Statistica Sinica **31** (2021), 301-332
doi:https://doi.org/10.5705/ss.202017.0424

ENVELOPES FOR ELLIPTICAL MULTIVARIATE LINEAR REGRESSION

Liliana Forzani and Zhihua Su

Universidad Nacional del Litoral and University of Florida

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

Key words and phrases: Elliptical multivariate linear regression, envelopes, reduced-rank regression.

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 vector $X \in \mathbb{R}^p$. It can be formulated as

$$Y = \mu_Y + \beta(X - \mu_X) + \epsilon, \tag{1.1}$$

where $\beta \in \mathbb{R}^{r\times p}$ is the coefficient matrix, and 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. Here, associations between responses are not considered; thus, the efficiency of the estimation can be improved by incorporating such dependencies. The envelope model, introduced in a seminal paper Cook, Li and Chiaromonte (2010) in the context of (1.1), uses sufficient dimension-reduction techniques to identify the part of Y that is immaterial to the changes

Corresponding author: Zhihua Su, Department of Statistics, University of Florida, Gainsville, FL 32603, USA. E-mail: zhihuasu@stat.u.edu.

in X. This immaterial part is removed from any subsequent analysis, making the estimation more efficient. Another method that considers the associations between responses is the reduced-rank regression (Anderson (1951), Anderson (1999), Izenman (1975), Reinsel and Velu (1998), Stoica and Viberg (1996)). A 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 $d \leq \min(r, p)$, resulting in fewer parameters and, thus, estimators that are more efficient.

The envelope model and the reduced-rank regression both use dimension-reduction techniques to improve the estimation efficiency, but they adopt different perspectives and make different assumptions. In practice, determining which method is more efficient for a given problem is not always straightforward. In response to this problem, the reduced-rank envelope was proposed recently in Cook, Forzani and Zhang (2015), which combines the advantages of both, while making it more efficient than either of the individual methods. However, the estimation of the reduced-rank envelope model uses 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 which case, alternative distributions (or methodologies) need to be considered. One choice is the family of elliptically contoured distributions, which includes the normal, Student-t, power exponential, and contaminated normal distributions, among others. These may have heavier or lighter tails than the normal distribution, and are more adaptive to the data. Elliptical multivariate linear regression models have been studied extensively in the statistical literature; see, for example, Bura and Forzani (2015), Cysneiros and Paula (2004), Díaz-García, Galea Rojas and Leiva-Sánchez (2003), Díaz-García and Gutiérrez-Jáimez (2007), Fernandez and Steel (1999), Gabriel (2004), Gómez, Gómez-Villegas and Marín (2003), Galea, Riquelme and Paula (2000), Kowalski et al. (1999), Lange, Little and Taylor (1989), Lemonte and Patriota (2011), Liu (2000), Liu (2002), Osorio, Paula and Galea (2007), Russo, Paula and Aoki (2009), Savalli, Paula and Cysneiros (2006), and Welsh and Richardson (1997), among others. In particular, Lemonte and Patriota (2011) 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, including nonlinear regressions, a mixed-effects model with nonlinear fixed effects, and errors-in-variables models. Bias correction for the maximum likelihood estimator (MLE) and adjustments of the likelihood-ratio statistics are also derived for this general model (see Melo, Ferrari and Patriota (2018), Melo, Ferrari and Patriota (2017)). Elliptical distributions can also be used to consider robustness in a multivariate linear regression, as in Dutta and Genton (2017), García Ben, Martínez and Yohai (2006), Kudraszow and Maronna (2011), Maronna, Martin and Yohai (2006), Rousseeuw et al. (2004), and Zhao, Lian and Ma (2017), among others. Nevertheless, the envelope model has yet to be implemented in the context of an elliptical multivariate regression. Furthermore, few studies have examined reduced-rank regressions beyond the normal case. Thus far, the only attempts to extend a reduced-rank regression to a nonnormal case have used M-estimators or other robust estimators, including some from the elliptical class. For example, Zhao, Lian and Ma (2017) develops a robust estimator for a reduced-rank regression, and proposes a novel rank-based estimation procedure using Wilcoxon scores. Although the reduced-rank estimator in Zhao, Lian and Ma (2017) allows a general error distribution, we aim to further improve the efficiency of the estimator using an MLE and the envelope method in the context of elliptical multivariate linear regressions.

The goal of this study is to derive a reduced-rank regression estimator, envelope estimator, and reduced-rank envelope estimator for an elliptical multivariate linear regression. Because the 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 latter. We examine the asymptotic properties and efficiency gains of the reduced-rank regression, envelope model, and reduced-rank envelope model, and demonstrate their effectiveness using 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 for elliptical multivariate linear regressions, and Section 3 describes the most used elliptically contoured distributions in such regressions. In Section 4, we derive the 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 presents the asymptotic properties of the estimators, and demonstrates the efficiency gains without the normality assumptions. Section 6 discusses how to select the rank and dimension for the reduced-rank envelope model. The simulation results are presented in Section 7, and examples are given in Section 8. All proofs are included in the online Supplementary Material.

2. Models

We consider the following elliptical multivariate linear regression model:

$$Y = \mu_Y + \beta(X - \mu_X) + \epsilon, \quad \epsilon \sim EC_r(0, \Sigma, g_{Y|X}), \tag{2.1}$$

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

$$f_Z(z) = |\Sigma_Z|^{-1/2} g_Z\left[(z - \mu_Z)^T \Sigma_Z^{-1} (z - \mu_Z)\right], \tag{2.2}$$

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 (2.1) the *standard model* in the following discussion. Based on (2.1), $Y \mid X$ 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 the 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 Fang and Zhang (1990), p.65). Note that, in general, $\mathrm{var}(Y \mid X)$ depends on X, except for the normal errors with constant variance.

A reduced-rank regression assumes that the rank of the coefficient β in model (2.1) is at most $d \leq \min(p, r)$. As a result, we have

$$\beta = AB, \ A \in \mathbb{R}^{r\times d}, \ \ B \in \mathbb{R}^{d\times p}, \ \mathrm{rank}(A) = \mathrm{rank}(B) = d, \tag{2.3}$$

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, because $AB = (AU)(U^{-1}B) := A^*B^*$, for any invertible U. In the case of normally distributed errors with a constant covariance matrix, the MLE of β for (2.3) and its asymptotic distribution are derived in Anderson (1999), Reinsel and Velu (1998), and Stoica and Viberg (1996), by imposing various constraints on A and B for identifiability. Because the goal is to estimate β, rather than A and/or B, Cook, Forzani and Zhang (2015) derived an estimator for β that does not impose any constraints on A and B, other than requiring that the rank of β be equal to d. It has been shown that when ϵ follows a multivariate normal distribution with a constant covariance matrix, the reduced-rank regression may yield an estimator for β that is more efficient than the ordinary least square (OLS) estimator. Note that, under normality, the OLS estimator is the MLE for the standard model.

The envelope model Cook, Li and Chiaromonte (2010) provides another way to obtain an efficient estimator. Let span(β) denote the subspace spanned by the columns of β. Under model (2.1), if span(β) is contained in the span of m ($m < r$) eigenvectors of the error covariance matrix Σ, but 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 $\mathbb{R}^r$ spanned by some eigenvectors of Σ, and let 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; that is, (Γ, Γ_0) is an orthogonal matrix. Because span$(\beta) \subseteq \mathcal{E}_\Sigma(\beta) = \text{span}(\Gamma)$, there exists $\xi \in \mathbb{R}^{u\times p}$ such that $\beta = \Gamma\xi$. Because $\mathcal{E}_\Sigma(\beta)$ is spanned by the eigenvectors of Σ, there exist $\Omega \in \mathbb{R}^{u\times u}$ and $\Omega_0 \in \mathbb{R}^{(r-u)\times(r-u)}$, such that $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$. Here, ξ includes the coordinates of β with respect to Γ, and Ω and Ω_0 contain the coordinates of Σ with respect to Γ and Γ_0, respectively. To summarize, if β and Γ satisfy the conditions

$$\beta = \Gamma\xi, \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \tag{2.4}$$

we refer to (2.1) as an envelope model of dimension u. Based on (2.4), $\mathcal{E}_\Sigma(\beta)$ provides a link between β and Σ: the variation in ϵ can be decomposed into $\Gamma\Omega\Gamma^T$, which is material to the estimation β, and $\Gamma_0\Omega_0\Gamma_0^T$, which is immaterial to the estimation of β. Using this decomposition, Cook, Li and Chiaromonte (2010) showed that the envelope estimator of β in a normal setting is at least as efficient as the OLS estimator, asymptotically. The improvement in 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 a normal distribution of the error term, Cook, Forzani and Zhang (2015) presented a novel unified framework for the reduced-rank regression and the envelope model, called the reduced-rank envelope model, which obtains estimators that are more efficient than those of the individual methods. Cook, Forzani and Zhang (2015) assumed that β and Σ follow the envelope structure (2.4), and that 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 rank $d \leq \min(r, p)$. Then, model (2.1) is called the reduced-rank envelope model when

$$\beta = AB = \Gamma\eta B, \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \tag{2.5}$$

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; that is, (Γ, Γ_0) is an orthogonal matrix. The reduced-rank envelope model performs a dimension reduction on two levels: The first level, $\beta = AB$, assumes we have a reduced-rank regression. The second level, $\beta = \Gamma\eta B$, assumes that β intersects only u eigenvectors of the covariance matrix Σ. When $u = r$, $\Gamma = I_r$; then, (2.5) degenerates to the usual reduced-rank regression in (2.3). When $d = \min(u, p)$, (2.5) reduces to the envelope model in (2.4). Finally, the reduced-rank envelope model in (2.5) is equivalent to the standard model in (2.1) when $d = u = r$. Cook, Forzani and Zhang (2015) obtained the MLEs of β and Σ, as well as their asymptotic distributions under the normality assumption. Note that for the reduced-rank regression, envelope model, and reduced-rank envelope model, the constituent parameters A, B, Γ, Γ_0, ξ, η, Ω, and 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 Cook, Forzani and Zhang (2015) when studying the asymptotic distribution of the identifiable parameters β and Σ.

3. Examples

In this section, we present three scenarios in which elliptically contoured distributions are used in regressions.

3.1. The data matrix follows an elliptically contoured distribution

A case that is commonly studied in the literature is that in which 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 the 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 ones. Under this assumption, and using Theorem 2.8

from Gupta, Varga and Bodnar (2013), 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 using a heteroscedastic structure; see Gupta, Varga and Bodnar (2013) for further details. Examples of this distribution include the matrix variate symmetric Kotz-type distribution, Pearson type-II distribution, Pearson type-VII distribution, symmetric Bessel distribution, symmetric logistic distribution, and symmetric stable law, among others. Of the aforementioned distributions, the most common is the normal distribution with a nonconstant variance Fang, Kotz and Ng (1990).

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 $M = 1_n\mu_Y^T + (I_n - (1/n)1_n 1_n^T)\mathbb{X}\beta^T$ and η a diagonal matrix. The diagonal elements of η are denoted by η_{ii}, where $\eta_{ii} > 0$, 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 $g_i(t) = (2\pi\eta_{ii})^{-r/2} e^{-t/(2\eta_{ii})}$.

Note that we consider only the error structure in which the covariance matrices are proportional and the heteroscedasticity depends on g, but not the scale parameter. The general nonconstant 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 are jointly elliptically contoured distributed

Sometimes, $(X^T, Y^T)^T$ jointly follows an elliptically contoured distribution, or can be transformed to ellipticity (e.g., Cook and Nachtsheim (1994)). 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

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

where $g(\cdot) \geq 0$ and Φ is a $(p + r) \times (p + r)$ positive-definite matrix. Following Bura and Forzani (2015), if we partition Φ as

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_X & \Phi_Y \\ \Phi_Y^T & \Sigma_Y \end{pmatrix}, \tag{3.1}$$

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 Gupta, Varga and Bodnar (2013)). 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 the multivariate t-distribution as an example. Suppose $Z \in R^k$ 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

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

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 (3.1). 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 $g(t) = (\nu + p + t)^{-(p+r+\nu)/2}$, $h(X) = [\nu + (X - \mu_X)^T\Phi_{11}^{-1}(X - \mu_X)]/(\nu + p)$, 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 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 $N_r(\mu_{Y|X}, k_i\Sigma)$, for $i = 1, \ldots, m$, with weights $p_1, \ldots, p_m$. Then, its density function is given by

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

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 for modeling 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

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

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

4. Estimation

Under the standard model (2.1), 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, Fang and Anderson (1990)). 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 using an iterative re-weighted least squares (IRLS) algorithm; see Bura and Forzani (2015) for a discussion of its properties. When $g_{Y|X}$ is unknown, Bura and Forzani (2015) derived an estimator of β and investigated its properties.

The goal of this section is to derive the MLEs for the reduced-rank regression, envelope model, and 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 $\mathcal{E}_\Sigma(\beta)$. The procedures used to select 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 Cook, Forzani and Zhang (2015) 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 in (2.1), the parameter vector is $h = \left(\mathrm{vec}^T(\beta), \mathrm{vech}^T(\Sigma)\right)^T$. We do not consider μ_X or μ_Y, because the estimators are asymptotically independent of the estimators of β and Σ. Here ψ denotes the parameter vector of the reduced-rank regression in (2.3), δ denotes that of the envelope model in (2.4), and ϕ denotes that of the reduced-rank envelope model (2.5). Then,

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

We use $N(v)$ to denote the number of parameters in the 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 fewer parameters than the standard model, because $N(h) - N(\psi) = (p-d)(r-d) \geq 0$; the reduced-rank envelope model has even fewer parameters than the reduced-rank regression does, because $N(\psi) - N(\phi) = (r-u)d \geq 0$. On the other hand, compared with the standard model, the number of parameters is reduced by $p(r-u) \geq 0$ when using the envelope model, and is further reduced by $(p-d)(u-d) \geq 0$ when using the reduced-rank envelope model.

Remark 1. If the model assumption holds, fewer parameters often results in better estimation efficiency and, thus, improved prediction accuracy. However, if the model assumption does not hold, having fewer parameters introduces bias, but may still reduce the variance of the estimator. In this case, we have a trade-off between the bias and the variance reduction.

4.2. MLEs

Assume that $Y \mid X$ follows an elliptically contoured distribution $EC_r(0, \Sigma, g_{Y|X})$, with density given by (2.2). Let (X_i, Y_i) be n independent samples of (X, Y), for $i = 1, \dots, n$, 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

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

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

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

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

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

If the weights W_i are positive and are known, we can 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 a normal distribution. With this idea in mind, we propose the following IRLS algorithm, for $W_i \geq 0$. The estimator obtained from this algorithm is equivalent to the MLE estimator (See del Pino (1989) and Green (1984)).

1. Obtain 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 estimators.
   (b) Using the data $(\sqrt{W_i}X_i, \sqrt{W_i}Y_i)$, update the estimators of β and Σ as if the data follow a normal distribution.

The estimator in Step 2(b) is obtained from the fast envelope estimation algorithm developed in Cook, Forzani and Su (2016), which is implemented in the R package `Renvlp` Lee and Su (2018). In addition to the standard model, this algorithm can also be used for the reduced-rank regression, envelope model, and reduced-rank envelope model. For example, consider the reduced-rank envelope model:

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

Note that the term $\partial h/\partial \phi^T$ is the same for both likelihoods, and that h is a function of β and Σ. We can estimate the reduced-rank envelope estimator using the preceding algorithm, except that 2(b) changes to "Using 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 a normal distribution." The reduced-rank regression and the envelope model follow the same procedure. For completeness, the online Supplementary Material includes 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 distributions.

Normal with nonconstant variance

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

Normal mixture distribution

Suppose $Y \mid X$ follows a normal mixture distribution from m normal distributions $N_r(\mu_{Y|X}, k_i\Sigma)$, for $i = 1, \ldots, m$, with weights $p_1, \ldots, p_m$. From the discussion 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 $(X^T, Y^T)^T$ follows a joint multivariate t-distribution $t_{p+r}((\mu_X^T, \mu_Y^T)^T, \Phi, \nu)$, with Φ following the structure in (3.1). Based on the discussion in Section 3.2, $Y \mid X$ follows the t-distribution $t_r(\mu_Y + \Phi_{21}\Phi_{11}^{-1}(X - \mu_X), (\nu + (x-\mu_X)^T \Phi_{11}^{-1}(x - \mu_X))/(\nu + p)\Phi_{22.1}, \nu + p)$. 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 $\Sigma_{Y|X} = (\nu + (x - \mu_X)^T \Phi_{11}^{-1}(x - \mu_X))\Phi_{22.1}/(\nu + p)$.

Conditional t-distribution

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

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

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

Note that all weights are positive. For illustration purposes, the constants η_{ii} in the normal distribution with a nonconstant variance, and the k_i 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 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 (2.1). 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 Fang and Zhang (1990)). Note that c_X can vary across 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 follow a normal distribution. If c_{X_i} is unknown, we estimate it using $\hat{c}_{X_i} = [Y_i - \widehat{\mu}_Y - \widehat{\beta}(X_i - \widehat{\mu}_X)]^T \widehat{\Sigma}^{-1} [Y_i - \widehat{\mu}_Y - \widehat{\beta}(X_i - \widehat{\mu}_X)]$. According to Bura and Forzani (2015), 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 summarizes the preceding discussion.

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

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 estimates of μ_X, μ_Y, β, and Σ, respectively.

    (b) Using 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 follow a normal distribution.

Note that this algorithm is similar to that 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 refer to $\hat{c}_{X_i}^{-1}$ as *approximate weights* in subsequent discussions. The $\sqrt{n}$-consistency of the estimator of β, obtained using these *approximate weights*, follows similarly to Theorem 4 from Bura and Forzani (2015).

5. Asymptotics

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

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 Henderson and Searle (1979) that connects the vector operator vec and the vector half operator vech, as follows: $\mathrm{vec}(S) = E_r\,\mathrm{vech}(S)$ and $\mathrm{vech}(S) = C_r \mathrm{vec}(S)$, for any $r \times r$ symmetric matrix S. Let $U = \Sigma^{-1/2}[Y - \mu_Y - \beta(X - \mu_X)]$, $N_X = E[(g'(U^TU)/g(U^TU))^2 U^TU|X]/r$, and $M_X = E[(g'(U^TU)/g(U^TU))^2 (U^TU)^2|X]/[r(r+2)]$. We define $\widetilde{\Sigma}_X = E(N_X X X^T)$ and $M = E(M_X)$ if X is random and the expectations exist, and $\widetilde{\Sigma}_X = \lim_{n\to\infty}(1/n)\sum_{i=1}^n N_{X_i} X_i X_i^T$ and $M = \lim_{n\to\infty}(1/n)\sum_{i=1}^n M_{X_i}$ 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 require g, such that the above quantities are finite, and that the MLE for model (2.1) exists, is consistent, and is asymptotically normal (e.g., see the conditions for elliptical distributions in Miao and Wu (1996), Bilodeau and Brenner (1999), Bai and Wu (1993), Kent and Tyler (1991), Kudraszow and Maronna (2011), Maronna (1976), Miao and Wu (1996), and Zhao, Lian and Ma (2017) or, more generally, the conditions on Theorems 5.23, 5.31, 5.39, 5.41, or 5.42 from van der Vaart (2000)). 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 $J_\Sigma = 2ME_r^T(\Sigma^{-1}\otimes\Sigma^{-1})E_r + (M - 1/4)E_r^T\mathrm{vec}(\Sigma^{-1})\,\mathrm{vec}^T(\Sigma^{-1})E_r$. Detailed calculations are included in the Supplementary Material. 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., Cook, Li and Chiaromonte (2010)).

Proposition 1 gives the asymptotic variance of the MLEs of β under the standard model (2.1), reduced-rank regression (2.3), envelope model (2.4), and reduced-rank envelope model (2.5). Suppose $\widehat{\theta}$ 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 model* (2.1) *holds; that is, the error vector* ϵ *follows the elliptically contoured distribution* $EC_r(0, \Sigma, g)$. *Suppose the MLE of* β *under the standard model* (2.1), $\widehat{\beta}_{\mathrm{std}}$, *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), *for* $i = 1, \ldots, n$, *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* (5.1). *If models* (2.3), (2.4), *or* (2.5) *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* (5.2), (5.3), *and* (5.4), *respectively.*

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{std}})] = \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma, \tag{5.1}$$

$$\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RR}})] = \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes (\Sigma - M_A), \tag{5.2}$$

$$\begin{aligned}\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_E)] = {} & \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T + \frac{1}{4}(\xi^T \otimes \Gamma_0)[\xi\widetilde{\Sigma}_X\xi^T \otimes \Omega_0^{-1} + M(\Omega \otimes \Omega_0^{-1} \\ & + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1}(\xi \otimes \Gamma_0^T),\end{aligned} \tag{5.3}$$

$$\begin{aligned}\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})] = {} & \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Sigma - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes [\Sigma - \Gamma\eta(\eta^T\Omega^{-1}\eta)^{-1}\eta^T\Gamma^T] \\ & - \frac{1}{4}M_B \otimes \Gamma_0\Omega_0\Gamma_0^T + \frac{1}{4}(B^T\eta^T \otimes \Gamma_0)[\eta B\widetilde{\Sigma}_X B^T\eta^T \otimes \Omega_0^{-1} \\ & + M(\Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_u \otimes I_{r-u})]^{-1} \\ & (\eta B \otimes \Gamma_0^T),\end{aligned} \tag{5.4}$$

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 2. The asymptotic variance does not depend on the choices of A, B, Γ, ξ, or η, because 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 3. Note 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, which 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. Because $\widetilde{\Sigma}_X^{-1} - M_B$ 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 that the reduced-rank regression estimator is at least 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. Note 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,

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

To compare the reduced-rank envelope estimator and the reduced-rank regression estimator, note that

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

Therefore, the reduced-rank envelope estimator is at least as efficient as the reduced-rank regression estimator. Finally, comparing the envelope estimator and the reduced-rank envelope estimator, we have

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

$$= \frac{1}{4}\widetilde{\Sigma}_X^{-1} \otimes \Gamma\Omega\Gamma^T - \frac{1}{4}(\widetilde{\Sigma}_X^{-1} - M_B) \otimes (\Gamma\Omega\Gamma^T - \Gamma\Omega^{1/2}P_{\Omega^{-1/2}\eta}\Omega^{1/2}\Gamma^T),$$

where $P_{\Omega^{-1/2}\eta}$ denotes the projection matrix onto the space spanned by the columns of $\Omega^{-1/2}\eta$. We have $\mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_E)] \geq \mathrm{avar}[\sqrt{n}\mathrm{vec}(\widehat{\beta}_{\mathrm{RE}})]$, because $\widetilde{\Sigma}_X^{-1} - M_B$ 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 among the compared models.

6. Selections of rank and envelope dimension

For the reduced-rank regression, we choose d using the same sequential test as in Cook, Forzani and Zhang (2015). 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 $\widehat{\Sigma}_X^{1/2}\widehat{\beta}_{\mathrm{std}}^T\widehat{\Sigma}_{Y|X}^{-1/2}$, $\widehat{\Sigma}_X$ 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 an information criterion such as the AIC or BIC to select the dimension u. The criterion requires the log-likelihood function. Here, we use the actual log-likelihood if g is known. If g is unknown, we substitute the normal log-likelihood with the following 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 $u = u_0$, for $0 \leq u_0 \leq r$. Then, u is chosen to minimize $-2l_u + kN(\delta)$, where $N(\delta)$ is the number of parameters in the envelope model at dimension u (see Section 4.1), and k is the penalty, which takes the value 2 in the AIC and $\log(n)$ in the 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 compute the information criterion, the AIC or BIC, for $u = d_0, \ldots, r$, in the same way as we did for the envelope model. We select u that minimizes the information criterion.

7. Simulations

In this section, we report the results from numerical experiments used to compare the performance of the estimators derived under the elliptically contoured distribution, normal likelihood, and approximate weights. The simulation in Section 7.1 employs the envelope model, and that in Section 7.2 uses 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 refer to the envelope model derived in Cook, Li and Chiaromonte (2010) as the basic envelope model, and to the reduced-rank envelope model derived in Cook, Forzani and Zhang (2015) as the basic reduced-rank envelope model.

7.1. Envelope model

In this simulation, we investigate the estimation performance of our estimators in the context of the envelope model. We set $p = 5$, $r = 20$, and $u = 4$. The predictors are generated independently from a uniform $(0, 5)$ distribution, (Γ, Γ_0) is obtained by orthogonalizing an $r \times r$ matrix of independent uniform $(0, 1)$ variates, and the elements in ξ are independent standard normal variates. The errors are generated from a 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 is zero. The sample size varies, from 100, 200, 400, 800, to 1,600. For each sample size, we generated 10,000 replications. For each data set, we computed the OLS estimator, basic envelope estimator, envelope estimator with exact weights (computed from the true g), and envelope estimator derived using 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. However, the envelope estimators with exact weights and approximate weights achieve even better efficiency. The right panel indicates that the basic envelope model can be similar, or even less efficient than the OLS estimator, whereas the envelope estimators with exact weights or approximate weights are always more efficient than the OLS estimator. For example, at sample size 1,600, 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 approximate weights are used, the ratios of the estimation standard deviation of the OLS estimator over the envelope

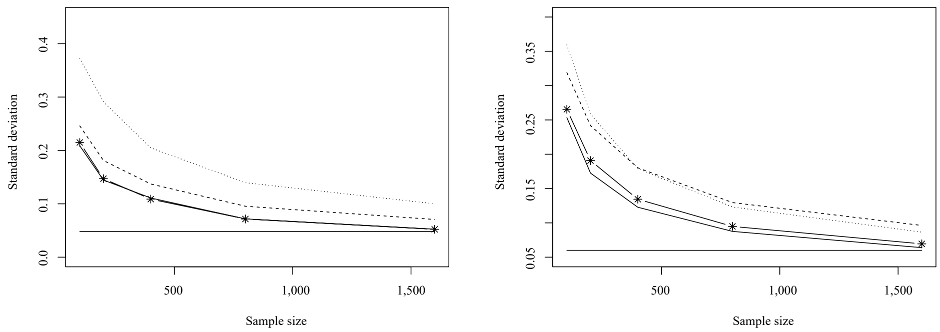

Figure 1. Estimation standard deviation versus sample size for two randomly selected elements in $\boldsymbol{\beta}$. Line — marks the envelope estimator with approximate weights, line $-*-$ marks the envelope estimator with exact weights, line - - marks the basic envelope estimator, and line $\cdots$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

estimator range from 1.301 to 3.467, with an average of 1.903. The performance of the envelope estimator with approximate weights is very similar to that with exact weights, as shown in Figure 1. At times, it may even be more efficient than the envelope estimator with exact weights, because it is data adaptive, as indicated in the right panel. Figure 1 also confirms the asymptotic distribution derived in Section 5, and that the envelope estimator with exact weights is $\sqrt{n}$-consistent. We computed the bootstrap standard deviation for each estimator, and found that it is a good approximation to the actual estimation standard deviation. This result is not shown in the figure for readability.

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

The results in Figure 1 are based on a known dimension of the envelope subspace. However, the dimension u is usually unknown in practice. Therefore, we examine the performance of 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 the AIC and BIC do not select the true dimension, we find that they always overestimate the dimension. This will cause a loss of efficiency, but it does not introduce bias into the estimation. When exact weights are used, the BIC is a consistent se-

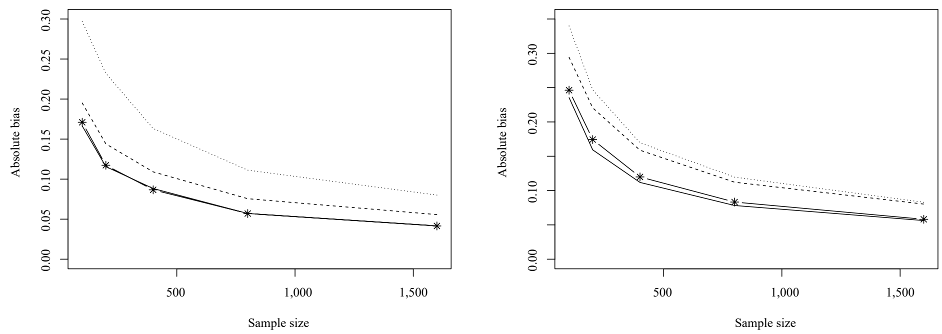

Figure 2. Average absolute bias versus sample size for two randomly selected elements in $\boldsymbol{\beta}$. The line types are the same as in Figure 1.

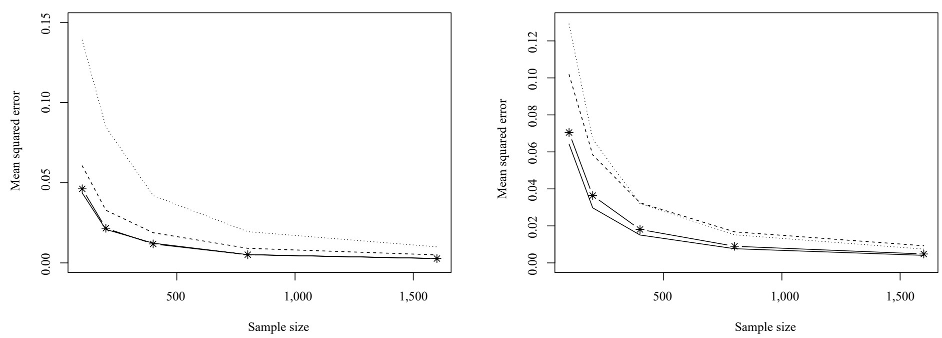

Figure 3. Average MSE versus sample size for two randomly selected elements in $\boldsymbol{\beta}$. The line types are the same as in Figure 1.

lection criterion. The AIC is too conservative, and selects a bigger dimension in most cases. When approximate weights are used, the BIC tends to overestimate the dimension of the envelope subspace. However, we can still achieve efficiency gains and have a smaller MSE than that of 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, owing to the consistency of the BIC.

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, which shows that the estimation standard deviations and MSEs of

Table 1. Percentage selection of the true dimension.

	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%
$n =$ 1,600	14.4%	98.1%	33.8%	42.6%

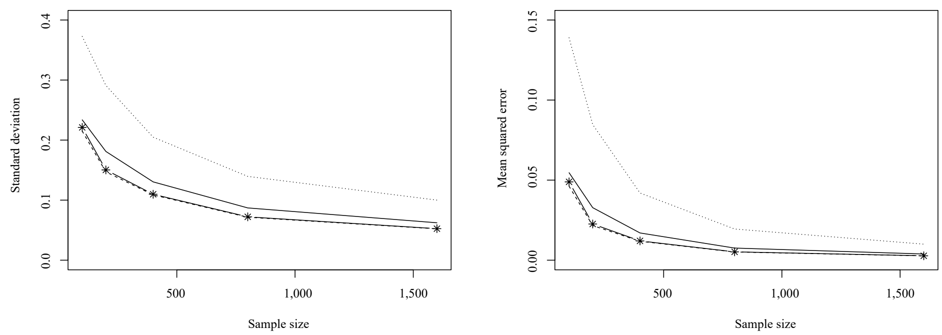

Figure 4. Estimation standard deviations and MSEs for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. Line $-*-$ marks the envelope estimator with dimension selected by BIC using exact weights, line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

the basic envelope estimator and those of the envelope estimators with "exact" weights (i.e., weights computed from a t-distribution) and approximate weights are almost indistinguishable. In this example, using weights derived from a t-distribution or approximate weights does not cause a notable loss of efficiency in the normal case. This may be because the approximate weights are computed from data, and therefore are data adaptive. Although the "exact" weights depend on the error distribution, they also have 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.

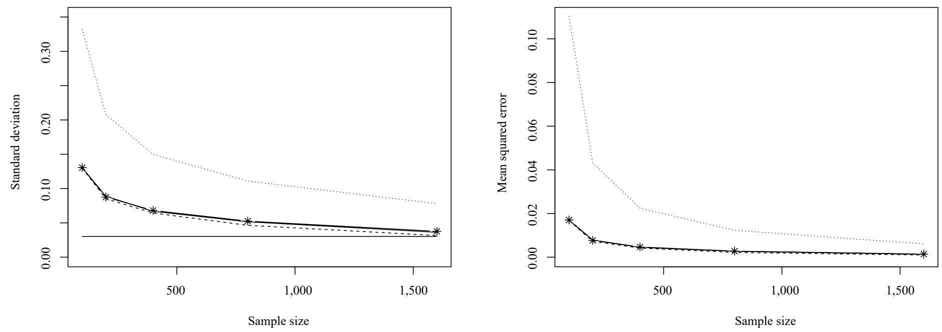

Figure 5. Estimation standard deviations and MSEs for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 1.

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) is obtained by normalizing an $r \times r$ matrix of independent uniform $(0, 1)$ variates. The elements in η and B are 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 are $N(0, 1)$ variates. The elements in the predictor vector X are independent uniform $(0, 1)$ variates. The errors are generated from a normal mixture distribution of two normal distributions $N(0, 2\Sigma)$ and $N(0, 0.1\Sigma)$, with probability 0.5 and 0.5, respectively. We varied the sample size from 100, 200, 400, 800, to 1,600. For each sample size, we generated 10,000 replications and computed the OLS estimator, basic reduced-rank envelope estimator, and reduced-rank envelope estimator with exact 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. Note 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

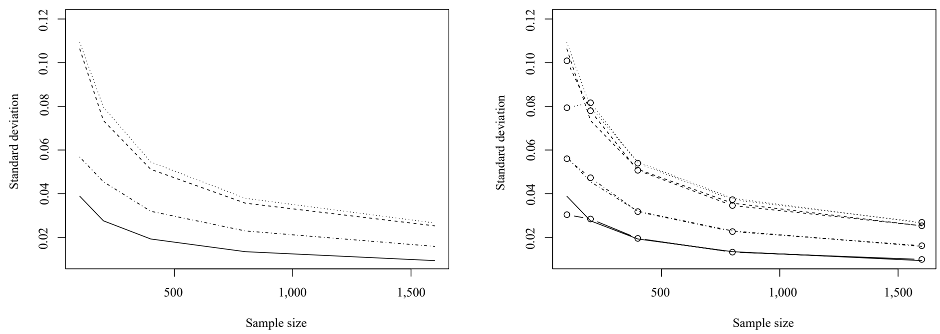

Figure 6. Estimation standard deviation and bootstrap standard deviation for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Estimation standard deviation only. Right panel: Estimation standard deviation with bootstrap standard deviation imposed. Line — marks the reduced-rank envelope 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.

average of 3.98. This indicates that correctly specifying the structure of the error distribution offers efficiency gains in the 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 efficient than the basic reduced-rank envelope estimator or the OLS estimator. We also computed the bootstrap standard deviation of the estimators from 10,000 residual bootstraps; see 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.

We investigate 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 find that for all estimators, the estimation standard deviation is the major component of the MSE. Therefore, the MSEs follow a similar trend to that of the estimation standard deviation. From the absolute bias plot, we find that the OLS estimator and the basic reduced-rank envelope estimator are more biased than the reduced-rank envelope estimators with true and approximate

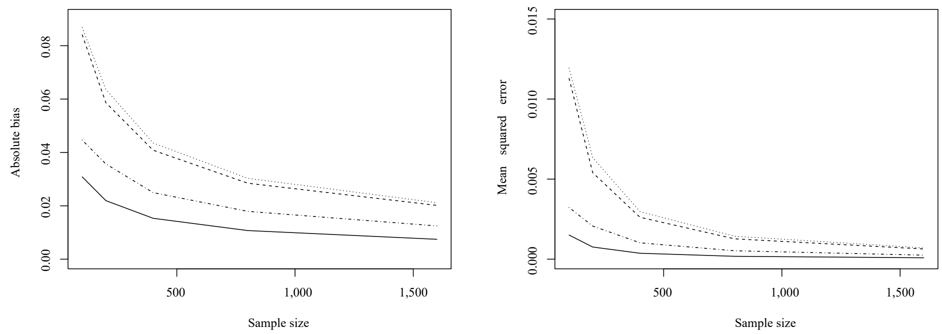

Figure 7. Average absolute bias and MSE for a randomly selected element in $\boldsymbol{\beta}$. Left panel: Bias versus sample size. Right panel: MSE versus sample size. The line types are the same as in Figure 6.

weights. Figures 6 and 7 together show that we obtain a less biased and more efficiency estimator by considering the error distribution.

Now, we examine the performance of the sequential test, AIC, and BIC discussed in Section 6 in the selection of d and u. We use the same context as that for Figures 6 and 7, and compute the percentage of times that a particular criterion selects the true dimension (out of 200 replications). The significance level for the sequential test is 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 exact weights are used, the BIC performs better because it is a consistent selection criterion. The AIC tends to be conservative and always selects a bigger dimension. When approximate weights are used, the AIC and BIC tend to overestimate the dimension of the envelope subspace. Overestimation causes a loss of efficiency, but it retains useful information. Based on this result, we use the BIC to choose u in applications. Next, 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 the BIC using approximate weights loses some efficiency compared with the estimator with known dimension and exact weights. However, 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 Supplementary Material.

Table 2. Percentage selection of the true dimension.

	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%
$n =$ 1,600	98.8%	83.5%	99.8%	26.0%	29.3%

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 envelope model, and the other under the reduced-rank envelope model.

We first generated the data from the envelope model in (2.4), where $p = 5$, $r = 5$, $u = 3$, and $n = 25$. The predictors are independent uniform $(0, 4)$ random variates. The coefficients have the structure $\beta = \Gamma\xi$, where elements in ξ are independent standard normal random variates, and (Γ, Γ_0) are obtained by orthogonalizing an $r\times r$ matrix of uniform $(0, 1)$ variates. The errors are 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 five-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 from the training data. The average prediction error for 50 random splits was calculated for the OLS estimator, basic envelope estimator, envelope estimator with exact weights, and envelope estimator with approximate weights. The results appear in Figure 8. The average prediction error for the OLS estimator is 8.34. Note that the basic envelope estimator always has a larger prediction error than that of 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 worse performance in terms of prediction. The predictor error for the envelope estimator with exact weights achieves its minimum 7.49 at the $u = 3$. Compared with the OLS estimator, the envelope estimator with exact weights reduces the prediction error by 10.2%. The estimator with approximate weights achieves a minimum prediction error of 7.21 at $u = 3$, which is a 14.8% reduction over the OLS estimator. In this

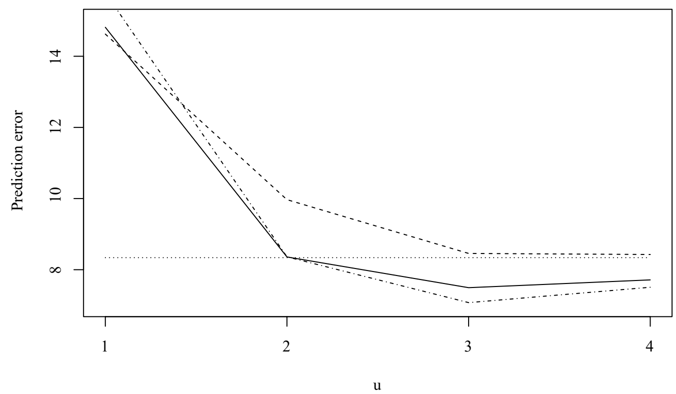

Figure 8. Prediction error versus u. Line — marks the envelope estimator with exact weights, line —·— marks the reduced-rank envelope estimator with approximated weights, line - - marks the basic reduced-rank envelope estimator, and line $\cdots$ marks the OLS estimator.

example, the estimator with approximate weights gives a better prediction than the estimator with exact weights. This might be because we have a small sample size and because the approximate weights are more adaptive to the data.

In the second numerical study, data were simulated from the reduced-rank envelope model given in (2.5). We set $p = 5$, $r = 10$, $d = 2$, $u = 3$, and $n = 30$. The predictors are independent uniform $(0, 1)$ random variates, and the errors are 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 the elements in A are standard normal random variates. The regression coefficient β has the structure $\beta = \Gamma\eta B$, where elements in B and η are independent standard normal random variates, and (Γ, Γ_0) is 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, and reduced-rank envelope estimators with true and approximate weights for u, from d to $r-1$. The prediction errors are calculated based on five-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 with the OLS estimator, the basic reduced-rank

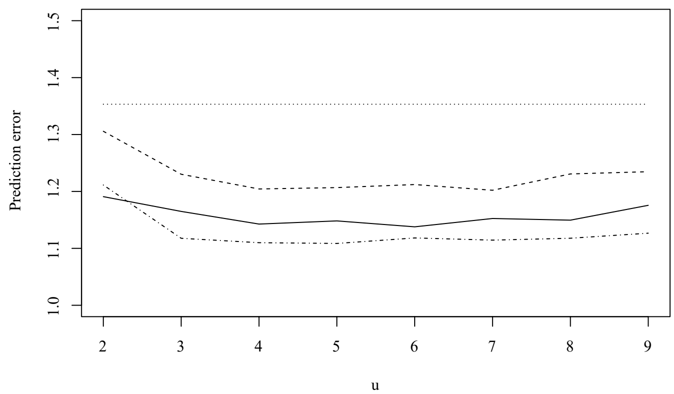

Figure 9. Prediction error versus u. Line — marks the envelope estimator with exact weights, line $-\cdot-$ marks the reduced-rank envelope estimator with approximate weights, line - - marks the basic reduced-rank envelope estimator, and line $\cdots$ marks the OLS estimator.

envelope estimator reduced the prediction error by 11.1%. The reduced-rank envelope estimator with exact weights achieves a minimum prediction error 1.14 at $u = 6$, which is a 15.6% reduction compared with the OLS estimator. The reduced-rank envelope estimator with approximate weights achieves a minimum prediction error of 1.11 at $u = 5$, which is a 17.8% reduction compared with the OLS estimator. In this numerical study, although the basic envelope estimator shows better prediction performance than that of 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 approximate weights.

8. Examples

8.1. Concrete slump test data

The slump flow of concrete depends on the components of the concrete. This data set contains 103 records on various mix proportions Yeh (2007), 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. These are ingredients of concrete, and are measured in kilograms per cubic meter concrete. The output variables are slump, flow, and 28-day compressive strength. We use the first 78 records as a training set, and the new 25 records as the testing set. The prediction error of the OLS estimator is 25.0. We fit the basic envelope model to the data, and the BIC suggests $u = 2$. 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 of an efficiency gain for 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 are adaptive to the data, and give good estimations and prediction results. We fit the data using the reduced-rank envelope estimator with approximate weights. The sequential test selected $d = 2$ and the BIC suggested $u = 2$. Thus, 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 confirmed by the prediction performance. The prediction error is 12.27 for the envelope estimator with approximate weight. This is a 51% reduction compared with the prediction error of the OLS estimator, and a 49% reduction compared with the basic envelope estimator. This example shows that considering the error structure of the data achieves efficiency gains and better prediction performance.

8.2. Vehicle data

The vehicle data examined here contain measurements for various characteristics of 30 vehicles from different manufacturers, including Audi, Dodge, Honda, and so on. The data are available in the R package *plsdepot* Sanchez (2016), and are used to illustrate the partial least squares regression methods. Following Sanchez (2016), we use price in dollars (USD), insurance risk rating, fuel consumption (miles per gallon) in a city, and fuel consumption on highways 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 five-fold cross-validation with 50 random splits to evaluate the prediction performance. We scale the data so that all variables have unit standard deviation because the range of the response variables is relatively wide. For example, price in dollars ranges from 5,348 to 37,038, whereas the fuel consumption in a city ranges from 15 to 38. If the original scale is used, the prediction error is dominated by the 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 the BIC suggested $u = 3$. The prediction error is 1.52, which is a 10.6% reduction compared with 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 with 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. In addition, the ratios of bootstrap standard deviations of the OLS estimator versus those of 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 standard deviations of the two estimators are similar. However, because the basic reduced-rank envelope estimator has a larger bias, due to the misspecification of the error structure, the reduced-rank envelope estimator with approximate weights gives better prediction performance.

Supplementary Material

The online Supplementary Material contains proofs for the results presented here, as well as additional simulation results.

Acknowledgements

We are grateful to the associate editor and two referees for their helpful comments and suggestions. This work was supported, in part, by grant DMS-1407460 from the US National Science Foundation.

References

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

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

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.

Bilodeau, M. and Brenner, D. (1999) *Theory of Multivariate Statistics.* Springer.

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

Cook, R. D., Forzani, L. and Su, Z. (2016). A note on fast envelope estimation. *J. Multivariate*

Anal. **150**, 42–54.

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

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.

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

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.

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

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.

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.

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

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

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

Fang, K. and Zhang, Y. (1990). *Generalized Multivariate Analysis.* Science Press, Beijing and Springer-Verlag, Berlin.

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

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

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

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.

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

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 (Statistical Methodology)* **46**, 149–192.

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

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.

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

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

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.

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

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.

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

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

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

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

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

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

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**, 1416–1428.

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**, 44–68.

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

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.

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

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

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.

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

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.

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.

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

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

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

Zhao, W., Lian, H. and Ma, S. (2017). Robust reduced-rank modeling via rank regression. *J.*

Statist. Plann. Inference **180**, 1–12.

Liliana Forzani

Facultad de Ingeniería Química, Universidad Nacional del Litoral, CONICET-UNL, Santa Fe, Argentina.

E-mail: liliana.forzani@gmail.com

Zhihua Su

Department of Statistics, University of Florida, Gainsville, FL 32603, USA.

E-mail: zhihuasu@stat.ufl.edu

(Received October 2017; accepted April 2019)