

SUPPLEMENT TO “A BAYESIAN APPROACH FOR ENVELOPE MODELS”

BY KSHITIJ KHARE*, SUBHADIP PAL, AND ZHIHUA SU†

University of Florida

A. REJECTION SAMPLING PROCEDURE TO SAMPLE FROM $GB_{2,2}(\mathbf{A}, \mathbf{B})$ DISTRIBUTION

Suppose that the matrix \mathbf{Z} has a $GB_{2,2}(\mathbf{A}, \mathbf{B})$ distribution. As \mathbf{Z} is an orthogonal matrix, it can be written in the form

$$\mathbf{Z} = \begin{bmatrix} \cos(\theta^*) & s \sin(\theta^*) \\ \sin(\theta^*) & -s \cos(\theta^*) \end{bmatrix}$$

where $\theta^* \in [-\pi, \pi)$ and $s \in \{+1, -1\}$. Using transformations, it can be easily shown that θ^* and s are independent, $P(s = 1) = P(s = -1) = 1/2$, and the marginal density of θ^* is proportional to

$$h_1(\theta^*) = e^{-(\mathbf{A}_{11} + \mathbf{B}_{22})\cos^2(\theta^*) - (\mathbf{A}_{22} + \mathbf{B}_{11})\sin^2(\theta^*) - (\mathbf{A}_{12} + \mathbf{A}_{21} - \mathbf{B}_{12} - \mathbf{B}_{21})\sin(\theta^*) \cos(\theta^*)}$$

for every $\theta^* \in [-\pi, \pi)$. Let us consider a transformation $\theta^* \rightarrow (\theta, s_1)$ defined by

$$\theta = \begin{cases} \theta^* & \text{if } -\frac{\pi}{2} \leq \theta^* < \frac{\pi}{2}, \\ \theta^* + \pi & \text{if } -\pi \leq \theta^* < -\frac{\pi}{2}, \\ \theta^* - \pi & \text{if } \frac{\pi}{2} \leq \theta^* < \pi, \end{cases}$$

and

$$s_1 = \begin{cases} 1 & \text{if } -\frac{\pi}{2} \leq \theta^* \leq \frac{\pi}{2}, \\ -1 & \text{otherwise.} \end{cases}$$

Using the fact that $h_1(\theta^*) = h_1(\theta^* - \pi)$ if $0 \leq \theta^* \leq \pi/2$, and $h_1(\theta^*) = h_1(\theta^* + \pi)$ if $-\pi/2 \leq \theta^* \leq 0$, it can be easily shown that θ and s_1 are independent, $P(s_1 = 1) = P(s_1 = -1) = 1/2$, and the marginal density of θ is proportional to

$$e^{-(\mathbf{A}_{11} + \mathbf{B}_{22})\cos^2(\theta) - (\mathbf{A}_{22} + \mathbf{B}_{11})\sin^2(\theta) - (\mathbf{A}_{12} + \mathbf{A}_{21} - \mathbf{B}_{12} - \mathbf{B}_{21})\sin(\theta) \cos(\theta)}$$

for every $\theta \in [-\pi/2, \pi/2)$. Now, consider transformation $\theta \rightarrow \omega$ defined by

$$\omega = \begin{cases} \cos(\theta) & \text{if } 0 \leq \theta < \frac{\pi}{2}, \\ -\cos(\theta) & \text{if } -\frac{\pi}{2} \leq \theta < 0. \end{cases}$$

*Supported by NSF Grant DMS-1106084.

†Supported by NSF Grant DMS-1407460.

The Jacobian of this transformation is given by $\frac{1}{\sqrt{1-\omega^2}}$. Hence the density of ω is proportional to

$$h(\omega) = \frac{1}{\sqrt{1-\omega^2}} e^{a\omega^2 + b\sqrt{\omega^2(1-\omega^2)}} I_{[\omega \geq 0]} + \frac{1}{\sqrt{1-\omega^2}} e^{a\omega^2 - b\sqrt{\omega^2(1-\omega^2)}} I_{[\omega < 0]},$$

where $a := (\mathbf{A}_{22} + \mathbf{B}_{11} - \mathbf{A}_{11} - \mathbf{B}_{22})$ and $b := (\mathbf{B}_{12} + \mathbf{B}_{21} - \mathbf{A}_{12} - \mathbf{A}_{21})$. Note that if $\mathbf{Z} \sim GB_{2,2}(\mathbf{A}, \mathbf{B})$, then $\tilde{\mathbf{Z}} \sim GB_{2,2}(\mathbf{B}, \mathbf{A})$, where $\tilde{\mathbf{Z}}$ is obtained from \mathbf{Z} by switching its two columns. Hence, without loss of generality, we can assume $a \leq 0$. We now consider the following cases to obtain an upper bound for $h(\omega)$.

- $a < 0$ and $b < 0$: Using the inequality $e^{-x} \leq \frac{0.573}{x^{0.223}}$ on $[0, 1]$, we get

$$h(\omega) \leq \frac{\beta^2}{(ab)^\gamma} (\omega^2)^{-\frac{3}{2}\gamma} (1-\omega^2)^{-\frac{\gamma+1}{2}} I_{[\omega \geq 0]} + \frac{\beta e^{\frac{|b|}{2}}}{|a|^\gamma} (\omega^2)^{-\gamma} (1-\omega^2)^{-\frac{1}{2}} I_{[\omega < 0]}.$$

- $a < 0$ and $b > 0$: In this case we get

$$h(\omega) \leq \frac{\beta e^{\frac{b}{2}}}{|a|^\gamma} (\omega^2)^{-\gamma} (1-\omega^2)^{-\frac{1}{2}} I_{[\omega \geq 0]} + \frac{\beta^2}{|ab|^\gamma} (\omega^2)^{-\frac{3}{2}\gamma} (1-\omega^2)^{-\frac{\gamma+1}{2}} I_{[\omega < 0]}$$

Hence ω can be sampled using a rejection sampler with a proposal distribution that is a mixture of the square root of a Beta distribution and the negative square root of another Beta distribution. Hence, based on the above arguments, a sample of the matrix \mathbf{Z} from the $GB_{2,2}(\mathbf{A}, \mathbf{B})$ distribution can be obtained as follows.

- Generate s and s_1 independently, both taking values from $\{-1, 1\}$ with equal probability.
- Generate ω from the rejection sampler presented above.
- If $\omega > 0$,

$$\mathbf{Z} = s_1 \begin{bmatrix} \omega & s \sqrt{1-\omega^2} \\ \sqrt{1-\omega^2} & -s \omega \end{bmatrix}$$

- If $\omega < 0$,

$$\mathbf{Z} = s_1 \begin{bmatrix} \omega & -s \sqrt{1-\omega^2} \\ -\sqrt{1-\omega^2} & -s \omega \end{bmatrix}$$

In practice, we have observed, that for some values of a, b the rejection sampler can be quite inefficient. In such cases we directly sample θ by using a discrete approximation of the distribution of θ over the range $[-\pi, \pi)$.

B. A PROCEDURE TO GET THE EMPIRICAL PRIOR

Let $\hat{\Sigma}_{\mathbf{Y}}$ be the sample covariance matrix \mathbf{Y} , $\hat{\Sigma}_{\text{res}}$ be the sample covariance matrix of the residuals from the OLS fit of \mathbf{Y} on \mathbf{X} , and $\hat{\beta}_{\text{ols}}$ be the OLS estimator of β . Following [14], for a given u , define the objective function as

$$f_{\text{obj}}(\mathbf{H}) = \log |\mathbf{H}^T \hat{\Sigma}_{\text{res}} \mathbf{H}| + \log |\mathbf{H}^T \hat{\Sigma}_{\mathbf{Y}}^{-1} \mathbf{H}|,$$

where \mathbf{H} is $r \times u$ semi-orthogonal matrix. Notice that $f_{obj}(\mathbf{H})$ depends on \mathbf{H} only through its span. By [5], the envelope subspace is spanned by the eigenspace of Σ in population. Therefore, we go through all C_u^r (r choose u) combinations of u eigenvectors of Σ to find the combination that minimizes f_{obj} , and we collect the u eigenvectors in this combination as the columns of $\tilde{\Gamma}$. Other eigenvectors form columns of $\tilde{\Gamma}_0$. If the number C_u^r is too large, we start with a random combination, and sequentially update each element in it by searching through the remaining eigenvectors. Repeat the sequential search several times. After we have $(\tilde{\Gamma}, \tilde{\Gamma}_0)$, $\tilde{\Omega} = \tilde{\Gamma}^T \hat{\Sigma}_{\text{res}} \tilde{\Gamma}$ and $\tilde{\Omega}_0 = \tilde{\Gamma}_0^T \hat{\Sigma}_{\mathbf{Y}} \tilde{\Gamma}_0$. If $\tilde{\Omega}$ and $\tilde{\Omega}_0$ are not diagonal matrices, we perform a spectral decomposition to both matrices $\tilde{\Omega} = \mathbf{T} \mathbf{\Lambda} \mathbf{T}^T$ and $\tilde{\Omega}_0 = \mathbf{T}_0 \mathbf{\Lambda}_0 \mathbf{T}_0^T$, where $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_0$ are diagonal matrices with the diagonal elements arranged in decreasing order. Then $\Gamma^* = \tilde{\Gamma} \mathbf{T}$, $\Gamma_0^* = \tilde{\Gamma}_0 \mathbf{T}_0$, ω^* are diagonal elements in $\mathbf{\Lambda}$, ω_0^* are diagonal elements in $\mathbf{\Lambda}_0$ and $\eta^* = \Gamma^{*T} \hat{\beta}_{\text{ols}}$. Now, $((\eta^*, (\Gamma^*, \Gamma_0^*), \omega^*, \omega_0^*))$ can be used to obtain the empirical prior as described in Section 3.1. For the case $n < r$, the matrix $\hat{\Sigma}_{\mathbf{Y}}$ is computationally singular. Hence, $f_{obj}(\mathbf{H})$ is not well-defined. To deal with this situation, we use another method to get $\tilde{\Gamma}$. Let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be eigenvectors of $\hat{\Sigma}_{\mathbf{Y}}$, then the columns of $\tilde{\Gamma}$ are the u eigenvectors that maximize the norm of $\mathbf{v}_i^T \hat{\beta}_{\text{ols}}$. Having $\tilde{\Gamma}$, $((\eta^*, (\Gamma^*, \Gamma_0^*), \omega^*, \omega_0^*))$ can be obtained in the same way as above.

C. SAMPLING FROM THE POSTERIOR DENSITY

C.1. Simulating from the full conditional posterior density of μ . It follows by (11) and (12) (see proof of Theorem 1 in Section 5.2) that the conditional posterior density of μ given $((\Gamma, \Gamma_0), \eta, \omega, \omega_0)$ is

$$\pi(\mu \mid \eta, (\Gamma, \Gamma_0), \omega, \omega_0, \mathbb{Y}) \propto e^{-\frac{1}{2} \text{tr}\{n(\mu - \bar{\mathbf{Y}})^T (\mathbf{\Gamma} \mathbf{\Omega} \mathbf{\Gamma}^T + \mathbf{\Gamma}_0 \mathbf{\Omega}_0 \mathbf{\Gamma}_0^T)^{-1} (\mu - \bar{\mathbf{Y}})\}},$$

which is a multivariate normal distribution $N_r(\bar{\mathbf{Y}}, \Sigma/n)$, recall that $\Sigma = \mathbf{\Gamma} \mathbf{\Omega} \mathbf{\Gamma}^T + \mathbf{\Gamma}_0 \mathbf{\Omega}_0 \mathbf{\Gamma}_0^T$. Therefore we can simulate from the posterior distribution of μ by making a draw from this multivariate normal distribution.

C.2. Simulating from the full conditional posterior density of η . Firstly, note that from (11) and (13) (see proof of Theorem 1 in Section 5.2) the full conditional posterior density of η given $((\Gamma, \Gamma_0), \omega, \omega_0)$ is given by

$$(C.1) \quad \pi(\eta \mid (\Gamma, \Gamma_0), \omega, \omega_0, \mathbb{Y}) \propto e^{-\frac{1}{2} \text{tr}(\mathbf{\Omega}^{-1} (\eta - \mathbf{\Gamma}^T \tilde{\mathbf{e}}) (\mathbb{X}^T \mathbb{X} + \mathbf{C}) (\eta - \mathbf{\Gamma}^T \tilde{\mathbf{e}})^T)},$$

where $\tilde{\mathbf{e}} = (\mathbb{Y}_c^T \mathbb{X} + \mathbf{eC}) (\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1}$. It follows from (C.1) that we can simulate from the full conditional posterior distribution of η given $((\Gamma, \Gamma_0), \omega, \omega_0)$ by making a draw from the $MN_{u,p}(\mathbf{\Gamma}^T \tilde{\mathbf{e}}, \mathbf{\Omega}, (\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1})$ distribution.

C.3. Simulating from the posterior density of $((\Gamma, \Gamma_0), \omega, \omega_0)$. We now focus on simulating from the posterior density of $((\Gamma, \Gamma_0), \omega, \omega_0)$. It follows from (14) (see proof of

Theorem 1 in Section 5.2) that the posterior density of $((\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0)$ is given by

$$\begin{aligned} \pi((\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0 \mid \mathbb{Y}) &\propto \frac{n^{\frac{r}{2}} |\boldsymbol{\Omega}|^{-\frac{n-1}{2}} |\boldsymbol{\Omega}_0|^{-\frac{n-1}{2}} |(\mathbb{X}^T \mathbb{X} + \mathbf{C})|^{-\frac{u}{2}}}{(\sqrt{2\pi})^{r(n-1)-up}} e^{-\frac{1}{2} \text{tr}(\boldsymbol{\Omega}^{-1} \mathbf{\Gamma}^T \tilde{\mathbf{G}} \mathbf{\Gamma})} \\ &e^{-\frac{1}{2} \text{tr}(\boldsymbol{\Omega}_0^{-1} \mathbf{\Gamma}_0^T (\mathbb{Y}_c^T \mathbb{Y}_c) \mathbf{\Gamma}_0)} \prod_{i=1}^u \omega_i^{-\alpha-1} e^{-\frac{\lambda}{\omega_i}} \prod_{i=1}^{r-u} \omega_{0,i}^{-\alpha_0-1} e^{-\frac{\lambda_0}{\omega_{0,i}}} \\ &e^{-\frac{1}{2} \text{tr}(\mathbf{D}^{-1} \mathbf{O}^T \mathbf{G} \mathbf{O})}, \end{aligned} \quad (\text{C.2})$$

where $\tilde{\mathbf{G}} = \mathbb{Y}_c^T \mathbb{Y}_c + \mathbf{e} \mathbf{C} \mathbf{e}^T - \tilde{\mathbf{e}} (\mathbb{X}^T \mathbb{X} + \mathbf{C}) \tilde{\mathbf{e}}^T$. Although it is not possible to directly simulate from this density, we will derive a Gibbs sampling based procedure to generate approximate samples from this density. To achieve this objective, we need to derive various conditional densities associated with the density in (C.2).

Note that the entries of $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$ are arranged in decreasing order. Let $1 \leq i \leq u$ be arbitrarily fixed. Let $\boldsymbol{\omega}^{-i}$ denote the $u-1$ dimensional vector obtained by removing the i^{th} entry of $\boldsymbol{\omega}$. It follows from (16), (17) (see proof of Theorem 1) and (C.2) that the conditional posterior density of ω_i given $(\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}^{-i}, \boldsymbol{\omega}_0$ is given by

$$\pi(\omega_i \mid (\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}^{-i}, \boldsymbol{\omega}_0, \mathbb{Y}) \propto \omega_i^{-\frac{n+2\alpha+1}{2}} e^{-\frac{(\mathbf{\Gamma}^T \tilde{\mathbf{G}} \mathbf{\Gamma})_{ii} + 2\lambda}{2\omega_i}} \mathbf{1}_{\omega_i \in (\omega_{i+1}, \omega_{i-1})}. \quad (\text{C.3})$$

Here $\omega_0 = \infty$, $\omega_{u+1} = 0$ and $(\mathbf{\Gamma}^T \tilde{\mathbf{G}} \mathbf{\Gamma})_{ii}$ is the i^{th} diagonal element of the matrix $\mathbf{\Gamma}^T \tilde{\mathbf{G}} \mathbf{\Gamma}$. It follows from (C.3) that we can simulate from the conditional posterior density of ω_i given $(\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}^{-i}, \boldsymbol{\omega}_0$ by making a draw from the

$$\text{truncated-Inverse-Gamma} \left(\frac{n+2\alpha-1}{2}, \frac{(\mathbf{\Gamma}^T \tilde{\mathbf{G}} \mathbf{\Gamma})_{ii} + 2\lambda}{2}, \omega_{i+1}, \omega_{i-1} \right)$$

distribution.

Let $1 \leq i \leq r-u$ be arbitrarily fixed. Let $\boldsymbol{\omega}_0^{-i}$ denote the $r-u-1$ dimensional vector obtained by removing the i^{th} entry of $\boldsymbol{\omega}_0$. It follows from (16), (18) (see proof of Theorem 1) and (C.2) that the conditional posterior density of $\omega_{0,i}$ given $(\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0^{-i}$ is given by

$$\pi(\omega_{0,i} \mid (\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0^{-i}, \mathbb{Y}) \propto \omega_{0,i}^{-\frac{n+2\alpha_0+1}{2}} e^{-\frac{(\mathbf{\Gamma}_0^T (\mathbb{Y}_c^T \mathbb{Y}_c) \mathbf{\Gamma}_0)_{ii} + 2\lambda_0}{2\omega_{0,i}}} \mathbf{1}_{\omega_{0,i} \in (\omega_{0,i+1}, \omega_{0,i-1})} \quad (\text{C.4})$$

Here $\omega_{0,0} = \infty$, $\omega_{0,r-u+1} = 0$ and $(\mathbf{\Gamma}_0^T (\mathbb{Y}_c^T \mathbb{Y}_c) \mathbf{\Gamma}_0)_{ii}$ is the i^{th} diagonal element of the matrix $\mathbf{\Gamma}_0^T (\mathbb{Y}_c^T \mathbb{Y}_c) \mathbf{\Gamma}_0$. It follows from (C.4) that we can simulate from the conditional posterior density of $\omega_{0,i}$ given $(\mathbf{\Gamma}, \mathbf{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0^{-i}$ by making a draw from the truncated-Inverse-Gamma $\left(\frac{n+2\alpha_0-1}{2}, \frac{(\mathbf{\Gamma}_0^T (\mathbb{Y}_c^T \mathbb{Y}_c) \mathbf{\Gamma}_0)_{ii} + 2\lambda_0}{2}, \omega_{0,i+1}, \omega_{0,i-1} \right)$ distribution.

Let $\mathbf{O}_{\cdot i}$ and $\mathbf{O}_{\cdot j}$ denote the i^{th} and j^{th} columns of the matrix $\mathbf{O} = [\mathbf{\Gamma} : \mathbf{\Gamma}_0]$ respectively, where $1 \leq i < j \leq r$. Let $\mathbf{O}^{-(i,j)}$ denote the matrix obtained by removing the i^{th} and j^{th} columns of \mathbf{O} .

- **Case 1:** If $1 \leq i \leq u$ and $u + 1 \leq j \leq r$, it follows from (C.2) that the conditional posterior density (on the space $S_{r,2}^+$) of $[\mathbf{O}_i : \mathbf{O}_j]$ given $(\boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbf{O}^{-(i,j)})$ is proportional to

$$(C.5) \quad e^{-\frac{1}{2}\mathbf{O}_i^T\left(\frac{\tilde{\mathbf{G}}}{\omega_i} + \frac{\mathbf{G}}{\mathbf{D}_{ii}}\right)\mathbf{O}_i - \frac{1}{2}\mathbf{O}_j^T\left(\frac{\mathbb{Y}_c^T\mathbb{Y}_c}{\omega_{0,j-u}} + \frac{\mathbf{G}}{\mathbf{D}_{jj}}\right)\mathbf{O}_j} 1_{[\mathbf{O}_i : \mathbf{O}_j]^T \mathbf{O}^{-(i,j)} = 0}.$$

Note that the above density is invariant under arbitrary changes to the signs of the columns of the matrix $[\mathbf{O}_i : \mathbf{O}_j]$. Hence, if we simulate from the density (on the space $S_{r,2}$) which is proportional to

$$(C.6) \quad e^{-\frac{1}{2}\mathbf{O}_i^T\left(\frac{\tilde{\mathbf{G}}}{\omega_i} + \frac{\mathbf{G}}{\mathbf{D}_{ii}}\right)\mathbf{O}_i - \frac{1}{2}\mathbf{O}_j^T\left(\frac{\mathbb{Y}_c^T\mathbb{Y}_c}{\omega_{0,j-u}} + \frac{\mathbf{G}}{\mathbf{D}_{jj}}\right)\mathbf{O}_j} 1_{[\mathbf{O}_i : \mathbf{O}_j]^T \mathbf{O}^{-(i,j)} = 0},$$

and then change the signs of the simulated matrix to ensure that the maximum entry (in absolute value) in each column is positive, the resulting matrix will correspond to a simulation from the density in (C.5). Let $\mathbf{N} \in S_{r,2}$ be such that $\mathbf{N}^T \mathbf{O}^{-(i,j)} = 0$. Let $\mathbf{Z} = [\mathbf{Z}_1 : \mathbf{Z}_2]$ be a random matrix (on the space $S_{2,2}$) with density proportional to

$$(C.7) \quad e^{-\frac{1}{2}\mathbf{Z}_1^T \mathbf{N}^T \left(\frac{\tilde{\mathbf{G}}}{\omega_i} + \frac{\mathbf{G}}{\mathbf{D}_{ii}}\right) \mathbf{N} \mathbf{Z}_1 - \frac{1}{2}\mathbf{Z}_2^T \mathbf{N}^T \left(\frac{\mathbb{Y}_c^T\mathbb{Y}_c}{\omega_{0,j-u}} + \frac{\mathbf{G}}{\mathbf{D}_{jj}}\right) \mathbf{N} \mathbf{Z}_2}.$$

It follows that the density of $\mathbf{N}\mathbf{Z}$ (on the space $S_{r,2}$) is given by (C.6). Hence, we can simulate from the conditional posterior density of $[\mathbf{O}_i : \mathbf{O}_j]$ given $(\boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbf{O}^{-(i,j)})$ by simulating from the

$$GB_{2,2} \left(\frac{1}{2} \mathbf{N}^T \left(\frac{\tilde{\mathbf{G}}}{\omega_i} + \frac{\mathbf{G}}{\mathbf{D}_{ii}} \right) \mathbf{N}, \frac{1}{2} \mathbf{N}^T \left(\frac{\mathbb{Y}_c^T \mathbb{Y}_c}{\omega_{0,j-u}} + \frac{\mathbf{G}}{\mathbf{D}_{jj}} \right) \mathbf{N} \right)$$

distribution, multiplying the resulting matrix by \mathbf{N} , and then changing the signs of the columns so that the maximum entry (in absolute value) in each column is positive.

- **Case 2:** If $1 \leq i < j \leq u$ it follows from (C.2) that the conditional posterior density (on the space $S_{r,2}^+$) of $[\mathbf{O}_i : \mathbf{O}_j]$ given $(\boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbf{O}^{-(i,j)})$ is proportional to

$$(C.8) \quad e^{-\frac{1}{2}\mathbf{O}_i^T\left(\frac{\tilde{\mathbf{G}}}{\omega_i} + \frac{\mathbf{G}}{\mathbf{D}_{ii}}\right)\mathbf{O}_i - \frac{1}{2}\mathbf{O}_j^T\left(\frac{\tilde{\mathbf{G}}}{\omega_j} + \frac{\mathbf{G}}{\mathbf{D}_{jj}}\right)\mathbf{O}_j} 1_{[\mathbf{O}_i : \mathbf{O}_j]^T \mathbf{O}^{-(i,j)} = 0}$$

By a similar analysis as in the previous case, we can simulate from the conditional posterior density of $[\mathbf{O}_i : \mathbf{O}_j]$ given $(\boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbf{O}^{-(i,j)})$ by simulating from the

$$GB_{2,2} \left(\frac{1}{2} \mathbf{N}^T \left(\frac{\tilde{\mathbf{G}}}{\omega_i} + \frac{\mathbf{G}}{\mathbf{D}_{ii}} \right) \mathbf{N}, \frac{1}{2} \mathbf{N}^T \left(\frac{\tilde{\mathbf{G}}}{\omega_j} + \frac{\mathbf{G}}{\mathbf{D}_{jj}} \right) \mathbf{N} \right)$$

distribution, multiplying the resulting matrix by \mathbf{N} , and then changing the signs of the columns so that the maximum entry (in absolute value) in each column is positive.

- **Case 3:** If $u+1 \leq i < j \leq r$, it follows from (C.2) that the conditional posterior density (on the space $S_{r,2}^+$) of $[\mathbf{O}_{\cdot i} : \mathbf{O}_{\cdot j}]$ given $(\boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbf{O}^{-(i,j)})$ is proportional to

$$(C.9) \quad e^{-\frac{1}{2} \mathbf{O}_{\cdot i}^T \left(\frac{\mathbb{Y}_c^T \mathbb{Y}_c}{\omega_{0,i-u}} + \frac{\mathbf{G}}{\mathbf{D}_{ii}} \right) \mathbf{O}_{\cdot i} - \frac{1}{2} \mathbf{O}_{\cdot j}^T \left(\frac{\mathbb{Y}_c^T \mathbb{Y}_c}{\omega_{0,j-u}} + \frac{\mathbf{G}}{\mathbf{D}_{jj}} \right) \mathbf{O}_{\cdot j}} 1_{[\mathbf{O}_{\cdot i} : \mathbf{O}_{\cdot j}]^T \mathbf{O}^{-(i,j)} = 0}$$

By a similar analysis as in the previous case, we can simulate from the conditional posterior density of $[\mathbf{O}_{\cdot i} : \mathbf{O}_{\cdot j}]$ given $(\boldsymbol{\omega}, \boldsymbol{\omega}_0, \mathbf{O}^{-(i,j)})$ by simulating from the

$$GB_{2,2} \left(\frac{1}{2} \mathbf{N}^T \left(\frac{\mathbb{Y}_c^T \mathbb{Y}_c}{\omega_{0,i-u}} + \frac{\mathbf{G}}{\mathbf{D}_{ii}} \right) \mathbf{N}, \frac{1}{2} \mathbf{N}^T \left(\frac{\mathbb{Y}_c^T \mathbb{Y}_c}{\omega_{0,j-u}} + \frac{\mathbf{G}}{\mathbf{D}_{jj}} \right) \mathbf{N} \right)$$

distribution, multiplying the resulting matrix by \mathbf{N} , and then changing the signs of the columns so that the maximum entry (in absolute value) in each column is positive.

Using this analysis we now specify a systematic scan Gibbs sampling procedure to generate approximate samples from the posterior density of $((\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0)$ (specified in (C.2)). We start at a given initial value of the parameters, and repeat the following steps.

- For $i = 1, 2, \dots, u$, update ω_i by sampling from the conditional density in (C.3).
- For $i = 1, 2, \dots, r-u$, update $\omega_{0,i}$ by sampling from the conditional density in (C.4).
- For every pair (i, j) such that $1 \leq i < j \leq r$, depending on where i and j lie, update $\mathbf{O}_{\cdot i}$ and $\mathbf{O}_{\cdot j}$ by sampling from the conditional density in (C.5), (C.8) or (C.9).

Alternatively, one can also use a random scan version of the Gibbs sampler as follows. For every iteration, the first two steps in the algorithm above are the same. However, at the final step, we randomly choose a pair (i, j) , and updates the i^{th} and j^{th} columns of $\mathbf{O} = [\boldsymbol{\Gamma} : \boldsymbol{\Gamma}_0]$ by sampling from the appropriate conditional distribution. It follows easily that the Markov chains corresponding to both the random scan and systematic scan Gibbs samplers described above have the density in (C.2) as a stationary density. Theorem 2 shows that both the random scan and the systematic scan versions of the Gibbs sampler for the generalized matrix Bingham distribution are in fact Harris ergodic. This provides theoretical guarantee that the Gibbs sampling algorithms provide approximate samples from the density in (C.2).

D. ITERATIVE ALGORITHM FOR COMPUTING POSTERIOR MODE

In this section, we provide a method to compute the posterior mode. Note that to obtain the posterior mode, we need to maximize the function

$$l = \log\{\pi((\boldsymbol{\mu}, \boldsymbol{\eta}, (\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0), \boldsymbol{\omega}, \boldsymbol{\omega}_0) \mid \mathbb{Y})\}.$$

Since \mathbf{X} is centered, it follows from (11) that irrespective of the values of the other parameters, l is maximized at $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}$. Substitute $\boldsymbol{\mu} = \bar{\mathbf{Y}}$ in l , and denoting the resulting partially

maximized version of l by l_1 , we get

$$\begin{aligned}
l_1 &= \text{constant} - \frac{n}{2} \log |\mathbf{\Omega}| - \frac{n}{2} \log |\mathbf{\Omega}_0| - \\
&\quad \frac{1}{2} \text{tr}\{(\mathbb{Y}_c - \mathbb{X}\boldsymbol{\eta}^T \mathbf{\Gamma}^T)(\mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^T + \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^T)^{-1}(\mathbb{Y}_c - \mathbb{X}\boldsymbol{\eta}^T \mathbf{\Gamma}^T)^T\} \\
&\quad - \frac{1}{2} \text{tr}\{\mathbf{\Omega}^{-1}(\boldsymbol{\eta} - \mathbf{\Gamma}^T \mathbf{e})\mathbf{C}(\boldsymbol{\eta} - \mathbf{\Gamma}^T \mathbf{e})^T\} - \frac{1}{2} \text{tr}(\mathbf{D}^{-1}\mathbf{O}^T \mathbf{G}\mathbf{O}) \\
&\quad - (\alpha + 1) \log |\mathbf{\Omega}| - (\alpha_0 + 1) \log |\mathbf{\Omega}_0| - \text{tr}(\lambda\mathbf{\Omega}^{-1}) - \text{tr}(\lambda_0\mathbf{\Omega}_0^{-1}) \\
&= \text{constant} - \frac{n + 2\alpha + 2}{2} \log |\mathbf{\Omega}| - \frac{n + 2\alpha_0 + 2}{2} \log |\mathbf{\Omega}_0| - \text{tr}(\lambda\mathbf{\Omega}^{-1}) - \text{tr}(\lambda_0\mathbf{\Omega}_0^{-1}) \\
&\quad - \frac{1}{2} \text{tr}\{(\mathbb{Y}_c \mathbf{\Gamma} - \mathbb{X}\boldsymbol{\eta}^T)\mathbf{\Omega}^{-1}(\mathbb{Y}_c \mathbf{\Gamma} - \mathbb{X}\boldsymbol{\eta}^T)^T\} - \frac{1}{2} \text{tr}(\mathbb{Y}_c \mathbf{\Gamma}_0 \mathbf{\Omega}_0^{-1} \mathbf{\Gamma}_0^T \mathbb{Y}_c^T) \\
&\quad - \frac{1}{2} \text{tr}\{\mathbf{\Omega}^{-1}(\boldsymbol{\eta} - \mathbf{\Gamma}^T \mathbf{e})\mathbf{C}(\boldsymbol{\eta} - \mathbf{\Gamma}^T \mathbf{e})^T\} - \frac{1}{2} \text{tr}(\mathbf{D}^{-1}\mathbf{O}^T \mathbf{G}\mathbf{O}).
\end{aligned}$$

Note that

$$\frac{\partial l}{\partial \boldsymbol{\eta}} = -\mathbf{\Omega}^{-1}(\boldsymbol{\eta}\mathbb{X}^T - \mathbf{\Gamma}^T \mathbb{Y}_c^T)\mathbb{X} - \mathbf{\Omega}^{-1}(\boldsymbol{\eta} - \mathbf{\Gamma}^T \mathbf{e})\mathbf{C}.$$

Let $\hat{\boldsymbol{\eta}} = \mathbf{\Gamma}^T(\mathbb{Y}_c^T \mathbb{X} + \mathbf{e}\mathbf{C})(\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1}$. Also, if

$$\begin{aligned}
\mathbf{A} &= (\mathbb{Y}_c \mathbf{\Gamma} - \mathbb{X}\hat{\boldsymbol{\eta}}^T)^T(\mathbb{Y}_c \mathbf{\Gamma} - \mathbb{X}\hat{\boldsymbol{\eta}}^T) + (\hat{\boldsymbol{\eta}} - \mathbf{\Gamma}^T \mathbf{e})\mathbf{C}(\hat{\boldsymbol{\eta}} - \mathbf{\Gamma}^T \mathbf{e})^T + 2\lambda \mathbf{I}_u \\
&= \mathbf{\Gamma}^T \left\{ [\mathbb{Y}_c - \mathbb{X}(\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1}(\mathbb{Y}_c^T \mathbb{X} + \mathbf{e}\mathbf{C})^T]^T [\mathbb{Y}_c - \mathbb{X}(\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1}(\mathbb{Y}_c^T \mathbb{X} + \mathbf{e}\mathbf{C})^T] \right. \\
&\quad \left. + [(\mathbb{Y}_c^T \mathbb{X} + \mathbf{e}\mathbf{C})(\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1} - \mathbf{e}]\mathbf{C}[(\mathbb{Y}_c^T \mathbb{X} + \mathbf{e}\mathbf{C})(\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1} - \mathbf{e}]^T + 2\lambda \mathbf{I}_r \right\} \mathbf{\Gamma}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{B} &= \mathbf{\Gamma}_0^T \mathbb{Y}_c^T \mathbb{Y}_c \mathbf{\Gamma}_0 + 2\lambda_0 \mathbf{I}_{r-u} \\
&= \mathbf{\Gamma}_0^T (\mathbb{Y}_c^T \mathbb{Y}_c + 2\lambda_0 \mathbf{I}_r) \mathbf{\Gamma}_0,
\end{aligned}$$

then

$$\frac{\partial l}{\partial \omega_i} = -\frac{n + 2\alpha + 2}{2} \frac{1}{\omega_i} + \frac{1}{2} \frac{(\mathbf{A})_{ii}}{\omega_i^2},$$

and

$$\frac{\partial l}{\partial \omega_{0,i}} = -\frac{n + 2\alpha_0 + 2}{2} \frac{1}{\omega_{0,i}} + \frac{1}{2} \frac{(\mathbf{B})_{ii}}{\omega_{0,i}^2}.$$

It follows that given $\mathbf{O} = [\mathbf{\Gamma} \quad \mathbf{\Gamma}_0]$, the expression l_1 is maximized at $\hat{\boldsymbol{\eta}} = \mathbf{\Gamma}^T(\mathbb{Y}_c^T \mathbb{X} + \mathbf{e}\mathbf{C})(\mathbb{X}^T \mathbb{X} + \mathbf{C})^{-1}$, $\hat{\omega}_{0,ii} = (\mathbf{B})_{ii}/(n + 2\alpha_0 + 2)$, for $i = 1, \dots, r - u$, and $\hat{\omega}_{ii} = (\mathbf{A})_{ii}/(n + 2\alpha + 2)$, for $i = 1, \dots, u$. Here we did not consider the fact that the ω_i 's and $\omega_{0,ii}$'s have to be in decreasing order. However, note that l_1 is invariant to permuting the entries of $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$ (given that the same permutation is applied to the columns of $\mathbf{\Gamma}$ and $\mathbf{\Gamma}_0$). Hence, we can simply order the $\hat{\omega}_{ii}$'s and $\hat{\omega}'_{0,ii}$ s, and then permute the columns of $\mathbf{\Gamma}$ and $\mathbf{\Gamma}_0$ accordingly. By

substituting the maximas for $\boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{\omega}_0$ in l_1 , and denoting the resulting partially maximized version by l_2 , we get

$$l_2 = \text{constant} - \frac{n + 2\alpha + 2}{2} \sum_{i=1}^u \log \frac{(\mathbf{A})_{ii}}{n + 2\alpha + 2} - \frac{n + 2\alpha_0 + 2}{2} \sum_{i=1}^{r-u} \log \frac{(\mathbf{B})_{ii}}{n + 2\alpha_0 + 2} - \frac{1}{2} \text{tr}(\mathbf{D}^{-1} \mathbf{O}^T \mathbf{G} \mathbf{O}).$$

Note that l_2 is a function of $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$. The optimization with respect to $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ can be performed using MATLAB toolbox *sg_min 2.4.3* by Lippert (<http://web.mit.edu/~ripper/www/sgmin.html>).

E. DISCRETIZATION APPROACH FOR TRUNCATED INVERSE-GAMMA

Suppose $X \sim \text{Inverse-Gamma}(\alpha, \lambda, L, U)$ where $U > L \geq 0$. One can use a rejection sampling based approach or use the R package “*distr*” to sample from this distribution. However, none of the above mentioned methods work well when the Inverse-Gamma($\alpha, \lambda, 0, \infty$) distribution assigns a very small probability to the interval (L, U) . To overcome this problem, we sample from a discretized version of the Inverse-Gamma(α, λ, L, U) distribution. But this discretization method is valid only when (L, U) is a finite interval, i.e. $U < \infty$. To deal with the case $U = \infty$, we sample from a discretized version of the Inverse-Gamma(a, λ, L, ∞) distribution restricted to the finite interval $[L, \delta L]$ where $\delta = e \left\{ (1 + e) \frac{1-\epsilon}{\epsilon} \right\}^{\frac{1}{\alpha}}$ for any appropriately small $\epsilon \in (0, \frac{1}{2})$. The following lemma justifies the truncation by δL . Note that, for the case $U = \infty$, this discretization approach is required only when L is greater than $\frac{\lambda}{1+\alpha}$, the mode of a Inverse-Gamma($\alpha, \lambda, 0, \infty$) distribution.

Lemma 1 *Let $X \sim \text{Inverse-Gamma}(\alpha, \lambda, L, \infty)$ where $L > \frac{\lambda}{1+\alpha}$, then for any $\epsilon \in (0, \frac{1}{2})$, we have $P(X > \delta L) \leq \epsilon$.*

Proof Since $0 < \epsilon \leq \frac{1}{2}$ we get that,

$$(E.10) \quad \frac{1-\epsilon}{\epsilon} \geq 1 \implies \left(\frac{1-\epsilon}{\epsilon} \right) e^\alpha \geq 1 \implies e \left\{ (1+e) \frac{1-\epsilon}{\epsilon} \right\}^{\frac{1}{\alpha}} \geq \left\{ 1 + \frac{1-\epsilon}{\epsilon} e^{1+\alpha} \right\}^{\frac{1}{\alpha}}.$$

It follows by (E.10) and the definition of δ that

$$\left(e^{\frac{1+\alpha}{\delta}} + \frac{\epsilon}{1-\epsilon} \right) \frac{1}{\delta^\alpha} \leq \frac{\epsilon}{1-\epsilon}.$$

Using the fact that $L > \frac{\lambda}{1+\alpha}$ along with some simple algebraic manipulations, we get

$$\frac{e^{\frac{\lambda}{L\delta}} \frac{1}{(L\delta)^\alpha}}{\frac{1}{L^\alpha} \left(1 - \frac{\lambda}{\delta^\alpha} \right)} = \frac{\int_{\delta L}^{\infty} x^{-\alpha-1} dx}{e^{-\frac{\lambda}{\delta L}} \int_L^{\delta L} x^{-\alpha-1} dx} \leq \frac{\epsilon}{1-\epsilon}.$$

Since $e^{-\frac{\lambda}{L\delta}} < 1$, it follows that

$$\frac{\int_{\delta L}^{\infty} x^{-\alpha-1} e^{-\frac{\lambda}{x}} dx}{\int_L^{\delta L} x^{-\alpha-1} e^{-\frac{\lambda}{x}} dx} \leq \frac{\epsilon}{1-\epsilon}.$$

Finally, we note that

$$P(X > \delta L) = \frac{\int_{\delta L}^{\infty} x^{-\alpha-1} e^{-\frac{\lambda}{x}} dx}{\int_L^{\infty} x^{-\alpha-1} e^{-\frac{\lambda}{x}} dx} \leq \epsilon.$$

F. JEFFREY'S PRIOR FOR ENVELOPE MODELS

Since the parameters in \mathbf{O} are not independent, specifically, $\mathbf{O}\mathbf{O}^T = \mathbf{I}_r$. To compute the Jeffery's prior, we introduce an alternative parameterization which can write the parameters in the model as independent parameters. According to [1], we can write \mathbf{O} as $\mathbf{O} = \mathbf{O}_{12}\mathbf{O}_{13}\cdots\mathbf{O}_{1r}\mathbf{O}_{23}\cdots\mathbf{O}_{2r}\cdots\mathbf{O}_{r-1,r}$, where

$$\mathbf{O}_{ij} = \begin{matrix} & i & & j & \\ i & \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & \cos\theta_{ij} & 0 & -\sin\theta_{ij} & 0 \\ 0 & 0 & \mathbf{I} & 0 & 0 \\ 0 & \sin\theta_{ij} & 0 & \cos\theta_{ij} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I} \end{pmatrix} & & & \\ j & & & & \end{matrix}.$$

Now the parameters for the envelope model (4) are $\boldsymbol{\psi} = (\boldsymbol{\mu}^T, \theta_{12}, \dots, \theta_{r-1,r}, \text{vec}^T(\boldsymbol{\eta}), \boldsymbol{\omega}^T, \boldsymbol{\omega}_0^T)^T$, and they are on a product space. Let $\mathbf{K} = (\mathbf{I}_u, 0)^T \in \mathbb{R}^{r \times u}$, then $\boldsymbol{\beta} = \mathbf{O}\mathbf{K}\boldsymbol{\eta}$. The parameters under the standard model are $\boldsymbol{\xi} = (\boldsymbol{\mu}^T, \text{vec}^T(\boldsymbol{\beta}), \text{vech}^T(\boldsymbol{\Sigma}))^T$, where ‘‘vech’’ is the operator that stacks the lower triangle of a symmetric matrix into a vector. The Fisher information matrix under the standard model is

$$\mathbf{J} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{\mathbf{X}} \otimes \boldsymbol{\Sigma}^{-1} & 0 \\ 0 & 0 & \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{E}_r \end{pmatrix}.$$

Let $\mathbf{v}_{ij} \in \mathbb{R}^{r^2}$ be a vector of 0's, but with $-\sin\theta_{ij}$, $\cos\theta_{ij}$, $-\cos\theta_{ij}$ and $\sin\theta_{ij}$ on the $(i-1)r + i$ th, $(i-1)r + j$ th, $(j-1)r + i$ th and $(j-1)r + j$ th elements. Let $\mathbf{d}_i \in \mathbb{R}^{r^2}$ be a vector of 0's but having 1 on the $(i-1)r + i$ th element. We define matrices $\mathbf{A}_{ij}, \mathbf{B}_{ij}$ by the equation $\mathbf{O} = \mathbf{A}_{ij}\mathbf{O}_{ij}\mathbf{B}_{ij}$. The gradient matrix $\mathbf{G} = d\xi/d\boldsymbol{\psi}$ is

$$\begin{pmatrix} \mathbf{I}_r & 0 & 0 & 0 \\ 0 & (\boldsymbol{\eta}^T \mathbf{K}^T \mathbf{B}_{11}^T \otimes \mathbf{A}_{11})\mathbf{v}_{11} \cdots (\boldsymbol{\eta}^T \mathbf{K}^T \mathbf{B}_{r-1,r}^T \otimes \mathbf{A}_{r-1,r})\mathbf{v}_{r-1,r} & \mathbf{I}_u \otimes \mathbf{O}\mathbf{K} & 0 \\ 0 & 2\mathbf{C}_r(\boldsymbol{\Sigma}\mathbf{O}\mathbf{B}_{11}^T \otimes \mathbf{A}_{11})\mathbf{v}_{11} \cdots 2\mathbf{C}_r(\boldsymbol{\Sigma}\mathbf{O}\mathbf{B}_{r-1,r}^T \otimes \mathbf{A}_{r-1,r})\mathbf{v}_{r-1,r} & 0 & \mathbf{G}_{44} \end{pmatrix},$$

where $\mathbf{G}_{44} = \mathbf{C}_r(\mathbf{O} \otimes \mathbf{O})\mathbf{d}_1 \cdots \mathbf{C}_r(\mathbf{O} \otimes \mathbf{O})\mathbf{d}_r$. Then the Fisher information under the envelope model is $\mathbf{G}^T \mathbf{J} \mathbf{G}$. While explicit form for the Fisher information is present, the determinant of the Fisher information, and hence the Jeffrey's prior, has an intractable form. The main focus in this paper is to derive a flexible class of priors that have clear interpretation and reflect prior information through hyper-parameters. While invariance to reparameterization is a useful property to pursue, we find it tangential to our goal at this point.

G. COMPARISON OF THE BAYESIAN ENVELOPE MODEL AND BAYESIAN WISHART STANDARD MODEL

Here we compare the performance of the Bayesian envelope model versus the Bayesian Wishart standard model (using Inverse-Wishart prior for Σ and a matrix-normal prior for $\beta \mid \Sigma$). We use the uniform Haar prior for the Bayesian envelope model. We use a uniform improper prior for both $\beta \mid \Sigma$ and Σ for the Bayesian prior model. We note that this prior is different than the uniform Haar prior for the envelope model with $u = r$. Table 1 displays the ratios of estimation variance and average squared errors using the same data that used for generating Table 2. We notice that the ratios in Table 1 are similar to those in Table 2 for all sample sizes, which suggests that the efficiency gains from the Bayesian envelope model are quite stable to different versions of the Bayesian standard models.

	n = 30		n=100		n=200	
	<i>Ratio_{Var}</i>	<i>Ratio_{MSE}</i>	<i>Ratio_{Var}</i>	<i>Ratio_{MSE}</i>	<i>Ratio_{Var}</i>	<i>Ratio_{MSE}</i>
$\beta_{1,1}$	2.695	2.656	3.552	3.528	3.342	3.354
$\beta_{1,2}$	1.977	1.978	2.053	2.057	2.418	2.418
$\beta_{1,3}$	2.756	2.749	3.108	3.108	3.785	3.779
$\beta_{1,4}$	3.515	3.505	3.332	3.330	3.950	3.940
$\beta_{1,5}$	2.815	2.793	3.528	3.517	3.785	3.787
$\beta_{2,1}$	2.167	2.148	2.190	2.207	1.995	1.982
$\beta_{2,2}$	1.850	1.855	1.426	1.437	1.666	1.684
$\beta_{2,3}$	2.209	2.205	2.374	2.364	2.528	2.545
$\beta_{2,4}$	2.307	2.299	2.331	2.329	2.404	2.405
$\beta_{2,5}$	2.039	2.037	2.352	2.360	1.850	1.842

TABLE 1

Ratio of estimation variance and average squared error in estimation of β . $Ratio_{Var}$ denotes the ratio of estimation variance of Bayesian Wishart standard model versus the Bayesian envelope model. $Ratio_{MSE}$ denotes the ratio of average squared error of Bayesian Wishart standard model versus the Bayesian envelope model.

H. NUMERICAL EXPERIMENTS ON THE PERFORMANCE OF THE BAYESIAN ENVELOPE MODEL WHEN $\omega_{TRUE} = (20, 16)^T$, $\omega_{0,TRUE} = (0.5, 1, 2)^T$ AND $\omega_{TRUE} = (1, 1)^T$, $\omega_{0,TRUE} = (1, 1, 1)^T$

The simulation settings for Table 2 is the same as those in Table 2, but with $\omega_{true} = (20, 16)^T$ and $\omega_{0,true} = (0.5, 1, 2)^T$. By [5], in this case, the non-Bayesian envelope model is still more efficient than the non-Bayesian standard model, but the efficiency gain is not as large as those displayed in Table 2. From the results in Table 2, we notice that the Bayesian envelope model also follows the same phenomenon in this situation.

The simulation settings for Table 3 is the same as those in Table 2, expect $\omega_{true} = (1, 1)^T$ and $\omega_{0,true} = (1, 1, 1)^T$. For the non-Bayesian envelope model, this is a special case where the non-Bayesian envelope model is as efficient as the non-Bayesian standard model. From the results in Table 3, we notice that the Bayesian envelope model also is about as efficient as the Bayesian standard model.

	n = 30		n=100		n=200	
	$Ratio_{Var}$	$Ratio_{MSE}$	$Ratio_{Var}$	$Ratio_{MSE}$	$Ratio_{Var}$	$Ratio_{MSE}$
$\beta_{1,1}$	1.003	1.004	0.995	1.003	1.012	1.012
$\beta_{1,2}$	0.994	0.999	1.092	1.089	1.022	1.022
$\beta_{1,3}$	1.040	1.042	1.033	1.008	1.030	1.034
$\beta_{1,4}$	0.999	0.996	1.068	1.069	0.998	1.007
$\beta_{1,5}$	1.013	1.011	0.988	0.989	0.992	0.994
$\beta_{2,1}$	1.049	1.051	1.004	1.005	1.002	1.005
$\beta_{2,2}$	1.042	1.035	1.063	1.065	1.098	1.066
$\beta_{2,3}$	1.035	1.078	1.075	1.073	1.167	1.162
$\beta_{2,4}$	0.973	1.102	1.135	1.138	1.033	1.036
$\beta_{2,5}$	1.001	0.997	0.975	0.976	1.026	1.026

TABLE 2

Ratio of estimation variance and average squared error in estimation of β . $Ratio_{Var}$ denotes the ratio of estimation variance of the Bayesian standard model versus the Bayesian envelope model. $Ratio_{MSE}$ denotes the ratio of average squared error of the Bayesian standard model versus the Bayesian envelope model.

$$(\omega_{true} = (20, 16)^T, \omega_{0,true} = (0.5, 1, 2)^T)$$

	n = 30		n=100		n=200	
	$Ratio_{Var}$	$Ratio_{MSE}$	$Ratio_{Var}$	$Ratio_{MSE}$	$Ratio_{Var}$	$Ratio_{MSE}$
$\beta_{1,1}$	1.031	1.018	1.016	1.022	1.013	1.009
$\beta_{1,2}$	0.994	0.979	1.008	1.009	1.005	1.013
$\beta_{1,3}$	0.969	0.938	0.992	0.987	1.016	1.025
$\beta_{1,4}$	1.009	1.005	0.997	0.990	1.002	1.000
$\beta_{1,5}$	1.024	1.013	1.019	1.024	0.999	1.000
$\beta_{2,1}$	1.020	1.021	1.006	1.003	1.007	1.009
$\beta_{2,2}$	0.994	0.973	1.024	1.008	0.985	0.978
$\beta_{2,3}$	0.937	0.911	0.993	0.995	0.996	0.998
$\beta_{2,4}$	1.050	1.036	1.003	1.004	1.004	1.003
$\beta_{2,5}$	1.008	0.982	1.003	1.002	1.011	1.008

TABLE 3

Ratio of estimation variance and average squared error in estimation of β . $Ratio_{Var}$ denotes the ratio of estimation variance of the Bayesian standard model versus the Bayesian envelope model. $Ratio_{MSE}$ denotes the ratio of average squared error of the Bayesian standard model versus the Bayesian envelope model.

$$(\omega_{true} = (1, 1)^T, \omega_{0,true} = (1, 1, 1)^T)$$

I. COMPARISON OF THE BAYESIAN ENVELOPE MODEL AND BAYESIAN WISHART STANDARD MODEL WITH SMALL SAMPLE SIZE

The simulation settings of Table 4 are the same as those that produced Table 3. But the Bayesian Wishart standard model has a different prior. We use a uniform improper prior for $\beta \mid \Sigma$ and an inverse Wishart prior for Σ . We notice that ratios of estimation variance and average squared errors are about the same in the two tables, which suggests that the efficiency gains achieved by the Bayesian envelope model are still stable with different versions of the Bayesian Wishart standard model when the sample size is small.

	$\beta_{6,2}$	$\beta_{9,10}$	$\beta_{19,8}$	$\beta_{15,6}$	$\beta_{10,9}$	$\beta_{48,9}$	$\beta_{3,5}$	$\beta_{15,2}$	$\beta_{17,9}$	$\beta_{34,2}$
<i>Ratio_{MSE}</i>	4.391	5.423	5.103	7.917	8.873	7.650	5.412	2.711	8.543	3.918
<i>Ratio_{Var}</i>	4.391	5.383	5.101	9.914	8.840	7.968	5.357	3.129	8.627	3.935

TABLE 4

Ratio of estimation variance and average squared error for ten randomly selected elements in β , with $n = 30$ and $r = 50$. $Ratio_{Var}$ denotes the ratio of estimation variance of the Bayesian Wishart standard model versus the Bayesian envelope model. $Ratio_{MSE}$ denotes the ratio of average squared error of the Bayesian Wishart standard model versus the Bayesian envelope model.

REFERENCES

- [1] Anderson, T.W., Olkin, I. and Underhill, L. G. (1987). Generation of random orthogonal matrices. *SIAM Journal on Scientific and Statistical Computing*, **8**, 625–629.
- [2] Asmussen, S. and Glynn, P. W. (2011). A new proof of convergence of MCMC via the ergodic theorem, *Statistics & Probability Letters* **81**, 1482–1485.
- [3] Bingham, C. (1974). An antipodally symmetric distribution on the sphere, *Annals of Statistics* **2**, 1201–1225.
- [4] Cook, R.D., Helland, I. S. and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B*, **75**, 851–877.
- [5] Cook, R.D., Li, B. and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression, *Statistica Sinica*, **20**, 927–960.
- [6] Cook, R. D. and Su, Z. (2013). Scaled envelopes: Scale invariant and efficient estimation method in multivariate linear regression. *Biometrika*, **100**, 921–938.
- [7] Conway, J. (1990). *A Course in Functional Analysis*. New York: Springer.
- [8] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2004). *Bayesian Data Analysis*, Chapman and Hall/CRC Press.
- [9] Hoff, P.D. (2009). Simulation of the matrix Bingham von Mises Fisher Distribution, with applications to multivariate and relational Data, *Journal of Computational and Graphical Statistics* **18**, 438–456.
- [10] Mao, K. and Liang, F. and Mukherjee, S. (2010). Supervised dimension reduction using Bayesian mixture modeling. *International Conference on Artificial Intelligence and Statistics*, 501–508.
- [11] Reich, B. J. and Bondell, H. D. and Li, L. (2011). Sufficient dimension reduction via Bayesian mixture modeling. *Biometrics* **67**, 886–895.
- [12] Rossi, P.E., Allenby, G.M. and McCulloch, R. (2012). *Bayesian Statistics and Marketing*, John Wiley & Sons.
- [13] Sameh, A.H. and Wisniewski, J.A. (1982). A trace minimization algorithm for the generalized eigenvalue problem, *SIAM Journal on Numerical Analysis* **19**, 1243–1259.
- [14] Su, Z. and Cook, R.D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98**, 133–146.

- [15] Su, Z. and Cook, R.D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika* **99**, 687–702.
- [16] Su, Z. and Cook, R.D. (2013). Estimation of multivariate means with heteroscedastic error using envelope models. *Statistica Sinica* **23**, 213–230.
- [17] Schott, J.R. (2013). On the likelihood ratio test for envelope models in multivariate linear regression. *Biometrika*, **100**, 531–537.
- [18] Tokdar, S. T., Zhu, Y. M. and Ghosh, J. K. (2010). Bayesian density regression with Logistic Gaussian process and subspace projection. *Bayesian Analysis*, **5**, 319–344.

DEPARTMENT OF STATISTICS
102 GRIFFIN-FLOYD HALL
UNIVERSITY OF FLORIDA
GAINESVILLE, FL 32605