Supplementary materials for: Envelopes for elliptical multivariate linear regression

Liliana Forzani[*] and Zhihua Su[†‡]

A Technical details in Section 4.2: Derivatives of m_i with respect to β and Σ

Using equation (119) from [13], we get

$$\frac{\partial m_i}{\partial \beta} = -2\Sigma^{-1}(Y_i - \mu_Y - \beta(X_i - \mu_X))(X_i - \mu_X)^T.$$

On the other hand, for $F(X) \in \mathbb{R}^{m\times p}$ and $G(X) \in \mathbb{R}^{p\times q}$, (see [6]), we have

$$\frac{\partial \mathrm{vec}(F(X)G(X))}{\partial \mathrm{vec}^T(X)} = (G(X)^T \otimes I_m)\frac{\partial \mathrm{vec}(F(X))}{\partial \mathrm{vec}^T(X)} + (I_q \otimes F(X))\frac{\partial \mathrm{vec}(G(X))}{\partial \mathrm{vec}^T(X)}.$$

Now take $X = \Sigma$, $F(\Sigma) = \Sigma$ and $G(\Sigma) = \Sigma^{-1}$, then

$$0 = (\Sigma^{-1} \otimes I_r)\frac{\partial \mathrm{vec}(\Sigma)}{\partial \mathrm{vec}^T(\Sigma)} + (I_r \otimes \Sigma)\frac{\partial \mathrm{vec}(\Sigma^{-1})}{\partial \mathrm{vec}^T(\Sigma)},$$

[*] Liliana Forzani is Professor, Facultad de Ingeniería Química, Universidad Nacional del Litoral, CONICET-UNL, Santa Fe, Argentina. Email: liliana.forzani@gmail.com.

[†] Zhihua Su is Associate Professor, Department of Statistics, University of Florida, Gainsville, FL 32611-8545. E-mail: zhihuasu@stat.ufl.edu.

[‡] Author names are listed alphabetically.

which yields

$$\frac{\partial \text{vec}(\Sigma^{-1})}{\partial \text{vec}^T(\Sigma)} = -(\Sigma^{-1} \otimes \Sigma^{-1}).$$

Therefore

$$\frac{\partial m_i}{\partial \text{vec}(\Sigma)} = -\left[(\Sigma^{-1} \otimes \Sigma^{-1})\text{vec}\big((Y_i - \mu_Y - \beta(X_i - \mu_X))(Y_i - \mu_Y - \beta(X_i - \mu_X)^T)\big)\right]^T,$$

or equivalently

$$\frac{\partial m_i}{\partial \Sigma} = -\Sigma^{-1}[Y_i - \mu_Y - \beta(X_i - \mu_X)][Y_i - \mu_Y - \beta(X_i - \mu_X)^T]\Sigma^{-1}.$$

B Proof of Proposition 1

We calculate the asymptotic distributions for the MLE estimator of model (1), the reduced-rank regression estimator, the envelope estimator and the reduced-rank envelope estimator.

The MLE estimator of model (1)**:** $\text{avar}\{\sqrt{n}\text{vec}(\widehat{\beta}_{\text{std}})\}$

From [1], [2], [8], [9], [10], [11], or [15] to compute the asymptotic variance of the estimator we need to compute the information matrix.

Without loss of generality we assume that $\mu_Y = 0$ and $\mu_X = 0$. The log-likelihood function for the elliptical linear multivariate regression (1) is given by

$$L \;=\; -\frac{1}{2}\log|\Sigma| + \log g\left[(Y - \beta X)^T \Sigma^{-1} (Y - \beta X)\right]. \tag{1}$$

The Fisher information matrix for $(\text{vec}^T(\beta),\ \text{vech}^T(\Sigma))$ is given by

$$J_h = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

with

$$\begin{aligned} J_{11} &= E\left(\frac{\partial L}{\partial \text{vec}^T(\beta)} \frac{\partial L}{\partial \text{vec}(\beta)}\right), \\ J_{12} &= E\left(\frac{\partial L}{\partial \text{vec}^T(\beta)} \frac{\partial L}{\partial\, \text{vech}(\Sigma)}\right), \quad J_{21} = J_{12}^T, \\ J_{22} &= E\left(\frac{\partial L}{\partial\, \text{vech}^T(\Sigma)} \frac{\partial L}{\partial\, \text{vech}(\Sigma)}\right). \end{aligned}$$

Let $U = \Sigma^{-1/2}(Y - \beta X)$. We will prove that

$$A \equiv E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 UU^T \middle| X\right] = N_X I_r, \tag{2}$$

$$B \equiv E\left(U \frac{g'(U^TU)}{g(U^TU)} \middle| X\right) = 0, \tag{3}$$

$$C \equiv E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 \text{vec}(UU^T)U^T \middle| X\right] = 0, \tag{4}$$

$$D \equiv E\left[\frac{g'(U^TU)}{g(U^TU)}(UU^T) \middle| X\right] = -\frac{1}{2} I_r, \tag{5}$$

$$E \equiv E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 \text{vec}(UU^T)\text{vec}^T(UU^T) \middle| X\right] = M_X(I_{r^2} + K_{rr}) + M_X \text{vec}(I_r)\text{vec}^T(I_r), \tag{6}$$

where $N_X = E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 U^TU \middle| X\right]/r$, $M_X = E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 (U^TU)^2 \middle| X\right]/(r(r+2))$ and $K_{rr} \in \mathbb{R}^{r^2 \times r^2}$ denotes a commutation matrix that for an arbitrary matrix $A \in \mathbb{R}^{r\times r}$, $\text{vec}(A^T) = K_{rr}\text{vec}(A)$.

Using (2)-(6), we have

$$
\begin{aligned}
J_{11} &= 4(\widetilde{\Sigma}_X \otimes \Sigma^{-1}), &(7)\\
J_{12} &= 0, &(8)\\
J_{22} &= 2ME_r^T(\Sigma^{-1}\otimes\Sigma^{-1})E_r + (M-\frac{1}{4})E_r^T\mathrm{vec}(\Sigma^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r, &(9)
\end{aligned}
$$

where $\widetilde{\Sigma}_X = E(N_X XX^T)$ if X is random and $\widetilde{\Sigma}_X = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n N_{X_i}X_iX_i^T$ if X is fixed (assuming the limit is finite); $M = E(M_X)$ if X is random and $M = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n M_{X_i}$ if X is fixed (assuming the limit is finite).

We first prove (7)-(9) given (2)-(6), and then establish (2)-(6).

Based on (1), the derivatives of L with respect to β and Σ^{-1} are given by

$$
\begin{aligned}
\frac{\partial L}{\partial \mathrm{vec}^T(\beta)} &= -2\frac{g'(U^TU)}{g(U^TU)}(X\otimes\Sigma^{-1/2})U,\\
\frac{\partial L}{\partial\,\mathrm{vech}^T(\Sigma^{-1})} &= \frac{1}{2}E_r^T\mathrm{vec}(\Sigma) + E_r^T\frac{g'(U^TU)}{g(U^TU)}(\Sigma^{1/2}\otimes\Sigma^{1/2})\mathrm{vec}(UU^T).
\end{aligned}
$$

Since

$$
\frac{\partial \mathrm{vec}(\Sigma^{-1})}{\mathrm{vec}^T(\Sigma)} = -(\Sigma^{-1}\otimes\Sigma^{-1}),
$$

we get

$$
\frac{\partial L}{\partial\,\mathrm{vech}^T(\Sigma)} = -\frac{1}{2}E_r^T\mathrm{vec}(\Sigma^{-1}) - E_r^T\frac{g'(U^TU)}{g(U^TU)}(\Sigma^{-1/2}\otimes\Sigma^{-1/2})\mathrm{vec}(UU^T).
$$

Then

$$
\begin{aligned}
J_{11} &= 4E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 (X\otimes\Sigma^{-1/2})UU^T(X^T\otimes\Sigma^{-1/2})\right]\\
&= 4E\left((X\otimes\Sigma^{-1/2})A(X^T\otimes\Sigma^{-1/2})\right)\\
&= 4E\left(N_X(XX^T\otimes\Sigma^{-1})\right) = 4(\widetilde{\Sigma}_X\otimes\Sigma^{-1}),\\
J_{12} &= 2E\left(\frac{g'(U^TU)}{g(U^TU)}(X\otimes\Sigma^{-1/2})U\left[\mathrm{vec}^T(UU^T)(\Sigma^{-1/2}\otimes\Sigma^{-1/2})\frac{g'(U^TU)}{g(U^TU)}E_r + \frac{1}{2}\mathrm{vec}^T(\Sigma^{-1})E_r\right]\right)\\
&= 2E\left((X\otimes\Sigma^{-1/2})C^T(\Sigma^{-1/2}\otimes\Sigma^{-1/2})E_r + (X\otimes\Sigma^{-1/2})B\frac{1}{2}\mathrm{vec}^T(\Sigma^{-1})E_r\right)\\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
J_{22} &= E\left(\left[\frac{1}{2}E_r^T\mathrm{vec}(\Sigma^{-1}) + E_r^T\frac{g'(U^TU)}{g(U^TU)}(\Sigma^{-1/2}\otimes\Sigma^{-1/2})\mathrm{vec}(UU^T)\right]\right.\\
&\qquad \left.\left[\mathrm{vec}^T(UU^T)(\Sigma^{-1/2}\otimes\Sigma^{-1/2})\frac{g'(U^TU)}{g(U^TU)}E_r + \frac{1}{2}\mathrm{vec}^T(\Sigma^{-1})E_r\right]\right)\\
&= \frac{1}{4}E_r^T\mathrm{vec}(\Sigma^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r + \frac{1}{2}E_r^T\mathrm{vec}(\Sigma^{-1})\mathrm{vec}^T(D)(\Sigma^{-1/2}\otimes\Sigma^{-1/2})E_r\\
&\qquad + \frac{1}{2}E_r^T(\Sigma^{-1/2}\otimes\Sigma^{-1/2})\mathrm{vec}(D)\mathrm{vec}^T(\Sigma^{-1})E_r + E_r^T(\Sigma^{-1/2}\otimes\Sigma^{-1/2})E(\Sigma^{-1/2}\otimes\Sigma^{-1/2})E_r\\
&= -\frac{1}{4}E_r^T\mathrm{vec}(\Sigma^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r\\
&\qquad + E_r^T(\Sigma^{-1/2}\otimes\Sigma^{-1/2})E\left[M_X(I_{r^2}+K_{rr}) + M_X\mathrm{vec}(I_r)\mathrm{vec}^T(I_r)\right](\Sigma^{-1/2}\otimes\Sigma^{-1/2})E_r\\
&= 2ME_r^T(\Sigma^{-1}\otimes\Sigma^{-1})E_r + (M-\frac{1}{4})E_r^T\mathrm{vec}(\Sigma^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r,
\end{aligned}
$$

from which we establish (7) - (9). Now we prove (2)-(6).

Proof of (2): Since the distribution of U given X is symmetric, for any orthogonal matrix V, U and

VU should have the same distribution. As a consequence,

$$
\begin{aligned}
A &= E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 UU^T \middle| X\right] \\
&= E\left[\left(\frac{g'(U^TV^TVU)}{g(U^TV^TVU)}\right)^2 VUU^TV^T \middle| X\right] \\
&= VE\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 \mathrm{vec}(U)\mathrm{vec}^T(U) \middle| X\right] V^T \\
&= VAV^T.
\end{aligned}
$$

Using Proposition 2.14 of [7], we have

$$A = N_X I_r.$$

To find N_X, notice that $\mathrm{tr}(A) = rN_X$. So we have

$$N_X = \frac{1}{r}\,\mathrm{tr}(A) = \frac{1}{r}E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 U^TU \middle| X\right].$$

Proof of (3): We prove (3) using the same technique as that in the proof of (2). For any $r \times r$ orthogonal matrix V,

$$B = VE\left(U\frac{g'(U^TU)}{g(U^TU)} \middle| X\right) = VB.$$

Let e_i denote the vector of all zeros except that its i element is one. Take V to be the permutation matrix such that $Ve_i = e_j$, $Ve_j = e_i$ and $Ve_k = e_k$ for $k = 1, \ldots, k \neq i, j$. Then B must be proportional to a vector of 1's. In other words, $B = c1_r$, where 1_r is the vector of r one's. Then for any orthogonal matrix V, we have $c = c\sum_{j=1}^{r} v_{ij}$ for $i = 1, \ldots, r$. Therefore $c = 0$ and (3) follows.

Proof of (4): Using the same reasoning, for any orthogonal matrix V,

$$C = (V \otimes V)CV^T. \tag{10}$$

We first take V to have the following form

$$V = \begin{pmatrix} 1 & 0 \\ 0 & V_{-i} \end{pmatrix},$$

where V_{-i} is any $(r-1) \times (r-1)$ orthogonal matrix. Then we have $Ve_1 = e_1$. Let c_1 denote the first column of C. Then we have

$$c_1 \quad = \quad (V \otimes V)c_1.$$

Let M_1 be an $r \times r$ matrix such that $\text{vec}(M_1) = c_1$. Then we have

$$\text{vec}(M_1) = (V \otimes V)\text{vec}(M_1) = \text{vec}(VM_1V^T).$$

If we partition the matrix M_1 the same way as we partition V. Then by Proposition 2.14 of [7], M_1 must have the following structure

$$M_1 = \begin{pmatrix} d_1 & 0 \\ 0 & f_1 I_{r-1} \end{pmatrix},$$

where d_1 and f_1 are constants. Similarly, we take V such that $V^T e_i = e_i$. Then we can construct M_i such that $\text{vec}(M_i) = c_i$, where c_i is the ith column of C. By the previous discussion M_i must be a diagonal matrix where the ith diagonal element is d_i and the rest are f_i.

Now we want to prove $d_i = f_i$ for all i. Let V be a permutation matrix such that for $i \neq j$,

$V^T e_i = e_j$, $V^T e_j = e_i$ and $Ve_k = e_k$ for $k \neq i, j$. Then we have

$$c_i = (V \otimes V)c_j, \qquad \text{vec}(M_i) = \text{vec}(VM_jV^T).$$

So $M_i = VM_jV^T$. Because of the structure of V, we then have $d_i = d_j = f_i = f_j \equiv c$ for all $1 \leq i, j \leq r$, where c is a constant. Therefore $M_i = cI_r$ for all $i = 1, \ldots, r$. This implies that $C = c\,\text{vec}(I_r)1_r^T$. Using (10), we have

$$c\,\text{vec}(I_r)1_r^T = c(V \otimes V)\left(\text{vec}(I_r)1_r^T\right)V^T = c\,\text{vec}(VV^T)1_r^TV^T = c\,\text{vec}(I_r)1_r^TV^T.$$

Pre-multiplying both sides by $\text{vec}^T(I_r)$ we get

$$cr1_r^T = cr1_r^TV^T$$

for any V orthogonal. This implies $c = 0$ and therefore $C = 0$.

Proof of (5): Again for any orthogonal matrix V, we have $D = VDV^T$. By Proposition 2.14 of [7]

$$D = H_X I_r$$

with $H_X = E\left[\frac{g'(U^TU)}{g(U^TU)}(U^TU)\Big| X\right]/r$. Using equation (2.12) of [12], we get $H_X = -1/2$.

Proof of (6): By the definition of E, for any orthogonal matrix V,

$$E = (V \otimes V)E(V^T \otimes V^T).$$

Using Proposition 4.1 of [4], we get $E = cI_{r^2} + aK_{rr} + 2d\text{vec}(I_r)\text{vec}^T(I_r)$, where a, c and d are constants. Since E is symmetric,

$$E = K_{rr}E = cK_{rr} + aI_{r^2} + 2d\mathrm{vec}(I_r)\mathrm{vec}^T(I_r).$$

Therefore $c(K_{rr} - I_{r^2}) = a(K_{rr} - I_{r^2})$, which implies that $a = c$ and

$$E = c(I_{r^2} + K_{rr}) + 2d\mathrm{vec}(I_r)\mathrm{vec}^T(I_r). \tag{11}$$

Now we compute c and d. Taking trace of (11) on both sides and using the fact that $\mathrm{tr}(A^TB) = \mathrm{vec}^T(A)\mathrm{vec}(B)$ for any matrices $A \in \mathbb{R}^{a\times b}$ and $B \in \mathbb{R}^{a\times b}$, we have

$$E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 (U^TU)^2 \middle| X\right] \quad = \quad cr(r+1) + 2dr.$$

Now pre-multiply (11) by $\mathrm{vec}^T(I)$ and post-multiply (11) by $\mathrm{vec}(I)$, and take the trace. We have

$$E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 (U^TU)^2 \middle| X\right] = \quad 2cr + 2dr^2.$$

As a consequence, $2d = c$. Then

$$r(r+2)c \quad = \quad E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 (U^TU)^2 \middle| X\right].$$

Let M_x denote c, and we have (6).

Since the reduced-rank regression, the envelope and the reduced-rank envelope models are over-parameterized, we will apply Proposition 4.1 from [14] to prove the asymptotic distribution for $\widehat{\beta}_{\mathrm{RR}}$, $\widehat{\beta}_E$ and $\widehat{\beta}_{\mathrm{RE}}$ as in [5] and [6]. To apply Proposition 4.1 of [14], we will check the assumptions first. Along the discussion, we will match Shapiro's notations in our context. Let us call $F((\beta_{std}, \Sigma_{std}), (\beta, \Sigma)) = L(\widehat{\beta}_{std}, \widehat{\Sigma}_{std}) - L(\beta, \Sigma)$ where L is the likelihood function. Then F satisfies the four conditions for F in Section 3 in [14]. The function g defined by Shapiro in (2.1) are the functions g_1, g_2 and g_3 defined in (12) under our context. Let $h = (\mathrm{vec}^T(\beta), \ \mathrm{vech}^T(\Sigma))^T$, $\psi = \big(\big(\mathrm{vec}^T(A), \mathrm{vec}^T(B), \ \mathrm{vech}^T(\Sigma)\big)^T$, $\delta = (\mathrm{vec}^T(\xi), \mathrm{vec}^T(\Gamma), \ \mathrm{vech}^T(\Omega), \ \mathrm{vech}^T(\Omega_0))^T$ and $\phi = (\mathrm{vec}^T(B), \mathrm{vec}^T(\eta), \mathrm{vec}^T(\Gamma), \ \mathrm{vech}^T(\Omega), \ \mathrm{vech}^T(\Omega_0))^T$ denote the parameters in the standard model, the reduced-rank regression, the envelope model and the reduced-rank envelope model respectively. We have

$$h = g_1(\psi) = \begin{pmatrix} \mathrm{vec}(AB) \\ \mathrm{vech}(\Sigma) \end{pmatrix}, \qquad h = g_2(\delta) = \begin{pmatrix} \mathrm{vec}(\Gamma\xi) \\ \mathrm{vech}(\Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T) \end{pmatrix},$$
$$h = g_3(\phi) = \begin{pmatrix} \mathrm{vec}(\Gamma\eta B) \\ \mathrm{vech}(\Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T) \end{pmatrix}. \tag{12}$$

It is obvious that g_1, g_2 and g_3 are all twice continuous differentiable. Therefore all the assumptions of Shapiro's Proposition 4.1 are satisfied, and we can get the asymptotic distribution of each of the estimators using Proposition 4.1 from [14]. Furthermore, the asymptotic variance of the estimator of reduced-rank regression, the envelope model or the reduced-rank envelope model is given by $H(H^T J_h H)^{\dagger} H^T$, where J_h is the Fisher information under the standard model, and H is the gradient matrix, which equals to $\partial h/\partial^T \psi$, $\partial h/\partial^T \delta$ and $\partial h/\partial^T \phi$ under the reduced-rank regression, the envelope model and the reduced-rank envelope model respectively.

Next we calculate the asymptotic variance of the reduced-rank regression estimator, the envelope estimator and the reduced-rank envelope estimator in details as follows.

The reduced-rank regression estimator: $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})]$

To estimate $h = (\mathrm{vec}^T(\beta),\ \mathrm{vech}^T(\Sigma))^T$, the constituent parameters of the reduced-rank regression are $\psi = \big(\mathrm{vec}^T(A), \mathrm{vec}^T(B),\ \ \mathrm{vech}^T(\Sigma)\big)^T$. Since $\beta = AB$, the gradient matrix $H = \partial h / \partial^T \psi$ is

$$H = \begin{pmatrix} B^T \otimes I_r & I_p \otimes A & 0 \\ 0 & 0 & I_{r(r+1)/2} \end{pmatrix} = \begin{pmatrix} h_1 & 0 \\ 0 & I_{r(r+1)/2} \end{pmatrix}, \tag{13}$$

with $h_1 = (B^T \otimes I_r,\ \ I_p \otimes A)$. Using Proposition 4.1 in [14], $\mathrm{avar}[\sqrt{n}h(\widehat{\psi})]$ is given by

$$H(H^T J_h H)^\dagger H^T \quad = \quad \begin{pmatrix} \frac{1}{4} h_1 [h_1^T (\widetilde{\Sigma}_X \otimes \Sigma^{-1}) h_1]^\dagger h_1^T & 0 \\ 0 & J_\Sigma^{-1} \end{pmatrix},$$

where $\dagger$ denotes the Moore-Penrose generalized inverse. We can write $h_1 = H_1 H_2$, where

$$H_1 = \begin{pmatrix} B^T \otimes I & (I - B^T (B \widetilde{\Sigma}_X B^T)^{-1} B \widetilde{\Sigma}_X) \otimes A \end{pmatrix}, \quad H_2 = \begin{pmatrix} I_{rd} & (B\widetilde{\Sigma}_X B^T)^{-1} B \widetilde{\Sigma}_X \otimes A \\ 0 & I_{pd} \end{pmatrix}.$$

Since H_2 is of full rank, $h_1 [h_1^T (\widetilde{\Sigma}_X \otimes \Sigma^{-1}) h_1]^\dagger h_1^T = H_1 [H_1^T (\widetilde{\Sigma}_X \otimes \Sigma^{-1}) H_1]^\dagger H_1^T$. Now

$$H_1^T (\widetilde{\Sigma}_X \otimes \Sigma^{-1}) H_1 \quad = \quad \begin{pmatrix} B\widetilde{\Sigma}_X B^T \otimes \Sigma^{-1} & 0 \\ 0 & (\widetilde{\Sigma}_X - \widetilde{\Sigma}_X M_B \widetilde{\Sigma}_X) \otimes A^T \Sigma^{-1} A \end{pmatrix} \tag{14}$$

where $M_B = B^T (B\widetilde{\Sigma}_X B^T)^{-1} B$. Notice that both $B\widetilde{\Sigma}_X B^T \otimes \Sigma^{-1}$ and $A^T \Sigma^{-1} A$ are invertible. Since $M_B \widetilde{\Sigma}_X M_B = M_B$, $(\widetilde{\Sigma}_X - \widetilde{\Sigma}_X M_B \widetilde{\Sigma}_X)^\dagger = \widetilde{\Sigma}_X^{-1} - M_B$. Therefore,

$$\Big[H_1^T (\widetilde{\Sigma}_X \otimes \Sigma^{-1}) H_1\Big]^\dagger = \begin{pmatrix} (B\widetilde{\Sigma}_X B^T)^{-1} \otimes \Sigma & 0 \\ 0 & (\widetilde{\Sigma}_X^{-1} - M_B) \otimes (A^T \Sigma^{-1} A)^{-1} \end{pmatrix}. \tag{15}$$

Let $M_A = A(A^T\Sigma^{-1}A)^{-1}A^T$. Since $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})]$ correspond to the upper left block of $H_1[H_1^T(\widetilde{\Sigma}_X \otimes \Sigma^{-1})H_1]^{\dagger}H_1^T$, we have

$$
\begin{aligned}
\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})] &= \frac{1}{4}M_B \otimes \Sigma + \frac{1}{4}(I_p - M_B\widetilde{\Sigma}_X)(\widetilde{\Sigma}_X^{-1} - M_B)(I_p - M_B\widetilde{\Sigma}_X) \otimes M_A \\
&= \frac{1}{4}M_B \otimes \Sigma + \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes M_A \\
&= \frac{1}{4}[\widetilde{\Sigma}_X^{-1} - (\widetilde{\Sigma}_X^{-1} - M_B)] \otimes \Sigma + \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes M_A \\
&= \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes (\Sigma - M_A). \qquad (16)
\end{aligned}
$$

The envelope estimator: $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_E)]$

Under the envelope model, the constituent parameters are $\delta = (\mathrm{vec}^T(\xi), \mathrm{vec}^T(\Gamma),\ \mathrm{vech}^T(\Omega),\ \mathrm{vech}^T(\Omega_0))^T$. Since $\beta = \Gamma\xi$, $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$, the gradient matrix $H = \partial h/\partial^T\delta$ is

$$
\begin{pmatrix}
I_p \otimes \Gamma & \xi^T \otimes I_r & 0 & 0 \\
0 & 2C_r(\Gamma\Omega \otimes I_r - \Gamma \otimes \Gamma_0\Omega_0\Gamma_0^T) & C_r(\Gamma \otimes \Gamma)E_u & C_r(\Gamma_0 \otimes \Gamma_0)E_{r-u}
\end{pmatrix}.
$$

By Proposition 4.1 in [14], $\mathrm{avar}[\sqrt{n}h(\widehat{\delta})] = H(H^TJ_hH)^{\dagger}H^T$. Again we write $H = H_1H_2$, where

$$
H_1 = \begin{pmatrix}
I_p \otimes \Gamma & \xi^T \otimes \Gamma_0 & 0 & 0 \\
0 & 2C_r(\Gamma\Omega \otimes \Gamma_0 - \Gamma \otimes \Gamma_0\Omega_0) & C_r(\Gamma \otimes \Gamma)E_u & C_r(\Gamma_0 \otimes \Gamma_0)E_{r-u}
\end{pmatrix},
$$

and

$$
H_2 = \begin{pmatrix}
I_{pu} & \xi^T \otimes \Gamma^T & 0 & 0 \\
0 & I_u \otimes \Gamma_0^T & 0 & 0 \\
0 & 2C_u(\Omega \otimes \Gamma^T) & I_{(u(u+1)/2} & 0 \\
0 & 0 & 0 & I_{(r-u)(r-u+1)/2}
\end{pmatrix}.
$$

Since H_2 is full rank, then the asymptotic variance is $\mathrm{avar}[\sqrt{n}h(\widehat{\delta})] = H_1(H_1^T J_h H_1)^\dagger H_1^T$. Now

$$H_1^T J_h \;=\; \begin{pmatrix} 4(\widetilde{\Sigma}_X \otimes \Omega^{-1}\Gamma^T) & 0 \\ 4(\xi\widetilde{\Sigma}_X \otimes \Omega_0^{-1}\Gamma_0^T) & 2M(\Gamma^T \otimes \Omega_0^{-1}\Gamma_0^T - \Omega^{-1}\Gamma^T \otimes \Gamma_0^T)E_r \\ 0 & 2ME_u^T(\Omega^{-1}\Gamma^T \otimes \Omega^{-1}\Gamma^T)E_r + (M-\frac{1}{4})E_u^T\mathrm{vec}(\Omega^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r \\ 0 & 2ME_{r-u}^T(\Omega_0^{-1}\Gamma_0^T \otimes \Omega_0^{-1}\Gamma_0^T)E_r + (M-\frac{1}{4})E_{r-u}^T\mathrm{vec}(\Omega_0^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r \end{pmatrix},$$

and

$$H_1^T J_h H_1 = \begin{pmatrix} 4(\widetilde{\Sigma}_X \otimes \Omega^{-1}) & 0 & 0 & 0 \\ 0 & S_{22} & 0 & 0 \\ 0 & 0 & S_{33} & S_{34} \\ 0 & 0 & S_{43} & S_{44} \end{pmatrix},$$

where

$$\begin{aligned} S_{22} &= 4(\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1}) + 4M(\Gamma^T \otimes \Omega_0^{-1}\Gamma_0^T - \Omega^{-1}\Gamma^T \otimes \Gamma_0^T)E_rC_r(\Gamma\Omega \otimes \Gamma_0 - \Gamma \otimes \Gamma_0\Omega_0) \\ &= 4(\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1}) + 4M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u}). \end{aligned}$$

The asymptotic variance of $\widehat{\beta}_E$ does not depend on S_{33}, S_{34}, S_{43} and S_{44}, since it is equal to the upper left block of $H_1(H_1^T J_h H_1)^\dagger H_1^T$:

$$\begin{aligned} \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_E)] &= \frac{1}{4}(I_p \otimes \Gamma)(\widetilde{\Sigma}_X^{-1} \otimes \Omega)(I_p \otimes \Gamma^T) + (\xi^T \otimes \Gamma_0)S_{22}^{-1}(\xi \otimes \Gamma_0^T) \\ &= \frac{1}{4}(\widetilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T) \\ &\quad + \frac{1}{4}(\xi^T \otimes \Gamma_0)[\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\xi \otimes \Gamma_0^T). \end{aligned}$$

The reduced-rank envelope estimator: $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})]$

The constituent parameters of the reduced-rank envelope model are $\phi = (\mathrm{vec}^T(B), \mathrm{vec}^T(\eta), \mathrm{vec}^T(\Gamma),\ \mathrm{vech}^T(\Omega), \mathrm{vech}^T(\Omega_0))^T$. Since $\beta = \Gamma\eta B$, $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$, the gradient matrix $H = \partial h/\partial^T\phi$ is

$$H = \begin{pmatrix} B^T\eta^T \otimes I_r & B^T \otimes \Gamma & I_p \otimes \Gamma\eta & 0 & 0 \\ 2C_r(\Gamma\Omega \otimes I_r - \Gamma \otimes \Gamma_0\Omega_0\Gamma_0^T) & 0 & 0 & C_r(\Gamma \otimes \Gamma)E_u & C_r(\Gamma_0 \otimes \Gamma_0)E_{r-u} \end{pmatrix}. \tag{17}$$

Again by Proposition 4.1 of [14], $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})] = H(H^TJ_hH)^\dagger H^T$. We write $H = H_1H_2$, where

$$H_1 = \begin{pmatrix} B^T\eta^T \otimes \Gamma_0 & B^T \otimes \Gamma & (I_p - M_B\Sigma_X) \otimes \Gamma\eta & 0 & 0 \\ 2C_r(\Gamma\Omega \otimes \Gamma_0 - \Gamma \otimes \Gamma_0\Omega_0) & 0 & 0 & C_r(\Gamma \otimes \Gamma)E_u & C_r(\Gamma_0 \otimes \Gamma_0)E_{r-u} \end{pmatrix}, \tag{18}$$

and

$$H_2 = \begin{pmatrix} I_u \otimes \Gamma_0^T & 0 & 0 & 0 & 0 \\ \eta^T \otimes \Gamma^T & I_{ud} & (B\Sigma_XB^T)^{-1}B\Sigma_X \otimes \eta & 0 & 0 \\ 0 & 0 & I_{pd} & 0 & 0 \\ 2C_u(\Omega \otimes \Gamma^T) & 0 & 0 & I_{r(r+1)/2} & 0 \\ 0 & 0 & 0 & 0 & I_{(r-u)(r-u+1)/2} \end{pmatrix}. \tag{19}$$

Since H_2 is of full rank, $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})] = H_1(H_1^TJ_hH_1)^\dagger H_1^T$. First we calculate $H_1^TJ_hH_1$. We write H_1 as $H_1 = (H_{11}, H_{12}, H_{13}, H_{14}, H_{15})$. Since

$$J_hH_{11} = \begin{pmatrix} 4\widetilde{\Sigma}_XB^T\eta^T \otimes \Gamma_0\Omega_0^{-1} \\ 4ME_r^T(\Gamma \otimes \Gamma_0\Omega_0^{-1} - \Gamma\Omega^{-1} \otimes \Gamma_0) \end{pmatrix}, \tag{20}$$

we have $H_{1i}^T J_h H_{11} = 0$ for $i = 2, 3, 4, 5$. Because

$$J_h(H_{12}, H_{13}) = 4 \begin{pmatrix} \widetilde{\Sigma}_X B^T \otimes \Gamma\Omega^{-1} & (\widetilde{\Sigma}_X - \widetilde{\Sigma}_X M_B \widetilde{\Sigma}_X) \otimes \Gamma\Omega^{-1}\eta \\ 0 & 0 \end{pmatrix}, \qquad (21)$$

then $H_{1i}^T J_h H_{12} = 0$ for $i = 3, 4, 5$, and $H_{1i}^T J_h H_{13} = 0$ for $i = 4, 5$. Therefore

$$H_1^T J_h H_1 = \begin{pmatrix} H_{11}^T J_h H_{11}^T & 0 & 0 & 0 \\ 0 & H_{12}^T J_h H_{12}^T & 0 & 0 \\ 0 & 0 & H_{13}^T J_h H_{13}^T & 0 \\ 0 & 0 & 0 & (H_{14}, H_{15})^T J_h (H_{14}^T, H_{15}^T)^T \end{pmatrix},$$

and

$$(H_1^T J_h H_1)^\dagger = \begin{pmatrix} (H_{11}^T J_h H_{11}^T)^\dagger & 0 & 0 & 0 \\ 0 & (H_{12}^T J_h H_{12}^T)^\dagger & 0 & 0 \\ 0 & 0 & (H_{13}^T J_h H_{13}^T)^\dagger & 0 \\ 0 & 0 & 0 & [(H_{14}, H_{15})^T J_h (H_{14}^T, H_{15}^T)^T]^\dagger \end{pmatrix}.$$

As a consequence,

$$\begin{aligned} \mathrm{avar}\{\sqrt{n} h(\widehat{\phi})\} &= H_{11}(H_{11}^T J_h H_{11})^\dagger H_{11}^T + H_{12}(H_{12}^T J_h H_{12})^\dagger H_{12}^T + H_{13}(H_{13}^T J_h H_{13})^\dagger H_{13}^T \\ &\quad + (H_{14}, H_{15})((H_{14}, H_{15})^T J_h (H_{14}^T, H_{15}^T)^T)^\dagger (H_{14}^T, H_{15}^T). \end{aligned}$$

The asymptotic covariance of $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})]$ corresponds to the upper left block of $\mathrm{avar}[\sqrt{n} h(\widehat{\phi})]$, and $(H_{14}, H_{15})[(H_{14}, H_{15})^T J_h (H_{14}^T, H_{15}^T)^T]^\dagger (H_{14}^T, H_{15}^T)$ does not contribute to it. So we only consider the upper left block of $\sum_{i=1}^{3} H_{1i}(H_{1i}^T J_h H_{1i})^\dagger H_{1i}^T$. The upper left block of $H_{11}(H_{11}^T J_h H_{11})^\dagger H_{11}^T$ is

$$\begin{aligned}
& (B^T\eta^T \otimes \Gamma_0)(H_{11}^T J_h H_{11})^\dagger(\eta B \otimes \Gamma_0^T) \\
= \; & \frac{1}{4}(B^T\eta^T \otimes \Gamma_0)[\eta B\widetilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} - 2I_{u(r-u)} + \Omega^{-1} \otimes \Omega_0)]^{-1}(\eta B \otimes \Gamma_0^T).
\end{aligned}$$

The upper left block of $H_{12}(H_{12}^T J_h H_{12})^\dagger H_{12}^T$ is

$$(B^T \otimes \Gamma)(H_{12}^T J_h H_{12})^\dagger(B \otimes \Gamma^T) = \frac{1}{4}(B^T \otimes \Gamma)\{(B\widetilde{\Sigma}_X B^T)^{-1} \otimes \Omega\}(B \otimes \Gamma^T) = \frac{1}{4}M_B \otimes \Gamma\Omega\Gamma^T.$$

And the upper left block of $H_{13}(H_{13}^T J_h H_{13})^\dagger H_{13}^T$ is

$$\begin{aligned}
& \frac{1}{4}[(I_p - M_B\widetilde{\Sigma}_X) \otimes \Gamma\eta](\Sigma_X \otimes \eta^T\Omega^{-1}\eta)^{-1}[(I_p - \widetilde{\Sigma}_X M_B) \otimes \eta^T\Gamma^T] \\
= \; & \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T.
\end{aligned}$$

Hence $\mathrm{avar}\{\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})\}$ is given by

$$\begin{aligned}
& \frac{1}{4}(B^T\eta^T \otimes \Gamma_0)[\eta B\widetilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} - 2I_{u(r-u)} + \Omega^{-1} \otimes \Omega_0)]^{-1}(\eta B \otimes \Gamma_0^T) \\
+ \; & \frac{1}{4}M_B \otimes \Gamma\Omega\Gamma^T + \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T. \qquad (22)
\end{aligned}$$

C Additional simulation results for Section 7.2

We repeated the simulation with the same setting as in Figure 6 of the paper, but the errors were generated from the multivariate normal distribution $N(0, 2\Sigma)$. The comparison of the estimation standard deviation and MSE are displayed in Figure 1. Under this setting, the basic reduced-rank envelope estimator is the MLE. The reduced-rank envelope estimator with approximate weights has about the same efficiency as the MLE, since the weights are adaptive to the data. The reduced-rank envelope estimator with normal mixture weights (computed from two normal distributions $N(0, 2\Sigma)$ and $N(0, 0.1\Sigma)$ with probability 0.5 and 0.5) loses some efficiency because of the wrong weights. For example, at sample size 100, the ratios of estimation standard deviations of the reduced-rank envelope estimator with normal mixture weights versus the basic reduced-rank envelope estimator range from 1.04 to 1.17 with an average of 1.08. The MSE shows a similar pattern as the estimation standard deviation. The bias of the three reduced-rank envelope estimators is about the same, while the absolute value of the bias of the OLS estimator is slightly larger. Since bias is not a major component of the MSE, we did not include it in Figure 1.

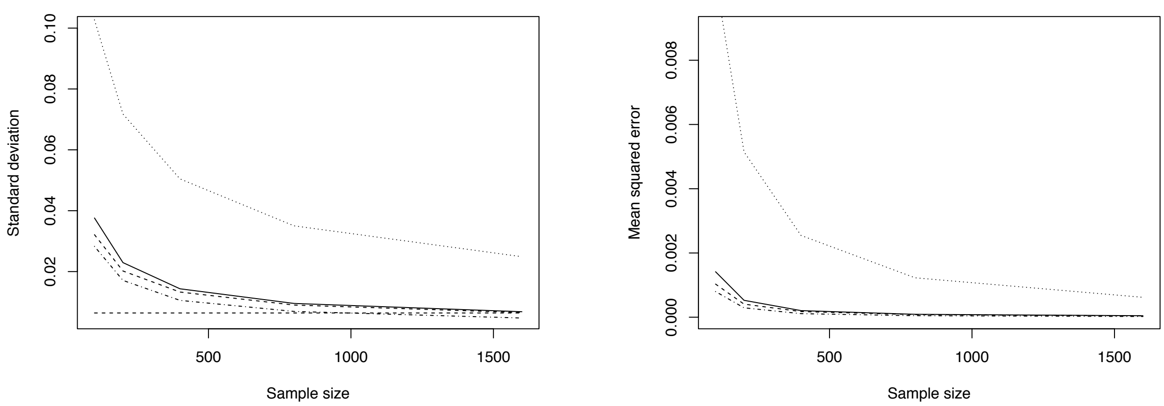

Figure 1: Estimation standard deviation and MSE for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 6. The horizontal solid line at the bottom of the left panel marks the asymptotic standard deviation of the basic envelope estimator.

References

[1] Bai, Z., and Wu, Y. (1993). Recursive algorithm for M-estimates of regression coefficients and scatter parameters in linear models. *Sankhyā: The Indian Journal of Statistics, Series B*, **55**, 199-

218.

[2] Bilodeau, M. and Brenner, M. (1999). *Theory of multivariate statistics*, Springer.

[3] Cook, R. D., Li, B. and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression (with discussion). *Statist. Sinica*, **20**, 927–1010.

[4] Cook, R. D. and Forzani, L. (2011). On the mean and variance of the generalized inverse of a singular Wishart matrix. *Electron. J. Stat.*, **5**, 146–158.

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

[6] Cook, R. D., Li, B. and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression (with discussion). *Statist. Sinica*, **20**, 927–1010.

[7] Eaton, M. L. (2007). *Multivariate Statistics: A Vector Space Approach.* Beachwood, Ohio: Institute of Mathematical Statistics.

[8] Kent, J.T., and D.E. Tyler (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.*, **19**, 2102–2019.

[9] Kudraszow, N. L. and Maronna, R. A. (2011). Estimates of MM type for the multivariate linear model. *J. Multivariate Anal.*, **102**, 1280–1292.

[10] Maronna, R. A. (1976). Robust estimators of Multivariate location and scatter. *Ann. Statist.*, **4**, 51–67.

[11] Miao, B. Q, and Wu, Y. (1996). Limiting behavior of recursive M-estimators in multivariate linear regression models. *J. Multivariate Anal.*, **59**, 60–80

[12] Mitchell, A. F. S. (1989). The information matrix, skewness tensor and α-connections for the general multivariate elliptic distribution. *Ann. Inst. Statist. Math.*, **41**, 289–304.

[13] Petersen K.B. and Pedersen M.S. (2008). *The matrix cookbook*. Desktop reference. http://www.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf

[14] Shapiro, A. (1986). Asymptotic theory of over-parameterized structural models. *J. Amer. Statist. Assoc.*, **81**, 142–149.

[15] van der Vaart, A.W. (2000). *Asymptotic Statistics*, Cambridge University Press.