行政院國家科學委員會專題研究計畫 成果報告

多序類相關與多列相關之最大削減概似估計
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2118-M-004-006-
執 行 期 間 ：96 年 08 月 01 日至 97 年 08 月 31 日
執 行 單 位 ：國立政治大學統計學系

計 畫 主 持 人 ：鄭宗記

計畫參與人員： 碩士班研究生-兼任助理人員：任嘉珩

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 97 年 12 月 01 日

Maximum Trimmed Likelihood Estimation of Polychoric and Polyserial Correlations

Tsung-Chi Cheng*

Abstract

In this project we apply the maximum trimmed likelihood (MTL) approach (Hadi and Luceño 1997) to obtain the robust estimators of polychoric and polyserial correlations. The breakdown property of the resulting estimator is discussed. The forward search algorithm (Atkinson 1994) is adapted to compute the proposed MTL estimates. A real dataset is also used to illustrate the method and results of the detection of the outliers.

Keywords: Breakdown point, maximum trimmed likelihood estimator, polychoric correlation, polyserial correlation.

1 Introduction

The detection of multiple outliers in multivariate data has been a particularly intractable problem. There is a number of approaches for their identification, which essentially requires a robust estimation of multivariate location and shape. A difficulty is that most estimation procedures are known to break down when the fraction of contamination is greater than $1/(p+1)$, where p is the dimension of the data. Both the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) estimators provide a high breakdown of the robust estimation of multivariate location and shape (Rousseeuw and Leroy 1987). Moreover, Butler *et al.* (1993) show that the MCD estimator has better theoretical properties than the MVE. Woodruff and Rocke (1994) give empirical results which show that the MCD is preferred over the MVE in their applications. Croux and Haesbroeck (1999) discuss other statistical properties of robustness about MCD.

Rather than directly trimming the data, Hadi and Luceño (1997) present the trimmed likelihood estimator, which is based on trimming the likelihood function. They refer to this

*Department of Statistics, National Chengchi University, 64 ZhihNan Road, Section 2, Taipei 11605, Taiwan. E-mail: chengt@nccu.edu.tw

method as the *maximum trimmed likelihood* (MTL) method and the corresponding estimator as the maximum trimmed likelihood estimator (MTLE). Müller and Neykov (2003) discuss the relationships of the least trimmed squares (LTS) estimator and MTLE for a generalized linear model. Cheng (2005) combines both robust and diagnostic approaches to obtain the robust regression transformation, in which LTS and MTLE are also linked together.

Most robust estimations focus on the data only with continuous variables as discussed above. There are relatively few works available about robustness and outliers under a categorical data analysis (e.g. Bartlett and Lewis (1994), Basu and Basu (1998), Shane and Simonoff (2001)). In behavioral and psychological studies, data are often mixed with continuous and polytomous (ordinal) variables. A simple approach to analyzing this kind of data is to assign integral values to each category and proceed with the analysis as if the data have been measured on an interval scale with the desired distribution. However, this may lead to erroneous results (Song and Lee, 2001). Several approaches to dealing with this problem have been explored and proposed in the last three decades (see Olsson (1979), Lee and Poon (1986), Muthén (1987), Poon and Lee (1987) Lee, Poon and Bentler (1995), Song, J.-Q. and Lee (2001), Song, X.-Y. and Lee (2003) and among others). Lee and Xu (2003) propose a method to detect influential observation for a factor analysis with continuous and ordinal variables.

Cheng and Biswas (2008) apply the MTL approach to obtain the robust estimators of multivariate location and shape, especially for data mixed with continuous and categorical variables. Their model is aspired from the general location model of Olkin and Tate (1961). In this project we further apply the MTL approach to obtain the robust estimation of polychoric and polyserial correlations. The forward search algorithm of Atkinson (1994) is adapted to compute the proposed estimates.

2 Polychoric and polyserial correlations

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be continuous vectors of dimension p and q, respectively. It is assumed that $(\boldsymbol{x}^T, \boldsymbol{y}^T)$ is distributed a multivariate normal distribution, denoted by

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \sim MN\left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{bmatrix} \boldsymbol{C}_{xx} & \boldsymbol{C}_{xy} \\ \boldsymbol{C}_{yx} & \boldsymbol{R}_{yy} \end{bmatrix} \right),$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are $p \times 1$ and $q \times 1$ mean vectors of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{C}_{xx}$ is the $p \times p$ covariance matrix of $\boldsymbol{x}$, $\boldsymbol{C}_{xy} = \boldsymbol{C}_{yx}^T$ is $p \times q$ covariance matrix of $(\boldsymbol{x}, \boldsymbol{y})$, and $\boldsymbol{R}_{yy}$ is the correlation matrix of $\boldsymbol{y}$. It is noted that if $\boldsymbol{C}_{xx}$ is the correlation matrix of $\boldsymbol{x}$, then $\boldsymbol{C}_{xy}$ will store the correlation matrix of $(\boldsymbol{x}, \boldsymbol{y})$.

Let $\{\alpha_{i,1} = -\infty, \alpha_{i,2}, \cdots, \alpha_{i,m(i)}, \alpha_{i,m(i)+1} = \infty\}$ be the thresholds corresponding to the ith variable, $i = 1, \cdots, p$, the relations of the observable polytomous vector $\boldsymbol{z}$ with the latent continuous vector $\boldsymbol{y}$ are given by

$$Z_i = k(i), \qquad \text{if } \alpha_{i,k(i)} \leq Y_i < \alpha_{i,k(i)+1}$$

for $i = 1, 2, \cdots, p$, $k(i) = 1, 2, \cdots, m(i)$.

Note that $\rho_{ij} = \rho_{ji}$, $i, j = 1, 2, \cdots, n$, $i < j$ are off diagonal elements of $\boldsymbol{R}_{yy}$. The estimates of elements in $\boldsymbol{C}_{xy}$ based on random observations of $\boldsymbol{x}$ and $\boldsymbol{z}$ are called the polyserial correlations and the estimates of ρ_{ij} are called the polychoric correlations.

2.1 Maximum likelihood estimation

Poon and Lee (1987) show the MLE of the polyserial and polychoric correlations. Following their presentation, let $p(\boldsymbol{x})$ and $p(\boldsymbol{x}, \boldsymbol{z})$ be the probability density function of $\boldsymbol{X}$ and $(\boldsymbol{X}, \boldsymbol{Z})$, respectively. Moreover, the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ is $MN(\boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \boldsymbol{C}_{yy.x})$ with $\boldsymbol{C}_{yy.x} = \boldsymbol{R}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}$. let $\boldsymbol{c}_i^T$ be the ith row of $\boldsymbol{C}_{yx}$, and let $[diag(\boldsymbol{C}_{yy.x})]^{-1/2}$ denote the diagonal matrix with its (i,i)th entry equal to $(1 - \boldsymbol{c}_i^T\boldsymbol{C}_{xx}^{-1}\boldsymbol{c}_i)^{-1/2}$. Then $[diag(\boldsymbol{C}_{yy.x})]^{-1/2}[\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)]$ has a multivariate normal distribution with mean vector $\boldsymbol{0}$ and correlation matrix $\boldsymbol{R}$, where $\boldsymbol{R} = [diag(\boldsymbol{C}_{yy.x})]^{-1/2}\boldsymbol{C}_{yy.x}[diag(\boldsymbol{C}_{yy.x})]^{-1/2}$. It can be shown that

$$\begin{aligned} P(\boldsymbol{z}|\boldsymbol{x}) &= P(Z_1 = k(1), \cdots, Z_p = k(p)|\boldsymbol{X} = \boldsymbol{x}) \\ &= (-1)^p \sum_{i(1)=0}^{1} \cdots \sum_{i(n)=0}^{1} (-1)^{\sum_{u=1}^{p} i(u)} \\ &\quad \times \Phi_p[(\alpha_{1,k(1)+i(1)} - \boldsymbol{c}_1^T\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)(1 - \boldsymbol{c}_1^T\boldsymbol{C}_{xx}^{-1}\boldsymbol{c}_1)^{1/2}, \cdots; \boldsymbol{R}], \end{aligned}$$

where $\Phi_p(\beta_1, \beta_2, \cdots, \beta_p; \boldsymbol{R})$ is equal to

$$\int_{-\infty}^{\beta_1} \cdots \int_{-\infty}^{\beta_n} (2\pi)^{-n/2}|\boldsymbol{R}|^{-1/2} \exp\left(-\frac{\boldsymbol{y}^T\boldsymbol{R}^{-1}\boldsymbol{y}}{2}\right) dy_n \cdots dy_1.$$

Here the jth component Z_j of z takes the value from $1, 2, 3, \ldots m(j)$ and $i(z)$ denotes the index of the particular observation with $Z = z$. Thus $i(u)$ takes the value from the sequence $1, 2, \ldots f(z)$ with $f(z)$ the total number of observations with $Z = z$. Clearly,

$$\sum_{k(1)=1}^{m(1)} \cdots \sum_{k(p)=1}^{m(p)} f(k(1), \ldots, k(p)) = n.$$

We then let

$$\begin{aligned}
a_{i,1} &= -\infty, \qquad a_{i,m(i)+1} = \infty, \\
a_{i,k(i)} &= (\alpha_{i,k(i)} + \boldsymbol{c}_i^T \boldsymbol{C}_{xx}^{-1} \boldsymbol{\mu}_x)(1 - \boldsymbol{c}_i^T \boldsymbol{C}_{xx}^{-1} \boldsymbol{c}_i)^{-1/2}, \\
\boldsymbol{b}_i &= -\boldsymbol{C}_{xx}^{-1} \boldsymbol{c}_i (1 - \boldsymbol{c}_i^T \boldsymbol{C}_{xx}^{-1} \boldsymbol{c}_i)^{-1/2}, \\
r_{ij} &= (\rho_{ij} - \boldsymbol{c}_i^T \boldsymbol{C}_{xx}^{-1} \boldsymbol{c}_j)[(1 - \boldsymbol{c}_i^T \boldsymbol{C}_{xx}^{-1} \boldsymbol{c}_i)(1 - \boldsymbol{c}_j^T \boldsymbol{C}_{xx}^{-1} \boldsymbol{c}_j)]^{-1/2},
\end{aligned}$$

for $i, j = 1, \cdots, p, i < j$ and $k(i) = 2, 3, \cdots, m(i)$. Consider the one-to-one transformation by its inverse, which leads to the following results

$$\begin{aligned}
\alpha_{i,k(i)} &= (a_{i,k(i)} + \boldsymbol{b}_i^T \boldsymbol{x})(1 + \boldsymbol{b}_i^T \boldsymbol{C}_{xx} \boldsymbol{b}_i)^{-1/2}, \\
\boldsymbol{c}_i &= -\boldsymbol{C}_{xx} \boldsymbol{b}_i (1 + \boldsymbol{b}_i^T \boldsymbol{C}_{xx} \boldsymbol{b}_i)^{-1/2}, \\
\rho_{ij} &= (r_{ij} + \boldsymbol{b}_i^T \boldsymbol{C}_{xx} \boldsymbol{b}_j)[(1 + \boldsymbol{b}_i^T \boldsymbol{C}_{xx} \boldsymbol{b}_i)(1 + \boldsymbol{b}_j^T \boldsymbol{C}_{xx} \boldsymbol{b}_j)]^{-1/2}.
\end{aligned}$$

The parameters of interest are denoted by $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_x, \boldsymbol{C}_{xx}; \boldsymbol{b}_i; r_{ij}, i, j = 1, \cdots, p, i < j; a_{i,k(i)}, i = 1, \cdots, p, k(i) = 2, 3, \cdots, m(i)\}$. If $p(\boldsymbol{x})$ is the q-dimensional multivariate normal density function, then the likelihood function of $\boldsymbol{\Theta}$ is given by

$$\begin{aligned}
L(\boldsymbol{\Theta}) &= \prod_{i=1}^{n} p(\boldsymbol{x}_i) \prod_{i=1}^{n} p(\boldsymbol{z}_i | \boldsymbol{x}_i) \\
&= (2\pi)^{-qn/2} |\boldsymbol{C}_{xx}|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}_x)^T \boldsymbol{C}_{xx}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_x) \right\} \times \\
&\quad \prod_{k(1)=1}^{m(1)} \cdots \prod_{k(p)=1}^{m(p)} \prod_{f=1}^{f(k)} \left[(-1)^p \sum_{i(1)=0}^{1} \cdots \sum_{i(p)=0}^{1} (-1)^{\sum_{u=1}^{p} i(u)} \Phi_p(a_{1,k(1)+i(1)} + \boldsymbol{b}_1^T \boldsymbol{x}_{k,(f)}, \cdots; \boldsymbol{R}) \right].
\end{aligned}$$

The log-likelihood is then

$$\log L(\boldsymbol{\Theta}) = -[\log L_1(\boldsymbol{\Theta}_1) + \log L_2(\boldsymbol{\Theta}_2)] \tag{1}$$

where $\boldsymbol{\Theta}_1 = \{\boldsymbol{\mu}_x, \boldsymbol{C}_{xx}\}$ and $\boldsymbol{\Theta}_2 = \{\boldsymbol{b}_i; r_{ij}, i, j = 1, \cdots, p, i < j; a_{i,k(i)}, i = 1, \cdots, p, k(i) = 2, 3, \cdots, m(i)\}\}$; and

$$\begin{aligned}
\log L_1(\boldsymbol{\Theta}_1) &= \frac{1}{2} \left\{ qn \log(2\pi) + n \log |\boldsymbol{C}_{xx}| + \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}_x)^T \boldsymbol{C}_{xx}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_x) \right\}; \\
\log L_2(\boldsymbol{\Theta}_2) &= -\sum_{k(1)=1}^{m(1)} \cdots \sum_{k(p)=1}^{m(p)} \sum_{f=1}^{f(k)} \left[\log(-1)^p \sum_{i(1)=0}^{1} \cdots \sum_{i(p)=0}^{1} \right. \\
&\quad \left. \times (-1)^{\sum_{u=1}^{p} i(u)} \Phi_p(\cdots, a_{i,k(i)+i(i)} + \boldsymbol{b}_i^T \boldsymbol{x}_{k,(f)}, \cdots; \boldsymbol{R}) \right]
\end{aligned}$$

2.2 The maximum trimmed likelihood estimator

The trimmed likelihood principle is based on trimming the likelihood function rather than directly trimming the data, which was introduced independently by Hadi and Luceño (1997) and Vandev and Neykov (1998). It is always possible to order and trim observations according to their contributions to the likelihood function, because the likelihood is scalar-valued. For any given value of θ,

$$l(\theta; x_1) \geq l(\theta; x_2) \geq \cdots \geq l(\theta; x_n), \tag{2}$$

where $l(\theta; x_i) = \ln f(x_i; \theta)$ is the contribution of the ith observation to the log likelihood function. Therefore, the ML estimator maximizes the log likelihood function as

$$\sum_{i=1}^{n} l(\theta; x_i).$$

Instead of summing up all values of the log likelihood function for each observation, the trimmed likelihood approach considers to maximize the following objective function:

$$\sum_{i=a}^{b} w_i l(\theta; x_i), \tag{3}$$

where $a \leq b$, $(a, b) \in \{1, 2, \cdots, n\}$, and $w_i \geq 0$ are weights. The estimator $\theta(a, b, w)$ is obtained by maximizing (3). The resulting method is called as the *maximum trimmed likelihood* (MTL) method and $\hat{\theta}(a, b, w)$ is the maximum trimmed likelihood estimator (MTLE).

Neykov *et al.* (2007) give the combinatorial representation of MTLE (3) as follows:

$$\max_{\theta} \sum_{i=1}^{h} w_i l(\theta; x_i) = \max_{\theta} \max_{\mathcal{H} \in H} \sum_{i \in \mathcal{H}} w_i l(\theta; x_i) = \max_{\mathcal{H} \in H} \max_{\theta} \sum_{i \in \mathcal{H}} w_i l(\theta; x_i),$$

where H is the set of all h-subsets of the set $\{1, \cdots, n\}$. Therefore, it follows that all possible $\binom{n}{h}$ partitions of the data have to be fitted by the MLE, and the MTLE is given by the partition with the maximum log-likelihood.

Hadi and Luceño (1997) show that this trimming likelihood principle produces many existing estimators, such as MLE, least median squares (LMS), LTS, and minimum volume ellipsoid (MVE) estimators. Vandev and Neykov (1993) present the relation of MTLE with the minimum covariance determinant (MCD) estimator for multivariate data. Moreover, Vandev (1993) and Vandev and Neykov (1998) proposed more general classes of estimators based on the concept of trimming, which accommodate several kinds of estimators. The

breakdown point properties of MTLE are then unified by these authors (see also Müller and Neykov (2003) and Dimova and Neykov (2004), in which MTLE is applied to the generalized linear models). Neykov *et al.* (2007) apply the MTLE to the robust estimation for a finite mixture of distributions. Cheng and Biswas (2008) extend the MTL method to the general location model for multivariate data mixed with continuous and categorical variables. Čížek (2008) generalizes the concept of trimming likelihood approach to several applications and provides with its asymptotic properties.

In order to study the breakdown properties of general estimators such as LMS and LTS, Vandev (1993) develops a d-fullness technique. He shows that their breakdown point is not less than $(n-h)/n$ if h is within the range of values $(n+d)/2 \leq h \leq (n-d)$ for some constant d which depends upon the density considered. A finite set $\Gamma = \{\gamma_i : \Theta \to R; i = 1, \cdots, n\}$ of functions is called d-full if for every $\{i_1, \cdots, i_d\} \subset \{1, \cdots n\}$ the function γ given by $\gamma(\theta) := \max\{\gamma_{i_h}(\theta), h = 1, 2 \cdots, d\}$ is sub-compact (Muller and Neykov, 2003). The breakdown point can be exemplified by the range of values of h by using d-fullness. A recommendable choice of h is $[(n+d+1)/2]$ because then the breakdown point of MTLE is maximized.

The d-fullness technique allows the statistician to choose the tuning parameter h according to the expected percent of outliers in data. Müller (1995) disregards the assumption that the observations are in general position (Rousseeuw and Leroy, 1987) for the case of experimental design. Müller and Neykov (2003) and Dimova and Neykov (2004) relax the compactness condition required by Vandev (1993) and further present a generalization of the result for case (3). Its breakdown point depends on the quantity $\mathcal{N}(X)$ introduced by Müller (1995). $\mathcal{N}(X)$ provides the maximum number of explanatory variables lying in a subspace. The breakdown point for LTS is then determined by $\mathcal{N}(X) = \max_{0 \neq \beta \in R^p} \text{card}\{i \in \{1, \cdots, n\}; \boldsymbol{x}_i^T \boldsymbol{\beta} = 0\}$. If the explanatory variables are in general position then $\mathcal{N}(X) = p-1$ which is the minimum value for $\mathcal{N}(X)$. Müller and Neykov (2003) show the connection between d-fullness and $\mathcal{N}(X)$.

The breakdown points of MVE and MCD have been discussed by Vandev and Neykov (1993). For computational aspect, Neykov and Müller (2003) propose a fast computing algorithm for MTLE, which is analogous to the C-step for LTS and MCD of Rousseeuw and van Driessen (1999, 2006).

3 MTLE for Polychoric and polyserial correlations

We then extend the MTL approach to the estimation of polychoric and polyserial correlations for (1). According to the combinatorial representation of Neykov *et al.* (2007), let $\boldsymbol{\Theta}_h = (\boldsymbol{\Theta}_{1h}, \boldsymbol{\Theta}_{2h})$ denote the parameters for a specific value of h. If H denotes the subset with h cases and the corresponding data are denoted by $\boldsymbol{X}_h$ and $\boldsymbol{Z}_h$, then the objective function of MTLE evaluated at h is then

$$L(\boldsymbol{\Theta}_h) \;=\; \prod_{i\in H} p(\boldsymbol{x}_i) \prod_{i\in H} p(\boldsymbol{z}_i|\boldsymbol{x}_i). \tag{4}$$

An analog of the log-likelihood of (4) is then

$$\log L(\boldsymbol{\Theta}_h) \;=\; -[\log L_1(\boldsymbol{\Theta}_{1h}) + \log L_2(\boldsymbol{\Theta}_{2h})] \tag{5}$$

where

$$\begin{aligned}
\log L_1(\boldsymbol{\Theta}_{1h}) &= \frac{1}{2}\left\{ qh\log(2\pi) + h\log|\boldsymbol{C}_{xxh}| + \sum_{i\in H}(\boldsymbol{x}_i - \boldsymbol{\mu}_{xh})^T \boldsymbol{C}_{xxh}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_{xh}) \right\}; \\
\log L_2(\boldsymbol{\Theta}_{2h}) &= -\sum_{k(1)=1}^{m(1)} \cdots \sum_{k(p)=1}^{m(p)} \sum_{f=1}^{f(k)} \left[\log(-1)^p \sum_{i(1)=0}^{1} \cdots \sum_{i(p)=0}^{1} \right. \\
&\quad \left. \times (-1)^{\sum_{u=1}^{p} i(u)} \Phi_p(\cdots, a_{i,k(i)+i(i)} + \boldsymbol{b}_{ih}^T \boldsymbol{x}_{k,(f)}, \cdots; \boldsymbol{R}_h) \right].
\end{aligned}$$

Here $i(u)$ takes the value from the sequence $1, 2, \ldots f(z)$ with $f(z)$ the total number of observations with $Z = z$ corresponding to H and hence $\sum_{k(1)=1}^{m(1)} \cdots \sum_{k(p)=1}^{m(p)} f(k(1), \ldots, k(p)) = h$. The difficulty here is to find the subset $\mathcal{H}$, which corresponds to the MTLE estimator $\hat{\boldsymbol{\Theta}}_h$ and the subscript h is still used for brevity.

3.1 Breakdown point

According to those studies about MTLE (Vandev and Neykov, 1993; Müller and Neykov, 2003; Dimova and Neykov, 2004; Neykov *et al.*, 2007), let

$$\begin{aligned}
\gamma(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) := \max_{i\in H} \left\{ -\frac{1}{2}\log|\boldsymbol{C}_{xx}| + (\boldsymbol{x}_i - \boldsymbol{\mu}_x)^T \boldsymbol{C}_{xx}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_x) \right. \\
\left. + \sum_{k(i)=1}^{m(i)} \left[\log(-1)^p \sum_{i(1)=0}^{1} \cdots \sum_{i(p)=0}^{1} (-1)^{\sum_{u=1}^{p} i(u)} \Phi_p(\cdots, a_{i,k(i)+i(i)} + \boldsymbol{b}_i^T \boldsymbol{x}_i, \cdots; \boldsymbol{R}) \right] - K \right\}
\end{aligned}$$

be sub-compact for all $H \in \{1, \cdots, n\}$ with cardinality $\mathcal{N}(X) + 1$ and $K \in R$. Here $\mathcal{N}(\mathcal{X})$ is defined as the maximum value between $\mathcal{N}_1(X)$ and $\mathcal{N}_2(X)$, where $\mathcal{N}_1(X)$ is discussed in Vandev and Neykov (1993) and $\mathcal{N}_2(X) = \max_{0 \neq \beta \in R^p} \text{card}\{i \in \{1, \cdots, n\}; \boldsymbol{x}_i^T \boldsymbol{\beta} = 0\}$. Setting

$$\gamma_2(\boldsymbol{\Theta}_2) := \gamma_2(\boldsymbol{\Theta}_2(\boldsymbol{\Theta}_1)) = \max_{i \in H} \sum_{k(i)=1}^{m(i)} \Bigg[\log(-1)^p \sum_{i(1)=0}^{1} \cdots \sum_{i(p)=0}^{1} (-1)^{\sum_{u=1}^{p} i(u)} \Phi_p(\cdots, a_{i,k(i)+i(i)} + \boldsymbol{b}_i^T \boldsymbol{x}_i, \cdots; \boldsymbol{R}) \Big],$$

we see that γ_2 is a sub-compact function. Hence, there exits a compact set $\Theta_2 \subsetneq R^p$, such that $\{\boldsymbol{\Theta}_2; \gamma_2(\boldsymbol{\Theta}_2(\boldsymbol{\Theta}_1)) \leq C\} \subset \Theta_2$. Also,

$$\gamma_1(\boldsymbol{\Theta}_1) := \max_{i \in H} \left\{ -\frac{1}{2} \log |\boldsymbol{C}_{xx}| + (\boldsymbol{x}_i - \boldsymbol{\mu}_x)^T \boldsymbol{C}_{xx}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_x) \right\}$$

is sub-compact so that $\{\boldsymbol{\Theta}_1; \gamma_1(\boldsymbol{\Theta}_1) \leq C\} \subset \Theta_1$ for some compact set $\Theta_1 \subsetneq R^q$. We then have

$$\begin{aligned} &\{(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2(\boldsymbol{\Theta}_1)) \in R^q \times R^p; \gamma(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2(\boldsymbol{\Theta}_1)) \leq C\} \\ &\subset \{(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2(\boldsymbol{\Theta}_1)) \in R^q \times R^p; \gamma_2(\boldsymbol{\Theta}_2(\boldsymbol{\Theta}_1)) \leq C \text{ and } \gamma_1(\boldsymbol{\Theta}_1) \leq C\} \subset \Theta_1 \times \Theta_2. \end{aligned}$$

The breakdown point of MTLE $(\boldsymbol{\Theta}_{1h}, \boldsymbol{\Theta}_{1h})$ is not less than $(1/n) \min\{n-h+1; h-\mathcal{N}(X)\}$ and its lower bound attains the maximum value of $(1/n)[(n-\mathcal{N}(X)+1)/2]$ if $[(n+\mathcal{N}(X)+1)/2] \leq h \leq [(n + \mathcal{N}(X) + 2)/2]$, together with the above discussion and Theorem 1 of Müller and Neykov (2003).

3.2 The forward search algorithm

To obtain the resulting estimates of the previous subsection, we apply the forward search algorithm of Atkinson (1994) starts with a randomly selected subset of observations. The observations of the subset are incremented in such a way that outliers are unlikely to be included. The algorithm can be briefly summarized as follows.

- **(F0)** Choose m observations (e.g. $m = p + 1$, the so-called elemental set) from the dataset.
- **(F1)** Obtain the ML estimates based on the subset, compute the values of the log-likelihood for all observations, and order the log-likelihoods.
- **(F2)** Calculate the value of the objective criterion.

- **(F3)** Choose $m + s$ (usually $s = 1$) cases with the smallest squared distances of (F1) as the new subset, and return to step (F1).

- **(F4)** Iterate steps (F1) to (F3) until the size of the subset equals n.

We call steps (F0) to (F4) a one forward search. There are two ways for obtaining the initial subset of step (F0). The first one is the original version of Atkinson (1994), in which the forward searches are run 100 times and each initial subset is randomly chosen from the data. The other adapted version is to first get a subset which is intended to be outliers and then only one forward search is performed (see Atkinson and Riani (2000)).

3.3 Real data illustration

The data set can be obtained from http://kdd.ics.uci.edu/databases/coil/coil.html. The purpose of collecting these data is to protect rivers and streams by monitoring chemical concentrations and algae communities. Recent years have been characterized by increasing concern at the impact man is having on the environment. The impact on the environment of toxic waste, from a wide variety of manufacturing processes, is well known. In temperate climates across the world summers are characterized by numerous reports excessive summer algae growth resulting in poor water clarity, mass deaths of river fish from reduced oxygen levels and the closure of recreational water facilities on account of the toxic effects of this annual algal bloom. During the research study water quality samples were taken from sites on different European rivers of a period of approximately one year. These samples were analyzed for various chemical substances including: nitrogen in the form of nitrates, nitrites and ammonia, phosphate, pH, oxygen, chloride. In parallel, algae samples were collected to determine the algae population distributions. It is well known that the dynamics of the algae community is determined by external chemical environment with one or more factors being predominant.

The first 8 values for this data set are 8 chemical concentrations, denoted by $V_i, i = 1, \cdots, 8$, which should be relevant for the algae population distribution, denoted by V_9, and there are two ordinal variables, the river size and the fluid velocity. The following table shows the estimated correlation matrix of these data, in which the first line is calculated by MLE approach and the second line is using MTLE at $h = [0.75n]$. It is clear to see the difference between classical and robust methods.

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	Size	Velocity
V_1	1	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Polyserial	Polyserial
V_2	0.0184	1	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Polyserial	Polyserial
	-0.0388										
V_3	0.2180	-0.2305	1	Pearson	Pearson	Pearson	Pearson	Pearson	Pearson	Polyserial	Polyserial
	0.2454	-0.2699									
V_4	-0.1743	0.1056	0.1101	1	Pearson	Pearson	Pearson	Pearson	Pearson	Polyserial	Polyserial
	-0.0085	0.0769	0.2634								
V_5	-0.0766	-0.1925	0.0933	0.0688	1	Pearson	Pearson	Pearson	Pearson	Polyserial	Polyserial
	-0.1683	-0.0640	0.2949	0.3545							
V_6	0.1792	-0.3530	0.4108	0.0892	0.0811	1	Pearson	Pearson	Pearson	Polyserial	Polyserial
	0.2429	-0.3239	0.4726	0.4188	0.0497						
V_7	0.1408	-0.4167	0.4474	0.0738	0.1872	0.8336	1	Pearson	Pearson	Polyserial	Polyserial
	0.2316	-0.4399	0.5366	0.3998	0.2877	0.8849					
V_8	0.2770	-0.0358	0.1663	0.3335	0.1053	0.0529	0.2570	1	Pearson	Polyserial	Polyserial
	0.3877	-0.3319	0.3027	0.1656	0.0480	0.3406	0.5075				
V_9	-0.1611	0.1588	0.2037	0.3102	0.0664	0.0613	0.1327	0.0068	1	Polyserial	Polyserial
	-0.1630	0.1947	0.1767	0.2336	0.2750	0.0519	0.1355	-0.0936			
Size	0.2961	-0.1385	0.0905	-0.0507	0.0467	0.1969	0.2011	0.0833	0.1571	1	Polychoric
	0.2997	-0.0822	-0.0386	-0.0916	0.1229	0.0350	0.1260	0.2098	0.1643		
Velocity	-0.2776	0.2907	-0.3049	-0.0302	-0.3882	-0.2816	-0.4238	-0.3060	-0.0680	-0.5129	1
	-0.3653	0.0717	-0.2679	-0.1333	-0.2763	-0.1590	-0.2300	-0.3859	-0.0093	-0.4141	

4 Conclusion

In this project, the main contribution is to derive the breakdown point of MTLE for multivariate polychoric and polyserial correlations. In addition, the forward search algorithm is used to find the solution and the whole proposed procedure is applied to a real data set.

References

Atkinson, A. C. (1994) "Fast Very Robust Methods for the Detection of Multiple Outliers," *Journal of the American Statistical Association*, **89**, 1329-1339.

Atkinson, A. C. and Riani, M. (2000) *Robust Diagnostic and Regression Analysis*, New York: Springer-Verlag.

Bar-Hen, A. and Daudin, J. J. (1995) "Generalization of the Mahalanobis Distance in the Mixed Case," *Journal of Multivariate Analysis*, **53**, 332-342.

Barnett,V. and Lewis, T. (1994) *Outliers in Statistical Data*, 3rd ed., New York: Wiley.

Basu, A. and Basu, S. (1998) "Penalized Minimum Disparity Methods for Multinomial Models," *Statistica Sinica*, 841-860.

Bedrick, E. J., Lapidus, J. and Powell, J. F. (2000) "Estimating the Mahalanobis Distance from Mixed Continuous and Discrete Data," *Biometrics*, **56**, 394-401.

Butler, R. W., Davies, P. L. and Jhun, M. (1993) "Asymptotics for the Minimum Covariance Determinant Estimator," *The Annals of Statistics*, **21**, 1385-1400.

Cheng, T.-C. (2005) "Robust Regression Diagnostics With Data Transformations," *Computational Statistics and Data Analysis*, **49**, 875-891.

Cheng, T.-C. and Biswas, A., 2008. Maximum trimmed likelihood estimator for multivariate mixed continuous and categorical data. Computational Statistics and Data Analysis 52, 2042-2065.

Croux, C. and Haesbroeck, G. (1999) "Influence function and efficiency of the minimum covariance determinant scatter matrix estimator," *Journal of Multivariate Analysis*, **71**, 161-190.

de Leon, A. R. and Carriére, K. C. (2005) "A Generalized Mahalanobis Distance for Mixed Data," *Journal of Multivariate Analysis*, **92**, 174-185.

Fisher, L. D. and van Bell, G. (1993) *Biostatistics: A Methodology for the Health Science*, New York: Wiley.

Hadi, A. S. and Luceño, A. (1997) "Maximum Trimmed Likelihood Estimators: a Unified Approach, Examples, and Algorithms," *Computational Statistics & Data Analysis*, **25**, 251-272.

Hawkins, D. M. (1994) "The Feasible Solution Algorithm for the Minimum Covariance Determinant Estimator in Multivariate Data," *Computational Statistics & Data Analysis*, **17**, 197-210.

Hubert, M. and Rousseeuw, P. J. (1997) "Robust Regression With Both Continuous and Binary Regressors," *Journal of Statistical Planning and Inference*, **57**, 153-163.

Koepsel, T. D., Inui, T. S., and Farewell, V. T. (1981) "Factors Affecting Perforation in Acute Appendicitis," *Surgery, Gynecology and Obstetrics*, **153**, 508-510.

Lee, S.-Y. and Poon, W.-Y. (1986) "Maximum Likelihood Estimation of Polyserial Correlations," *Psychometrika*, **51**, 113-121.

Lee, S.-Y. and Poon, W.-Y. and Bentler, P. M. (1995) "A Two-stage Estimation of Structural Equation Models With Continuous and Polytomous Variables," *British Journal of Mathematical and Statistical Psychology*, **48**, 339-358.

Lee, S.-Y. and Xu, L. (2003) "Case-deletion Diagnostics or Factor Analysis Models With Continuous and Ordinal categorical Data," *Sociological Methods & Research*, **31**, 389-419.

Maronna, R. A. and Yohai, V. J. (2000) "Robust Regression With Both Continuous and Categorical Predictors," *Journal of Statistical Planning and Inference*, **89**, 197-214.

Müller, C.H., 1995. Breakdown point for designed experiments. Journal of Statistical and Planning Inference 45, 413-427.

Müller, C. H., Neykov, N. (2003) "Breakdown Points of Trimmed Likelihood Estimators and Related Estimators in Generalized Linear Models," *Journal of Statistical Planning and Inference*, **116**, 503-519.

Neykov, N.M., Filzmoser, P., Dimova, R., Neytchev, P.N., 2007. Robust fitting of mixtures using the trimmed likelihood estimator. Computational Statistics and Data Analysis 52, 299-308.

Neykov, N.M., Neytchev, P.N., 1990. A robust alternative of the maximum likelihood estimators. Proceedings of Computational Statistics 90, short communications, Dubrovnik, Yugoslavia, pp. 99-100.

Muthén, B. (1987) "LISCOMP: Analysis of Linear Statistical Equation With a Comprehensive Measurement Model," Mooresville, IN: Scientific Software Inc.

Olkin, I. and Tate, R. F. (1961) "Multivariate Correlation Models with Mixed Discrete and Continuous Variables," *Annals of Mathematical Statistics*, **32**, 448-465.

Olsson, U. (1979) "Maximum Likelihood Estimation of the Polychoric Correlation Coefficient," *Psychometrika*, **44**, 443-460.

Rousseeuw, P. J. and Leroy, A. M. (1987) *Robust Regression and Outlier Detection*, New York: John Wiley.

Poon, W.-Y. and Lee, S.-Y. (1987) "Maximum Likelihood Estimation of Multivariate Polyserial and Polychoric Correlations," *Psychometrika*, **52**, 409-430.

Shane, K. V. and Simonoff, J. S. (2001) "A Robust Approach to Categorical Data Analysis," *Journal of Computational and Graphical Statistics*, **10**, 135-157.

Song, J.-Q. and Lee, S.-Y. (2001) "Bayesian Estimation and Model Selection of Multivariate Linear Model with Polyomous Variables," *Multivariate Behavioral Research*, **37**, 453-477.

Song, X.-Y. and Lee, S.-Y. (2003) "Full Maximum Likelihood Estimation of Polychoric and Polyserial Correlations with Missing Data," *Multivariate Behavioral Research*, **38**, 57-79.

Woodruff, D. L. and Rocke, D. M. (1994) "Computable Robust Estimation of Multivariate Location and Shape in High Dimension Using Compound Estimators", *Journal of the American Statistical Association*, **89**, 888-896.

Vandev, D.L., 1993. A note on breakdown point of the least median squares and least trimmed squares. Statistics and Probability Letters 16, 117-119.

Vandev, D.L. and Neykov, N.M., 1993. Robust maximum likelihood in the Gaussian case, In: New Directions in Data Analysis and Robustness, Morgenthaler, S., Ronchetti, E. and Stahel, W.A. (eds.), Birkhauser Verlag, pp. 259-264.

Vandev, D.L., Neykov, N.M., 1998. About regression estimators with high breakdown point. Statistics 32, 111-129.

Joint Statistical Meetings（JSM）為美國統計學會所舉辦，是北美統計學界的年度盛事，歷年來每次與會人數皆超過五千人。參與者涵蓋產官學界中，各種統計理論及應用之專家、學者與使用者。本次會議依主題區分，共計超過五百個場次，每一場次約有 3 至 7 篇之論文發表；內容涵蓋各項統計理論及其相關之應用。會議相關訊息，可參考會議之網址 http://www.amstat.org/meetings/jsm/2008/。

在此次議程中，個人於會議之第一天即擔任一個場次的 Session Chair 之工作，並於會議最後一天發表論文一篇。在此會議期間，亦參與聆聽許多場次的論文發表，由於同時間發表論文甚多，主要選擇與個人之研究興趣有關，包括遺漏值分析、穩健估計分析及其他統計熱門問題，藉此得知當前的研究方向與成果。另外，亦參與有關「統計教學」部分的場次，數位學者發表其在教學上之改進及相關作法，令人映象深刻，提供個人未來在統計教學上的參考，多所助益。

由於國內統計學者，大多於美國獲得博士學位；因此，此項會議向來也是國內統計學者所重視，並為年度大事之一。此次與會之國內參與者，來自中研院統計所，國家衛生研究院、相關研究機構及各大學統計系所之學者及博士班學生有 20 人以上。平日大家忙於教學研究，在台時間很少互動；藉此機會不僅得與國外學者有所交流，亦與國內統計學者論及國內統計之研究環境與方向，及各單位之差異等，為另一收穫。

不少與會者目前仍為博士班學生，國內大學亦有博士班學生參與並發表論文，值得本校借鏡，並鼓勵本校博士班學生多多參與類似國際學術活動。另外，華人在美國統計學界，向來居重要地位，這些年來中國大陸旅居美國之學者愈來愈多。以此次會議為例，原來自中國大陸的華人學者在會場中之比例甚高，這是值得國內統計學界該注意、並需持續努力，以維持國際上之學術領先地位。

個人過去指導碩士班學生林虹妤，畢業於本校統計系所後，兩年前完成 University of Denver 商學院之學位，與其夫婿目前皆於丹佛市工作。其抽空領我至丹佛大學一遊，特別進入該校商學院與法學院參觀，為此次會議之額外收穫。

攜回資料包括會議議程紙本及光碟片。