Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/020201

On Multivariate Rank and Generalized Rank Regression

Majeda Salman[1] *and Joseph W. McKean*[2]

[1]University of Bahrain, Kingdom of Bahrain
[2]Western Michigan University, USA

Received: 7 Jun. 2012, Revised: 21 Sep. 2012, Accepted: 23 Sep. 2012
Published online: 1 Jul. 2013

Abstract: Multivariate regression estimates based on ranks and generalized ranks are proposed. These estimates are based on a transformation and retransformation technique that uses Tyler's (1987) M-estimator of scatter. The proposed estimates are obtained by retransforming the componentwise rank-based estimate due to Davis and McKean (1993) and a componentwise generalized rank estimate. Asymptotic properties of the estimates are established under some regularity conditions. It is shown that both estimates have a multivariate normal limiting distribution. The influence function of the retransformed generalized rank estimate has a bounded influence in both factor and response spaces. It is shown through a simulation study that the transformed-retransformed R and GR estimates are highly efficient compared to the componentwise R, GR and least absolute deviations estimates. Also, it is shown that the new estimates perform better than the least squares estimate when the errors have a heavy tailed distribution. An example illustrating the estimation procedures is presented.

Keywords: Asymptotic distributions; Efficiency; Robust; Regression; GR-estimators; R-estimators; Least squares; Simulations; Wilcoxon.

1 Introduction

Suppose we have a matrix of response variables **Y** which follow the multivariate linear model

$$\mathbf{Y} = \mathbb{X}\mathscr{B} + \boldsymbol{\mathcal{E}}, \tag{1.1}$$

where $\mathbb{X}$ is an $n \times d$ matrix of regression coefficients, $\mathscr{B}$ is an $p \times d$ matrix of regression parameters, and $\boldsymbol{\mathcal{E}}$ is an $n \times d$ matrix of random errors whose rows have covariance matrix Σ. We wish to estimate and make inference on the parameters $\mathscr{B}$.

The least squares estimate of $\mathscr{B}$, $\widehat{\mathscr{B}}_{LS} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbf{Y}$, has two advantages besides being easy to compute, it is affine equivariant. It is equivariant under constant shifts and multiplication by arbitrary nonsingular matrices. It is the optimal estimator of $\mathscr{B}$ when the distribution of the errors is multivariate normal. However, it is not robust if the errors have a heavy tailed distribution.

There are several approaches in the literature competing with the LS method and producing estimators that are robust and more efficient. One approach is to use a robust fit on each component separately. Rao (1988) used this approach where the robust fit was based on the least absolute deviations (LAD) estimators. Davis and McKean (1993) developed a rank-based theory for the multivariate linear model in a manner similar to its development for the univariate linear model; see Hettmansperger and McKean (2011). The estimates are component-wise R estimates for general score functions, including the sign (LAD) and Wilcoxon scores. Besides estimation, their analysis includes confidence regions and tests of general linear multivariate hypotheses. However, the efficiency of these estimates slips when the variables are highly correlated. We extend their theory to positive breakdown estimates (see Theorem 4.4). Hence, the analysis includes a large family of estimates, including highly efficient and positive breakdown estimates.

Maronna and Morgethaler (1986) proposed the covariance estimation approach to estimate the parameters of a univariate linear regression model. The data are summarized by a covariance matrix of the concatenated vector of explanatory variables and response variable. A robust estimate of the covariance matrix leads to a robust regression

* Corresponding author e-mail: salmanmajda@yahoo.com

estimate. Ollila, Hettmansperger and Oja (2002) used a similar approach in multivariate linear regression. They showed under some conditions that the sign covariance matrix (SCM) regression estimate is a consistent estimate of $\mathscr{B}$, affine equivariant, and asymptotically normal. It has a bounded influence function in both the $\boldsymbol{x}$ and $\boldsymbol{y}$ spaces.

Chakraborty (1996) proposed an extension of LAD based on the transformation and retransformation technique. He computed his transformation matrix from a subset of the data which must be chosen. He called his estimate, TREMMER, (Transformation-Retransformation Estimate in Multivariate Median Regression). In his paper, he proved that the TREMMER is asymptotically normal and highly efficient relative to LAD especially when the correlation between the response variables increases. The choice of the optimal α was based on minimizing the asymptotic generalized variance of the TREMMER. Chakraborty and Chauduri (1997) applied the TREMMER algorithm to Davis and McKean (1993) componentwise rank regression estimate that is based on the Wilcoxon scores. They proved that their estimate inherits its asymptotic normality and robustness from the Wilcoxon estimate. The transformation matrix was chosen such that the asymptotic generalized variance of $\sqrt{n}\,(\widehat{\mathscr{B}} - \mathscr{B})$ is minimum.

In this paper, we propose robust transformation-retransformation estimates for a general multivariate linear model. The procedure is based on three steps, (Section 2). First, the matrix of responses is transformed using Tyler's (1987) multivariate scatter matrix. Then the Davis and McKean estimates are obtained on the transformed observations. In the third step, these estimates are retransformed. General scores can be used for the R estimates; hence, these estimates can be optimized if knowledge of the error distribuion is known. We shall call this type of estimate transformed-retransformed R (TRR) estimates. The estimates using the GR estimates are called the transformed-retransformed GR (TRGR) estimates. We show the TRR and TRGR estimates satisfy some equivariance properties and obtain their asymptotic distributions in Sections 2 and 4. The TRR estimates have bounded influence functions in the $\boldsymbol{y}$ space and possess good efficiency properties. The TRGR estimates have positive breakdown and have a bounded influence function in both the $\boldsymbol{x}$ and $\boldsymbol{y}$ spaces. The efficieny of either estimate does not slip for highly correlated data. Unlike Chakraborty's (1996) estimate, the transformation is based on all the data. Hence, no subset of the data has to be chosen. These estimates offer the user a large class of estimates from which to choose including highly efficient estimates and positive breakdown estimates all of which are quickly computed.

Consistent estimators of the asymptotic standard deviations are available. Also, because of the quick computation bootstrap estimates of the standard errors can be used for moderate sized data sets. We discuss both in the paper and compare them on an example in Section 3. Section 5 presents the results of simulation studies which demonstrated the high efficiency of the new estimates relative to LAD. In particular, our transformation-retransformation rank estimate performs as well as or better than the estimate of Chakraborty and Chauduri (1997).

2 Transformation-Retransformation R and GR Estimators

In this section, we describe the transformed-retransformed R (TRR) and GR (TRGR) estimators. We consider the multivariate linear model

$$\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \mathscr{B}_1'\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i, \quad i = 1,\ldots,n, \tag{2.1}$$

where $\boldsymbol{y}_i \in \Re^d$ is a vector of response variables, $\boldsymbol{x}_i \in \Re^p$ is a vector of constant regressors, $\boldsymbol{\beta}_0 \in \Re^d$ is an unknown vector of intercepts, $\mathscr{B}_1$ is a $p \times d$ matrix of unknown regression coefficients, and $\boldsymbol{\varepsilon}_i \in \Re^d$ is a vector of random errors. The random errors $\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n$ are assumed to be independent and identically distributed with $E[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix. Let

$$\mathbf{Y} = \begin{pmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_n' \end{pmatrix}, \mathbf{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}, \boldsymbol{\mathcal{E}} = \begin{pmatrix} \boldsymbol{\varepsilon}_1' \\ \vdots \\ \boldsymbol{\varepsilon}_n' \end{pmatrix},$$

$$\mathbb{X} = \left(\mathbf{1}_n\ \mathbf{X}\right), \mathscr{B} = \begin{pmatrix} \boldsymbol{\beta}_0' \\ \mathscr{B}_1 \end{pmatrix}.$$

Then we can write model (2.1) in a matrix form as

$$\mathbf{Y} = \mathbb{X}\mathscr{B} + \boldsymbol{\mathcal{E}}. \tag{2.2}$$

The algorithm for the transformed-retransformed R (TRR) estimator is:

1. **Transformation Step.** Fit Model (2.1) using LS and obtain the LS residuals $\widehat{\boldsymbol{\varepsilon}}_1, \widehat{\boldsymbol{\varepsilon}}_2, \ldots, \widehat{\boldsymbol{\varepsilon}}_n$. Then obtain the transformation matrix $\hat{A} = \hat{A}(\widehat{\boldsymbol{\varepsilon}}_1, \widehat{\boldsymbol{\varepsilon}}_2, \ldots, \widehat{\boldsymbol{\varepsilon}}_n)$, as described in Section 2.1 and use it to get the transformed response variables $\boldsymbol{z}_i = \hat{A}\boldsymbol{y}_i$ for $i = 1,\ldots,n$.

2. **R-Estimation Step.** Obtain the component-wise R estimate, $\widehat{\mathscr{B}}_R = \left(\widehat{\boldsymbol{\beta}}_{0,s}\ \widehat{\mathscr{B}}'_{1,\varphi}\right)'$, on the data $(\boldsymbol{x}_1, \boldsymbol{z}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{z}_n)$, as described in Section 2.2.
3. **Retransformation Step.** Retransform $\widehat{\mathscr{B}}_R$ to obtain the TRR estimate $\widehat{\mathscr{B}}_{TRR} = \widehat{\mathscr{B}}_R(\hat{A}')^{-1}$.

The algorithm for the transformed-retransformed GR (TRGR) estimator is the same as that of the TRR estimator, except the R-estimation step is replaced by the GR-estimation step, as described in Section 2.3.

We briefly describe the transformation and R-estimation steps. More details are given in Section 4. The LS estimator of $\mathscr{B}$ is of course

$$\widehat{\mathscr{B}}_{LS} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbf{Y}.$$

The LS estimator is a quick computation. Furthermore, $\widehat{\mathscr{B}}_{LS}$ is affine equivariant.

2.1 Transformation Step

The transformation matrix $\hat{A}$ is a data-driven nonsingular matrix that was proposed by Tyler (1987). Given the LS residuals $\widehat{\boldsymbol{\varepsilon}}_1, \widehat{\boldsymbol{\varepsilon}}_2, \ldots, \widehat{\boldsymbol{\varepsilon}}_n$, $\hat{A}$ is the unique upper triangular positive definite matrix with a one in the upper left hand element that solves

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A\widehat{\boldsymbol{\varepsilon}}_i}{\|A\widehat{\boldsymbol{\varepsilon}}_i\|}\right)\left(\frac{A\widehat{\boldsymbol{\varepsilon}}_i}{\|A\widehat{\boldsymbol{\varepsilon}}_i\|}\right)' = \frac{1}{d}I \tag{2.3}$$

Equation (2.3) shows that the transformation matrix $\hat{A}$ is chosen so that the sample variance-covariance matrix of the unit-transformed vectors is $1/d$ times the identity. In other words, the unit vectors of the transformed residuals have the variance covariance structure of a random variable that is uniformly distributed on the unit d-sphere. Besides being nonsingular, $\hat{A}$ satisfies the affine equivariance property

$$D'\hat{A}_D'\hat{A}_D D = c_0\hat{A}'\hat{A}, \tag{2.4}$$

for a fixed nonsingular $d \times d$ matrix D, where $\hat{A}_D$ is the matrix $\hat{A}$ calculated on the transformed observations $D\widehat{\boldsymbol{\varepsilon}}_i$, $\hat{A}$ is the computed matrix on the residuals $\widehat{\boldsymbol{\varepsilon}}_i$, and c_0 is a positive scalar that may depend on D and the $\widehat{\boldsymbol{\varepsilon}}_i$'s.

For the location problem, Tyler (1987) showed that $\hat{A}$ is unique if the sample is drawn from a continuous distribution and $n > d(d-1)$. He also proved that $\hat{A}$ is consistent. In Section A.2, we show that $\hat{A} = \hat{A}(\widehat{\boldsymbol{\varepsilon}}_1, \widehat{\boldsymbol{\varepsilon}}_2, \ldots, \widehat{\boldsymbol{\varepsilon}}_n)$ is a consistent estimator. Discussion of an iterative procedure which is quick and easy for computing $\hat{A}$, (Randles (2000)), is given in Appendix A.1.

2.2 R-Estimation Step.

Davis and McKean (1993) developed a rank-based theory for the multivariate linear model in a manner similar to its development for the univariate linear model; see also Chapter 6 of Hettmansperger and McKean (2011). The estimate of $\mathscr{B}$ was obtained by first estimating the regression coefficient matrix $\mathscr{B}_1$ by minimizing for $j = 1, \ldots, d$ the dispersion functions

$$D(\mathscr{B}_1^{(j)}) = \sum_{i=1}^{n} a(R(\mathbf{Y}_i^{(j)} - \boldsymbol{x}_i'\mathscr{B}_1^{(j)}))(\mathbf{Y}_i^{(j)} - \boldsymbol{x}_i'\mathscr{B}_1^{(j)}), \tag{2.5}$$

where $a(i)$ are scores such that $a(1) \le a(2) \le \cdots \le a(n)$ and $\sum a(i) = 0$. The scores are generated by a score generating function φ as $a(i) = \varphi(i/(n+1))$. The most widely used scores are the Wilcoxon scores which can be generated by $\varphi(u) = 12^{1/2}(u - (1/2))$. The ranks are component-wise rankings on $\mathbf{Y}_i^{(j)} - \boldsymbol{x}_i'\mathscr{B}_1^{(j)}$, $j = 1, \ldots, d$. The intercept vector $\boldsymbol{\beta}_0$ is then computed as a location estimate of the residuals for each component. Thus the problem of estimating $\mathscr{B}$ is reduced to estimating $\mathscr{B}^{(j)}$ for each column separately. Under certain conditions, Davis and McKean showed that $\widehat{\mathscr{B}}_R = \left(\widehat{\boldsymbol{\beta}}_0\ \widehat{\mathscr{B}}'_{1,\varphi}\right)'$ is a highly efficient asymptotically normal estimator; see Section 4.2 for details. In the case of Wilcoxon scores, $\widehat{\mathscr{B}}_{1,\varphi}$ has an asymptotic relative efficiency (ARE) of 95% relative to the LS estimate.

2.3 GR-Estimation Step.

Model (2.2) can be written as

$$\mathbf{Y} = \mathbf{1}_n \boldsymbol{\alpha}_0' + \mathbf{X}_c \mathscr{B}_1 + \boldsymbol{\varepsilon}, \tag{2.6}$$

where $\boldsymbol{\alpha}_0' = \boldsymbol{\beta}_0' + \overline{\boldsymbol{x}}' \mathscr{B}_1$. The model in (2.6) is the concatenation of the d univariate linear models $\mathbf{Y}^{(k)} = \mathbf{1}_n \boldsymbol{\alpha}_0^{(k)} + \mathbf{X}_c \mathscr{B}_1^{(k)} + \boldsymbol{\varepsilon}^{(k)}$, $k = 1, \ldots, d$. Consider the function

$$\|\boldsymbol{u}\|_{GR} = \sum_{i<j} b_{ij} |u_i - u_j|, \tag{2.7}$$

where the weights b_{ij} are functions of the x_{ij}'s and are assumed to be positive and symmetric, i.e., $b_{ij} \equiv b_{ji}$. Note that this function is the Wilcoxon pseudonorm if the weights $b_{ij} \equiv 1$. The componentwise GR-estimate of $\mathscr{B}_1$ is a matrix $\widehat{\mathscr{B}}_{1,GR} = \left(\widehat{\mathscr{B}}_{1,GR}^{(1)} \cdots \widehat{\mathscr{B}}_{1,GR}^{(d)} \right)$, where $\widehat{\mathscr{B}}_{1,GR}^{(k)}$ minimizes $D_{GR}(\mathscr{B}_1^{(k)}) = \|\mathbf{Y}^{(k)} - \mathbf{X}_c \mathscr{B}_1^{(k)}\|_{GR}$ which is a continuous, nonnegative and convex function of $\mathscr{B}_1^{(k)}$. For $k = 1, \ldots, d$ the negative of the gradient of $D_{GR}(\mathscr{B}_1^{(k)})$ is given by

$$\boldsymbol{S}_{GR}^{(k)}(\mathscr{B}_1^{(k)}) = \sum_{i<j} b_{ij} (\boldsymbol{x}_i - \boldsymbol{x}_j) \operatorname{sgn}((\mathbf{Y}_i^{(k)} - \mathbf{Y}_j^{(k)}) - (\boldsymbol{x}_i - \boldsymbol{x}_j)' \mathscr{B}_1^{(k)}).$$

Define the statistic $\boldsymbol{S}_{GR}(\mathscr{B}_1) = \left(\boldsymbol{S}_{GR}^{(1)}(\mathscr{B}_1^{(1)}) \cdots \boldsymbol{S}_{GR}^{(d)}(\mathscr{B}_1^{(d)}) \right)$. Then the componentwise GR-estimate of $\mathscr{B}_1$ solves the estimating equations $\boldsymbol{S}_{GR}(\mathscr{B}_1) \doteqdot \mathbf{0}$ by solving the equations $\boldsymbol{S}_{GR}^{(k)}(\mathscr{B}_1^{(k)}) \doteqdot \mathbf{0}$, for $k = 1, \ldots, d$.

We have chosen to use high breakdown weights of the form $b_{ij} = b_i b_j$, where b_i is defined by

$$b_i = \min\{1, c/\sqrt{(\boldsymbol{x}_i - \mathbf{v})' \mathbf{V}^{-1} (\boldsymbol{x}_i - \mathbf{v})}\}^{\alpha}, \tag{2.8}$$

and $(\mathbf{v}, \mathbf{V})$ are the minimum covariance determinant (MCD) estimates of location and scatter; see Rousseeuw and Van Driessen (1999). In our work we set α at 1 and the parameter c at the 95th percentile of the χ^2 distribution with p degrees of freedom.

The univariate GR-estimates were proposed by Sievers (1983) and further developed by Naranjo and Hettmansperger (1994). In Section 4.3, we establish the asymptotic normality of these component-wise GR estimators under Model (2.1). Using the fast MCD algorithm of Rousseeuw and Van Driessen (1999) for the weights and a simple Gauss-Netwon algorithm to obtain the estimates, these GR estimates are quickly computed.

2.4 Equivariance Properties of the TRR and TRGR Estimators

The next theorem establishes the equivariance properties of the TRR estimator. A similar theorem is true for the TRGR estimator. Simply replace $\widehat{\mathscr{B}}_{TRR}$ by $\widehat{\mathscr{B}}_{TRGR}$ in the theorem.

Lemma 2.1. *The estimate $\widehat{\mathscr{B}}_{TRR} = \widehat{\mathscr{B}}_{TRR}(\boldsymbol{x}, \boldsymbol{y})$ satisfy the properties*

1. $\boldsymbol{y}$- scale equivariance. If k is a non zero scalar and $\mathbf{b}$ is a $d \times 1$ constant vector, then

$$\widehat{\mathscr{B}}_{TRR}(\boldsymbol{x}, k\boldsymbol{y} + \mathbf{b}) = \begin{pmatrix} k\widehat{\boldsymbol{\beta}}_{0,TRR}' + \mathbf{b}' \\ k\widehat{\mathscr{B}}_{1,TRR} \end{pmatrix}$$

2. regression equivariance. For any G a $p \times d$ matrix

$$\widehat{\mathscr{B}}_{TRR}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{0,TRR}' \\ \widehat{\mathscr{B}}_{1,TRR} - G \end{pmatrix}$$

3. $\boldsymbol{x}$- affine equivariance. For any fixed $p \times p$ nonsingular matrix W and a $p \times 1$ constant vector $\boldsymbol{c}$

$$\widehat{\mathscr{B}}_{TRR}(W\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y}) = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{0,TRR}' - \boldsymbol{c}'(W^{-1})' \widehat{\mathscr{B}}_{1,TRR} \\ (W^{-1})' \widehat{\mathscr{B}}_{1,TRR} \end{pmatrix}.$$

3 Example

The data of this example was collected by the Biological Sciences Division of Indian Statistical Institute, Calcutta and consists of systolic and diastolic blood pressures of 40 Marwari females resting at Burrabazar area of Calcutta and their ages; see Chakraborty and Chaudhuri (1997). There is a linear relationship between blood pressure and age. Also, there is a positive correlation between systolic and diastolic blood pressures. Let y_1 denote the systolic blood pressure and y_2 denote the diastolic blood pressure. Let x denote age. Then upon fitting this data set to model 2.1 with $d = 2$ and $p = 1$ we get the estimates

$$\widehat{\mathscr{B}}_{TRR} = \begin{pmatrix} 102.64(8.08, 5.62) & 73.35(4.13, 3.23) \\ 0.86(0.18, 0.21) & 0.35(0.10, 0.11) \end{pmatrix},$$

and

$$\widehat{\mathscr{B}}_{TRGR} = \begin{pmatrix} 102.67(6.18, 6.10) & 73.35(3.52, 3.23) \\ 0.86(0.17, 0.23) & 0.35(0.10, 0.11) \end{pmatrix}.$$

The first number in parenthesis is the estimated standard errors while the second number is the bootstrap standard error (based on 10,000 bootstrap replications). The estimated standard errors based on asymptotic theory and based on the bootstrap are very similar. Note that the R and GR fits are almost the same.

As a comparison, the estimate obtained by Chakraborty and Chaudhuri (1997) is

$$\widehat{\mathscr{B}}_{\text{Chk}} = \begin{pmatrix} 100.64 & 74.4 \\ 0.8 & 0.32 \end{pmatrix}.$$

Also, Chakraborty and Chaudhuri computed the standard errors of their estimate using a bootstrap technique. They obtained as standard errors of the coefficients of age for y_1 and y_2 the values 0.20 and 0.11 respectively. Their results are quite similar to those of the TRR estimate.

4 Theory

In this section, we obtain the asymptotic distributions of the of the TRR estimator $\widehat{\mathscr{B}}_{TRR}$ and the TRGR estimator $\widehat{\mathscr{B}}_{TRGR}$. They both depend on the consistency of the transformation matrix $\hat{A}$ which we show first.

The assumptions for the theory are:

A1. The rows of $\boldsymbol{\varepsilon}$ are iid with an absolutely continuous joint distribution function F and a continuous joint density function.
A2. The marginal distribution function F_j has a unique median at 0 and a differentiable density f_j with finite Fisher information.
A3.$\mathbf{X}$ and $\mathbf{X}_c$ are of full column rank.
A4. Huber's condition holds for $\mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'$.
A5.$\text{Cov}(\varphi(F_j(\varepsilon_{ij}), \varphi(F_{j'}(\varepsilon_{ij'})) = s_{jj'} < \infty$, for $j, j' = 1, \ldots, d$. Also, $\boldsymbol{S} = (s_{jj'})$ is positive definite.
A6.$\lim_{n\to\infty} \bar{\mathbf{X}}' = \mathbf{v}'$, where $\bar{\mathbf{X}}' = n^{-1}\mathbf{1}_n'\mathbf{X}$.
A7.$\lim_{n\to\infty} n^{-1/2}(\mathbf{X}_c'\mathbf{X}_c)^{1/2} = \mathbf{V}^{1/2}$, where $\mathbf{V}^{1/2}$ is finite positive definite.
A8. For $d > 1$, $E\left(\frac{1}{\|A\boldsymbol{\varepsilon}\|}\right) < \infty$.

The major assumption on the design matrix is Huber's condition (A4), which is the assumption required for LS asymptotic theory. This assumption and the others in A1-A7 are the same as required for the multivariate R regression estimators of Davis and McKean (1993). For $d > 1$ assumption A.8 is needed for the consistency of the transformation matrix $\hat{A}$. This holds for many elliptical multivariate distributions, including the multivariate normal.

4.1 Consistency of $\hat{A}$

Consider Model (2.1). Recall that our transformation matrix $\hat{A}$ is a function of the LS residuals $\widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_n$ and that it is the unique solution of equation (2.3). For the true errors of Model (2.1), $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$, Tyler (1987) showed that there exists unique A such that

$$E\left(\frac{A\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'A'}{\|A\boldsymbol{\varepsilon}\|^2}\right) = \frac{1}{d}I,$$

where $\boldsymbol{\varepsilon}$ is the $n \times d$ matrix of true errors. Let A^* solve the equation,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A\boldsymbol{\varepsilon}_i}{\|A\boldsymbol{\varepsilon}_i\|}\right)\left(\frac{A\boldsymbol{\varepsilon}_i}{\|A\boldsymbol{\varepsilon}_i\|}\right)' = \frac{1}{d}I \tag{4.1}$$

Then Tyler (1987) showed that $\|A^* - A\|_F = o_p(1)$, where $\|.\|_F$ denotes the Frobenius norm on matrices. Then, as proved in Appendix A.2, the same result holds for our estimate $\hat{A}$; i.e.,

Theorem 4.1. *Under conditions A4 and A7–A8,* $\|\hat{A} - A\|_F = o_p(1)$.

4.2 Asymptotic Normality of $\widehat{\mathscr{B}}_{TRR}$

Consider Model (2.1). Let $\hat{A}$ be the transformation matrix based on the LS residuals for Model (2.1). As in Section 2, denote the transformed responses by $\boldsymbol{z}_i = \hat{A}\boldsymbol{y}_i$ for $i = 1,\dots,n$. Suppose that a score function $\varphi(u)$ has been specified. Assume that $\varphi(u)$ is a square integrable, nondecreasing function defined on $(0,1)$ and is standardized so the $\int\varphi(u)\,du = 0$ and $\int\varphi^2(u)\,du = 1$. Denote the R scores by $a(i) = \varphi(i/(n+1))$. Define the $n \times d$ matrix of scored ranks by

$$A_S(Z - X\mathscr{B}_1) = \left[a\left(R\left(z_i^{(j)} - x_i'\beta^{(j)}\right)\right)\right], \tag{4.2}$$

where $R\left(z_i^{(j)} - x_i'\beta^{(j)}\right)$ denotes the rank of $z_i^{(j)} - x_i'\beta^{(j)}$ among the n residuals $z_1^{(j)} - x_1'\beta^{(j)},\ \dots,\ z_n^{(j)} - x_n'\beta^{(j)}$, i.e., component-wise rankings. Then the R-estimator of the transformed variables is

$$\widehat{\mathscr{B}}_{1,\varphi} = \text{Argmin}\, D(\mathscr{B}_1) = \text{Argmin}\,\text{tr}\,(\mathbf{Z} - \mathbf{X}\mathscr{B}_1)'\mathbf{A}_S(\mathbf{Z} - \mathbf{X}\mathscr{B}_1). \tag{4.3}$$

The negative of the gradient of $D(\mathscr{B}_1)$ is

$$\mathbf{L}(\mathscr{B}_1) = \mathbf{X}'\mathbf{A}_S(\mathbf{Z} - \mathbf{X}\mathscr{B}_1). \tag{4.4}$$

Then equivalently, $\widehat{\mathscr{B}}_{1,\varphi}$ solves the equation $\mathbf{L}(\mathscr{B}_1) = \mathbf{0}$. Once $\widehat{\mathscr{B}}_{1,\varphi}$ is obtained, we estimate the vector of intercept parameters $\widehat{\beta}_{0,s}$ by component-wise location estimations based on the residuals. In this paper, we will only consider the median of the residuals. Hence, for $j = 1,2,\dots,d$

$$\widehat{\beta}_{0,s}^{(j)} = \text{med}_{1\le i\le n}\{Y_i^{(j)} - \mathbf{x}_i'\widehat{\mathscr{B}}_{1,\varphi}^{(j)}\}, \tag{4.5}$$

and let $\widehat{\beta}_{0,s} = \left(\widehat{\beta}_{0,s}^{(1)} \dots \widehat{\beta}_{0,s}^{(d)}\right)$. Stacking the intercept estimators and the regression estimators together and then transforming back, we get the TRR estimator Stacking the intercept estimators and the regression estimators together and then transforming back, we get the TRR estimator

$$\widehat{\mathscr{B}}_{TRR} = \begin{pmatrix}\widehat{\beta}_{0,s}' \\ \widehat{\mathscr{B}}_{1,\varphi}\end{pmatrix}(\hat{A}')^{-1}. \tag{4.6}$$

As theorem 4.2 shows $\widehat{\mathscr{B}}_{TRR}$ has an asymptotic normal distribution. To state the asymptotic covariance structure, we need some additional notation.

If D is an $m \times n$ matrix then by vec(D) we mean the $mn \times 1$ vector formed by stacking the columns of D. Let A be an $m_2 \times n_2$ matrix and let B be an $m_1 \times n_1$; then $A \otimes B$ denote the left direct product of A and B, see Graybill (1983). Define the scale parameter τ_j by $\tau_j^{-1} = \int_0^1 \varphi(u)\varphi(u, f_j)\,du$, where $\varphi(u, f_j) = -\frac{f_j'(F_j^{-1}(u))}{f_j(F_j^{-1}(u))}$. Also, let $\tau_j^* = \frac{1}{2f_j(0)}$. Let $\mathbf{T}, \mathbf{T}^*$ be a $d \times d$ diagonal matrices whose j^{th} diagonal element is τ_j, τ_j^* respectively. The assumption on f_j to have finite Fisher information ensures that τ_j is finite. Let $\boldsymbol{S} = (s_{jj'})$ where $s_{jj'} = \text{Cov}(\varphi(F_j(\varepsilon_{ij}), \varphi(F_{j'}(\varepsilon_{ij'}))$ and let $\boldsymbol{S}^* = (s_{jj'}^*)$ where $s_{jj'}^* = \Pr(\varepsilon_{ij} < 0, \varepsilon_{ij'} < 0) + \Pr(\varepsilon_{ij} > 0, \varepsilon_{ij'} > 0) - \Pr(\varepsilon_{ij} < 0, \varepsilon_{ij'} > 0) - \Pr(\varepsilon_{ij} > 0, \varepsilon_{ij'} < 0)$.

Theorem 4.2. *Under assumptions A1-A8*

$$\sqrt{n}\,\text{vec}(\widehat{\mathscr{B}}_{TRR} - \mathscr{B})' \xrightarrow{\mathscr{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix}\boldsymbol{S}_{11,TRR} & \boldsymbol{S}_{12,TRR} \\ \boldsymbol{S}_{21,TRR} & \boldsymbol{S}_{22,TRR}\end{pmatrix}\right) \tag{4.7}$$

where

$$\begin{aligned}\boldsymbol{S}_{11,TRR} &= A^{-1}(\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + (\mathbf{v}'\mathbf{V}^{-1}\mathbf{v})\mathbf{T}\boldsymbol{S}\mathbf{T})(A^{-1})' \\ \boldsymbol{S}_{12,TRR} &= A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes \mathbf{v}'\mathbf{V}^{-1} \\ \boldsymbol{S}_{21,TRR} &= A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes \mathbf{V}^{-1}\mathbf{v} \\ \boldsymbol{S}_{22,TRR} &= A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes \mathbf{V}^{-1},\end{aligned}$$

where $\mathbf{v}$ *and* $\mathbf{V}$ *are defined by Assumptions A6 and A7, respectively.*

The proof of this theorem is given in the appendix, Section A.3. Using the results from Peters and Randles(1990), a basic contiguity result was established by Davis and McKean (1993). This was used to obtain the asymptotic properties of the gradient of the dispersion function for the true and local distributions of the errors. This lead to asymptotic linearity and quadraticity results from which the proof of Theorem 4.2 follows. Also it induces the following influence function of $\widehat{\mathscr{B}}_{TRR}$ as:

Corollary 4.1.*The influence function of* $\widehat{\mathscr{B}}_{1,TRR}$ *is given by*

$$\boldsymbol{\Omega}(\boldsymbol{x}_0,\boldsymbol{y}_0,\widehat{\mathscr{B}}_{1,TRR}) = \mathbf{V}^{-1}\boldsymbol{x}_0\left(\varphi(F_1(A\boldsymbol{y}_0)_1)\ \cdots\ \varphi(F_d(A\boldsymbol{y}_0)_d)\right)\mathbf{T}(A^{-1})' \tag{4.8}$$

Note that, given A, the influence function of the TRR estimate is bounded in the $\boldsymbol{y}$-space but not in the $\boldsymbol{x}$-space.

4.3 Asymptotic Normality of $\widehat{\mathscr{B}}_{TRGR}$

Define the weight matrix $\mathbf{W} = (w_{ij})_{n\times n}$ as

$$w_{ij} = \begin{cases} -\frac{1}{n}b_{ij} & i \neq j \\ \frac{1}{n}\sum_{k\neq i} b_{ik} & i = j. \end{cases}$$

Then $\mathbf{W}$ is symmetric and its rows sum to zero. In addition to conditions A1–A4 and A6–A7 needed for the asymptotic theory of the R-estimate we need to assume the following

B1.$\lim_{n\to\infty}\frac{1}{n}\mathbf{X}'\mathbf{W}\mathbf{X} = \mathbf{C}$, $\mathbf{C} > \mathbf{0}$.
B2.$\lim_{n\to\infty}\frac{1}{n}\mathbf{X}'\mathbf{W}^2\mathbf{X} = \mathbf{E}$, $\mathbf{E} > \mathbf{0}$.
B3.WX satisfies Huber's condition.

Asymptotic properties of $\widehat{\mathscr{B}}_{TRGR}$ follow from the corresponding properties of the componentwise estimate $\widehat{\mathscr{B}}_{GR}$. Let $\boldsymbol{S} = (s_{jj'})$ where $s_{jj'} = \mathrm{Cov}(2F_j(\mathbf{Y}_1^{(j)}), 2F_{j'}(\mathbf{Y}_1^{(j')}))$. Also, let $\boldsymbol{S}^*$, $\mathbf{T}$ and $\mathbf{T}^*$ be as defined in Section 4.2. For the TRGR estimate, we first establish the asymptotic distribution of the component-wise GR estimate.

Theorem 4.3.

$$\begin{aligned}&\sqrt{n}\,\mathrm{vec}\left(\widehat{\boldsymbol{\beta}}_{0,GR} - \boldsymbol{\beta}_0\ \widehat{\mathscr{B}}'_{1,GR} - \mathscr{B}'_1\right) \\ &\xrightarrow{\mathscr{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix} \mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + 3(\overline{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\overline{\boldsymbol{x}})\mathbf{T}\boldsymbol{S}\mathbf{T} & -\mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\overline{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1} \\ -\mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\overline{\boldsymbol{x}} & \mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\end{pmatrix}\right).\end{aligned} \tag{4.9}$$

The asymptotic distribution of the TRGR estimator is given in the next theorem.

Theorem 4.4.

$$\sqrt{n}\,\mathrm{vec}(\widehat{\mathscr{B}}_{TRGR} - \mathscr{B})' \xrightarrow{\mathscr{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix}\boldsymbol{S}_{11,TRGR} & \boldsymbol{S}_{12,TRGR} \\ \boldsymbol{S}_{21,TRGR} & \boldsymbol{S}_{22,TRGR}\end{pmatrix}\right) \tag{4.10}$$

where

$$\begin{aligned}\boldsymbol{S}_{11,TRGR} &= A^{-1}(\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + 3(\overline{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\overline{\boldsymbol{x}})\mathbf{T}\boldsymbol{S}\mathbf{T})(A^{-1})' \\ \boldsymbol{S}_{12,TRGR} &= -A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes 3\overline{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1} \\ \boldsymbol{S}_{21,TRGR} &= -A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\overline{\boldsymbol{x}} \\ \boldsymbol{S}_{22,TRGR} &= A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\end{aligned}$$

The proof of this theorem proceeds similar to that of the asymptotic distribution of the TRR estimator. The proof is given in Appendix **??**. Part of the proof results in an asymptoic representation of the TRGR estimator which leads to its influence function.

Corollary 4.2.

$$\sqrt{n}\,\mathrm{vec}(\widehat{\mathscr{B}}_{1,TRGR}-\mathscr{B}_1)\xrightarrow{\mathscr{D}} N_{pd}(\mathbf{0},3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\otimes A^{-1}\mathbf{TST}(A^{-1})'). \tag{4.11}$$

The next corollary shows that $\widehat{\mathscr{B}}_{1,TRGR}$ has a bounded influence in both $\boldsymbol{x}$-space and $\boldsymbol{y}$-space.

Corollary 4.3.*The influence function of* $\widehat{\mathscr{B}}_{1,TRGR}$ *is*

$$\begin{aligned} IF(\boldsymbol{x}_0,\boldsymbol{y}_0,\widehat{\mathscr{B}}_{1,TRGR}) &= \sqrt{12}\,\mathbf{C}^{-1}\int(\boldsymbol{x}-\boldsymbol{x}_0)\,dM(\boldsymbol{x}) \\ &\times\left(F_1(\boldsymbol{y}_0^{(1)}-\boldsymbol{\alpha}_0^{(1)}-\boldsymbol{x}_0'\mathscr{B}_1^{(1)})-\tfrac{1}{2}\cdots F_d(\boldsymbol{y}_0^{(d)}-\boldsymbol{\alpha}_0^{(d)}-\boldsymbol{x}_0'\mathscr{B}_1^{(d)})-\tfrac{1}{2}\right) \\ &\times\mathbf{T}(A^{-1})'. \end{aligned}$$

Given the matrix A, with a proper choice of weights, the influence function of the TRGR estimator is bounded in both the $\boldsymbol{y}$- and the $\boldsymbol{x}$-spaces.

5 Simulations

In our simulation study a comparison is made between the performance of $\widehat{\mathscr{B}}_{TRR}$ and $\widehat{\mathscr{B}}_{TRGR}$ and the other procedures in the literature. In this study we used model (2.1) with $d=2$ and $p=1$. That is we used the multivariate linear model

$$\begin{pmatrix} y_{i1}\\ y_{i2}\end{pmatrix}=\begin{pmatrix}\beta_{01}\\ \beta_{02}\end{pmatrix}+x_i\begin{pmatrix}\beta_{11}\\ \beta_{12}\end{pmatrix}+\begin{pmatrix}\varepsilon_{i1}\\ \varepsilon_{i2}\end{pmatrix}. \tag{5.1}$$

The parameter matrix $\mathscr{B}$ was set to zero. The regressors x_i were generated as a random sample from $N(0,1)$ and the independent errors from elliptically symmetric distributions, i.e. distributions having a density proportional to

$$(\det\boldsymbol{\Sigma})^{-1/2}h(\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}). \tag{5.2}$$

From this class of distributions we included in the study the bivariate normal, bivariate contaminated normal, bivariate t with 3 degrees of freedom and bivariate Cauchy. The study also covered the case where the errors have the elliptical bivariate Laplace distribution. This distribution has the spherical density

$$h(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})=\frac{1}{2\pi}\exp(-\sqrt{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}). \tag{5.3}$$

Further, we used the covariance matrix

$$\boldsymbol{\Sigma}=\begin{pmatrix}1&\rho\\ \rho&1\end{pmatrix} \tag{5.4}$$

where we chose values for ρ between 0.00 and 0.95.

We considered the LS, LAD, R (Wilcoxon scores), TRR (Wilcoxon scores), and TRGR estimates. The TRR and TRGR estimates were computed as described in the algorithm of Section 2. The Wicoxon estimates were computed by the RGLM algorithm (see Hettmansperger and McKean, (2011), and the LAD estimate was computed by the algorithm of Armstrong and Kung (1978).

In this study $\widehat{\mathscr{B}}_{TRR}$ and $\widehat{\mathscr{B}}_{TRGR}$ were compared to the LS, LAD and the corresponding componentwise estimate. The finite sample efficiencies were computed as the fourth root of the ratios of the generalized variances of the estimates. See Bickel 1964. The study was run for 3000 Monte Carlo replications and for a sample size $n=30$. A similar simulation study was conducted by Chakraborty (1997) and Oja (2002).

From our results, tables 1–8, we observe that the performance of the componentwise estimators decreases as the correlation among the response variables increases. This is true regardless of the distribution of the errors. In contrast, relatively, the TRR and the TRGR estimators are increasingly more efficient than the component-wise estimators as ρ increases across all distributions. Note that the TRR estimator is more efficient than the LAD estimator for all the bivariate laplace distributions. However, $\widehat{\mathscr{B}}_{TRR}$ performance is better than $\widehat{\mathscr{B}}_{TRGR}$ because the matrix

$(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^2\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}$ is positive semi-definite; see Hettmansperger and McKean (2011). The LS estimator performs best for multivariate normal errors. However, for every other distribution the TRR estimator was more efficient than the LS estimator over all situations. The same is true for the TRGR estimator.

Compared to Chakraborty's result our estimate has a higher efficiency for heavy tailed distributions and has almost similar efficiency as Oja's estimate for these distributions.

In order to study the effect of the initial estimate on the proposed estimators we ran the same simulation study but using robust initial estimates like MCD and WTLMCD. The results showed that the TRR and the TRGR were not affected by the LS initial estimation step except when the errors have Cauchy distribution.

Table 1: $\widehat{\mathscr{B}}_{TRR}$ estimated relative efficiencies under bivariate normal errors

ρ	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{R}$)
0.00	0.809	1.215	0.986
0.20	0.798	1.240	1.000
0.50	0.791	1.330	1.031
0.75	0.799	1.577	1.132
0.80	0.804	1.684	1.176
0.85	0.799	1.758	1.227
0.90	0.791	1.985	1.290
0.95	0.805	2.377	1.431

Table 2: $\widehat{\mathscr{B}}_{TRGR}$ estimated relative efficiencies under bivariate normal errors

ρ	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{GR}$)
0.00	0.787	1.181	0.984
0.20	0.770	1.197	0.999
0.50	0.769	1.293	1.040
0.75	0.770	1.520	1.140
0.80	0.776	1.625	1.188
0.85	0.781	1.719	1.251
0.90	0.773	1.940	1.310
0.95	0.782	2.310	1.452

Table 3: $\widehat{\mathscr{B}}_{TRR}$ estimated relative efficiencies under bivariate laplace errors

ρ	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{R}$)
0.00	1.119	1.104	1.001
0.20	1.117	1.104	1.005
0.50	1.114	1.178	1.050
0.75	1.127	1.413	1.180
0.80	1.105	1.459	1.167
0.85	1.109	1.584	1.235
0.90	1.103	1.750	1.300
0.95	1.105	2.084	1.455

Table 4: $\widehat{\mathscr{B}}_{TRGR}$ estimated relative efficiencies under bivariate laplace errors

ρ	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{GR}$)
0.00	1.090	1.075	1.000
0.20	1.088	1.075	1.010
0.50	1.071	1.132	1.054
0.75	1.095	1.373	1.185
0.80	1.073	1.417	1.178
0.85	1.075	1.536	1.247
0.90	1.071	1.699	1.320
0.95	1.079	2.035	1.479

Table 5: $\widehat{\mathscr{B}}_{TRR}$ estimated relative efficiencies under bivariate $t(3)$ errors

ρ	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{R}$)
0.00	1.676	1.082	0.993
0.20	1.677	1.089	0.992
0.50	1.700	1.133	1.007
0.75	1.757	1.367	1.147
0.80	1.766	1.441	1.178
0.85	1.631	1.592	1.242
0.90	1.798	1.792	1.332
0.95	1.899	2.117	1.481

Table 6: $\widehat{\mathscr{B}}_{TRGR}$ estimated relative efficiencies under bivariate $t(3)$ errors

ρ	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{GR}$)
0.00	1.627	1.050	0.991
0.20	1.628	1.058	0.989
0.50	1.661	1.108	1.013
0.75	1.701	1.323	1.151
0.80	1.697	1.384	1.176
0.85	1.573	1.535	1.257
0.90	1.741	1.735	1.337
0.95	1.842	2.055	1.494

Table 7: $\widehat{\mathscr{B}}_{TRR}$ estimated relative efficiencies under bivariate cauchy errors

ρ	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRR}$,$\widehat{\mathscr{B}}_{R}$)
0.00	1941.31	0.741	0.842
0.20	1450.32	0.659	0.752
0.50	940.462	0.606	0.668
0.75	2233.78	0.939	1.019
0.80	1935.99	0.917	0.946
0.85	1136.22	1.092	1.115
0.90	1266.34	1.208	1.207
0.95	2378.34	1.435	1.272

Table 8: $\widehat{\mathscr{B}}_{TRGR}$ estimated relative efficiencies under bivariate cauchy errors

ρ	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LS}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{LAD}$)	ARE($\widehat{\mathscr{B}}_{TRGR}$,$\widehat{\mathscr{B}}_{GR}$)
0.00	1903.08	0.726	0.840
0.20	1451.27	0.659	0.779
0.50	1065.54	0.686	0.773
0.75	2221.51	0.934	1.026
0.80	1937.74	0.918	0.959
0.85	1132.01	1.088	1.112
0.90	1239.29	1.182	1.206
0.95	2339.96	1.412	1.300

Table 9: Finite Sample Efficiencies of TRR Relative to LAD

		Initial Estimator		
Distribution	rho	LS	MCD	WTLMCD
Normal	0.75	1.564	1.538	1.556
	0.80	1.620	1.601	1.604
	0.85	1.764	1.763	1.748
	0.90	1.962	1.923	1.932
	0.95	2.306	2.283	2.2890
Laplace	0.75	1.521	1.485	1.509
	0.80	1.572	1.556	1.560
	0.85	1.725	1.708	1.708
	0.90	1.905	1.865	1.870
	0.95	2.253	2.234	2.230
t_3	0.75	1.364	1.369	1.380
	0.80	1.485	1.482	1.486
	0.85	1.571	1.570	1.578
	0.90	1.798	1.824	1.821
	0.95	2.110	2.140	2.122
Cauchy	0.75	0.972	1.172	1.183
	0.80	1.00	1.223	1.228
	0.85	1.157	1.332	1.340
	0.90	1.249	1.470	1.472
	0.95	1.516	1.898	1.905

Table 10: Finite Sample Efficiencies of TRGR Relative to LAD

		Initial Estimator		
Distribution	rho	LS	MCD	WTLMCD
Normal	0.75	1.521	1.485	1.509
	0.80	1.572	1.556	1.560
	0.85	1.725	1.708	1.708
	0.90	1.905	1.865	1.870
	0.95	2.253	2.234	2.230
Laplace	0.75	1.336	1.335	1.328
	0.80	1.461	1.442	1.438
	0.85	1.524	1.511	1.498
	0.90	1.706	1.702	1.693
	0.95	2.043	2.027	2.034
t_3	0.75	1.314	1.311	1.325
	0.80	1.438	1.430	1.436
	0.85	1.532	1.526	1.531
	0.90	1.731	1.754	1.756
	0.95	2.038	2.059	2.053
Cauchy	0.75	0.938	1.138	1.141
	0.80	0.986	1.178	1.185
	0.85	1.120	1.328	1.342
	0.90	1.246	1.458	1.465
	0.95	1.522	1.875	1.882

6 Conclusion

In this article, we have proposed estimators for multivariate linear models based on ranks and generalized ranks. They are transformation and retransformation type estimators. The matrix of responses is first transformed using Tyler's (1987) multivariate scatter matrix, based on residuals. Then the rank-based estimates of Davis and McKean (1993) are obtained on the transformed data. These are then retransformed to obtain the final estimates. Chakraborty and Chauduri's (1997) transformed-retransformed estimates depend on a preselected subset of the data. Our estimates, however, use all the data in the transformation step, no preselection is necessary. We have introduced both a highly efficient estimator (TRR) which bounds influence in the response space and a bounded influence estimator (TRGR) which bounds influence in both the response and factor spaces.

We developed the asymptotic distribution theory for both the TRR and the TRGR estimators. The theory results in asymptotic variances and covariances for which consistent robust estimators are available. The asymptotic theory depends on the consistency of Tyler's (1987) scatter matrix based on residuals which we also proved.

We presented the results of a Monte Carlo study over a variety of error distributions and correlation structures. These studies confirmed previous results of the slippage of efficiency of componentwise robust estimators as correlation increases. The TRR estimator showed high efficiency over all the situations. It was more efficient than the LS estimates over all situations other than the multivariate normal. It was much more efficient for the heavy tailed error distributions including an elliptical multivariate Cauchy. It was more efficient than the LAD estimator over all situations including a multivariate Laplace except for the multivariate Cauchy; however, even here it was more efficient at the high correlation situations. The TRGR estimator although less efficient than the TRR estimator still displayed good empirical efficiencies in the study.

We presented the results of several examples showing the practicality of our estimators. We confirmed the consistent estimates of the asymptotic standard errors of the estimators with bootstrap (using 3000 replications) estimates of the variance-covariance matrices. The study also showed that the LS initial estimation step has a little effect on the efficinecy of the estimators.

In summary, we recommend the use of the TRR and TRGR estimators for multivariate linear models. The TRR estimator is a highly efficient estimator while the TRGR estimator is a bonded influence estimator. They are quickly and efficiently computed using R/SPLUS routines or standard fortran routines. There is only one concern about the proposed estimates which is they are not fully affine equivariant as they lack rotation equivariance.

Appendix

A.1 Computing $\hat{A}$

The following steps gives the computational algorithm for $\hat{A}$ on the LS residuals $\hat{\boldsymbol{\varepsilon}}_{i,LS}$.

Step I. Compute

$$\boldsymbol{S}_0 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\boldsymbol{\varepsilon}}_{i,LS}}{\|\hat{\boldsymbol{\varepsilon}}_{i,LS}\|}\right)\left(\frac{\hat{\boldsymbol{\varepsilon}}_{i,LS}}{\|\hat{\boldsymbol{\varepsilon}}_{i,LS}\|}\right)' \tag{A.1.1}$$

and form $\hat{A}_0 = \mathrm{Chol}(\boldsymbol{S}_0^{-1})$, where $\mathrm{Chol}(M)$ denotes the upper triangular Cholesky factorization of the positive definite matrix M, divided by the upper-left element of that upper triangular matrix.

Step II. At the t^th iteration, form

$$\hat{A}_{dt} = \hat{A}_{t-1}\hat{A}_{t-2}\cdots\hat{A}_0, \tag{A.1.2}$$

and

$$\boldsymbol{S}_t = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{A}_{dt}\hat{\boldsymbol{\varepsilon}}_{i,LS}}{\|\hat{A}_{dt}\hat{\boldsymbol{\varepsilon}}_{i,LS}\|}\right)\left(\frac{\hat{A}_{dt}\hat{\boldsymbol{\varepsilon}}_{i,LS}}{\|\hat{A}_{dt}\hat{\boldsymbol{\varepsilon}}_{i,LS}\|}\right)'. \tag{A.1.3}$$

Step III. If $\|\boldsymbol{S}_t - \frac{1}{d}I\|$ is sufficiently small, then stop and set $\hat{A} = \hat{A}_{dt}$. If not, then compute $\hat{A}_t = \mathrm{Chol}(\boldsymbol{S}_t^{-1})$ and go back to step II.

A.2 Consistency of $\hat{A}$

In this section of the appendix we obtain the proof of Theorem 4.1 which shows that the matrix $\hat{A}$ based on LS residuals is a consistent estimator of the matrix A defined in expression (4.1).

Before we establish the consistency of $\hat{A}$ we need to introduce the following notation. Let $W_n = \mathbf{X}'\mathbf{X}$ where in the following context $\mathbf{X} = \mathbf{X}_c$. Also, let $\boldsymbol{x}_{ni} = W_n^{-1/2}\boldsymbol{x}_i$, $\mathscr{B}_{1,n} = W_n^{1/2}\mathscr{B}_1$, $\boldsymbol{y}_{ni} = \boldsymbol{y}_i$ and $\boldsymbol{\varepsilon}_{ni} = \boldsymbol{\varepsilon}_i$. Now, recall model (2.1)

$$\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \mathscr{B}_1'\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i, \quad i = 1,\ldots,n.$$

For simplicity we may assume wlog that $\boldsymbol{\beta}_0 = \mathbf{0}$. Then in terms of the new notation we can write this model as

$$\boldsymbol{y}_{ni} = \mathscr{B}_{1,n}'\boldsymbol{x}_{ni} + \boldsymbol{\varepsilon}_{ni}, \quad i = 1,\ldots,n.$$

Let $\mathscr{B}^*_{1,n,LS}$ be the LS estimate of $\mathscr{B}_{1,n}$ and $\widehat{\mathscr{B}}_{1,LS}$ be the estimate of $\mathscr{B}_1$. Assuming the true parameter $\mathscr{B}_1 = \mathbf{0}$, we can write the residuals $\hat{\boldsymbol{\varepsilon}}_{ni,LS}$ for $i = 1,\dots,n$ as

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}_{ni,LS} &= \boldsymbol{y}_{ni} - \hat{\boldsymbol{y}}_{ni,LS}\\ &= \mathscr{B}'_{1,n}\boldsymbol{x}_{ni} + \boldsymbol{\varepsilon}_{ni} - (\mathscr{B}^*_{1,n,LS})'\boldsymbol{x}_{ni}\\ &= \boldsymbol{\varepsilon}_{ni} - (\mathscr{B}^*_{1,n,LS})'\boldsymbol{x}_{ni}.\end{aligned}$$

Further let $d^*_n = \max_{1\le i\le n} \|\boldsymbol{x}_{ni}\|$, then by A4

$$\lim_{n\to\infty} d^*_n = \lim_{n\to\infty} \max_{1\le i\le n} \boldsymbol{x}'_{ni}\boldsymbol{x}_{ni} = \lim_{n\to\infty} \max_{1\le i\le n} \boldsymbol{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_i = 0.$$

Lemma A.2.1. Under A7,

$$\mathscr{B}^*_{1,n,LS} = O_p(1).$$

We need the following definitions on the norm of matrices; see Golub et al (1983).

Definition A.2.1. For an $m\times n$ matrix $\mathbf{C}$ define

1.The Frobenius norm of $\mathbf{C}$

$$\|\mathbf{C}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}^2\right)^{1/2}$$

2.The 2-norm of $\mathbf{C}$

$$\|\mathbf{C}\| = \|\mathbf{C}\|_2 = \sup_{\boldsymbol{u}\neq 0}\frac{\|\mathbf{C}\boldsymbol{u}\|}{\|\boldsymbol{u}\|},$$

where $\|\boldsymbol{u}\| = (u_1^2+\cdots+u_d^2)^{1/2}$.

The 2-norm and Frobenius norms are related through the inequality $\|C\| \le \|C\|_F$. Further, recall the inequality $\|\mathbf{CD}\|_F \le \|\mathbf{C}\|_F\|\mathbf{D}\|_F$. Using these results and some linear algebra, the following lemma can be proved.

Lemma A.2.2. For any $\boldsymbol{a}, \boldsymbol{b} \in \Re^d$, $\boldsymbol{b} \neq \mathbf{0}$, $\boldsymbol{a} \neq \boldsymbol{b}$ and $\boldsymbol{a} \perp (\boldsymbol{b}-\boldsymbol{a})$

$$\left\|\frac{(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{a})'}{\|\boldsymbol{b}-\boldsymbol{a}\|^2} - \frac{\boldsymbol{b}\boldsymbol{b}'}{\|\boldsymbol{b}\|^2}\right\|_F \le 4\frac{\|\boldsymbol{a}\|}{\|\boldsymbol{b}\|}.$$

Proof of Theorem 4.1 Let

$$\boldsymbol{S}_n(A) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A\hat{\boldsymbol{\varepsilon}}_{ni,LS}}{\|A\hat{\boldsymbol{\varepsilon}}_{ni,LS}\|}\right)\left(\frac{A\hat{\boldsymbol{\varepsilon}}_{ni,LS}}{\|A\hat{\boldsymbol{\varepsilon}}_{ni,LS}\|}\right)' \tag{A.2.4}$$

and

$$\boldsymbol{S}(A) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A\boldsymbol{\varepsilon}_i}{\|A\boldsymbol{\varepsilon}_i\|}\right)\left(\frac{A\boldsymbol{\varepsilon}_i}{\|A\boldsymbol{\varepsilon}_i\|}\right)' \tag{A.2.5}$$

be as defined by (2.3) when the average is taken over the LS residuals and the true errors, respectively. Tyler proved the consistency of $\hat{A}$ when it is computed on a random sample. For the argument of his proof to be applied to our case in which we compute $\hat{A}$ on the LS residuals we only need to show that

$$\|\boldsymbol{S}_n(A_{dt}) - \boldsymbol{S}(A_{dt})\|_F = o_p(1).$$

In his paper, Tyler mentioned that application of his Theorems 2.1 and 2.2 to a continuous population insures the existance of a unique matrix A such that A_{dt} of the algorithm converges to A and $\boldsymbol{S}(A) = \frac{1}{d}I$. Thus, $\forall\eta > 0$ and $i = 1,\dots,n$ we have for large t

$$\|A_{dt}\boldsymbol{\varepsilon}_{ni}\| \ge \|A\boldsymbol{\varepsilon}_{ni}\| - \eta.$$

Since $d^*_n \to 0$, we can choose a sequence of positive constants (v_n) such that

$$\lim_{n\to\infty} v_n = \infty, \qquad \lim_{n\to\infty} v_n d^*_n = 0. \tag{A.2.6}$$

Recall that $\hat{\boldsymbol{\varepsilon}}_{ni,LS} = \boldsymbol{\varepsilon}_{ni} - (\mathscr{B}^*_{1,n,LS})'\mathbf{x}_{ni}$ and note that the fitted values $(\mathscr{B}^*_{1,n,LS})'\mathbf{x}_{ni}$ are orthogonal to the residuals $\hat{\boldsymbol{\varepsilon}}_{ni}$. Now using the result of the last lemma we have

$$\begin{aligned}
\|\boldsymbol{S}_n(A_{dt}) - \boldsymbol{S}(A_{dt})\|_F &= \left\| \frac{1}{n}\sum_{i=1}^n \frac{A_{dt}\hat{\boldsymbol{\varepsilon}}_{ni,LS}\hat{\boldsymbol{\varepsilon}}'_{ni,LS}A'_{dt}}{\|A_{dt}\hat{\boldsymbol{\varepsilon}}_{ni,LS}\|^2} - \sum_{i=1}^n \frac{A_{dt}\boldsymbol{\varepsilon}_{ni}\boldsymbol{\varepsilon}'_{ni}A'_{dt}}{\|A_{dt}\boldsymbol{\varepsilon}_{ni}\|^2}\right\|_F \\
&\le \frac{1}{n}\sum_{i=1}^n \left\| \frac{A_{dt}\hat{\boldsymbol{\varepsilon}}_{ni,LS}\hat{\boldsymbol{\varepsilon}}'_{ni,LS}A'_{dt}}{\|A_{dt}\hat{\boldsymbol{\varepsilon}}_{ni,LS}\|^2} - \frac{A_{dt}\boldsymbol{\varepsilon}_{ni}\boldsymbol{\varepsilon}'_{ni}A'_{dt}}{\|A_{dt}\boldsymbol{\varepsilon}_{ni}\|^2}\right\|_F \\
&\le \frac{4}{n}\sum_{i=1}^n \frac{\|A_{dt}(\mathscr{B}^*_{1,n,LS})'\mathbf{x}_{ni}\|}{\|A_{dt}\boldsymbol{\varepsilon}_{ni}\|} \\
&\le 4\|A_{dt}\|_F\,\|\mathscr{B}^*_{1,n,LS}\|_F\, d^*_n \frac{1}{n}\sum_{i=1}^n \frac{1}{\|A_{dt}\boldsymbol{\varepsilon}_{ni}\|} \\
&\le 4(\|A\|_F + \eta)d^*_n v_n \frac{1}{n}\sum_{i=1}^n \frac{1}{\|A\boldsymbol{\varepsilon}_{ni}\| - \eta} \\
&\quad + 4(\|A\|_F + \eta)d^*_n I(\|\mathscr{B}^*_{1,n,LS}\|_F \ge v_n)\frac{1}{n}\sum_{i=1}^n \frac{1}{\|A\boldsymbol{\varepsilon}_{ni}\| - \eta}
\end{aligned}$$

where $I(.)$ denotes the indicator function. Now let $\eta \to \mathbf{0}$, since $E(\|A\boldsymbol{\varepsilon}\|^{-1}) < \infty$ and $d^*_n v_n \to 0$, from Strong Law of Large Numbers the first term converges to 0 in probability. The same is true for the second term because $\mathscr{B}^*_{1,n,LS} = O_p(1)$, $d^*_n \to 0$ and $E(\|A\boldsymbol{\varepsilon}\|^{-1}) < \infty$.

A.3 Asymptotic Normality of $\widehat{\mathscr{B}}_{TRR}$

Recall that by (2.1), the multivariate linear model can be written as

$$\mathbf{Y} = \mathbf{1}_n\boldsymbol{\beta}'_0 + \mathbf{X}\mathscr{B}_1 + \boldsymbol{\mathcal{E}}. \tag{A.3.7}$$

For convenience, we introduce the following transformation of Model (A.3.7). Let $\mathbf{P}_1$ be the projection matrix for the space spanned by $\mathbf{1}_n$. Let $\mathbf{I}_n$ be the $n \times n$ identity matrix. Consider the following notation

N1.$\mathbf{X}_c = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}$.
N2.$\mathbf{C} = \mathbf{X}_c(\mathbf{X}'_c\mathbf{X}_c)^{-1/2}$.
N3.$\boldsymbol{\Delta} = (\mathbf{X}'_c\mathbf{X}_c)^{1/2}\mathscr{B}_1$.
N4.$\boldsymbol{\alpha}'_0 = \boldsymbol{\beta}'_0 + n^{-1}\mathbf{1}'_n\mathbf{X}(\mathbf{X}'_c\mathbf{X}_c)^{-1/2}\boldsymbol{\Delta}$.

Under this notation we can express model (A.3.7) as

$$\mathbf{Y} = \mathbf{1}_n\boldsymbol{\alpha}'_0 + \mathbf{C}\boldsymbol{\Delta} + \boldsymbol{\mathcal{E}} \tag{A.3.8}$$

To obtain the asymptotic distribution of $\widehat{\mathscr{B}}_{TRR}$ we can assume without loss of generality that the true regression parameters are zero; i.e., $\boldsymbol{\beta}_0 = \mathbf{0}$ and that $\mathscr{B}_1 = \mathbf{0}$. The following theorem due to Davis and McKean (1993) gives the asymptotic distribution of $\widehat{\mathscr{B}}_R$.

Theorem A.3.1. *Under assumptions* A1-A7

$$\sqrt{n}\,\mathrm{vec}[\widehat{\boldsymbol{\beta}}_{0,s}, \widehat{\mathscr{B}}'_{1,\varphi}] \xrightarrow{\mathscr{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix} \mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + \mathbf{v}'\mathbf{V}^{-1}\mathbf{v} & \mathbf{TST}\otimes\mathbf{v}'\mathbf{V}^{-1} \\ \mathbf{TST}\otimes\mathbf{V}^{-1}\mathbf{v} & \mathbf{TST}\otimes\mathbf{V}^{-1}\end{pmatrix}\right). \tag{A.3.9}$$

Proof of Theorem 4.2 Since the transformed-retransformed estimate $\widehat{\mathscr{B}}_{TRR} = \widehat{\mathscr{B}}_R(\hat{A}^{-1})'$ we have

$$\mathrm{vec}(\widehat{\mathscr{B}}'_{TRR}) = (\hat{A}_n^{-1}\otimes I_{p+1})\,\mathrm{vec}(\widehat{\mathscr{B}}'_R).$$

Let

$$M^* = \begin{pmatrix} 1 & -\overline{\mathbf{x}}'(n\mathbf{X}'_c\mathbf{X}_c)^{-1/2} \\ \mathbf{0}_p & (n\mathbf{X}'_c\mathbf{X}_c)^{-1/2}\end{pmatrix}.$$

It follows from Theorem A.3.1 that

$$
\begin{aligned}
&\operatorname{cov}(\sqrt{n}\operatorname{vec}\widehat{\mathscr{B}}'_{TRR}) \\
&= (\hat{A}_n^{-1}\otimes I_{p+1})\operatorname{cov}(\sqrt{n}\operatorname{vec}\widehat{\mathscr{B}}'_R)((\hat{A}_n^{-1})'\otimes I_{p+1}) \\
&= (\hat{A}_n^{-1}\otimes I_{p+1})(I_d\otimes M^*)\begin{pmatrix}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* & & & \\ & \mathbf{TST} & & \\ & & \ddots & \\ & & & \mathbf{TST}\end{pmatrix}(I_d\otimes M^{*\prime})((\hat{A}_n^{-1})'\otimes I_{p+1}) \\
&= (I_d\otimes M^*)(\hat{A}_n^{-1}\otimes I_{p+1})\begin{pmatrix}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* & & & \\ & \mathbf{TST} & & \\ & & \ddots & \\ & & & \mathbf{TST}\end{pmatrix}((\hat{A}_n^{-1})'\otimes I_{p+1})(I_d\otimes M^{*\prime}) \\
&= (I_d\otimes M^*)\begin{pmatrix}\hat{A}_n^{-1}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*(\hat{A}_n^{-1})' & & & \\ & \hat{A}_n^{-1}\mathbf{TST}(\hat{A}_n^{-1})' & & \\ & & \ddots & \\ & & & \hat{A}_n^{-1}\mathbf{TST}(\hat{A}_n^{-1})'\end{pmatrix}(I_d\otimes M^{*\prime})
\end{aligned}
$$

As $\hat{A}_n \xrightarrow{p} A$ we obtain the asymptotic distribution of $\widehat{\mathscr{B}}_{TRR}$ under the assumption $\mathscr{B} = \mathbf{0}$ as stated in the theorem.

A.4 Asymptotic Normality of $\widehat{\mathscr{B}}_{TRGR}$

Lemma A.4.1. Let $\overline{\boldsymbol{S}_{GR}(\mathbf{0})} = n^{-3/2}\boldsymbol{S}_{GR}(\mathbf{0})$. Further, let

$$
\begin{aligned}
\overline{\boldsymbol{S}_{GR}}^*(\mathbf{0}) &= \left(\overline{\boldsymbol{S}_{GR}^{(1)}}^*(\mathbf{0}) \cdots \overline{\boldsymbol{S}_{GR}^{(d)}}^*(\mathbf{0})\right) \\
&= \frac{n}{n^{3/2}}\mathbf{X}'\mathbf{W}\begin{pmatrix}2F_1(\mathbf{Y}_1^{(1)}) & \cdots & 2F_d(\mathbf{Y}_1^{(d)}) \\ \vdots & \cdots & \vdots \\ 2F_1(\mathbf{Y}_n^{(1)}) & \cdots & 2F_d(\mathbf{Y}_n^{(d)})\end{pmatrix}.
\end{aligned}
\tag{A.4.10}
$$

Then $\overline{\boldsymbol{S}^*_{GR}(\mathbf{0})} - \overline{\boldsymbol{S}_{GR}(\mathbf{0})} \xrightarrow{p} \mathbf{0}$.

Proof. See Hettmansperger and McKean (2011).

Theorem A.4.1. *Under assumptions A1, B1–B3 and assuming that $\mathscr{B}_1 = \mathbf{0}$,*

$$\overline{\boldsymbol{S}_{GR}(\mathbf{0})} \xrightarrow{\mathscr{D}} N_{p,d}(\mathbf{0},\mathbf{E},\boldsymbol{S})$$

*Proof.*By lemma A.4.1 we only need to show that $\overline{\boldsymbol{S}_{GR}}^*(\mathbf{0}) \xrightarrow{\mathscr{D}} N_{p,d}(\mathbf{0},\mathbf{E},\boldsymbol{S})$. Let $\mathbf{M}$ be any $p\times d$ matrix. Then by Theorem 19.15 of Arnold (1981) we need to show that $\operatorname{tr}\mathbf{M}'\overline{\boldsymbol{S}_{GR}}^*(\mathbf{0}) \xrightarrow{\mathscr{D}} N_1(\mathbf{0},\operatorname{tr}\mathbf{M}'\mathbf{E}\mathbf{M}\boldsymbol{S})$. Now, let $\mathbf{F}'_n = \left(\mathbf{F}(\mathbf{Y}_1) \ldots \mathbf{F}(\mathbf{Y}_n)\right)$ where, for $i = 1,\ldots,n$,

$$\mathbf{F}'(\mathbf{Y}_i) = \left(2F_1(\mathbf{Y}_i^{(1)}) - 1,\ \ldots,\ 2F_d(\mathbf{Y}_i^{(d)}) - 1\right).$$

Then we have $E[\mathbf{F}(\mathbf{Y}_i)] = \mathbf{0}$ and $\operatorname{Cov}(\mathbf{F}(\mathbf{Y}_i)) = \boldsymbol{S}$. Finally, let $\mathbf{B}_n = (1/\sqrt{n})\mathbf{W}_n\mathbf{X}_n\mathbf{M}$. Then

$$
\begin{aligned}
\operatorname{tr}\mathbf{M}'\overline{\boldsymbol{S}_{GR}}^*(\mathbf{0}) &= \operatorname{tr}\frac{1}{\sqrt{n}}\mathbf{M}'\mathbf{X}'\mathbf{W}_n\mathbf{F}_n \\
&= \operatorname{tr}\mathbf{B}'_n\mathbf{F}_n,
\end{aligned}
$$

and

$$\begin{aligned} \operatorname{tr} \mathbf{B}_n \boldsymbol{S} \mathbf{B}_n' &= \operatorname{tr} \frac{1}{n} \mathbf{W}_n \mathbf{X}_n \mathbf{M} \boldsymbol{S} \mathbf{M}' \mathbf{X}_n' \mathbf{W}_n \\ &= \operatorname{tr} (\mathbf{M} \boldsymbol{S} \mathbf{M}')(\frac{1}{n} \mathbf{X}_n' \mathbf{W}_n^2 \mathbf{X}_n) \\ &\longrightarrow \operatorname{tr} (\mathbf{M} \boldsymbol{S} \mathbf{M}') \mathbf{E} < \infty. \end{aligned}$$

Also, we have

$$\begin{aligned} m(\mathbf{B}_n) &= m(\frac{1}{\sqrt{n}} \mathbf{W}_n \mathbf{X}_n \mathbf{M}) \\ &\leq \frac{1}{\sqrt{n}} p\, m(\mathbf{W}_n \mathbf{X}_n)\, m(\mathbf{M}) \\ &\longrightarrow 0 \end{aligned}$$

by conditions B2 and B3. Thus by theorem 19.16 of Arnold (1981) we have $\operatorname{tr} \mathbf{B}_n' \mathbf{F}_n \xrightarrow{\mathscr{D}} N_1(\mathbf{0}, \operatorname{tr} \mathbf{M} \boldsymbol{S} \mathbf{M}' \mathbf{E})$. Therefore, again by theorem 19.15 of Arnold, $\overline{\boldsymbol{S}_{GR}{}^*(\mathbf{0})} = \mathbf{B}_n' \mathbf{F}_n \xrightarrow{\mathscr{D}} N_{p,d}(\mathbf{0}, \mathbf{E}, \boldsymbol{S})$.

The approximating quadratics for $j = 1, \ldots, d$ are given by

$$Q(\mathscr{B}_1^{(j)}) = \frac{n}{2\sqrt{3}\,\tau_j} \mathscr{B}_1^{(j)\prime} \mathbf{X}' \mathbf{W} \mathbf{X} \mathscr{B}_1^{(j)} - \mathscr{B}_1^{(j)} \boldsymbol{S}_{GR}^{(j)}(\mathbf{0}) + D_{GR}^{(j)}(\mathbf{0}).$$

The vector $\mathscr{B}_1^{(j)}$ that minimizes $Q(\mathscr{B}_1^{(j)})$ is $\widetilde{\mathscr{B}}_{1,GR}^{(j)} = \frac{\sqrt{3}\tau_j}{n} (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \boldsymbol{S}_{GR}^{(j)}(\mathbf{0})$. Let $\widetilde{\mathscr{B}}_{1,GR} = \left(\widetilde{\mathscr{B}}_{1,GR}^{(1)} \cdots \widetilde{\mathscr{B}}_{1,GR}^{(d)} \right)$. Then $\widetilde{\mathscr{B}}_{1,GR} = (\sqrt{3}/n)(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \boldsymbol{S}_{GR}(\mathbf{0}) \mathbf{T}$.

Theorem A.4.2

$$\sqrt{n}\, \widehat{\mathscr{B}}_{1,GR} \xrightarrow{\mathscr{D}} N_{p,d}(\mathbf{0}, 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}, \mathbf{T}\boldsymbol{S}\mathbf{T}).$$

Proof. The result follows from the fact that

$$\begin{aligned} \sqrt{n}\, \widetilde{\mathscr{B}}_{1,GR} &= \sqrt{3}\, (\frac{1}{n} \mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \overline{\boldsymbol{S}_{GR}(\mathbf{0})} \mathbf{T} \\ &\xrightarrow{\mathscr{D}} N_{p,d}(\mathbf{0}, 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}, \mathbf{T}\boldsymbol{S}\mathbf{T}) \end{aligned}$$

and the result, $\sqrt{n}\, (\widetilde{\mathscr{B}}_{1,GR} - \widehat{\mathscr{B}}_{1,GR}) \xrightarrow{p} \mathbf{0}$, proved in Hettmansperger and McKean (2011).

To estimate the intercept vector $\boldsymbol{\alpha}_0$, $\widehat{\boldsymbol{\alpha}}_{0,GR}^{(j)}$ for $j = 1, \ldots, d$, is taken to be the median of the residuals of the j$^{\text{th}}$ column, i. e.

$$\widehat{\boldsymbol{\alpha}}_{0,GR}^{(j)} = \underset{1 \leq i \leq n}{\operatorname{med}} (\mathbf{Y}_i^{(j)} - \boldsymbol{x}_{c,i}' \widehat{\mathscr{B}}_{1,GR}^{(j)}).$$

For $j = 1, \ldots, d$ let

$$\boldsymbol{S}_1(\mathbf{Y}^{(j)} - \boldsymbol{\alpha}_0^{(j)} \mathbf{1}_n - \mathbf{X}_c \widehat{\mathscr{B}}_{1,GR}^{(j)}) = \sum_{i=1}^{n} \operatorname{sgn}(\mathbf{Y}_i^{(j)} - \boldsymbol{\alpha}_0^{(j)} - \boldsymbol{x}_{c,i}' \widehat{\mathscr{B}}_{1,GR}^{(j)}).$$

Then $\widehat{\boldsymbol{\alpha}}_{0,GR}^{(j)}$ solves

$$\boldsymbol{S}_1(\mathbf{Y}^{(j)} - \boldsymbol{\alpha}_0^{(j)} \mathbf{1}_n - \mathbf{X}_c \widehat{\mathscr{B}}_{1,GR}^{(j)}) \doteq \mathbf{0}.$$

Lemma A.4.2. *If assumptions A1, A2 and A3 hold and* $\boldsymbol{\Phi}^* = (\operatorname{sgn}(\varepsilon_{ij}))$ then for all $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr(\| n^{1/2} \widehat{\boldsymbol{\alpha}}_{0,GR}' \mathbf{T}^* - n^{-1/2} \mathbf{1}_n' \boldsymbol{\Phi}^* \| > \varepsilon) = 0.$$

Proof. It is enough to show for all $j = 1, \ldots, d$ that

$$\lim_{n\to\infty} \Pr(|n^{1/2}\widehat{\boldsymbol{\alpha}}_{0,GR}^{(j)}\tau_j^{*-1} - n^{-1/2}\mathbf{1}_n'\boldsymbol{\Phi}^{*(j)}| > \varepsilon) = 0.$$

Note that

$$\begin{aligned}\mathbf{1}_n'\boldsymbol{\Phi}^{*(j)} &= \boldsymbol{S}_1(\mathbf{Y}^{(j)}) \\ &= \sum_{i=1}^{n} \mathrm{sgn}(\mathbf{Y}_i^{(j)}).\end{aligned}$$

Now

$$\begin{aligned}&\Pr(|n^{1/2}\widehat{\boldsymbol{\alpha}}_{0,GR}^{(j)}\tau_j^{*-1} - n^{-1/2}\boldsymbol{S}_1(\mathbf{Y}^{(j)})| > \varepsilon) \\ &\quad \le \Pr(|n^{1/2}\widehat{\boldsymbol{\alpha}}_{0,GR}^{(j)}\tau_j^{*-1} - n^{-1/2}\boldsymbol{S}_1(\mathbf{Y}^{(j)} - \widehat{\alpha}^{(j)}\mathbf{1}_n - \mathbf{X}_c\widehat{\mathscr{B}}_{1,GR}^{(j)}) \\ &\qquad\quad - n^{-1/2}\boldsymbol{S}_1(\mathbf{Y}^{(j)} - \mathbf{X}_c\widehat{\mathscr{B}}_{1,GR}^{(j)})| > \varepsilon/3) \\ &\quad + \Pr(|n^{-1/2}\boldsymbol{S}_1(\mathbf{Y}^{(j)} - \mathbf{X}_c\widehat{\mathscr{B}}_{1,GR}^{(j)}) - n^{-1/2}\boldsymbol{S}_1(\mathbf{Y}^{(j)})| > \varepsilon/3) \\ &\quad + \Pr(|n^{-1/2}\boldsymbol{S}_1(\mathbf{Y}^{(j)} - \widehat{\alpha}^{(j)}\mathbf{1}_n - \mathbf{X}_c\widehat{\mathscr{B}}_{1,GR}^{(j)})| > \varepsilon/3)\end{aligned}$$

The first and second terms go to 0 by theorem 3.5.9 and lemma 3.5.8 of Hettmansperger and McKean (2011).

It follows from theorem 18 of Lehman (1975) that $(\frac{1}{\sqrt{n}}\mathbf{1}_n'\boldsymbol{\Phi}^*)' \xrightarrow{\mathscr{D}} N_d(\mathbf{0}, \boldsymbol{S}^*)$. Thus, $\sqrt{n}\,\widehat{\boldsymbol{\alpha}}_{0,GR}' \xrightarrow{\mathscr{D}} N_d(\mathbf{0}, \mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*)$.

Theorem A.4.3

$$\sqrt{n}\,\mathrm{vec}\left(\widehat{\boldsymbol{\alpha}}_{0,GR}\;\widehat{\mathscr{B}}_{1,GR}'\right) \xrightarrow{\mathscr{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{TST}\otimes 3\mathbf{C}^{-1}\mathbf{EC}^{-1}\end{pmatrix}\right). \tag{A.4.11}$$

Proof of Theorem 4.3 The result is obtained by noting that

$$\begin{aligned}\sqrt{n}\,\mathrm{vec}\left(\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{0,GR}' \\ \widehat{\mathscr{B}}_{1,GR}\end{pmatrix}'\right) &= \mathrm{vec}\left(\begin{pmatrix}1 & -\bar{\boldsymbol{x}}' \\ \mathbf{0} & I_p\end{pmatrix}\begin{pmatrix}\sqrt{n}\,\widehat{\boldsymbol{\alpha}}_{0,GR}' \\ \sqrt{n}\,\widehat{\mathscr{B}}_{1,GR}\end{pmatrix}\right)' \\ &= \mathrm{vec}\left(\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_{0,GR}\;\widehat{\mathscr{B}}_{1,GR}'\right)\begin{pmatrix}1 & -\bar{\boldsymbol{x}}' \\ \mathbf{0} & I_p\end{pmatrix}'\right) \\ &= \left(I_d \otimes \begin{pmatrix}1 & -\bar{\boldsymbol{x}}' \\ \mathbf{0} & I_p\end{pmatrix}\right)\mathrm{vec}\left(\sqrt{n}\,\widehat{\boldsymbol{\alpha}}_{0,GR}\;\sqrt{n}\,\widehat{\mathscr{B}}_{1,GR}'\right).\end{aligned}$$

Now the result follows from theorem A.4.3.

Proof of Theorem 4.4 Similar to the proof of Theorem 4.2. The affine invariance of the weights, the consistency of $\hat{A}$ and Theorem 4.3 gives the above asymptotic distribution of $\widehat{\mathscr{B}}_{TRGR}$.

References

[1] Armstrong, R. D. and Kung, D. S. (1978), Least absolute value estimates for a simple linear regression problem, *Applied Statistics*, **27**, 363-366.
[2] Arnold, Steven F. (1981), *The Theory of Linear Models and Multivariate Analysis*, New York: John Wiley & Sons Inc.
[3] Bickel, P. J. (1964), On some alternative estimates for the shift in the p-variate one sample problem, *The Annals of Mathematical Statistics*, **35**, 1079-1090.
[4] Chakraborty, B. (1996), On multivariate median regression, Calcutta: Indian Statistical Institute, Theoretical Statistics & Mathematical Unit, 9.
[5] Chakraborty, B. and Chauduri, P. (1997), On multivariate rank regression, 399,414, Ed: Y. Dodge, *IMS Lecture Notes: L_1-Statistical Procedures and Related Topics*, Heyward, CA: IMS.

[6] Davis, J. B. and McKean, J. W. (1993), Rank based methods for multivariate linear models, *Journal of the American Statistical Association*, **88**, 245-251.
[7] Davis, J. B. and McKean, J. W. (1993), Rank based methods for multivariate linear models, *Journal of the American Statistical Association*, **88**, 245-251.
[8] Graybill, F. A. (1983), *Matrices with Applications in Statistics*, Belmont, CA: Wadsworth.
[9] Hájek, J. and Šidák, Z. (1967), *Theory of Rank Tests*, New York: Academic Press.
[10] Hettmansperger, T. P. and McKean, J. W. (2011), *Robust and Nonparametric Statistical Methods 2nd Ed.*, Boca Raton, FL: Chapman-Hall.
[11] Johnson, R. A. and Wichern, D. W. (1998), *Applied Multivariate Statistical Analysis, 4th Ed.*, Upper Saddle River, New Jersey: Prentice Hall.
[12] Lehmann, E. L. (1975), *Nonparametric Statistical Methods Based on Ranks*, San Francisco: Holden-Day.
[13] Maronna, R. A. and Morgethaler, S. (1986), Robust regression through robust covariances, *Communications in Statistics, Theory Meth.*, **15**, 1347-1366.
[14] Naranjo, J. D. and Hettmansperger, T. P. (1994), Bounded influence rank regression, *Journal of the Royal Statistical Society. Series B*, **56**, 209-220.
[15] Peters, D. and Randles, R. H. (1990), A multivariate signed-rank test for the one-sampled location problem, *Journal of the American Statistical Association*, **85**, 552-557.
[16] Ollila, E., Hettmansperger, T. P. and Oja, H. (2002), Estimates of regression coefficients based on sign covariance matrix, *Journal of the Royal Statistical Society, B*, **64**, 447-466.
[17] Randles, R. H. (2000), A simpler, affine-invariant, multivariate distribution-free sign test, *Journal of the American Statistical Association*, **95**, 1263-1268.
[18] Rao, C. R. (1988), Methodology based on L_1-norm in statistical inference, *Sankhya, A*, **50**, 289-313.
[19] Rousseeuw, P. and Van Driessen, K. (1999), A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, **41**, 212-223.
[20] Sievers, G. L. (1983), A weighted dispersion function for estimation in linear models, *Communications in Statistics, Theory and Methods*, **12**, 1161-1179.
[21] Tyler, D. E. (1987), A distribution-free M-estimator of scatter, *Annals of Statistics*, **15**, 234-251.
[22] Terpstra, J. and McKean, J. W. (2005), Rank-Based Analyses of Linear Models using R, *Journal of Statistical Software*, **14**, Issue 7.
[23] Kloke, J.D. and McKean, J.W. (2012), Rfit: Rank-based estimation for linear models, *The R Journal*, Tentatively accepted.