Statistica Sinica Preprint No: SS-2017-0424	
Title	Envelopes for elliptical multivariate linear regression
Manuscript ID	SS-2017-0424
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0424
Complete List of Authors	Liliana Forzani and Zhihua Su
Corresponding Author	Zhihua Su
E-mail	zhihuasu@stat.ufl.edu
Notice: Accepted version subject to English editing.	

Envelopes for elliptical multivariate linear regression

Liliana Forzani* and Zhihua Su†‡

Abstract

We incorporate the idea of reduced rank envelope [7] to elliptical multivariate linear regression to improve the efficiency of estimation. The reduced rank envelope model takes advantage of both reduced rank regression and envelope model, and is an efficient estimation technique in multivariate linear regression. However, it uses the normal log-likelihood as its objective function, and is most effective when the normality assumption holds. The proposed methodology considers elliptically contoured distributions and it incorporates this distribution structure into the modeling. Consequently, it is more flexible and its estimator outperforms the estimator derived for the normal case. When the specific distribution is unknown, we present an estimator that performs well as long as the elliptically contoured assumption holds.

Keywords: envelopes; elliptical multivariate linear regression; reduced rank regression

*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.

1 Introduction

The multivariate linear regression model studies the conditional distribution of a stochastic response vector $Y \in \mathbb{R}^r$ as a linear function of the predictor $X \in \mathbb{R}^p$. It can be formulated as

$$Y = \mu_Y + \beta(X - \mu_X) + \epsilon, \tag{1}$$

where $\beta \in \mathbb{R}^{r\times p}$ is the coefficient matrix, the error vector ϵ is independent of X and follows a normal distribution with mean zero and covariance matrix Σ. The standard method of estimation fits a linear regression model for each response independently. The association among the response is not used and the efficiency of the estimation can be improved by considering these dependences. The envelope model is introduced in a seminal paper [8] in the context of (1) and the key idea is to identify the part of Y that is immaterial to the changes in X by using sufficient dimension reduction techniques. This immaterial part is removed from subsequent analysis, making the estimation more efficient. Another method that considers the association among the response is the reduced rank regression ([1], [2], [26], [41], [46]). Reduced rank regression assumes that the rank of the matrix $\beta \in \mathbb{R}^{r\times p}$ is less than or equal to d, where $d \leq \min(r, p)$. It has less number of parameters and therefore more efficient estimators can be obtained.

The envelope model and the reduced rank regression both use dimension reduction techniques to improve the estimation efficiency, but they have different perspectives and make different assumptions. In practice, it may take considerable effort to find out which method is more efficient for a given problem. The reduced rank envelope was recently proposed in [7], which combines the advantage of both methods and is more efficient than both methods. However, the estimation of the reduced rank envelope model takes the normal likelihood function as the objective function, and is most effective when the normality assumption holds.

It is well-known that the normality assumption is not always reasonable in many applications, and alternative distributions (or methodologies) have to be considered to fit the distribution of the data better. One choice is the family of elliptically contoured distributions, which includes the classical normal distribution and many important distributions such as Student-t, power exponential, contaminated normal, etc. They can have heavier or lighter tails than the normal distribution, and are more adaptive to the data. Elliptical multivariate linear regression models have been extensively studied in the statistical literature, see for example [5], [10], [12], [13], [18], [19],[22], [20], [28], [30], [32], [33], [34], [40], [43], [45] and [48] among others. In particular, [32] introduces a general elliptical multivariate regression model in which the mean vector and the scale matrix have parameters in common. Then they unify several elliptical models, such as nonlinear regressions, mixed-effects model with nonlinear fixed effects, errors-in-variables models, etc. Bias correction for the maximum likelihood estimator and adjustments of the likelihood-ratio statistics are also derived for this general model (see [38], [37]). The elliptical distributions can also be used as the basis to consider robustness in multivariate linear regression, as in [14], [21], [29], [36], [42], [50] and many others. Nevertheless the envelope model under the context of elliptical multivariate regression has not yet been implemented. There is also not much literature about reduced rank regression beyond the normal case. The only attempt to extend reduced rank regression to the non-normal case is through M-estimators or other robust estimators which include some of the elliptical class. For example, [50] develops a robust estimator in reduced rank regression and proposes a novel rank-based estimation procedure using Wilcoxon scores. While the reduced rank estimator in [50] allows a general error distribution, we aim to further improve the efficiency of the reduced rank estimator using maximum likelihood estimators (MLE) and envelope methods in the context of elliptical multivariate linear regression.

The goal of this paper is to derive the reduced rank regression estimator, envelope estimator and reduced rank envelope estimator for elliptical multivariate linear regression. Since both reduced rank regression and the envelope model are special cases of the reduced rank envelope model, we present a unified approach that focuses on the reduced rank envelope model. The asymptotic properties and efficiency gains of the reduced rank regression, envelope model and reduced rank envelopes model will be studied, and we will demonstrate their effectiveness in simulations and real data examples.

The rest of this paper is organized as follows. In Section 2, we introduce the reduced rank regression, envelope model and reduced rank envelope model under the elliptical multivariate linear regression. Section 3 describes the most used elliptically contoured distributions in elliptical multivariate linear regression. In Section 4, we derive the maximum likelihood estimators (MLE) for the models considered in Section 2, and propose a weighted least square estimator when the error distribution is unknown but elliptically contoured. Section 5 studies the asymptotic properties of the estimators, and demonstrates the efficiency gains without the normality assumptions. Section 6 discusses the selection of the rank and dimension in the reduced rank envelope model. The simulation results are presented in Section 7, and examples are given in Section 8 for illustration. Proofs are included in the Online Supplement.

2 Models

We consider the following elliptical multivariate linear regression model given by

$$Y = \mu_Y + \beta(X - \mu_X) + \epsilon, \quad \epsilon \sim EC_r(0, \Sigma, g_{Y|X}), \tag{2}$$

where $Y \in \mathbb{R}^r$ denotes the response vector, $X \in \mathbb{R}^p$ denotes the predictor vector, $\beta \in \mathbb{R}^{r\times p}$, and $\sim$ denotes equal in distribution. If a random vector $Z \in \mathbb{R}^m$ follows an elliptically contoured distribution $EC_m(\mu_Z, \Sigma_Z, g_Z)$ with density, then the density function is given by

$$f_Z(z) \quad = |\Sigma_Z|^{-\frac{1}{2}} g_Z\left[(z-\mu_Z)^T \Sigma_Z^{-1}(z-\mu_Z)\right], \tag{3}$$

where $\mu_Z \in \mathbb{R}^m$ is the location parameter; $\Sigma_Z \in \mathbb{R}^{m\times m}$ is a positive definite scale matrix; $g_Z(\cdot) \geq 0$ is a real-valued function and $\int_0^\infty u^{m/2-1} g_Z(u)du < \infty$. We call (2) the *standard model* in the following discussions. Based on (2), $Y \mid X$ follows the elliptically contoured distribution $EC_r(\mu_{Y|X}, \Sigma, g_{Y|X})$ where $\mu_{Y|X} = \mu_Y + \beta(X - \mu_X)$. When the conditional expectation and variance exist, $\mathrm{E}(Y \mid X) = \mu_{Y|X}$ and $\mathrm{var}(Y \mid X) = c_X\Sigma$, where $c_X = \mathrm{E}(Q^2)/r$ and $Q^2 = (Y - \mu_{Y|X})^T\Sigma^{-1}(Y - \mu_{Y|X})$ (see Corollary 2 in [17], p. 65). Notice that in general $\mathrm{var}(Y \mid X)$ depends on X, except for the normal errors with constant variance.

The reduced rank regression assumes that the rank of the coefficient β in model (2) is at most $d \leq \min(p, r)$. As a consequence

$$\beta = AB,\ A \in \mathbb{R}^{r\times d},\ \ B \in \mathbb{R}^{d\times p},\ \mathrm{rank}(A) = \mathrm{rank}(B) = d, \tag{4}$$

for some $A \in \mathbb{R}^{r\times d}$ and $B \in \mathbb{R}^{d\times p}$. Note that A and B are not identifiable since $AB = (AU)(U^{-1}B) := A^*B^*$ for any invertible U. If the errors are normally distributed with a constant covariance matrix, the MLE of β for (4) and its asymptotic distribution were derived in [2], [41] and [46], by imposing various constraints on A and B for identifiability. Since the goal is to estimate β, rather than A and/or B, recently [7] derived the estimator of β without imposing any constraints on A and B other than requiring the rank of β is equal to d. It has been shown that when ϵ follows the multivariate normal distribution with a constant covariance matrix, the reduced rank regression has the potential to yield more efficient estimator for β than the ordinary least square (OLS) estimator. Note that under normality, the OLS estimator is the MLE.

The envelope model [8] is another way to get efficient estimator. Let $\mathrm{span}(\beta)$ denote the subspace spanned by the columns of β. Under model (2), if $\mathrm{span}(\beta)$ is contained in the span of m ($m < r$) eigenvectors of the error covariance matrix Σ, not necessarily the leading eigenvectors, then the envelope estimator of β is expected to be more efficient than the OLS estimator. More specifically, let $\mathcal{S}$ be a subspace of $\mathbb{R}^r$ that is spanned by some eigenvectors of Σ, and $\mathrm{span}(\beta) \subseteq \mathcal{S}$. The intersection of all such $\mathcal{S}$ is called the Σ-envelope of β, which is denoted by $\mathcal{E}_\Sigma(\beta)$. Let u be the dimension of $\mathcal{E}_\Sigma(\beta)$. Then $u \leq r$. Take $\Gamma \in \mathbb{R}^{r\times u}$ to be an orthonormal basis of $\mathcal{E}_\Sigma(\beta)$ and $\Gamma_0 \in \mathbb{R}^{r\times(r-u)}$ to be a completion of Γ, i.e., (Γ, Γ_0) is an orthogonal matrix. Since $\mathrm{span}(\beta) \subseteq \mathcal{E}_\Sigma(\beta) = \mathrm{span}(\Gamma)$, there exists a $\xi \in \mathbb{R}^{u\times p}$ such that $\beta = \Gamma\xi$. Because that $\mathcal{E}_\Sigma(\beta)$ is spanned by eigenvectors of Σ, there exist $\Omega \in \mathbb{R}^{u\times u}$ and $\Omega_0 \in \mathbb{R}^{(r-u)\times(r-u)}$ that $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$. Here ξ carries the coordinates of β with respect to Γ and Ω and Ω_0 carry the coordinates of Σ with respect to Γ and Γ_0. To summarize, if β and Γ satisfy the following conditions

$$\beta = \Gamma\xi \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \tag{5}$$

we called (2) an envelope model of dimension u. Based on (5), $\mathcal{E}_\Sigma(\beta)$ provides a link between β and Σ: The variation of the ϵ can be decomposed to one part $\Gamma\Omega\Gamma^T$ that is material to the estimation β and the other part $\Gamma_0\Omega_0\Gamma_0^T$ that is immaterial to the estimation of β. By realizing this decomposition, in the normal setting, [8] showed that the envelope estimator of β is more or at least as efficient as the OLS estimator asymptotically. The efficiency can be substantial especially if the immaterial variation $\|\Gamma_0\Omega_0\Gamma_0^T\|$ is substantially larger than the material variation $\|\Gamma\Omega\Gamma^T\|$, where $\|\cdot\|$ denotes the spectral norm of a matrix.

Under normal distribution for the error term, [7] presented a novel unified framework of the reduced rank regression and the envelope model called the reduced rank envelope model, which obtains more efficient estimators compared to either of them. [7] assumed that β and Σ follows the envelope structure (5) and at the same time the coordinate ξ has a reduced rank structure $\xi = \eta B$, where $\eta \in \mathbb{R}^{u\times d}$ and $B \in \mathbb{R}^{d\times p}$ with the rank $d \leq \min(r, p)$. Then model (2) is called the reduced rank envelope model when

$$\beta = AB = \Gamma\eta B, \;\; \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \tag{6}$$

where $B \in \mathbb{R}^{d\times p}$ has rank d, $\eta \in \mathbb{R}^{u\times d}$, $\Omega \in \mathbb{R}^{u\times u}$ and $\Omega_0 \in \mathbb{R}^{(r-u)\times(r-u)}$ are positive-definite matrices, and $\Gamma_0 \in \mathbb{R}^{r\times(r-u)}$ is a completion of Γ, i.e. (Γ, Γ_0) is an orthogonal matrix. The reduced rank envelope model performs dimension reduction in two levels: The first level $\beta = AB$ assumed that we have a reduced rank regression. The second level $\beta = \Gamma\eta B$ is based on the assumption that β only intersects u eigenvectors of the covariance matrix Σ. When $u = r$, $\Gamma = I_r$, then (6) degenerates to the usual reduced rank regression (4). When $d = \min(u, p)$, then (6) reduces to an envelope model (5). Finally, the reduced rank envelope model (6) is equivalent to the standard model (2) if $d = u = r$. [7] obtained the MLEs of β and Σ, as well as their asymptotic distributions under the normality assumption. It should be noted that for the reduced rank regression, the envelope model and the reduced rank envelope model, the constituent parameters A, B, Γ, Γ_0, ξ, η, Ω, Ω_0 are not unique. Hence, they are not identifiable. Nevertheless, β and Σ are unique. No additional constraints are imposed on the constituent parameters in [7] when studying the asymptotic distribution of the identifiable parameters β and Σ.

3 Examples

In this section we present three scenarios where the elliptically contoured distributions are used in regression.

3.1 The data matrix is elliptically contoured distributed

A case that is commonly studied in the literature is that the data matrix follows a matrix elliptically contoured distribution. A $p \times q$ random matrix Z follows a matrix elliptically contoured distribution $EC_{p,q}(M, A \otimes B, \Psi)$ if and only if $\mathrm{vec}(Z^T)$ follows an elliptically contoured distribution $EC_{pq}(\mathrm{vec}(M^T), A \otimes B, \Psi)$, where $\otimes$ denotes Kronecker product, and vec denotes the vector operator that stacks the columns of a matrix into a vector.

Let $\mathbb{X} = (X_1^T, \ldots, X_n^T)^T \in \mathbb{R}^{n\times p}$ and $\mathbb{Y} = (Y_1^T, \ldots, Y_n^T)^T \in \mathbb{R}^{n\times r}$ be data matrices such that $\mathbb{Y} \mid \mathbb{X}$ follows a matrix elliptically contoured distribution $EC_{n,r}(M, \eta \otimes \Sigma, g)$ with $M = 1_n\mu_Y^T + (\mathbb{X} - 1_n\mu_X^T)\beta^T$, where 1_n denotes an n dimensional vector of 1's. Under this assumption, by using Theorem 2.8 from [24], we have $Y_i \mid \mathbb{X} \sim Y_i \mid X_i \sim EC_r(\mu_Y + \beta(X_i - \mu_X), \eta_{ii}\Sigma, g)$. This allows the errors to be modeled with a heteroscedastic structure. More properties of the matrix elliptically contoured distribution are discussed in [24]. Examples of this distribution include matrix variate symmetric Kotz Type distribution, Pearson Type II distribution, Pearson Type VII distribution, symmetric Bessel distribution, symmetric Logistic distribution, symmetric stable law, etc. Among these distributions, the most common one is the normal with non-constant variance [16].

As an example of the matrix elliptically contoured distribution, we consider that $\mathbb{Y} \mid \mathbb{X}$ follows a matrix normal distribution $N_{n\times r}(M, \eta \otimes \Sigma)$ with $M = 1_n\mu_Y^T + (I_n - \frac{1}{n}1_n1_n^T)\mathbb{X}\beta^T$ and η being a diagonal matrix. The diagonal elements of η are denoted by η_{ii} and $\eta_{ii} > 0$ for $i = 1, \ldots, n$. Then

$$Y_i = \mu_Y + \beta(X_i - \mu_X) + \epsilon,$$

where $\epsilon \sim N(0, \eta_{ii}\Sigma)$. Therefore $Y_i \mid X_i$ follows a normal distribution with mean $\mu_Y + \beta(X_i - \mu_X)$ and covariance matrix $\eta_{ii}\Sigma$. In other words, it is an elliptically contoured distribution $EC_r(\mu_Y + \beta(X_i - \mu_X), \Sigma, g_i)$ with $g_i(t) = (2\pi\eta_{ii})^{-r/2}e^{-\frac{t}{2\eta_{ii}}}$.

Note that we only considered the error structure that the covariance matrices are proportional, and the heteroscedasticity only depends on g, not the scale parameter. The general non-constant covariance structure is not included because we would need a general $rn \times rn$ scale matrix instead of $\eta \otimes \Sigma$ in the matrix elliptically contoured distribution.

3.2 X and Y is jointly elliptically contoured distributed

Sometimes $(X^T, Y^T)^T$ jointly follows an elliptically contoured distribution or it can be transformed to ellipticity (e.g., [9]). Suppose $(X^T, Y^T)^T$ follows the distribution $EC_{p+r}((\mu_X^T, \mu_Y^T)^T, \Phi, g)$, then its density function is

$$f_{X,Y}(x,y) = |\Phi|^{-\frac{1}{2}} g\left[\left\{(x^T, y^T) - (\mu_X^T, \mu_Y^T)\right\} \Phi^{-1} \left\{(x^T, y^T) - (\mu_X^T, \mu_Y^T)\right\}^T\right]$$

where $g(\cdot) \geq 0$ and Φ is a $(p+r) \times (p+r)$ positive definite matrix. Following [5], if we partition Φ as

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_X & \Phi_Y \\ \Phi_Y^T & \Sigma_Y \end{pmatrix}, \tag{7}$$

then X and Y are marginally elliptically contoured distributed where X follows $EC_p(\mu_X, \Sigma_X, g)$ and Y follows $EC_r(\mu_Y, \Sigma_Y, g)$ (Theorem 2.8 from [24]). The conditional distribution of $Y \mid X$ is also elliptically contoured

$$Y \mid X \sim EC_r(\mu_{Y|X}, \Phi_{22.1}, g_{Y|X}),$$

where $\mu_{Y|X} = \mu_Y + \Phi_{21}\Phi_{11}^{-1}(X - \mu_X)$, $\Phi_{22.1} = \Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{21}^T$ and $g_{Y|X}(t) = g(t + m(X))/g(m(X))$ with $m(X) = (X - \mu_X)^T \Phi_{11}^{-1}(X - \mu_X)$. Note that $\mu_{Y|X}$ is linear in X and $\Phi_{22.1}$ is a constant.

Now we use multivariate t-distribution as an example. Suppose that $Z \in R^k$ follows a multivariate t-distribution $t_k(\mu, \Sigma, \nu)$, where ν denotes the degrees of freedom. The density function of Z is given by

$$f_Z(z) = \frac{\Gamma(\frac{\nu+k}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\nu\pi^{k/2}} \frac{1}{\sqrt{|\Sigma|}} \left(1 + \frac{(z-\mu)^T\Sigma^{-1}(z-\mu)}{\nu}\right)^{-\frac{k+\nu}{2}}.$$

Suppose that $(X^T, Y^T)^T \sim t_{p+r}\big((\mu_X^T, \mu_Y^T)^T, \Phi, \nu\big)$ with Φ following the structure in (7). Then

$$Y \mid X \sim t_r\left(\mu_Y + \Phi_{21}\Phi_{11}^{-1}(X-\mu_X), \frac{\nu + (x-\mu_X)^T\Phi_{11}^{-1}(x-\mu_X)}{\nu+p}\Phi_{22.1}, \nu+p\right).$$

Equivalently, $Y \mid X \sim EC_r(\mu_{Y|X}, \Phi_{22.1}, g_{Y|X})$ with $\mu_{Y|X} = \mu_Y + \Phi_{21}\Phi_{11}^{-1}(X-\mu_X)$, $\Phi_{22.1} = \Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{21}^T$ and $g_{Y|X}(t) = c_{\nu,p,r}g[t/h(X)]/h(X)^{r/2}$, where $g(t) = (\nu+p+t)^{-\frac{p+r+\nu}{2}}$, $h(X) = [\nu + (X-\mu_X)^T\Phi_{11}^{-1}(X-\mu_X)]/(\nu+p)$ and $c_{\nu,p,r}$ is the normalizing constant.

3.3 Y given X follows an elliptically contoured distribution

It is also reasonable to assume that the error vector ϵ follows an elliptically contoured distribution, in other words, Y given X follows an elliptically contoured distribution. We will present two examples, $Y \mid X$ follows a normal mixture distribution and $Y \mid X$ follows a conditional t-distribution.

We say $Y \mid X$ follows a normal mixture distribution if its density function is a convex linear combination of normal density functions. Suppose $Y \mid X$ follows a normal mixture distribution from m normal distributions $N_r(\mu_{Y|X}, k_i\Sigma)$, $i = 1, \ldots, m$ with weights $p_1, \ldots, p_m$, then its density function is given by

$$f_{Y|X}(y) = \sum_{i=1}^m p_i k_i^{-\frac{r}{2}} \frac{1}{(2\pi)^{r/2}|\Sigma|^{1/2}} e^{-\frac{1}{2k_i}(y-\mu_{Y|X})^T\Sigma^{-1}(y-\mu_{Y|X})},$$

where $k_i > 0, p_i > 0$ for $i = 1, \ldots, r$ and $\sum_{i=1}^m p_i = 1$. Equivalently $Y \mid X$ follows an elliptically contoured distribution $EC_r(\mu_{Y|X}, \Sigma, g)$ with $g(t) = \sum_{i=1}^m p_i k_i^{-r/2}(2\pi)^{-r/2}|\Sigma|^{-1/2}e^{-t/(2k_i)}$.

The t-distribution is useful to model heavy tails. As discussed in Section 3.2, $Y \mid X$ follows a

t-distribution $t_r(\mu_Y + \beta(X - \mu_X), \Sigma, \nu)$ if its density function takes the form

$$f_{Y|X}(y) = \frac{\Gamma(\frac{\nu+r}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\nu\pi^{r/2}} \frac{1}{\sqrt{|\Sigma|}} \left(1 + \frac{[y - \mu_Y + \beta(X - \mu_X)]^T \Sigma^{-1} [y - \mu_Y + \beta(X - \mu_X)]}{\nu}\right)^{-\frac{r+\nu}{2}}.$$

Equivalently, $Y \mid X$ follows an elliptically contoured distribution $EC_r(\mu_Y + \beta(X - \mu_X), \Sigma, g)$ with $g(t) = c_{\nu,r}(1 + \frac{1}{\nu}t)^{-\frac{\nu+r}{2}}$, where $c_{\nu,r} = \nu^{-1}\pi^{-r/2}\Gamma(\frac{\nu+r}{2})/\Gamma(\frac{\nu}{2})$ is a normalizing constant.

4 Estimation

Under the standard model (2), if the errors $(\epsilon_1, \ldots, \epsilon_n)$ jointly follow a matrix elliptically contoured distribution, the OLS estimator of β is its MLE (See Chapter 9, [15]). When the errors $(\epsilon_1, \ldots, \epsilon_n)$ do not jointly follow a matrix elliptically contoured distribution but $g_{Y|X}$ is known, an estimator of β can be computed via an iterative re-weighted least squares algorithm and its properties are studied in [5]. When $g_{Y|X}$ is unknown, [5] derived an estimator of β and investigated its properties.

The goal of this section is to derive the MLEs for the reduced rank regression, the envelope model and the reduced rank envelope model for given d, u and $g_{Y|X}$, where d is the rank of β and u is the dimension of the envelope $\mathcal{E}_\Sigma(\beta)$. Procedures for selecting d and u are discussed in Section 6. Note that when the errors $(\epsilon_1, \ldots, \epsilon_n)$ jointly follow a matrix elliptically contoured distribution, the estimators of β obtained by [7] are the MLEs of the corresponding models.

4.1 Parametrization for the different models

Let vech denote the vector half operator that stacks the lower triangle of a matrix to a vector. Then under the standard model (2), the parameter vector is $h = \left(\mathrm{vec}^T(\beta),\ \mathrm{vech}^T(\Sigma)\right)^T$. We did not consider μ_X or μ_Y because the estimators are asymptotically independent to the estimators of β and Σ. We use ψ to denote the parameter vector of the reduced rank regression (4), δ for the envelope model (5) and ϕ for the reduced rank envelope model (6). Then

$$h = \begin{pmatrix} \mathrm{vec}(\beta) \\ \mathrm{vech}(\Sigma) \end{pmatrix}, \quad \psi = \begin{pmatrix} \mathrm{vec}(A) \\ \mathrm{vec}(B) \\ \mathrm{vech}(\Sigma) \end{pmatrix}, \quad \delta = \begin{pmatrix} \mathrm{vec}(\Gamma) \\ \mathrm{vec}(\xi) \\ \mathrm{vech}(\Omega) \\ \mathrm{vech}(\Omega_0) \end{pmatrix}, \quad \phi = \begin{pmatrix} \mathrm{vec}(\Gamma) \\ \mathrm{vec}(\eta) \\ \mathrm{vec}(B) \\ \mathrm{vech}(\Omega) \\ \mathrm{vech}(\Omega_0) \end{pmatrix}.$$

We use $N(v)$ to denote the number of parameters in a parameter vector v. Then $N(h) = pr + r(r+1)/2$, $N(\psi) = (r-d)d + pd + r(r+1)/2$, $N(\delta) = pu + r(r+1)/2$ and $N(\phi) = (u-d)d + pd + r(r+1)/2$. The reduced rank regression has less parameters than the standard model since $N(h) - N(\psi) = (p-d)(r-d) \geq 0$, and the reduced rank envelope model has even less parameters than the reduced rank regression as $N(\psi) - N(\phi) = (r-u)d \geq 0$. On the other hand, compared to the standard model, the number of parameters is reduced by $p(r-u) \geq 0$ by using the envelope model and it is further reduced by $(p-d)(u-d) \geq 0$ by using the reduced rank envelope model.

Remark: If the model assumption holds, less parameters often result in an improvement of estimation efficiency. And the improved efficiency often leads to an improved prediction accuracy. However, if the model assumption does not hold, having less parameters will introduce bias but may still reduce the variance of the estimator. Then it is a bias-variance trade-off on if the increase in bias or reduction in variance dominates.

4.2 Maximum likelihood estimators

Assume that $Y \mid X$ follows an elliptically contoured distribution $EC_r(0, \Sigma, g_{Y|X})$ with density given by (3). Let (X_i, Y_i) be n independent samples of (X, Y), $i = 1, \ldots, n$, and let $m_i = [Y_i - \mu_Y - \beta(X_i - \mu_X)]^T \Sigma^{-1} [Y_i - \mu_Y - \beta(X_i - \mu_X)]$. The log-likelihood function is given by

$$l = -\frac{n}{2} \log |\Sigma| + \sum_{i=1}^{n} \log g(m_i),$$

where we denote $g_{Y|X}$ as g from now on. Taking the derivative of the log-likelihood function with respect to β and Σ and setting to zero, we have

$$\frac{\partial l}{\partial \beta} = -\frac{1}{2} \sum_{i=1}^{n} W_i \frac{\partial m_i}{\partial \beta} = 0, \qquad \frac{\partial l}{\partial \Sigma} = -\frac{n}{2} \Sigma^{-1} - \frac{1}{2} \sum_{i=1}^{n} W_i \frac{\partial m_i}{\partial \Sigma} = 0,$$

where $W_i = -2g'(m_i)/g(m_i)$. If $Y \mid X$ followed a normal distribution, the log-likelihood function would be

$$l_2 = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^{n} m_i.$$

Taking the derivative of l_2 with respect to β and Σ and setting to zero, we have

$$\frac{\partial l_2}{\partial \beta} = -\frac{1}{2} \sum_{i=1}^{n} \frac{\partial m_i}{\partial \beta} = 0, \qquad \frac{\partial l_2}{\partial \Sigma} = -\frac{n}{2} \Sigma^{-1} - \frac{1}{2} \sum_{i=1}^{n} \frac{\partial m_i}{\partial \Sigma} = 0.$$

If the weights W_i are positive and they were known, we could transform the data to $(\sqrt{W_i} X_i, \sqrt{W_i} Y_i)$ and solve for β and Σ as if the data follow the normal distribution. With this idea in mind, we propose the following iterative re-weighted least squares (IRLS) algorithm when $W_i \geq 0$. The obtained estimator from this algorithm is equivalent to the MLE estimator (See [11] and [23]).

1. Get the initial values for β and Σ
2. Repeat the following until convergence
   (a) Compute $W_i = -2g'(m_i)/g(m_i)$ with β and Σ being the current estimator
   (b) With the data $(\sqrt{W_i}X_i, \sqrt{W_i}Y_i)$, update the estimators of β and Σ as if the data follow normal distribution.

The estimator in Step 2 (b) is obtained from a fast envelope estimation algorithm developed in [6], which is implemented in the R package `Renvlp` [31]. This algorithm can be used not only for the standard model, but also for the reduced rank regression, the envelope model and the reduced rank envelope model. Take the reduced rank envelope model as an example, we have

$$\frac{\partial l}{\partial \phi^T} = \frac{\partial l}{\partial h^T}\frac{\partial h}{\partial \phi^T}, \qquad \frac{\partial l_2}{\partial \phi^T} = \frac{\partial l_2}{\partial h^T}\frac{\partial h}{\partial \phi^T}.$$

Notice that the term $\partial h/\partial \phi^T$ is the same for both likelihoods, and h is a function of β and Σ. We can estimate the reduced rank envelope estimator using the preceding algorithm except that 2(b) is changed to "With the data $(\sqrt{W_i}X_i, \sqrt{W_i}Y_i)$, update the reduced rank envelope estimators of β and Σ as if the data follow normal distribution." The reduced rank regression and the envelope model can follow the same procedure. For completeness we include in the Online Supplement the derivatives of m_i with respect to the parameters β and Σ.

4.3 Weights

We now give the weights for some commonly used elliptically contoured distribution.

Normal with non-constant variance

If $Y_i \mid X_i$ follows the normal distribution $N(\mu_{Y|X}, \eta_{ii}\Sigma)$ with $\eta_{ii} > 0$, $i = 1, \ldots, n$. Then $W_i = 1/\eta_{ii}$.

Normal mixture distribution

Suppose $Y \mid X$ follows a normal mixture distribution from m normal distributions $N_r(\mu_{Y|X}, k_i\Sigma)$, $i = 1, \ldots, m$ with weights $p_1, \ldots, p_m$. From the discussions in Section 3.3, the weights are given by

$$W(t_i) = \frac{\sum_{j=1}^m p_j k_j^{-r/2-1} e^{-t_i/2k_j}}{\sum_{j=1}^m p_j k_j^{-r/2} e^{-t_i/2k_j}},$$

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

Multivariate t-distribution

Suppose that $(X^T, Y^T)^T$ follows a joint multivariate t-distribution $t_{p+r}((\mu_X^T, \mu_Y^T)^T, \Phi, \nu)$ with Φ following the structure in (7). Based on the discussion in Section 3.2, $Y \mid X$ follows the t-distribution $t_r(\mu_Y + \Phi_{21}\Phi_{11}^{-1}(X - \mu_X), \frac{\nu + (x-\mu_X)^T \Phi_{11}^{-1}(x-\mu_X)}{\nu+p}\Phi_{22.1}, \nu + p)$. After some straightforward calculations,

$$W_i(t_i) = \frac{p + r + \nu}{\nu + (X_i - \mu_X)^T \Phi_{11}^{-1}(X_i - \mu_X) + t_i},$$

where $t_i = [Y_i - \mu_Y - \Phi_{21}\Phi_{11}^{-1}(X_i - \mu_X)]^T \Sigma_{Y|X}^{-1} [Y_i - \mu_Y - \Phi_{21}\Phi_{11}^{-1}(X_i - \mu_X)]$ and $\Sigma_{Y|X} = \frac{\nu + (x-\mu_X)^T \Phi_{11}^{-1}(x-\mu_X)}{\nu+p}\Phi_{22.1}$.

Conditional t-distribution

Suppose that $Y \mid X$ follows a t-distribution with $t_r(\mu_{Y|X}, \Sigma, \nu)$, then

$$W(t_i) = \frac{\nu + r}{\nu + t_i},$$

where $t_i = (Y_i - \mu_{Y|X})^T \Sigma^{-1} (Y_i - \mu_{Y|X})$.

Notice that all these weights are positive. For illustration purpose, the constants η_{ii}'s in normal with non-constant variance and k_i's in normal mixture distribution are fixed and known in the calculation of the weights. If they are unknown, or more generally if g is unknown, Section 4.4 presents an algorithm to estimate the weights.

4.4 Weighted least square estimators

The IRLS algorithm in Section 4.2 requires the knowledge of g, which may not be available in practice. In this section we propose an algorithm for the case when g is unknown.

Suppose that the model has the structure in (2), then we have $\mathrm{var}(Y \mid X) = c_X \Sigma$, where $c_X = \mathrm{E}(Q^2)/r$ and $Q^2 = [Y - \mu_Y - \beta(X - \mu_X)]^T \Sigma^{-1} [Y - \mu_Y - \beta(X - \mu_X)]$ (see Corollary 2 in [17]). Notice that c_X can be different across the observations. We use c_{X_i} to denote c_X for the ith observation. If c_{X_i} is known, then we can transform the data to $(c_{X_i}^{-1/2} X_i, c_{X_i}^{-1/2} Y_i)$ and estimate the parameters as if the data follows the normal distribution. If c_{X_i} is unknown, we estimate it by $\hat{c}_{X_i} = [Y_i - \widehat{\mu}_Y - \widehat{\beta}(X_i - \widehat{\mu}_X)]^T \widehat{\Sigma}^{-1} [Y_i - \widehat{\mu}_Y - \widehat{\beta}(X_i - \widehat{\mu}_X)]$. According to [5], the resulting estimators of β and Σ are robust to a moderate departure from normality. Let $\bar{X}$ and $\bar{Y}$ denote the sample mean of X and Y. The following algorithm summarized the preceding discussion.

1. Get the initial values for β and Σ from the corresponding model, i.e. reduced rank regression, envelope model or reduced rank envelope model. Set the initial values of μ_X and μ_Y as $\bar{X}$ and $\bar{Y}$.

2. Repeat the following until convergence

   (a) Compute $\hat{c}_{X_i} = [Y_i - \widehat{\mu}_Y - \widehat{\beta}(X_i - \widehat{\mu}_X)]^T \widehat{\Sigma}^{-1} [Y_i - \widehat{\mu}_Y - \widehat{\beta}(X_i - \widehat{\mu}_X)]$, where $\widehat{\mu}_X$, $\widehat{\mu}_Y$, $\widehat{\beta}$ and $\widehat{\Sigma}$ are the estimates of μ_X, μ_Y, β and Σ.

(b) With the data $(\hat{c}_{X_i}^{-1/2}X_i, \hat{c}_{X_i}^{-1/2}Y_i)$, update the estimators of β, Σ, μ_X and μ_Y under the corresponding model as if the data follows normal distribution.

Note that this algorithm is similar to the algorithm discussed in Section 4.2 except that we are using $\hat{c}_{X_i}^{-1}$ as weights instead of using the exact weights computed from the knowledge of g. We call $\hat{c}_{X_i}^{-1}$ as *approximate weights* in subsequent discussions. The $\sqrt{n}$-consistency of the estimator of β obtained by using these *approximate weights* follows similarly to Theorem 4 from [5].

5 Asymptotics

In this section, we present the asymptotic distribution for the MLEs of β: the standard estimator $\widehat{\beta}_{\text{std}}$, the reduced rank regression estimator $\widehat{\beta}_{\text{RR}}$, the envelope model estimator $\widehat{\beta}_E$ and reduced rank envelope estimator $\widehat{\beta}_{\text{RE}}$.

Without loss of generality, we assume that $\mu_X = 0$ and the predictors are centered in the sample. Let C_r and E_r denote the contraction and expansion matrix defined in [25] that connects the vector operator vec and the vector half operator vech as follows $\text{vec}(S) = E_r \,\text{vech}(S)$ and $\text{vech}(S) = C_r\text{vec}(S)$ for any $r \times r$ symmetric matrix S. Let $U = \Sigma^{-1/2}[Y - \mu_Y - \beta(X - \mu_X)]$, $N_X = E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 U^TU \middle| X\right]/r$ and $M_X = E\left[\left(\frac{g'(U^TU)}{g(U^TU)}\right)^2 (U^TU)^2 \middle| X\right]/[r(r+2)]$. We define $\widetilde{\Sigma}_X = E(N_X XX^T)$ and $M = E(M_X)$ if X is random and the expectations exist, $\widetilde{\Sigma}_X = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n N_{X_i}X_iX_i^T$ and $M = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n M_{X_i}$ if X is fixed when the limits are finite. We further assume that $\widetilde{\Sigma}_X$ is positive definite and $M > 0$. For the rest of the section we ask g such that the above quantities are finites and that the maximum likelihood estimator for model (2) exists, is consistent and asymptotically normal (See for example the conditions for elliptical distributions on [39], [4], [3], [27], [29], [35], [39], [50] or more generally conditions on Theorems 5.23, 5.31, 5.39, 5.41 or 5.42 from [47]. Then, the Fisher information for $h = \left(\mathrm{vec}^T(\beta),\ \mathrm{vech}^T(\Sigma)\right)^T$ is given by

$$J_h = \begin{pmatrix} J_\beta & 0 \\ 0 & J_\Sigma \end{pmatrix}$$

with $J_\beta = 4\widetilde{\Sigma}_X \otimes \Sigma^{-1}$ and $J_\Sigma = 2ME_r^T(\Sigma^{-1} \otimes \Sigma^{-1})E_r + (M - \frac{1}{4})E_r^T\mathrm{vec}(\Sigma^{-1})\mathrm{vec}^T(\Sigma^{-1})E_r$. Detailed calculations are included in the Online Supplement. When ϵ follows a normal distribution, we have $N_X = M = 1/4$ and J_h has the same form as in the literature (e.g. [8]).

Proposition 1 gives the asymptotic variance of the MLEs of β under the standard model (2), the reduced rank regression (4), the envelope model (5) and the reduced rank envelope model (6). Suppose that $\widehat{\theta}$ is an estimator of θ. We write $\mathrm{avar}(\sqrt{n}\widehat{\theta}) = V$ if $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, V)$, where $\xrightarrow{d}$ denotes convergence in distribution.

Proposition 1 *Suppose that model* (2) *holds, i.e. the error vector* ϵ *follows the elliptically contoured distribution* $EC_r(0, \Sigma, g)$. *Suppose that the MLE of* β *under the standard model* (2), $\widehat{\beta}_{\mathrm{std}}$, *exists and* $\mathrm{vec}(\widehat{\beta}_{\mathrm{std}})$ *is* $\sqrt{n}$ *consistent and asymptotically normally distributed with asymptotic variance equal to the inverse of the Fisher information matrix* J_β. *We further assume that* $(X_i, Y_i),\ i = 1, \ldots, n$ *are independent and identical copies of* (X, Y). *Then* $\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{std}} - \beta)$ *is asymptotically normally distributed with mean zero and variance given by (8). If models* (4), (5) *or* (6) *hold, then* $\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}} - \beta)$, $\sqrt{n}\mathrm{vec}(\widehat{\beta}_E - \beta)$ *and* $\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}} - \beta)$ *are asymptotically normally* *distributed with mean zero and variance given by* (9)*,* (10) *and* (11) *respectively.*

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{std}})] = \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma, \quad (8)$$

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})] = \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes (\Sigma - M_A), \quad (9)$$

$$\begin{aligned}\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_E)] &= \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T \\ &+ \frac{1}{4}(\xi^T \otimes \Gamma_0)[\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\xi \otimes \Gamma_0^T), \quad (10)\end{aligned}$$

$$\begin{aligned}\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})] &= \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes [\Sigma - \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T] - \frac{1}{4}M_B \otimes \Gamma_0\Omega_0\Gamma_0^T \quad (11) \\ &+ \frac{1}{4}(B^T\eta^T \otimes \Gamma_0)[\eta B\widetilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\eta B \otimes \Gamma_0^T\end{aligned}$$

where $M_A = A(A^T\Sigma^{-1}A)^{-1}A^T$ *and* $M_B = B^T(B\widetilde{\Sigma}_X B^T)^{-1}B$*.*

Remark 1 *The asymptotic variance does not depend on the choices of* A*,* B*,* Γ*,* ξ *or* η*, since the values of the terms* $\xi^T\xi$*,* M_A*,* M_B*,* $\Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T$ *and* $B^T\eta^T\eta B$ *are unique.*

Remark 2 *Notice that* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})]$ *coincides with* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})]$ *when* $u = r$*, and* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})]$ *coincides with* $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_E)]$ *when* $d = \min(u, p)$*. This is consistent with the structure of the reduced rank envelope model: the reduced rank envelope model degenerates to the reduced rank regression when* $u = r$ *and to the envelope model when* $d = \min(u, p)$*.*

Now we compare the efficiency of the models. Since $\widetilde{\Sigma}_X^{-1} - M_B$ and $\Sigma - M_A$ are both semi-positive definite, $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{std}})] - \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})]$ is semi-positive definite. This implies the reduced rank regression estimator is more efficient than or as efficient as the standard estimator when the reduced rank regression model holds.

Next we prove that the envelope estimator is asymptotically at least as efficient as the standard estimator. Notice that $\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u}$ is semi-positive definite. Then

$$
\begin{aligned}
& (\xi^T \otimes \Gamma_0)[\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\xi \otimes \Gamma_0^T) \\
\leq \; & (\xi^T \otimes \Gamma_0)(\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1})^{-1}(\xi \otimes \Gamma_0^T) = \xi^T(\xi\widetilde{\Sigma}_X\xi^T)^{-1}\xi \otimes \Gamma_0\Omega_0\Gamma_0^T \leq \widetilde{\Sigma}_X^{-1} \otimes \Gamma_0\Omega_0\Gamma_0^T.
\end{aligned}
$$

Therefore

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

To compare the reduced rank envelope estimator and the reduced rank regression estimator, notice that

$$
\begin{aligned}
& 4\big\{\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})] - \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})]\big\} \\
= \; & M_B \otimes \Gamma_0\Omega_0\Gamma_0^T - (B^T\eta^T \otimes \Gamma_0)[\eta B\widetilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\eta B \otimes \Gamma \\
\geq \; & M_B \otimes \Gamma_0\Omega_0\Gamma_0^T - (B^T\eta^T \otimes \Gamma_0)(\eta B\widetilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1})^{-1}(\eta B \otimes \Gamma_0^T) \\
= \; & B^T(B\widetilde{\Sigma}_X B^T)^{-1}B \otimes \Gamma_0\Omega_0\Gamma_0^T - B^T\eta^T(\eta B\widetilde{\Sigma}_X B^T\eta^T)^{-1}\eta B \otimes \Gamma_0\Omega_0\Gamma_0^T \\
= \; & B^T(B\widetilde{\Sigma}_X B^T)^{-1/2}(I_d - P_{(B\widetilde{\Sigma}_X B^T)^{1/2}\eta^T})(B\widetilde{\Sigma}_X B^T)^{-1/2}B \otimes \Gamma_0\Omega_0\Gamma_0^T \geq 0.
\end{aligned}
$$

Therefore the reduced rank envelope estimator is more efficient than or as efficient as the reduced rank regression estimator. Finally, comparing the envelope estimator and the reduced rank envelope estimator, we have

$$
\begin{aligned}
& \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes [\Sigma - \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T] - \frac{1}{4}M_B \otimes \Gamma_0\Omega_0\Gamma_0^T \\
= \; & \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes [\Gamma\Omega\Gamma^T - \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T] \\
= \; & \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes (\Gamma\Omega\Gamma^T - \Gamma\Omega^{1/2}P_{\Omega^{-1/2}\eta}\Omega^{1/2}\Gamma^T),
\end{aligned}
$$

where $P_{\Omega^{-1/2}\eta}$ denotes the projection matrix onto the space spanned by the columns of $\Omega^{-1/2}\eta$. We have $\text{avar}[\sqrt{n}\text{vec}(\widehat{\beta}_E)] \geq \text{avar}[\sqrt{n}\text{vec}(\widehat{\beta}_{\text{RE}})]$ since $\widetilde{\Sigma}_X^{-1} - M_B$ and $\Gamma\Omega\Gamma^T - \Gamma\Omega^{1/2}P_{\Omega^{-1/2}\eta}\Omega^{1/2}\Gamma^T$ are both semi-positive definite. Therefore the reduced rank envelope model yields the most efficient estimator compared to all the other models.

6 Selections of rank and envelope dimension

For the reduce rank regression, we choose d using the same sequential test as in [7]. To test the null hypothesis $d = d_0$, the test statistic is $T(d_0) = (n - p - 1)\sum_{i=d_0+1}^{\min(p,r)} \lambda_i^2$, where λ_i is the ith largest eigenvalue of the matrix $\widehat{\Sigma}_X^{1/2}\widehat{\beta}_{\text{std}}^T\widehat{\Sigma}_{Y|X}^{-1/2}$, $\widehat{\Sigma}_X$ denotes the sample covariance matrix of X and $\widehat{\Sigma}_{Y|X}$ denotes the sample covariance matrix of the residuals from the OLS fit of Y on X. The reference distribution is a chi-squared distribution with degrees of freedom $(p - d_0)(r - d_0)$. We start with $d_0 = 0$, and increase d_0 if the null hypothesis is rejected. We choose the smallest d_0 that is not rejected. For the envelope model, we can apply information criterion such as AIC or BIC to select the dimension u. The information criterion requires the log likelihood function. We use the actual log likelihood if g is known. If g is unknown, we substitute the normal log likelihood with approximate weights

$$l_{u_0} = -\frac{1}{2}\sum_{i=1}^{n}\log|c_{X_i}\widehat{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}[Y_i - \bar{Y} - \widehat{\beta}(X_i - \bar{X})]^T\widehat{\Sigma}^{-1}[Y_i - \bar{Y} - \widehat{\beta}(X_i - \bar{X})],$$

where $\widehat{\beta}$ and $\widehat{\Sigma}$ are the envelope estimators obtained using the algorithm in Section 4.4 with $u = u_0$, $0 \leq u_0 \leq r$. Then u is chosen as the one that minimizes $-2l_u + kN(\delta)$, where $N(\delta)$ is the number of parameter of the envelope at dimension u (see Section 4.1) and k is the penalty which takes 2 in AIC and $\log(n)$ in BIC.

The reduced rank envelope model has two parameters d and u. We first choose d using the same sequential test as in the reduced rank regression. If d is chosen to be r, then we have $u = d = r$. If d is chosen to be $d_0 < r$, we then compute the information criterion, AIC or BIC, for $u = d_0, \cdots, r$, the same way as for the envelope model. We pick the u that minimizes the information criterion.

7 Simulations

In this section, we report results from the numerical experiments to compare the performance of the estimators derived under the elliptically contoured distribution, the estimators derived using the normal likelihood and the estimators derived using the approximate weights. The simulation in Section 7.1 is in the context of the envelope model and the simulation in Section 7.2 is in the context of the reduced rank envelope model. Sections 7.1 and 7.2 focus on estimation performance and Section 7.3 focuses on prediction performance. For simplicity, we call the envelope model derived in [8] as basic envelope model, and the reduced rank envelope model derived in [7] as basic reduced rank envelope model.

7.1 Envelope model

In this simulation, we investigate the estimation performance of our estimators in the context of envelope model. We fixed $p = 5$, $r = 20$ and $u = 4$. The predictors were generated independently from uniform $(0, 5)$ distribution, (Γ, Γ_0) was obtained by orthogonalizing an r by r matrix of independent uniform $(0, 1)$ variates, and elements in ξ were independent standard normal variates. The errors were generated from the multivariate t-distribution with mean 0, degrees of freedom 5 and $\Sigma = \sigma^2\Gamma\Gamma^T + \sigma_0^2\Gamma_0\Gamma_0^T$, where $\sigma = 2$ and $\sigma_0 = 5$. The intercept μ_Y was zero. The sample size was varied from 100, 200, 400, 800 and 1600. For each sample size, we generated 10000 replications. For each data set, we computed the OLS estimator, the basic envelope estimator, the envelope estimator with exact weights (the weights are computed from the true g) and the envelope estimator derived with the approximate weights. The estimation standard deviations of two randomly selected elements in β are displayed in Figure 1. In the left panel, the basic envelope model is more efficient than the OLS estimator. But the envelope estimators with exact weights and approximate weights achieve even more efficiency gains. The right panel indicates that the basic envelope model can be similar or even less efficient than the OLS estimator, while the envelope estimators with the exact weights or approximate weights are always more efficient than the OLS estimator. For example, at sample size 1600, the ratios of the estimation standard deviation of the OLS estimator versus that of the basic envelope estimator for all elements in β range from 0.800 to 2.701, with an average of 1.372. The ratios of the OLS estimator over the envelope estimator with exact weights range from 1.111 to 3.536, with an average of 1.823. If the approximate weights are used, the ratios of the estimation standard deviation of the OLS estimator over the envelope estimator range from 1.301 to 3.467, with an average of 1.903. The performance of the envelope estimator with approximate weights are very similar to the estimator with exact weights, as also demonstrated in Figure 1. Sometimes it can be even a little more efficient than the envelope estimator with exact weights since it is data adaptive, as indicated in the right panel. Figure 1 also confirms the asymptotic distribution derived in Section 5, and the envelope estimator with exact weights is $\sqrt{n}$-consistent. We have computed the bootstrap standard deviation for each estimator, and found that it is a good approximation to the actual estimation standard deviation. The results are not shown in the figure for better visibility.

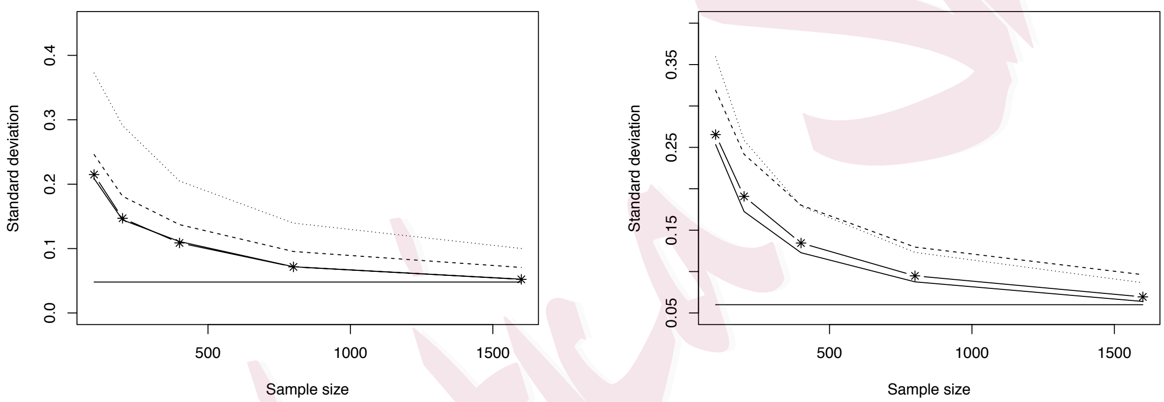

Figure 1: Estimation standard deviation versus sample size for two randomly selected elements in $\boldsymbol{\beta}$. Line — marks the envelope estimator with approximate weights, line $- * -$ marks the envelope estimator with exact weights, line - - marks the basic envelope estimator and line $\cdots$ marks the OLS estimator. The horizontal solid line at the bottom marks the asymptotic standard deviation of the envelope estimator with exact weights.

The average of absolute bias and MSE of the estimators in Figure 1 are included in Figure 2 and Figure 3 respectively. We notice that the estimation variance is the main component of the MSE. And the pattern of the MSE in Figure 3 is similar to the pattern of the estimation standard deviation in Figure 1.

The results in Figure 1 are based on known dimension of the envelope subspace. However, the dimension u is usually unknown in applications. Therefore we look into the performance of

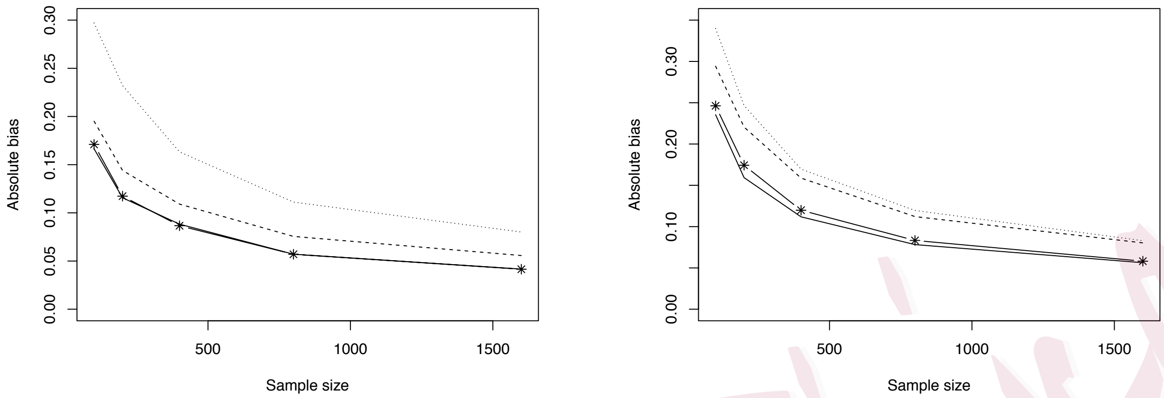

Figure 2: Average absolute bias versus sample size for two randomly selected elements in $\boldsymbol{\beta}$. The line types are the same as in Figure 1.

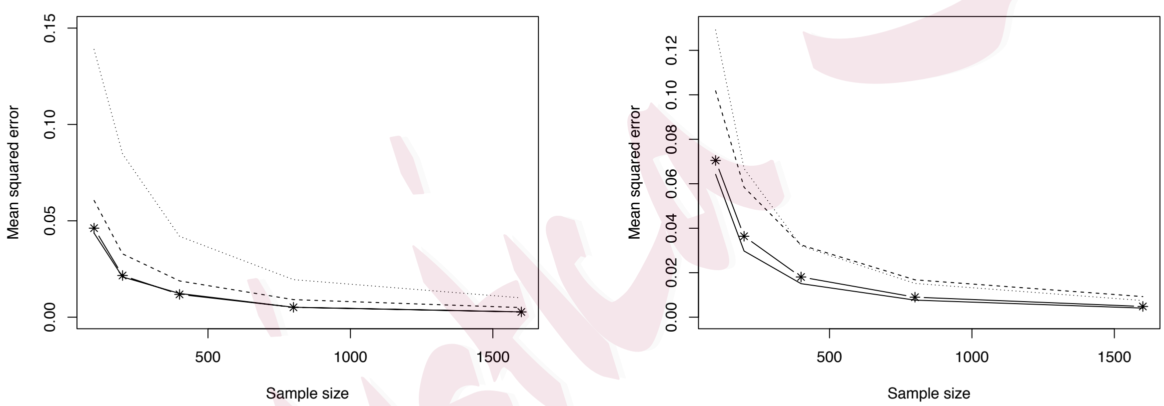

Figure 3: Average MSE versus sample size for two randomly selected elements in $\boldsymbol{\beta}$. The line types are the same as in Figure 1.

the dimension selection criteria discussed in Section 6. For the 200 replications, we computed the fraction that a criterion selects the true dimension. The results are summarized in Table 1. When AIC or BIC are not selecting the true dimension, we find that they always overestimate the dimension. This will cause a loss of some efficiency gains, but it does not introduce bias in estimation. When the exact weights are used, BIC is a consistent selection criterion. AIC is too conservative and selects a bigger dimension most of the time. When the approximate weights are

	Exact Weights		Approximate Weights	
	AIC	BIC	AIC	BIC
$n = 100$	14.4%	81.6%	10.0%	39.8%
$n = 200$	14.2%	90.2%	14.7%	26.0%
$n = 400$	13.9%	95.1%	21.3%	31.9%
$n = 800$	13.7%	96.3%	28.0%	36.4%
$n = 1600$	14.4%	98.1%	33.8%	42.6%

Table 1: Fraction of the time that selects the true dimension.

used, BIC tends to overestimate the dimension of the envelope subspace, but we can still achieve efficiency gains and have a smaller MSE than the standard model, as indicated in Figure 4. When exact weights are used, the estimation standard deviation and MSE of the envelope estimator are very close to those of the envelope estimator with known dimension, due to the consistency of BIC.

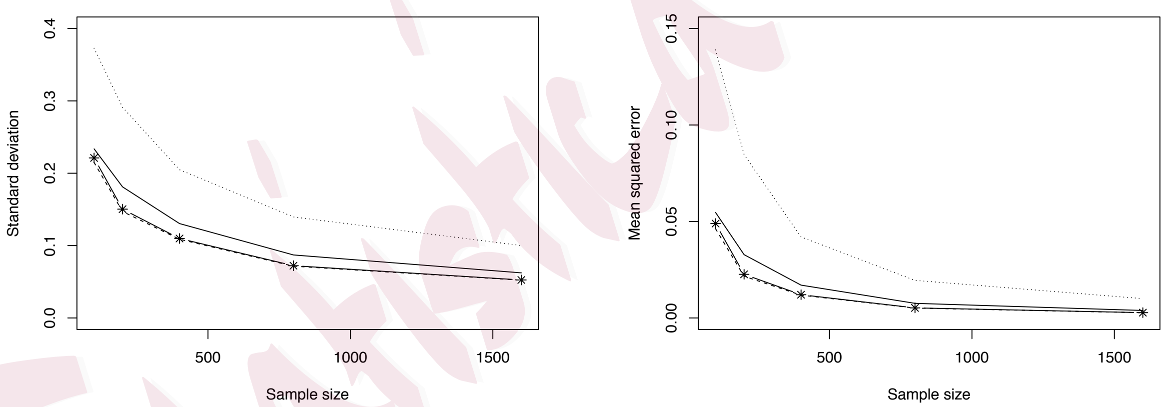

Figure 4: Estimation standard deviations and MSEs for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. Line $- * -$ marks the envelope estimator with dimension selected by BIC using exact weights, line — marks the envelope estimator with dimension selected by BIC using approximate weights, line - - marks the envelope estimator with known dimension and exact weights, and line $\cdots$ marks the OLS estimator.

We also investigate the performance of our estimators under normality. We repeated the simulations with the same settings except that the errors were generated from a multivariate normal distribution. The results are summarized in Figure 5. From the plot, we notice that the estimation standard deviations and MSEs of the basic envelope estimator and the envelope estimators with "exact" weights (i.e., the weights computed from t distribution) and approximate weights are almost indistinguishable. In this example, using the weights derived from t distribution or approximate weights does not cause a notable loss of efficiency in the normal case. This may be because that the approximate weights are computed from data, and therefore are data adaptive. For the "exact" weights, although it depends on the error distribution, it also has a data-dependent part (see Section 4.3). Therefore, these estimators do not lose much efficiency when the true distribution is normal. The performance of the dimension selection criteria is similar to that in Table 1, except that the BIC with "exact" weights selects the true dimension less frequently and the BIC with approximate weights selects the true dimension slightly more frequently.

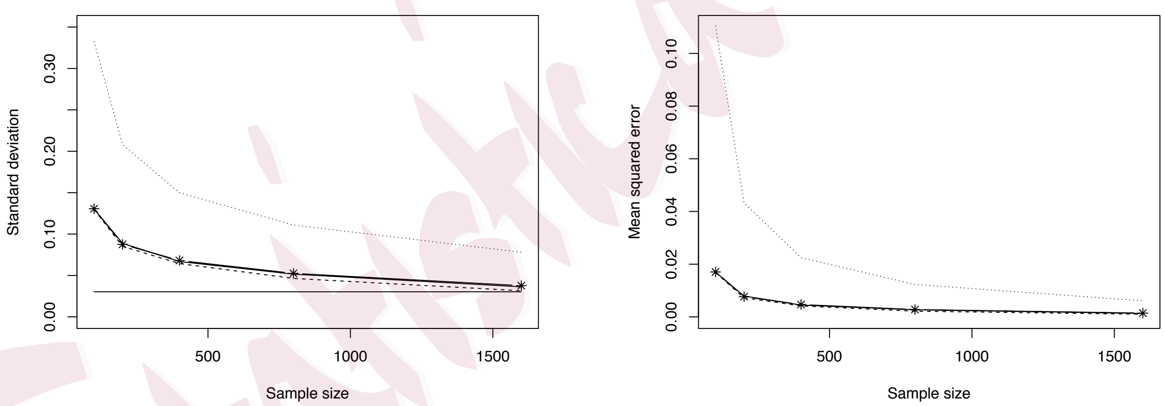

Figure 5: Estimation standard deviations and MSEs for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 1.

7.2 Reduced rank envelope model

This simulation studies the estimation performance of different estimators in the context of the reduced rank envelope model. We set $r = 10$, $p = 5$, $d = 2$ and $u = 3$. The matrix (Γ, Γ_0) was obtained by normalizing an r by r matrix of independent uniform $(0, 1)$ variates. The elements in η and B were standard normal variates, $\Sigma = \sigma^2 \Gamma A A^T \Gamma^T + \sigma_0^2 \Gamma_0 \Gamma_0^T$, where $\sigma = 0.4$, $\sigma_0 = 0.1$ and elements in A were $N(0, 1)$ variates. The elements in the predictor vector X are independent uniform $(0, 1)$ variates. The errors are generated from normal mixture distribution of two normal distributions $N(0, 2\Sigma)$ and $N(0, 0.1\Sigma)$ with probability 0.5 and 0.5. We varied the sample size from 100, 200, 400, 800 and 1600. For each sample size, we generated 10000 replications and computed the OLS estimator, the basic reduced rank envelope estimator and the reduced rank envelope estimator with exact weights (weights derived from the true error distribution) and approximate weights (Section 4.4). The estimation standard deviation for a randomly chosen element in β is displayed in the left panel of Figure 6. We notice that the basic reduced rank envelope estimator does not gain much efficiency compared to the OLS estimator. For example, with sample size 100, the standard deviation ratios of the OLS estimator versus the basic reduced rank envelope estimator range from 0.94 to 3.26 with an average of 1.42. The reduced rank envelope estimator computed from the exact weights obtains the most efficiency gains. When the sample size is 100, the ratios of the OLS estimator versus the reduced envelope estimator with exact weights range from 2.37 to 12.00 with an average of 3.98. This indicates that correctly specifying the structure of the error distribution will provide efficiency gains in estimation. However, we do not know the exact weights in practice. Figure 6 shows that the estimator computed from the approximate weights still provides substantial efficiency gains. The ratios of the OLS estimator versus the reduced envelope estimator with approximate weights range from 1.63 to 8.79 with an average of 2.77. Although the estimator with approximate weights is not as efficient as the estimator with exact weights, it is still more efficiency than the basic reduced rank envelope estimator or the OLS estimator. We also computed the bootstrap standard deviation of the estimators from 10000 residual bootstraps, and included the results in the right panel of Figure 6. The bootstrap standard deviation seems to be a good estimator of the actual estimation standard deviation. Therefore we compare the efficiency of different estimators using bootstrap standard deviations in applications.

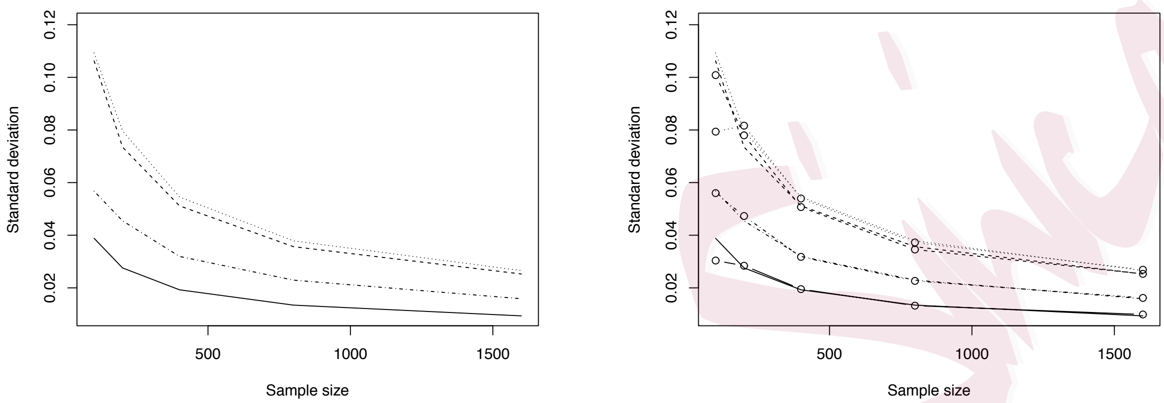

Figure 6: Estimation standard deviation and bootstrap standard deviation for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation only. Right panel: Estimation standard deviation with bootstrap standard deviation imposed. Line — marks the reduced rank envelope estimator with exact weights, line $- \cdot -$ marks the reduced rank envelope estimator with approximate weights, line - - marks the basic reduced rank envelope estimator and line $\cdots$ marks the OLS estimator. The lines with circles mark the bootstrap standard deviations for the corresponding estimator.

We investigated the bias and the MSE of the estimators. The results are summarized in Figure 7. Comparing the scale of the estimation standard deviation and the bias, we notice that for all estimators, the estimation standard deviation is the major component of MSE. Therefore the MSEs follow a similar trend as the estimation standard deviation. From the absolute bias plot, we notice that the OLS estimator and the basic reduced rank envelope estimator are more biased than the reduced rank envelope estimators with true and approximate weights. Figures 6 and 7 together show that we obtain a less biased and more efficiency estimator by considering the error distribution.

Now we look into the performance of the sequential test, AIC and BIC discussed in Section 6 in the selection of d and u. We used the same context as that generated Figures 6 and 7, and computed

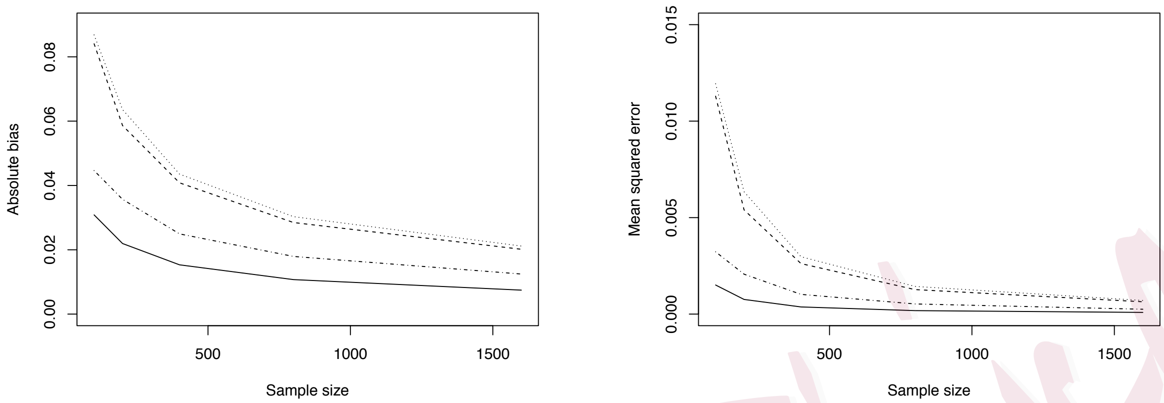

Figure 7: Average absolute bias and MSE for a randomly selected element in β. Left panel: Bias versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 6.

the fraction that a criterion selects the true dimension (out of 200 replications). The significance level for the sequential test was set at 0.01. The results are summarized in Table 2. The fraction that the sequential test chooses the true d approaches 99% as the sample size becomes large. When the exact weights are used, BIC performs better since it is a consistent selection criterion. AIC tends to be conservative and always selects a bigger dimension. When the approximate weights are used, AIC and BIC tend to overestimate the dimension of the envelope subspace. Overestimation causes a loss of efficiency, but it retains the useful information. Based on this result, we use BIC to choose u in applications. We compared the estimators with known and selected dimension as we did in Figure 4 of Section 7.1. The pattern is the same as in Figure 4, the reduced rank envelope estimator with dimension selected by BIC using approximate weights loses some efficiency compared to the estimator with known dimension and exact weights, but it is still notably more efficient than the estimator with the basic reduced rank envelope estimator.

We repeated the simulation with the same setting as in Figure 6, but the errors were generated from the multivariate normal distribution $N(0, 2\Sigma)$. The results are included in the Online

	Selection of d	Exact weights		Approximate weights	
	Sequential test	AIC	BIC	AIC	BIC
$n = 100$	96.3%	71.3%	97.1%	11.4%	12.2%
$n = 200$	98.1%	77.0%	99.0%	18.1%	19.0%
$n = 400$	98.6%	79.3%	99.7%	24.9%	26.6%
$n = 800$	98.9%	81.6%	99.8%	28.2%	31.1%
$n = 1600$	98.8%	83.5%	99.8%	26.0%	29.3%

Table 2: Fraction of the time that selects the true dimension.

Supplement.

7.3 Prediction

By modeling the error distribution, the efficiency gains in estimation often lead to improvements in prediction accuracy. In this section, we report the results of two numerical studies on prediction performance, one under the context of the envelope model and the other one under the context of the reduced rank envelope model.

We first generated the data from the envelope model (5). We set $p = 5$, $r = 5$, $u = 3$ and $n = 25$. The predictors were independent uniform $(0, 4)$ random variates. The coefficients had the structure $\beta = \Gamma\xi$, where elements in ξ were independent standard normal random variates and (Γ, Γ_0) were obtained by orthogonalizing an $r \times r$ matrix of uniform $(0, 1)$ variates. The errors were generated from the multivariate t-distribution with mean 0, degrees of freedom 5 and $\Sigma = \sigma^2\Gamma\Gamma^T + \sigma_0^2\Gamma_0\Gamma_0^T$, where $\sigma = 0.9$ and $\sigma_0 = 2$. We used 5-fold cross validation to evaluate the prediction error, and the experiment was repeated for 50 random splits. The prediction error was computed as $\sqrt{(Y - \hat{Y})^T(Y - \hat{Y})}$, where $\hat{Y}$ was the predicted value based on the estimators calculated form the training data. The average prediction error for 50 random splits were calculated for the OLS estimator, basic envelope estimator, the envelope estimator with exact weights and the envelope estimator with approximate weights. Results were summarized in Figure 8. The average prediction error for the OLS estimator is 8.34. We notice that the basic envelope estimator always has a larger prediction error than the OLS estimator for all u and its prediction error at $u = 3$ is 8.46. This indicates that by misspecifying the error distribution, we can also have a worse performance on prediction. The predictor error for the envelope estimator with exact weights achieves its minimum 7.49 at the $u = 3$. Compared to the OLS estimator, the envelope estimator with exact weights reduces the prediction error by 10.2%. The estimator with approximate weights achieves its minimum prediction error 7.21 at $u = 3$, which is a 14.8% reduction compare with the OLS estimator. In this example, the estimator with approximate weights gives a better prediction than the estimator with exact weights. This might be explained by the fact that we have a small sample size and the approximate weights are more adaptive to the data.

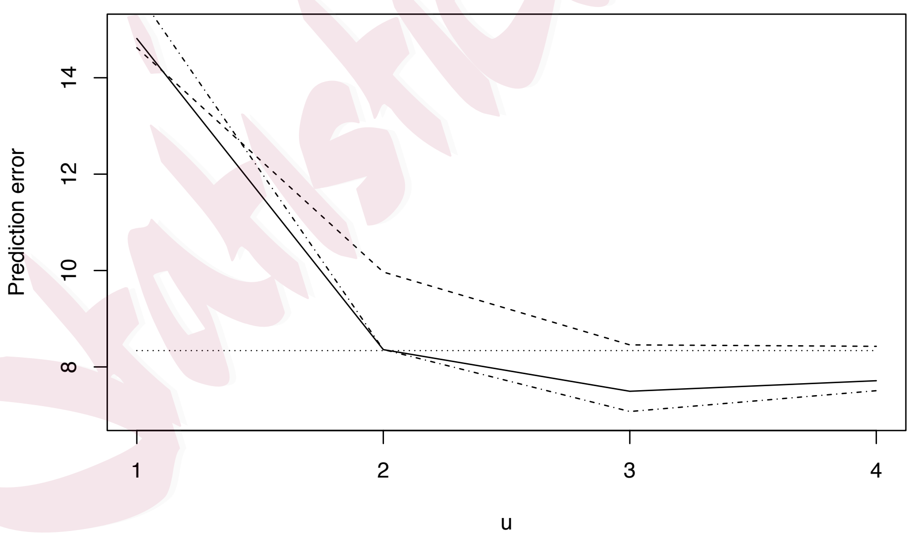

Figure 8: Prediction error versus u. Line — marks the envelope estimator with exact weights, line $- \cdot -$ marks the reduced rank envelope estimator with approximated weights, line - - marks the basic reduced rank envelope estimator and line $\cdots$ marks the OLS estimator.

In the second numerical study, data were simulated from the reduced rank envelope model (6). We set $p = 5$, $r = 10$, $d = 2$, $u = 3$ and $n = 30$. The predictors were independent uniform $(0, 1)$ random variates, and the errors were normal mixture random variates from two normal populations $N(0, 2\Sigma)$ and $N(0, 0.1\Sigma)$ with probability 0.5 and 0.5. Here Σ has the structure $\Sigma = \sigma^2\Gamma AA^T\Gamma^T + \sigma_0^2\Gamma_0\Gamma_0^T$, where $\sigma = 0.4$, $\sigma_0 = 0.1$ and elements in A were standard normal random variates. The regression coefficients β has the structure $\beta = \Gamma\eta B$, where elements in B and η were independent standard normal random variates, and (Γ, Γ_0) was obtained by normalizing an $r \times r$ matrix of independent uniform $(0, 1)$ random variates. We computed the prediction errors of the OLS estimator, basic reduced rank envelope estimator, the reduced rank envelope estimators with true and approximate weights for u from d to $r - 1$. The prediction errors were calculated based on 5-fold cross validation with 50 random splits of the data. The results are included in Figure 9. The prediction error of the OLS estimator is 1.35. The basic reduced rank envelope estimator achieves its minimum prediction error 1.20 at $u = 7$, although the prediction errors for $u \geq 3$ are all quite close. Compared to the OLS estimator, the basic reduced rank envelope estimator reduced the prediction error by 11.1%. The reduced rank envelope estimator with exact weights achieves it minimum prediction error 1.14 at $u = 6$, which is a 15.6% reduction compared to the OLS estimator. The reduced rank envelope estimator with approximate weights achieves it minimum prediction error 1.11 at $u = 5$, which is a 17.8% reduction compared to the OLS estimator. In this numerical study, although the basic envelope estimator shows better prediction performance than the OLS estimator; by taking the error distribution into account, we can further improve the prediction performance.

From the simulation results, it seems that when the true $g_{Y|X}$ is unknown, it is best to use the approximate weights.

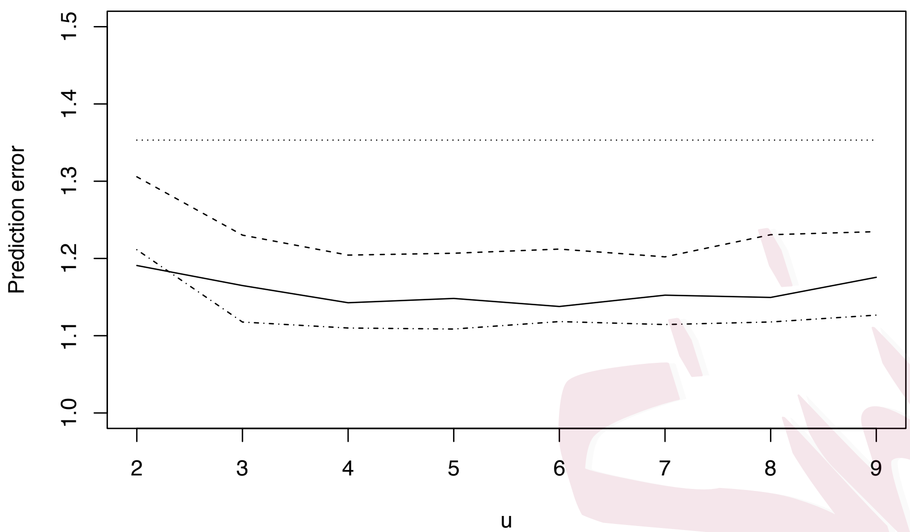

Figure 9: Prediction error versus u. Line — marks the envelope estimator with exact weights, line $-\cdot-$ marks the reduced rank envelope estimator with approximate weights, line - - marks the basic reduced rank envelope estimator and line $\cdots$ marks the OLS estimator.

8 Examples

8.1 The concrete slump test data

The slump flow of concrete depends on the components of the concrete. This dataset contains 103 records on various mix proportions [49], where the initial data set included 78 records and 25 new records were added later. The input variables are cement, fly ash, slag, water, super plasticizer, coarse aggregate and fine aggregate. They are ingredients of concrete and are measured in kilo per cubic meter concrete. The output variables are slump, flow and 28-day compressive strength. We use the first 78 records as training set and the new 25 records as testing set. The prediction error of the OLS estimator is 25.0. We fit the basic envelope model to the data, and BIC suggested $u = 2$. The bootstrap standard deviation ratios of the OLS estimator versus the basic envelope estimator range from 0.985 to 1.087 with an average of 1.028. This indicates that the basic envelope model does not yield much efficiency gains in this data. The prediction error for the basic envelope estimator is 24.2, which is quite close to that of the OLS estimator. From the discussion in Section 7, we find that when the error distribution is unknown, the approximate weights is adaptive to the data and gives good estimation and prediction results. We fit the data with the reduced rank envelope estimator with approximate weights. The sequential test selected $d = 2$ and BIC inferred $u = 2$. So the reduced rank envelope estimator degenerates to the envelope estimator with approximate weights. The bootstrap standard deviation ratios of the OLS estimator versus the envelope estimator with approximate weights range from 4.925 to 118.2 with an average of 55.57, which suggests a substantial efficiency gain. This is also confirmed by prediction performance: The prediction error is 12.27 for the envelope estimator with approximate weight. This is a 51% reduction compared to the prediction error of the OLS estimator, and a 49% reduction compared to the basic envelope estimator. This example shows that by considering the error structure of the data, we achieve efficiency gains and also obtain better prediction performance.

8.2 Vehicle data

The vehicle data contains measurements for various characteristics for 30 vehicles from different makers, e.g. Audi, Dodge, Honda, etc. The data is found in the R package *plsdepot* [44], and is used as an example to illustrate methods for partial least squares regression. Following [44], we use price in dollars, insurance risk rating, fuel consume (miles per gallon) in city and fuel consume in highway as responses. The predictors are indicators for turbo aspiration, vehicles with two doors and hatchback body-style, car length, width and height, curb weight, engine size, horsepower and peak revolutions per minute. This data set does not come with a natural testing set, so we used

5-fold cross validation with 50 random splits to evaluate the prediction performance. We scale the data so that all variables have unit standard deviation. This is because the range of the response variables are very different, for example, price in dollars ranges from 5348 to 37038 while the fuel consume in city ranges from 15 to 38. If the original scale is used, the prediction error is dominated by price in dollars. The prediction error for the OLS estimator is 1.70. Then we fit the reduced rank envelope estimator with approximate weights. The sequential test selected $d = 2$ and BIC suggested $u = 3$. The prediction error is 1.52, which is a 10.6% reduction compared to that of the OLS estimator. The basic reduced rank envelope estimator with $u = 3$ and $d = 2$ has prediction error 1.64, which is a 3.5% reduction compared to that of the OLS estimator. The bootstrap standard deviation ratios of the OLS estimator versus the basic reduced rank envelope estimator range from 0.919 to 1.844 with an average of 1.277. And the bootstrap standard deviation ratios of the OLS estimator versus the reduced rank envelope estimator with approximate weights range from 0.862 to 1.734 with an average of 1.289. In this case, the estimation standard deviation of the two estimators are similar. However, since the basic reduced rank envelope estimator has a larger bias due to misspecification of the error structure, the reduced rank envelope estimator with approximate weights gives a better prediction performance.

Acknowledgement

We are grateful to the associate editor and two referees for comments that helped us greatly improve the paper. This work was supported in part by grant DMS-1407460 from the US National Science Foundation.

Technical details

The proofs of the results stated in these paper as well as some additional simulation results are available online in the Online Supplement.

References

[1] Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22**, 327–351.

[2] Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions. *Ann. Statist.* **27**, 1141–1154.

[3] 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.

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

[5] Bura, E. and Forzani, L. (2015). Sufficient reductions in regressions with elliptically contoured inverse predictors. *J. Amer. Statist. Assoc.*, **110**, 420–434.

[6] Cook, R. D., Forzani, L. and Su, Z. (2016). A note on fast envelope estimation. *J. Multivariate Anal.*, **150**, 42–54.

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

[8] 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.

[9] Cook, R. D. and Nachtsheim, C. J. (1994). Reweighting to achieve elliptically contoured covariates in regression. *J. Amer. Statist. Assoc.*, **89**, 592–599.

[10] Cysneiros, F. J. A. and Paula, G. A. (2004). One-sided test in linear models with multivariate t-distribution. *Commun Stat Simul Comput.*, **33**, 747–771.

[11] del Pino, G. (1989). The unifying role of iterative generalized least squares in statistical algorithms. *Statistical Science*, **4(4)**, 394–408.

[12] Díaz-García, J. A., Galea Rojas, M. and Leiva-Sánchez, V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Comm. Statist. Theory Methods*, **32**, 625–641.

[13] Díaz-García, J. A and Gutiérrez-Jáimez, R. (2007). The distribution of residuals from a general elliptical linear model. *J. Statist. Plann. Inference*, **137**, 2347–2354.

[14] Dutta, S., & Genton, M. G. (2017). Depth-weighted robust multivariate regression with application to sparse data. *Canad. J. of Statist.*, **45(2)**, 164–184.

[15] Fang, K. and Anderson, T. W. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*. New York: Allerton Press.

[16] Fang, K., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. New York: Chapman and Hall.

[17] Fang, K. and Zhang, Y. (1990). *Generalized multivariate analysis*. Beijing: Science Press and Berlin: Springer-Verlag.

[18] Fernandez, C., and Steel, M. F. J. (1999). Multivariate student-t regression models: pitfalls and inference. *Biometrika*, **86**, 153–167.

[19] Gabriel F. (2004). *Generalized Elliptical Distributions: Theory and Applications.* PhD thesis, Universität zu Köln.

[20] Galea, M., Riquelme, M., and Paula, G. A. (2000). Diagnostic methods in elliptical linear regression models. *Braz. J. Prob. Stat.*, **14**, 167–184.

[21] García Ben, M., Martínez, E. and Yohai, V. J. (2006). Robust estimation for the multivariate linear model based on a τ-scale. *J. Multivariate Anal.*, **97**, 1600–1622.

[22] Gómez, E., Gómez-Villegas, M. A. and Marín, J. M. (2003). A survey on continuous elliptical vector distributions. *Rev. Mat. Complut*, **16**, 345–361.

[23] Green, P. J. (1984), Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives, *Journal of the Royal Statistical Society. Series B (Methodological)*, **46(2)**, 149–192.

[24] Gupta, A. K., Varga, T. and Bodnar, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*. New York: Springer.

[25] Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canad. J. of Statist.*, **7**, 65–81.

[26] Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *J. Multivariate Anal.*, **5**, 248–264.

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

[28] Kowalski, J., Mendoza-Blanco, J. R., Tu, X. M. and Gleser, L. J. (1999). On the difference in inference and prediction between the joint and independent f-error models for seemingly unrelated regressions. *Comm. Statist. Theory Methods*, **28**, 2119–2140.

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

[30] Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t-distribution. *J. Amer. Statist. Assoc.*, **84**, 881–896.

[31] Lee, M. and Su, Z. (2018). *Renvlp: Computing Envelope Estimators*. R package version 2.5.

[32] Lemonte, A. J. and Patriota, A. G. (2011). Multivariate elliptical models with general parameterization. *Statistical Methodology*, **8**, 389–400.

[33] Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers*, **41**, 211–224.

[34] Liu, S. (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra Appl* ., **354**, 159–174.

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

[36] Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods.* Chichester: John Wiley & Sons, Ltd.

[37] Melo, T. F. N., Ferrari, S. L. P. and Patriota, A. G. (2017) Improved hypothesis testing in a general multivariate elliptical model. *J. Stat. Comput. Simul.* **87(7)**, 1416–1428.

[38] Melo, T. F. N., Ferrari, S. L. P. and Patriota, A. G. (2018) Improved estimation in a general multivariate elliptical model. *Braz. J. Probab. Stat.* **32(1)**, 44–68.

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

[40] Osorio, F., Paula, G.A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Comput. Statist. Data Anal.*, **51**, 4354–4368.

[41] Reinsel, G. C. and Velu, R. P. (1998). *Multivariate reduced rank Regression: Theory and Applications*. New York: Springer.

[42] Rousseeuw, P. J., Van Aelst, S., Driessen, K. V. and Gulló, J. A. (2004). Robust multivariate regression. *Technometrics*, **46**, 293–305.

[43] Russo, C. M., Paula, G. A. and Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Comput. Statist. Data Anal.*, **53**, 4143–4156.

[44] Sanchez, G. (2016). *plsdepot: Partial Least Squares (PLS) Data Analysis Methods*. R package version 0.1.17.

[45] Savalli, C., Paula, G. A. and Cysneiros, F. J. A. (2006). Assessment of variance components in elliptical linear mixed models. *Stat. Modelling*, **6**, 59–76.

[46] Stoica, P. and Viberg, M. (1996). Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. *IEEE Trans. Signal Process.* **44**, 3069–3078.

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

[48] Welsh, A. T. and Richardson, A. M. (1997). Approaches to the robust estimation of mixed models. *Handbook of Statistics*, **15**, 343–384.

[49] Yeh, I. C. (2007). Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cement and Concrete Composites*, **29(6)**, 474–480.

[50] Zhao, W., Lian, H. and Ma, S. (2017). Robust reduced rank modeling via rank regression. *J. Statist. Plann. Inference*, **180**, 1–12.