UvA ECONOMETRICS

Discussion Paper: 2004/02

TR multivariate conditional median estimation

Jan G. de Gooijer and Ali Gannoun

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

TR Multivariate Conditional Median Estimation

Jan G. De Gooijer[1] and Ali Gannoun[2]

[1] Department of Quantitative Economics
University of Amsterdam
Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
Telephone: +31–20–525 4244; Fax: +31–20–525 4349
e-mail: j.g.degooijer@uva.nl

[2] Laboratoire de Probabilités et Statistique
Université Montpellier II
Place Eugène Bataillon, 34095 Montpellier Cédex 5, France
Telephone: +33-4-67-14-3578; Fax: +33-4-67-14-4974
e-mail: gannoun@math.univ-montp2.fr

Abstract

An affine equivariant version of the nonparametric spatial conditional median (SCM) is constructed, using an adaptive transformation-retransformation (TR) procedure. The relative performance of SCM estimates, computed with and without applying the TR–procedure, are compared through simulation. Also included is the vector of coordinate conditional, kernel-based, medians (VCCMs). The methodology is illustrated via an empirical data set. It is shown that the TR–SCM estimator is more efficient than the SCM estimator, even when the amount of contamination in the data set is as high as 25%. The TR–VCCM- and VCCM estimators lack efficiency, and consequently should not be used in practice.

Key Words: Spatial conditional median; kernel; retransformation; robust; transformation.

1 Introduction

Let $\{(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)\}$ be independent replicates of a random vector $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^p \times \mathbb{R}^s$ where $p \geqslant 1, s \geqslant 2$, and $n > p + s$. "Classical" nonparametric regression analysis focuses on the problem of estimating the conditional mean function of the s-variate response variables $\boldsymbol{Y}$ given values $\boldsymbol{X} = \boldsymbol{x}$ of the explanatory variables. However, it is well-known that outliers can have an arbitrarily large effect on estimates of the conditional mean. The conditional median is an obvious alternative since it is quite resistant to outliers. Interestingly, as compared to the many papers published on the multivariate *unconditional* median (see, e.g., Small (1990) for a survey), relative little attention has been paid to the estimation of multivariate *conditional* medians. One approach is to generalize the notion of univariate conditional median to the multivariate case in the same manner as some authors (e.g. Bickel (1964) and Barnett (1976)) consider the vector of unconditional medians formed by the coordinate univariate medians. It is known that the resulting estimator, called vector coordinate conditional median (VCCM), satisfies two fundamental properties, namely: (i) it has a breakdown point of 50%; and (ii) under suitable regularity conditions it is $\sqrt{n}$-consistent and asymptotically normally distributed. But the VCCM, like its unconditional counter-part, suffers from obscurity with respect to a third important property, namely (iii) the estimator is not affine equivariant. Thus, when the scales of the data vectors are measured in different units, or when they have different degrees of statistical variation, the resulting VCCM estimates cannot be easily interpreted.

An alternative notion of multivariate conditional median has been proposed by Berlinet, Cadre and Gannoun (2001a, b). These authors generalized a notion of spatial median originally given by Weber (1909) and studied by, among others, Haldane (1948), and Kemperman (1987), to the conditional case; see Subsection 2.1 for some details. Following Chaudhuri (1992) it can be shown that the resulting estimator, called spatial conditional median (SCM), is equivariant under orthogonal transformations, but that it is nonequivariant under arbitrary affine transformations. A solution to this problem follows from the work of Chakraborty and Chaudhuri (1996). They use a data-driven transformation and retransformation (TR) procedure for converting a nonequivariant multivariate *unconditional* median into an equivariant estimate of multivariate location. Within this context, Gannoun, Saracco, Yuan and Bonney (2003) introduced a TR–SCM estimator and studied its asymptotic properties. Our objective is to explore the efficiency of VCCM and SCM estimates, obtained with- and without the use of the TR–procedure, in moderately sized sample situations through real and simulated data.

The plan of the paper is as follows. In Section 2, we introduce the SCM estimator. Also, we consider briefly the VCCM estimator. In Section 3, we describe the TR–procedure when applied to multivariate conditional medians. Section 4 contains results of the simulation study. To motivate the need of the TR–procedure a real-data example is given in Section 5.

2 Multivariate Conditional Median

2.1 Spatial conditional median (SCM)

Let $\| \cdot \|$ denote a strictly convex norm on $\mathbb{R}^s$, i.e. $\| \alpha + \beta \| < \| \alpha \| + \| \beta \|$ whenever α and β are not proportional. For a fixed $\boldsymbol{x} \in \mathbb{R}^p$, denote by $Q(\cdot|\boldsymbol{x})$ the probability measure of $\boldsymbol{Y}$ conditionally on $\boldsymbol{X} = \boldsymbol{x}$, and assume that the support of $Q(\cdot|\boldsymbol{x})$ is not included in a straight line. Given this set-up, Berlinet *et al.* (2001a, b) define the SCM of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ as the vector $\boldsymbol{\theta}(\boldsymbol{x})$ which assumes the infimum

$$\varphi(\boldsymbol{\theta}(\boldsymbol{x})) = \inf_{\boldsymbol{\theta} \in \mathbb{R}^s} \varphi(\boldsymbol{\theta}, \boldsymbol{x}) \tag{1}$$

where for $\boldsymbol{\theta} \in \mathbb{R}^s$,

$$\begin{aligned} \varphi(\boldsymbol{\theta}, \boldsymbol{x}) &= E(\| \boldsymbol{Y} - \boldsymbol{\theta} \| - \| \boldsymbol{Y} \| \,|\boldsymbol{X} = \boldsymbol{x}) \\ &= \int_{\mathbb{R}^s} \Big(\| \boldsymbol{y} - \boldsymbol{\theta} \| - \| \boldsymbol{y} \| \Big) Q(d\boldsymbol{y}|\boldsymbol{x}) \end{aligned}$$

exists and is unique; see, e.g., Kemperman (1987).

A consistent nonparametric estimator of $\boldsymbol{\theta}(\boldsymbol{x})$ can be introduced as follows. First define $F_n(\cdot|\boldsymbol{x})$, a nonparametric estimate of $F(\cdot|\boldsymbol{x})$ the conditional distribution function of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$, by

$$F_n(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{n} \mathbf{1}_{(\boldsymbol{Y}_i \leqslant \boldsymbol{y})} K\{(\boldsymbol{x} - \boldsymbol{X}_i)/h_n\} \Big/ \sum_{i=1}^{n} K\{(\boldsymbol{x} - \boldsymbol{X}_i)/h_n\}, \quad \boldsymbol{y} \in \mathbb{R}^s, \tag{2}$$

where $\mathbf{1}_{(\boldsymbol{Y}_i \leqslant \boldsymbol{y})} = \mathbf{1}_{(Y_{i,1} \leqslant y_1)} \times \ldots \times \mathbf{1}_{(Y_{i,s} \leqslant y_s)}$, if $\boldsymbol{y} = (y_1, \ldots, y_s) \in \mathbb{R}^s$, with $\mathbf{1}_A$ the indicator function for set A. $K(\cdot)$ is a strictly positive density function on $\mathbb{R}^p$ (the kernel), and h_n (the bandwidth) is a sequence of real positive numbers such that as $n \to \infty$, $h_n \to 0$ and $nh_n^p \to \infty$. Let $Q_n(\cdot|\boldsymbol{x})$ be the estimate of $Q(\cdot|\boldsymbol{x})$, defined for any Borel set $V \subset \mathbb{R}^s$ by $Q_n(V|\boldsymbol{x}) = \int_V F_n(d\boldsymbol{y}|\boldsymbol{x})$. Then, for $\boldsymbol{\theta} \in \mathbb{R}^s$, the natural estimate of $\varphi(\boldsymbol{\theta}, \boldsymbol{x})$, denoted by $\varphi_n(\boldsymbol{\theta}, \boldsymbol{x})$, can be defined as

$$\begin{aligned} \varphi_n(\boldsymbol{\theta}, \boldsymbol{x}) &= \int_{\mathbb{R}^s} \Big(\| \boldsymbol{y} - \boldsymbol{\theta} \| - \| \boldsymbol{y} \| \Big) Q_n(d\boldsymbol{y}|\boldsymbol{x}) \\ &= \sum_{i=1}^{n} \Big(\| \boldsymbol{Y}_i - \boldsymbol{\theta} \| - \| \boldsymbol{Y}_i \| \Big) K\{(\boldsymbol{x} - \boldsymbol{X}_i)/h_n\} \Big/ \sum_{i=1}^{n} K\{(\boldsymbol{x} - \boldsymbol{X}_i)/h_n\}. \end{aligned}$$

Finally, if we minimize $\varphi_n(\boldsymbol{\theta}, \boldsymbol{x})$ instead of $\varphi(\boldsymbol{\theta}, \boldsymbol{x})$, the minimizer is an estimator of $\boldsymbol{\theta}(\boldsymbol{x})$. Denoted by $\boldsymbol{\theta}_n(\boldsymbol{x})$, such an estimator is given by

$$\boldsymbol{\theta}_n^*(\boldsymbol{x}) = \underset{\boldsymbol{\theta}\in \mathrm{I\!R}^s}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(\parallel \mathbf{Y}_i - \boldsymbol{\theta} \parallel - \parallel \mathbf{Y}_i \parallel \Big) K\{(\boldsymbol{x} - \mathbf{X}_i)/h_n\}. \tag{3}$$

In the rest of the paper it is assumed that $E(\parallel \boldsymbol{Y} \parallel) < \infty$. Then (3) becomes

$$\boldsymbol{\theta}_n(\boldsymbol{x}) = \underset{\boldsymbol{\theta}\in \mathrm{I\!R}^s}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(\parallel \mathbf{Y}_i - \boldsymbol{\theta} \parallel\parallel \Big) K\{(\boldsymbol{x} - \mathbf{X}_i)/h_n\}. \tag{4}$$

Berlinet *et al.* (2001a, b) studied the asymptotic properties of (4). It is shown that if the norm used in the definitions of $\varphi(\cdot|\boldsymbol{x})$ and $\varphi_n(\cdot|\boldsymbol{x})$ is the Euclidean norm, $\boldsymbol{\theta}_n(\boldsymbol{x})$ exists and is unique. These authors also show that $\boldsymbol{\theta}_n(\boldsymbol{x})$ converges to $\boldsymbol{\theta}(\boldsymbol{x})$ on a compact subset $\mathcal{C} \subset \mathrm{I\!R}^p$. Cadre and Gannoun (2000) proved the asymptotic normality of (4) in the case the data are independently and identically distributed (i.i.d.).

2.2 Vector coordinate conditional median (VCCM)

Let $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_s)^T$ denote a $p \times 1$ and an $s \times 1$ column vector, respectively. Then, the vector of coordinate conditional medians (VCCMs) is defined as

$$\boldsymbol{\mu}(\boldsymbol{x}) = (\mu_1(\boldsymbol{x}), \ldots, \mu_s(\boldsymbol{x}))^T,$$

where $\mu_j(\boldsymbol{x}) = F^{-1}(1/2|\boldsymbol{x})$ with $F(\cdot|\boldsymbol{x})$ the conditional distribution function of Y_j $(j = 1, \ldots, s)$ given $\boldsymbol{X} = \boldsymbol{x}$, i.e. the one-dimensional conditional median. An estimator $\hat{\mu}_j(\boldsymbol{x})$ of $\mu_j(\boldsymbol{x})$ can be obtained by estimating the conditional marginal median obtained from the set of observations $(\boldsymbol{X}_i, Y_{i,j})_{i=1,n}$. Note that, for ease of notation, we omitted the subscript n from $\hat{\mu}_j(\boldsymbol{x})$. Then the VCCM estimator is given by $\hat{\boldsymbol{\mu}}_n(\boldsymbol{x}) = (\hat{\mu}_1(\boldsymbol{x}), \ldots, \hat{\mu}_s(\boldsymbol{x}))^T$. In a nonparametric setting several methods exist for estimating conditional distributions. Here we will adopt the local constant (Nadaraya-Watson) kernel smoother of $F(\cdot|\boldsymbol{x})$ which is given by

$$F_n(y_j|\boldsymbol{x}) = \sum_{i=1}^{n} \mathbf{1}_{(Y_{i,j} \leqslant y_j)} K\{(\boldsymbol{x} - \mathbf{X}_i)/h_n\} \Big/ \sum_{i=1}^{n} K\{(\boldsymbol{x} - \mathbf{X}_i)/h_n\}, \quad y_j \in \mathrm{I\!R}. \tag{5}$$

The motivation for using (5) stems from the fact that asymptotic properties of the resulting conditional median $\hat{\mu}_j(\boldsymbol{x})$ are well-established. For instance, under mixing assumptions, the convergence of nonparametric estimates of the conditional median $\mu_j(\boldsymbol{x})$ was proved by Gannoun (1990) and Boente and Fraiman (1995). Sufficient conditions for the asymptotic normality of convergent estimates of $\mu_j(\boldsymbol{x})$, using (5) as an estimate of $F(\cdot|\boldsymbol{x})$ and irrespective of data dependence, are given by Berlinet, Gannoun and Matzner-Løber (2001).

3 TR Multivariate Conditional Median

3.1 TR–spatial conditional median (TR–SCM)

The basic principle of the data-driven TR–procedure goes as follows. Suppose that $(\boldsymbol{Y}_i)_{i=1,n} \in \mathrm{I\!R}^s$, and let S_n be the collection of all subsets of size $s+1$ of $\{1, 2, \ldots, n\}$. For a fixed $\alpha = \{i_0, i_1, \ldots, i_s\} \in S_n$, consider the $s \times s$ matrix $\boldsymbol{Y}(\alpha)$ defined by $\boldsymbol{Y}(\alpha) = [\boldsymbol{Y}_{i_1} - \boldsymbol{Y}_{i_0}, \ldots, \boldsymbol{Y}_{i_s} - \boldsymbol{Y}_{i_0}]$, where $\boldsymbol{Y}_{i_0}$ determines the origin and the lines joining that origin to the remaining s data points $\boldsymbol{Y}_{i_1}, \ldots, \boldsymbol{Y}_{i_s}$. The matrix $\boldsymbol{Y}(\alpha)$ transforms all observations $\boldsymbol{Y}_j$ $(1 \leqslant j \leqslant n, j \notin \alpha)$ into terms of the new coordinate system as $\boldsymbol{Z}_j^{(\alpha)} = \{\boldsymbol{Y}(\alpha)\}^{-1}\boldsymbol{Y}_j$. Since the vectors $\boldsymbol{Y}_i$ happen to be i.i.d. with a common probability distribution, which is assumed to be absolutely continuous w.r.t. the Lebesgue measure on $\mathrm{I\!R}^s$, the invertibility of the matrix $\boldsymbol{Y}(\alpha)$ is ensured with probability 1 for any choice of $\alpha \in S_n$. Clearly, if it is assumed that the SCM is a linear function of the regressors, the TR-procedure should be applied on both $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Now, given the new coordinate system, one can compute quantities like the VCCM and the SCM. For ease of exposition consider the latter estimator, i.e. given (4) the transformed SCM is given by

$$\boldsymbol{\theta}_n^{(\alpha)}(\boldsymbol{x}) = \underset{\boldsymbol{\theta} \in \mathrm{I\!R}^s}{\operatorname{argmin}} \sum_{j \notin \alpha}^{n} \Big(\parallel \boldsymbol{Z}_j^{(\alpha)} - \boldsymbol{\theta} \parallel \Big) K\{(\boldsymbol{x} - \boldsymbol{X}_j)/h_n\}. \tag{6}$$

Then, transforming all observations back in the original coordinate system, the TR–SCM estimator is given by

$$\hat{\boldsymbol{\theta}}_n^{(\alpha)}(\boldsymbol{x}) = \{\boldsymbol{Y}(\alpha)\}\boldsymbol{\theta}_n^{(\alpha)}(\boldsymbol{x}). \tag{7}$$

With an obvious modification in notation, the TR–procedure can also be applied to the VCCM estimator $\hat{\boldsymbol{\mu}}_n(\boldsymbol{x})$. In Section 4, the resulting TR–VCCM will be denoted by $\hat{\boldsymbol{\mu}}_n^{(\alpha)}(\boldsymbol{x})$. For i.i.d. observations, and a fixed $\alpha \notin S_n$, Gannoun *et al.* (2003) show that $\hat{\boldsymbol{\theta}}_n^{(\alpha)}(\boldsymbol{x})$ exists, unless the support of $\{\boldsymbol{Y}_j,\ j \notin \alpha\}$ given $\boldsymbol{X} = \boldsymbol{x}$ is included in a straight line. Moreover, these authors obtain the following results: (i) the TR–SCM estimator is affine equivariant; (ii) it is a consistent estimator of $\boldsymbol{\theta}(\boldsymbol{x})$ if the common conditional distribution of $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i = \boldsymbol{x}$ has an elliptically symmetric density function; (iii) $\hat{\boldsymbol{\theta}}_n^{(\alpha)}(\boldsymbol{x}) \to \boldsymbol{\theta}(\boldsymbol{x})$ a.s. (pointwise convergence); and (iv) $\hat{\boldsymbol{\theta}}_n^{(\alpha)}(\boldsymbol{x})$ is asymptotically normally distributed.

3.2 Selection of $\boldsymbol{Y}(\alpha)$ and h_n

Different choices of the subsets of indices α may result in different estimates of $\hat{\boldsymbol{\theta}}_n^{(\alpha)}(\boldsymbol{x})$. Let $\boldsymbol{\Sigma}_{\mathbf{YY}}$ denote the $s \times s$ positive definite variance-covariance matrix of $\boldsymbol{Y}$. In the case of i.i.d. observations, and using the asymptotic normality of the unconditional spatial median, Chakraborty and Chaudhury (1966) suggest to choose α such that $\{\boldsymbol{Y}(\alpha)\}^T \boldsymbol{\Sigma}_{\mathbf{YY}}^{-1} \boldsymbol{Y}(\alpha) \approx \lambda \boldsymbol{I}$, where $\lambda > 0$ is a constant and $\boldsymbol{I}$ is the $s \times s$ identity matrix. In practice, $\boldsymbol{\Sigma}_{\mathbf{YY}}$ is replaced by an affine equivariant estimate (up to a scalar multiple), say $\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}$, obtained from the data. For instance, in a linear regression context, $\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}$ may be taken as the variance-covariance matrix of the regression residuals. Chakraborty (2003, Appendix A) provides an adaptive algorithm for computing the TR matrix $\boldsymbol{Y}(\hat{\alpha})$ as an estimate of $\boldsymbol{Y}(\alpha)$.

Note that, when $n \gg s$, the search for an optimal transformation matrix can become computationally intensive. Moreover, Chakraborty's adaptive algorithm cannot be applied to non-i.i.d. data. A general method of selecting $\boldsymbol{Y}(\alpha)$ can be based on a training sample of size $n_0 \ll n$. Assume that $\boldsymbol{Y}_i$ is k-Markovian. Then, let $\boldsymbol{Y}_0$ be an alternative matrix of $\boldsymbol{Y}(\alpha)$ defined by

$$\begin{aligned}\boldsymbol{Y}_0 \;=\; & \frac{1}{(k+1)m} \sum_{u=0}^{m-1} \sum_{v=0}^{k} \Big(\boldsymbol{Y}_{u(s+1)(k+1)+v+k+2} - \boldsymbol{Y}_{u(s+1)(k+1)+v+1}, \boldsymbol{Y}_{u(s+1)(k+1)+v+2k+3} - \\ & \boldsymbol{Y}_{u(s+1)(k+1)+v+k+2}, \ldots, \boldsymbol{Y}_{u(s+1)(k+1)+v+sk+s+1} - \boldsymbol{Y}_{u(s+1)(k+1)+v+(s-1)k+s} \Big), \qquad (8)\end{aligned}$$

where $m = [n_0/(s+1)(k+1)]$ with $[\cdot]$ denoting the integer part. Thus, $\boldsymbol{Y}_0$ is an average of $(k+1)m$ i.i.d. random vectors, each with the same distribution as $\boldsymbol{Y}(\alpha)$ for all $\alpha \in S_n$. This implies that $\boldsymbol{Y}_0$ has approximately the same distribution as $\boldsymbol{Y}(\alpha)$. Obviously, for i.i.d. data, k should be set equal to zero in (8).

Also, the choice of the bandwidth h_n is a matter of concern. Various existing bandwidth selection techniques can be adapted for selecting h_n. The optimal bandwidth depends on $\boldsymbol{x}$, i.e. the amount of smoothing required to estimate different parts of $F(\cdot|\boldsymbol{x})$ may differ from what is optimal to estimate the whole conditional distribution function. Therefore a unique bandwidth is chosen for each multivariate conditional median. For simplicity, we employ a special case of the practical rule-of-thumb given by Yu and Jones (1998) for bandwidth selection in quantile regression estimation. In particular, first a primary bandwidth h_{mean}, suitable for conditional mean estimation, is selected. Then, it is adjusted according to the formula $h_{mean}(\pi/2)^{1/(p+4)}$.

4 Simulations

4.1 Design

All simulation results will be based on 500 replications. For the kernel $K(\cdot)$ we choose the bivariate standard normal density. The sample sizes are fixed at $n = 100$, and 200. For visual purpose, we restrict the choice of the dimensions to the case $p = 1$ and $s = 2$. Observations $(\boldsymbol{Y}_i, X_i)_{i=1,n}$ will be generated from the following distributions:

(a) Two centrally symmetric trivariate normal distributions $N_3(\mathbf{0}, \boldsymbol{\Sigma_i})$ $(i = 1, 2)$ with variance-covariance matrices

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & \sigma_{\mathrm{Y}_1,\mathrm{Y}_2} & 0.4 \\ \sigma_{\mathrm{Y}_1,\mathrm{Y}_2} & 1 & 0.2 \\ 0.4 & 0.2 & 1 \end{pmatrix}, \ \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & \sigma_{\mathrm{Y}_1,\mathrm{Y}_2} & 0.5 \\ \sigma_{\mathrm{Y}_1,\mathrm{Y}_2} & 1 & 0.1 \\ 0.5 & 0.1 & 1 \end{pmatrix}, \tag{9}$$

and with $\sigma_{\mathrm{Y}_1,\mathrm{Y}_2} = -0.8, -0.7, \ldots, 0.8$;

(b) Two $\beta\%$ contaminated asymmetric trivariate normal distributions $\mathrm{AN}_i = (1-\beta)N_3(\mathbf{0}, \boldsymbol{\Sigma}_i) + \beta N_3(\mathbf{2}, \boldsymbol{I})$, with $\mathbf{2} = (2, 2, 2)^T$, and $\beta = 0.10, 0.25$. The variance-covariance matrices $\boldsymbol{\Sigma}_i$ $(i = 1, 2)$ are given by (9) with $\sigma_{\mathrm{Y}_1,\mathrm{Y}_2} = -0.8, -0.4, \ldots, 0.8$.

Let $\boldsymbol{T}_n(x)$ be a bivariate conditional median estimator of $\boldsymbol{\theta}(x)$, where $\boldsymbol{T}_n(x)$ is either $\hat{\boldsymbol{\theta}}_n^{(\alpha)}(x)$, $\hat{\boldsymbol{\theta}}_n(x)$, $\hat{\boldsymbol{\mu}}_n^{(\alpha)}(x)$, or $\hat{\boldsymbol{\mu}}_n(x)$. Then, the variance-covariance matrix of $\boldsymbol{T}_n(x)$ is given by $\boldsymbol{\Sigma}(\boldsymbol{T}_n(x)) = (E[\boldsymbol{T}_n(x)] - \boldsymbol{\theta}(x))(E[\boldsymbol{T}_n(x)] - \boldsymbol{\theta}(x))^T$. It is well-known that for multivariate normal distributions, the conditional mean coincides with the conditional median. Thus, under (a), $\boldsymbol{\theta}(x)$ can be readily expressed in terms of the bivariate vectors $x \times (0.4, 0.2)^T$ and $x \times (0.5, 0.1)^T$, respectively. Similar expressions for the AN_i $(i = 1, 2)$ distributions cannot be established. However, if we adopt the point of view that contamination is an uncontrollable disturbance in the estimation of the conditional medians, it is natural to use the above expressions for $\boldsymbol{\theta}(x)$ also in the case the observations are generated by the AN_i distributions specified under (b). This implies that in both cases (a) and (b), an estimate of $\boldsymbol{\Sigma}(\boldsymbol{T}_n(x))$, will be computed as

$$\boldsymbol{S}(\boldsymbol{T}_n(x)) = \frac{1}{500} \sum_{j=1}^{500} (\boldsymbol{T}_n^j(x) - \boldsymbol{\theta}(x))(\boldsymbol{T}_n^j(x) - \boldsymbol{\theta}(x))^T, \tag{10}$$

where $\boldsymbol{T}_n^j(x)$ denotes the value of $\boldsymbol{T}_n(x)$ for the jth replication $(j = 1, \ldots, 500)$.

In the final evaluation, the performance of the conditional median estimators will be assessed by the following three relative errors (REs), as functions of (10):

$$\mathrm{RE}_n^{(1)}(x) = \frac{[tr\{\boldsymbol{S}(\hat{\boldsymbol{\mu}}_n^{(\alpha)}(x))\}]^{1/2}}{[tr\{\boldsymbol{S}(\hat{\boldsymbol{\mu}}_n(x))\}]^{1/2}}, \ \mathrm{RE}_n^{(2)}(x) = \frac{[tr\{\boldsymbol{S}(\hat{\boldsymbol{\theta}}_n^{(\alpha)}(x))\}]^{1/2}}{[tr\{\boldsymbol{S}(\hat{\boldsymbol{\theta}}_n(x))\}]^{1/2}}, \ \mathrm{RE}_n^{(3)}(x) = \frac{[tr\{\boldsymbol{S}(\hat{\boldsymbol{\mu}}_n(x))\}]^{1/2}}{[tr\{\boldsymbol{S}(\hat{\boldsymbol{\theta}}_n(x))\}]^{1/2}}.$$

Values of $\mathrm{RE}_n^{(1)}(x)$ ($\mathrm{RE}_n^{(2)}(x)$) less than one indicate that TR–VCCMs (TR–SCMs) are more efficient than VCCMs (SCMs). Similarly, values of $\mathrm{RE}_n^{(3)}(x)$ less than one provide an indication that VCCMs are more efficient than SCMs.

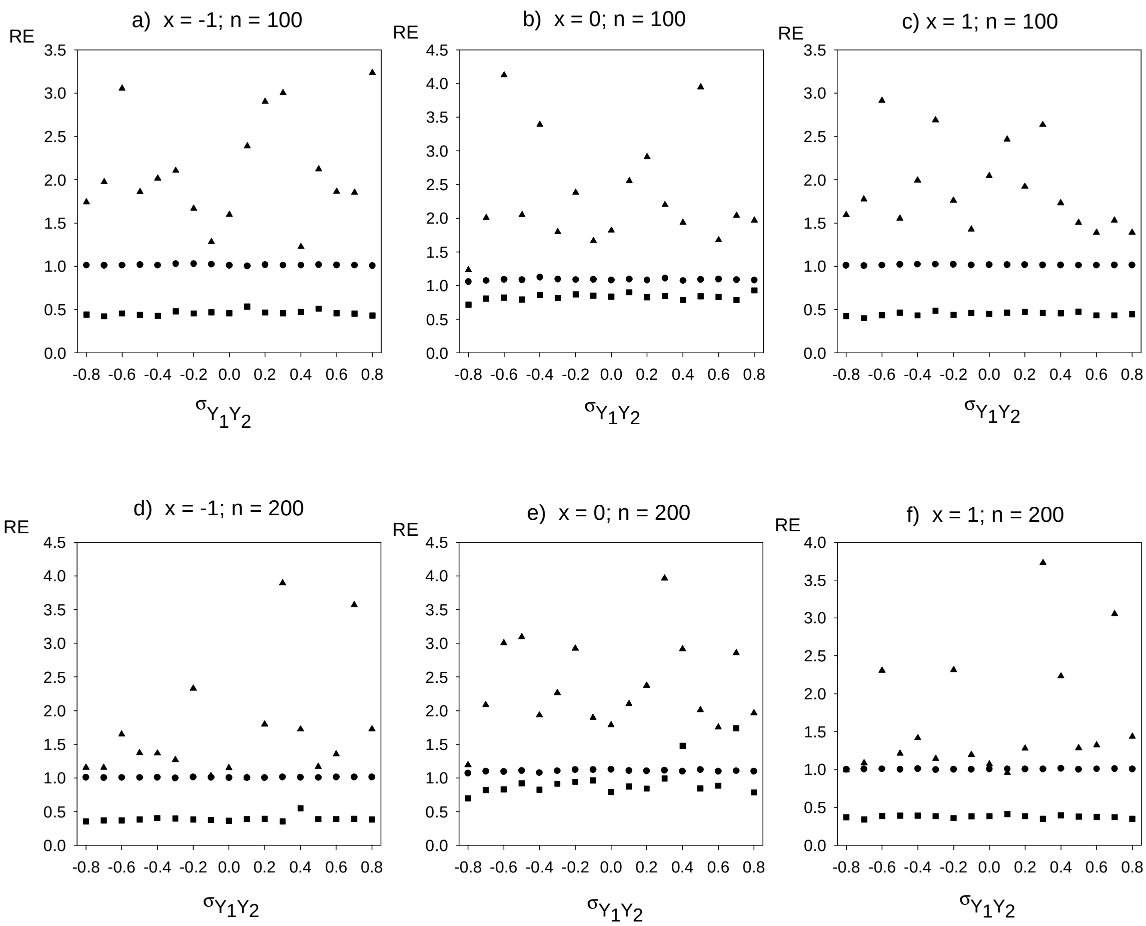

Figure 1: *RE-values for the centrally symmetric* $N_3(\mathbf{0}, \mathbf{\Sigma}_1)$ *distribution; black triangles are* $RE_n^{(1)}(x)$*-values, black squares are* $RE_n^{(2)}(x)$*-values, and black dots are* $RE_n^{(3)}(x)$*-values; 500 replications.*

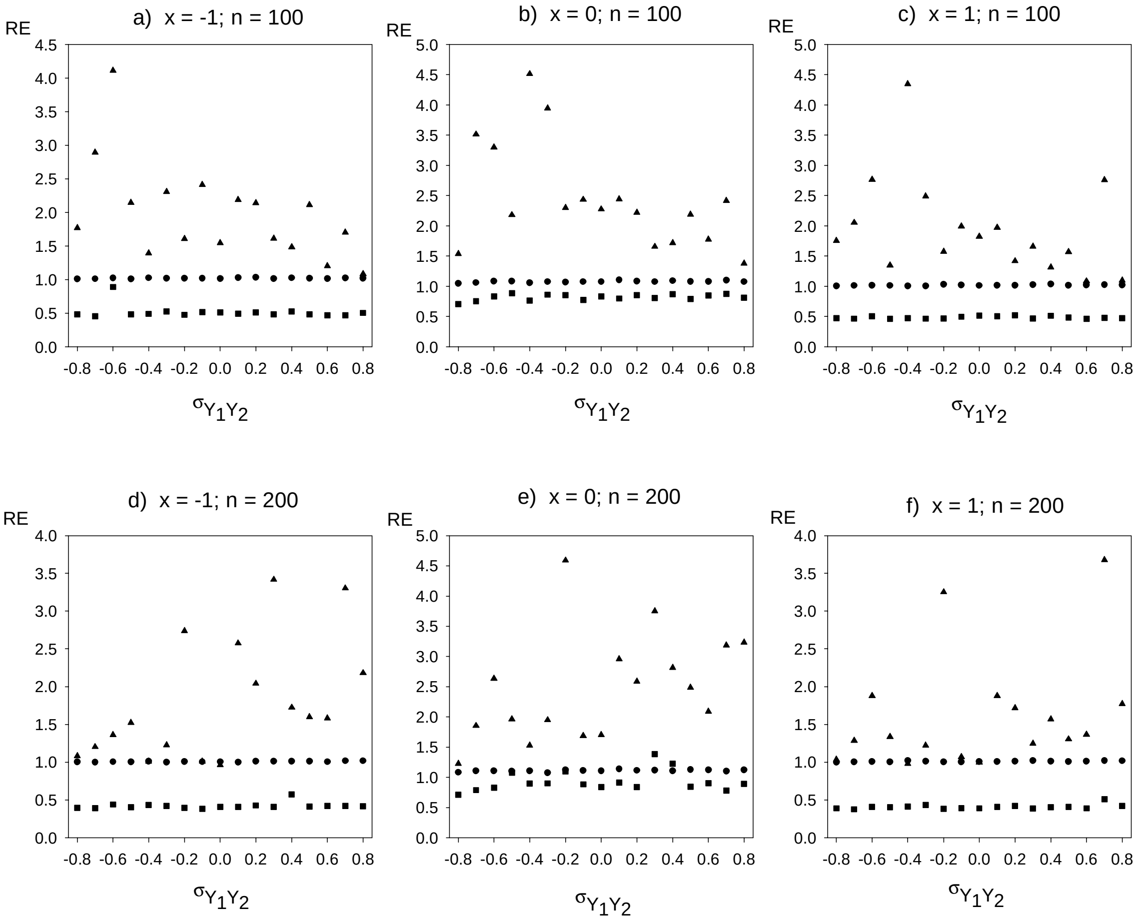

Figure 2: *RE-values for the centrally symmetric* $N_3(\mathbf{0}, \mathbf{\Sigma}_2)$ *distribution; black triangles are* $RE_n^{(1)}(x)$*-values, black squares are* $RE_n^{(2)}(x)$*-values, and black dots are* $RE_n^{(3)}(x)$*-values; 500 replications.*

4.2 Results

Figures 1 and 2 show RE-values for the centrally symmetric distributions $N_3(\mathbf{0}, \mathbf{\Sigma}_i)$ $(i = 1, 2)$. We see from the $\text{RE}_n^{(1)}$-values, plotted as black triangles, that the TR–VCCM estimator $\hat{\boldsymbol{\mu}}_n^{(\alpha)}(x)$ behaves badly as opposed to the VCCM estimator $\hat{\boldsymbol{\mu}}_n(x)$, irrespective of the sample sizes n, domain values x, and covariance values $\sigma_{\mathrm{Y}_1,\mathrm{Y}_2}$. In contrast, the TR–SCM estimator $\hat{\boldsymbol{\theta}}_n^{(\alpha)}(x)$ estimator is very efficient as opposed to the SCM estimator $\hat{\boldsymbol{\theta}}_n(x)$, as can be noted from the $\text{RE}_n^{(2)}$-values (black squares). In particular, this is the case for the domain points $x = -1$ and $x = 1$ with values of $\text{RE}_n^{(2)}$ given by approximately 0.5 ($n = 100$) and 0.4 ($n = 200$). At the true center $x = 0$, the efficiency of the TR–SCM estimator seems to decrease slightly. However, note that for all domain points the improvement in efficiency of the TR–SCM estimator over the SCM estimator remains almost constant over the whole range of values $\sigma_{\mathrm{Y}_1,\mathrm{Y}_2}$. Clearly, for conditional median estimation, the presence and size of the correlations among the responses are of less importance than deviations of the covariates from the true center of the distribution. Further, note that the $\text{RE}_n^{(3)}$-values (black dots) are approximately equal to unity. Hence, both the estimators $\hat{\boldsymbol{\mu}}_n(x)$ and $\hat{\boldsymbol{\theta}}_n(x)$ seem to perform equally well in terms of RE-values.

The question which arises is what price we have paid for the increase in efficiency. One would expect an increase in bias of the TR–SCM estimator. However, we observed that the differences in bias between TR–SCM and SCMs were negligible. Also, in almost all cases, the empirical standard deviations of the TR–SCMs turned out to be smaller than the corresponding estimated standard deviations for SCMs. Thus, the lack of affine behavior of the SCMs, may be the sole problem for its relative poor performance.

Table 1 shows $\text{RE}_n^{(1)}$- and $\text{RE}_n^{(2)}$-values for AN_1 with $n = 100$, 200. The RE-values for AN_2 showed qualitatively similar patterns, and hence have been omitted. At all domain points x the $\text{RE}_n^{(1)}$-values tend to decrease as β increases, with the largest reduction occurring at $x = -1$ and $x = 1$. The TR–VCCM estimator performs poorly as opposed to the VCCM estimator with all values larger than one. In general, the $\text{RE}_n^{(2)}$-values increase as the amount of contamination increases from 10% to 25%. This is particularly the case for $x = 1$. Clearly, this is due to the fact that the right-tail of the symmetric $N_3(\mathbf{0}, \mathbf{\Sigma}_1)$ distribution is corrupted by i.i.d. observations of the $N_3(\mathbf{2}, \boldsymbol{I})$ distribution. Still, in all cases, the TR–SCM estimator is more efficient that the SCM estimator. Chakraborty and Chaudhuri (1999) showed that the finite-sample breakdown point of the TR–multivariate *unconditional* median estimator is as high as $n^{-1}[(n-(p+s)+1)/2]$, where $[\cdot]$ denotes the integer part. This result also seems to hold for the TR–SCM estimator.

Table 1: $\mathrm{RE}_n^{(1)}$- and $\mathrm{R}_n^{(2)}$-values for different values of σ_{Y_1,Y_2}.

		$x=-1$		$x=0$		$x=1$	
σ_{Y_1,Y_2}	$\beta \times 100\%$	$\mathrm{RE}_n^{(1)}$	$\mathrm{RE}_n^{(2)}$	$\mathrm{RE}_n^{(1)}$	$\mathrm{RE}_n^{(2)}$	$\mathrm{RE}_n^{(1)}$	$\mathrm{RE}_n^{(2)}$
				$n=100$			
-0.8	10	3.079	0.466	3.012	0.693	2.916	0.405
	25	1.901	0.442	1.429	0.437	1.659	0.468
-0.4	10	1.359	0.478	2.386	0.787	1.757	0.413
	25	2.578	0.471	4.120	0.547	4.120	0.547
0.0	10	1.818	0.477	2.142	0.848	2.075	0.435
	25	1.722	0.473	1.340	0.589	1.340	0.589
0.4	10	1.857	0.465	3.178	0.856	2.500	0.461
	25	1.671	0.455	1.434	0.644	1.434	0.644
0.8	10	1.081	0.433	2.050	0.784	1.316	0.452
	25	1.429	0.437	1.345	0.707	1.345	0.707
				$n=200$			
-0.8	10	1.902	0.401	1.684	0.643	1.978	0.341
	25	1.402	0.381	1.784	0.942	1.417	0.478
-0.4	10	1.455	0.381	2.350	0.940	1.607	0.439
	25	1.948	0.434	2.644	0.995	2.172	0.574
0.0	10	4.304	0.402	2.437	0.857	3.944	0.334
	25	1.465	0.406	2.603	0.960	1.499	0.577
0.4	10	1.453	0.381	3.035	0.917	1.440	0.362
	25	1.214	0.377	2.267	0.927	2.685	0.682
0.8	10	1.959	0.361	1.969	0.810	1.987	0.388
	25	1.308	0.386	1.957	0.966	1.419	0.680

Hence, in summary, the simulation results indicate that both the TR–VCCM and VCCM estimators lack efficiency and therefore should not be used in practice. The data-driven TR–procedure is recommended for obtaining affine equivariant SCMs. In Section 5 the TR-procedure will be taken out of the controlled but unrealistic environment of simulated observations and tried on a real data set. However, before closing this section, it is interesting to note that the results on the SCM- and VCCM estimators are similar to results reported by Massé and Plante (2003) for the spatial *unconditional* median and the vector of coordinate *unconditional* medians. Using a Monte Carlo study, these authors show that the unconditional spatial median, in terms of accuracy and robustness, outperforms four other bivariate medians, with the vector of coordinate unconditional medians ranking fourth.

5 An Example

To illustrate the use of the TR–procedure in estimating multivariate conditional medians, we consider a data set analysed by Cheng and De Gooijer (2003) as a part of a study of diabetes among U.S. African-Americans. For $n = 378$ respondents, the response variables Y_1 and Y_2 are systolic and diastolic blood pressure, denoted by respectively BPS and BPD. Two covariates X were selected: Age with values in the range from 19 to 92 years, and the body mass index (BMI) with values ranging from 16.04 to 55.9 kg/m^2. To assess the efficiency of the TR–SCM

Table 2: $D_i(\hat{\boldsymbol\theta}^{(\alpha)}, \hat{\boldsymbol\theta})$-values ($i = 1, 2$) for the TR–SCM estimator relative to the SCM estimator.

Covariates	$D_1(\hat{\boldsymbol\theta}^{(\alpha)}, \hat{\boldsymbol\theta})$	$D_2(\hat{\boldsymbol\theta}^{(\alpha)}, \hat{\boldsymbol\theta})$
Age	0.953	0.908
BMI	0.995	0.989

estimator relative to the SCM estimator, we compute the following two ratios of errors

$$D_1(\hat{\boldsymbol\theta}^{(\alpha)}, \hat{\boldsymbol\theta}) = \sum_{i=1}^{378} \Big\{ \frac{\sqrt{\sum_{j=1}^{2} (\hat\theta_j^{(\alpha)}(x_i) - Y_j(x_i))^2}}{\sqrt{\sum_{j=1}^{2} (\hat\theta_j(x_i) - Y_j(x_i))^2}} \Big\}, \quad D_2(\hat{\boldsymbol\theta}^{(\alpha)}, \hat{\boldsymbol\theta}) = \sum_{i=1}^{378} \Big\{ \frac{\sum_{j=1}^{2} |\hat\theta_j^{(\alpha)}(x_i) - Y_j(x_i)|}{\sum_{j=1}^{2} |\hat\theta_j(x_i) - Y_j(x_i)|} \Big\},$$

where $\hat{\boldsymbol\theta}^{(\alpha)} = (\hat\theta_1^{(\alpha)}(x_i), \hat\theta_2^{(\alpha)}(x_i))^T$, $\hat{\boldsymbol\theta} = (\hat\theta_1(x_i), \hat\theta_2(x_i))^T$, and $Y_j(x_i)$ ($i = 1, \ldots, 378; j = 1, 2$) denotes the value of the jth response variable observed at the ith value of the covariate.

In Table 2 $D_i(\hat{\boldsymbol\theta}^{(\alpha)}, \hat{\boldsymbol\theta})$-values are reported for the covariates Age and BMI. We see that the TR–SCM estimator is more efficient than the SCM estimator with all values less than one. However, the improvement is modest. The reason is that the covariates have relatively low sample correlations with the response variables: 0.452 (0.000) for Age and BPS, 0.063 (0.220) for Age and BPD, 0.114 (0.026) for BMI and BPS, and 0.153 (0.003) for BMI and BPD, with p-values in parentheses. On the other hand, the sample correlation coefficient between BPS and BPD is large: 0.608 (0.000). Thus, when Age is a covariate the reduction in efficiency of the TR–SCM estimator is solely due to the association between Age and BPD. When BMI is a covariate both the association between BMI and BPS, and BMI and BPD contribute, albeit very little, to the better performance of the TR–SCM estimator.

References

Barnett, V. (1976), "The ordering of multivariate data" (with discussion), *Journal of the Royal Statistical Society*, A, 139, 318-354.

Berlinet, A., Gannoun, A. and Matzner-Løber, E. (2001), "Asymptotic normality of convergent estimates of conditional quantiles", *Statistics*, 35, 139-169.

Berlinet, A., Cadre, B. and Gannoun, A. (2001a), "Estimation of conditional L_1-median from dependent observations", *Statistics & Probability Letters*, 55, 353-358.

Berlinet, A., Cadre, B. and Gannoun, A. (2001b), "On the conditional L_1-median and its estimation", *Nonparametric Statistics*, 13, 631-645.

Bickel, P.J. (1964), "On some alternative estimates for shift in the p-variate one sample problem", *Annals of Mathematical Statistics*, 35, 1079-1090.

Boente, G. and Fraiman, R. (1995), "Asymptotic distribution of based on local means and local medians under dependance", *Journal of Multivariate Analysis*, 54, 77-90.

Cadre, B. and Gannoun, A. (2000), "Asymptotic normality of consistent estimate of the conditional L_1-median", *Annales de l'Institut de Statistique de l'Université de Paris*, 44, 13-33.

Chakraborty, B. and Chaudhuri, P. (1996), "On a transformation and re-transformation technique for constructing an affine equivariant multivariate median", *Proceedings of the American Mathematical Society*, 124, 2539-2547.

Chakraborty, B. and Chaudhuri, P. (1999), "A note on the robustness of multivariate medians", *Statistics & Probability Letters*, 45, 269-276.

Chakraborty, B. (2003), "On multivariate quantile regression", *Journal Statistical Planning and Inference*, 110, 109-132.

Chaudhuri, P. (1992), "Multivariate location estimation using extension of R-estimates through U-statistics type approach", *Annals of Statistics*, 20, 987-917.

Cheng, Y. and De Gooijer, J.G. (2003), "On the u-th geometric conditional quantile". Report.

Gannoun, A. (1990), "Estimation non paramétrique de la médiane conditionelle: médianogramme et méthode du noyau", *Publications de l'Institute de Statistique de l'Université de Paris*, XXXXV, 11-12.

Gannoun, A., Saracco, J. Yuan, A. and Bonney, G.E. (2003), "On adaptive transformation-retransformation estimate of conditional spatial median", *Communications in Statistics: Theory and Methods*, 32, 1981-2012.

Haldane, J.B.S. (1948), "Note on the median of a multivariate distribution", *Biometrika*, 35, 414-415.

Kemperman, J.H.D. (1987), "The median of finite measure of Banach space". In *Statistical Data Analysis Based on the L_1-norm and Related Methods*, Y. Dodge (ed.), Amsterdam: North-Holland, pp. 217-230.

Massé, J-C. and Plante, J-F. (2003), "A Monte Carlo study of the accuracy and robustness of ten bivariate location estimators", *Computational Statistics & Data Analysis*, 42, 1-26.

Small, C.G. (1990), "A survey of multivariate medians", *International Statistical Review*, 58, 263-267.

Weber, A. (1909), "Über den Standort der Industrien", Tübingen. (English translation by C.J. Freidrich (1929), *Alfred Weber's Theory of Location of Industries*, Chicago University Press).