

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, Gainesville, 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

$$Y = \mu_Y + \beta(X - \mu_X) + \epsilon, \quad (1)$$

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

33 of the reduced rank envelope model takes the normal likelihood function as the objective function,
34 and is most effective when the normality assumption holds.

35 It is well-known that the normality assumption is not always reasonable in many applications,
36 and alternative distributions (or methodologies) have to be considered to fit the distribution of the
37 data better. One choice is the family of elliptically contoured distributions, which includes the
38 classical normal distribution and many important distributions such as Student- t , power exponen-
39 tial, contaminated normal, etc. They can have heavier or lighter tails than the normal distribution,
40 and are more adaptive to the data. Elliptical multivariate linear regression models have been ex-
41 tensively studied in the statistical literature, see for example [5], [10], [12], [13], [18], [19],[22],
42 [20], [28], [30], [32], [33], [34], [40], [43], [45] and [48] among others. In particular, [32] in-
43 troduces a general elliptical multivariate regression model in which the mean vector and the scale
44 matrix have parameters in common. Then they unify several elliptical models, such as nonlinear
45 regressions, mixed-effects model with nonlinear fixed effects, errors-in-variables models, etc. Bias
46 correction for the maximum likelihood estimator and adjustments of the likelihood-ratio statistics
47 are also derived for this general model (see [38], [37]). The elliptical distributions can also be used
48 as the basis to consider robustness in multivariate linear regression, as in [14], [21], [29], [36],
49 [42], [50] and many others. Nevertheless the envelope model under the context of elliptical multi-
50 variate regression has not yet been implemented. There is also not much literature about reduced
51 rank regression beyond the normal case. The only attempt to extend reduced rank regression to
52 the non-normal case is through M-estimators or other robust estimators which include some of
53 the elliptical class. For example, [50] develops a robust estimator in reduced rank regression and
54 proposes a novel rank-based estimation procedure using Wilcoxon scores. While the reduced rank
55 estimator in [50] allows a general error distribution, we aim to further improve the efficiency of the

56 reduced rank estimator using maximum likelihood estimators (MLE) and envelope methods in the
57 context of elliptical multivariate linear regression.

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

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

75 2 Models

76 We consider the following elliptical multivariate linear regression model given by

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

77 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
 78 \sim denotes equal in distribution. If a random vector $Z \in \mathbb{R}^m$ follows an elliptically contoured
 79 distribution $EC_m(\mu_Z, \Sigma_Z, g_Z)$ with density, then the density function is given by

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

80 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; $g_Z(\cdot) \geq$
 81 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
 82 the following discussions. Based on (2), $Y | X$ follows the elliptically contoured distribution
 83 $EC_r(\mu_{Y|X}, \Sigma, g_{Y|X})$ where $\mu_{Y|X} = \mu_Y + \beta(X - \mu_X)$. When the conditional expectation and
 84 variance exist, $E(Y | X) = \mu_{Y|X}$ and $\text{var}(Y | X) = c_X \Sigma$, where $c_X = E(Q^2)/r$ and $Q^2 =$
 85 $(Y - \mu_{Y|X})^T \Sigma^{-1} (Y - \mu_{Y|X})$ (see Corollary 2 in [17], p. 65). Notice that in general $\text{var}(Y | X)$
 86 depends on X , except for the normal errors with constant variance.

87 The reduced rank regression assumes that the rank of the coefficient β in model (2) is at most
 88 $d \leq \min(p, r)$. As a consequence

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

89 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 $AB =$
90 $(AU)(U^{-1}B) := A^*B^*$ for any invertible U . If the errors are normally distributed with a con-
91 stant covariance matrix, the MLE of β for (4) and its asymptotic distribution were derived in [2],
92 [41] and [46], by imposing various constraints on A and B for identifiability. Since the goal is to
93 estimate β , rather than A and/or B , recently [7] derived the estimator of β without imposing any
94 constraints on A and B other than requiring the rank of β is equal to d . It has been shown that when
95 ϵ follows the multivariate normal distribution with a constant covariance matrix, the reduced rank
96 regression has the potential to yield more efficient estimator for β than the ordinary least square
97 (OLS) estimator. Note that under normality, the OLS estimator is the MLE.

98 The envelope model [8] is another way to get efficient estimator. Let $\text{span}(\beta)$ denote the
99 subspace spanned by the columns of β . Under model (2), if $\text{span}(\beta)$ is contained in the span of m
100 ($m < r$) eigenvectors of the error covariance matrix Σ , not necessarily the leading eigenvectors,
101 then the envelope estimator of β is expected to be more efficient than the OLS estimator. More
102 specifically, let \mathcal{S} be a subspace of \mathbb{R}^r that is spanned by some eigenvectors of Σ , and $\text{span}(\beta) \subseteq \mathcal{S}$.
103 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
104 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
105 $\Gamma_0 \in \mathbb{R}^{r \times (r-u)}$ to be a completion of Γ , i.e., (Γ, Γ_0) is an orthogonal matrix. Since $\text{span}(\beta) \subseteq$
106 $\mathcal{E}_\Sigma(\beta) = \text{span}(\Gamma)$, there exists a $\xi \in \mathbb{R}^{u \times p}$ such that $\beta = \Gamma\xi$. Because that $\mathcal{E}_\Sigma(\beta)$ is spanned by
107 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
108 carries the coordinates of β with respect to Γ and Ω and Ω_0 carry the coordinates of Σ with respect
109 to Γ and Γ_0 . To summarize, if β and Γ satisfy the following conditions

$$\beta = \Gamma\xi \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \quad (5)$$

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

117 Under normal distribution for the error term, [7] presented a novel unified framework of the
 118 reduced rank regression and the envelope model called the reduced rank envelope model, which
 119 obtains more efficient estimators compared to either of them. [7] assumed that β and Σ follows the
 120 envelope structure (5) and at the same time the coordinate ξ has a reduced rank structure $\xi = \eta B$,
 121 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
 122 rank envelope model when

$$\beta = AB = \Gamma\eta B, \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \quad (6)$$

123 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
 124 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
 125 rank envelope model performs dimension reduction in two levels: The first level $\beta = AB$ assumed
 126 that we have a reduced rank regression. The second level $\beta = \Gamma\eta B$ is based on the assumption
 127 that β only intersects u eigenvectors of the covariance matrix Σ . When $u = r$, $\Gamma = I_r$, then (6)
 128 degenerates to the usual reduced rank regression (4). When $d = \min(u, p)$, then (6) reduces to
 129 an envelope model (5). Finally, the reduced rank envelope model (6) is equivalent to the standard

130 model (2) if $d = u = r$. [7] obtained the MLEs of β and Σ , as well as their asymptotic distribu-
 131 tions under the normality assumption. It should be noted that for the reduced rank regression, the
 132 envelope model and the reduced rank envelope model, the constituent parameters $A, B, \Gamma, \Gamma_0, \xi,$
 133 η, Ω, Ω_0 are not unique. Hence, they are not identifiable. Nevertheless, β and Σ are unique. No ad-
 134 ditional constraints are imposed on the constituent parameters in [7] when studying the asymptotic
 135 distribution of the identifiable parameters β and Σ .

136 3 Examples

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

139 3.1 The data matrix is elliptically contoured distributed

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

145 Let $\mathbb{X} = (X_1^T, \dots, X_n^T)^T \in \mathbb{R}^{n \times p}$ and $\mathbb{Y} = (Y_1^T, \dots, Y_n^T)^T \in \mathbb{R}^{n \times r}$ be data matrices such that
 146 $\mathbb{Y} \mid \mathbb{X}$ follows a matrix elliptically contoured distribution $EC_{n,r}(M, \eta \otimes \Sigma, g)$ with $M = 1_n \mu_Y^T +$
 147 $(\mathbb{X} - 1_n \mu_X^T) \beta^T$, where 1_n denotes an n dimensional vector of 1's. Under this assumption, by using
 148 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
 149 the errors to be modeled with a heteroscedastic structure. More properties of the matrix elliptically

150 contoured distribution are discussed in [24]. Examples of this distribution include matrix variate
 151 symmetric Kotz Type distribution, Pearson Type II distribution, Pearson Type VII distribution,
 152 symmetric Bessel distribution, symmetric Logistic distribution, symmetric stable law, etc. Among
 153 these distributions, the most common one is the normal with non-constant variance [16].

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

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

157 where $\epsilon \sim N(0, \eta_{ii} \Sigma)$. Therefore $Y_i \mid X_i$ follows a normal distribution with mean $\mu_Y + \beta(X_i - \mu_X)$
 158 and covariance matrix $\eta_{ii} \Sigma$. In other words, it is an elliptically contoured distribution $EC_r(\mu_Y +$
 159 $\beta(X_i - \mu_X), \Sigma, g_i)$ with $g_i(t) = (2\pi\eta_{ii})^{-r/2} e^{-\frac{t}{2\eta_{ii}}}$.

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

164 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 trans-
 formed 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

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

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

166 Φ as

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

167 then X and Y are marginally elliptically contoured distributed where X follows $EC_p(\mu_X, \Sigma_X, g)$

168 and Y follows $EC_r(\mu_Y, \Sigma_Y, g)$ (Theorem 2.8 from [24]). The conditional distribution of $Y | X$ is

169 also elliptically contoured

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

170 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 $g_{Y|X}(t) = g(t +$

171 $m(X))/g(m(X))$ with $m(X) = (X - \mu_X)^T \Phi_{11}^{-1}(X - \mu_X)$. Note that $\mu_{Y|X}$ is linear in X and

172 $\Phi_{22.1}$ is a constant.

173 Now we use multivariate t -distribution as an example. Suppose that $Z \in R^k$ follows a multi-

174 variate t -distribution $t_k(\mu, \Sigma, \nu)$, where ν denotes the degrees of freedom. The density function of

175 Z is given by

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

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

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

176 Equivalently, $Y | X \sim EC_r(\mu_{Y|X}, \Phi_{22.1}, g_{Y|X})$ with $\mu_{Y|X} = \mu_Y + \Phi_{21}\Phi_{11}^{-1}(X - \mu_X)$, $\Phi_{22.1} =$
 177 $\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{21}^T$ and $g_{Y|X}(t) = c_{\nu,p,r}g[t/h(X)]/h(X)^{r/2}$, where $g(t) = (\nu + p + t)^{-\frac{\nu+r+\nu}{2}}$,
 178 $h(X) = [\nu + (X - \mu_X)^T\Phi_{11}^{-1}(X - \mu_X)]/(\nu + p)$ and $c_{\nu,p,r}$ is the normalizing constant.

179 3.3 Y given X follows an elliptically contoured distribution

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

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

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

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

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

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

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

192 Equivalently, $Y | X$ follows an elliptically contoured distribution $EC_r(\mu_Y + \beta(X - \mu_X), \Sigma, g)$

193 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.

194 4 Estimation

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

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

205 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 $h = (\text{vec}^T(\beta), \text{vech}^T(\Sigma))^T$. 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

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

206 We use $N(v)$ to denote the number of parameters in a parameter vector v . Then $N(h) = pr +$
 207 $r(r + 1)/2$, $N(\psi) = (r - d)d + pd + r(r + 1)/2$, $N(\delta) = pu + r(r + 1)/2$ and $N(\phi) =$
 208 $(u - d)d + pd + r(r + 1)/2$. The reduced rank regression has less parameters than the standard
 209 model since $N(h) - N(\psi) = (p - d)(r - d) \geq 0$, and the reduced rank envelope model has even
 210 less parameters than the reduced rank regression as $N(\psi) - N(\phi) = (r - u)d \geq 0$. On the other
 211 hand, compared to the standard model, the number of parameters is reduced by $p(r - u) \geq 0$ by
 212 using the envelope model and it is further reduced by $(p - d)(u - d) \geq 0$ by using the reduced
 213 rank envelope model.

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

219 **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) , $i = 1, \dots, n$, 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

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

220 where $W_i = -2g'(m_i)/g(m_i)$. If $Y \mid X$ followed a normal distribution, the log-likelihood function
 221 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

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

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

226 1. Get the initial values for β and Σ

227 2. Repeat the following until convergence

228 (a) Compute $W_i = -2g'(m_i)/g(m_i)$ with β and Σ being the current estimator

229 (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
230 normal distribution.

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

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

235 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
236 can estimate the reduced rank envelope estimator using the preceding algorithm except that 2(b)
237 is changed to “With the data $(\sqrt{W_i}X_i, \sqrt{W_i}Y_i)$, update the reduced rank envelope estimators of
238 β and Σ as if the data follow normal distribution.” The reduced rank regression and the envelope
239 model can follow the same procedure. For completeness we include in the Online Supplement the
240 derivatives of m_i with respect to the parameters β and Σ .

241 4.3 Weights

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

243 Normal with non-constant variance

244 If $Y_i | X_i$ follows the normal distribution $N(\mu_{Y|X}, \eta_{ii}\Sigma)$ with $\eta_{ii} > 0, i = 1, \dots, n$. Then $W_i =$
 245 $1/\eta_{ii}$.

246 **Normal mixture distribution**

247 Suppose $Y | X$ follows a normal mixture distribution from m normal distributions $N_r(\mu_{Y|X}, k_i\Sigma),$
 248 $i = 1, \dots, m$ with weights p_1, \dots, p_m . From the discussions in Section 3.3, the weights are given
 249 by

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

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

251 **Multivariate t -distribution**

252 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
 253 following the structure in (7). Based on the discussion in Section 3.2, $Y | X$ follows the t -
 254 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
 255 calculations,

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

256 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 $\Sigma_{Y|X} =$
 257 $\frac{\nu+(x-\mu_X)^T\Phi_{11}^{-1}(x-\mu_X)}{\nu+p}\Phi_{22.1}$.

258 **Conditional t -distribution**

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

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

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

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

265 4.4 Weighted least square estimators

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

268 Suppose that the model has the structure in (2), then we have $\text{var}(Y | X) = c_X \Sigma$, where
 269 $c_X = E(Q^2)/r$ and $Q^2 = [Y - \mu_Y - \beta(X - \mu_X)]^T \Sigma^{-1} [Y - \mu_Y - \beta(X - \mu_X)]$ (see Corollary 2
 270 in [17]). Notice that c_X can be different across the observations. We use c_{X_i} to denote c_X for the
 271 i th 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
 272 the parameters as if the data follows the normal distribution. If c_{X_i} is unknown, we estimate it by
 273 $\hat{c}_{X_i} = [Y_i - \hat{\mu}_Y - \hat{\beta}(X_i - \hat{\mu}_X)]^T \hat{\Sigma}^{-1} [Y_i - \hat{\mu}_Y - \hat{\beta}(X_i - \hat{\mu}_X)]$. According to [5], the resulting
 274 estimators of β and Σ are robust to a moderate departure from normality. Let \bar{X} and \bar{Y} denote the
 275 sample mean of X and Y . The following algorithm summarized the preceding discussion.

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

279 2. Repeat the following until convergence

280 (a) Compute $\hat{c}_{X_i} = [Y_i - \hat{\mu}_Y - \hat{\beta}(X_i - \hat{\mu}_X)]^T \hat{\Sigma}^{-1} [Y_i - \hat{\mu}_Y - \hat{\beta}(X_i - \hat{\mu}_X)]$, where $\hat{\mu}_X, \hat{\mu}_Y,$
 281 $\hat{\beta}$ and $\hat{\Sigma}$ are the estimates of μ_X, μ_Y, β and Σ .

282 (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
 283 corresponding model as if the data follows normal distribution.

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

288 5 Asymptotics

289 In this section, we present the asymptotic distribution for the MLEs of β : the standard estimator
 290 $\hat{\beta}_{\text{std}}$, the reduced rank regression estimator $\hat{\beta}_{\text{RR}}$, the envelope model estimator $\hat{\beta}_E$ and reduced rank
 291 envelope estimator $\hat{\beta}_{\text{RE}}$.

292 Without loss of generality, we assume that $\mu_X = 0$ and the predictors are centered in the
 293 sample. Let C_r and E_r denote the contraction and expansion matrix defined in [25] that connects
 294 the vector operator vec and the vector half operator vech as follows $\text{vec}(S) = E_r \text{vech}(S)$ and
 295 $\text{vech}(S) = C_r \text{vec}(S)$ for any $r \times r$ symmetric matrix S . Let $U = \Sigma^{-1/2}[Y - \mu_Y - \beta(X -$
 296 $\mu_X)]$, $N_X = E \left[\left(\frac{g'(U^T U)}{g(U^T U)} \right)^2 U^T U \middle| X \right] / r$ and $M_X = E \left[\left(\frac{g'(U^T U)}{g(U^T U)} \right)^2 (U^T U)^2 \middle| X \right] / [r(r+2)]$. We
 297 define $\tilde{\Sigma}_X = E(N_X X X^T)$ and $M = E(M_X)$ if X is random and the expectations exist, $\tilde{\Sigma}_X =$
 298 $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n N_{X_i} X_i X_i^T$ and $M = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n M_{X_i}$ if X is fixed when the limits are finite.
 299 We further assume that $\tilde{\Sigma}_X$ is positive definite and $M > 0$. For the rest of the section we ask g such
 300 that the above quantities are finites and that the maximum likelihood estimator for model (2) exists,
 301 is consistent and asymptotically normal (See for example the conditions for elliptical distributions
 302 on [39], [4], [3], [27], [29], [35], [39], [50] or more generally conditions on Theorems 5.23, 5.31,

303 5.39, 5.41 or 5.42 from [47]. Then, the Fisher information for $h = (\text{vec}^T(\beta), \text{vech}^T(\Sigma))^T$ is given

304 by

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

305 with $J_\beta = 4\tilde{\Sigma}_X \otimes \Sigma^{-1}$ and $J_\Sigma = 2ME_r^T(\Sigma^{-1} \otimes \Sigma^{-1})E_r + (M - \frac{1}{4})E_r^T \text{vec}(\Sigma^{-1})\text{vec}^T(\Sigma^{-1})E_r$.

306 Detailed calculations are included in the Online Supplement. When ϵ follows a normal distribution,

307 we have $N_X = M = 1/4$ and J_h has the same form as in the literature (e.g. [8]).

308 Proposition 1 gives the asymptotic variance of the MLEs of β under the standard model (2),

309 the reduced rank regression (4), the envelope model (5) and the reduced rank envelope model (6).

310 Suppose that $\hat{\theta}$ is an estimator of θ . We write $\text{avar}(\sqrt{n}\hat{\theta}) = V$ if $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$, where

311 \xrightarrow{d} denotes convergence in distribution.

312 **Proposition 1** *Suppose that model (2) holds, i.e. the error vector ϵ follows the elliptically con-*

313 *toured distribution $EC_r(0, \Sigma, g)$. Suppose that the MLE of β under the standard model (2),*

314 *$\hat{\beta}_{\text{std}}$, exists and $\text{vec}(\hat{\beta}_{\text{std}})$ is \sqrt{n} consistent and asymptotically normally distributed with asymp-*

315 *totic variance equal to the inverse of the Fisher information matrix J_β . We further assume that*

316 *(X_i, Y_i) , $i = 1, \dots, n$ are independent and identical copies of (X, Y) . Then $\sqrt{n}\text{vec}(\hat{\beta}_{\text{std}} - \beta)$ is*

317 *asymptotically normally distributed with mean zero and variance given by (8). If models (4), (5) or*

318 *(6) hold, then $\sqrt{n}\text{vec}(\hat{\beta}_{\text{RR}} - \beta)$, $\sqrt{n}\text{vec}(\hat{\beta}_E - \beta)$ and $\sqrt{n}\text{vec}(\hat{\beta}_{\text{RE}} - \beta)$ are asymptotically normally*

319 *distributed with mean zero and variance given by (9), (10) and (11) respectively.*

$$\text{avar}[\sqrt{n}\text{vec}(\hat{\beta}_{\text{std}})] = \frac{1}{4}\tilde{\Sigma}_X^{-1} \otimes \Sigma, \quad (9)$$

$$\text{avar}[\sqrt{n}\text{vec}(\hat{\beta}_{\text{RR}})] = \frac{1}{4}\tilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\tilde{\Sigma}_X^{-1} - M_B) \otimes (\Sigma - M_A), \quad (10)$$

$$\begin{aligned} \text{avar}[\sqrt{n}\text{vec}(\hat{\beta}_E)] &= \frac{1}{4}\tilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T \\ &\quad + \frac{1}{4}(\xi^T \otimes \Gamma_0)[\xi\tilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\xi \otimes \Gamma_0^T), \end{aligned} \quad (11)$$

$$\begin{aligned} \text{avar}[\sqrt{n}\text{vec}(\hat{\beta}_{\text{RE}})] &= \frac{1}{4}\tilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\tilde{\Sigma}_X^{-1} - M_B) \otimes [\Sigma - \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T] - \frac{1}{4}M_B \otimes \Gamma_0\Omega_0\Gamma_0^T \\ &\quad + \frac{1}{4}(B^T\eta^T \otimes \Gamma_0)[\eta B\tilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\eta B \otimes \Gamma_0^T) \end{aligned} \quad (12)$$

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

321 **Remark 1** *The asymptotic variance does not depend on the choices of A , B , Γ , ξ or η , since the*
 322 *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.*

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

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

331 Next we prove that the envelope estimator is asymptotically at least as efficient as the standard

332 estimator. Notice that $\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u}$ is semi-positive definite. Then

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

333 Therefore

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

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

335 that

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

336 Therefore the reduced rank envelope estimator is more efficient than or as efficient as the reduced

337 rank regression estimator. Finally, comparing the envelope estimator and the reduced rank envelope

338 estimator, we have

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

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

343 6 Selections of rank and envelope dimension

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

353 approximate weights

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

354 where $\widehat{\beta}$ and $\widehat{\Sigma}$ are the envelope estimators obtained using the algorithm in Section 4.4 with $u = u_0$,
355 $0 \leq u_0 \leq r$. Then u is chosen as the one that minimizes $-2l_u + kN(\delta)$, where $N(\delta)$ is the number
356 of parameter of the envelope at dimension u (see Section 4.1) and k is the penalty which takes 2 in
357 AIC and $\log(n)$ in BIC.

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

362 7 Simulations

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

371 7.1 Envelope model

372 In this simulation, we investigate the estimation performance of our estimators in the context of
373 envelope model. We fixed $p = 5$, $r = 20$ and $u = 4$. The predictors were generated independently
374 from uniform $(0, 5)$ distribution, (Γ, Γ_0) was obtained by orthogonalizing an r by r matrix of
375 independent uniform $(0, 1)$ variates, and elements in ξ were independent standard normal variates.
376 The errors were generated from the multivariate t -distribution with mean 0, degrees of freedom 5
377 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
378 size was varied from 100, 200, 400, 800 and 1600. For each sample size, we generated 10000
379 replications. For each data set, we computed the OLS estimator, the basic envelope estimator,
380 the envelope estimator with exact weights (the weights are computed from the true g) and the
381 envelope estimator derived with the approximate weights. The estimation standard deviations of
382 two randomly selected elements in β are displayed in Figure 1. In the left panel, the basic envelope
383 model is more efficient than the OLS estimator. But the envelope estimators with exact weights and
384 approximate weights achieve even more efficiency gains. The right panel indicates that the basic
385 envelope model can be similar or even less efficient than the OLS estimator, while the envelope
386 estimators with the exact weights or approximate weights are always more efficient than the OLS
387 estimator. For example, at sample size 1600, the ratios of the estimation standard deviation of the
388 OLS estimator versus that of the basic envelope estimator for all elements in β range from 0.800 to
389 2.701, with an average of 1.372. The ratios of the OLS estimator over the envelope estimator with
390 exact weights range from 1.111 to 3.536, with an average of 1.823. If the approximate weights
391 are used, the ratios of the estimation standard deviation of the OLS estimator over the envelope
392 estimator range from 1.301 to 3.467, with an average of 1.903. The performance of the envelope

393 estimator with approximate weights are very similar to the estimator with exact weights, as also
 394 demonstrated in Figure 1. Sometimes it can be even a little more efficient than the envelope
 395 estimator with exact weights since it is data adaptive, as indicated in the right panel. Figure 1 also
 396 confirms the asymptotic distribution derived in Section 5, and the envelope estimator with exact
 397 weights is \sqrt{n} -consistent. We have computed the bootstrap standard deviation for each estimator,
 398 and found that it is a good approximation to the actual estimation standard deviation. The results
 399 are not shown in the figure for better visibility.

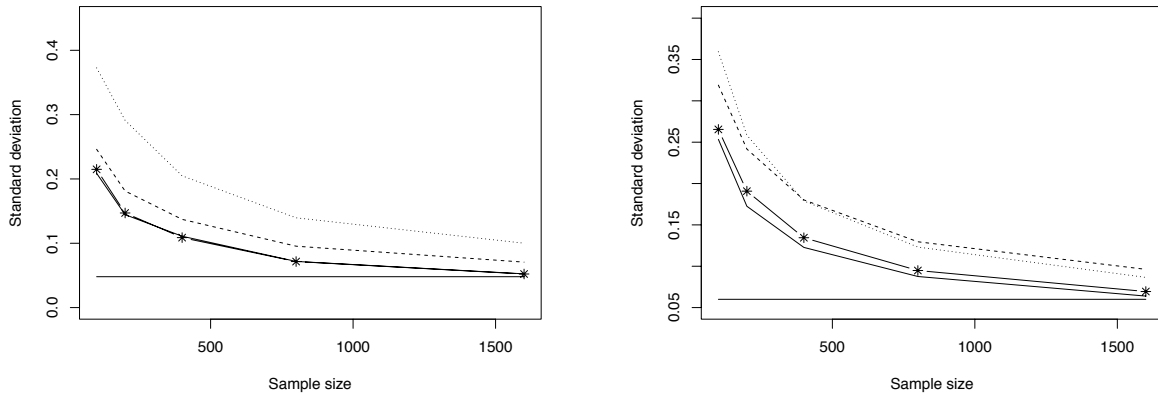


Figure 1: Estimation standard deviation versus sample size for two randomly selected elements in β . Line — marks the envelope estimator with approximate weights, line - * - marks the envelope estimator with exact weights, line - - marks the basic envelope estimator and line $\cdot \cdot \cdot$ marks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

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

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

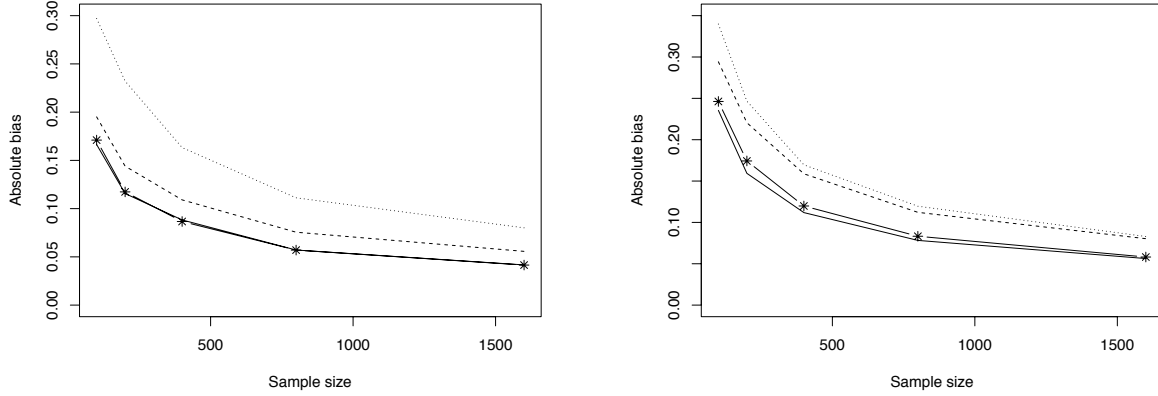


Figure 2: Average absolute bias versus sample size for two randomly selected elements in β . The line types are the same as in Figure 1.

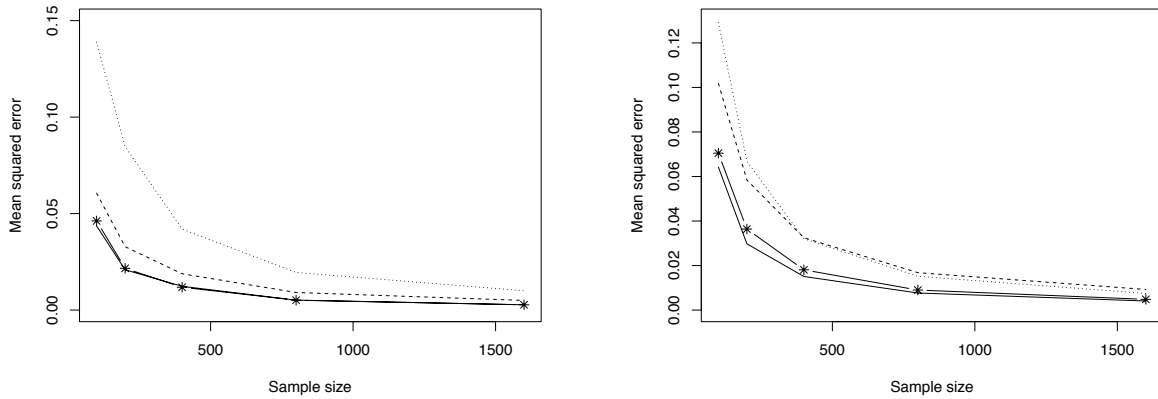


Figure 3: Average MSE versus sample size for two randomly selected elements in β . The line types are the same as in Figure 1.

406 the dimension selection criteria discussed in Section 6. For the 200 replications, we computed
 407 the fraction that a criterion selects the true dimension. The results are summarized in Table 1.
 408 When AIC or BIC are not selecting the true dimension, we find that they always overestimate
 409 the dimension. This will cause a loss of some efficiency gains, but it does not introduce bias in
 410 estimation. When the exact weights are used, BIC is a consistent selection criterion. AIC is too
 411 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.

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

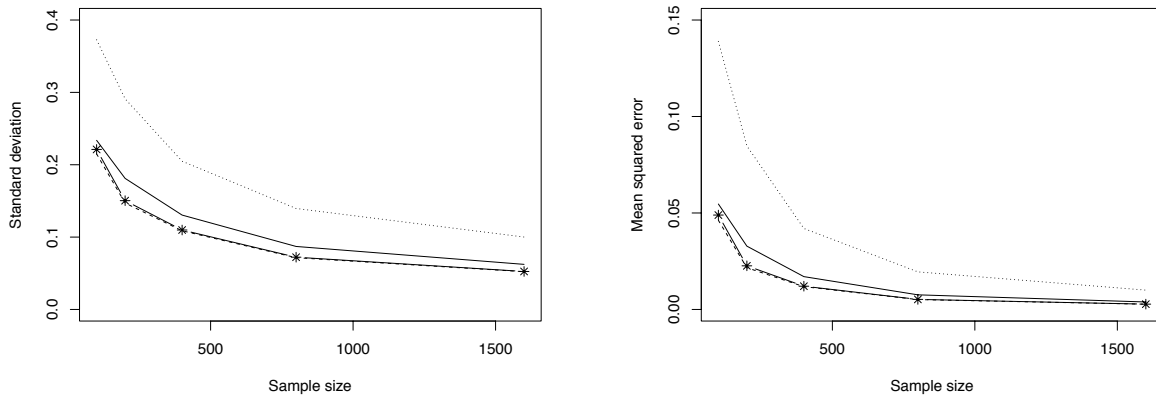
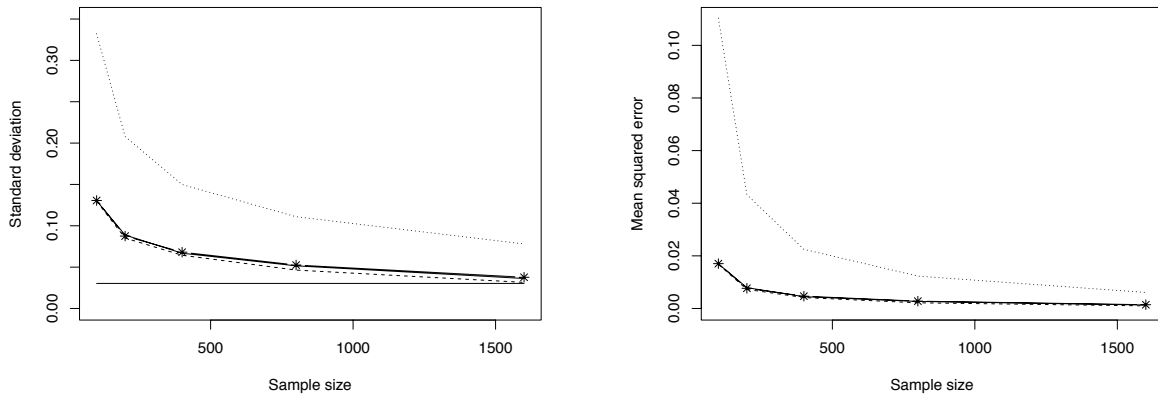


Figure 4: Estimation standard deviations and MSEs for a randomly selected element in β . Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. Line $- * -$ marks the envelope estimator with dimension selected by BIC using exact weights, line $—$ marks the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

416 We also investigate the performance of our estimators under normality. We repeated the sim-
417 ulations with the same settings except that the errors were generated from a multivariate normal
418 distribution. The results are summarized in Figure 5. From the plot, we notice that the estima-

419 tion standard deviations and MSEs of the basic envelope estimator and the envelope estimators
 420 with “exact” weights (i.e., the weights computed from t distribution) and approximate weights are
 421 almost indistinguishable. In this example, using the weights derived from t distribution or approx-
 422 imate weights does not cause a notable loss of efficiency in the normal case. This may be because
 423 that the approximate weights are computed from data, and therefore are data adaptive. For the
 424 “exact” weights, although it depends on the error distribution, it also has a data-dependent part
 425 (see Section 4.3). Therefore, these estimators do not lose much efficiency when the true distribu-
 426 tion is normal. The performance of the dimension selection criteria is similar to that in Table 1,
 427 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.



428 Figure 5: Estimation standard deviations and MSEs for a randomly selected element in β . Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 1.

429 7.2 Reduced rank envelope model

430 This simulation studies the estimation performance of different estimators in the context of the
 431 reduced rank envelope model. We set $r = 10$, $p = 5$, $d = 2$ and $u = 3$. The matrix (Γ, Γ_0) was

432 obtained by normalizing an r by r matrix of independent uniform $(0, 1)$ variates. The elements in
 433 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 $\sigma = 0.4$, $\sigma_0 = 0.1$
 434 and elements in A were $N(0, 1)$ variates. The elements in the predictor vector X are independent
 435 uniform $(0, 1)$ variates. The errors are generated from normal mixture distribution of two normal
 436 distributions $N(0, 2\Sigma)$ and $N(0, 0.1\Sigma)$ with probability 0.5 and 0.5. We varied the sample size
 437 from 100, 200, 400, 800 and 1600. For each sample size, we generated 10000 replications and com-
 438 puted the OLS estimator, the basic reduced rank envelope estimator and the reduced rank envelope
 439 estimator with exact weights (weights derived from the true error distribution) and approximate
 440 weights (Section 4.4). The estimation standard deviation for a randomly chosen element in β is
 441 displayed in the left panel of Figure 6. We notice that the basic reduced rank envelope estimator
 442 does not gain much efficiency compared to the OLS estimator. For example, with sample size 100,
 443 the standard deviation ratios of the OLS estimator versus the basic reduced rank envelope estimator
 444 range from 0.94 to 3.26 with an average of 1.42. The reduced rank envelope estimator computed
 445 from the exact weights obtains the most efficiency gains. When the sample size is 100, the ra-
 446 tios of the OLS estimator versus the reduced envelope estimator with exact weights range from
 447 2.37 to 12.00 with an average of 3.98. This indicates that correctly specifying the structure of the
 448 error distribution will provide efficiency gains in estimation. However, we do not know the exact
 449 weights in practice. Figure 6 shows that the estimator computed from the approximate weights still
 450 provides substantial efficiency gains. The ratios of the OLS estimator versus the reduced envelope
 451 estimator with approximate weights range from 1.63 to 8.79 with an average of 2.77. Although
 452 the estimator with approximate weights is not as efficient as the estimator with exact weights, it is
 453 still more efficiency than the basic reduced rank envelope estimator or the OLS estimator. We also
 454 computed the bootstrap standard deviation of the estimators from 10000 residual bootstraps, and

455 included the results in the right panel of Figure 6. The bootstrap standard deviation seems to be a
 456 good estimator of the actual estimation standard deviation. Therefore we compare the efficiency
 457 of different estimators using bootstrap standard deviations in applications.

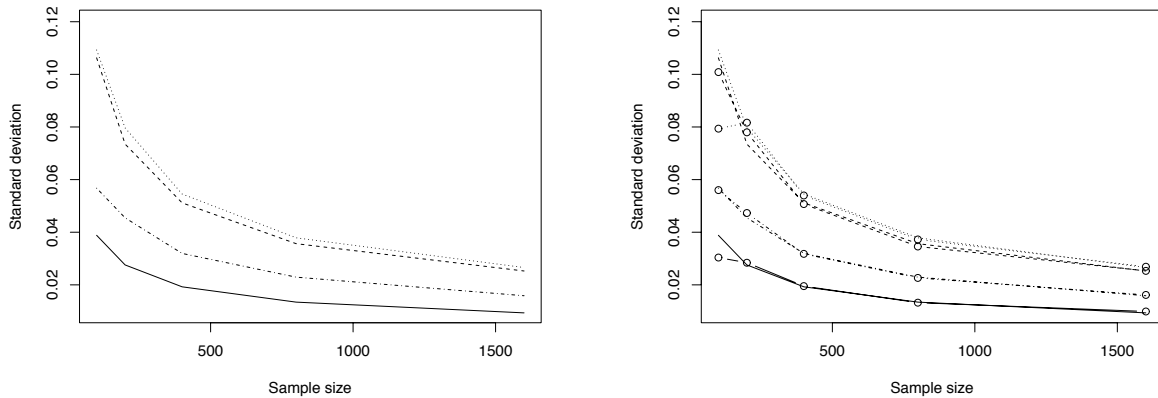


Figure 6: Estimation standard deviation and bootstrap standard deviation for a randomly selected element in β . Left panel: Estimation standard deviation only. Right panel: Estimation standard deviation with bootstrap standard deviation imposed. Line — marks the reduced rank envelope estimator with exact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

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

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

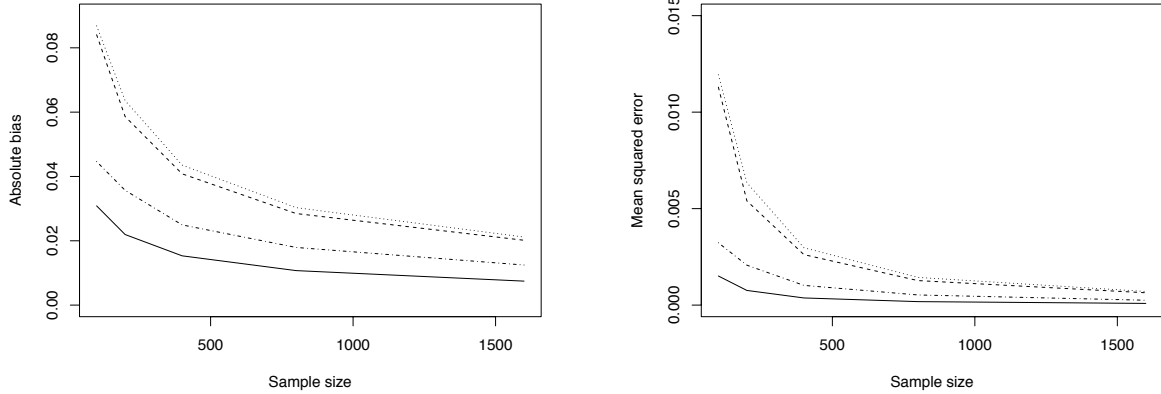


Figure 7: Average absolute bias and MSE for a randomly selected element in β . Left panel: Bias versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 6.

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

479 We repeated the simulation with the same setting as in Figure 6, but the errors were gener-
 480 ated 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.

481 Supplement.

482 7.3 Prediction

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

487 We first generated the data from the envelope model (5). We set $p = 5$, $r = 5$, $u = 3$ and
488 $n = 25$. The predictors were independent uniform $(0, 4)$ random variates. The coefficients had
489 the structure $\beta = \Gamma\xi$, where elements in ξ were independent standard normal random variates
490 and (Γ, Γ_0) were obtained by orthogonalizing an $r \times r$ matrix of uniform $(0, 1)$ variates. The
491 errors were generated from the multivariate t -distribution with mean 0, degrees of freedom 5 and
492 $\Sigma = \sigma^2\Gamma\Gamma^T + \sigma_0^2\Gamma_0\Gamma_0^T$, where $\sigma = 0.9$ and $\sigma_0 = 2$. We used 5-fold cross validation to evaluate
493 the prediction error, and the experiment was repeated for 50 random splits. The prediction error
494 was computed as $\sqrt{(Y - \hat{Y})^T(Y - \hat{Y})}$, where \hat{Y} was the predicted value based on the estimators
495 calculated from the training data. The average prediction error for 50 random splits were calculated
496 for the OLS estimator, basic envelope estimator, the envelope estimator with exact weights and the

497 envelope estimator with approximate weights. Results were summarized in Figure 8. The average
 498 prediction error for the OLS estimator is 8.34. We notice that the basic envelope estimator always
 499 has a larger prediction error than the OLS estimator for all u and its prediction error at $u = 3$
 500 is 8.46. This indicates that by misspecifying the error distribution, we can also have a worse
 501 performance on prediction. The predictor error for the envelope estimator with exact weights
 502 achieves its minimum 7.49 at the $u = 3$. Compared to the OLS estimator, the envelope estimator
 503 with exact weights reduces the prediction error by 10.2%. The estimator with approximate weights
 504 achieves its minimum prediction error 7.21 at $u = 3$, which is a 14.8% reduction compare with the
 505 OLS estimator. In this example, the estimator with approximate weights gives a better prediction
 506 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.

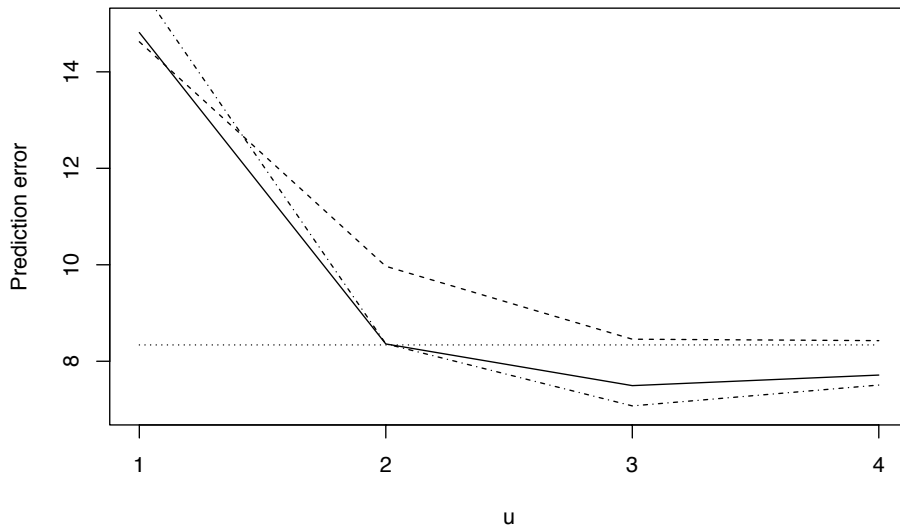


Figure 8: Prediction error versus u . Line — marks the envelope estimator with exact weights, line - · - marks the reduced rank envelope estimator with approximated weights, line - - marks the basic reduced rank envelope estimator and line · · · marks the OLS estimator.

507

508 In the second numerical study, data were simulated from the reduced rank envelope model
 509 (6). We set $p = 5$, $r = 10$, $d = 2$, $u = 3$ and $n = 30$. The predictors were independent
 510 uniform $(0, 1)$ random variates, and the errors were normal mixture random variates from two
 511 normal populations $N(0, 2\Sigma)$ and $N(0, 0.1\Sigma)$ with probability 0.5 and 0.5. Here Σ has the structure
 512 $\Sigma = \sigma^2 \Gamma A A^T \Gamma^T + \sigma_0^2 \Gamma_0 \Gamma_0^T$, where $\sigma = 0.4$, $\sigma_0 = 0.1$ and elements in A were standard normal
 513 random variates. The regression coefficients β has the structure $\beta = \Gamma \eta B$, where elements in B
 514 and η were independent standard normal random variates, and (Γ, Γ_0) was obtained by normalizing
 515 an $r \times r$ matrix of independent uniform $(0, 1)$ random variates. We computed the prediction errors
 516 of the OLS estimator, basic reduced rank envelope estimator, the reduced rank envelope estimators
 517 with true and approximate weights for u from d to $r - 1$. The prediction errors were calculated
 518 based on 5-fold cross validation with 50 random splits of the data. The results are included in
 519 Figure 9. The prediction error of the OLS estimator is 1.35. The basic reduced rank envelope
 520 estimator achieves its minimum prediction error 1.20 at $u = 7$, although the prediction errors
 521 for $u \geq 3$ are all quite close. Compared to the OLS estimator, the basic reduced rank envelope
 522 estimator reduced the prediction error by 11.1%. The reduced rank envelope estimator with exact
 523 weights achieves its minimum prediction error 1.14 at $u = 6$, which is a 15.6% reduction compared
 524 to the OLS estimator. The reduced rank envelope estimator with approximate weights achieves
 525 its minimum prediction error 1.11 at $u = 5$, which is a 17.8% reduction compared to the OLS
 526 estimator. In this numerical study, although the basic envelope estimator shows better prediction
 527 performance than the OLS estimator; by taking the error distribution into account, we can further
 528 improve the prediction performance.

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

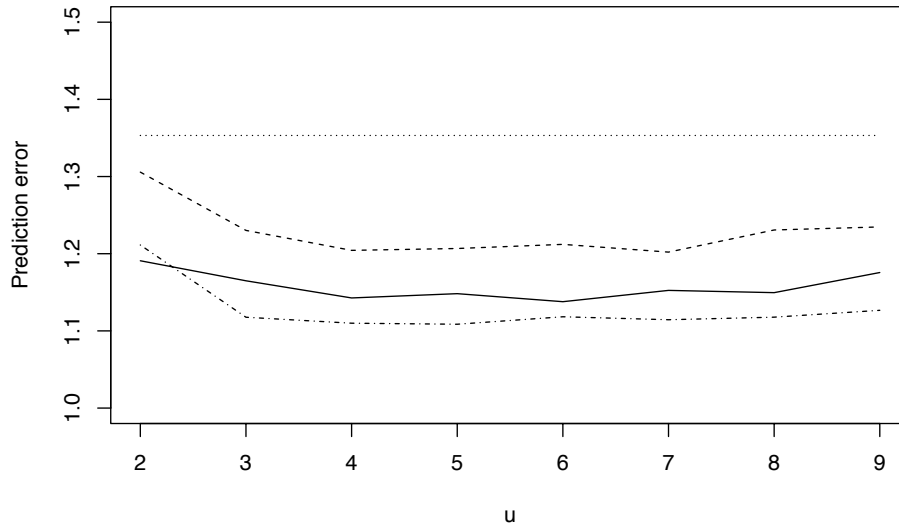


Figure 9: Prediction error versus u . Line — marks the envelope estimator with exact weights, line - - - marks the reduced rank envelope estimator with approximate weights, line - - marks the basic reduced rank envelope estimator and line ··· marks the OLS estimator.

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 enve-

542 lope model does not yield much efficiency gains in this data. The prediction error for the basic
543 envelope estimator is 24.2, which is quite close to that of the OLS estimator. From the discus-
544 sion in Section 7, we find that when the error distribution is unknown, the approximate weights
545 is adaptive to the data and gives good estimation and prediction results. We fit the data with the
546 reduced rank envelope estimator with approximate weights. The sequential test selected $d = 2$ and
547 BIC inferred $u = 2$. So the reduced rank envelope estimator degenerates to the envelope estimator
548 with approximate weights. The bootstrap standard deviation ratios of the OLS estimator versus the
549 envelope estimator with approximate weights range from 4.925 to 118.2 with an average of 55.57,
550 which suggests a substantial efficiency gain. This is also confirmed by prediction performance:
551 The prediction error is 12.27 for the envelope estimator with approximate weight. This is a 51%
552 reduction compared to the prediction error of the OLS estimator, and a 49% reduction compared
553 to the basic envelope estimator. This example shows that by considering the error structure of the
554 data, we achieve efficiency gains and also obtain better prediction performance.

555 **8.2 Vehicle data**

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

563 5-fold cross validation with 50 random splits to evaluate the prediction performance. We scale the
564 data so that all variables have unit standard deviation. This is because the range of the response
565 variables are very different, for example, price in dollars ranges from 5348 to 37038 while the
566 fuel consume in city ranges from 15 to 38. If the original scale is used, the prediction error is
567 dominated by price in dollars. The prediction error for the OLS estimator is 1.70. Then we fit the
568 reduced rank envelope estimator with approximate weights. The sequential test selected $d = 2$
569 and BIC suggested $u = 3$. The prediction error is 1.52, which is a 10.6% reduction compared
570 to that of the OLS estimator. The basic reduced rank envelope estimator with $u = 3$ and $d = 2$
571 has prediction error 1.64, which is a 3.5% reduction compared to that of the OLS estimator. The
572 bootstrap standard deviation ratios of the OLS estimator versus the basic reduced rank envelope
573 estimator range from 0.919 to 1.844 with an average of 1.277. And the bootstrap standard deviation
574 ratios of the OLS estimator versus the reduced rank envelope estimator with approximate weights
575 range from 0.862 to 1.734 with an average of 1.289. In this case, the estimation standard deviation
576 of the two estimators are similar. However, since the basic reduced rank envelope estimator has a
577 larger bias due to misspecification of the error structure, the reduced rank envelope estimator with
578 approximate weights gives a better prediction performance.

579 **Acknowledgement**

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

583 **Technical details**

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

586 **References**

- 587 [1] Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multi-
588 variate normal distributions. *Ann. Math. Statist.* **22**, 327–351.
- 589 [2] Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator
590 under general conditions. *Ann. Statist.* **27**, 1141–1154.
- 591 [3] Bai, Z., and Wu, Y. (1993). Recursive algorithm for M-estimates of regression coefficients
592 and scatter parameters in linear models. *Sankhyā: The Indian Journal of Statistics, Series B*,
593 **55**, 199-218.
- 594 [4] Bilodeau, M. and Brenner, M. (1999) *Theory of multivariate statistics*. Springer.
- 595 [5] Bura, E. and Forzani, L. (2015). Sufficient reductions in regressions with elliptically con-
596 toured inverse predictors. *J. Amer. Statist. Assoc.*, **110**, 420–434.
- 597 [6] Cook, R. D., Forzani, L. and Su, Z. (2016). A note on fast envelope estimation. *J. Multivariate*
598 *Anal.*, **150**, 42–54.
- 599 [7] Cook, R. D., Forzani, L., and Zhang, X. (2015). Envelopes and reduced rank regression.
600 *Biometrika*, **102(2)**, 439–456.

- 601 [8] Cook, R. D., Li, B. and Chiaromonte, F. (2010). Envelope models for parsimonious and
602 efficient multivariate linear regression (with discussion). *Statist. Sinica*, **20**, 927–1010.
- 603 [9] Cook, R. D. and Nachtsheim, C. J. (1994). Reweighting to achieve elliptically contoured
604 covariates in regression. *J. Amer. Statist. Assoc.*, **89**, 592–599.
- 605 [10] Cysneiros, F. J. A. and Paula, G. A. (2004). One-sided test in linear models with multivariate
606 t-distribution. *Commun Stat Simul Comput.*, **33**, 747–771.
- 607 [11] del Pino, G. (1989). The unifying role of iterative generalized least squares in statistical
608 algorithms. *Statistical Science*, **4(4)**, 394–408.
- 609 [12] Díaz-García, J. A., Galea Rojas, M. and Leiva-Sánchez, V. (2003). Influence diagnostics for
610 elliptical multivariate linear regression models. *Comm. Statist. Theory Methods*, **32**, 625–641.
- 611 [13] Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2007). The distribution of residuals from a gen-
612 eral elliptical linear model. *J. Statist. Plann. Inference*, **137**, 2347–2354.
- 613 [14] Dutta, S., & Genton, M. G. (2017). Depth-weighted robust multivariate regression with ap-
614 plication to sparse data. *Canad. J. of Statist.*, **45(2)**, 164–184.
- 615 [15] Fang, K. and Anderson, T. W. (1990). *Statistical Inference in Elliptically Contoured and*
616 *Related Distributions*. New York: Allerton Press.
- 617 [16] Fang, K., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*.
618 New York: Chapman and Hall.
- 619 [17] Fang, K. and Zhang, Y. (1990). *Generalized multivariate analysis*. Beijing: Science Press
620 and Berlin: Springer-Verlag.

- 621 [18] Fernandez, C., and Steel, M. F. J. (1999). Multivariate student- t regression models: pitfalls
622 and inference. *Biometrika*, **86**, 153–167.
- 623 [19] Gabriel F. (2004). *Generalized Elliptical Distributions: Theory and Applications*. PhD thesis,
624 Universität zu Köln.
- 625 [20] Galea, M., Riquelme, M., and Paula, G. A. (2000). Diagnostic methods in elliptical linear
626 regression models. *Braz. J. Prob. Stat.*, **14**, 167–184.
- 627 [21] García Ben, M., Martínez, E. and Yohai, V. J. (2006). Robust estimation for the multivariate
628 linear model based on a τ -scale. *J. Multivariate Anal.*, **97**, 1600–1622.
- 629 [22] Gómez, E., Gómez-Villegas, M. A. and Marín, J. M. (2003). A survey on continuous elliptical
630 vector distributions. *Rev. Mat. Complut*, **16**, 345–361.
- 631 [23] Green, P. J. (1984), Iteratively reweighted least squares for maximum likelihood estimation,
632 and some robust and resistant alternatives, *Journal of the Royal Statistical Society. Series B*
633 *(Methodological)*, **46(2)**, 149–192.
- 634 [24] Gupta, A. K., Varga, T. and Bodnar, T. (2013). *Elliptically Contoured Models in Statistics*
635 *and Portfolio Theory*. New York: Springer.
- 636 [25] Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some
637 uses in Jacobians and multivariate statistics. *Canad. J. of Statist.*, **7**, 65–81.
- 638 [26] Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *J. Multi-*
639 *variate Anal.*, **5**, 248–264.

- 640 [27] Kent, J.T., and D.E. Tyler (1991). Redescending M-estimates of multivariate location and
641 scatter. *Ann. Statist.*, **19**, 2102–2119.
- 642 [28] Kowalski, J., Mendoza-Blanco, J. R., Tu, X. M. and Gleser, L. J. (1999). On the difference
643 in inference and prediction between the joint and independent f-error models for seemingly
644 unrelated regressions. *Comm. Statist. Theory Methods*, **28**, 2119–2140.
- 645 [29] Kudraszow, N. L. and Maronna, R. A. (2011). Estimates of MM type for the multivariate
646 linear model. *Journal of Multivariate Analysis*, **102**, 1280–1292.
- 647 [30] Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using
648 the t-distribution. *J. Amer. Statist. Assoc.*, **84**, 881–896.
- 649 [31] Lee, M. and Su, Z. (2018). *Renvlp: Computing Envelope Estimators*. R package version 2.5.
- 650 [32] Lemonte, A. J. and Patriota, A. G. (2011). Multivariate elliptical models with general param-
651 eterization. *Statistical Methodology*, **8**, 389–400.
- 652 [33] Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers*, **41**, 211–224.
- 653 [34] Liu, S. (2002). Local influence in multivariate elliptical linear regression models. *Linear Al-*
654 *gebra Appl.*, **354**, 159–174.
- 655 [35] Maronna, R. A. (1976). Robust estimators of Multivariate location and scatter. *Ann. Statist.*,
656 **4**, 51–67.
- 657 [36] Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics: Theory and Meth-*
658 *ods*. Chichester: John Wiley & Sons, Ltd.

- 659 [37] Melo, T. F. N., Ferrari, S. L. P. and Patriota, A. G. (2017) Improved hypothesis testing in a
660 general multivariate elliptical model. *J. Stat. Comput. Simul.* **87(7)**, 1416–1428.
- 661 [38] Melo, T. F. N., Ferrari, S. L. P. and Patriota, A. G. (2018) Improved estimation in a general
662 multivariate elliptical model. *Braz. J. Probab. Stat.* **32(1)**, 44–68.
- 663 [39] Miao, B. Q, and Wu, Y. (1996) Limiting behavior of recursive M-estimators in multivariate
664 linear regression models. *J. Multivariate Anal.*, **59**, 60–80
- 665 [40] Osorio, F., Paula, G.A. and Galea, M. (2007). Assessment of local influence in elliptical linear
666 models with longitudinal structure. *Comput. Statist. Data Anal.*, **51**, 4354–4368.
- 667 [41] Reinsel, G. C. and Velu, R. P. (1998). *Multivariate reduced rank Regression: Theory and*
668 *Applications*. New York: Springer.
- 669 [42] Rousseeuw, P. J., Van Aelst, S., Driessen, K. V. and Gulló, J. A. (2004). Robust multivariate
670 regression. *Technometrics*, **46**, 293–305.
- 671 [43] Russo, C. M., Paula, G. A. and Aoki, R. (2009). Influence diagnostics in nonlinear mixed-
672 effects elliptical models. *Comput. Statist. Data Anal.*, **53**, 4143–4156.
- 673 [44] Sanchez, G. (2016). *plsdepot: Partial Least Squares (PLS) Data Analysis Methods*. R pack-
674 age version 0.1.17.
- 675 [45] Savalli, C., Paula, G. A. and Cysneiros, F. J. A. (2006). Assessment of variance components
676 in elliptical linear mixed models. *Stat. Modelling*, **6**, 59–76.
- 677 [46] Stoica, P. and Viberg, M. (1996). Maximum likelihood parameter and rank estimation in
678 reduced-rank multivariate linear regressions. *IEEE Trans. Signal Process.* **44**, 3069–3078.

- 679 [47] van der Vaart, A.W. (2000). *Asymptotic Statistics*, Cambridge University Press.
- 680 [48] Welsh, A. T. and Richardson, A. M. (1997). Approaches to the robust estimation of mixed
681 models. *Handbook of Statistics*, **15**, 343–384.
- 682 [49] Yeh, I. C. (2007). Modeling slump flow of concrete using second-order regressions and arti-
683 ficial neural networks. *Cement and Concrete Composites*, **29(6)**, 474–480.
- 684 [50] Zhao, W., Lian, H. and Ma, S. (2017). Robust reduced rank modeling via rank regression. *J.*
685 *Statist. Plann. Inference*, **180**, 1–12.