm{IW}_{p-m}(\widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}}, \widetilde{\nu}_{0,\boldsymbol{X}})$$

where

$$\widetilde{\nu}_{0,\boldsymbol{X}} = \nu_{0,\boldsymbol{X}} + n$$
$$\text{and } \widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}} = \boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}_0(\boldsymbol{A}).$$

6. The posterior conditional log density of $\boldsymbol{A}$ given the other parameters is given by

$$\begin{aligned}
&\log \pi(\boldsymbol{A} \mid \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbb{X}, \mathbb{Y}) \\
&\quad = \text{const.} - \frac{1}{2}\,\mathrm{trace}\left[(\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}} - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})\,\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\,(\mathbb{Y}_{\boldsymbol{\mu}_{\boldsymbol{Y}}} - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})^T\right] \\
&\qquad - \frac{1}{2}\,\mathrm{trace}\left[\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\right] \\
&\qquad - \frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\right] \\
&\qquad - \frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]. \qquad \text{(S25)}
\end{aligned}$$

Clearly, rejection sampling from this density will be inefficient, and a more reasonable strategy will be to use a Markov chain simulation technique such as Metropolis Hastings.

D.6 Simulation results for the Predictor Envelope Model

Now we perform a simulation study under the context of the predictor envelope model (10). We set $r = 3$, $p = 15$ and $m_{\text{true}} = 5$. The elements of $\boldsymbol{\mu}_{\boldsymbol{X},\text{true}}$ and $\boldsymbol{\mu}_{\boldsymbol{Y},\text{true}}$ were sampled independently from $\text{Unif}(-10, 10)$, and the elements of $\boldsymbol{\eta}_{\text{true}}$ and $\boldsymbol{A}_{\text{true}}$ were drawn independently from $\text{Unif}(5, 10)$ and $\text{Unif}(0, 5)$, respectively. The matrices $\boldsymbol{\Omega}_{\text{true}}$, $\boldsymbol{\Omega}_{0,\text{true}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X},\text{true}}$ were simulated independently from $5\mathrm{IW}_m(m + 2, 500\boldsymbol{I}_m)$, $\mathrm{IW}_{p-m}(p - m + 2, 0.1\boldsymbol{I}_{p-m})$ and $\mathrm{IW}_r(r + 1, 5\boldsymbol{I}_r)$. The sample size n was taken to be 50, 100, 200, 500 and 1000, and 200 datasets were generated for each sample size. On each dataset, we ran Algorithm D.1 (or its extension) to generate 20,000 MCMC samples (after discarding a burn-in of 10,000 iterations) from the posterior distribution associated with the Bayesian predictor envelope model for each m in $\{0, 1, \ldots, p\}$. Then we computed the approximate posterior probabilities $\Pr(m \mid \text{data})$. A summary of the u estimation is results is displayed in Table S12. The trend in Table S12 is similar to that in the response envelope model case.

Now we compare the MSE and the estimation variance. We used the MCMC posterior mean as the point estimator for the regression coefficient $\boldsymbol{\beta}$, and calculated its variance and MSE from the 200 replications in exactly the same way as in Section 7.1. This computation was carried out for $m = m_{\text{true}}, m_{\text{sel}}$ and p. The results are plotted in Figure S2, and a close look at the results for $\beta_{1,1}$ is taken in Figure S3. Again the results demonstrate the efficiency gains obtained by the Bayesian predictor envelope model for all sample sizes. For smaller sample sizes, the envelope estimator with selected dimension has larger estimation variances and MSEs than the envelope estimator with the true dimension, but is still more efficient than the standard estimator.

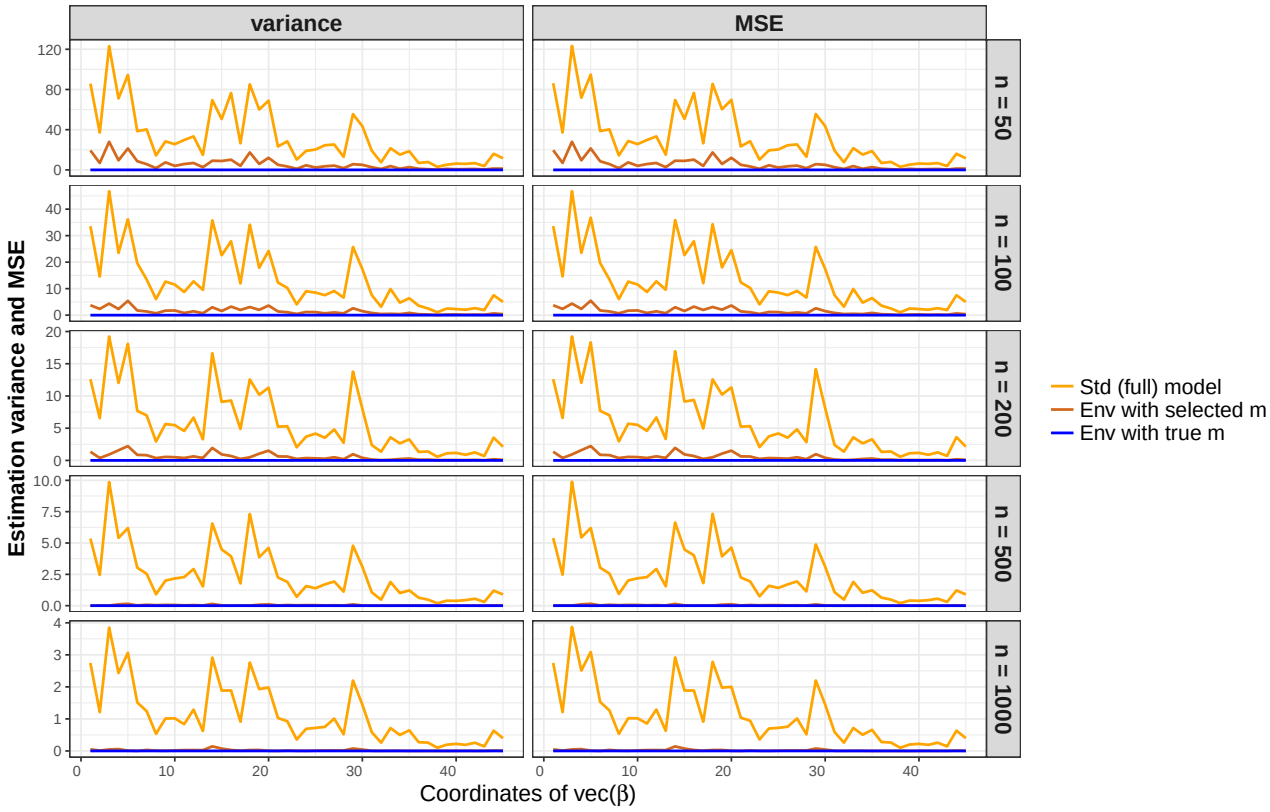
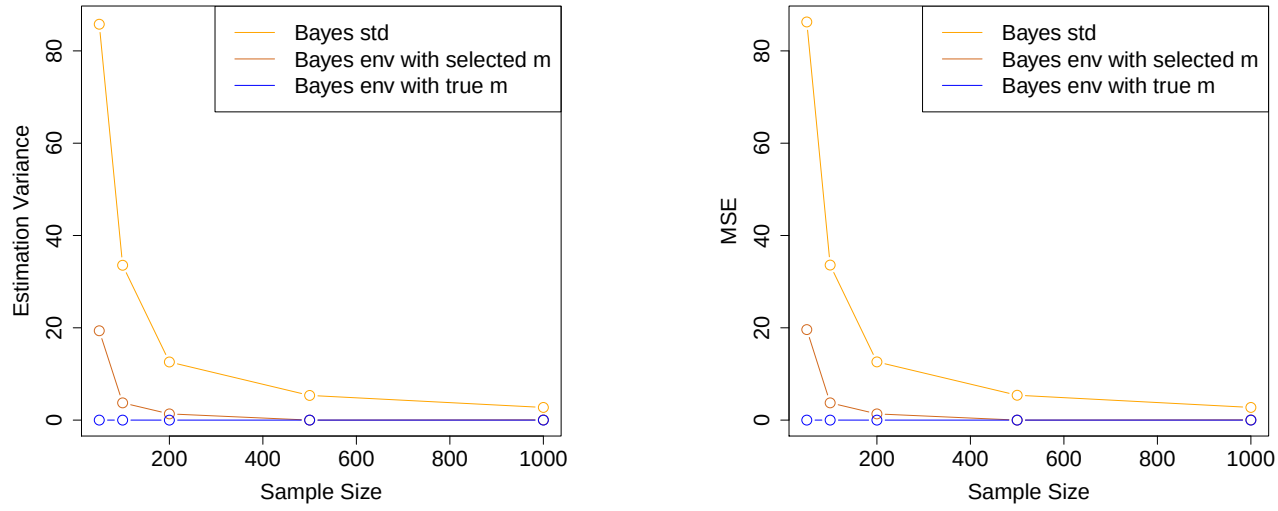


Figure S2: Comparison on estimation variance and MSE for the Bayesian predictor envelope model versus the Bayesian standard linear regression model.



(a) Estimation variances for $\beta_{1,1}$

(b) MSEs for $\beta_{1,1}$

Figure S3: MSE and estimation variances of $\beta_{1,1}$ at different sample sizes.

n	$0 \le u \le 4$	$u = 5$	$u = 6$	$u = 7$	$u = 8$	$9 \le u \le 15$
50	0.001 (0.032)	0.580 (0.384)	0.320 (0.362)	0.080 (0.195)	0.014 (0.066)	0.000 (0.003)
100	0.002 (0.038)	0.733 (0.335)	0.205 (0.293)	0.049 (0.124)	0.004 (0.049)	0.000 (0.000)
200	0.000 (0.000)	0.815 (0.288)	0.163 (0.264)	0.023 (0.104)	0.000 (0.000)	0.000 (0.000)
500	0.000 (0.000)	0.920 (0.184)	0.079 (0.184)	0.001 (0.002)	0.000 (0.000)	0.000 (0.000)
1000	0.000 (0.000)	0.916 (0.196)	0.084 (0.196)	0.000 (0.001)	0.000 (0.000)	0.000 (0.000)

Table S12: approximate posterior probabilities of envelope dimension.

E Various Technical Details and Additional Results on the Bayesian Envelope Probit Model and its Implementation

E.1 Log likelihood, log posterior density, and MCMC sampler for Bayesian envelope probit model

Let $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ be i.i.d. samples from the model (11); then the log-likelihood of the model parameters is given by

$$\begin{aligned}
&\log L(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}) = \text{const.} - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| \\
&- \frac{1}{2}\,\text{trace}\left[\left(\mathbb{X} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)\left(\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)^{-1}\left(\mathbb{X} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{X}}^T\right)^T\right] \\
&+ \sum_{i=1}^n Y_i\,\log\Phi\left(\mu_Y + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}})\right) + \sum_{i=1}^n (1-Y_i)\log\left[1 - \Phi\left(\mu_Y + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}})\right)\right].
\end{aligned} \tag{S26}$$

Under the Bayesian envelope probit model, the log posterior density associated with the prior distributions and the log likelihood (S26) is given by

$$\begin{aligned}
&\log\pi(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= \text{const.} + \sum_{i=1}^n Y_i\,\log\Phi\left(\mu_Y + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}})\right) + \sum_{i=1}^n (1-Y_i)\log\left[1 - \Phi\left(\mu_Y + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\boldsymbol{X}_i - \boldsymbol{\mu}_{\boldsymbol{X}})\right)\right] \\
&- \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\,\text{trace}\left[\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\left(\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\right] \\
&- \frac{1}{2}(\mu_Y - \alpha_Y)^2/\Xi_Y - \frac{1}{2}\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\alpha}_{\boldsymbol{X}}\right)^T\boldsymbol{\Xi}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\alpha}_{\boldsymbol{X}}\right) \\
&- \frac{1}{2}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right) - \frac{\nu_{\boldsymbol{X}} + m + 1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\,\text{trace}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right) \\
&- \frac{\nu_{0,\boldsymbol{X}} + (p-m) + 1}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\,\text{trace}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right) - \frac{1}{2}\,\text{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right].
\end{aligned} \tag{S27}$$

Let $\mathbb{U}^T = (U_1, \ldots, U_n)$ be the vector of augmented data. Then the joint posterior density of the parameters $\{\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}\}$ and the latent data $\mathbb{U}$ is given by

$$\begin{aligned}
&\pi(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{U} \mid \mathbb{X}, \mathbb{Y}) \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\mathbb{U} - \mathbf{1}_n\mu_Y - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)^T\left(\mathbb{U} - \mathbf{1}_n\mu_Y - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta}\right)\right]\right\} \times \prod_{i=1}^n \mathbb{1}\{U_i \ge 0\}^{Y_i}\mathbb{1}\{U_i < 0\}^{1-Y_i}
\end{aligned}$$

$$\times |\boldsymbol{\Omega}|^{-n/2}|\boldsymbol{\Omega}_0|^{-n/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\mathbb{X}_{\boldsymbol{\mu_X}}\left(\boldsymbol{\Gamma}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0\,\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)\mathbb{X}_{\boldsymbol{\mu_X}}^T\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2}(\mu_Y-\alpha_Y)^2/\Xi_Y-\frac{1}{2}\left(\boldsymbol{\mu_X}-\boldsymbol{\alpha_X}\right)^T\boldsymbol{\Xi}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{\mu_X}-\boldsymbol{\alpha_X}\right)\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right\}$$
$$\times |\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}+m+1)/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right)\right\}\times|\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}+(p-m)+1)/2}\exp\left\{-\frac{1}{2}\,\mathrm{trace}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right)\right\}$$
$$\times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\}. \tag{S28}$$

Algorithm E.1. One iteration of a data augmentation Metropolis-within-Gibbs MCMC sampler for updating $\{\boldsymbol{\mu_X}, \mu_Y, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}\}$ when $1 \le m \le p-1$. .

S.1 (**Data augmentation.**) Generate independent U_i from $\mathrm{TN}\left(\mu_Y+\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\boldsymbol{A})(\boldsymbol{X}_i-\boldsymbol{\mu_X}),1,Y_i\right)$, for $i=1,\ldots,n$, and let $\mathbb{U}=(U_1,\ldots,U_n)^T$.

S.2 Generate $\boldsymbol{\mu_Z}=(\boldsymbol{\mu}_{\boldsymbol{X}}^T,\mu_Y)^T$ from $\mathrm{N}_{p+1}\left(\widetilde{\boldsymbol{\alpha}}_{\boldsymbol{Z}},\widetilde{\boldsymbol{\Xi}}_{\boldsymbol{Z}}\right)$, where $\widetilde{\boldsymbol{\Xi}}_Z=\left(n\boldsymbol{\Delta}_{\boldsymbol{Z}}^{-1}+\boldsymbol{\Xi}_{\boldsymbol{Z}}^{-1}\right)^{-1}$, $\widetilde{\boldsymbol{\alpha}}_{\boldsymbol{Z}}=\widetilde{\boldsymbol{\Xi}}_{\boldsymbol{Z}}\left(n\boldsymbol{\Delta}_{\boldsymbol{Z}}^{-1}\overline{\boldsymbol{Z}}+\boldsymbol{\Xi}_{\boldsymbol{Z}}^{-1}\boldsymbol{\alpha}_{\boldsymbol{Z}}\right)$ and

$$\overline{\boldsymbol{Z}}=\begin{pmatrix}\overline{\boldsymbol{X}}\\ \overline{U}\end{pmatrix},\quad \boldsymbol{\Delta}_Z=\begin{pmatrix}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T(\boldsymbol{A}) & \boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\Omega}\boldsymbol{\eta}\\ \boldsymbol{\eta}^T\boldsymbol{\Omega}\boldsymbol{\Gamma}^T(\boldsymbol{A}) & 1+\boldsymbol{\eta}^T\boldsymbol{\Omega}\boldsymbol{\eta}\end{pmatrix},$$

where $\overline{U}=\mathbf{1}_n^T\mathbb{U}/n$. Define $\mathbb{U}_{\mu_Y}=\mathbb{U}-\mu_Y\mathbf{1}_n$.

S.3 Generate $\boldsymbol{\eta}$ from $\mathrm{N}_m\left(\widetilde{\boldsymbol{M}}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\widetilde{\boldsymbol{e}},\widetilde{\boldsymbol{M}}^{-1}\right)$, where $\widetilde{\boldsymbol{M}}=\boldsymbol{M}+\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu_X}}^T\mathbb{X}_{\boldsymbol{\mu_X}}\boldsymbol{\Gamma}(\boldsymbol{A})$ and $\widetilde{\boldsymbol{e}}=\boldsymbol{e}+\mathbb{X}_{\boldsymbol{\mu_X}}^T\mathbb{U}_{\mu_Y}$.

S.4 Generate $\boldsymbol{\Omega}$ from $\mathrm{IW}_m\left(\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}},\widetilde{\nu}_{\boldsymbol{X}}\right)$, where $\widetilde{\nu}_{\boldsymbol{X}}=\nu_{\boldsymbol{X}}+n$, and $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}}=\boldsymbol{\Psi}_{\boldsymbol{X}}+\boldsymbol{\Gamma}^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu_X}}^T\mathbb{X}_{\boldsymbol{\mu_X}}\boldsymbol{\Gamma}(\boldsymbol{A})$.

S.5 Generate $\boldsymbol{\Omega}_0$ from $\mathrm{IW}_{p-m}(\widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}},\widetilde{\nu}_{0,\boldsymbol{X}})$ where $\widetilde{\nu}_{0,\boldsymbol{X}}=\nu_{0,\boldsymbol{X}}+n$ and $\widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}}=\boldsymbol{\Psi}_{0,\boldsymbol{X}}+\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\mathbb{X}_{\boldsymbol{\mu_X}}^T\mathbb{X}_{\boldsymbol{\mu_X}}\boldsymbol{\Gamma}_0(\boldsymbol{A})$.

S.6 Generate Markov chain realizations for $\boldsymbol{A}$ from stationary density proportional to $\exp(\widetilde{h}(\boldsymbol{A}))$, where $\widetilde{h}(\boldsymbol{A})$ is the full conditional posterior density of $\boldsymbol{A}$ obtained from (S28).

Simple Metropolis steps to update $\boldsymbol{A}$ columnwise are described as follows. Let $\boldsymbol{a}_j\in\mathbb{R}^{p-m}$ denote the j-th column of $\boldsymbol{A}$, $j=1,\ldots,m$. Given the tuning parameter $\tau>0$, for $j=i_1,\ldots,i_m$, where $\{i_1,\ldots,i_m\}$ denotes a random permutation of $\{1,\ldots,m\}$, do the following:

(a) Generate $\boldsymbol{a}_j^*\sim\mathrm{N}_{p-m}(\boldsymbol{a}_j,\tau^2\boldsymbol{I}_{p-m})$. Replace the j-th column of $\boldsymbol{A}$ by $\boldsymbol{a}_j^*$ and call the resulting matrix $\boldsymbol{A}^*$. Calculate $\rho(\boldsymbol{A}^*,\boldsymbol{A})=\exp\left[\widetilde{h}(\boldsymbol{A}^*)-\widetilde{h}(\boldsymbol{A})\right]$.

(b) Perform a Bernoulli experiment with probability of success $\min[1,\rho(\boldsymbol{A}^*,\boldsymbol{A})]$. If a success is achieved, update $\boldsymbol{a}_j$ to $\boldsymbol{a}_j^*$; otherwise retain $\boldsymbol{a}_j$.

Once $\boldsymbol{A}$ has been updated, update $\boldsymbol{C_A}$, $\boldsymbol{D_A}$ and $\boldsymbol{\Sigma_X}$ by (9) and (11).

Remark E.1. Similar to the response and predictor envelope models, modifications of Algorithm D.1 to account for the cases $m=0$ and $m=p$ are straightforward. The Metropolis step S.6 is not needed as the parameter $\boldsymbol{A}$ is not present in the model in both cases. In addition, when $m=0$, then $\boldsymbol{\eta}=0,\boldsymbol{\Gamma}_0(\boldsymbol{A})=\boldsymbol{I}_r$ and $\boldsymbol{\Sigma_X}=\boldsymbol{\Omega}_0$, and steps S.3 and S.4 are not needed. On the other hand, when $m=p$, the model degenerates to the standard Bayesian probit regression model with $\boldsymbol{\Gamma}(\boldsymbol{A})=\boldsymbol{I}_r$ and $\boldsymbol{\Sigma_X}=\boldsymbol{\Omega}$, and step S.5 is to be skipped.

E.2 Maximum a Posteriori Estimation under the Envelope probit model

This section proposes an expectation conditional maximization (ECM, Meng and Rubin, 1993) algorithm for finding the posterior mode of the envelope probit model (11) by exploiting the latent data structure and the full conditional distributions.

Algorithm E.2. Computation of the MAP estimators $\{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}, \widehat{\mu}_Y, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Omega}}_0, \widehat{\boldsymbol{A}}\}$ for the parameters of the envelope probit model (11).

S.1 **E Step:** Impute the latent variables

$$\widetilde{U}_i = \begin{cases} \hat{\mu}_i + \dfrac{\phi(\hat{\mu}_i)}{1-\Phi(-\hat{\mu}_i)} & \text{if } Y_i = 1, \\ \hat{\mu}_i - \dfrac{\phi(\hat{\mu}_i)}{\Phi(-\hat{\mu}_i)} & \text{if } Y_i = 0, \end{cases}$$

where $\phi(\cdot)$ denote the standard normal density, and

$$\hat{\mu}_i = \widehat{\mu}_Y + \widehat{\boldsymbol{\eta}}^T \boldsymbol{\Gamma}^T(\widehat{\boldsymbol{A}})(X_i - \widehat{\boldsymbol{\mu}}_X), \quad \text{for } i = 1, \ldots, n.$$

S.2 **CM Steps:** Based on the imputed data $\widetilde{\mathbb{U}} = (\widetilde{U}_1, \ldots, \widetilde{U}_n)^T$ obtained from the E step, do the following

S.2.1 Compute $\hat{\boldsymbol{\mu}}_{\boldsymbol{Z}} = (\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}^T, \hat{\mu}_Y)^T = \tilde{\boldsymbol{\alpha}}_{\boldsymbol{Z}}$ where $\tilde{\boldsymbol{\alpha}}_{\boldsymbol{Z}} = \widetilde{\boldsymbol{\Xi}}_{\boldsymbol{Z}} \left(n\boldsymbol{\Delta}_{\boldsymbol{Z}}^{-1} \overline{\boldsymbol{Z}} + \boldsymbol{\Xi}_{\boldsymbol{Z}}^{-1} \boldsymbol{\alpha}_{\boldsymbol{Z}} \right)$ with $\widetilde{\boldsymbol{\Xi}}_Z = \left(n\boldsymbol{\Delta}_{\boldsymbol{Z}}^{-1} + \boldsymbol{\Xi}_{\boldsymbol{Z}}^{-1} \right)^{-1}$, and $\overline{\boldsymbol{Z}}^T = (\overline{\boldsymbol{X}}^T, \overline{U})$, with $\overline{U} = \mathbf{1}_n^T \mathbb{U}/n$. Define $\widetilde{\mathbb{U}}_{\hat{\mu}_Y} = \widetilde{\mathbb{U}} - \hat{\mu}_Y \mathbf{1}_n$ and $\widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}} = \mathbb{X} - \mathbf{1}_n \hat{\boldsymbol{\mu}}_{\boldsymbol{X}}^T$.

S.2.2 Define $\hat{\boldsymbol{\eta}} = \widetilde{\boldsymbol{M}}^{-1} \boldsymbol{\Gamma}^T(\hat{\boldsymbol{A}}) \tilde{\boldsymbol{e}}$, where $\widetilde{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{\Gamma}^T(\hat{\boldsymbol{A}}) \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}}^T \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}} \boldsymbol{\Gamma}(\hat{\boldsymbol{A}})$ and $\tilde{\boldsymbol{e}} = \boldsymbol{e} + \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}}^T \mathbb{U}_{\mu_Y}$.

S.2.3 Compute $\hat{\boldsymbol{\Omega}} = \widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}}/\widetilde{\nu}_{\boldsymbol{X}}$ and $\hat{\boldsymbol{\Omega}}_0 = \widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}}/\widetilde{\nu}_{0,\boldsymbol{X}}$ where $\widetilde{\nu}_{\boldsymbol{X}} = \nu_{\boldsymbol{X}} + n$, $\widetilde{\nu}_{0,\boldsymbol{X}} = \nu_{0,\boldsymbol{X}} + n$, $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{X}} = \boldsymbol{\Psi}_{\boldsymbol{X}} + \boldsymbol{\Gamma}^T(\hat{\boldsymbol{A}}) \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}}^T \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}} \boldsymbol{\Gamma}(\hat{\boldsymbol{A}})$ and $\widetilde{\boldsymbol{\Psi}}_{0,\boldsymbol{X}} = \boldsymbol{\Psi}_{0,\boldsymbol{X}} + \boldsymbol{\Gamma}_0^T(\hat{\boldsymbol{A}}) \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}}^T \widetilde{\mathbb{X}}_{\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}} \boldsymbol{\Gamma}_0(\hat{\boldsymbol{A}})$.

S.2.4 Maximize the full conditional (log) density of $\boldsymbol{A}$ given $\hat{\boldsymbol{\mu}}_{\boldsymbol{X}}, \hat{\mu}_Y, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Omega}}$, and $\hat{\boldsymbol{\Omega}}_0$, and call the maximizer $\hat{\boldsymbol{A}}$.

E.3 Proof of Theorem 5.1

Using arguments similar to the ones used in the proof of Theorem 3.2, ϕ-irreducibility and aperiodicity of the Markov chain can be established. Moreover, if $m = 0$ or $m = p$, the algorithm becomes a Gibbs sampler, which, coupled with ϕ-irreducibility, ensures Harris recurrence. In order to prove Harris recurrence when $1 \le m \le p-1$, we show that the joint density $\pi(\boldsymbol{\mu}_{\boldsymbol{Z}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_m, \mathbb{U} \mid \mathbb{X}, \mathbb{Y})$ as provided in (S28) is Lebesgue integrable with respect to any $1 \le k \le 5+m$ elements from the set $\{\boldsymbol{\mu}_{\boldsymbol{Z}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_m, \mathbb{U}\}$, where $\mu_{\boldsymbol{Z}}^T = (\boldsymbol{\mu}_{\boldsymbol{X}}^T, \mu_Y)$ and $\boldsymbol{a}_j$ denote the jth column of $\boldsymbol{A}$. From (S28), it follows that for an appropriate (finite, due to posterior propriety) constant C_0,

$$\begin{aligned}
&\pi(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A}, \mathbb{U} \mid \mathbb{X}, \mathbb{Y}) \\
&= C_0 \exp\left\{ -\frac{1}{2} \left[(\mathbb{U} - \mathbf{1}_n \mu_Y - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}} \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{\eta})^T (\mathbb{U} - \mathbf{1}_n \mu_Y - \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}} \boldsymbol{\Gamma}(\boldsymbol{A}) \boldsymbol{\eta}) \right] \right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-n/2} |\boldsymbol{\Omega}_0|^{-n/2} \exp\left\{ -\frac{1}{2} \operatorname{trace} \left[\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}} \left(\boldsymbol{\Gamma}\, \boldsymbol{\Omega}^{-1}\, \boldsymbol{\Gamma}^T(\boldsymbol{A}) + \boldsymbol{\Gamma}_0\, \boldsymbol{\Omega}_0^{-1}\, \boldsymbol{\Gamma}_0^T(\boldsymbol{A}) \right) \mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T \right] \right\} \\
&\quad \times \exp\left\{ -\frac{1}{2} (\mu_Y - \alpha_Y)^2 / \Xi_Y - \frac{1}{2} (\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\alpha}_{\boldsymbol{X}})^T \boldsymbol{\Xi}_{\boldsymbol{X}}^{-1} (\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\alpha}_{\boldsymbol{X}}) \right\}
\end{aligned}$$

$$
\begin{aligned}
&\times \exp\left\{-\frac{1}{2}\boldsymbol{M}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right\}\\
&\times |\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}+m+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right)\right\}\times|\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}+(p-m)+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right)\right\}\\
&\times\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\}\times\prod_{i=1}^{n}\mathbb{1}\{U_i\geq 0\}^{Y_i}\mathbb{1}\{U_i<0\}^{1-Y_i}\\
\leq& C_0\exp\left\{-\frac{1}{2}\left[(\mathbb{U}-\mathbf{1}_n\mu_Y-\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})^T(\mathbb{U}-\mathbf{1}_n\mu_Y-\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})\right]\right\}\\
&\times\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\left(\boldsymbol{\Gamma}(\boldsymbol{A})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma}^T(\boldsymbol{A})+\boldsymbol{\Gamma}_0(\boldsymbol{A})\,\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Gamma}_0^T(\boldsymbol{A})\right)\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}^T\right]\right\}\\
&\times\exp\left\{-\frac{1}{2}(\mu_Y-\alpha_Y)^2/\Xi_Y-\frac{1}{2}(\boldsymbol{\mu}_{\boldsymbol{X}}-\boldsymbol{\alpha}_{\boldsymbol{X}})^T\boldsymbol{\Xi}_{\boldsymbol{X}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{X}}-\boldsymbol{\alpha}_{\boldsymbol{X}})\right\}\\
&\times\exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right\}\\
&\times|\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}+m+n+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right)\right\}\times|\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}+(p-m)+n+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right)\right\}\\
&\times\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\}\\
=:& f_0(\boldsymbol{\mu}_{\boldsymbol{Y}},\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A},\mathbb{U}\mid\mathbb{X},\mathbb{Y}),\ \text{say,}
\end{aligned}
$$

and it will be enough to show that f_0 is integrable with respect to one or more parameter (and latent data) blocks. Now for any $\mathbb{U}\in\mathbb{R}^n$

$$
\exp\left\{-\frac{1}{2}\left[(\mathbb{U}-\mathbf{1}_n\mu_Y-\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})^T(\mathbb{U}-\mathbf{1}_n\mu_Y-\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})\right]\right\}\leq 1,
$$

and

$$
\int\exp\left\{-\frac{1}{2}\left[(\mathbb{U}-\mathbf{1}_n\mu_Y-\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})^T(\mathbb{U}-\mathbf{1}_n\mu_Y-\mathbb{X}_{\boldsymbol{\mu}_{\boldsymbol{X}}}\boldsymbol{\Gamma}(\boldsymbol{A})\boldsymbol{\eta})\right]\right\}d\mathbb{U}=(2\pi)^{n/2}.
$$

Therefore,

$$
f_0(\mu_Y,\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A},\mathbb{U}\mid\mathbb{X},\mathbb{Y})\leq f_1^{(0)}(\mu_Y,\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y})
$$

for all $\mathbb{U}\in\mathbb{R}^n$, and

$$
\int f_0(\mu_Y,\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A},\mathbb{U}\mid\mathbb{X},\mathbb{Y})\,d\mathbb{U}=f_1^{(1)}(\mu_Y,\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y}),
$$

where

$$
\begin{aligned}
&f_1^{(\xi)}(\mu_Y,\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}_0,\boldsymbol{A}\mid\mathbb{X},\mathbb{Y})\\
=&(2\pi)^{\xi n/2}C_0\exp\left\{-\frac{1}{2}(\mu_Y-\alpha_Y)^2/\Xi_Y-\frac{1}{2}(\boldsymbol{\mu}_{\boldsymbol{X}}-\boldsymbol{\alpha}_{\boldsymbol{X}})^T\boldsymbol{\Xi}_{\boldsymbol{X}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{X}}-\boldsymbol{\alpha}_{\boldsymbol{X}})\right\}\\
&\times\exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T\boldsymbol{M}\left(\boldsymbol{\eta}-\boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right\}\\
&\times|\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}}+n+m+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right)\right\}\\
&\times|\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}}+n+(p-m)+1)/2}\exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right)\right\}\\
&\times\exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A}-\boldsymbol{A}_0)^T\right]\right\}
\end{aligned}
$$

for $\xi = 0, 1$. Furthermore, for all $\boldsymbol{\eta} \in \mathbb{R}^m$

$$\exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T \boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right\} \leq 1$$

and

$$\int \exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)^T \boldsymbol{M}\left(\boldsymbol{\eta} - \boldsymbol{M}^{-1}\boldsymbol{\Gamma}^T(\boldsymbol{A})\boldsymbol{e}\right)\right\} d\boldsymbol{\eta} = (2\pi)^{m/2}|\boldsymbol{M}|^{1/2},$$

which implies that

$$f_1^{(\xi)}(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \leq f_2^{(\xi,0)}(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

for all $\boldsymbol{\eta} \in \mathbb{R}^m$, and

$$\int f_1^{(\xi)}(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})\, d\boldsymbol{\eta} = f_2^{(\xi,1)}(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$$

where

$$\begin{aligned}
&f_2^{(\xi,\kappa)}(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y}) \\
&= (2\pi)^{(\xi n + \kappa m)/2}|\boldsymbol{M}|^{\kappa/2} C_0 \exp\left\{-\frac{1}{2}(\mu_Y - \alpha_Y)^2/\Xi_Y - \frac{1}{2}\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\alpha}_{\boldsymbol{X}}\right)^T \boldsymbol{\Xi}_{\boldsymbol{X}}^{-1}\left(\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\alpha}_{\boldsymbol{X}}\right)\right\} \\
&\quad \times |\boldsymbol{\Omega}|^{-(\nu_{\boldsymbol{X}} + n + m + 1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}_{\boldsymbol{X}}\right)\right\} \\
&\quad \times |\boldsymbol{\Omega}_0|^{-(\nu_{0,\boldsymbol{X}} + n + (p-m) + 1)/2} \exp\left\{-\frac{1}{2}\operatorname{trace}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_{0,\boldsymbol{X}}\right)\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\operatorname{trace}\left[\boldsymbol{K}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)\boldsymbol{L}^{-1}(\boldsymbol{A} - \boldsymbol{A}_0)^T\right]\right\}
\end{aligned}$$

for $\kappa = 0, 1$. Thus, it is enough to show that for $\xi, \kappa = 0, 1$, $f_2^{(\xi,\kappa)}(\mu_Y, \boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{A} \mid \mathbb{X}, \mathbb{Y})$ is Lebesgue integrable with respect to any $1 \leq k \leq 3 + m$ elements of $\{\mu_{\boldsymbol{Z}}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_m\}$. The proof is completed by noting that for $\xi, \kappa = 0, 1$, $f_2^{(\xi,\kappa)}$ is the proportional to the product of independent $\mathrm{N}_p(\boldsymbol{\alpha}_{\boldsymbol{X}}, \boldsymbol{\Xi}_{\boldsymbol{X}})$, $\mathrm{N}(\alpha, \Xi_Y)$, $\mathrm{IW}_m(\boldsymbol{\Psi}_{\boldsymbol{X}}, \nu_{\boldsymbol{X}} + n)$, $\mathrm{IW}_{p-m}(\boldsymbol{\Psi}_{\boldsymbol{X}}, \nu_{\boldsymbol{X}} + n)$ and $\mathrm{MN}_{p-m,m}(\boldsymbol{A}_0, \boldsymbol{K}, \boldsymbol{L})$ densities for $\boldsymbol{\mu}_{\boldsymbol{X}}, \mu_Y, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0$ and $\boldsymbol{A}$ respectively.

□

E.4 Additional simulation results for the Bayesian Envelope Probit Model

In this section, we investigate the Bayesian envelope probit model (11). We generated data from model (11) with $p = 10$ and $m_{\text{true}} = 2$. We took $\boldsymbol{\mu}_{\boldsymbol{X},\text{true}} = 0$, $\boldsymbol{\mu}_{Y,\text{true}} = 1$, $\boldsymbol{\eta}_{\text{true}} = (1, 1)^T$ and $\boldsymbol{\Omega}_{\text{true}} = \boldsymbol{I}_2$. The matrix $\boldsymbol{\Omega}_{0,\text{true}}$ was a diagonal matrix with diagonal elements $0.002 + 0.014k$, for $k = 0, 1, \ldots, 7$. The elements of $\boldsymbol{A}_{\text{true}}$ were independent Unif$(0, 5)$ variates, and the sample size was taken to be 50, 100, 200, 500 and 1000. We generated 200 datasets for each sample size. On each dataset, we ran Algorithm E.1 (or its extension) to generate 10,000 MCMC samples (after discarding a burn-in of 40,000 iterations) from the posterior distribution associated with the Bayesian envelope probit model for $m = 2$ and $m = 10$ (full or standard model). We obtained the Bayes point estimator by taking the MCMC posterior mean for each model. The estimation variances and MSE's for the estimators of $\boldsymbol{\beta}$ are plotted side by side in Figure S4. A more detailed picture for β_1, the first element of $\boldsymbol{\beta} \in \mathbb{R}^p$, is shown in Figure S5. Both figures display noticeable efficiency gains obtained by the envelope probit model. In Figure S5, it appears that the line for the envelope probit model and the line for the standard model almost overlap with each other when the sample size is large. However,

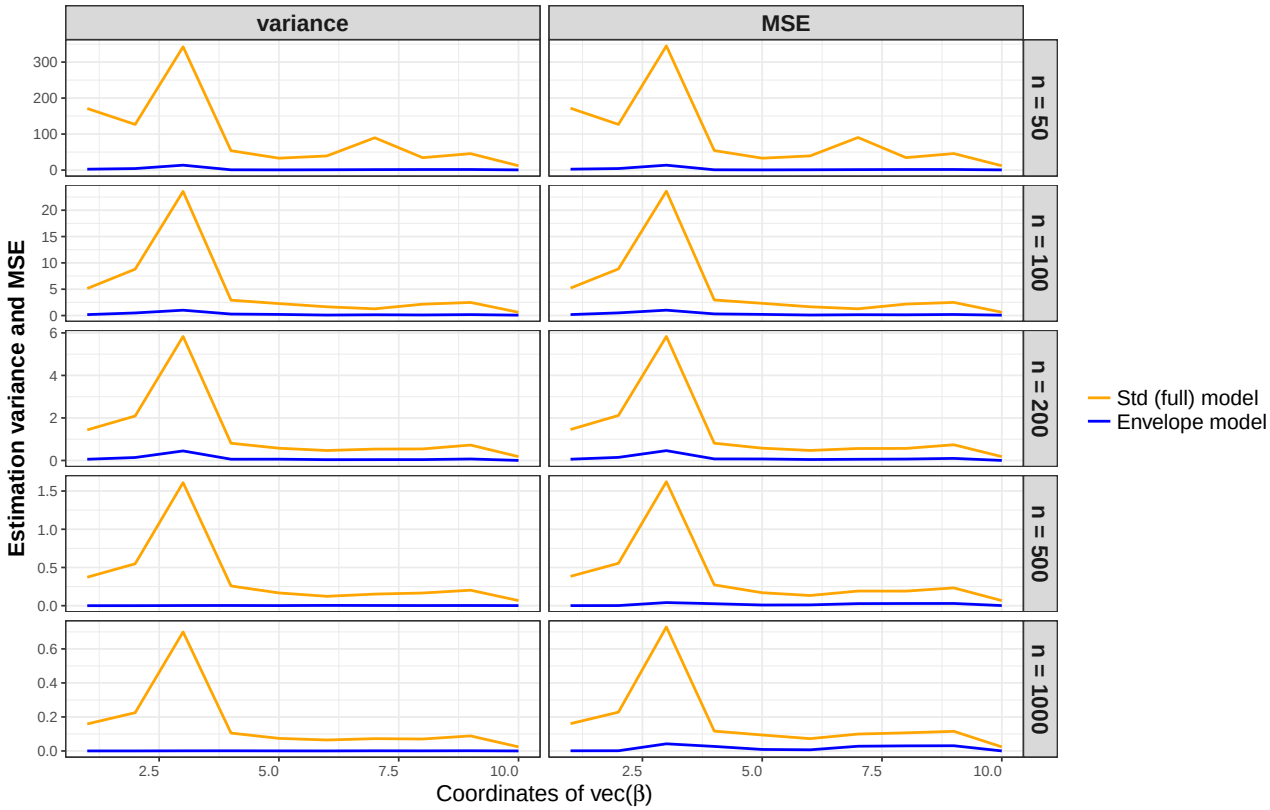
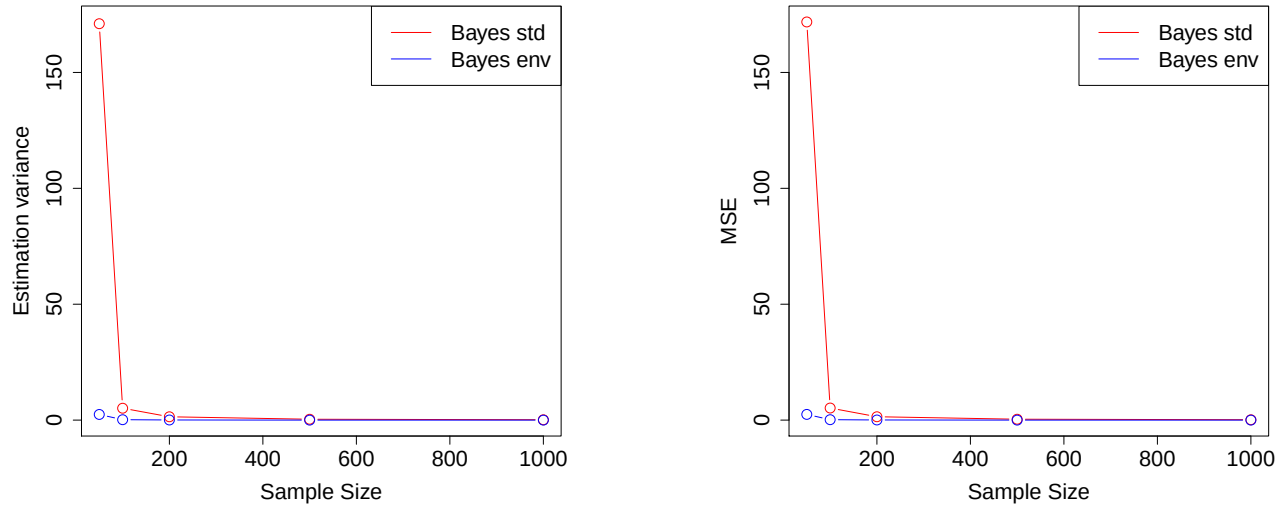


Figure S4: Comparison on estimation variance and MSE for the Bayesian envelope probit model versus the Bayesian probit model.



(a) Estimation variances for $\beta_{1,1}$

(b) MSEs for $\beta_{1,1}$

Figure S5: MSE and estimation variances of β_1 at different sample sizes.

they are actually different, as is indicated in Figure S4. This is because when the sample size is small, the standard probit model has very large estimation variance and MSE. This happens when $m = p$ or m is close to p. In such cases, the log-likelihood is sometimes infinite, which invalidates the approxiamte BIC-based $\Pr(m \mid \text{data})$ computation. Therefore we omitted the dimension estimation results in the current settings.

References

Andersson, S. A. and G. G. Wojnar (2004). Wishart distributions on homogeneous cones. *Journal of Theoretical Probability 17*(4), 781–818.

Ben-Israel, A. (1999). The change-of-variables formula using matrix volume. *SIAM Journal on Matrix Analysis and Applications 21*(1), 300–312.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017, January). Stan: A Probabilistic Programming Language. *Journal of Statistical Software 76*, 1–32.

Cook, R., I. Helland, and Z. Su (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society, Ser. B 75*(5), 851–877.

Cook, R. D., L. Forzani, and Z. Su (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis 150*, 42–54.

Cook, R. D., L. Forzani, and X. Zhang (2015). Envelopes and reduced-rank regression. *Biometrika 102*(2), 439–456.

Cook, R. D., B. Li, and F. Chiaromonte (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica 20*, 927–960.

Dutta, S. (2012). Multiplicative random walk metropolis-hastings on the real line. *Sankhya B 74*(2), 315–342.

Eck, D. J. and R. D. Cook (2017). Weighted envelope estimation to handle variability in model selection. *Biometrika 104*(3), 743–749.

Files, B. T., M. Strelioff, and R. Bonnevie (2019). Bayesian reduced-rank regression with stan. Technical report, ARMY RESEARCH LAB ADELPHI MD Playa Vista United States.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*. Springer series in statistics New York.

Frühwirth-Schnatter, S. (2011). *Dealing with Label Switching under Model Uncertainty*, Chapter 10, pp. 213–239. John Wiley & Sons Ltd.

Gelman, A., D. Lee, and J. Guo (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics 40*(5), 530–543.

Geyer, C. J. (1998). Markov chain Monte Carlo lecture notes.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika 82*(4), 711–732.

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the american statistical association 90*(430), 773–795.

Khare, K. and J. P. Hobert (2012). Geometric ergodicity of the gibbs sampler for bayesian quantile regression. *Journal of Multivariate Analysis 112*, 108–116.

Khare, K., S. Pal, and Z. Su (2017). A Bayesian approach for envelope models. *The Annals of Statistics 45*(1), 196–222.

Kucukelbir, A., D. Tran, R. Ranganath, A. Gelman, and D. M. Blei (2017). Automatic differentiation variational inference. *Journal of machine learning research*.

Lee, M. and Z. Su (2020). *Renvlp: Computing Envelope Estimators*. R package version 2.9.

Ma, Y. and L. Zhu (2013). Efficiency loss and the linearity condition in dimension reduction. *Biometrika 100*(2), 371–383.

Mardia, K. V. and P. E. Jupp (2009). *Directional statistics*, Volume 494. John Wiley & Sons.

Meng, X.-L. and D. B. Rubin (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika 80*(2), 267–278.

Peng, J., J. Zhu, A. Bergamaschi, W. Han, D.-Y. Noh, J. R. Pollack, and P. Wang (2010, March). Regularized Multivariate Regression for Identifying Master Predictors with Application to Integrative Genomics Study of Breast Cancer. *The annals of applied statistics 4*(1), 53–77.

Roberts, G. O. and J. S. Rosenthal (2006). Harris recurrence of metropolis-within-gibbs and trans-dimensional markov chains. *The Annals of Applied Probability 16*(4), 2123–2139.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Ser. B 64*(4), 583–639.

Su, Z., G. Zhu, X. Chen, and Y. Yang (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika 103*(3), 579–593.

Tierney, L. (1994, 12). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701–1728.