应用概率统计 第二十七卷
第四期 2011年8月

Chinese Journal of Applied Probability
and Statistics Vol.27 No.4 Aug. 2011

Robust Asymptotic Analysis for Mean and Covariance Structure Model *

Xia Yemao Liu Yingan

(*Department of Applied Mathematics, Nanjing Forestry University, Nanjing, 210037*)

Abstract

Mean and covariance structure model is widely applied in behavioral, educational, medical, social and psychological research. The classic maximum likelihood estimate is vulnerable to outliers and distributional deviation. In this paper, robust estimate based on minimizing the objective function is proposed, and M-ratio test based on the robust deviance is suggested to assess the model fit. Empirical results are illustrated by a real example.

Keywords: Mean and covariance structure model, goodness-of-fit, robust deviance.

AMS Subject Classification: 62H15, 62H25.

§1. Introduction

Mean and covariance structure model, in which mean vector and covariance matrix share the common parameter set, has been widely studied in understanding the underlying structure of multivariate data[1–4]. The common assumptions about the distribution of the observed individuals are normal and the classic maximum likelihood (ML) approach is employed to take statistical analysis. As is well known, ML estimate is an M-estimate[5, 6] of which the influence function (IF)[7] is not bounded. This means ML estimate is sensitive to outliers in data. In order to downweight the influence of distributional deviation and outliers, robust methods such as asymptotically distribution-free (ADF)[4] methods and robust ML procedures[8, 9] are proposed. Though, less dependent on distribution, the empirical evidence reported so far does not recommend their routine use, since they may lead to excessive computational burden and/or lack of robustness in dealing with mutimodality (see [10]). [11] illustrated that ADF can't always downweight the influence of outliers for Bollen's cloud data.

In this study, we define the robust estimate for the mean and structure model based on minimizing the objective function considered by Kent and Tyler (1991). With this

*The research was fully supported by the National Natural Science Foundation of China (10671032) and Nanjing Forestry University Foundation (163101004).

Received October 23, 2008. Revised May 1, 2011.

objective function, robust deviance[13] is defined and M-ratio statistic for assessing the adequacy of model is proposed. For clarity, the following notations will be used: for any symmetric matrix $\mathbf{A}$, $\mathrm{vec}(\mathbf{A})$ denotes the p^2 dimensional vector formed by stacking the columns of matrix $\mathbf{A}_{p\times p}$, while $\mathrm{vecs}(\mathbf{A})$ is a $p^* = p(p+1)/2$ dimensional vector formed by the non-duplicated elements of $\mathbf{A}$. The duplication matrix $\mathbf{D}_p(p^2\times p^*)$ is defined as $\mathbf{D}_p\mathrm{vecs}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$[14]; $\mathbf{D}_p^+ = (\mathbf{D}_p^T\mathbf{D}_p)^{-1}\mathbf{D}_p^T$ is the generalized inverse of $\mathbf{D}_p$. For any matrix $\mathbf{B}$, $\mathbf{B}^T$ means the transpose of $\mathbf{B}$ and $\|\mathbf{B}\|$ is the norm; $\mathbf{A}\otimes\mathbf{B}$ is the Kronecker product between the matrix $\mathbf{A}$ and $\mathbf{B}$. $\dot{\mathbf{h}}_\theta$ means the partial derivatives of any vector function $\mathbf{h}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. $\mathbf{I}_p$ represents the $p\times p$ identity matrix. All proofs of the related lemmas, theorems and corollaries are given in Appendix.

§2. M-estimate

Let $\{\boldsymbol{y}_i : i = 1,\dots,n\}$ be a p-dimensional i.i.d. random sample with mean vector $\boldsymbol{\mu}_0$ and scale matrix $\boldsymbol{\Sigma}_0$. Suppose that $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ depend on an unknown q-dimensional parameter vector $\boldsymbol{\theta}_0$ such that $\boldsymbol{\mu}_0 = \boldsymbol{\mu}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, in which $\boldsymbol{\theta}_0$ lies in the parameter space $\boldsymbol{\Theta}\subseteq(\mathbb{R}^q)$. We treat $\boldsymbol{\theta}$ as a vector of mathematical variables which can take values in $\boldsymbol{\Theta}$. The following mild regularity conditions will be assumed throughout this paper.

Assumptions A

(a) The vector $\boldsymbol{\theta}_0$ is an interior point in the parameter space $\boldsymbol{\Theta}$. The matrix $\boldsymbol{\Sigma}_0$ is positive definite.

(b) The model is identification in the sense that $\boldsymbol{\mu}(\boldsymbol{\theta}_0) = \boldsymbol{\mu}(\boldsymbol{\theta}^*)$, $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\theta}^*)$ implies $\boldsymbol{\theta}_0 = \boldsymbol{\theta}^*$.

(c) All partial derivatives of the first three orders of $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are continuous and bounded in a neighborhood of $\boldsymbol{\theta}_0$.

(d) $\dot{\boldsymbol{\mu}}_\theta$ and $\dot{\boldsymbol{\sigma}}_\theta$ are full rank in a neighborhood of $\boldsymbol{\theta}_0$, where $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \mathrm{vecs}(\boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

To assess the mean and covariance structure, we define M-estimate $\widehat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_0$ by minimizing

$$\mathscr{L}_M(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^n\big\{\rho(\boldsymbol{y}_i-\boldsymbol{\mu}(\boldsymbol{\theta}))^T\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}(\boldsymbol{\theta})) + \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})|\big\}, \tag{2.1}$$

for some rho function $\rho(s)$, $s\ge 0$. (2.1) is a direct extension of (1.1) in [12] which is used to define the redescending M-estimates of multivariate location and scatter matrix without structure. Surely, when $\exp\{-\rho(\|\boldsymbol{y}_i\|^2)\}$ is integrable over R^p, (2.1) can be regarded as the negative log-likelihood from an elliptically symmetric distribution[15] with location $\boldsymbol{\mu}(\boldsymbol{\theta})$ and scale matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

If ρ is differentiable, then setting the derivatives of (2.1) with respect to $\boldsymbol{\theta}$ to zeros gives the following estimating equation (2.2).

$$\boldsymbol{G}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{y}_i, \boldsymbol{\theta}) = \mathbf{0} \tag{2.2}$$

with $\boldsymbol{g}(\boldsymbol{y}_i,\boldsymbol{\theta}) = \dot{\boldsymbol{\mu}}_\theta^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} u(s_i)[\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\theta})] + \dot{\boldsymbol{\sigma}}_\theta^T \boldsymbol{\Gamma}_N(\boldsymbol{\theta})^{-1}[u(s_i)\mathbf{s}_i - \boldsymbol{\sigma}(\boldsymbol{\theta})]$ and

$$\mathbf{s}_i = \mathrm{vecs}\{(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\theta}))(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\theta}))^T\}, \qquad \boldsymbol{\Gamma}_N(\boldsymbol{\theta}) = 2\mathbf{D}_p^+\{\boldsymbol{\Sigma}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta})\}\mathbf{D}_p^{+T},$$

in which $u(s) = \mathrm{d}\rho(s)/\mathrm{d}s$ and $s_i = s(\boldsymbol{y}_i,\boldsymbol{\theta}) = (\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{\theta}))$.

In general, the estimator defined by minimum of (2.1) is not equivalent to that through the implicit estimate equation (2.2). For the multivariate location and scatter problem, [12] provided mild conditions on $\rho(s)$ and $u(s)$ as well as on the observed data to ensure the existence and uniqueness of the estimator for the finite sample. However, as pointed out by [5], the uniqueness of the estimators may be unrealistic for a specific data. For our problems, in order to establish the consistent and the asymptotic normality of the estimation of structure parameters, we consider the following assumptions:

Assumptions B

(i) $\rho(s)$ is twice continuously differentiable and $u(s) = \mathrm{d}\rho(s)/\mathrm{d}s$.

(ii) $s^{1/2}u(s)$ and $su(s)$are bounded and $\dot{u}(s)s$ and $\dot{u}(s)s^2$ are also bounded.

(iii) $\mathsf{E}_{\theta_0}[\boldsymbol{g}(\boldsymbol{y}_i,\boldsymbol{\theta}_0)] = \mathbf{0}$, $\mathbf{Q}(\boldsymbol{\theta}_0) = \mathsf{Cov}_{\theta_0}[\boldsymbol{g}(\boldsymbol{y}_i,\boldsymbol{\theta}_0)] > \mathbf{0}$, and $\mathbf{M}(\boldsymbol{\theta}_0) = \mathsf{E}_{\theta_0}[\dot{\boldsymbol{g}}_\theta(\boldsymbol{y}_i,\boldsymbol{\theta}_0)]$ is non-singular.

Theorem 2.1 Under Assumption A and B, there is a sequence $\widehat{\boldsymbol{\theta}}_n$ such that

$$\mathbf{G}_n(\widehat{\boldsymbol{\theta}}_n) = o_p(\mathbf{1}), \qquad \widehat{\boldsymbol{\theta}}_n \xrightarrow{P_{\theta_0}} \boldsymbol{\theta}_0, \text{ and } \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}_0)),$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \mathbf{M}(\boldsymbol{\theta}_0)^{-1}\mathbf{Q}(\boldsymbol{\theta}_0)\mathbf{M}(\boldsymbol{\theta}_0)^{-T}$.

Proof See A1 in Appendix. □

The asymptotical covariance matrix of $\widehat{\boldsymbol{\theta}}_n$ can be estimated by $\boldsymbol{\Omega}(\widehat{\boldsymbol{\theta}}_n)/n$ or via $\mathbf{M}_n^{-1}\mathbf{Q}_n\mathbf{M}_n^{-T}$, where $\mathbf{M}_n$ and $\mathbf{Q}_n$ are the consistent estimators of $\mathbf{M}(\boldsymbol{\theta}_0)$ and $\mathbf{Q}(\boldsymbol{\theta}_0)$, respectively.

An useful heuristic tool to assess the effect of the observations on the estimates is the influence function (IF)[17]. For our problem, the IF of $\widehat{\boldsymbol{\theta}}_n$ is given by

$$\mathrm{IF}_\theta(\boldsymbol{y}) = -\mathbf{Q}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{g}(\boldsymbol{y},\boldsymbol{\theta}_0) \tag{2.3}$$

in which $\boldsymbol{g}(\boldsymbol{y},\boldsymbol{\theta})$ is given in (2.2). Note that

$$\boldsymbol{g}(\boldsymbol{y},\boldsymbol{\theta}) = \dot{\boldsymbol{\mu}}_\theta^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1/2} s^{1/2} u(s)\mathbf{x} + \frac{1}{2}\dot{\boldsymbol{\Sigma}}_\theta^T \{\boldsymbol{\Sigma}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta})\}^{-1/2} \mathrm{vec}[su(s)\mathbf{x}\mathbf{x}^T - \mathbf{I}],$$

where $\mathbf{x} = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))/\sqrt{s(\boldsymbol{y},\boldsymbol{\theta})}$ satisfying $\|\mathbf{x}\| = 1.0$ and $\dot{\boldsymbol{\Sigma}}_\theta = \partial\boldsymbol{\Sigma}/\partial\boldsymbol{\theta}$. Based on the Assumption A and B, it can be seen clearly that $\mathrm{IF}_\theta(\boldsymbol{y})$ is bounded.

§3. Robust Inference

3.1 Goodness-of-fit Test

To assess the hypothesized model structure, we consider the following null hypothesis $H_0 : \boldsymbol{\mu}_0 = \boldsymbol{\mu}(\boldsymbol{\theta}_0)$, $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. If H_0 is rejected, one may conclude that $\boldsymbol{\mu}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ are not agreed with the data. Let $\boldsymbol{\beta} = (\boldsymbol{\mu}^T, \boldsymbol{\sigma}^T)^T$ be the saturated model parameters, and $\widehat{\boldsymbol{\beta}}_n$ be an estimate of $\boldsymbol{\beta}_0$ which minimizes

$$\mathscr{L}_M^*(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^n \{\rho((\boldsymbol{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})) + \log|\boldsymbol{\Sigma}|\}.$$

Similar to (2.2), under that ρ is differentiable, $\widehat{\boldsymbol{\beta}}_n$ satisfies the following equation $\boldsymbol{G}_n^*(\boldsymbol{\beta}) = \sum\limits_{i=1}^n \boldsymbol{g}^*(\boldsymbol{y}_i,\boldsymbol{\beta}) = \mathbf{0}$, in which

$$\boldsymbol{g}^*(\boldsymbol{y}_i,\boldsymbol{\beta}) = \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_N^{-1} \end{bmatrix} \begin{bmatrix} u(s_i)(\boldsymbol{y}_i - \boldsymbol{\mu}) \\ u(s_i)\boldsymbol{s}_i - \boldsymbol{\sigma} \end{bmatrix}. \tag{3.1}$$

The consistence and asymptotical normality of $\widehat{\boldsymbol{\beta}}_n$ can be obtained based on the similar argument in Theorem 2.1. Specifically, the asymptotical covariance of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is $\boldsymbol{\Omega}^*(\boldsymbol{\beta}_0) = \mathbf{M}_0^{*-1}\mathbf{Q}_0^*\mathbf{M}_0^{*-T}$ with $\mathbf{M}_0^* = -\mathsf{E}_{\beta_0}[\dot{\boldsymbol{g}}_\beta^*(\boldsymbol{y}_i,\boldsymbol{\beta}_0)]$ and $\mathbf{Q}_0^* = \mathsf{Cov}_{\beta_0}[\boldsymbol{g}^*(\boldsymbol{y}_i,\boldsymbol{\beta}_0)]$.

Let $\widetilde{\boldsymbol{\beta}}_n = \boldsymbol{\beta}(\widehat{\boldsymbol{\theta}}_n)$, and $\widehat{\boldsymbol{\Delta}} = \partial\boldsymbol{\beta}/\partial\boldsymbol{\theta}^T$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n$. The following theorem establishes the Wald-type test statistic for H_0.

Theorem 3.1 Under the Assumptions A and B, and the null hypothesis $H_0 : \boldsymbol{\mu}_0 = \boldsymbol{\mu}(\boldsymbol{\theta}_0)$, $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$,

$$T_n = n(\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n)^T \widehat{\mathbf{W}}_n (\widehat{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_n) \xrightarrow{L} \chi^2(p + p^* - q),$$

where $\widehat{\mathbf{W}}_n = \mathbf{W}(\widehat{\boldsymbol{\theta}}_n) = \widehat{\boldsymbol{\Delta}}^{\perp}\{\widehat{\boldsymbol{\Delta}}^{\perp T}\widehat{\boldsymbol{\Omega}}_n^*\widehat{\boldsymbol{\Delta}}^{\perp}\}^{-1}\widehat{\boldsymbol{\Delta}}^{\perp T}$; $\widehat{\boldsymbol{\Omega}}_n^*$ is a consistent estimator of $\boldsymbol{\Omega}_0^*$ and $\widehat{\boldsymbol{\Delta}}^{\perp}$ is a $(p + p^*) \times (p + p^* - q)$ matrix of which columns are orthogonal to those of $\widehat{\boldsymbol{\Delta}}$.

Proof See A2 in Appendix. □

3.2 M-ratio Test

The function $\mathscr{L}_M(\boldsymbol{\theta})$ defined in (2.1) allows us to develop robust tools for further inference and model selection. To avoid introducing too much notation, we deal with the general case in this section. However, in the context of mean-covariance structure model, one should keep in mind that $\boldsymbol{\theta}$ in this section may be $\boldsymbol{\beta}$ for the statured model and $\boldsymbol{\vartheta}$ may be structured parameters. Assume that $\boldsymbol{\Xi}$ is a r $(r \leq q)$ dimensional parametric space and $\boldsymbol{\pi}: \boldsymbol{\Xi} \mapsto \boldsymbol{\Theta}$ is a vector-valued function such that $\boldsymbol{\pi}(\boldsymbol{\vartheta})$ is continuously differentiable with respect to $\boldsymbol{\vartheta}$. The matrix $\mathbf{A}(\boldsymbol{\vartheta}) = \dot{\boldsymbol{\pi}}_\vartheta$ is full rank of the column. The true value $\boldsymbol{\vartheta}_0$ corresponds to $\boldsymbol{\theta}_0$ which lies in the interior of $\boldsymbol{\Xi}$. A simple example is $\boldsymbol{\pi}(\boldsymbol{\vartheta}) = (\boldsymbol{\vartheta}^T, \mathbf{0}^T)^T_{q\times 1}$. We are interested in testing the null hypothesis $H_0: \boldsymbol{\theta}_0 = \boldsymbol{\pi}(\boldsymbol{\vartheta}_0)$.

Let $\widehat{\boldsymbol{\theta}}_n$ be the solution of equation (2.1) under the complete model, and $\widehat{\boldsymbol{\vartheta}}_n$ be the estimator of $\boldsymbol{\vartheta}_0$ under the reduction model with parametric space $\boldsymbol{\Xi}$ using the same procedure. Let $D_M(\boldsymbol{\theta}) = -2\mathscr{L}_M(\boldsymbol{\theta})$, a robust quasi-deviance[13] which describes the quality of a fit. Denote $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\pi}(\widehat{\boldsymbol{\vartheta}})$, we consider a robust measure of discrepancy between two nested models:

$$W_n = D_M(\widehat{\boldsymbol{\theta}}_n) - D_M(\widetilde{\boldsymbol{\theta}}_n) = 2[\mathscr{L}_M(\widetilde{\boldsymbol{\theta}}_n) - \mathscr{L}_M(\widehat{\boldsymbol{\theta}}_n)]. \tag{3.2}$$

The following theorem establishes the asymptotical distribution of W_n under the null hypothesis.

Theorem 3.2 Under the assumptions in Theorem 2.1 and $H_0: \boldsymbol{\theta}_0 = \boldsymbol{\pi}(\boldsymbol{\vartheta}_0)$,

$$W_n \xrightarrow{L} \sum_{j=1}^{q-r} \lambda_j z_j^2, \tag{3.3}$$

where z_j, $j = 1, \ldots, (q-r)$ are i.i.d. standard normal variables, and λ_j, $j = 1, \ldots, (q-r)$ are the positive eigenvalues of matrix $\mathbf{H}_0$ where

$$\mathbf{H}_0 = \mathbf{Q}_0[\mathbf{M}_0^{-1} - \mathbf{A}_0(\mathbf{A}_0^T\mathbf{M}_0\mathbf{A}_0)^{-1}\mathbf{A}_0^T],$$

and $\mathbf{Q}_0$ and $\mathbf{M}_0$ is given in Theorem 2.1.

Proof See A3 in Appendix. □

The above $\mathbf{H}_0$ is evaluated at $\boldsymbol{\theta}_0 = \boldsymbol{\pi}(\boldsymbol{\vartheta}_0)$. A natural replacement of $\boldsymbol{\vartheta}_0$ is the M-estimate $\widehat{\boldsymbol{\vartheta}}_n$. Under the normal assumption with $\rho(s) = s$, $\mathbf{H}_0$ is an idempotent matrix with $(q-r)$ unity eigenvalues. This gives (3.3) the central chi-square distribution with $(q-r)$ degrees of freedom. However, the asymptotical distribution of W_n is a linear combination of (central) χ^2 random variables with 1 degree of freedom when $\rho(s) \neq s$, even though the data is normal. One can approximate this linear combination by a scaled chi-square variable[17]. Specifically, let $\widehat{\mathbf{H}} = \mathbf{H}(\widehat{\boldsymbol{\vartheta}})$, and $\widehat{\lambda}_k$ be the positive eigenvalues of $\widehat{\mathbf{H}}$, then a rescaled statistic of (3.3) is given

$$W_n^R = W_n/\widehat{c}, \tag{3.4}$$

where $\widehat{c} = \sum_{k=1}^{q-r} \widehat{\lambda}_k/(q-r) = \mathrm{trace}(\widehat{\mathbf{H}})/(q-r)$. Using $\chi^2(q-r)$ to approximate the distribution of W_n^R represents an improvement.

§4. Real Example

In this section, we present some results of a real example to illustrate the performances of the proposed procedure. The open and closed data set[18] consists of $n = 88$ cases and $p = 5$ variables. The five topics are mechanics, vector, algebra, analysis, and statistics. First two topics were tested with closed book exams and the last three topics were tested with open book exams. The data was reanalyzed by [19] to take influence analysis under a two-factor model and consequently the 81st case is identified as the most influential point. To formulate the problem, for case i, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i5})^T$ denote the observed vector, and $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2})^T$ be the factor vector of which ω_{i1} is related to the first two observed variables and ω_{i2} to the last three observed variables. The proposed model is given by $\boldsymbol{y}_i = \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$, in which $\boldsymbol{\Lambda}$ is a 5×2 factor loading matrix and $\boldsymbol{\epsilon}_i$ is the unique error with zero mean vector and covariance matrix $\boldsymbol{\Upsilon} = \mathrm{diag}\{\psi_{11}, \ldots, \psi_{55}\}$. Further, it assumed that $\boldsymbol{\omega}_i$ is independent of $\boldsymbol{\epsilon}_i$ and has mean zeros and covariance matrix $\boldsymbol{\Phi} > 0$. For model identification (see [1]), we fixed some parameters in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ as follows: $\lambda_{12} = \lambda_{22} = \lambda_{31} = \lambda_{41} = \lambda_{51} = 0$, $\phi_{11} = \phi_{22} = 1$. Hence, the unknown parametric vector $\boldsymbol{\theta}$ are formed by the free parameters contained in $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Upsilon}$. Based on these settings, the covariance matrix of $\boldsymbol{y}_i$ is given by $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda} + \boldsymbol{\Upsilon}$. Because the true distribution is unknown, we assume the data is from the normal distribution for the ML analysis.

For the robust analysis, we consider the following weight function $u(s) = cI\{s \le s_0\} + c(\nu + s_0)/(\nu + s)I\{s > s_0\}$ for some $\nu > 0$, $c > 0$ and $s_0 \ge 0$. This weight function is normal in the middle and has t-type tail. If $\nu = 0$ with $c = 1$, $u(s)$ reduces to the Huber type weight functions. Note that the weight function $u(s)$ also satisfies the M$_1$-M$_4$ conditions of [6] with $u_1(\sqrt{s}) = u_2(s) = u(s)$. Hence, M-estimate for the mean and covariance matrix without any structure can be obtained by using iterative reweighting algorithm. The corresponding objective function is given by $\rho(s) = csI\{s \le s_0\} + c((\nu + s_0)\log((\nu + s)/(\nu + s_0)) + s_0)I\{s > s_0\}$. We take $s_0 = 11.071$ and $s_0 = 9.236$, corresponding to the 10% percentile and 5% percentile of chi-square with freedom 5, respectively. Based on the previous analysis, we choose smaller value $\nu = 1$ since the proposal procedure with larger values of ν gives the performance as that ML method. The values of c are chosen to meet Assumption B in Section 2 which gives $c^{-1} = 0.977$ and $c^{-1} = 0.952$, respectively. MLE, MLE with the 81st case removed (RMLE) and M-estimate are represented in Table 1 together with corresponding standard errors. The statistic T_n and M-ratio statistics W_n, W_n^R with freedom 4 for testing $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ are calculated. For the ML estimate, W_n is asymptotically center chi-square distribution, so the correction term is one. By comparing the fit indices with critical values $\chi_4^2(0.05)^{-1} = 9.488$, all the fit indices are far from significant, and the two factor model is not rejected by any of the fitting methods. There exists a slightly difference between T_n and W_n^R, and W_n^R is always smaller than T_n. In order to see the effect of an outlier on different methods, we created a artificial outlier by multiplying the score of the last case with 6. The corresponding results of test statistics are given in the last three rows in Table 1. The likelihood ratio test statistic and T_n under the MLE failed completely while the M-ratio statistic and T_n based on M-estimate still give the correct model assessment.

Table 1 Parameters estimate and test statistics for open and closed data

					M			
Para.	MLE	SD	RMLE	SD	5%	SD	10%	SD
λ_{11}	12.178	(1.823)	11.384	(1.789)	11.731	(1.834)	11.833	(1.832)
λ_{21}	10.328	(1.365)	9.648	(1.286)	10.084	(1.360)	10.018	(1.356)
λ_{32}	9.716	(0.917)	9.848	(0.916)	9.580	(0.921)	9.692	(0.920)
λ_{42}	12.021	(1.396)	12.013	(1.409)	11.996	(1.416)	12.046	(1.416)
λ_{52}	12.562	(1.640)	12.524	(1.658)	12.778	(1.669)	12.645	(1.659)
ψ_{11}	153.99	(31.191)	161.137	(30.168)	159.249	(31.317)	158.835	(31.35)
ψ_{22}	64.204	(17.838)	59.644	(15.600)	63.032	(17.382)	63.636	(17.406)
ψ_{33}	17.199	(6.904)	15.895	(6.665)	17.337	(6.712)	16.840	(6.809)
ψ_{44}	82.594	(16.253)	85.386	(16.362)	83.038	(16.456)	84.264	(16.563)
ψ_{55}	136.575	(24.016)	140.836	(24.466)	135.047	(24.330)	136.951	(24.318)
ϕ_{12}	0.817	(0.072)	0.87940	(0.069)	0.846	(0.072)	0.838	(0.072)
W_n	2.625	–	3.040	–	2.374	–	2.472	–
W_n^R	2.625	–	3.040	–	2.561	–	2.572	–
T_n	2.755	–	3.177	–	2.764	–	2.738	–
W_n	26.257	–	–	–	3.018	–	3.140	–
W_n^R	26.257	–	–	–	3.255	–	3.266	–
T_n	27.405	–	–	–	3.559	–	3.475	–

§5. Discussion

For the mean and covariance structure analysis, more attentions are given about the fit of a model, that is, assessing whether the mean and covariance matrix are some function of unknown parameter. The ML and GLS methods under the common normal theory easily breakdown when data are exposed to non-normal, outliers or model mis-specification. In this paper, robust version of likelihood ratio tests for mean and covariance structure models are introduced based on general M-estimates. M-ratio test is defined on the basis of robust quasi-deviance, a robust goodness-of-fit measure. They are a valuable complement to a classical techniques and are more reliable in the presence of outlying points and other deviations from the assumed model.

Appendix

A.1 Proof of Theorem 2.1

The proof of existence and consistence of estimator relies on the following lemma[20].

Lemma A.1 Let $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n$ be a i.i.d. random sample. $\boldsymbol{g}(\boldsymbol{y},\boldsymbol{\theta})$ is a stochastic function of $\boldsymbol{\theta}$ which is continuous with respect to $\boldsymbol{\theta}$. Define $G_n(\boldsymbol{\theta})=n^{-1}\sum\limits_{i=1}^{n}\boldsymbol{g}(\boldsymbol{y}_i,\boldsymbol{\theta})$. if (i) $\boldsymbol{G}_n(\boldsymbol{\theta}_0)\longrightarrow\mathbf{0}$ with probability 1; (ii) In a neighborhood of $\boldsymbol{\theta}_0$, $\dot{\boldsymbol{G}}_{n\theta}(\boldsymbol{\theta})$ converges uniformly to a non-stochastic function which is not singular at $\boldsymbol{\theta}_0$, then, with probability 1, there are zeros $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{G}_n(\boldsymbol{\theta})$ such that $\widehat{\boldsymbol{\theta}}_n\xrightarrow{P_{\theta_0}}\boldsymbol{\theta}_0$.

Based on the Assumption A and B, we only need to show that $\dot{\boldsymbol{G}}_\theta(\boldsymbol{\theta})$ converges uniformly in a neighborhood of $\boldsymbol{\theta}_0$. This can be completed by showing that $\dot{\boldsymbol{g}}_\theta(\boldsymbol{y},\boldsymbol{\theta})$ is bounded by an integrable function (see [20]). Let $\mathbf{z}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\boldsymbol{y}-\boldsymbol{\mu}(\boldsymbol{\theta}))$ and $\dot{\boldsymbol{\Sigma}}_\theta=\partial\boldsymbol{\Sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. By the matrix differential,

$$\begin{aligned}
\dot{\boldsymbol{g}}_\theta(\boldsymbol{y},\boldsymbol{\theta}) \quad = \quad & -u(s)\dot{\boldsymbol{\mu}}_\theta^T\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mu}}_\theta+u(s)\ddot{\boldsymbol{\mu}}_{\theta\theta}^T[\mathbf{I}_q\otimes\mathbf{z}]-u(s)\dot{\boldsymbol{\Sigma}}_\theta^T[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{z}]\dot{\boldsymbol{\mu}}_\theta\\
&-2\dot{u}(s)\dot{\boldsymbol{\mu}}_\theta^T[\mathbf{z}\mathbf{z}^T]\dot{\boldsymbol{\mu}}_\theta-\dot{u}(s)\dot{\boldsymbol{\Sigma}}_\theta^T[\mathbf{z}\otimes\mathbf{z}\mathbf{z}^T]\dot{\boldsymbol{\mu}}_\theta-\frac{1}{2}\ddot{\boldsymbol{\Sigma}}_{\theta\theta}^T[\mathrm{vec}\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_p]\\
&+\dot{\boldsymbol{\Sigma}}_\theta^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\dot{\boldsymbol{\Sigma}}_\theta+\frac{1}{2}u(s)\ddot{\boldsymbol{\Sigma}}_{\theta\theta}^T[\mathrm{vec}(\mathbf{z}\mathbf{z}^T)\otimes\mathbf{I}_p]\\
&-\frac{1}{2}u(s)\dot{\boldsymbol{\Sigma}}_\theta^T[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{z}\mathbf{z}^T]\dot{\boldsymbol{\Sigma}}_\theta-\frac{1}{2}u(s)\dot{\boldsymbol{\Sigma}}_\theta^T[\mathbf{z}\mathbf{z}^T\otimes\boldsymbol{\Sigma}^{-1}]\dot{\boldsymbol{\Sigma}}_\theta\\
&-\frac{1}{2}u(s)\dot{\boldsymbol{\mu}}_\theta^T[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{z}^T]\dot{\boldsymbol{\Sigma}}_\theta-\frac{1}{2}u(s)\dot{\boldsymbol{\mu}}_\theta^T[\mathbf{z}^T\otimes\boldsymbol{\Sigma}^{-1}]\dot{\boldsymbol{\Sigma}}_\theta\\
&-\dot{u}(s)\dot{\boldsymbol{\mu}}_\theta^T[\mathbf{z}\mathbf{z}^T\otimes\mathbf{z}^T]\dot{\boldsymbol{\Sigma}}_\theta-\frac{1}{2}\dot{u}(s)\dot{\boldsymbol{\Sigma}}_\theta^T[\mathbf{z}\mathbf{z}^T\otimes\mathbf{z}\mathbf{z}^T]\dot{\boldsymbol{\Sigma}}_\theta.
\end{aligned}$$

According to the property of norms of matrices and norms of vectors: $\|\mathbf{AB}\|\le\|\mathbf{A}\|\cdot\|\mathbf{B}\|$,

$\|\mathbf{A}\otimes\mathbf{B}\| = \|\mathbf{A}\|\cdot\|\mathbf{B}\|$, $\|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$. Based on the Assumption A and B, we can conclude that $\dot{\boldsymbol{g}}_\theta(\boldsymbol{y},\boldsymbol{\theta})$ is bounded.

By the Assumption A and B, we have $\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{G}_n(\boldsymbol{\theta}_0) + \dot{\boldsymbol{G}}_n(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + R_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, where $\boldsymbol{R}_n \longrightarrow \mathbf{0}$ as $\widehat{\boldsymbol{\theta}}_n \longrightarrow \boldsymbol{\theta}_0$; Hence, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = (\dot{\boldsymbol{G}}_n(\boldsymbol{\theta}_0))^{-1}(\boldsymbol{G}_n(\boldsymbol{\theta}_0) + \boldsymbol{R}_n)$. Note that $\boldsymbol{G}_n(\boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_0))$ and $\boldsymbol{G}_n(\boldsymbol{\theta}_0) \longrightarrow \mathbf{M}(\boldsymbol{\theta}_0)$ in probability. We complete the proof of asymptotic normality by the well-known Slutsky theorem.

A.2 Proof of Theorem 3.1

The proof follows the similar routine in [16]. We outlined it briefly for completeness. Let $\mathbf{M}_0 = \mathbf{M}(\boldsymbol{\theta}_0)$, $\mathbf{Q}_0 = \mathbf{Q}(\boldsymbol{\theta}_0)$, $\mathbf{M}_0^* = \mathbf{M}(\boldsymbol{\beta}_0)$, $\mathbf{Q}_0^* = \mathbf{Q}(\boldsymbol{\beta}_0)$, and $\boldsymbol{\Delta}_0 = \partial\boldsymbol{\beta}/\partial\boldsymbol{\theta}_0$. Under the null hypothesis, we have $\mathbf{G}_n(\boldsymbol{\theta}_0) = \boldsymbol{\Delta}_0^T\mathbf{G}_n^*(\boldsymbol{\beta}_0)$ and $\mathbf{M}_0 = \boldsymbol{\Delta}_0^T\mathbf{M}_0^*\boldsymbol{\Delta}_0$, $\mathbf{Q}_0 = \boldsymbol{\Delta}_0^T\mathbf{Q}_0^*\boldsymbol{\Delta}_0$. Based on the Assumption A and B, for any $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, it can be shown that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\mathbf{M}_0^{-1}\frac{1}{\sqrt{n}}\mathbf{G}_n(\boldsymbol{\theta}_0) + o_p(\mathbf{1}).$$

Similarly, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = -\mathbf{M}_0^{*-1}(1/\sqrt{n})\mathbf{G}_n^*(\boldsymbol{\beta}_0) + o_p(\mathbf{1})$. So, under the null hypothesis, $\sqrt{n}(\boldsymbol{\beta}(\widehat{\boldsymbol{\theta}}_n) - \boldsymbol{\beta}(\boldsymbol{\theta}_0)) = \boldsymbol{\Delta}_0\mathbf{M}_0^{-1}\boldsymbol{\Delta}_0^T\mathbf{M}_0^*\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(\mathbf{1})$, which gives

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}(\widehat{\boldsymbol{\theta}}_n)) = (\mathbf{I} - \boldsymbol{\Delta}_0\mathbf{M}_0^{-1}\boldsymbol{\Delta}_0^T\mathbf{M}_0^*)\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(\mathbf{1}).$$

By that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}_0^*)$, and $\boldsymbol{\Delta}_0^{\perp T}\boldsymbol{\Delta}_0 = \mathbf{0}$, we have

$$\boldsymbol{\Delta}^{\perp T}\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}(\widehat{\boldsymbol{\theta}}_n)) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Delta}_0^{\perp T}\boldsymbol{\Omega}_0^*\boldsymbol{\Delta}_0^{\perp}).$$

Hence, based on the Cochran theorem, under the null hypothesis,

$$T_n = n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}(\widehat{\boldsymbol{\theta}}_n))^T\widehat{\mathbf{W}}_n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}(\widehat{\boldsymbol{\theta}}_n)) \xrightarrow{L} \chi^2(p + p^* - q).$$

A.3 Proof of Theorem 3.2

Let $\widetilde{D}_M(\boldsymbol{\vartheta}) = D_M(\boldsymbol{\pi}(\boldsymbol{\vartheta}))$, then,

$$\frac{\partial\widetilde{D}_M}{\partial\boldsymbol{\vartheta}_0^T} = \mathbf{A}(\boldsymbol{\vartheta}_0)^T\frac{\partial D_M}{\partial\boldsymbol{\theta}_0^T}. \tag{A.1}$$

Obviously, $\widetilde{\mathbf{Q}}(\boldsymbol{\vartheta}_0) = \mathbf{A}(\boldsymbol{\vartheta}_0)^T\mathbf{Q}_0\mathbf{A}(\boldsymbol{\vartheta}_0)$, $\widetilde{\mathbf{M}}(\boldsymbol{\vartheta}_0) = \mathbf{A}(\boldsymbol{\vartheta}_0)^T\mathbf{M}_0\mathbf{A}(\boldsymbol{\vartheta}_0)$. Since

$$D_M(\widehat{\boldsymbol{\theta}}_n) - D_M(\boldsymbol{\theta}_0) = \frac{1}{2}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T\frac{\partial^2 D_M}{n\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^T}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(\mathbf{1}), \tag{A.2}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\Big(\frac{\partial^2 D_M}{n\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^T}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial D_M}{\partial\boldsymbol{\theta}_0} + o_p(\mathbf{1}), \qquad \frac{\partial^2 D_M}{n\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^T} \xrightarrow{P_{\theta_0}} \mathbf{M}_0. \tag{A.3}$$

We obtain

$$D_M(\widehat{\boldsymbol{\theta}}_n)-D_M(\boldsymbol{\theta}_0)\stackrel{e}{=}\frac{1}{2}\Big[\frac{1}{\sqrt{n}}\frac{\partial D_M}{\partial\boldsymbol{\theta}_0^T}\Big]\mathbf{M}_0^{-1}\Big[\frac{1}{\sqrt{n}}\frac{\partial D_M}{\partial\boldsymbol{\theta}_0}\Big]. \tag{A.4}$$

Similarly, following the above step and noting (A.1),

$$\widetilde{D}_M(\widehat{\boldsymbol{\vartheta}}_n)-\widetilde{D}_M(\boldsymbol{\vartheta}_0)\stackrel{e}{=}\frac{1}{2}\Big[\frac{1}{\sqrt{n}}\frac{\partial\widetilde{D}_M}{\partial\boldsymbol{\vartheta}_0^T}\Big]\widetilde{\mathbf{M}}_0^{-1}\Big[\frac{1}{\sqrt{n}}\frac{\partial\widetilde{D}_M}{\partial\boldsymbol{\vartheta}_0}\Big]. \tag{A.5}$$

Thus, based on Equation (A.1) to (A.5), and under the null hypothesis,

$$\begin{aligned}D_M(\widehat{\boldsymbol{\theta}}_n)-D_M(\widetilde{\boldsymbol{\theta}}_n) &= [D_M(\widehat{\boldsymbol{\theta}}_n)-D_M(\boldsymbol{\theta}_0)]-[\widetilde{D}_M(\widehat{\boldsymbol{\vartheta}_n})-\widetilde{D}_M(\boldsymbol{\vartheta}_0)]\\ &\stackrel{e}{=} \frac{1}{2}\Big[\frac{1}{\sqrt{n}}\frac{\partial D_M}{\partial\boldsymbol{\theta}_0^T}\Big]\mathbf{H}_0\Big[\frac{1}{\sqrt{n}}\frac{\partial D_M}{\partial\boldsymbol{\theta}_0}\Big],\end{aligned}$$

where $\mathbf{H}_0=\mathbf{M}_0^{-1}-\mathbf{A}_0(\mathbf{A}_0^T\mathbf{M}_0\mathbf{A}_0)^{-1}\mathbf{A}_0^T$. Further, for any $\boldsymbol{\theta}_0$,

$$\mathbf{z}_n=\mathbf{Q}_0^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial D_M}{\partial\boldsymbol{\theta}_0}\xrightarrow{L}N_p(\mathbf{0},\mathbf{I}).$$

Thus, $W_n=\mathbf{z}_n^T\mathbf{Q}_0^{1/2}\mathbf{H}_0\mathbf{Q}_0^{1/2}\mathbf{z}_n+o_p(\mathbf{1})$. Note that the rank of the matrix $\mathbf{Q}_0^{1/2}\mathbf{H}_0\mathbf{Q}_0^{1/2}$ equal to rank of $\mathbf{H}_0$ is $q-r$, then,

$$W_n=\sum_{i=1}^{q-r}\lambda_i z_i^2+o_p(\mathbf{1}),$$

where $\lambda_1,\lambda_2,\ldots,\lambda_{q-r}$ are the $q-r$ positive eigenvalues of matrix $\mathbf{Q}_0\mathbf{H}_0$. We complete the proof of Theorem 3.2.

References

[1] Bollen, K.A., *Structural Equations with Latent Variable*, New York: Wiley, 1989.

[2] Bentler, P.M. and Dudgeon, P., Covariance structure analysis: statistical practice, theory, and directions, *Annual Review of Psychology*, **47**(1996), 541–570.

[3] Vonesh, E.F., Wang, H. and Majumdar, D., Generalized least squares, Taylor series linearization, and Fisher's scoring in multivariate nonlinear regression, *Journal of the American Statistics Association*, **96**(2001), 282–291.

[4] Browne, M.W., Asymptotically distribution-free methods for the analysis of covariance structures, *British Journal of mathematical and statistical Psychology*, **37**(1984), 62–83.

[5] Huber, P.J., *Robust Statistics*, New York: Wiley, 1981.

[6] Maronna, R.A., Robust M-estimate of multivariate location and scatter, *Annals of Statistics*, **4**(1976), 51–67.

[7] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley, 1986.

[8] Lee, S.Y. and Xia, Y.M., Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data, *Psychometrika*, **71**(2006), 565–585.

[9] Xia, Y.M., Song, X.Y. and Lee, S.Y., Robust model fitting for the nonlinear structural equation model under normal theory, *British Journal of Mathematical and Statistical Psychology*, DOI:10.1348/000711008X345966, 2008.

[10] Muthén, B. and Kaplan, D., A comparison of some methodolodies for the factor analysis of non-normal Likert variables, *British Journal of Mathematical and Statistical Psychology*, **38**(1992), 171–189.

[11] Yuan, K.H. and Bentler, P.M., Robust mean and covariance structure analysis through iteratively reweighted least squares, *Psychometrika*, **65**(2000), 43–58.

[12] Kent, J.T. and Tyler, D.E., Redescending M-estimate of multivariate location and scatter, *The Annals of Statistics*, **19**(1991), 2102–2119.

[13] Cantoni, E. and Ronchetti, E., Robust inference generalized linear model, *Journal of the Ameriacan Statistics Association*, **96**(2001), 1022–1030.

[14] Magnus, J.R. and Neudecker, M.H., *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: Wiley, 1999.

[15] Fang, K.-T., Kotz, S. and Ng, K.W., *Symmetric Multivariate and Related Distributions*, London, England: Chapman and Hall, 1990.

[16] Yuan, K.H. and Bentler, P.M., Robust mean and covariance structure analysis, *British Journal of Mathematical and Statistical Psychology*, **51**(1998), 63–88.

[17] Yuan, K.H., Marshall, L.L. and Bentler, P.M., A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers, *Psychometrika*, **67**(2002), 95–122.

[18] Mardia, K.V., Measures of multivariate skewness and kurtosis with application, *Biometrika*, **57**(1970), 519–530.

[19] Tanaka, Y., Watadani, S. and Moon, S.H., Influence in covariance structure analysis: with application to confirmatory factor analysis, *Communication in Statistical – Theory and Method*, **20**(1991), 3805–3821.

[20] Ferguson, T., *A Course in Large Sample Threory*, London: Chapman and Hall, 1996.

均值方差结构模型的渐近稳健推断

夏业茂　　刘应安

(南京林业大学应用数学系, 南京, 210037)

均值方差模型广泛应用于行为、教育、医学、社会和心理学的研究. 经典的极大似然估计对于异常点和分布扰动易受影响. 本文基于目标函数最小化给出稳健估计, 并基于稳健偏差提出模型拟合.

关 键 词: 均值方差模型, 拟合优度检验, 稳健偏差.

学 科 分 类 号: O212.5.