Affine Equivariant Multivariate Rank-Based and Generalized Rank Regression

Majeda Salman
Western Michigan University

AFFINE EQUIVARIANT MULTIVARIATE RANK-BASED AND GENERALIZED RANK REGRESSION

by

Majeda Salman

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics

Western Michigan University
Kalamazoo, Michigan
June 2004

AFFINE EQUIVARIANT MULTIVARIATE RANK-BASED AND GENERALIZED RANK REGRESSION

Majeda Salman, Ph.D.

Western Michigan University, 2004

Affine equivariant estimates for the regression coefficient matrix of the multivariate linear model 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 non-equivariant 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. Based on the new estimates quadratic procedures for testing the general hypothesis $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$ are developed. Examples illustrating the estimation and testing procedures are presented.

ACKNOWLEDGMENTS

I am very thankful to Dr. Joseph W. McKean for being my dissertation advisor. His guiddance, encouragement and support have made this work possible. I would like to thank my family for all of their support, encouragement and believing in me.

Majeda Salman

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

CHAPTER I

INTRODUCTION

This dissertation investigates procedures for estimating regression parameters of a multivariate linear model. A method similar to the transformation and retransformation approach of Chakraborty et. al. (1997) which uses Tyler's M-estimator of scatter is proposed.

Consider the multivariate linear model

$$\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\mathcal{B}}_1'\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, n, \tag{1.1}$$

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, $\boldsymbol{\mathcal{B}}_1$ 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

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

$$\mathbb{X} = \begin{pmatrix} \mathbf{1}_n & \mathbf{X} \end{pmatrix}, \boldsymbol{\mathcal{B}} = \begin{pmatrix} \beta_0' \\ \boldsymbol{\mathcal{B}}_1 \end{pmatrix}.$$

Then we can write model (1.1) in a matrix form as

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{E}}. \tag{1.2}$$

We wish to estimate and test hypothesis concerning $\boldsymbol{\mathcal{B}}$.

1.1 LS Estimation

The least squares estimate of $\boldsymbol{\mathcal{B}}$ is

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

Besides being easy to compute, $\widehat{\boldsymbol{\mathcal{B}}}_{LS}$ is affine equivariant. In other words it is equivariant under constant shifts and multiplication by arbitrary nonsingular matrices. It is the optimal estimator of $\boldsymbol{\mathcal{B}}$ when the distribution of the errors is multivariate normal. However, it is not robust if the errors have a heavy tailed distribution.

1.2 Robust Estimation

There are mainly three approaches in the literature competing with the LS method and producing estimators that are robust and more efficient.

1.2.1 Componentwise Methods

Let $\mathbf{Y}^{(j)}$, $\boldsymbol{\mathcal{B}}^{(j)} = \left(\beta_0^{(j)},\ \boldsymbol{\mathcal{B}}_1^{(j)\prime}\right)'$, $\boldsymbol{\mathcal{E}}^{(j)}$ denote the j$^{\text{th}}$ columns of the matrices $\mathbf{Y}$, $\boldsymbol{\mathcal{B}}$, and $\boldsymbol{\mathcal{E}}$ respectively. Extracting the j$^{\text{th}}$ column from each side of (1.2) gives the d equations

$$\mathbf{Y}^{(j)} = \mathbf{1}_n\beta_0^{(j)} + \mathbf{X}\boldsymbol{\mathcal{B}}_1^{(j)} + \boldsymbol{\mathcal{E}}^{(j)}, \quad \text{for } j = 1, \ldots, d. \tag{1.3}$$

The componentwise estimation treats each of the d response variables separately, even though the $\mathbf{Y}^{(j)}$ are correlated. It finds $\widehat{\boldsymbol{\mathcal{B}}}$ by computing $\widehat{\boldsymbol{\mathcal{B}}}^{(j)}$ for $j = 1, \ldots, d$ using univariate techniques. The main disadvantage of these procedures is that they are not affine equivariant.

The method of least absolute deviations (LAD) finds $\boldsymbol{\mathcal{B}}^{(j)}$ that

$$\text{minimize} \sum_{i=1}^{n} |\mathbf{Y}_i^{(j)} - \beta_0^{(j)} - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_1^{(j)}|.$$

Bassett and Koenker (1978) investigated LAD in linear models. There are many good algorithms that compute LAD estimates. See, for example, Armstrong and Kung (1978). Rao (1988) used the univariate least absolute deviation regression in a multivariate regression setup.

Another approach to generalize LAD estimation in the multivariate setup is due to Bia, Mai and Rao (1990) who extended the notion of spatial median (see Haldane (1948), Brown (1983)) to the multivariate regression problem to obtain an estimate of $\boldsymbol{\mathcal{B}}$ that minimizes $\sum_{i=1}^{n} ||\boldsymbol{y}_i - \boldsymbol{\mathcal{B}}'\boldsymbol{x}_i||$. This estimate is affine equivariant only under orthogonal transformation of the response variable.

Rank regression was proposed by Jurečková (1971). Jaeckel (1972) proposed a measure of distance based on ranks that provides geometry for estimation and hypothesis testing similar to least squares; see also McKean and Schrader (1980). 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 by Hettmansperger and McKean (1983). The estimate of $\boldsymbol{\mathcal{B}}$ was obtained by first estimating the regression coefficient matrix $\boldsymbol{\mathcal{B}}_1$ by minimizing for $j = 1, \ldots, d$ the dispersion functions

$$D(\boldsymbol{\mathcal{B}}_1^{(j)}) = \sum_{i=1}^{n} a(R(\mathbf{Y}_i^{(j)} - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_1^{(j)}))(\mathbf{Y}_i^{(j)} - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_1^{(j)}), \tag{1.4}$$

where $a(i)$ are scores such that $a(1) \leq a(2) \leq \cdots \leq a(n)$ and $\sum a(i) = 0$. The scores are generated by a score generating function φ by $a(i) = \varphi(i/(n+1))$. The most widely used scores are wilcoxon scores which can be generated by $\varphi(u) = 12^{1/2}(u - .5)$. The ranks R were computed on the i^{th} element of the vector $\mathbf{Y}^{(j)}$. The intercept vector $\boldsymbol{\beta}_0$ is then computed as a location estimate of the residuals for each component. Thus the problem of estimating $\boldsymbol{\mathcal{B}}$ is reduced to estimating $\boldsymbol{\mathcal{B}}^{(j)}$ for each column separately. Under certain conditions, Davis and McKean showed that $\widehat{\boldsymbol{\mathcal{B}}}_R = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_0 & \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}' \end{pmatrix}'$ is asymptotically normal and $\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}$ has an asymptotic relative efficiency (ARE) of 95% relative to the LS estimate.

1.2.2 Covariance Estimation

Maronna et. al. (1986) proposed this 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 estimate. Ollila, Hettmansperger and Oja (2002) used a similar approach in multivariate linear regression. To be able to describe their work, let F be the joint distribution function of the independent and dependent variables. Let

$$\mathbf{Z} = \begin{pmatrix} z_1 & \cdots & z_n \end{pmatrix}' = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \\ \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_n \end{pmatrix}' = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \end{pmatrix}$$

be the data matrix. The errors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are iid with $\text{Cov}(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_i) = 0$ and covariance matrix $\Psi = \text{Cov}(\boldsymbol{\varepsilon}_i)$. Let

$$E_F(\boldsymbol{z}_i) = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \text{Cov}(\boldsymbol{z}_i) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then the regression coefficient matrix is given by

$$\boldsymbol{\mathcal{B}} = \begin{pmatrix} \boldsymbol{\beta}_0' \\ \boldsymbol{\mathcal{B}}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_2' - \boldsymbol{\mu}_1' \Sigma_{11}^{-1} \Sigma_{12} \\ \Sigma_{11}^{-1} \Sigma_{12} \end{pmatrix}.$$

The multivariate empirical sign vector of $\boldsymbol{z}$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{S}(\boldsymbol{z}; \boldsymbol{\theta})$, is defined to be the gradient of

$$d(\boldsymbol{z}; \boldsymbol{\theta}) = \frac{1}{\binom{n}{k-1}} \sum_I \text{abs} \left\{ \det \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \boldsymbol{\theta} & \boldsymbol{z}_{i_1} & \cdots & \boldsymbol{z}_{i_{k-1}} & z \end{pmatrix} \right\}$$

with respect to $\boldsymbol{z}$ where I denotes the set of subscripts $\{i_1 < \ldots < i_n\}$ from $1, \ldots, n$. Oja's median (1983) solves $\sum \boldsymbol{S}(\boldsymbol{z}_i; \boldsymbol{\theta}) = 0$. Let $\widehat{\boldsymbol{\theta}}$ denotes the Oja median and let $\widehat{\boldsymbol{S}}_i = \boldsymbol{S}(\boldsymbol{z}_i; \widehat{\boldsymbol{\theta}})$, then the sign covariance matrix (SCM) is given by $\widehat{D} = \text{ave}_i\{\widehat{\boldsymbol{S}}_i \widehat{\boldsymbol{S}}_i^T\}$. Let

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \widehat{\boldsymbol{\theta}}_1 \\ \widehat{\boldsymbol{\theta}}_2 \end{pmatrix} \quad \text{and} \quad \widehat{D} = \begin{pmatrix} \widehat{D}_{11} & \widehat{D}_{12} \\ \widehat{D}_{21} & \widehat{D}_{22} \end{pmatrix}$$

be the sample Oja median and the sample SCM respectively. Ollila, Oja and Hettmansperger defined the SCM regression estimate $\widehat{\boldsymbol{\mathcal{B}}}_{SCM}$ as

$$\widehat{\boldsymbol{\mathcal{B}}}_{SCM} = \begin{pmatrix} \widehat{\beta}_0' \\ \widehat{\boldsymbol{\mathcal{B}}}_1 \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\theta}}_2' + \widehat{\boldsymbol{\theta}}_1' \widehat{D}_{12} \widehat{D}_{22}^{-1} \\ \widehat{D}_{12} \widehat{D}_{22}^{-1} \end{pmatrix}.$$

They showed under some conditions that $\widehat{\boldsymbol{\mathcal{B}}}_{SCM}$ is a consistent estimate of $\boldsymbol{\mathcal{B}}$, affine equivariant, and asymptotically normal. It has a bounded influence function in both $\boldsymbol{x}$ and $\boldsymbol{y}$ spaces. The ARE of $\widehat{\boldsymbol{\mathcal{B}}}_{SCM}$ relative to LAD increases as the correlation among the response variables increases.

1.2.3 Transformation-Retransformation Technique

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. To illustrate his method, let

$$A_n = \{\, \boldsymbol{a} : \boldsymbol{a} \subset \{1, 2, \ldots, n\} \quad \text{and} \quad \#\{i : i \in \boldsymbol{a}\} = k \,\}$$

$$B_n = \{\, \boldsymbol{b} : \boldsymbol{b} \subset \{1, 2, \ldots, n\} \quad \text{and} \quad \#\{i : i \in \boldsymbol{b}\} = d \,\}$$

and

$$S_n = \{\, \alpha = \boldsymbol{a} \cup \boldsymbol{b} : \boldsymbol{a} \in A_n,\ \boldsymbol{b} \in B_n, \boldsymbol{a} \cap \boldsymbol{b} = \phi \,\}$$

For a fixed $\alpha = \{\, i_1, \ldots, i_k, j_1, \ldots, j_d \,\}$, Chakraborty constructed the column matrices

$$W(\alpha) = \begin{pmatrix} \boldsymbol{x}_{i_1} & \cdots & \boldsymbol{x}_{i_k} \end{pmatrix} \quad \text{and} \quad Z(\alpha) = \begin{pmatrix} \boldsymbol{y}_{i_1} & \cdots & \boldsymbol{y}_{i_k} \end{pmatrix}.$$

Then he defined his transformation matrix to be the column matrix

$$E(\alpha) = \begin{pmatrix} \boldsymbol{y}_{j_1} - Z(\alpha)W^{-1}(\alpha)\boldsymbol{x}_{j_1} & \cdots & \boldsymbol{y}_{j_d} - Z(\alpha)W^{-1}(\alpha)\boldsymbol{x}_{j_d} \end{pmatrix}.$$

Next Chakraborty transformed all the $\boldsymbol{y}_i$'s such that $i \notin \alpha$ to $\boldsymbol{z}_i = E^{-1}(\alpha)\boldsymbol{y}_i$. $\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$ was then computed on $(\boldsymbol{x}_i, \boldsymbol{z}_i)$ with $i \notin \alpha$. He called the retransformed estimate, $\widehat{\boldsymbol{\mathcal{B}}} = \widehat{\boldsymbol{\mathcal{B}}}_{LAD}\, E'(\hat{\alpha})$, TREMMER (Transformation-Retransformation Estimate in Multivariate Median Regression). In his paper, he proved that the TREMMER is indeed affine equivariant, 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) modified the TREMMER algorithm and applied it to Davis and McKean (1993) componentwise rank regression estimate that is based on the Wilcoxon scores. Here they defined the transformed-retransformed estimate as $\widehat{\boldsymbol{\mathcal{B}}} = E(\alpha)\,\widehat{\boldsymbol{\mathcal{B}}}_{WIL}$, where $\widehat{\boldsymbol{\mathcal{B}}}_{WIL}$ is the Wilcoxon estimate computed on the regressor $\boldsymbol{x}_i$ and the transformed response $\boldsymbol{z}_i$ with $i \notin \alpha$. They proved that their estimate is affine equivariant under nonsingular transformation of the $\boldsymbol{y}_i$'s. It inherits its asymptotic normality and robustness from $\widehat{\boldsymbol{\mathcal{B}}}_{WIL}$. They also showed that the transformation matrix $E(\alpha)$ yields an estimate $\widehat{\boldsymbol{\mathcal{B}}}$ such that the asymptotic generalized variance

of $\sqrt{n}\,(\widehat{\mathcal{B}} - \mathcal{B})$ is minimum. From their simulation results, they concluded that this estimate has a high ARE relative to LAD and its performance is even much better than the TREMMER.

In this paper new estimates of the multivariate regression coefficient matrix $\mathcal{B}$ are proposed. The transformation-retransformation approach of Chakraborty (1996) together with a directional transformation matrix due to Tyler (1987) are used. The transformation matrix is computed such that the sample variance-covariance matrix of the unit transformed data is $1/d$ times the identity matrix. In chapter II, Davis and McKean (1993) rank regression estimate is described and based on it a new affine equivariant estimate is proposed. The asymptotic distribution of this estimate is established and its influence function is obtained. A generalized rank multivariate regression estimate is proposed in chapter III. This estimate is affine equivariant and has a bounded influence in both $\boldsymbol{x}$ and $\boldsymbol{y}$ spaces. Finally, several examples are presented in chapter IV to illustrate our procedures of estimation and testing. Chapter IV also includes simulation studies which demonstrated the high efficiency of the new estimates relative to LAD. In particular, our transformation-retransformation rank estimate performs as good as or better than the estimate of Chakraborty and Chauduri (1997).

CHAPTER II

A MULTIVARIATE RANK REGRESSION ESTIMATE

2.1 Definition of the Estimate

Since the new regression estimate $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ is a modification of Davis and McKean (1993) multivariate rank regression estimator, it is important to describe this estimate in more details. Further, we need to study the transformation matrix needed for this modification.

2.1.1 The Componentwise Estimate

In their approach, Davis and McKean (1993) thought of the multivariate linear model as a series of concatenations of univariate linear models. Recall that by (1.1), the multivariate linear model can be written as

$$\mathbf{Y} = \mathbf{1}_n\boldsymbol{\beta}_0' + \mathbf{X}\boldsymbol{\mathcal{B}}_1 + \boldsymbol{\mathcal{E}}. \tag{2.1}$$

For convenience, we introduce the following transformation of Model (2.1). 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. $\mathbf{\Delta} = (\mathbf{X}_c'\mathbf{X}_c)^{1/2}\boldsymbol{\mathcal{B}}_1$.

N4. $\boldsymbol{\alpha}_0' = \boldsymbol{\beta}_0' + n^{-1}\mathbf{1}_n'\mathbf{X}(\mathbf{X}_c'\mathbf{X}_c)^{-1/2}\mathbf{\Delta}$.

Under this notation we can express model (2.1) as

$$\mathbf{Y} = \mathbf{1}_n\boldsymbol{\alpha}_0' + \mathbf{C}\mathbf{\Delta} + \boldsymbol{\mathcal{E}} \tag{2.2}$$

which can be seen as a concatenation of the d univariate linear models

$$\mathbf{Y}^{(j)} = \mathbf{1}_n\boldsymbol{\alpha}_0^{(j)} + \mathbf{C}\mathbf{\Delta}^{(j)} + \boldsymbol{\mathcal{E}}^{(j)}, \quad \text{for } j = 1, \ldots, d. \tag{2.3}$$

Davis and McKean defined the $n \times p$ score matrix

$$\begin{aligned} A(\mathbf{Y} - \mathbf{C}\mathbf{\Delta}) &= [\, a(R(\mathbf{Y}_i^{(j)} - \mathbf{c}_i'\mathbf{\Delta}^{(j)}))\,] \\ &= \begin{pmatrix} A^{(1)} & \cdots & A^{(d)} \end{pmatrix}, \end{aligned}$$

where the ranks were taken within columns. In other words $\mathbf{Y}_i^{(j)} - \mathbf{c}_i'\mathbf{\Delta}^{(j)}$ was ranked among $\mathbf{Y}_1^{(j)} - \mathbf{c}_1'\mathbf{\Delta}^{(j)}, \ldots, \mathbf{Y}_n^{(j)} - \mathbf{c}_n'\mathbf{\Delta}^{(j)}$. The scores were generated by $a(i) = \varphi(i/(n+1))$, $0 < \varphi(u) < 1$, $\int \varphi(u)\,du = 0$, and $\int \varphi^2(u)\,du = 1$. The dispersion function is then given by

$$\begin{aligned} D(\mathbf{\Delta}) &= \operatorname{tr}(\mathbf{Y} - \mathbf{C}\mathbf{\Delta})'A(\mathbf{Y} - \mathbf{C}\mathbf{\Delta}) \\ &= \sum_{j=1}^{d}(\mathbf{Y}^{(j)} - \mathbf{C}\mathbf{\Delta}^{(j)})'\,A^{(j)} \\ &= \sum_{j=1}^{d} D_j(\mathbf{\Delta}). \end{aligned} \tag{2.4}$$

The matrix of the negatives of the partial derivatives is

$$\begin{aligned} L(\mathbf{\Delta}) &= \mathbf{C}'A(\mathbf{Y} - \mathbf{C}\mathbf{\Delta}) \\ &= \begin{pmatrix} \mathbf{C}'A^{(1)} & \cdots & \mathbf{C}'A^{(d)} \end{pmatrix}. \end{aligned} \tag{2.5}$$

The columns of (2.5) represent the estimating equations for the d concatenated univariate linear models in (2.3). The componentwise R-estimate defined by Davis and McKean then minimizes $D(\boldsymbol{\Delta})$ or solves $L(\boldsymbol{\Delta}) = \mathbf{0}$.

The intercept vector $\boldsymbol{\alpha}_0$ was estimated using signed rank statistics as the solution of the d equations

$$\sum_{i=1}^{n} a^+(R[\mathbf{Y}_i^{(j)} - \boldsymbol{\alpha}_0^{(j)} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}_\varphi^{(j)}])\,\mathrm{sgn}(\mathbf{Y}_i^{(j)} - \boldsymbol{\alpha}_0^{(j)} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}_\varphi^{(j)}) = 0, \quad j = 1, \ldots, d \qquad (2.6)$$

where $a^+(i) = \varphi^+(i/(n+1))$ and $\varphi^+(u) = \varphi(.5(u+1))$. We consider here only the case $a^+(i) = 1$. That is the intercept estimate, $\widehat{\alpha}_0^{(j)}$, is the median of the residuals in the j$^{\text{th}}$ column of $\mathbf{Y} - \mathbf{C}\widehat{\boldsymbol{\Delta}}_\varphi$.

The following assumptions are needed for the theory given by Davis and McKean (1993):

A1. The rows of $\boldsymbol{\mathcal{E}}$ 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. $\mathrm{Cov}(\varphi(F_j(\varepsilon_{ij}), \varphi(F_{j'}(\varepsilon_{ij'})) = s_{jj'} < \infty$, 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.

2.1.2 The Transformation Matrix

The transformation matrix $\hat{A}$ needed for the new estimate is a data-determined nonsingular matrix that was proposed by Tyler (1987). For a random sample $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ from a continuous distribution the matrix $\hat{A}$ is the unique upper triangular positive definite matrix with a one in the upper left hand element that solves

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A(\boldsymbol{u}_i - E[\boldsymbol{u}_i])}{\|A(\boldsymbol{u}_i - E[\boldsymbol{u}_i])\|}\right)\left(\frac{A(\boldsymbol{u}_i - E[\boldsymbol{u}_i])}{\|A(\boldsymbol{u}_i - E[\boldsymbol{u}_i])\|}\right)' = \frac{1}{d}I \tag{2.7}$$

Equation (2.7) 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 errors 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

$$D'\hat{A}_{D\boldsymbol{u}}'\hat{A}_{D\boldsymbol{u}}D = c_0\,\hat{A}_{\boldsymbol{u}}'\hat{A}_{\boldsymbol{u}} \tag{2.8}$$

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

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 and asymptotically normal.

Before giving the algorithm that computes $\hat{A}$ we need the following definitions on the norm of matrices; see Golub et al (1983).

Definition 2.1.1. For an $m \times n$ matrix C define

1. The Frobenius norm of C

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

2. The 2-norm of C

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

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

The two norms are related through the inequality $\|C\| \leq \|C\|_F$.

In the following steps we describe the algorithm that computes $\hat{A}$ on a random sample $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ where $E[\boldsymbol{u}_i] = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{u}_i) = \boldsymbol{\Sigma}$, $\boldsymbol{\Sigma} > \boldsymbol{0}$.

Step I. Compute

$$\boldsymbol{S}_0 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|} \right) \left(\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|} \right)'$$

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$^{\text{th}}$ iteration, form

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

and

$$\boldsymbol{S}_t = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\hat{A}_{dt} \boldsymbol{u}_i}{\|\hat{A}_{dt} \boldsymbol{u}_i\|} \right) \left(\frac{\hat{A}_{dt} \boldsymbol{u}_i}{\|\hat{A}_{dt} \boldsymbol{u}_i\|} \right)'.$$

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.

Using model (1.1), the matrix A solves the equation

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A(\boldsymbol{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\mathcal{B}}_1'\boldsymbol{x}_i)}{\|A(\boldsymbol{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\mathcal{B}}_1'\boldsymbol{x}_i)\|}\right)\left(\frac{A(\boldsymbol{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\mathcal{B}}_1'\boldsymbol{x}_i)}{\|A(\boldsymbol{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\mathcal{B}}_1'\boldsymbol{x}_i)\|}\right)' = \frac{1}{d}I. \tag{2.9}$$

Equivalently, $\hat{A}$ is the solution to the equation

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A\boldsymbol{\varepsilon}_i}{\|A\boldsymbol{\varepsilon}_i\|}\right)\left(\frac{A\boldsymbol{\varepsilon}_i}{\|A\boldsymbol{\varepsilon}_i\|}\right)' = \frac{1}{d}I. \tag{2.10}$$

Thus to be able to compute $\hat{A}$ we need an initial estimate to the true errors. Since the LS estimate $\widehat{\boldsymbol{\mathcal{B}}}_{LS}$ is affine equivariant it is appropriate to compute $\hat{A}$ on the LS residuals $\hat{\boldsymbol{\varepsilon}}_{1,LS}, \ldots, \hat{\boldsymbol{\varepsilon}}_{n,LS}$.

2.1.3 Estimation

The first step in estimation is to transform the response variables $\boldsymbol{y}_i$'s to $\boldsymbol{z}_i = \hat{A}\boldsymbol{y}_i$ for $i = 1, \ldots, n$. Next the R-estimate, $\widehat{\boldsymbol{\mathcal{B}}}_R = \left(\widehat{\boldsymbol{\beta}}_{0,s} \;\; \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}'\right)'$, is computed on $(\boldsymbol{x}_1, \boldsymbol{z}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{z}_n)$. The last step is to retransform $\widehat{\boldsymbol{\mathcal{B}}}_R$ to obtain the new estimate $\widehat{\boldsymbol{\mathcal{B}}}_{TRR} = \widehat{\boldsymbol{\mathcal{B}}}_R(\hat{A}')^{-1}$.

2.2 Affine Equivariance of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$

In this section the affine equivariance of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ is established. See Maronna et. al. (1986) and Oja et. al. (2002).

Lemma 2.2.1. *The estimate $\widehat{\boldsymbol{\mathcal{B}}}_{TRR} = \widehat{\boldsymbol{\mathcal{B}}}_{TRR}(\boldsymbol{x}, \boldsymbol{y})$ is affine equivariant in the sense it has the properties*

1. $\boldsymbol{y}$- affine equivariance. If D is a fixed $d \times d$ nonsingular matrix and $\boldsymbol{b}$ is a $d \times 1$ constant vector, then

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

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

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

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

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

Proof. In the following proofs we assume that $\mathbf{X}$ is centered.

(1) $\boldsymbol{y}$- affine equivariance. Let $\widehat{\boldsymbol{\mathcal{B}}}_{LS} = \widehat{\boldsymbol{\mathcal{B}}}_{LS}(\boldsymbol{x}, \boldsymbol{y})$, $\hat{\boldsymbol{\varepsilon}}_{LS} = \hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x}, \boldsymbol{y})$ and $\hat{A} = \hat{A}_{\hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x},\boldsymbol{y})}$. Transform the response variables $\boldsymbol{y}_i \to D\boldsymbol{y}_i + \boldsymbol{b}$. Then the data matrix of the transformed response variables is

$$\begin{pmatrix} (D\boldsymbol{y}_1 + \boldsymbol{b})' \\ \vdots \\ (D\boldsymbol{y}_n + \boldsymbol{b})' \end{pmatrix} = \mathbf{Y}D' + \mathbf{1}_n \boldsymbol{b}'$$

The LS regression on $(\boldsymbol{x}_1, D\boldsymbol{y}_1), \ldots, (\boldsymbol{x}_n, D\boldsymbol{y}_n)$ gives the estimate

$$\begin{aligned}
\widehat{\boldsymbol{\mathcal{B}}}_{LS}(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b}) &= (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'(\mathbf{Y}D' + \mathbf{1}_n\boldsymbol{b}') \\
&= \widehat{\boldsymbol{\mathcal{B}}}_{LS}D' + \begin{pmatrix} n^{-1} & 0 \\ 0 & (\mathbf{X}'\mathbf{X})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_n' \\ \mathbf{X}' \end{pmatrix} \mathbf{1}_n\boldsymbol{b} \\
&= \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{0,LS}'D' + \boldsymbol{b}' \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}D' \end{pmatrix}.
\end{aligned}$$

Consequently, the LS residuals for $i = 1, \ldots, n$ of this model are

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}}_{i,LS}(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b}) &= (D\boldsymbol{y}_i + \boldsymbol{b}) - \widehat{\boldsymbol{\beta}}_{0,LS}(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b}) - \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b})\boldsymbol{x}_i \\
&= (D\boldsymbol{y}_i + \boldsymbol{b}) - (D\widehat{\boldsymbol{\beta}}_{0,LS} + \boldsymbol{b}) - D\widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'\boldsymbol{x}_i \\
&= D\hat{\boldsymbol{\varepsilon}}_{i,LS}.
\end{aligned}$$

By the uniqueness of Tyler's $\hat{A}$, there exists a positive constant c_0 such that

$$\hat{A}_{\hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})}D = c_0\hat{A}$$

or

$$D'\hat{A}'_{\hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})}\hat{A}_{\hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})}D = c_0^2\,\hat{A}'\hat{A}.$$

Letting $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{x}, \boldsymbol{y}) = \hat{A}\boldsymbol{y}$, we see for all $i = 1, \ldots, n$ that

$$\begin{aligned}
\boldsymbol{z}_i(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b}) &= \hat{A}_{\hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})}(D\boldsymbol{y}_i + \boldsymbol{b}) \\
&= c_0\hat{A}D^{-1}(D\boldsymbol{y}_i + \boldsymbol{b}) \\
&= c_0(\boldsymbol{z}_i + \hat{A}D^{-1}\boldsymbol{b}).
\end{aligned}$$

Now the R regression coefficient matrix is obtained by minimizing for $j = 1, \ldots, d$

the dispersion functions

$$\begin{aligned}
D(\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})) &= \sum_{i=1}^{n} a(R[\boldsymbol{z}_i^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})]) \\
&\qquad \times [\boldsymbol{z}_i^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})] \\
&= c_0 \sum_{i=1}^{n} a(R[\boldsymbol{z}_i^{(j)} - \frac{1}{c_0}\,\boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})]) \\
&\qquad \times [\boldsymbol{z}_i^{(j)} - \frac{1}{c_0}\,\boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})]
\end{aligned}$$

because $\sum_{i=1}^{n} a(i) = 0$ and the ranks are invariant under scalar translation and multiplication by a positive constant. It follows then that the estimate of the slopes matrix,

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) = c_0\,\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}$$

and the estimate of the intercept vector,

$$\begin{aligned}
\widehat{\beta}_{0,s}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) &= \operatorname*{med}_{1\le i\le n}\,[\boldsymbol{z}_i(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) - \boldsymbol{x}_i'\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(j)}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})] \\
&= \operatorname*{med}_{1\le i\le n}\,[c_0(\boldsymbol{z}_i^{(j)} - \boldsymbol{x}_i'\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(j)}) + c_0(\hat{A}D^{-1}\boldsymbol{b})_j] \\
&= c_0\,[\widehat{\boldsymbol{\beta}}_{0,s}^{(j)} + (\hat{A}D^{-1}\boldsymbol{b})_j].
\end{aligned}$$

Stacking the intercept components gives

$$\widehat{\boldsymbol{\beta}}_{0,s}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) = c_0(\widehat{\boldsymbol{\beta}}_{0,s} + \hat{A}D^{-1}\boldsymbol{b}).$$

On the retransformation step of $\widehat{\boldsymbol{\mathcal{B}}}_R(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})$ we get

$$\begin{aligned}
\widehat{\boldsymbol{\mathcal{B}}}_{TRR}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b}) &= \widehat{\boldsymbol{\mathcal{B}}}_R(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})(\hat{A}'_{\hat{\boldsymbol{\varepsilon}}_{LS}(\boldsymbol{x}, D\boldsymbol{y}+\boldsymbol{b})})^{-1} \\
&= c_0 \begin{pmatrix} \widehat{\boldsymbol{\beta}}'_{0,s} + (\hat{A}D^{-1}\boldsymbol{b})' \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \end{pmatrix} ((c_0\hat{A}D^{-1})')^{-1} \\
&= \widehat{\boldsymbol{\mathcal{B}}}_R(\hat{A}')^{-1}D' + \begin{pmatrix} \boldsymbol{b}' \\ \boldsymbol{0} \end{pmatrix} \\
&= \begin{pmatrix} \widehat{\boldsymbol{\beta}}'_{0,TRR}D' + \boldsymbol{b}' \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,TRR}D' \end{pmatrix}.
\end{aligned}$$

$\boldsymbol{y}$-Affine equivariance is an important property of the estimate. It enables us to transform models that have errors with elliptical distributions to the ones with spherical distribution errors. Performing then a componentwise regression on the later models is legitimate.

(2) regression affine equivariance. The data matrix of the response variables upon transforming $\boldsymbol{y}_i \to \boldsymbol{y}_i - G'\boldsymbol{x}_i$ is

$$\begin{pmatrix} (\boldsymbol{y}_1 - G'\boldsymbol{x}_1)' \\ \vdots \\ (\boldsymbol{y}_n - G'\boldsymbol{x}_n)' \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}'_1 \\ \vdots \\ \boldsymbol{y}'_n \end{pmatrix} - \begin{pmatrix} \boldsymbol{x}'_1 G \\ \vdots \\ \boldsymbol{x}'_n G \end{pmatrix} = \mathbf{Y} - \mathbf{X}G.$$

Consequently, the LS estimate becomes

$$\widehat{\mathcal{B}}_{LS}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'(\mathbf{Y} - \mathbf{X}G)$$

$$= \begin{pmatrix} n^{-1}\mathbf{1}_n' \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} \end{pmatrix} (\mathbf{Y} - \mathbf{X}G)$$

$$= \widehat{\mathcal{B}}_{LS} - \begin{pmatrix} \mathbf{0} \\ G \end{pmatrix}$$

$$= \begin{pmatrix} \widehat{\beta}_{0,LS}' \\ \widehat{\mathcal{B}}_{1,LS} - G \end{pmatrix}.$$

For $i = 1, \ldots, n$ the LS residuals are

$$\hat{\varepsilon}_{i,LS}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = (\boldsymbol{y}_i - G'\boldsymbol{x}_i) - \widehat{\beta}_{0,LS}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) - \widehat{\mathcal{B}}_{1,LS}'(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})\boldsymbol{x}_i$$

$$= \boldsymbol{y}_i - G'\boldsymbol{x}_i - \widehat{\beta}_{0,LS} - (\widehat{\mathcal{B}}_{1,LS} - G)'\boldsymbol{x}_i$$

$$= \hat{\varepsilon}_{i,LS}.$$

This implies that $\hat{A}_{\hat{\varepsilon}_{LS}(\boldsymbol{x},\boldsymbol{y}-G'\boldsymbol{x})} = \hat{A}$. Now, in the transformation step, for all $i = 1, \ldots, n$

$$\boldsymbol{z}_i(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = \hat{A}_{\hat{\varepsilon}_{LS}(\boldsymbol{x},\boldsymbol{y}-G'\boldsymbol{x})}(\boldsymbol{y}_i - G'\boldsymbol{x}_i)$$

$$= \hat{A}(\boldsymbol{y}_i - G'\boldsymbol{x}_i)$$

$$= \boldsymbol{z}_i - \hat{A}G'\boldsymbol{x}_i.$$

Let $(\hat{A}G')_{r_j} = \text{j}^{\text{th}}$ row of $\hat{A}G'$. To find $\widehat{\mathcal{B}}_{1,\varphi}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})$ we minimize, for $j =$

$1, \ldots, d$, the dispersion functions

$$\begin{aligned} D(\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})) &= \sum_{i=1}^{n} a(R[\boldsymbol{z}_i^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})]) \\ &\qquad \times (\boldsymbol{z}_i^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})) \\ &= \sum_{i=1}^{n} a(R[\boldsymbol{z}_i^{(j)} - \boldsymbol{x}_i'(\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) + (\hat{A}G')_{r_j}')]) \\ &\qquad (\boldsymbol{z}_i^{(j)} - \boldsymbol{x}_i'(\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) + (\hat{A}G')_{r_j}')). \end{aligned}$$

This implies that

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(j)}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(j)} - (\hat{A}G')_{r_j}'.$$

Hence,

$$\left(\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(1)} \cdots \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(d)}\right)_{(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})} = \left(\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(1)} \cdots \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}^{(d)}\right) - \left((\hat{A}G')_{r_1}' \cdots (\hat{A}G')_{r_d}'\right).$$

Equivalently,

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} - \begin{pmatrix} (\hat{A}G')_{r_1} \\ \vdots \\ (\hat{A}G')_{r_d} \end{pmatrix}'$$

or

$$\begin{aligned} \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) &= \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} - (\hat{A}G')' \\ &= \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} - G\hat{A}'. \end{aligned}$$

Now, for $j = 1, \ldots, d$, the intercepts are

$$\begin{aligned}
\widehat{\beta}^{(j)}_{0,s}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) &= \operatorname*{med}_{1\leq i\leq n}(\boldsymbol{z}^{(j)}_i(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) - \boldsymbol{x}'_i\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})) \\
&= \operatorname*{med}_{1\leq i\leq n}(\boldsymbol{z}^{(j)}_i - (\hat{A}G')_{r_j}\boldsymbol{x}_i - \boldsymbol{x}'_i(\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi} - (\hat{A}G')'_{r_j})) \\
&= \operatorname*{med}_{1\leq i\leq n}(\boldsymbol{z}^{(j)}_i - \boldsymbol{x}'_i\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi}) \\
&= \widehat{\beta}^{(j)}_{0,s}.
\end{aligned}$$

As a result

$$\widehat{\boldsymbol{\mathcal{B}}}_R(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = \begin{pmatrix} \widehat{\boldsymbol{\beta}}'_{0,s} \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} - GA' \end{pmatrix}.$$

The retransformation of $\widehat{\boldsymbol{\mathcal{B}}}_R(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})$ gives

$$\begin{aligned}
\widehat{\boldsymbol{\mathcal{B}}}_{TRR}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) &= \widehat{\boldsymbol{\mathcal{B}}}_R(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x})(\hat{A}^{-1})' \\
&= \begin{pmatrix} \widehat{\boldsymbol{\beta}}'_{0,s} \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} - G\hat{A}' \end{pmatrix}(\hat{A}^{-1})' \\
&= \left\{\begin{pmatrix} \widehat{\boldsymbol{\beta}}'_{0,s} \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ G\hat{A}' \end{pmatrix}\right\}(\hat{A}^{-1})' \\
&= \begin{pmatrix} \widehat{\boldsymbol{\beta}}'_{0,TRR} \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,TRR} - G \end{pmatrix}.
\end{aligned}$$

(3) $\boldsymbol{x}$- affine equivariance. Consider the linear model based on the transformed data points $(W\boldsymbol{x}_1 + \boldsymbol{c}, \boldsymbol{y}_1), \ldots, (W\boldsymbol{x}_n + \boldsymbol{c}, \boldsymbol{y}_n)$

$$\begin{aligned}
\boldsymbol{y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\mathcal{B}}'_1(W\boldsymbol{x}_i + \boldsymbol{c}) + \boldsymbol{\varepsilon}_i \\
&= (\boldsymbol{\beta}_0 + \boldsymbol{\mathcal{B}}'_1\boldsymbol{c}) + \boldsymbol{\mathcal{B}}'_1W\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i \\
&= \boldsymbol{\beta}^*_0 + \boldsymbol{\mathcal{B}}'_1W\boldsymbol{x}_i + \boldsymbol{\varepsilon}_i.
\end{aligned}$$

For the design matrix $\mathbb{X} = \begin{pmatrix} \mathbf{1}_n & \mathbf{X}W' \end{pmatrix}$, the LS estimate is

$$\begin{aligned}
\widehat{\boldsymbol{\mathcal{B}}}_{LS}(W\boldsymbol{x}, \boldsymbol{y}) &= (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbf{Y} \\
&= \begin{pmatrix} n^{-1}\mathbf{1}_n' \\ (W')^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \end{pmatrix} \mathbf{Y} \\
&= \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{0,LS}' \\ (W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,LS} \end{pmatrix}.
\end{aligned}$$

Hence,

$$\widehat{\boldsymbol{\beta}}_{0,LS}^{*}(W\boldsymbol{x}, \boldsymbol{y}) = \widehat{\boldsymbol{\beta}}_{0,LS}$$

This implies that

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{0,LS}(W\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y}) &= \widehat{\boldsymbol{\beta}}_{0,LS}^{*} - \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'(W\boldsymbol{x}, \boldsymbol{y})\,\boldsymbol{c} \\
&= \widehat{\boldsymbol{\beta}}_{0,LS} - (W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,LS}\,\boldsymbol{c}.
\end{aligned}$$

Consequently,

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

The residuals of this LS fit for $i = 1, \ldots, n$ are

$$\begin{aligned}
\hat{\boldsymbol{\varepsilon}}_{i,LS}(W\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y}) &= \boldsymbol{y}_i - \widehat{\boldsymbol{\beta}}_{0,LS}(W\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y}) - \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'(W\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y})(W\boldsymbol{x}_i + \boldsymbol{c}) \\
&= \boldsymbol{y}_i - \widehat{\boldsymbol{\beta}}_{0,LS} + \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'W^{-1}\mathbf{c} - \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'W^{-1}(W\boldsymbol{x}_i + \mathbf{c}) \\
&= \boldsymbol{y}_i - \widehat{\boldsymbol{\beta}}_{0,LS} - \widehat{\boldsymbol{\mathcal{B}}}_{1,LS}'\boldsymbol{x}_i \\
&= \hat{\boldsymbol{\varepsilon}}_{i,LS}.
\end{aligned}$$

Thus the transformation matrix remains unchanged, i.e. $\hat{A}_{\hat{\boldsymbol{\varepsilon}}_{LS}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})} = \hat{A}$. This implies for $i = 1, \ldots, n$ that

$$\begin{aligned} \boldsymbol{z}_i(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) &= \hat{A}_{\hat{\boldsymbol{\varepsilon}}_{LS}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})}\boldsymbol{y}_i \\ &= \hat{A}\boldsymbol{y}_i \\ &= \boldsymbol{z}_i. \end{aligned}$$

The rank multivariate regression estimate on the pairs $(W\boldsymbol{x}_1+\boldsymbol{c},\boldsymbol{y}_1), \ldots, (W\boldsymbol{x}_n+\boldsymbol{c},\boldsymbol{y}_n)$ solves for $\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})$ by minimizing the dispersion functions

$$\begin{aligned} D(\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) &= \sum_{i=1}^{n} a(R[\boldsymbol{z}_i^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) - (W\boldsymbol{x}_i+\boldsymbol{c})'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})]) \\ &\qquad \times (\boldsymbol{z}_i^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) - (W\boldsymbol{x}_i+\boldsymbol{c})'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) \\ &= \sum_{i=1}^{n} a(R[\boldsymbol{z}_i^{(j)} - \boldsymbol{x}_i'W'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})]) \\ &\qquad \times (\boldsymbol{z}_i^{(j)} - \boldsymbol{x}_i'W'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) \end{aligned}$$

because the ranks sum to zero and they are invariant under scalar translation. Hence, for $j = 1, \ldots, d$

$$\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)} = W'\boldsymbol{\mathcal{B}}_{1,\varphi}^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}).$$

So

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} = W'\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}).$$

Further, for $j = 1, \ldots, d$ the intercepts are

$$\begin{aligned}
\widehat{\beta}^{(j)}_{0,s}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) &= \operatorname*{med}_{1\le i\le n}(\boldsymbol{z}_i^{(j)}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) - (W\boldsymbol{x}_i+\boldsymbol{c})'\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) \\
&= \operatorname*{med}_{1\le i\le n}(\boldsymbol{z}_i^{(j)} - (W\boldsymbol{x}_i+\boldsymbol{c})'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi}) \\
&= \operatorname*{med}_{1\le i\le n}(\boldsymbol{z}_i^{(j)} - \boldsymbol{x}_i'\boldsymbol{\mathcal{B}}^{(j)}_{1,\varphi}) - \boldsymbol{c}'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi} \\
&= \widehat{\beta}^{(j)}_{0,s} - \boldsymbol{c}'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}^{(j)}_{1,\varphi}.
\end{aligned}$$

Therefore,

$$\widehat{\beta}'_{0,s}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) = \widehat{\beta}'_{0,s} - \boldsymbol{c}'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi}.$$

It follows then that

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

Finally on retransforming $\widehat{\boldsymbol{\mathcal{B}}}_R(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})$, we get

$$\begin{aligned}
\widehat{\boldsymbol{\mathcal{B}}}_{TRR}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) &= \widehat{\boldsymbol{\mathcal{B}}}_R(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})(\hat{A}^{-1}_{\hat{\boldsymbol{\varepsilon}}_{LS}(W\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})})' \\
&= \begin{pmatrix} \widehat{\beta}'_{0,s} - \boldsymbol{c}'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \\ (W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \end{pmatrix}(\hat{A}^{-1})' \\
&= \begin{pmatrix} \widehat{\beta}'_{0,s} \\ (W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \end{pmatrix}(\hat{A}^{-1})' - \begin{pmatrix} \boldsymbol{c}'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \\ \boldsymbol{0} \end{pmatrix}(\hat{A}^{-1})' \\
&= \begin{pmatrix} \widehat{\beta}'_{0,TRR} - \boldsymbol{c}'(W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,TRR} \\ (W')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,TRR} \end{pmatrix}.
\end{aligned}$$

$\square$

2.3 Consistency of $\hat{A}$

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 $\boldsymbol{x}_{ni} = W_n^{-1/2}\boldsymbol{x}_i$, $\boldsymbol{\mathcal{B}}_{1,n} = W_n^{1/2}\boldsymbol{\mathcal{B}}_1$, $\boldsymbol{y}_{ni} = \boldsymbol{y}_i$ and $\boldsymbol{\varepsilon}_{ni} = \boldsymbol{\varepsilon}_i$. Now, recall model (1.1)

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

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

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

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

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

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 2.3.1. *Under A7,*

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

Proof. Let

$$\begin{aligned}\mathbf{Z}_n' &= \begin{pmatrix} \boldsymbol{x}_{n1} & \cdots & \boldsymbol{x}_{nn} \end{pmatrix} \\ &= \begin{pmatrix} W_n^{-1/2}\boldsymbol{x}_1 & \cdots & W_n^{-1/2}\boldsymbol{x}_n \end{pmatrix} \\ &= W_n^{-1/2}\mathbf{X}'.\end{aligned}$$

Then

$$\begin{aligned}\boldsymbol{\mathcal{B}}_{1,n,LS}^* &= (\mathbf{Z}_n'\mathbf{Z}_n)^{-1}\mathbf{Z}_n'\mathbf{Y}_n \\ &= (W_n^{-1/2}\mathbf{X}'\mathbf{X}W_n^{-1/2})^{-1}W_n^{-1/2}\mathbf{X}'\mathbf{Y}_n \\ &= W_n^{1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= W_n^{1/2}\widehat{\boldsymbol{\mathcal{B}}}_{1,LS} \\ &= (\frac{1}{n}W_n)^{1/2}\sqrt{n}\,\widehat{\boldsymbol{\mathcal{B}}}_{1,LS} \\ &= O_p(1)\end{aligned}$$

because $\sqrt{n}\,\widehat{\boldsymbol{\mathcal{B}}}_{1,LS} = O_p(1)$ and $\frac{1}{n}W_n \to \mathbf{V}$ by A7. □

We also need to prove the following simple lemma.

Lemma 2.3.2. *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. If $\boldsymbol{a} \perp (\boldsymbol{b}-\boldsymbol{a})$ then we can write

$$\|\boldsymbol{b}\|^2 = \|\boldsymbol{b}-\boldsymbol{a}\|^2 + \|\boldsymbol{a}\|^2.$$

Further, $\|\boldsymbol{b}-\boldsymbol{a}\| \geq 0$ implies that $\|\boldsymbol{a}\|/\|\boldsymbol{b}\| \leq 1$. Now

$$\begin{aligned}\left\| \frac{(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{a})'}{\|\boldsymbol{b}-\boldsymbol{a}\|^2} - \frac{\boldsymbol{b}\boldsymbol{b}'}{\|\boldsymbol{b}\|^2} \right\|_F &= \left\| \frac{\|\boldsymbol{b}\|^2(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{a})' - \|\boldsymbol{b}-\boldsymbol{a}\|^2\boldsymbol{b}\boldsymbol{b}'}{\|\boldsymbol{b}\|^2\|\boldsymbol{b}-\boldsymbol{a}\|^2} \right\|_F \\ &= \left\| \frac{\{\|\boldsymbol{b}-\boldsymbol{a}\|^2 + \|\boldsymbol{a}\|^2\}(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{a})' - \|\boldsymbol{b}-\boldsymbol{a}\|^2\boldsymbol{b}\boldsymbol{b}'}{\|\boldsymbol{b}\|^2\|\boldsymbol{b}-\boldsymbol{a}\|^2} \right\|_F\end{aligned}$$

It is easy to show that,

$$\|AB\|_F \leq \|A\|_F\,\|B\|_F, \quad A \in \Re^{m\times n}, B \in \Re^{n\times q}.$$

Hence, after some simplifications, we get

$$\begin{aligned}\left\| \frac{(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{a})'}{\|\boldsymbol{b}-\boldsymbol{a}\|^2} - \frac{\boldsymbol{b}\boldsymbol{b}'}{\|\boldsymbol{b}\|^2} \right\|_F &\leq 2\frac{\|\boldsymbol{a}\|}{\|\boldsymbol{b}\|} + 2\frac{\|\boldsymbol{a}\|^2}{\|\boldsymbol{b}\|^2} \\ &\leq 4\frac{\|\boldsymbol{a}\|}{\|\boldsymbol{b}\|}.\end{aligned}$$

$\square$

In his paper, Tyler (1987) mentioned that application of his theorems to a continuous population rather than a sample insures that there exists unique A such that

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

Now, in order to prove the consistency of $\hat{A}$ computed on the LS residuals we need the following conditions to be satisfied

A8. $E\left(\frac{1}{\|\varepsilon\|}\right) < \infty$.

A9. $E\left(\frac{1}{\|A\varepsilon\|}\right) < \infty$.

Now let

$$S_0^* = \frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_{ni}\varepsilon_{ni}'}{\|\varepsilon_{ni}\|^2}$$

and

$$S_t^* = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{A}_{dt}\varepsilon_{ni}\varepsilon_{ni}'\hat{A}_{dt}'}{\|\hat{A}_{dt}\varepsilon_{ni}\|^2}.$$

Lemma 2.3.3. *Under conditions A4 and A7–A9*

$$\|S_0 - S_0^*\|_F = o_p(1) \tag{2.11}$$

and

$$\|S_t - S_t^*\|_F = o_p(1) \tag{2.12}$$

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

$$\lim_{n\to\infty} v_n = \infty, \quad \lim_{n\to\infty} v_n d_n^* = 0. \tag{2.13}$$

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

$$\begin{aligned}
\|S_0 - S_0^*\|_F &= \left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\varepsilon}_{ni}\hat{\varepsilon}_{ni}'}{\|\hat{\varepsilon}_{ni}\|^2} - \sum_{i=1}^{n}\frac{\varepsilon_{ni}\varepsilon_{ni}'}{\|\varepsilon_{ni}\|^2}\right\|_F \\
&\le \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\hat{\varepsilon}_{ni}\hat{\varepsilon}_{ni}'}{\|\hat{\varepsilon}_{ni}\|^2} - \frac{\varepsilon_{ni}\varepsilon_{ni}'}{\|\varepsilon_{ni}\|^2}\right\|_F \\
&\le \frac{4}{n}\sum_{i=1}^{n}\frac{\|(\mathcal{B}_{1,n,LS}^*)'x_{ni}\|_F}{\|\varepsilon_{ni}\|} \\
&\le \frac{4}{n}\sum_{i=1}^{n}\frac{\|\mathcal{B}_{1,n,LS}^*\|_F\|x_{ni}\|}{\|\varepsilon_{ni}\|} \\
&\le 4d_n^* v_n \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\|\varepsilon_{ni}\|} + 4d_n^* I(\|\mathcal{B}_{1,n,LS}^*\|_F \ge v_n)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\|\varepsilon_{ni}\|},
\end{aligned}$$

where $I(.)$ denotes the indicator function. Since $E(\|\varepsilon\|^{-1}) < \infty$ and $d_n^* v_n \to 0$, from SLLN the first term converges to 0 in probability. The same is true for the second term because $\mathcal{B}_{1,n,LS}^* = O_p(1)$, $d_n^* \to 0$ and $E(\|\varepsilon\|^{-1}) < \infty$. It is enough to show that $\|S_1 - S_1^*\|_F = o_p(1)$ since (2.12) follows by a similar argument. From the algorithm, we

have $\hat{A}_0 = \text{Chol}(\boldsymbol{S}_0^{-1})$. On the other hand, there exists a unique $A_0 = \text{Chol}((\boldsymbol{S}_0^*)^{-1})$. Consequently, (2.11) and the continuity of the Cholesky decomposition implies that

$$\|\hat{A}_0 - A_0\|_F = o_p(1).$$

Thus, for all $i = 1, \ldots, n$

$$\begin{aligned}\|\hat{A}_0\boldsymbol{\varepsilon}_{ni} - A_0\boldsymbol{\varepsilon}_{ni}\|_F &\le \|\hat{A}_0 - A_0\|_F \, \|\boldsymbol{\varepsilon}_{ni}\| \\ &\xrightarrow{p} 0.\end{aligned}$$

Hence, $\forall \eta > 0$ and $i = 1, \ldots, n$

$$\|\hat{A}_0\boldsymbol{\varepsilon}_{ni}\| \ge \|A_0\boldsymbol{\varepsilon}_{ni}\| - \eta.$$

Consequently,

$$\begin{aligned}\|\boldsymbol{S}_1 - \boldsymbol{S}_1^*\|_F &= \left\| \frac{1}{n}\sum_{i=1}^n \frac{\hat{A}_0\hat{\boldsymbol{\varepsilon}}_{ni}\hat{\boldsymbol{\varepsilon}}_{ni}'\hat{A}_0'}{\|\hat{A}_0\hat{\boldsymbol{\varepsilon}}_{ni}\|^2} - \sum_{i=1}^n \frac{\hat{A}_0\boldsymbol{\varepsilon}_{ni}\boldsymbol{\varepsilon}_{ni}'\hat{A}_0'}{\|\hat{A}_0\boldsymbol{\varepsilon}_{ni}\|^2} \right\|_F \\ &\le \frac{1}{n}\sum_{i=1}^n \left\| \frac{\hat{A}_0\hat{\boldsymbol{\varepsilon}}_{ni}\hat{\boldsymbol{\varepsilon}}_{ni}'\hat{A}_0'}{\|\hat{A}_0\hat{\boldsymbol{\varepsilon}}_{ni}\|^2} - \frac{\hat{A}_0\boldsymbol{\varepsilon}_{ni}\boldsymbol{\varepsilon}_{ni}'\hat{A}_0'}{\|\hat{A}_0\boldsymbol{\varepsilon}_{ni}\|^2} \right\|_F \\ &\le \frac{4}{n}\sum_{i=1}^n \frac{\|\hat{A}_0(\boldsymbol{\mathcal{B}}_{1,n,LS}^*)'\boldsymbol{x}_{ni}\|_F}{\|\hat{A}_0\boldsymbol{\varepsilon}_{ni}\|^2} \\ &\le 4\|\hat{A}_0\|_F \, \|\boldsymbol{\mathcal{B}}_{1,n,LS}^*\|_F d_n^* \frac{1}{n}\sum_{i=1}^n \frac{1}{\|\hat{A}_0\boldsymbol{\varepsilon}_{ni}\|^2} \\ &\le 4(\|A_0\|_F + \eta)d_n^* v_n \frac{1}{n}\sum_{i=1}^n \frac{1}{\|A_0\boldsymbol{\varepsilon}_{ni}\| - \eta} \\ &\quad + 4(\|A_0\|_F + \eta)d_n^* I(\|\boldsymbol{\mathcal{B}}_{1,n,LS}^*\|_F \ge v_n)\frac{1}{n}\sum_{i=1}^n \frac{1}{\|A_0\boldsymbol{\varepsilon}_{ni}\| - \eta}.\end{aligned}$$

Letting $\eta \to \mathbf{0}$, the first term converges to 0 in probability follows from the SLLN. The same is true for the second term. This proves that $\|\boldsymbol{S}_1 - \boldsymbol{S}_1^*\|_F = o_p(1)$. The proof for (2.12) follows by noting that at the t^{th} step, the matrix

$$\hat{A}_{dt} = \hat{A}_{t-1}\hat{A}_{t-2}\cdots\hat{A}_0$$

converges in probability to

$$A_{dt} = A_{t-1}A_{t-2}\cdots A_0.$$

$\square$

This last lemma proves that although the LS residuals are correlated, $\|\boldsymbol{S}_t - \boldsymbol{S}_t^*\|_F = o_p(1)$. Consequently, the consistency of $\hat{A}$ follows from Tyler's Theorem. He showed that the sequence $(\hat{A}_{dt})_t$ converges *almost surely* to a matrix A such that

$$E\left[\frac{A\varepsilon\varepsilon' A'}{\|A\varepsilon\|^2}\right] = \frac{1}{d}I.$$

Theorem 2.3.1 (Tyler, 1987). *If the sample $\{\varepsilon_i, i = 1, \ldots, n\}$ represents a random sample from a continuous distribution, then $\hat{A}_{dt} \to A$ almost surely.*

2.4 Asymptotic Normality

The asymptotic properties of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ are inherited from the componentwise rank regression estimate $\widehat{\boldsymbol{\mathcal{B}}}_R$. Therefore we need first to present the asymptotic results of the later estimate as given by Davis and McKean (1993). For this we need to introduce some notations and terminology.

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 the left direct product, see Graybill (1983), of A and B, written as $A \otimes B$ is a matrix

C of size $m_1m_2 \times n_1n_2$ defined by

$$C = \begin{pmatrix} b_{11}A & b_{12}A & \cdots & b_{1n_1}A \\ b_{21}A & b_{22}A & \cdots & b_{2n_1}A \\ \vdots & \vdots & \vdots & \vdots \\ b_{m_11}A & b_{m_12}A & \vdots & b_{m_1n_1}A \end{pmatrix}.$$

Now, 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$^{\text{th}}$ diagonal element is τ_j, τ_j^* respectively. The assumption on f_j to have finite Fisher information ensures that τ_j is finite. Recall that $\boldsymbol{S} = (s_{jj'})$ where

$$s_{jj'} = \mathrm{Cov}(\varphi(F_j(\varepsilon_{ij}),\ \varphi(F_{j'}(\varepsilon_{ij'}))$$

and let $\boldsymbol{S}^* = (s_{jj'}^*)$ where

$$s_{jj'}^* = \Pr(\varepsilon_{ij} < 0, \varepsilon_{ij'} < 0) + \Pr(\varepsilon_{ij} > 0, \varepsilon_{ij'} > 0) - \Pr(\varepsilon_{ij} < 0, \varepsilon_{ij'} > 0) - \Pr(\varepsilon_{ij} > 0, \varepsilon_{ij'} < 0).$$

Further, let

$$\boldsymbol{\Phi} = \Big(\varphi(F_j(\varepsilon_{ij}))\Big)$$

and

$$\boldsymbol{\Phi}^* = \Big(\mathrm{sgn}(\varepsilon_{ij})\Big).$$

Let

$$\hat{A}_n = \begin{pmatrix} \hat{\boldsymbol{a}}'_{n1} \\ \vdots \\ \hat{\boldsymbol{a}}'_{nd} \end{pmatrix} \quad \text{and} \quad A_0 = \begin{pmatrix} \boldsymbol{a}'_{01} \\ \vdots \\ \boldsymbol{a}'_{0d} \end{pmatrix},$$

where $\hat{\boldsymbol{a}}'_{nj}$ and $\boldsymbol{a}'_{0j}$ denote the jth rows of $\hat{A}_n$ and A_0, respectively, and A_0 satisfies 2.7. Now consider the dispersion function of the transformed model

$$\begin{aligned} D_{\hat{A}_n}(\boldsymbol{\Delta}) &= \text{tr }\, ((\mathbf{Y} - \mathbf{C}\boldsymbol{\Delta})\hat{A}'_n)' A((\mathbf{Y} - \mathbf{C}\boldsymbol{\Delta})\hat{A}'_n) \\ &= \sum_{j=1}^{d}\sum_{i=1}^{n} a(R((\boldsymbol{y}'_i - \boldsymbol{c}'_i\boldsymbol{\Delta})\hat{\boldsymbol{a}}_{nj}))(\boldsymbol{y}'_i - \boldsymbol{c}'_i\boldsymbol{\Delta})\hat{\boldsymbol{a}}_{nj} \\ &= \sum_{j=1}^{d} D_{\hat{A}_n,j}(\boldsymbol{\Delta}). \end{aligned}$$

The negative of the gradient of $D_{\hat{A}_n}(\boldsymbol{\Delta})$ is

$$\begin{aligned} L_{\hat{A}_n}(\boldsymbol{\Delta}) &= -\frac{\partial}{\partial \boldsymbol{\Delta}} D_{\hat{A}_n}(\boldsymbol{\Delta}) \\ &= -\sum_{j=1}^{d} \frac{\partial}{\partial \boldsymbol{\Delta}} D_{\hat{A}_n,j}(\boldsymbol{\Delta}) \\ &= \sum_{j=1}^{d}\sum_{i=1}^{n} a(R((\boldsymbol{y}'_i - \boldsymbol{c}'_i\boldsymbol{\Delta})\hat{\boldsymbol{a}}_{nj}))\, \boldsymbol{c}_i\hat{\boldsymbol{a}}'_{nj} \\ &= \sum_{j=1}^{d} \mathbf{C}' A^{(j)}((\mathbf{Y} - \mathbf{C}\boldsymbol{\Delta})\hat{A}'_n)\, \hat{\boldsymbol{a}}'_{nj} \\ &= \mathbf{C}' A((\mathbf{Y} - \mathbf{C}\boldsymbol{\Delta})\hat{A}'_n)\hat{A}_n. \end{aligned}$$

To obtain the asymptotic distribution of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ we need to prove several claims. By the affine equivariance of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ we can assume without loss of generality that $\boldsymbol{\Sigma} = I_d$. This implies that the true $A_0 = I_d$. For $j = 1, \ldots, d$ let

$$p_{0,j} = \prod_{i=1}^{n} f_j(y_{ij}) \tag{2.14}$$

and

$$p_{\Delta,j} = \prod_{i=1}^{n} f_j(y_{ij} - \boldsymbol{c}_i'\boldsymbol{\Delta}\boldsymbol{\delta}_j), \tag{2.15}$$

where $\boldsymbol{\delta}_j \in \Re^d$ has one in the jth position and zero otherwise. Then as proved in Hettmansperger and McKean (1998) page 401, $p_{\Delta,j}$ is contiguous to $p_{0,j}$. For the next part we need the following lemma due to Peters and Randles (1990).

Lemma 2.4.1. *Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be independent and identically distributed as $\mathbf{X}$, where $\mathbf{X}$ has a distribution function $F_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)'$. Assume that $g(\mathbf{X}, \boldsymbol{\alpha})$ is a real valued function with $P(g(\mathbf{X}, \boldsymbol{\alpha}) \leq t) = G_{\boldsymbol{\alpha}}(t)$ and that $G_{\boldsymbol{\alpha}}(t)$ is continuous in t. Let $\widehat{\boldsymbol{\alpha}}_n = (\widehat{\alpha}_{n1}, \ldots, \widehat{\alpha}_{nd})$ be estimator of $\boldsymbol{\alpha}$. Suppose that*

(i) $n^{1/2}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}) = O_p(1)$, *as* $n \to \infty$,

(ii) $\sup_{\boldsymbol{s} \in \beta} |g(\mathbf{X}, \boldsymbol{\alpha} + n^{-1/2}\boldsymbol{s}) - g(\mathbf{X}, \boldsymbol{\alpha})| = o_p(1)$, *as $n \to \infty$, for any bounded sphere β in $\Re^d$.*

Let

$$G_{n,\widehat{\boldsymbol{\alpha}}_n}(t) = \frac{1}{n}\sum_{i=1}^{n} I[g(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_n) \leq t]$$

denote the empirical distribution function of the $g(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_n)$'s. Then for every positive integer i,

$$G_{n,\widehat{\boldsymbol{\alpha}}_n}(g(\mathbf{X}_i, \widehat{\boldsymbol{\alpha}}_n)) - G_{\boldsymbol{\alpha}}(g(\mathbf{X}_i, \boldsymbol{\alpha})) = o_p(1),$$

as $n \to \infty$.

Lemma 2.4.2.

$$\mathbf{C}'A(\mathbf{Y}\hat{A}_n') - \mathbf{C}'A(\mathbf{Y}A_0') \xrightarrow{p_0} 0. \tag{2.16}$$

Proof. Under the null hypothesis $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are independent and identically distributed. Let l_{kj} denote the kjth element of the left-hand side. Then

$$l_{kj} = \sum_{i=1}^{n} c_{ik}\, \varphi\left(\frac{R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj})}{n+1}\right) - \sum_{i=1}^{n} c_{ik}\, \varphi\left(\frac{R(\boldsymbol{y}_i'\boldsymbol{a}_{0j})}{n+1}\right)$$

We need to show that $l_{kj} \xrightarrow{p} 0$. Let G_j denote the distribution function of the jth column of $\mathbf{Y}A_0'$. We will first show that for each $i = 1, \ldots, n$

$$G_{nj}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj}) - G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j}) \xrightarrow{p} 0, \tag{2.17}$$

where

$$G_{nj}(t) = \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj} \le t),$$

so that

$$G_{nj}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj}) = \frac{1}{n}R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj}).$$

To apply lemma 2.4.1 to (2.17) we need to prove the following conditions

$$\sqrt{n}(\hat{\boldsymbol{a}}_{nj} - \boldsymbol{a}_{0j}) = O_p(1) \tag{2.18}$$

$$\sup_{\boldsymbol{s}\in\beta} |\boldsymbol{y}_i'(\boldsymbol{a}_{0j} + n^{-1/2}\boldsymbol{s}) - \boldsymbol{y}_i'\boldsymbol{a}_{0j}| = o_p(1), \tag{2.19}$$

where β is a bounded sphere in $\Re^d$. Condition (2.18) is satisfied because Tyler (1987) proved that

$$\sqrt{n}\,(\hat{A}_n - A_0) = O_p(1).$$

For condition (2.19), let $M > 0$ be a bound on the vectors of β. Then for $\boldsymbol{s} \in \beta$ we have

$$\begin{aligned}
|\boldsymbol{y}_i'(\boldsymbol{a}_{0j} + n^{-1/2}\boldsymbol{s}) - \boldsymbol{y}_i'\boldsymbol{a}_{0j}| &= |n^{-1/2}\boldsymbol{y}_i'\boldsymbol{s}| \\
&\le n^{-1/2}\|\boldsymbol{y}_i\|\,\|\boldsymbol{s}\| \\
&\le n^{-1/2}M\|\boldsymbol{y}_i\|.
\end{aligned}$$

Thus,

$$\sup_{s \in \beta} |\boldsymbol{y}_i'(\boldsymbol{a}_{0j} + n^{-1/2}\boldsymbol{s}) - \boldsymbol{y}_i'\boldsymbol{a}_{0j}| \xrightarrow{p} 0.$$

It follows now by lemma 2.4.1 that for each $i = 1, \ldots, n$

$$\frac{1}{n} R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj}) - G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j}) \xrightarrow{p} 0. \tag{2.20}$$

Next we claim that

$$l_{kj}^* = \sum_{i=1}^{n} c_{ik}\, \varphi\left(\frac{R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj})}{n+1}\right) - \sum_{i=1}^{n} c_{ik}\, \varphi\left(G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j})\right) \xrightarrow{p} 0 \tag{2.21}$$

Let $w_{ik} = \sqrt{n}\, c_{ik}$. Then since $\mathbf{C}'\mathbf{C} = I_p$ we have

$$\frac{1}{n} \sum_{i=1}^{n} w_{ik}^2 = \sum_{i=1}^{n} c_{ik}^2 = 1. \tag{2.22}$$

Also,

$$\frac{\max_{1 \le i \le n} w_{ik}^2}{\sum_{i=1}^{n} w_{ik}^2} = \max_{1 \le i \le n} c_{ik}^2 \longrightarrow 0. \tag{2.23}$$

Consequently,

$$\begin{aligned} E[{l_{kj}^*}^2] &= \frac{1}{n} E\left[\left(\sum_{i=1}^{n} w_{ik} \left\{\varphi\left(\frac{R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj})}{n+1}\right) - \varphi\left(G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j})\right)\right\}\right)^2\right] \\ &\le \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} w_{ik}^2\right) E\left[\left(\varphi\left(\frac{R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj})}{n+1}\right) - \varphi\left(G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j})\right)\right)^2\right], \end{aligned}$$

where the inequality is in Hájek and Šidák (1967) page 160. Now by (2.20) and the continuity of φ we have

$$\left(\varphi\left(\frac{R(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj})}{n+1}\right) - \varphi\left(G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j})\right)\right)^2 \xrightarrow{p} 0.$$

Moreover, φ is bounded and thus by the Lebesgue Dominated Convergence the expected value of the above expression goes to 0 for sufficiently large n. Consequently, by (2.22)

we have $E[l_{kj}^{*\,2}] \to 0$ as $n \to \infty$. Therefore, $l_{kj}^* \xrightarrow{p} 0$. In a similar way it can be shown, see Hettmansperger and McKean (1998) page 400, that

$$l_{kj}^{**} = \sum_{i=1}^{n} c_{ik}\, \varphi\left(\frac{R(\boldsymbol{y}_i'\boldsymbol{a}_{0j})}{n+1}\right) - \sum_{i=1}^{n} c_{ik}\, \varphi\left(G_j(\boldsymbol{y}_i'\boldsymbol{a}_{0j})\right) \xrightarrow{p} 0 \tag{2.24}$$

Hence, we can conclude from (2.21) and (2.24) that $l_{kj} \xrightarrow{p} 0$. □

Let

$$L_{A_0}(\boldsymbol{\Delta}) = \mathbf{C}'A((\mathbf{Y} - C\boldsymbol{\Delta})A_0')A_0.$$

Then by lemma 2.4.2 and the consistency of $\hat{A}_n$ we have the following results.

Corollary 2.4.1.

$$L_{\hat{A}_n}(\mathbf{0}) - L_{A_0}(\mathbf{0}) \xrightarrow{p_0} 0.$$

Corollary 2.4.2.

$$L_{\hat{A}_n}(\boldsymbol{\Delta}) - L_{A_0}(\boldsymbol{\Delta}) \xrightarrow{p_0} 0.$$

Proof. By contiguity we have

$$L_{\hat{A}_n}(\mathbf{0}) - L_{A_0}(\mathbf{0}) \xrightarrow{p_{\boldsymbol{\Delta}}} 0.$$

Hence, for all $\epsilon > 0$

$$P_0(\|L_{\hat{A}_n}(\boldsymbol{\Delta}) - L_{A_0}(\boldsymbol{\Delta})\| \geq \epsilon) = P_{\boldsymbol{\Delta}}(L_{\hat{A}_n}(\mathbf{0}) - L_{A_0}(\mathbf{0})\| \geq \epsilon)$$
$$\to 0 \quad \text{as} \quad n \to \infty.$$

□

Theorem 2.4.1.

$$L_{\hat{A}_n}(\mathbf{0}) \xrightarrow{D} N_{p,d}(\mathbf{0}, I_p, \boldsymbol{S}).$$

Proof. By Theorem 2.1 of Davis and McKean (1993) we have

$$L(\mathbf{0}) \xrightarrow{D} N_{p,d}(\mathbf{0}, I_p, \boldsymbol{S}).$$

The result now follows from corollary 2.4.1. □

Theorem 2.4.2. *For all $\epsilon > 0$ and $\boldsymbol{\Delta} \in \Re^{p\times d}$*

$$\lim_{n\to\infty} P(\|L_{\hat{A}_n}(\boldsymbol{\Delta}) - (L_{\hat{A}_n}(\mathbf{0}) - \boldsymbol{\Delta}\mathbf{T}^{-1})\| \geq \epsilon) = 0.$$

Proof. Fix j, $j = 1, \ldots, d$. Then by Theorem A.3.1 of Hettmansperger and McKean (1998) we have for all $\epsilon > 0$ and for all $\boldsymbol{\Delta}$

$$\lim_{n\to\infty} P(\|L^{(j)}(\boldsymbol{\Delta}) - (L^{(j)}(\mathbf{0}) - \tau_j^{-1}\boldsymbol{\Delta}\boldsymbol{\delta}_j)\| \geq \epsilon) = 0. \tag{2.25}$$

Now, we have

$$\begin{aligned} P(\|L^{(j)}_{\hat{A}_n}(\boldsymbol{\Delta}) - (L^{(j)}_{\hat{A}_n}(\mathbf{0}) - \tau_j^{-1}\boldsymbol{\Delta}\boldsymbol{\delta}_j)\| \geq \epsilon) \leq\ & P(\|L^{(j)}_{\hat{A}_n}(\boldsymbol{\Delta}) - L^{(j)}(\boldsymbol{\Delta})\| \geq \epsilon/3) \\ & + P(\|L^{(j)}_{\hat{A}_n}(\mathbf{0}) - L^{(j)}(\mathbf{0})\| \geq \epsilon/3) \\ & + P(\|L^{(j)}(\boldsymbol{\Delta}) - (L^{(j)}(\mathbf{0}) - \tau_j^{-1}\boldsymbol{\Delta}\boldsymbol{\delta}_j)\| \geq \epsilon/3). \end{aligned}$$

The result now follows from (2.25), corollary 2.4.1 and corollary 2.4.2. □

Corollary 2.4.3. *For all $\epsilon > 0$ and $\boldsymbol{\Delta} \in \Re^{p\times d}$*

$$\lim_{n\to\infty} P(\|L_{\hat{A}_n}(\boldsymbol{\Delta}) - (L_{\hat{A}_n}(\mathbf{0})\hat{A}_n - \boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1}\hat{A}_n)\| \geq \epsilon) = 0.$$

Proof. Given any $\epsilon > 0$ and $\boldsymbol{\Delta} \in \Re^{p\times d}$ we have

$$\begin{aligned} & P(\|L_{\hat{A}_n}(\boldsymbol{\Delta}) - (L_{\hat{A}_n}(\mathbf{0})\hat{A}_n - \boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1}\hat{A}_n)\| \geq \epsilon) \\ & \leq P(\|L_{\hat{A}_n}(\boldsymbol{\Delta}) - (L_{\hat{A}_n}(\mathbf{0}) - \boldsymbol{\Delta}\mathbf{T}^{-1})\| \geq \epsilon/3) + P(\|L_{\hat{A}_n}(\mathbf{0})(\hat{A}_n - I_d)\| \geq \epsilon/3) \\ & \quad + P(\|\boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1}\hat{A}_n - \boldsymbol{\Delta}\mathbf{T}^{-1}\| \geq \epsilon/3). \end{aligned}$$

By Theroem 2.4.2 the first term on the right-hand side goes to 0 in probability. Also by Theorem 2.4.1, $L_{\hat{A}_n}(\mathbf{0})$ converges in distribution and hence is bounded in probability. This implies that the second term also goes to 0 in probabilty since $\hat{A}_n \xrightarrow{p} I_d$. The last term goes to 0 in probability again by the consistency of $\hat{A}_n$. □

An approximation of $D_{\hat{A}_n}(\boldsymbol{\Delta})$ is the quadratic function

$$\begin{aligned}
Q_{\hat{A}_n}(\boldsymbol{\Delta}) &= \operatorname{tr}\left(\frac{1}{2}\hat{A}_n\boldsymbol{\Delta}'\boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1} - \hat{A}_n\boldsymbol{\Delta}'L_{\hat{A}_n}(\mathbf{0})\right) + D_{\hat{A}_n}(\mathbf{0}) \\
&= \sum_{j=1}^{d}\left(\frac{1}{2}\tau_j^{-1}\hat{\boldsymbol{a}}_{nj}'\boldsymbol{\Delta}'\boldsymbol{\Delta}\hat{\boldsymbol{a}}_{nj} - \hat{\boldsymbol{a}}_{nj}'\boldsymbol{\Delta}'L_{\hat{A}_n}^{(j)}(\mathbf{0}) + D_{\hat{A}_n,j}(\mathbf{0})\right) \\
&= \sum_{j=1}^{d} Q_{\hat{A}_n,j}(\boldsymbol{\Delta}).
\end{aligned}$$

The gradient of $Q_{\hat{A}_n}(\boldsymbol{\Delta})$ is given by

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\Delta}}Q_{\hat{A}_n}(\boldsymbol{\Delta}) &= \sum_{j=1}^{d}\frac{\partial}{\partial\boldsymbol{\Delta}}Q_{\hat{A}_n,j}(\boldsymbol{\Delta}) \\
&= \sum_{j=1}^{d}\left(\tau_j^{-1}\boldsymbol{\Delta}\hat{\boldsymbol{a}}_{nj}\hat{\boldsymbol{a}}_{nj}' - L_{\hat{A}_n}^{(j)}(\mathbf{0})\hat{\boldsymbol{a}}_{nj}'\right) \\
&= \sum_{j=1}^{d}\left(\tau_j^{-1}\boldsymbol{\Delta}\hat{\boldsymbol{a}}_{nj} - L_{\hat{A}_n}^{(j)}(\mathbf{0})\right)\hat{\boldsymbol{a}}_{nj}' \\
&= \begin{pmatrix}\tau_1^{-1}\boldsymbol{\Delta}\hat{\boldsymbol{a}}_{n1} - L_{\hat{A}_n}^{(1)}(\mathbf{0}) & \cdots & \tau_d^{-1}\boldsymbol{\Delta}\hat{\boldsymbol{a}}_{nd} - L_{\hat{A}_n}^{(d)}(\mathbf{0})\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{a}}_{n1}' \\ \vdots \\ \hat{\boldsymbol{a}}_{nd}'\end{pmatrix} \\
&= \left\{\begin{pmatrix}\boldsymbol{\Delta}\hat{\boldsymbol{a}}_{n1} & \cdots & \boldsymbol{\Delta}\hat{\boldsymbol{a}}_{nd}\end{pmatrix}\begin{pmatrix}\tau_1^{-1} & & \\ & \ddots & \\ & & \tau_d^{-1}\end{pmatrix} - L_{\hat{A}_n}(\mathbf{0})\right\}\hat{A}_n \\
&= \left\{\boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1} - L_{\hat{A}_n}(\mathbf{0})\right\}\hat{A}_n.
\end{aligned}$$

Theorem 2.4.3. *The following expressions are equivalent*

(i) $\lim_{n\to\infty} P(\|L_{\hat{A}_n}(\boldsymbol{\Delta}) - (L_{\hat{A}_n}(\mathbf{0})\hat{A}_n - \boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1}\hat{A}_n)\| \geq \epsilon) = 0$

(ii) *Asymptotic linearity*

$$\lim_{n\to\infty} P(\sup_{\|\boldsymbol{\Delta}\|\leq c} \|L_{\hat{A}_n}(\boldsymbol{\Delta}) - (L_{\hat{A}_n}(\mathbf{0})\hat{A}_n - \boldsymbol{\Delta}\hat{A}_n'\mathbf{T}^{-1}\hat{A}_n)\| \geq \epsilon) = 0 \tag{2.26}$$

(iii) *Asymptotic quadraticity*

$$\lim_{n\to\infty} P(\sup_{\|\boldsymbol{\Delta}\|\leq c} |D_{\hat{A}_n}(\boldsymbol{\Delta}) - Q_{\hat{A}_n}(\boldsymbol{\Delta})| \geq \epsilon) = 0 \tag{2.27}$$

for arbitrary $c > 0$ and $\epsilon > 0$.

Proof. The proof is based on the convexity of $D_{\hat{A}_n}(\boldsymbol{\Delta})$ and $Q_{\hat{A}_n}(\boldsymbol{\Delta})$ and the fact that $\hat{A}_n \to I_d$ almost surely. The details of the proof are similar to the univariate case, see Hettmansperger and McKean (1998) page 414. □

Because $Q_{\hat{A}_n}$ is a quadratic function it follows from differentiation that it is minimized by $\boldsymbol{\Delta}^*$ such that

$$\nabla Q_{\hat{A}_n}(\boldsymbol{\Delta}^*) = \mathbf{0},$$

or

$$\left\{\boldsymbol{\Delta}^*\hat{A}_n'\mathbf{T}^{-1} - L_{\hat{A}_n}(\mathbf{0})\right\}\hat{A}_n = \mathbf{0}. \tag{2.28}$$

That is

$$\boldsymbol{\Delta}^* = L_{\hat{A}_n}(\mathbf{0})\mathbf{T}(\hat{A}_n')^{-1}. \tag{2.29}$$

Theorem 2.4.4.

$$\boldsymbol{\Delta}^* \xrightarrow{D} N_{p,d}(\mathbf{0}, I_p, \mathbf{T}\boldsymbol{S}\mathbf{T}).$$

Proof. By 2.29 and Theorem 2.4.1, the column covariance of $\mathbf{\Delta}^*$ is $(\hat{A}_n)^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(\hat{A}_n')^{-1}$ which converges to $\mathbf{T}\boldsymbol{S}\mathbf{T}$ as $n \to \infty$ because $(\hat{A}_n)^{-1} \xrightarrow{p} I_d$. □

Since $Q_{\hat{A}_n}$ is a local approximation to $D_{\hat{A}_n}$, it would seem that their minimizing values are close. The following theorem proves this claim.

Theorem 2.4.5.

$$\widehat{\mathbf{\Delta}} - \mathbf{\Delta}^* \xrightarrow{p} 0$$

Proof. Let $\epsilon > 0$ and $\delta > 0$ be given. Since $\mathbf{\Delta}^*$ converges in distribution, it is bounded in probability. Thus there exists a $c_0 > 0$ such that

$$P[\|\mathbf{\Delta}^*\| \geq c_0] < \delta/2, \tag{2.30}$$

for sufficiently large n. Let

$$T = \min\{Q_{\hat{A}_n}(\mathbf{\Delta}) : \|\mathbf{\Delta} - \mathbf{\Delta}^*\| = \epsilon\} - Q_{\hat{A}_n}(\mathbf{\Delta}^*) \tag{2.31}$$

Note that $T > 0$ because $\mathbf{\Delta}^*$ is the unique minimizer of $Q_{\hat{A}_n}$. By the asymptotic quadraticity we have

$$P\left[\max_{\|\mathbf{\Delta}\| \leq c_0 + \epsilon} |D_{\hat{A}_n}(\mathbf{\Delta}) - Q_{\hat{A}_n}(\mathbf{\Delta})| \geq T/2\right] \leq \delta/2 \tag{2.32}$$

for sufficiently large n. Let $\mathbf{\Delta}$ be arbitrary and on the ring $\|\mathbf{\Delta} - \mathbf{\Delta}^*\| = \epsilon$ where $\|\mathbf{\Delta}^*\| < c_0$. Then $\|\mathbf{\Delta}\| < c_0 + \epsilon$. Combining (2.30) and (2.32) we can assert with probability greater than $1 - \delta$ that for sufficiently large n, we have

$$|D_{\hat{A}_n}(\mathbf{\Delta}) - Q_{\hat{A}_n}(\mathbf{\Delta})| < T/2 \quad \text{and} \quad \|\mathbf{\Delta}^*\| < c_0. \tag{2.33}$$

This implies with probability greater than $1 - \delta$ that for sufficiently large n,

$$D_{\hat{A}_n}(\mathbf{\Delta}) < Q_{\hat{A}_n}(\mathbf{\Delta}) + T/2 \quad \text{and} \quad \|\mathbf{\Delta}^*\| < c_0 \tag{2.34}$$

Also, we can conclude with probability greater than $1-\delta$ that for sufficiently large n we have

$$D_{\hat{A}_n}(\mathbf{\Delta}) > Q_{\hat{A}_n}(\mathbf{\Delta}) - T/2 \quad \text{and} \quad \|\mathbf{\Delta}^*\| < c_0 \tag{2.35}$$

Now it follows from (2.34) and (2.35) that

$$\begin{aligned} D_{\hat{A}_n}(\mathbf{\Delta}) &> Q_{\hat{A}_n}(\mathbf{\Delta}) - T/2 \\ &> \min\{Q_{\hat{A}_n}(\mathbf{\Delta}) : \|\mathbf{\Delta} - \mathbf{\Delta}^*\| = \epsilon\} - T/2 \\ &= T/2 + Q_{\hat{A}_n}(\mathbf{\Delta}^*) \\ &> D_{\hat{A}_n}(\mathbf{\Delta}^*) \end{aligned}$$

Thus we have $D_{\hat{A}_n}(\mathbf{\Delta}) > D_{\hat{A}_n}(\mathbf{\Delta}^*)$ for $\|\mathbf{\Delta}-\mathbf{\Delta}^*\| = \epsilon$. But since $D_{\hat{A}_n}$ is convex we must also have $D_{\hat{A}_n}(\mathbf{\Delta}) > D_{\hat{A}_n}(\mathbf{\Delta}^*)$ for $\|\mathbf{\Delta} - \mathbf{\Delta}^*\| > \epsilon$. Since $D_{\hat{A}_n}(\mathbf{\Delta}^*) \geq \min D_{\hat{A}_n}(\mathbf{\Delta}) = D_{\hat{A}_n}(\widehat{\mathbf{\Delta}})$ we nust have $\widehat{\mathbf{\Delta}}$ lie inside the disk $\|\mathbf{\Delta} - \mathbf{\Delta}^*\| = \epsilon$ with probability greater than $1-2\delta$. That is $P[\|\widehat{\mathbf{\Delta}} - \mathbf{\Delta}^*\| \leq \epsilon] \geq 1 - 2\delta$. □

Let $\widehat{\mathbf{\Delta}}_\varphi = \widehat{\mathbf{\Delta}}\hat{A}_n'$ be the estimate of the slope matrix in (2.2) under the transformed model. The following theorem gives the asymptotic distribution of $\widehat{\mathbf{\Delta}}_\varphi$.

Theorem 2.4.6.

$$\widehat{\mathbf{\Delta}}_\varphi \xrightarrow{D} N_{p,d}(\mathbf{0}, I_p, \mathbf{T}S\mathbf{T}).$$

Proof. Note that for all $\epsilon > 0$ we have

$$\begin{aligned} P(\|\mathbf{\Delta}^*\hat{A}_n' - \widehat{\mathbf{\Delta}}_\varphi\| > \epsilon) &= P(\|\mathbf{\Delta}^*\hat{A}_n' - \widehat{\mathbf{\Delta}}\hat{A}_n'\| > \epsilon) \\ &= P(\|\mathbf{\Delta}^*(\hat{A}_n' - I_d)\| > \epsilon/3) + P(\|\mathbf{\Delta}^* - \widehat{\mathbf{\Delta}}\| > \epsilon/3) \\ &+ P(\|\widehat{\mathbf{\Delta}}(I_d - \hat{A}_n')\| > \epsilon/3). \end{aligned}$$

Since $\boldsymbol{\Delta}^*$ is bounded in probability, so is $\widehat{\boldsymbol{\Delta}}$ by Theorem 2.4.5. Further $\hat{A}_n' \xrightarrow{p} I_d$. Consequently, we have $\boldsymbol{\Delta}^*\hat{A}_n' - \widehat{\boldsymbol{\Delta}}_\varphi \xrightarrow{p} 0$. Also, by (2.28) we have $\boldsymbol{\Delta}^*\hat{A}_n' = L_{\hat{A}_n}(\mathbf{0})\mathbf{T}$. Now the result follows immediately from Theorem 2.4.1. □

To solve for the intercept vector of the transformed model we minimize for $j = 1, \ldots, d$ the function

$$L_1^{(j)}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} |\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}|. \tag{2.36}$$

On taking the partial derivative with respect to $\boldsymbol{\alpha}$ we get

$$\frac{\partial}{\partial\boldsymbol{\alpha}} L_1^{(j)}(\boldsymbol{\alpha}) = -\left(\sum_{i=1}^{n} \operatorname{sgn}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})\right)\hat{\boldsymbol{a}}_{nj}.$$

Solving the equation $\frac{\partial}{\partial\boldsymbol{\alpha}} L_1^{(j)}(\boldsymbol{\alpha}) = \mathbf{0}$ we get

$$\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj} = \operatorname*{med}_{1\le i\le n}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}).$$

Consequently,

$$\widehat{\boldsymbol{\alpha}}'\hat{A}_n' = \begin{pmatrix} \operatorname*{med}_{1\le i\le n}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{n1} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{n1}) & \cdots & \operatorname*{med}_{1\le i\le n}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nd} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nd}) \end{pmatrix}.$$

Lemma 2.4.3. *Let* $\boldsymbol{\Phi}^*_{\hat{A}_n} = (\operatorname{sgn}(\boldsymbol{\varepsilon}_i'\hat{\boldsymbol{a}}_{nj}))$. *Then*

$$n^{-1/2}\mathbf{1}_n'(\boldsymbol{\Phi}^*_{\hat{A}_n} - \boldsymbol{\Phi}^*) \xrightarrow{p} 0. \tag{2.37}$$

Proof. Fix j, $j = 1, \ldots, d$. Let

$$l = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \operatorname{sgn}(\boldsymbol{\varepsilon}_i'\hat{\boldsymbol{a}}_{nj}) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \operatorname{sgn}(\varepsilon_{ij}).$$

Then

$$\begin{aligned} E[l^2] &= \frac{1}{n} E\left[\left(\sum_{i=1}^{n}(\operatorname{sgn}(\boldsymbol{\varepsilon}_i'\hat{\boldsymbol{a}}_{nj}) - \operatorname{sgn}(\varepsilon_{ij}))\right)^2\right] \\ &= E\left[(\operatorname{sgn}(\boldsymbol{\varepsilon}_1'\hat{\boldsymbol{a}}_{nj}) - \operatorname{sgn}(\boldsymbol{\varepsilon}_1'\boldsymbol{\delta}_j))^2\right] \end{aligned}$$

Because the sign function is continuous almost everywhere and $\hat{\boldsymbol{a}}_{nj} \xrightarrow{p} \boldsymbol{\delta}_j$ we get

$$(\operatorname{sgn}(\boldsymbol{\varepsilon}_1'\hat{\boldsymbol{a}}_{nj}) - \operatorname{sgn}(\boldsymbol{\varepsilon}_1'\boldsymbol{\delta}_j))^2 \xrightarrow{p} 0, \quad \text{and} \quad n \to \infty,$$

also we have ,

$$|\operatorname{sgn}(\boldsymbol{\varepsilon}_1'\hat{\boldsymbol{a}}_{nj}) - \operatorname{sgn}(\boldsymbol{\varepsilon}_1'\boldsymbol{\delta}_j)| \leq 2.$$

Thus by the Lebesgue Dominated Convergence Theorem we conclude that $E[l^2] \to 0$ as $n \to \infty$. Therefore, $l \xrightarrow{p} 0$. □

For the next results we need to introduce the following notation. Fix $j \in \{1, \ldots, d\}$. Let

$$S_1(\mathbf{Y}^{(j)} - \alpha_j \mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}^{(j)}) = \sum_{i=1}^{n} \operatorname{sgn}(y_{ij} - \alpha_j - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}^{(j)})$$

and

$$S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) = \sum_{i=1}^{n} \operatorname{sgn}(\boldsymbol{y}_i'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj} - \boldsymbol{c}_i'\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})$$

Lemma 2.4.4. *For any $\epsilon > 0$ and any $\boldsymbol{\alpha} \in \Re^d$,*

$$\lim_{n\to\infty} P[n^{1/2}|S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) - S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj})\mathbf{1}_n| \geq \epsilon] = 0 \quad (2.38)$$

Proof. Given $\epsilon > 0$ we have

$$P\left[n^{1/2}|S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) - S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj})\mathbf{1}_n| \geq \epsilon\right]$$
$$\leq P\Big[|n^{1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})$$
$$-n^{1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j - n^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\delta}_j\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j)| \geq \epsilon/3\Big]$$
$$+P\left[n^{1/2}|S_1(\mathbf{Y}\boldsymbol{\delta}_j - n^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\delta}_j\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j) - S_1(\mathbf{Y}\boldsymbol{\delta}_j - n^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\delta}_j\mathbf{1}_n)| \geq \epsilon/3\right]$$
$$+P\left[|n^{1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j - n^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\delta}_j\mathbf{1}_n) - n^{1/2}S_1(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n)| \geq \epsilon/3\right].$$

By Lemma 3.5.8 of Hettmansperger and McKean (1998) the second term on the right-hand side goes to 0. A similar argument to the proof of Lemma 2.4.3 shows that the first and last terms go to 0. □

Theorem 2.4.7. *For any $\epsilon > 0$ and $c > 0$,*

$$\lim_{n\to\infty} P[\sup_{\|\boldsymbol{\alpha}\|\leq c} |n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})$$
$$- n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj} + \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\tau_j^{*-1}| \geq \epsilon] = 0.$$

Proof. For any $\boldsymbol{\alpha}$ we have

$$|n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})$$
$$- n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) + \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\tau_j^{*-1}|$$
$$\leq |n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})$$
$$- n^{-1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j - n^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\delta}_j\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j)|$$
$$+ |n^{-1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j - n^{-1/2}\boldsymbol{\alpha}'\boldsymbol{\delta}_j\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j) - n^{-1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j) + \boldsymbol{\alpha}'\boldsymbol{\delta}_j\tau_j^{*-1}|$$
$$+ |n^{-1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j) - n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j)| + |\boldsymbol{\alpha}'(\hat{\boldsymbol{a}}_{nj} - \boldsymbol{\delta}_j)\tau_j^{*-1}|.$$

We can apply Theorem 3.5.9 of Hettmansperger and McKean (1998) to the second term on the right-hand side. An argument similar to the proof of Lemma 2.4.3 can be applied to the first and third terms. The consistency of $\hat{A}_n$ can be used for the last term. □

Let $\boldsymbol{\alpha} = \mathbf{0}$ in Lemma 2.38, and using the result of Lemma 2.4.3 we have $n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j) - n^{-1/2}S_1(\mathbf{Y}\boldsymbol{\delta}_j)$ goes to 0 in probaility. Thus, the sequence $n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j)$ converges in distribution to $N(0,1)$.

Lemma 2.4.5. *The random variable $n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj}$ is bounded in probability.*

Proof. Since $n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j)$ converges in distribution, it is bounded in probability. Thus given $\epsilon > 0$, there exists a $c < 0$ such that

$$P[n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\boldsymbol{\delta}_j) < c] < \epsilon/2. \tag{2.39}$$

Take $c^* = \tau_j^*(c-\epsilon)$. Since $\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj}$ is the solution of $S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj})$ and $S_{1,\hat{A}_n}$ is montone nonincreasing we have $n^{1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj} < c^*$ implies $n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}c^*\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) \leq 0$. Using the above linearaity result we have

$$\begin{aligned}
P[n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj} < c^*] &\leq P[n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}c^*\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) \leq 0] \\
&\leq P[\,|n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}c^*\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) \\
&\qquad - n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) + c^*\tau_j^{*-1}| \geq \epsilon] \\
&\quad + P[n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - -\mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) - c^*\tau_j^{*-1} < \epsilon].
\end{aligned}$$

By Theorem 2.4.7 the first term on the right-hand side can be made less than $\epsilon/2$ for large n. The second term is less than $\epsilon/2$ by (2.39). This proves that $n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj}$ is bounded in probabilty from below. A similar argument shows that $n^{1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}$ is bounded above in probability. □

Lemma 2.4.6. *For any* $\epsilon > 0$

$$\lim_{n\to\infty} P[\|n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{A}_n'\mathbf{T}^{*-1} - n^{-1/2}\mathbf{1}_n'\boldsymbol{\Phi}^*_{\hat{A}_n}\| \geq \epsilon] = 0.$$

Proof. The asymptotic linearity result of Theorem 2.4.7 can be written as

$$n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) = n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj}) - \boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\tau_j^{*-1} + o_p(1),$$

uniformly for all $\|\boldsymbol{\alpha}\| \leq c$ and for $c > 0$. since $\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj}$ is a solution to $S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj} - n^{-1/2}\boldsymbol{\alpha}'\hat{\boldsymbol{a}}_{nj}\mathbf{1}_n - \mathbf{C}\widehat{\boldsymbol{\Delta}}\hat{\boldsymbol{a}}_{nj}) = 0$ and $n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj}$ is bounded in probability we get

$$n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{\boldsymbol{a}}_{nj} = \tau_j^* n^{-1/2}S_{1,\hat{A}_n}(\mathbf{Y}\hat{\boldsymbol{a}}_{nj}) + o_p(1).$$

Thus, we have

$$n^{1/2}\widehat{\boldsymbol{\alpha}}' = n^{-1/2}\mathbf{1}_n'\boldsymbol{\Phi}^*_{\hat{A}_n}\mathbf{T}^* + o_p(1).$$

□

Lemma 2.4.7. *For any* $\epsilon > 0$

$$\lim_{n\to\infty} P(\|n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{A}_n'\mathbf{T}^{*-1} - n^{-1/2}\mathbf{1}_n'\boldsymbol{\Phi}^*\| \geq \epsilon) = 0. \tag{2.40}$$

Proof. Let $\epsilon > 0$ be given. Then

$$P(\|n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{A}_n'\mathbf{T}^{*-1} - n^{-1/2}\mathbf{1}_n'\boldsymbol{\Phi}^*\| \geq \epsilon) \leq P(\|n^{1/2}\widehat{\boldsymbol{\alpha}}'\hat{A}_n'\mathbf{T}^{*-1} - n^{-1/2}\mathbf{1}_n'\boldsymbol{\Phi}^*_{\hat{A}_n}\| \geq \epsilon/2)$$
$$+ P(\|n^{-1/2}\mathbf{1}_n'(\boldsymbol{\Phi}^*_{\hat{A}_n} - \boldsymbol{\Phi}^*)\| \geq \epsilon/2).$$

The first term goes to 0 by lemma 2.4.6. The second term goes to 0 by 2.37. □

Let $\widehat{\boldsymbol{\alpha}}_s = \hat{A}_n\widehat{\boldsymbol{\alpha}}$. Then by lemma 2.4.7 and theorem 18 of Lehmann (1975) page 390, we have

Corollary 2.4.4.

$$n^{1/2}\widehat{\boldsymbol{\alpha}}_s \xrightarrow{D} N_d(\mathbf{0}, \mathbf{T}^* \boldsymbol{S}^* \mathbf{T}^*) \tag{2.41}$$

Now recall that

$$\begin{aligned} \widehat{\boldsymbol{\mathcal{B}}}_R &= \begin{pmatrix} \widehat{\beta}'_{0,s} \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,\varphi} \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\frac{1}{n}\mathbf{X}(\mathbf{X}_c'\mathbf{X}_c)^{-1/2} \\ \mathbf{0}_p & (\mathbf{X}_c'\mathbf{X}_c)^{-1/2} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\alpha}}_s' \\ \widehat{\boldsymbol{\Delta}}_{\varphi} \end{pmatrix}. \end{aligned}$$

The following theorem due to Davis and McKean (1993) gives the asymptotic distribution of $\widehat{\boldsymbol{\mathcal{B}}}_R$.

Theorem 2.4.8. *Under assumptions A1-A7*

$$\sqrt{n}\,\mathrm{vec}[\widehat{\beta}_{0,s}, \widehat{\boldsymbol{\mathcal{B}}}'_{1,\varphi}] \xrightarrow{\mathcal{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix} \mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + \mathbf{v}'\mathbf{V}^{-1}\mathbf{v} & \mathbf{T}\boldsymbol{S}\mathbf{T} \otimes \mathbf{v}'\mathbf{V}^{-1} \\ \mathbf{T}\boldsymbol{S}\mathbf{T} \otimes \mathbf{V}^{-1}\mathbf{v} & \mathbf{T}\boldsymbol{S}\mathbf{T} \otimes \mathbf{V}^{-1} \end{pmatrix}\right). \tag{2.42}$$

Since the transformed-retransformed estimate $\widehat{\boldsymbol{\mathcal{B}}}_{TRR} = \widehat{\boldsymbol{\mathcal{B}}}_R(\hat{A}^{-1})'$ we have

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

Let

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

It follows from Theorem 2.4.8 that

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

As $\hat{A}_n \xrightarrow{P} I_d$ we obtain the asymptotic distribution of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ under the assumptions $\boldsymbol{\Sigma} = I_d$ and $\boldsymbol{\mathcal{B}} = \mathbf{0}$ as

$$\sqrt{n}\,\mathrm{vec}\widehat{\boldsymbol{\mathcal{B}}}'_{TRR} \xrightarrow{D} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix} \mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + \mathbf{v}'\mathbf{V}^{-1}\mathbf{v} & \mathbf{T}\boldsymbol{S}\mathbf{T} \otimes \mathbf{v}'\mathbf{V}^{-1} \\ \mathbf{T}\boldsymbol{S}\mathbf{T} \otimes \mathbf{V}^{-1}\mathbf{v} & \mathbf{T}\boldsymbol{S}\mathbf{T} \otimes \mathbf{V}^{-1} \end{pmatrix}\right). \tag{2.43}$$

Theorem 2.4.9. *Under assumptions A1-A9*

$$\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \boldsymbol{\mathcal{B}})' \xrightarrow{D} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{S}_{11,TRR} & \boldsymbol{S}_{12,TRR} \\ \boldsymbol{S}_{21,TRR} & \boldsymbol{S}_{22,TRR} \end{pmatrix}\right) \tag{2.44}$$

where

$$\mathcal{S}_{11,TRR} = A_0^{-1}(\mathbf{T}^*\mathcal{S}^*\mathbf{T}^* + (\mathbf{v}'\mathbf{V}^{-1}\mathbf{v})\mathbf{T}\mathcal{S}\mathbf{T})(A_0^{-1})'$$

$$\mathcal{S}_{12,TRR} = A_0^{-1}\mathbf{T}\mathcal{S}\mathbf{T}(A_0^{-1})' \otimes \mathbf{v}'\mathbf{V}^{-1}$$

$$\mathcal{S}_{21,TRR} = A_0^{-1}\mathbf{T}\mathcal{S}\mathbf{T}(A_0^{-1})' \otimes \mathbf{V}^{-1}\mathbf{v}$$

$$\mathcal{S}_{22,TRR} = A_0^{-1}\mathbf{T}\mathcal{S}\mathbf{T}(A_0^{-1})' \otimes \mathbf{V}^{-1}$$

Corollary 2.4.5. *The influence function of $\widehat{\mathcal{B}}_{1,TRR}$ is given by*

$$\boldsymbol{\Omega}(\boldsymbol{x}_0, \boldsymbol{y}_0, \widehat{\mathcal{B}}_{1,TRR}) = \mathbf{V}^{-1}\boldsymbol{x}_0 \begin{pmatrix} \varphi(F_1(A\boldsymbol{y}_0)_1) & \cdots & \varphi(F_d(A\boldsymbol{y}_0)_d) \end{pmatrix} \mathbf{T}\,(A^{-1})' \tag{2.45}$$

2.5 A Quadratic Test

Let $\mathbf{H}$ be an $r \times (p+1)$ constant matrix of rank r, and let $\mathbf{K}$ be a $d \times s$ constant matrix of rank s. In this section we consider a test of the general hypothesis $H_0 : \mathbf{H}\mathcal{B}\mathbf{K} = \mathbf{0}$. For this end we need the following lemma, see Muirhead (1982).

Lemma 2.5.1. *If $\mathbf{B}$ is $r \times m$, $\mathbf{X}$ is $m \times n$, and $\mathbf{C}$ is $n \times s$ then*

$$\mathrm{vec}(\mathbf{BXC}) = (\mathbf{B} \otimes \mathbf{C}')\,\mathrm{vec}(\mathbf{X}).$$

Theorem 2.5.1. *Under assumptions A1–A9 and $H_0 : \mathbf{H}\mathcal{B}\mathbf{K} = \mathbf{0}$,*

$$n\,\mathcal{L}_{TRR}'[(\mathbf{K}'A^{-1}\mathbf{T}^*\mathcal{S}^*\mathbf{T}^*(A^{-1})'\mathbf{K}\otimes\mathbf{H}_1\mathbf{H}_1') + (\mathbf{K}'A^{-1}\mathbf{T}\mathcal{S}\mathbf{T}(A^{-1})'\mathbf{K}\otimes N_{TRR})]^{-1}\mathcal{L}_{TRR} \xrightarrow{\mathcal{D}} \chi^2(rs), \tag{2.46}$$

where

$$\mathcal{L}_{TRR} = \mathrm{vec}(\mathbf{K}'\widehat{\mathcal{B}}_{TRR}\mathbf{H}'),$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1^{r\times 1} & \mathbf{H}_2^{r\times p} \end{pmatrix}$$

and

$$N_{TRR} = n(-\mathbf{H}_1\bar{\boldsymbol{x}}' + \mathbf{H}_2)(\mathbf{X}_c'\mathbf{X}_c)^{-1}(-\mathbf{H}_1\bar{\boldsymbol{x}}' + \mathbf{H}_2)'.$$

Proof. Based on lemma 2.5.1 we have

$$\begin{aligned}\boldsymbol{\mathcal{L}}_{TRR} &= \mathrm{vec}(\mathbf{K}'\widehat{\boldsymbol{\mathcal{B}}}'_{TRR}\mathbf{H}') \\ &= (\mathbf{K}' \otimes \mathbf{H})\,\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}'_{TRR}).\end{aligned}$$

It follows now from theorem 2.4.9 that

$$\sqrt{n}\,\boldsymbol{\mathcal{L}}_{TRR} \xrightarrow{\mathcal{D}} N_{rs}\left(\mathbf{0}, (\mathbf{K}' \otimes \mathbf{H})\begin{pmatrix} \boldsymbol{S}_{11,TRR} & \boldsymbol{S}_{12,TRR} \\ \boldsymbol{S}_{21,TRR} & \boldsymbol{S}_{22,TRR}\end{pmatrix}(\mathbf{K} \otimes \mathbf{H}')\right).$$

Further simplifications give the stated result. □

Let

$$V_{TRR} = (\mathbf{K}'A^{-1}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*(A^{-1})'\mathbf{K} \otimes \mathbf{H}_1\mathbf{H}_1') + (\mathbf{K}'A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})'\mathbf{K} \otimes N)$$

and

$$Q_{TRR} = n\,\boldsymbol{\mathcal{L}}'_{TRR}V_{TRR}^{-1}\boldsymbol{\mathcal{L}}_{TRR}.$$

Then we reject $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$ in favor of $H_A : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} \neq \mathbf{0}$ at asymptotic level α if $Q_{TRR} \geq \chi^2_\alpha(rs)$. Let

$$\boldsymbol{\mathcal{L}}_R = \mathrm{vec}(\mathbf{K}'\widehat{\boldsymbol{\mathcal{B}}}_R\mathbf{H}')$$

and

$$V_R = (\mathbf{K}'\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*\mathbf{K} \otimes \mathbf{H}_1\mathbf{H}_1') + (\mathbf{K}'\mathbf{T}\boldsymbol{S}\mathbf{T}\mathbf{K} \otimes N).$$

Then the quadratic test of $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$ based on $\widehat{\boldsymbol{\mathcal{B}}}_R$ was first obtained by Davis and McKean (19993) and it rejects $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$ at asymptotic level α if

$$Q_R = n\,\boldsymbol{\mathcal{L}}'_R V_R^{-1}\boldsymbol{\mathcal{L}}_R \geq \chi^2_\alpha(rs).$$

CHAPTER III

A MULTIVARIATE GR ESTIMATE

3.1 Componentwise GR Estimation

Recall the multivariate linear model given by (2.1),

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

Model (2.1) can be written as

$$\mathbf{Y} = \mathbf{1}_n \boldsymbol{\alpha}_0' + \mathbf{X}_c\boldsymbol{\mathcal{B}}_1 + \boldsymbol{\mathcal{E}}, \tag{3.1}$$

where $\boldsymbol{\alpha}_0' = \boldsymbol{\beta}_0' + \bar{\boldsymbol{x}}'\boldsymbol{\mathcal{B}}_1$. Model (3.1) is the concatenation of the d univariate linear models

$$\mathbf{Y}^{(k)} = \mathbf{1}_n \boldsymbol{\alpha}_0^{(k)} + \mathbf{X}_c\boldsymbol{\mathcal{B}}_1^{(k)} + \boldsymbol{\mathcal{E}}^{(k)}, \quad k = 1, \ldots, d. \tag{3.2}$$

Consider the function

$$\|\boldsymbol{u}\|_{GR} = \sum_{i<j} b_{ij}|u_i - u_j|, \tag{3.3}$$

where the b_{ij} are some functions of the x's and assumed to be positive and symmetric i. e. $b_{ij} \equiv b_{ji}$. Note that the function defined by (3.3) is the Wilcoxon pseudonorm if the weights $b_{ij} \equiv 1$. The componentwise GR-estimate of $\boldsymbol{\mathcal{B}}_1$ is a matrix

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,GR} = \begin{pmatrix} \widehat{\boldsymbol{\mathcal{B}}}_{1,GR}^{(1)} & \cdots & \widehat{\boldsymbol{\mathcal{B}}}_{1,GR}^{(d)} \end{pmatrix},$$

where $\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}^{(k)}$ minimizes

$$D_{GR}(\boldsymbol{\mathcal{B}}_1^{(k)}) = \|\mathbf{Y}^{(k)} - \mathbf{X}_c\boldsymbol{\mathcal{B}}_1^{(k)}\|_{GR}$$

which is a continuous, nonnegative and convex function of $\boldsymbol{\mathcal{B}}_1^{(k)}$. For $k = 1, \ldots, d$ the negative of the gradient of $D_{GR}(\boldsymbol{\mathcal{B}}_1^{(k)})$ is given by

$$\begin{aligned} \boldsymbol{S}_{GR}^{(k)}(\boldsymbol{\mathcal{B}}_1^{(k)}) &= -\frac{\partial}{\partial \boldsymbol{\mathcal{B}}_1^{(k)}} D_{GR}(\boldsymbol{\mathcal{B}}_1^{(k)}) \\ &= \sum_{i<j} b_{ij}(\boldsymbol{x}_i - \boldsymbol{x}_j)\operatorname{sgn}((\mathbf{Y}_i^{(k)} - \mathbf{Y}_j^{(k)}) - (\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{\mathcal{B}}_1^{(k)}). \end{aligned}$$

Define the statistic

$$\boldsymbol{S}_{GR}(\boldsymbol{\mathcal{B}}_1) = \begin{pmatrix} \boldsymbol{S}_{GR}^{(1)}(\boldsymbol{\mathcal{B}}_1^{(1)}) & \cdots & \boldsymbol{S}_{GR}^{(d)}(\boldsymbol{\mathcal{B}}_1^{(d)}) \end{pmatrix}.$$

Then the componentwise GR-estimate of $\boldsymbol{\mathcal{B}}_1$ solves the estimating equations

$$\boldsymbol{S}_{GR}(\boldsymbol{\mathcal{B}}_1) \doteq \mathbf{0}$$

by solving for $k = 1, \ldots, d$ the equations $\boldsymbol{S}_{GR}^{(k)}(\boldsymbol{\mathcal{B}}_1^{(k)}) \doteq \mathbf{0}$. The univariate GR-estimates were proposed by Sievers (1983) and further developed by Naranjo and Hettmansperger (1994).

3.1.1 Asymptotic Theory

The asymptotic theory of the GR-estimate is obtained by first deriving an asymptotic linearity result. To this end 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. $\mathbf{WX}$ satisfies Huber's condition.

The following lemmas give the projection of $\boldsymbol{S}_{GR}^{(j)}(\mathbf{0})$, for their proofs see Hettmansperger and McKean (1998).

Lemma 3.1.1. *Let* $\overline{\boldsymbol{S}_{GR}^{(j)}(\mathbf{0})} = n^{-3/2}\boldsymbol{S}_{GR}^{(j)}(\mathbf{0})$, $j = 1,\ldots,d$. *Then the projection of* $\overline{\boldsymbol{S}_{GR}^{(j)}(\mathbf{0})}$ *is given by*

$$\overline{\boldsymbol{S}_{GR}^{(j)\,*}(\mathbf{0})} = \frac{2n}{n^{3/2}}\boldsymbol{S}_{GR}^{(j)\,*}(\mathbf{0}) = \frac{2n}{n^{3/2}}\sum_{k=1}^{n}\Big(\sum_{i=1}^{n} w_{ki}\boldsymbol{x}_i\Big)F_j(\mathbf{Y}_k^{(j)}).$$

Further, $\mathrm{Var}(\overline{\boldsymbol{S}_{GR}^{(j)\,*}(\mathbf{0})}) = (\frac{1}{3})n^{-1}\mathbf{X}'\mathbf{W}^2\mathbf{X}$.

Lemma 3.1.2. *Under assumptions A1, B1–B3, and for all* $j = 1,\ldots,d$

$$\overline{\boldsymbol{S}_{GR}^{(j)\,*}(\mathbf{0})} - \overline{\boldsymbol{S}_{GR}^{(j)}(\mathbf{0})} \xrightarrow{\mathcal{P}} \mathbf{0}.$$

Let $\overline{\boldsymbol{S}_{GR}^{*}(\mathbf{0})} = \left(\overline{\boldsymbol{S}_{GR}^{(1)\,*}(\mathbf{0})} \quad \cdots \quad \overline{\boldsymbol{S}_{GR}^{(d)\,*}(\mathbf{0})}\right)$. Then by lemma 3.1.2 we have the result

Corollary 3.1.1.

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

In order to prove the asymptotic normality of $\overline{\boldsymbol{S}_{GR}^{*}(\mathbf{0})}$ we need to state the following theorems. See Arnold (1981).

Theorem 3.1.1. *Let* $\mathbf{U}_1, \mathbf{U}_2, \ldots$ *be a sequence of* $m \times q$ *random matrices. Then* $\mathbf{U}_n \xrightarrow{\mathcal{D}} N_{m,q}(\boldsymbol{\theta}, \boldsymbol{\Xi}, \boldsymbol{\Sigma})$ *if and only if*

$$\mathrm{tr}\,\mathbf{M}'\mathbf{U}_n \xrightarrow{\mathcal{D}} N_1(\mathrm{tr}\,\mathbf{M}'\boldsymbol{\theta}, \mathrm{tr}\,\boldsymbol{\Xi}\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}')$$

for all $m \times q$ matrices $\mathbf{M}$.

Theorem 3.1.2. *Let $\boldsymbol{e}_n' = \begin{pmatrix} \boldsymbol{f}_1 & \cdots & \boldsymbol{f}_n \end{pmatrix}$ where $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots$ is a sequence of independently, identically distributed random vectors in $\Re^d$ such that*

$$E[\boldsymbol{f}_i] = \mathbf{0} \quad \textit{and} \quad \mathrm{Cov}(\boldsymbol{f}_i) = \mathbf{\Sigma}.$$

Let $\mathbf{B}_1, \mathbf{B}_2, \ldots$ be a sequence of constant matrices such that $\mathbf{B}_n$ is $n \times r$, $\mathrm{tr}\,\mathbf{B}_n\mathbf{\Sigma}\mathbf{B}_n' \to c$ and $m(\mathbf{B}_n) = \max_{i,j}|b_{ij}| \to 0$. Then

$$\mathrm{tr}\,\mathbf{B}_n'\boldsymbol{e}_n \xrightarrow{\mathcal{D}} N_1(0, c).$$

Proof. The proof follows using the Lindeberg-Feller and bounded convergence theorems. Let $m_n = m(\mathbf{B}_n\mathbf{\Sigma}\mathbf{B}_n')$. Then $m_n \leq r^2 m(\mathbf{\Sigma})[m(\mathbf{B}_n)]^2 \to 0$. Let $\mathbf{B}_n' = \begin{pmatrix} \boldsymbol{b}_{n1} & \cdots & \boldsymbol{b}_{nn} \end{pmatrix}$ and $W_{ni} = \boldsymbol{b}_{ni}'\boldsymbol{f}_i$. Since the $\boldsymbol{f}_i$ are independent the W_{ni} are independent. Also,

$$E[W_{ni}] = 0, \quad \mathrm{Var}(W_{ni}) = \boldsymbol{b}_{ni}'\mathbf{\Sigma}\boldsymbol{b}_{ni},$$

$$\sum_{i=1}^{n} \mathrm{Var}(W_{ni}) = \sum_{i=1}^{n} \boldsymbol{b}_{ni}'\mathbf{\Sigma}\boldsymbol{b}_{ni} = \mathrm{tr}\,\mathbf{B}_n'\mathbf{\Sigma}\mathbf{B}_n \to c,$$

$$\max_{1 \leq i \leq n} \mathrm{Var}(W_{ni}) = \max_{1 \leq i \leq n} \boldsymbol{b}_{ni}'\mathbf{\Sigma}\boldsymbol{b}_{ni} = m_n \to 0.$$

Hence, by the Lindeberg-Feller theorem

$$\mathrm{tr}\,\mathbf{B}_n'\boldsymbol{e}_n = \sum_{i=1}^{n} \boldsymbol{b}_{ni}'\boldsymbol{f}_i = \sum_{i=1}^{n} W_{ni}$$

is asymptotically normal if we show that for all $\epsilon > 0$

$$C_n = E\sum_{i=1}^{n} W_{ni}^2 I_\epsilon(W_{ni}) \to 0.$$

where

$$I_\epsilon(W_{ni}) = \begin{cases} 1 & \text{if } |W_{ni}| > \epsilon \\ 0 & \text{if } |W_{ni}| \leq \epsilon. \end{cases}$$

By the Cauchy-Schwartz inequality

$$\begin{aligned} W_{ni}^2 &= (\boldsymbol{b}_{ni}'\boldsymbol{f}_i)^2 \qquad (3.4) \\ &= (\,(\boldsymbol{\Sigma}^{1/2}\boldsymbol{b}_{ni})'(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{f}_i)\,)^2 \\ &= \|\boldsymbol{\Sigma}^{1/2}\boldsymbol{b}_{ni}\|^2\,\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{f}_i\|^2 \\ &= (\boldsymbol{b}_{ni}'\boldsymbol{\Sigma}\boldsymbol{b}_{ni})(\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i). \end{aligned}$$

Thus

$$\begin{aligned} C_n &= E[\sum_{i=1}^{n} W_{ni}^2 I_\epsilon(W_{ni})] \\ &\le \sum_{i=1}^{n}(\boldsymbol{b}_{ni}'\boldsymbol{\Sigma}\boldsymbol{b}_{ni})E[\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i I_\epsilon(W_{ni})] \end{aligned}$$

Now using (3.4) we have

$$\begin{aligned} \epsilon^2 &< |w_{ni}|^2 \\ &\le (\boldsymbol{b}_{ni}'\boldsymbol{\Sigma}\boldsymbol{b}_{ni})(\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i) \\ &\le m_n \boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i. \end{aligned}$$

So $|W_{ni}| > \epsilon$ implies that $\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i > \epsilon^2/m_n$. Thus $I_\epsilon(W_{ni}) \le I_{\epsilon^2/m_n}(\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i)$. Let $U = \boldsymbol{f}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_1$. Then,

$$\begin{aligned} C_n &\le \sum_{i=1}^{n}(\boldsymbol{b}_{ni}'\boldsymbol{\Sigma}\boldsymbol{b}_{ni})E[\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i I_{\epsilon^2/m_n}(\boldsymbol{f}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{f}_i)] \\ &= \sum_{i=1}^{n}(\boldsymbol{b}_{ni}'\boldsymbol{\Sigma}\boldsymbol{b}_{ni})E[U I_{\epsilon^2/m_n}(U)] \\ &= \operatorname{tr}\mathbf{B}_n\boldsymbol{\Sigma}\mathbf{B}_n' E[U I_{\epsilon^2/m_n}(U)] \\ &\le cE[U I_{\epsilon^2/m_n}(U)]. \end{aligned}$$

Also, $UI_{\epsilon^2/m_n}(U) \leq U$ and $E[U] = d$. Therefore, by the bounded convergence theorem

$$\lim_{n\to\infty} E[UI_{\epsilon^2/m_n}(U)] = E[\lim_{n\to\infty} UI_{\epsilon^2/m_n}(U)] = 0.$$

Hence, C_n is trapped between 0 and a sequence which is converging to 0 and must also converge to 0. $\square$

According to lemma 3.1.1 the result for the projection is the matrix

$$\begin{aligned}\overline{\boldsymbol{S}_{GR}{}^*(\mathbf{0})} &= \left(\overline{\boldsymbol{S}_{GR}^{(1)}{}^*(\mathbf{0})} \ \cdots \ \overline{\boldsymbol{S}_{GR}^{(d)}{}^*(\mathbf{0})}\right) \\ &= \frac{n}{n^{3/2}}\mathbf{X}'\mathbf{W}\begin{pmatrix} 2F_1(\mathbf{Y}_1^{(1)}) & \cdots & 2F_d(\mathbf{Y}_1^{(d)}) \\ \vdots & \cdots & \vdots \\ 2F_1(\mathbf{Y}_n^{(1)}) & \cdots & 2F_d(\mathbf{Y}_n^{(d)}) \end{pmatrix}\end{aligned} \tag{3.5}$$

Let $\boldsymbol{S} = (s_{jj'})$ where

$$\begin{aligned} s_{jj'} &= \mathrm{Cov}(2F_j(\mathbf{Y}_1^{(j)}), 2F_{j'}(\mathbf{Y}_1^{(j')})) \\ &= 4E[F_j(\mathbf{Y}_1^{(j)})F_{j'}(\mathbf{Y}_1^{(j')})] - 1. \end{aligned}$$

Note that $s_{jj} = 1/3$.

Theorem 3.1.3. *Under assumptions A1, B1–B3 and assuming that $\boldsymbol{\mathcal{B}}_1 = \mathbf{0}$,*

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

Proof. By corollary 3.1.1 we only need to show that

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

Let $\mathbf{M}$ be any $p \times d$ matrix. We want to show that

$$\mathrm{tr}\, \mathbf{M}'\overline{\boldsymbol{S}_{GR}{}^*(\mathbf{0})} \xrightarrow{\mathcal{D}} N_1(\mathbf{0}, \mathrm{tr}\, \mathbf{M}'\mathbf{E}\mathbf{M}\boldsymbol{S}).$$

Now, let

$$\mathbf{F}_n = \begin{pmatrix} \mathbf{F}'(\mathbf{Y}_1) \\ \vdots \\ \mathbf{F}'(\mathbf{Y}_n) \end{pmatrix}$$

where for $i = 1, \dots, n$

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

Then we have $E[\mathbf{F}(\mathbf{Y}_i)] = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{F}(\mathbf{Y}_i)) = \boldsymbol{S}$. Finally, let $\mathbf{B}_n = (1/\sqrt{n})\mathbf{W}_n\mathbf{X}_n\mathbf{M}$. Then by (3.5)

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

and

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

Also, we have

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

by conditions B2 and B3. Thus by theorem 3.1.2 we have

$$\mathrm{tr}\,\mathbf{B}_n'\mathbf{F}_n \xrightarrow{\mathcal{D}} N_1(\mathbf{0}, \mathrm{tr}\,\mathbf{M}\boldsymbol{S}\mathbf{M}'\mathbf{E}).$$

Therefore, by theorem 3.1.1,

$$\overline{\boldsymbol{S}_{GR}{}^{*}(\mathbf{0})} = \mathbf{B}_n' \mathbf{F}_n \xrightarrow{\mathcal{D}} N_{p,d}(\mathbf{0}, \mathbf{E}, \boldsymbol{S})$$

□

Now we state the following results about the asymptotic linearity and quadraticity of $\widehat{\boldsymbol{B}}_{1,GR}^{(j)}$. See Hettmansperger and McKean (1998) for its derivation. As in chapter II, $\tau_j^{-1} = \sqrt{12}\int f_j^2(x)\,dx$ and $\mathbf{T}$ is the $d \times d$ diagonal matrix whose j$^{\text{th}}$ diagonal element is τ_j.

Theorem 3.1.4. *Under the assumptions A1, B1–B3 and for a fixed* $j \in \{1, \ldots, d\}$

$$n^{-3/2}\boldsymbol{S}_{GR}^{(j)}(\boldsymbol{B}_1^{(j)}) = n^{-3/2}\boldsymbol{S}_{GR}^{(j)}(\mathbf{0}) - \sqrt{n}\,(\sqrt{3}\,\tau_j)^{-1}\mathbf{C}\boldsymbol{B}_1^{(j)} + o_p(1)$$

uniformly for all $\boldsymbol{B}_1^{(j)}$ *such that* $\sqrt{n}\|\boldsymbol{B}_1^{(j)}\| \le r$*, for any* $r > 0$.

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

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

The vector $\boldsymbol{B}_1^{(j)}$ that minimizes $Q(\boldsymbol{B}_1^{(j)})$ is

$$\widetilde{\boldsymbol{B}}_{1,GR}^{(j)} = \frac{\sqrt{3}\,\tau_j}{n}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\boldsymbol{S}_{GR}^{(j)}(\mathbf{0}). \tag{3.6}$$

Lemma 3.1.3. *For all* $j = 1, \ldots, d$,

$$\sqrt{n}\,(\widetilde{\boldsymbol{B}}_{1,GR}^{(j)} - \widehat{\boldsymbol{B}}_{1,GR}^{(j)}) \xrightarrow{P} \mathbf{0}.$$

Theorem 3.1.5.

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

Proof. Let

$$\widetilde{\boldsymbol{\mathcal{B}}}_{1,GR} = \begin{pmatrix} \widetilde{\boldsymbol{\mathcal{B}}}_{1,GR}^{(1)} & \cdots & \widetilde{\boldsymbol{\mathcal{B}}}_{1,GR}^{(d)} \end{pmatrix}.$$

Then by (3.6)

$$\widetilde{\boldsymbol{\mathcal{B}}}_{1,GR} = \frac{\sqrt{3}}{n}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \begin{pmatrix} \boldsymbol{S}_{GR}^{(1)}(\mathbf{0}) & \cdots & \boldsymbol{S}_{GR}^{(d)}(\mathbf{0}) \end{pmatrix} \begin{pmatrix} \tau_1 & & \\ & \ddots & \\ & & \tau_d \end{pmatrix}$$

Thus,

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

Now the result follows from lemma 3.1.3. □

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

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

It follows then that

$$\widehat{\boldsymbol{\beta}}_{0,GR}' = \widehat{\boldsymbol{\alpha}}_{0,GR}' - \bar{\boldsymbol{x}}'\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}.$$

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

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

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

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

As in chapter II, let $\tau_j^{*-1} = 2f_j(0)$ and $\mathbf{T}^*$ be the $d \times d$ diagonal matrix whose j$^{\text{th}}$ element is τ_j^*.

Lemma 3.1.4. *If assumptions A1, A2 and A3 hold and*

$$\boldsymbol{\Phi}^* = (\mathrm{sgn}(\epsilon_{ij}))$$

then for all $\epsilon > 0$,

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

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

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

Note that

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

when the true parameters are zero. Now

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

The first and second terms go to 0 by theorem 3.5.9 and lemma 3.5.8 of Hettmansperger and McKean (1998). $\square$

Now recall from chapter II the matrix $\boldsymbol{S}^* = (s^*_{jj'})$ where

$$s^*_{jj'} = \Pr(\varepsilon_{ij} < 0, \varepsilon_{ij'} < 0) + \Pr(\varepsilon_{ij} > 0, \varepsilon_{ij'} > 0) - \Pr(\varepsilon_{ij} < 0, \varepsilon_{ij'} > 0) - \Pr(\varepsilon_{ij} > 0, \varepsilon_{ij'} < 0).$$

It follows from theorem 18 of Lehman (1975) that

$$(\frac{1}{\sqrt{n}}\mathbf{1}_n'\mathbf{\Phi}^*)' \xrightarrow{\mathcal{D}} N_d(\mathbf{0}, \boldsymbol{S}^*).$$

Thus,

$$\sqrt{n}\,\widehat{\boldsymbol{\alpha}}_{0,GR}' \xrightarrow{\mathcal{D}} N_d(\mathbf{0}, \mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*). \tag{3.7}$$

Theorem 3.1.6.

$$\sqrt{n}\,\mathrm{vec}\begin{pmatrix}\widehat{\boldsymbol{\alpha}}_{0,GR} & \widehat{\boldsymbol{\mathcal{B}}}_{1,GR}'\end{pmatrix} \xrightarrow{\mathcal{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\end{pmatrix}\right). \tag{3.8}$$

Proof. The result follows from 3.7 and theorem 3.1.5. Note that $\mathbf{X}'\mathbf{W}\mathbf{X} = \mathbf{X}_c'\mathbf{W}\mathbf{X}_c$ and $\mathbf{X}'\mathbf{W}^2\mathbf{X} = \mathbf{X}_c'\mathbf{W}^2\mathbf{X}_c$ because the rows and columns of $\mathbf{W}$ sum to zero. □

Corollary 3.1.2.

$$\sqrt{n}\,\mathrm{vec}\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{0,GR}-\boldsymbol{\beta}_0 & \widehat{\boldsymbol{\mathcal{B}}}_{1,GR}'-\boldsymbol{\mathcal{B}}_1'\end{pmatrix}$$
$$\xrightarrow{\mathcal{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + 3(\bar{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\bar{\boldsymbol{x}})\mathbf{T}\boldsymbol{S}\mathbf{T} & -\mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\bar{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1} \\ -\mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\bar{\boldsymbol{x}} & \mathbf{T}\boldsymbol{S}\mathbf{T}\otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\end{pmatrix}\right). \tag{3.9}$$

Proof. The result is obtained by noting that

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

Now the result follows from theorem 3.1.6 . □

The influence function of the GR-estimate for the univariate case was derived by Naranjo and Hettmansperger (1994). The following theorem is an extension of their result to the componentwise multivariate regression.

Theorem 3.1.7. *Let M denote the distribution function of $\boldsymbol{x}$ and $b(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be the weight function. Then the influence function of $\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}$ is*

$$IF(\boldsymbol{x}_0, \boldsymbol{y}_0, \widehat{\boldsymbol{\mathcal{B}}}_{1,GR}) = \sqrt{12}\,\mathbf{C}^{-1}\int(\boldsymbol{x}-\boldsymbol{x}_0)\,dM(\boldsymbol{x})$$
$$\times\left(F_1(\boldsymbol{y}_0^{(1)}-\boldsymbol{\alpha}_0^{(1)}-\boldsymbol{x}_0'\boldsymbol{\mathcal{B}}_1^{(1)})-\tfrac{1}{2}\quad\cdots\quad F_d(\boldsymbol{y}_0^{(d)}-\boldsymbol{\alpha}_0^{(d)}-\boldsymbol{x}_0'\boldsymbol{\mathcal{B}}_1^{(d)})-\tfrac{1}{2}\right)\mathbf{T}.$$

Proof. This result follows immediately from theorem 5.7.1 of Hettmansperger and McKean (1998) which gives for $j = 1, \ldots, d$ the following result

$$\mathrm{IF}\,(\boldsymbol{x}_0, \boldsymbol{y}_0^{(j)}, \widehat{\boldsymbol{\mathcal{B}}}_{1,GR}^{(j)}) = \tau_j\sqrt{12}\,(F(\boldsymbol{y}_0^{(j)}-\boldsymbol{\alpha}_0^{(j)}-\boldsymbol{x}_0'\boldsymbol{\mathcal{B}}_1^{(j)})-\frac{1}{2})\mathbf{C}^{-1}\int(\boldsymbol{x}-\boldsymbol{x}_0)\,dM(\boldsymbol{x}).$$

□

Note that the influence function is bounded in the $\boldsymbol{y}$-space and by a proper choice of the weight function it is also bounded in the $\boldsymbol{x}$-space.

3.1.2 A Quadratic Test

Using the asymptotic distribution of $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$ the following theorem gives a quadratic test for the hypothesis $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$.

Theorem 3.1.8. *Under $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$,*

$$n\,\boldsymbol{\mathcal{L}}_{GR}'[(\mathbf{K}'\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*\mathbf{K}\otimes\mathbf{H}_1\mathbf{H}_1') + (\mathbf{K}'\mathbf{T}\boldsymbol{S}\mathbf{T}\mathbf{K}\otimes M)]^{-1}\boldsymbol{\mathcal{L}}_{GR}\xrightarrow{\mathcal{D}}\chi^2(rs),$$

where

$$\boldsymbol{\mathcal{L}}_{GR} = \text{vec}(\mathbf{K}'\widehat{\boldsymbol{\mathcal{B}}}_{GR}\mathbf{H}'),$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1^{r\times 1} & \mathbf{H}_2^{r\times p} \end{pmatrix}$$

and

$$M = n(-\mathbf{H}_1\bar{\boldsymbol{x}}' + \mathbf{H}_2)(\mathbf{X}_c'\mathbf{W}\mathbf{X}_c)^{-1}(\mathbf{X}_c'\mathbf{W}^2\mathbf{X}_c)(\mathbf{X}_c'\mathbf{W}\mathbf{X}_c)^{-1}(-\mathbf{H}_1\bar{\boldsymbol{x}}' + \mathbf{H}_2)'.$$

Proof. The proof follows from lemma 2.5.1 and corollary 3.1.2. □

Let

$$V_{GR} = (\mathbf{K}'\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*\mathbf{K} \otimes \mathbf{H}_1\mathbf{H}_1') + (\mathbf{K}'\mathbf{T}\boldsymbol{S}\mathbf{T}\mathbf{K} \otimes M).$$

Then it follows from theorem 3.1.8 that we reject $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$ in favor of $H_A : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} \neq \mathbf{0}$ at asymptotic level α if

$$Q_{GR} = n\,\boldsymbol{\mathcal{L}}_{GR}'V_{GR}^{-1}\boldsymbol{\mathcal{L}}_{GR} \geq \chi^2_\alpha(rs).$$

3.2 Transformation-Retransformation GR Estimation

Like the other componentwise estimates of $\boldsymbol{\mathcal{B}}$, the GR-estimate $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$, lacks the affine equivariance property. Applying the transformation-retransformation technique to $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$ produces an estimate, $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$, that is affine equivariant, has a bounded influence in both factor and response spaces and is highly efficient compared to componentwise estimates especially when the response variables are strongly correlated.

3.2.1 Estimation

The estimation of $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ consists of three steps. First, the response variables $\boldsymbol{y}_i$ are transformed to $\boldsymbol{z}_i = \hat{A}\boldsymbol{y}_i$. Next, the componentwise estimate $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$ is computed on $(\boldsymbol{x}_1, \boldsymbol{z}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{z}_n)$. Finally, $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$ is retransformed into $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR} = \widehat{\boldsymbol{\mathcal{B}}}_{GR}(\hat{A}')^{-1}$.

The weights for the ith case, used in the componentwise estimation of $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$, are defined as $b_{ij} = b_i b_j$, where

$$b_i = \min\{1, c/\sqrt{(\boldsymbol{x}_i - \mathbf{v})'\mathbf{V}^{-1}(\boldsymbol{x}_i - \mathbf{v})}\}^{\alpha}, \tag{3.10}$$

and $(\mathbf{v}, \mathbf{V})$ are the minimum covariance determinant (MCD) estimates of location and scatter. [See Rousseeuw and Van Driessen (1999) for a fast MCD algorithm]. The MCD finds h observations (out of n) whose classical covariance matrix has the lowest determinant. The MCD estimate of location is then the average of these h points, and the MCD estimate of scatter is their covariance matrix. 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.

3.2.2 Affine Equivariance of $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$

Lemma 3.2.1. $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR} = \widehat{\boldsymbol{\mathcal{B}}}_{TRGR}(\boldsymbol{x}, \boldsymbol{y})$ *has the affine equivariance properties*

1. $\boldsymbol{y}$*- affine equivariance.*

$$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b}) = \begin{pmatrix} \widehat{\beta}'_{0,TRGR}D' + \boldsymbol{b}' \\ \widehat{\boldsymbol{\mathcal{B}}}_{1,TRGR}D' \end{pmatrix}$$

where D is any fixed $d \times d$ nonsingular matrix and $\boldsymbol{b}$ is a $d \times 1$ constant vector.

2. *regression affine equivariance.*

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

where G is any $p \times d$ matrix.

3. *$\boldsymbol{x}$- affine equivariance.*

$$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}(Q\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y}) = \begin{pmatrix} \widehat{\beta}'_{0,TRGR} - \boldsymbol{c}'(Q^{-1})'\widehat{\boldsymbol{\mathcal{B}}}_{1,TRGR} \\ (Q^{-1})'\widehat{\boldsymbol{\mathcal{B}}}_{1,TRGR} \end{pmatrix}$$

where Q is a $p \times p$ nonsingular matrix and $\mathbf{c}$ is a $p \times 1$ constant vector.

Proof. The proof of this lemma is similar to the proof of lemma 2.2.1. The results follow by showing the following simple properties of $\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}$

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}(\boldsymbol{x}, D\boldsymbol{y} + \boldsymbol{b}) = c_0\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}$$

for some constant $c_0 > 0$,

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}(\boldsymbol{x}, \boldsymbol{y} - G'\boldsymbol{x}) = \widehat{\boldsymbol{\mathcal{B}}}_{1,GR} - G\hat{A}'$$

and

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}(Q\boldsymbol{x} + \boldsymbol{c}, \boldsymbol{y}) = (Q')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}.$$

To show the last property of $\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}$ it is easily seen from (3.10) that the weights $b_i(Q\boldsymbol{x} + \boldsymbol{c}) = b_i(\boldsymbol{x}) = b_i$ and thus $\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}$ for $k = 1, \ldots, d$ minimizes the dispersion

function

$$
\begin{aligned}
D_{GR}(\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) &= \sum_{i<j} b_{ij}(Q\boldsymbol{x}+\boldsymbol{c})\,|(\mathbf{Y}_i^{(k)} - (Q\boldsymbol{x}_i+\boldsymbol{c})'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) \\
&\qquad - (\mathbf{Y}_j^{(k)} - (Q\boldsymbol{x}_j+\boldsymbol{c})'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}))| \\
&= \sum_{i<j} b_{ij}\,|(\mathbf{Y}_i^{(k)} - (Q\boldsymbol{x}_i)'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) \\
&\qquad - (\mathbf{Y}_j^{(k)} - (Q\boldsymbol{x}_j)'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}))| \\
&= \sum_{i<j} b_{ij}\,|(\mathbf{Y}_i^{(k)} - \boldsymbol{x}_i'Q'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y})) \\
&\qquad - (\mathbf{Y}_j^{(k)} - \boldsymbol{x}_j'Q'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}))|
\end{aligned}
$$

Thus, we have for all $k = 1, \ldots, d$,

$$\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR} = Q'\widehat{\boldsymbol{\mathcal{B}}}^{(k)}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}).$$

Hence,

$$\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}(Q\boldsymbol{x}+\boldsymbol{c},\boldsymbol{y}) = (Q')^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{1,GR}.$$

□

3.2.3 Asymptotic Normality

Asymptotic properties of $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ follows from the corresponding properties of the componentwise estimate $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$.

Theorem 3.2.1.

$$\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRGR} - \boldsymbol{\mathcal{B}})' \xrightarrow{\mathcal{D}} N_{(p+1)d}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{S}_{11,TRGR} & \boldsymbol{S}_{12,TRGR} \\ \boldsymbol{S}_{21,TRGR} & \boldsymbol{S}_{22,TRGR} \end{pmatrix}\right) \tag{3.11}$$

where

$$\boldsymbol{S}_{11,TRGR} = A^{-1}(\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + 3(\bar{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\bar{\boldsymbol{x}})\mathbf{T}\boldsymbol{S}\mathbf{T})(A^{-1})'$$

$$\boldsymbol{S}_{12,TRGR} = -A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes 3\bar{\boldsymbol{x}}'\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}$$

$$\boldsymbol{S}_{21,TRGR} = -A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}\bar{\boldsymbol{x}}$$

$$\boldsymbol{S}_{22,TRGR} = A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1}$$

Proof. Similar to $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ the consistency of $\hat{A}$ and corollary 3.1.2 gives the above asymptotic distribution of $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$. □

Corollary 3.2.1.

$$\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{1,TRGR} - \boldsymbol{\mathcal{B}}_1) \xrightarrow{\mathcal{D}} N_{pd}(\mathbf{0}, 3\mathbf{C}^{-1}\mathbf{E}\mathbf{C}^{-1} \otimes A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})'). \tag{3.12}$$

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

Corollary 3.2.2. *The influence function of $\widehat{\boldsymbol{\mathcal{B}}}_{1,TRGR}$ is*

$$\begin{aligned} IF(\boldsymbol{x}_0, \boldsymbol{y}_0, \widehat{\boldsymbol{\mathcal{B}}}_{1,TRGR}) &= \sqrt{12}\,\mathbf{C}^{-1} \int (\boldsymbol{x} - \boldsymbol{x}_0)\, dM(\boldsymbol{x}) \\ &\times \left(F_1(y_0^{(1)} - \alpha_0^{(1)} - \boldsymbol{x}_0'\boldsymbol{\mathcal{B}}_1^{(1)}) - \tfrac{1}{2} \quad \cdots \quad F_d(y_0^{(d)} - \alpha_0^{(d)} - \boldsymbol{x}_0'\boldsymbol{\mathcal{B}}_1^{(d)}) - \tfrac{1}{2} \right) \\ &\times \mathbf{T}(A^{-1})'. \end{aligned}$$

Proof. This result follows from theorem 3.1.7 and and theorem 3.2.1. □

3.2.4 A Quadratic Test

Based on the asymptotic distribution of $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ we can obtain obtain a test for the hypothesis $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$.

Theorem 3.2.2. *Under* $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$,

$$n\,\boldsymbol{\mathcal{L}}'_{TRGR}[(\mathbf{K}'A^{-1}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*(A^{-1})'\mathbf{K}\otimes\mathbf{H}_1\mathbf{H}_1')+(\mathbf{K}'A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})'\mathbf{K}\otimes M)]^{-1}\boldsymbol{\mathcal{L}}_{TRGR} \xrightarrow{\mathcal{D}} \chi^2(rs),$$

where

$$\boldsymbol{\mathcal{L}}_{TRGR} = \mathrm{vec}(\mathbf{K}'\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}\mathbf{H}'),$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1^{r\times 1} & \mathbf{H}_2^{r\times p} \end{pmatrix}$$

and

$$M = n(-\mathbf{H}_1\bar{\boldsymbol{x}}' + \mathbf{H}_2)(\mathbf{X}_c'\mathbf{W}\mathbf{X}_c)^{-1}(\mathbf{X}_c'\mathbf{W}^2\mathbf{X}_c)(\mathbf{X}_c'\mathbf{W}\mathbf{X}_c)^{-1}(-\mathbf{H}_1\bar{\boldsymbol{x}}' + \mathbf{H}_2)'.$$

Proof. the proof follows from lemma 2.5.1 and theorem 3.2.1. □

Let

$$V_{TRGR} = (\mathbf{K}'A^{-1}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*(A^{-1})'\mathbf{K}\otimes\mathbf{H}_1\mathbf{H}_1') + (\mathbf{K}'A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})'\mathbf{K}\otimes M).$$

Then according to theorem 3.2.2 we reject $H_0 : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} = \mathbf{0}$ in favor of $H_A : \mathbf{H}\boldsymbol{\mathcal{B}}\mathbf{K} \neq \mathbf{0}$ at asymptotic level α if

$$Q_{TRGR} = n\,\boldsymbol{\mathcal{L}}'_{TRGR}V_{TRGR}^{-1}\boldsymbol{\mathcal{L}}_{TRGR} \geq \chi^2_\alpha(rs).$$

3.2.5 A Diagnostic for Comparing $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$

The transformed-retransformed R-estimate, $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$, is robust against outlying $\boldsymbol{y}$-values but it is sensitive to points that are outlying in $\boldsymbol{x}$-space. However, the transformed retransformed GR-estimate, $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$, is robust in both $\boldsymbol{y}$ and $\boldsymbol{x}$ spaces. McKean

et al. (1996) proposed for the univariate case the following diagnostic for the overall difference between the R- and GR- fits

$$\text{TDBETAS}_R = (\widehat{\beta}_R - \widehat{\beta}_{GR})' A_R^{-1} (\widehat{\beta}_R - \widehat{\beta}_{GR})$$

where

$$A_R = \begin{pmatrix} \tau_s^2/n & \mathbf{0} \\ 0 & \tau^2(\mathbf{X}_c'\mathbf{X}_c)^{-1} \end{pmatrix}$$

is the covariance matrix of $\widehat{\beta}_R$. The difference in the betas was standardized by A_R rather than by the covariance matrix A_D of the difference $\widehat{\beta}_R - \widehat{\beta}_{GR}$ because A_D is singular as the intercepts of the two procedures were computed as the medians of the residuals and thus they have the same asymptotic representations. To obtain a benchmark for $TDBETAS$ they replaced τ_s by τ in A_R and then decided to flag the ith observation whenever

$$\frac{|\widehat{y}_{i,R} - \widehat{y}_{i,GR}|}{\widehat{\tau}\sqrt{h_{ii}}} > 2\sqrt{(p+1)/n}.$$

They used this result to get the benchmark $4(p+1)^2/n$ for the $TDBETAS_R$.

The univariate diagnostic $TDBETAS_R$ can be extended to the multivariate case to obtain a diagnostic and criteria for the overall difference between the transformation-retransformation R- and GR- fits. As the intercept vectors of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ were computed using the median of the residuals, the covariance matrix of the difference $\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ is singular. Thus, we define the following measure of the total difference in the fits of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ as

$$TDBETAS_{TRR} = (\text{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})')' A_{TRR}^{-1} \text{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})' \tag{3.13}$$

where

$$A_{TRR} = \begin{pmatrix} A^{-1}[\frac{1}{n}\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^* + \bar{\boldsymbol{x}}'(\mathbf{X}_c'\mathbf{X}_c)^{-1}\bar{\boldsymbol{x}}\mathbf{T}\boldsymbol{S}\mathbf{T}](A^{-1})' & A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes \bar{\boldsymbol{x}}'(\mathbf{X}_c'\mathbf{X}_c)^{-1} \\ A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes (\mathbf{X}_c'\mathbf{X}_c)^{-1}\bar{\boldsymbol{x}} & A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes (\mathbf{X}_c'\mathbf{X}_c)^{-1} \end{pmatrix}$$

is the covariance matrix of $\widehat{\boldsymbol{B}}_{TRR}'$. To derive a benchmark for $TDBETAS_{TRR}$ we need the following lemma. See Muirhead (1982).

Lemma 3.2.2.

$$\mathrm{tr}\,(\mathbf{B}\mathbf{X}'\mathbf{C}\mathbf{X}\mathbf{D}) = (\mathrm{vec}(\mathbf{X}))'(\mathbf{C} \otimes \mathbf{B}'\mathbf{D}')\mathrm{vec}(\mathbf{X}).$$

Now to obtain the benchmark for the $TDBETAS_{TRR}$ we let $\mathbb{X} = \begin{pmatrix} \mathbf{1}_n & \mathbf{X}_c \end{pmatrix}$ and replace $\mathbf{T}^*\boldsymbol{S}^*\mathbf{T}^*$ by $\mathbf{T}\boldsymbol{S}\mathbf{T}$ in the expression of A_{TRR}. Then we have the following approximation of A_{TRR}

$$A_{TRR} \doteqdot \begin{pmatrix} A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes [\frac{1}{n} + \bar{\boldsymbol{x}}'(\mathbf{X}_c'\mathbf{X}_c)^{-1}\bar{\boldsymbol{x}}] & A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes \bar{\boldsymbol{x}}'(\mathbf{X}_c'\mathbf{X}_c)^{-1} \\ A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes (\mathbf{X}_c'\mathbf{X}_c)^{-1}\bar{\boldsymbol{x}} & A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes (\mathbf{X}_c'\mathbf{X}_c)^{-1} \end{pmatrix}$$

$$= A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes \begin{pmatrix} \frac{1}{n} + \bar{\boldsymbol{x}}'(\mathbf{X}_c'\mathbf{X}_c)^{-1}\bar{\boldsymbol{x}} & \bar{\boldsymbol{x}}'(\mathbf{X}_c'\mathbf{X}_c)^{-1} \\ (\mathbf{X}_c'\mathbf{X}_c)^{-1}\bar{\boldsymbol{x}} & (\mathbf{X}_c'\mathbf{X}_c)^{-1} \end{pmatrix}$$

$$= A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes (\mathbb{X}'\mathbb{X})^{-1}.$$

Then it follows from lemma 3.2.2 that

$$\begin{aligned}
TDBETAS_{TRR} &= (\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})')' A_{TRR}^{-1} \mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})' \\
&= (\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})')' [A^{-1}\mathbf{T}\boldsymbol{S}\mathbf{T}(A^{-1})' \otimes (\mathbb{X}'\mathbb{X})^{-1}]^{-1} \\
&\qquad \times \mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})' \\
&= (\mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})')' [A'(\mathbf{T}\boldsymbol{S}\mathbf{T})^{-1} A \otimes \mathbb{X}'\mathbb{X}] \mathrm{vec}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})' \\
&= \mathrm{tr}\,\{\mathbb{X}'\mathbb{X}(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR}) A'(\mathbf{T}\boldsymbol{S}\mathbf{T})^{-1} A(\widehat{\boldsymbol{\mathcal{B}}}_{TRR} - \widehat{\boldsymbol{\mathcal{B}}}_{TRGR})'\} \\
&= \mathrm{tr}\,\{\mathbb{X}(\widehat{\boldsymbol{\mathcal{B}}}_{R} - \widehat{\boldsymbol{\mathcal{B}}}_{GR})(\mathbf{T}\boldsymbol{S}\mathbf{T})^{-1}(\widehat{\boldsymbol{\mathcal{B}}}_{R} - \widehat{\boldsymbol{\mathcal{B}}}_{GR})'\mathbb{X}'\} \\
&= \mathrm{tr}\,\{(\widehat{\mathbf{Y}}_R - \widehat{\mathbf{Y}}_{GR})'(\widehat{\mathbf{Y}}_R - \widehat{\mathbf{Y}}_{GR})(\mathbf{T}\boldsymbol{S}\mathbf{T})^{-1}\}.
\end{aligned}$$

Let

$$\begin{aligned}
\mathbf{D} &= (\widehat{\mathbf{Y}}_R - \widehat{\mathbf{Y}}_{GR})(\mathbf{T}\boldsymbol{S}\mathbf{T})^{-1/2} \\
&= \begin{pmatrix} \mathbf{D}^{(1)} & \cdots & \mathbf{D}^{(d)} \end{pmatrix}.
\end{aligned}$$

Then using the fact that average value of h_{ii} is $(p+1)/n$ we get

$$\begin{aligned}
TDBETAS_{TRR} &= \mathrm{tr}\,(\mathbf{D}'\mathbf{D}) \\
&= \sum_{j=1}^{d} \mathbf{D}^{(j)\prime}\mathbf{D}^{(j)} \\
&> \sum_{j=1}^{d}\sum_{i=1}^{n} \left(\frac{\widehat{y}_{i,R}^{(j)} - \widehat{y}_{i,GR}^{(j)}}{\widehat{\tau}_j} \right)^2 \\
&= \sum_{j=1}^{d} (p+1)\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\widehat{y}_{i,R}^{(j)} - \widehat{y}_{i,GR}^{(j)}}{\widehat{\tau}_j\sqrt{(p+1)/n}} \right)^2 \\
&> 4d(p+1)\frac{1}{n}\sum_{i=1}^{n} h_{ii} \\
&> 4d\frac{(p+1)^2}{n}.
\end{aligned}$$

CHAPTER IV

NUMERICAL RESULTS

4.1 Numerical Examples

The examples of this section illustrate the need of robust affine equivariant techniques for the analysis of a data set that has highly correlated response variables. The first two examples illustrate the estimation of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$. Example 4.1.1 compares $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ with the rank transformation-retransformation estimate of Chakraborty and Chaudhuri (1997) while example 4.1.2 compares $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ with $\widehat{\boldsymbol{\mathcal{B}}}_{SCM}$, Oja's estimate (2002).

The last two examples illustrate the use of the quadratic tests based on $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$. These examples also compares with the results obtained from the quadratic tests based on the componentwise estimates $\widehat{\boldsymbol{\mathcal{B}}}_{R}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$.

For comparison purposes, the finite sample variations of the estimates $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ have been computed using a bootstrap technique for the data sets of example 4.1.1 and example 4.1.2. The resampling variation of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ or $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ is obtained by resampling from the pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i = 1, \ldots, n$, computing the transformation matrix on the residuals of the initial LS fit of the bootstrap sample, transforming the bootstrap sample to $(\boldsymbol{x}_i, \boldsymbol{z}_i)$ where $\boldsymbol{z}_i = \hat{A}\boldsymbol{y}_i$, computing the componentwise estimates $\widehat{\boldsymbol{\mathcal{B}}}_{R}$ or $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$ on the transformed sample, and finally retransforming the later estimates to $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ or $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$.

The sample variance-covariance matrix of the column vector of an estimate $\widehat{\mathcal{B}}$ of $\mathcal{B}$ is computed from the bootstrap values as follows. Let $\mathcal{B}_i^*$ be the bootstrap estimates for $i = 1, \ldots, nb$. Let

$$\bar{\mathcal{B}}^* = \frac{1}{nb} \sum_{i=1}^{nb} \mathcal{B}_i^*.$$

Then the sample variance-covariance matrix of the column vector of $\widehat{\mathcal{B}}$ is

$$\mathrm{Cov}_{\mathrm{boot}}(\mathrm{vec}(\widehat{\mathcal{B}})) = \frac{1}{nb} \sum_{i=1}^{nb} \mathrm{vec}(\mathcal{B}_i^* - \bar{\mathcal{B}}^*)\,\mathrm{vec}(\mathcal{B}_i^* - \bar{\mathcal{B}}^*)'.$$

Example 4.1.1 *(Blood Pressure Data).* This data 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. 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 stands for the systolic blood pressure and y_2 stands for the diastolic blood pressure. Let x stands for age. Then upon fitting this data set to model 1.1 with $d = 2$ and $p = 1$ we get the estimates

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

and

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

The first number in parenthesis is the estimated standard errors while the second number is the bootstrap standard errors. These standard errors have been computed based on 10,000 bootstrap replications. Note that the R and GR fits are almost the same.

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

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

Also, Chakraborty and Chaudhuri computed the standard errors of their estimate using a bootstrap technique. However, they kept their transformation matrix fixed. They obtained as standard errors of the coefficients of age for y_1 and y_2 the values 0.20 and 0.11 respectively.

Example 4.1.2 *(Oxygen-Consumption Data).* The data set consists of 25 4-variate observations, namely the results on 4 measurements of oxygen consumption for 25 males measured by researchers interested in pulmonary function, see Johnson and Wichern (1998). The subjects were asked to run on a treadmill until exhaustion and then samples of air were collected at definite intervals and the gas contents analyzed. Let the explanatory variables be

$$x_1 = \text{ resting volume } O_2 \text{ (L/min)}$$

$$x_2 = \text{ resting volume } O_2 \text{ (mL/kg/min)},$$

and the response variables be

$$y_1 = \text{maximium volume } O_2 \text{ (L/min)}$$

$$y_2 = \text{maximum volume } O_2 \text{ (mL/kg/min)}.$$

The estimates $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ are given below together with the estimated and bootstrap standard errors. The bootstrap standard errors were computed on 10,000 bootstrap replications.

$$\widehat{\mathcal{B}}_{TRR} = \begin{pmatrix} 2.73(0.63, 0.62) & 38.39(8.31, 8.34) \\ 8.70(2.24, 2.44) & -10.15(29.28, 31.26) \\ -0.45(0.18, 0.19) & 3.04(2.31, 2.48) \end{pmatrix},$$

$$\widehat{\mathcal{B}}_{TRGR} = \begin{pmatrix} 2.70(0.67, 0.71) & 37.65(7.74, 9.49) \\ 8.45(2.44, 2.76) & -14.00(28.08, 35.63) \\ -0.42(0.19, 0.22) & 3.50(2.22, 2.82) \end{pmatrix}.$$

The SCM regression estimate of $\mathcal{B}$ obtained by Oja (2002) for this data set was

$$\widehat{\mathcal{B}}_{SCM} = \begin{pmatrix} 2.43(0.57) & 34.58(7.48) \\ 8.24(2.18) & -15.63(27.33) \\ -0.38(0.19) & 3.94(2.37) \end{pmatrix},$$

where the standard errors were obtained using 250 bootstrap samples. Note that $\widehat{\mathcal{B}}_{TRGR}$ is close to $\widehat{\mathcal{B}}_{SCM}$. The LS regression estimate of $\mathcal{B}$ is

$$\widehat{\mathcal{B}}_{LS} = \begin{pmatrix} 2.44(0.54) & 34.93(6.94) \\ 7.90(1.91) & -20.90(24.82) \\ -0.35(0.15) & 4.28(1.96) \end{pmatrix}$$

The difference in the fits between $\widehat{\mathcal{B}}_{TRR}$ and $\widehat{\mathcal{B}}_{LS}$ and the other two estimates might have been caused by the presence of outliers in the x-space, namely in the values of the regressor x_1 of the second component.

Example 4.1.3. The data set for this example was generated using model 1.1 with $d = 2$ and $p = 2$. Thus, the model used to generate the data was

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix}.$$

We set $\beta_{01} = \beta_{02} = 0$, $\beta_{11} = -2$, $\beta_{12} = 3$, and $\beta_{21} = \beta_{22} = 4$. For the errors, we first generated a matrix $E = \begin{pmatrix} E^{(1)} & E^{(2)} \end{pmatrix}$ of size 30×2 of independent observations from $N(0,1)$. Then for $i = 1, \ldots, n$ we defined the errors as

$$\epsilon_{i1} = 0.7E^{(1)} + 0.4E^{(2)},$$

and

$$\epsilon_{i2} = 0.5E^{(1)} + 0.8E^{(2)}.$$

The independent x's were generated as random samples of size $n = 30$ from $N(0,1)$. We obtained the following estimates. The estimated standard error of the estimates are given by the first number in the parenthesis. The bootstrap standard errors were computed using 10,1000 replications and are given by the second number in the parenthesis.

$$\widehat{\boldsymbol{\mathcal{B}}}_{TRR} = \begin{pmatrix} -0.11(0.18, 0.25) & -0.90(0.47, 0.47) \\ -1.90(0.17, 0.16) & 2.86(0.40, 0.31) \\ 3.95(0.13, 0.18) & 3.97(0.32, 0.26) \end{pmatrix},$$

$$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR} = \begin{pmatrix} -0.14(0.22, 0.27) & -0.98(0.46, 0.51) \\ -1.92(0.22, 0.22) & 2.82(0.40, 0.50) \\ 3.94(0.17, 0.20) & 3.94(0.32, 0.31) \end{pmatrix},$$

$$\widehat{\boldsymbol{\mathcal{B}}}_{LS} = \begin{pmatrix} -0.19(0.16) & -0.72(0.32) \\ -1.90(0.17) & 2.92(0.33) \\ 4.01(0.13) & 3.98(0.26) \end{pmatrix}.$$

Table 1 displays the results for the quadratic tests based on $\widehat{\boldsymbol{\mathcal{B}}}_R$, $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$, $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$, $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$

Table 1

Observed values and p values

Hypothesis	df	Q_R	Q_{TRR}	Q_{GR}	Q_{TRGR}	Q_{LH}
H_{01}	4	1212.73	2086.34	1288.54	14581.01	1613.42
p-value		0.000	0.000	0.000	0.000	0.000
H_{02}	2	231.28	272.37	240.02	672.24	363.40
p-value		0.000	0.000	0.000	0.000	0.000
H_{03}	1	0.03	4.11×10^{-3}	0.04	1.31×10^{-4}	0.02
p-value		0.872	0.949	0.83	0.991	0.884

and the Lawley Hotteling's test for the following hypothesis

$$H_{01} : \boldsymbol{y} \text{ is linearly related to } \boldsymbol{x},$$

$$H_{02} : y_1 \text{ and } y_2 \text{ have the same linear model},$$

$$H_{03} : \beta_{21} = \beta_{22}.$$

For testing H_{01} we used the matrices

$$H = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

while for the contrast test of H_{02} we used the same H as for H_{01} but we used $K = \begin{pmatrix} 1 & -1 \end{pmatrix}'$. The test of H_{03} was done using $H = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ and $K = \begin{pmatrix} 1 & -1 \end{pmatrix}'$.

Example 4.1.4. The data set of this example is the same as the one used for example 4.1.3 but the value of x_1 was changed for the 9th observation to 10,000 and the value of x_2 was changed for the 23rd observation to 100,000. These changes gave the following

estimates of $\boldsymbol{\mathcal{B}}$ and their standard errors

$$\widehat{\boldsymbol{\mathcal{B}}}_{TRR} = \begin{pmatrix} 2.41(0.71, 1.29) & 1.87(0.74, 1.39) \\ -2.00(4.2 \times 10^{-4}, 0.61) & 2.99(4.3 \times 10^{-4}, 0.65) \\ 0.04(4.2 \times 10^{-5}, 1.85) & 0.04(4.3 \times 10^{-5}, 1.87) \end{pmatrix},$$

$$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR} = \begin{pmatrix} -0.50(1343.34, 0.27) & -0.76(1099.13, 0.51) \\ -2.10(0.04, 0.22) & 2.92(0.03, 0.50) \\ 3.95(0.40, 0.20) & 3.89(0.33, 0.31) \end{pmatrix}$$

and

$$\widehat{\boldsymbol{\mathcal{B}}}_{LS} = \begin{pmatrix} 2.41(0.80) & 1.81(0.83) \\ -2.00(4.00 \times 10^{-4}) & 2.99(4.00 \times 10^{-4}) \\ 0.04(4.30 \times 10^{-5}) & 0.04(4.50 \times 10^{-5}) \end{pmatrix}.$$

Note that the R and LS procedures were fooled by the bad outliers while the GR procedure was less affected by the presence of high leverage points in the data set. It is also interesting to see the effect of the outlying x-values on the conclusion of the quadratic tests of the hypothesis H_{01}, H_{02} and H_{03}. Also, we would like to test here the hypothesis

$$H_{04} : \beta_{21} = 4.$$

For this test we used the matrix $H = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ and $K = \begin{pmatrix} 1 & 0 \end{pmatrix}'$. Table 2 gives the values of the test statistics and their p-values. The R and LS procedures got very large values for their test statistics when testing H_{01}, H_{02} and H_{04}. Only tests based on the GR procedure accepted the null hypothesis H_{04}.The componentwise GR estimate is much more robust than the affine equivariant transformed-retransformed GR estimate

Table 2

Observed values and p values of the contaminated data set

Hypothesis	df	Q_R	Q_{TRR}	Q_{GR}	Q_{TRGR}	Q_{LH}
H_{01}	4	8.33×10^8	9.78×10^8	2964.70	1702333	1.95×10^9
p-value		0.000	0.000	0.000	0.000	0.000
H_{02}	2	8.31×10^8	9.77×10^8	2905.13	344064	1.91×10^9
p-value		0.000	0.000	0.000	0.000	0.000
H_{03}	1	0.72	0.72	1.78×10^{-3}	0.45	1.33
p-value		0.397	0.398	0.966	0.50	0.25
H_{04}	1	8.86×10^9	8.73×10^9	0.02	0.03	8.48×10^9
p-value		0.000	0.000	0.895	0.869	0.000

since the later was affected by the initial LS fit used to compute the transfromation matrix.

4.2 Simulation Studies

In our first simulation study a comparison is made between the performance of $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ and the other procedures in the literature. In this study we used model (1.1) with $d = 2$ and $p = 1$. That is we used the multivariate linear model

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + x_i \begin{pmatrix} \beta_{11} \\ \beta_{12} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix}. \tag{4.1}$$

The parameter matrix $\boldsymbol{\mathcal{B}}$ 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

$$(\det \boldsymbol{\Sigma})^{-1/2} h(\boldsymbol{\varepsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}). \tag{4.2}$$

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

$$h(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \frac{1}{2\pi}\exp(-\sqrt{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}\,). \tag{4.3}$$

Further, we used the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{4.4}$$

where we chose values for ρ between 0.00 and 0.95. In this study $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ 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 Mont 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 4–19, we observed that the performance of the componentwise estimators decreases as the correlation among the response variables increases. Both estimates $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ perform better in higher dimensions compared to LAD and their componentwise estimates $\widehat{\boldsymbol{\mathcal{B}}}_{R}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{GR}$. However, $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ performance is better than $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ because the matrix $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}$ is positive semi-definite. See Hettmansperger and McKean (1998). $\widehat{\boldsymbol{\mathcal{B}}}_{LS}$ performs best for multivariate normal errors while it is not robust for the heavily tailed error distributions. 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 assess the performance of the benchmark of $TDBETAS_{TRR}$ we run another simulation study over 4 situations. We used model 4.1 of our first simulation study but for the distribution of the predictors, we chose a standard normal distribution and a contaminated normal distribution with percentage of contamination 25% and a contamination variance 25. The distribution of the errors were chosen to be the bivariate normal distribution and bivariate contaminated normal distribution with percentage of contamination 25% and with a contamination variance 25. The values of ρ, sample size n and the number of simulations for each situation were kept the same as they were in the first simulation study. Table 4 shows the number of times the diagnostic $TDBETAS_{TRR}$ showed a significant difference (value greater than the benchmark). As the correlation between the response variables increases the number of times a difference in the R-fit and GR-fit detected by the benchmark gets smaller. When the x's are normal the overall difference in the fits is flagged less than 5% of the time. However, when the x's are contaminated and the values of ρ are less than 0.50 the GR-fit is flagged different from the R-fit about 40% of the time.

Table 3

$TDBETAS_{TRR}$ between $\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ and $\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ fits

x-Dist.	$\boldsymbol{y}$-Dist.	ρ							
		0.00	0.20	0.50	0.75	0.80	0.85	0.90	0.95
N	N	58	74	70	68	58	50	35	22
N	CN	138	139	134	75	82	81	80	82
N	CN	1131	1174	1095	1072	1021	962	936	823
CN	CN	1283	1182	1167	1092	1058	1045	979	827

Table 4

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate normal errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
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 5

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate normal errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
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 6

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate laplace errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
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 7

$\widehat{\mathcal{B}}_{TRGR}$ estimated relative efficiencies under bivariate laplace errors

ρ	ARE($\widehat{\mathcal{B}}_{TRGR}$,$\widehat{\mathcal{B}}_{LS}$)	ARE($\widehat{\mathcal{B}}_{TRGR}$,$\widehat{\mathcal{B}}_{LAD}$)	ARE($\widehat{\mathcal{B}}_{TRGR}$,$\widehat{\mathcal{B}}_{GR}$)
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 8

$\widehat{\mathcal{B}}_{TRR}$ estimated relative efficiencies under bivariate $t(3)$ errors

ρ	ARE($\widehat{\mathcal{B}}_{TRR}$,$\widehat{\mathcal{B}}_{LS}$)	ARE($\widehat{\mathcal{B}}_{TRR}$,$\widehat{\mathcal{B}}_{LAD}$)	ARE($\widehat{\mathcal{B}}_{TRR}$,$\widehat{\mathcal{B}}_{R}$)
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 9

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate $t(3)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
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 10

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate cauchy errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
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 11

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate cauchy errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
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 12

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate $CN(25, .25)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
0.00	2.560	0.996	0.963
0.20	2.581	0.994	0.972
0.50	2.618	1.092	1.055
0.75	2.592	1.278	1.195
0.80	2.644	1.349	1.226
0.85	2.602	1.492	1.288
0.90	2.610	1.724	1.419
0.95	2.622	2.010	1.601

Table 13

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate $CN(25, .25)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
0.00	2.513	0.978	0.962
0.20	2.522	0.971	0.978
0.50	2.517	1.051	1.042
0.75	2.556	1.261	1.182
0.80	2.578	1.315	1.230
0.85	2.556	1.465	1.286
0.90	2.528	1.670	1.418
0.95	2.562	1.965	1.608

Table 14

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate $CN(25, .15)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
0.00	2.395	1.090	0.982
0.20	2.335	1.094	0.987
0.50	2.354	1.175	1.030
0.75	2.370	1.399	1.161
0.80	2.363	1.453	1.201
0.85	2.336	1.552	1.252
0.90	2.315	1.795	1.356
0.95	2.383	2.000	1.535

Table 15

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate $CN(25, .15)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
0.00	2.323	1.057	0.982
0.20	2.283	1.069	0.986
0.50	2.283	1.140	1.025
0.75	2.263	1.336	1.158
0.80	2.275	1.399	1.193
0.85	2.251	1.495	1.247
0.90	2.215	1.717	1.356
0.95	2.287	1.920	1.499

Table 16

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate $CN(64, .25)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
0.00	5.159	0.898	0.933
0.20	5.151	0.882	0.950
0.50	4.913	0.932	1.002
0.75	5.166	1.182	1.181
0.80	5.150	1.273	1.227
0.85	5.213	1.388	1.310
0.90	5.219	1.471	1.410
0.95	5.291	1.804	1.612

Table 17

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate $CN(64, .25)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
0.00	5.056	0.880	0.943
0.20	5.081	0.870	0.953
0.50	4.913	0.932	0.998
0.75	5.060	1.158	1.172
0.80	5.056	1.250	1.216
0.85	5.094	1.356	1.312
0.90	5.099	1.438	1.385
0.95	5.253	1.791	1.561

Table 18

$\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$ estimated relative efficiencies under bivariate $CN(64, .15)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{R}$)
0.00	4.740	1.020	0.955
0.20	4.803	1.020	0.963
0.50	4.940	1.117	1.027
0.75	4.962	1.306	1.137
0.80	5.075	1.415	1.219
0.85	4.997	1.478	1.251
0.90	4.882	1.672	1.354
0.95	4.875	1.950	1.523

Table 19

$\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$ estimated relative efficiencies under bivariate $CN(64, .15)$ errors

ρ	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LS}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{LAD}$)	ARE($\widehat{\boldsymbol{\mathcal{B}}}_{TRGR}$,$\widehat{\boldsymbol{\mathcal{B}}}_{GR}$)
0.00	4.685	1.008	0.959
0.20	4.686	0.995	0.969
0.50	4.820	1.090	1.030
0.75	4.742	1.248	1.125
0.80	4.854	1.353	1.194
0.85	4.867	1.439	1.256
0.90	4.639	1.589	1.342
0.95	4.686	1.874	1.511

CHAPTER V

SUMMARY, CONCLUSION, AND FUTURE RESEARCH

5.1 Summary and Conclusion

In this study the multivariate linear model was investigated. New estimates of the regression coefficient matrix were introduced. These estimates were proved to be equivariant under nonsingular transformation of the response variable. Asymptotic properties of the estimates were obtained under some regularity conditions. Comparisons between these estimates and the componentwise estimates, including the least absolute deviations estimate, were based on finite sample efficiencies. Further, quadratic tests for the general linear hypothesis were established. The test statistics have asymptotic chi squared distributions.

The simulation study presented shows that the new estimates that are based on the transformation-retransformation technique performs better than the componentwise estimates as the correlation between the response variables becomes stronger. The transformed-retransformed GR estimate is less efficient than the transformed-retransformed R estimate. However, the transformed-retransformed GR fits are the best for data sets that contain clusters of outlying values in factor space. Both estimates are easy to compute and are based on a transformation matrix which can be computed by a simple and fast algorithm that utilizes the whole data set.

5.2 Future Research

In the future, I would like to investigate more diagnostic techniques for the multivariate R and GR fits. Also, simulation experiments need to be conducted to study the performance of the quadratic tests based on these fits. Further, I would like to apply the transformation-retransformation technique to an HBR componentwise estimate of the matrix of regression coefficients of the multivariate linear model.

REFERENCES

Armstrong, R. D. & Kung, D. S. (1978). "Least Absolute Value Estimates for a Simple Linear Regression Problem". *Applied Statistics, 27*, 363–366. AS132.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* New York: John Wiley & Sons Inc.

Bai, Z. D., Chen, N. R., Miao, B. Q., & Rao, C. R. (1990). "Asymptotic Theory of Least Distances Estimate in Multivaraite Linear Models". *Statistics, 21*(4), 503–519.

Brown, B. M. (1983). "Statistical Use of Spatial Median". *Journal of the Royal Statistical Society. Series B, 45*, 25–30.

Chakraborty, B. (1996). "On Multivariate Median Regression". Technical Report 9/96, Indian Statistical Institute, Theoretical Statistics & Mathematical Unit, Calcutta.

Chakraborty, B. & Chauduri, P. (1997). "On Multivariate Rank Regression". *L_1-Statistical Procedures and Related Topics, 31*, 399–414. IMS Lecture Notes — Monograph Series.

Davis, J. B. & McKean, J. W. (1993). "Rank-Based Methods for Multivariate Linear Models". *Journal of the American Statistical Association, 88*(421), 245–251.

Golub, G. H. & Van Loan, C. F. (1983). *Matrix Computations* (3rd ed.). Baltimore, Maryland: John Hopkins.

Graybill, F. A. (1983). *Matrices with Applications in Statistics* (2nd ed.). Belmont, CA: Wadsworth.

Hájek, J. & Šidák, Z. (1967). *Theory of Rank Tests.* New York: Academic Press.

Haldane, J. B. S. (1948). "Note on the Median of a Multivariate Distribution". *Biometrika*, *35*, 414–415.

Hettmansperger, T. P. & McKean, J. W. (1983). "a Geometric Interpretations of Inferences Based on Ranks in the Linear Model". *Journal of the American Statistical Association*, *78*, 885–893.

Hettmansperger, T. P. & McKean, J. W. (1998). *Robust Nonparametric Statistical Methods.* London: Edward Arnold.

Jaeckel, L. A. (1972). "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals". *The Annals of Mathematical Statistics*, *43*, 1449–1458.

Johnson, R. A. & Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis.* Upper Saddle River: Prentice Hall.

Jurečková, J. (1971). "Nonparametric Estimate of Regression Coefficients". *The Annals of Mathematical Statistics*, *42*, 1328–1338.

Koenker, R. & Bassett, G. (1978). "Regression Quantiles". *Econometric*, *46*, 33–50.

Lehmann, E. L. (1975). "Nonparametric Statistical Methods Based on Ranks". *Nonparametrics.*

Maronna, R. A. & Morgethaler, S. (1986). "Robust Regression through Robust Covariances". *Communications in Statistics*, *15*, 1347–1366. Theory Meth.

McKean, J. W., Naranjo, J. D., & Sheather, S. J. (1996). "Diagnostics to Detect Differences in Robust Fits of Linear Models". *Computational Statistics*, *11*, 223–243.

McKean, J. W. & Schrader, R. M. (1980). "The Geometry of Robust Procedures in Linear Models". *Journal of the Royal Statistical Society. Series B. Methodological*, *42*(3), 366–371.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: John Wiley & Sons Inc.

Naranjo, J. D. & Hettmansperger, T. P. (1994). "Bounded Influence Rank Regression". *Journal of the Royal Statistical Society. Series B. Methodological*, *56*(1), 209–220.

Ollila, E., Hettmansperger, P., & Oja, H. (2002). "Estimates of Regression Coefficients Based on Sign Covariance Matrix". *Journal of the Royal Statistical Society*, *64*, 447–466.

Peters, D. & Randles, R. H. (1990). "A Multivariate signed-rank Test for the One-sampled Location Problem"'. *Journal of the American Statistical Association*, *85*, 552–557.

Rho, C. R. (1988). "Methodology based on l_1-norm in statistical inference". *Sankhya*, *50*, 289–313.

Rousseeuw, P. & Van Driessen, K. (1999). "A Fast Algorithm for the Minimum Covariance Determinant Estimator". *Technometrics*, *41*, 212–223.

Sievers, G. L. (1983). "A Weighted Dispersion Function for Estimation in Linear Models". *Communications in Statistics*, *12*(10), 1161–1179. Theory and Methods.

Tyler, D. E. (1987). "A Distribution-Free m-Estimator of Multivariate Scatter". *The Annals of Statistics*, *15*, 234–251.