Testing for elliptical symmetry of errors in the multivariate linear regression model

Yan Su^a, Shao-Yue Kang^b

School of Mathematics and Physics, North China Electric Power University, Baoding, 071003, China ^aemail: suyanhd@163.com, ^bemail:kangsy5@163.com

Keywords: Linear model; Errors; Elliptical symmetry; Goodness-of-fit; Bootstrap approximation

Abstract. This paper presents a characteristic test for testing the elliptical distribution of the errors in the multivariate linear regression model. We obtain the asymptotic spherical distribution of the transformed residuals of the regression model under the null hypothesis. Based on bootstrap approximation, an algorithm is given to estimate the critical values of the test. The test statistic possesses symmetry and then the test power can be enhanced. The test is practical to implement for arbitrary dimension of the errors.

Introduction

A multivariate linear model describes the relationship between a response vector y and a vector x of covariables. Let y_1, \dots, y_n be *n* independent observation vectors in \mathbb{R}^m , following the model

$$\mathbf{y}_{j}^{\prime} = \mathbf{x}_{j}^{\prime} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{j}^{\prime}, \, j = 1, \cdots, \mathbf{n}, \tag{1}$$

$$E(\varepsilon_j) = 0, \operatorname{Cov}(\varepsilon_j) = \Sigma, \qquad (2)$$

where the prime "'" denotes transpose, the design vectors $x_j \in \mathbb{R}^p$ are assumed to be nonrandom, β is an unknown $p \times m$ matrix of parameters called regression coefficients, $\varepsilon_j, j = 1, \dots, n$ are the *m*-vectors of errors, and Σ is an unknown $m \times m$ positive definite matrix.

The family of elliptically symmetric distributions is a natural extension of the family of multivariate normal distributions. It contains short-tailed and long-tailed distributions, including symmetric Kotz type distributions, symmetric multivariate Pearson Types VII and II distributions. The errors in a multivariate regression model can be assumed to have an elliptical distribution when the normality assumption fails.

In this paper, we consider the multivariate linear model with errors ε_j in (1)–(2) having an elliptical distribution. To avoid wrong conclusions in regression analysis, the distributional assumption on the errors should be checked. Let *F* be the unknown distribution of the errors ε_j and let F_0 be the elliptical distribution. We want to test the hypothesis

$$F = F_0. \tag{3}$$

Gamero, García and Mejías proposed a goodness-of-fit test for any fixed distribution of errors in multivariate linear models[1]. Su and Yang presented a goodness-of-fit test for uniformity on the surface of a unit sphere based on generalized inverse, the test possesses symmetry and has nice properties[2].

Let Ω_m denote the surface of a unit sphere centered at the origin in \mathbb{R}^m and let $U(\Omega_m)$ denote the uniform distribution on Ω_m . Let $\hat{\varepsilon}_j$ be the residuals of the multivariate linear regression model. The asymptotic null distribution of the transformed residuals is a spherical distribution. Based on a simple property for the spherical distribution, the goodness-of-fit test for the elliptical distribution of the errors ε_j in (1) can be translated into the goodness-of-fit test for $U(\Omega_m)$. We introduce a characterization-based test for the elliptical distribution of the errors ε_j in (1). The transformation

based on Cholesky decomposition leads to the transformed residuals whose joint distribution asymptotically does not depend on the unknown matrix Σ_m of the elliptical distribution. Hence, the critical values can be approximated by Monte Carlo and bootstrap samples.

The paper is organized as follows. In Section 2, we introduce the multivariate linear regression model and some lemmas. In Section 3, the generalized inverse-based test for the elliptical distribution of the errors is proposed. The asymptotic null spherical distribution of the transformed residuals is obtained. In Section 4, the algorithm to estimate the critical values is given. The conclusion and a possible extension of the obtained results are present in Section 5. The proofs of Theorem1 and Lemma2(b). are postponed to Appendix.

The multivariate linear model and some lemmas

Definition1^[3] Let $U^{(m)} \sim U(\Omega_m)$. An $m \times 1$ random vector ς is said to have a spherical distribution if ς has a stochastic representation $\varsigma \stackrel{d}{=} \kappa \cdot U^{(m)}$ for some random variable $\kappa \ge 0$, which is independent of $U^{(m)}$. Here $\stackrel{d}{=}$ signifies that the two sides have the same distribution.

Definition2^[3] Let A be an $m \times m$ matrix of rank m and $\Sigma_m = A A$. An $m \times 1$ random vector η is said to have an elliptical distribution with parameters $\mu(m \times 1)$ and Σ_m if

$$\eta = \mu + \kappa A U^{(d)}, \qquad (4)$$

where random variable $\kappa \ge 0$ is independent of $U^{(m)}$. We shall use the notation $\eta \sim EC_m(\mu, \Sigma_m)$. Let I_n denotes the $n \times n$ identity matrix and let

$$Y = (y_1, \dots, y_n), X = (x_1, \dots, x_n), \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$$
(5)

Then the multivariate linear model (1)-(2) takes the form

Y

$$= X\beta + \varepsilon, (6)$$

$$E[vec(\varepsilon')] = 0, \ \operatorname{Cov}[vec(\varepsilon')] = I_n \otimes \Sigma,$$
(7)

where *Y* and ε are $n \times m$ random matrices, *X* is a known $n \times p$ matrix, and β is an unknown $p \times m$ matrix. Here, the sign \otimes denotes the kronecker product of matrices.

The multivariate linear model (6) -(7) generalizes the multiple linear model (m = 1) by allowing a vector of observations, given by the rows of a matrix *Y*, to correspond to the rows of the design matrix *X*.

Lemma1^[3] Assume that $\eta \sim EC_m(\mu, \Sigma_m)$ with $E(\kappa^2) < \infty$, where κ is defined in (4). Let $Cov(\eta)$ denotes the covariance matrix of η . Then $Cov(\eta)$ exists and

$$E(\eta) = \mu, \operatorname{Cov}(\eta) = \Sigma = \frac{E(\kappa^2)}{m} \Sigma_m.$$

Lemma2 Let the model $Y = X\beta + \varepsilon$ be defined in (6). Let $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. ~ $EC_m(0, \Sigma_m)$, where $\varepsilon_1, \dots, \varepsilon_n$ are defined by (5) and $Cov(\varepsilon_1) = \Sigma$ exists. Let rank (X) = p and let

$$P_X = X(X'X)^{-1}X'.$$
 (8)

Let $\hat{\beta}$ be the least squares estimate of β , i.e., $\hat{\beta} = (X X)^{-1} X Y$. Let

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \cdots \hat{\varepsilon}_n)' = Y - X\hat{\beta}, \quad \hat{\Sigma} = \frac{1}{n - p}\hat{\varepsilon}'\hat{\varepsilon}, \quad \lim_{n \to \infty} \frac{1}{n}X'X = D,$$
(9)

where D is a positive definite matrix. Then (a).^[4]

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n)' = (I_n - P_X)\varepsilon, \qquad \hat{\beta} \xrightarrow{r} \beta, n \to \infty,$$
(10)

where \rightarrow denotes convergence in probability as $n \rightarrow \infty$.

(b).

$$\hat{\Sigma} \xrightarrow{P} \Sigma, n \to \infty, \tag{11}$$

Lemma3^[3] Assume that $\varphi \sim EC_m(0, \Sigma_m)$ with Rank $(\Sigma_m) = m$, B is an $m \times m$ matrix. Then

$$B' \varphi \sim EC_m(0, B' \Sigma_m B)$$
.

Lemma4^[5] Let $\varepsilon_i, i \le n$ be defined in (5). Let $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. $\sim EC_m(0, \Sigma_m)$ and let $\tilde{\Sigma} = \varepsilon' \varepsilon / (n-p)$. Let the Cholesky decomposition of $\tilde{\Sigma}$ be $\tilde{\Sigma} = [L(\tilde{\Sigma})][L(\tilde{\Sigma})]'$, where $L(\tilde{\Sigma})$ is lower triangular with positive diagonal elements. Let

$$w_i = [L(\tilde{\Sigma})]^{-1} \varepsilon_i, i = 1, \cdots, n, \quad W = (w_1, \cdots, w_n)^{'}.$$
(12)

Then the distribution of W does not depend on Σ_m .

Lemma5^[3] If An $m \times 1$ random vector ς has a spherical distribution then

$$\zeta / \| \zeta \| \sim U(\Omega_m),$$

where $\left\| \cdot \right\|$ denotes the Euclidean norm.

Lemma6^[2] Let $U^{(m)} = (U_1, \dots, U_m)' \sim U(\Omega_m)$ and let $G^{(m)} = (U_1^2, \dots, U_m^2)', \mu^{(m)} = (1/m, \dots, 1/m)'$. Let $U_i^{(m)} = (U_{1i}, \dots, U_{mi})', i = 1, \dots, n$ be i.i.d. $\sim U(\Omega_m)$ and let

$$Q_{jn} = \frac{1}{n} \sum_{i=1}^{n} U_{ji}^{2}, j = 1, \cdots, m, \quad V_{n}^{(m)} = (Q_{1n}, \cdots, Q_{mn})^{'}.$$
(13)

Then

(a). The covariance matrix of $G^{(m)}$ is an $m \times m$ matrix $\sigma^2 \cdot (a_{ij}) = \sigma^2 M$ with

$$\sigma^2 = \frac{2}{m^2(m+2)}, M = (a_{ij}), \tag{14}$$

$$a_{ii} = m - 1, i = 1, \dots, m, a_{ij} = -1, i, j = 1, \dots, m, i \neq j.$$

(b). rank (M) = m - 1 and $M^+ = (1/m^2)M$.

(c).

$$R_{n} = \sqrt{n} (V_{n}^{(m)} - \mu^{(m)}) \xrightarrow{d} N_{m} (0, \sigma^{2}M)$$

$$R_{n}^{'} \sigma^{-2} M^{-} R_{n} \xrightarrow{d} \chi_{m-1}^{2}, \quad n \to \infty, \quad \gamma = R_{n}^{'} (\sigma m)^{-2} M R_{n} \xrightarrow{d} \chi_{m-1}^{2}, \quad n \to \infty, \quad (15)$$

where \xrightarrow{a} denotes convergence in distribution as $n \to \infty$, χ^2_{m-1} is the chi-squared distribution with d-1 degrees of freedom.

Remark1 $Cov(U^{(m)})=m^{-1}I_m$ corresponds to the moment of inertia of $U^{(m)} \sim U(\Omega_m)$. Consider a system of *n* particles on Ω_m with unit mass. If the *n* particles are uniformly distributed on Ω_m , then the moment of inertia of the system about arbitrary unit vector should be nearly the same.

Goodness of fit test for the elliptical distribution of errors

Let Σ and $\hat{\Sigma}$ be defined in (7) and (9), respectively. Let the Cholesky decomposition of Σ , $\hat{\Sigma}$ and Σ_m be

$$\Sigma = [L(\Sigma)][L(\Sigma)]', \hat{\Sigma} = [L(\hat{\Sigma})][L(\hat{\Sigma})]', \Sigma_m = [L(\Sigma_m)][L(\Sigma_m)]',$$
(16)

respectively. Let L^{-1} be the inverse of L and let $\hat{\varepsilon}_i$ be defined in (9). Let

$$z_{i} = [L(\hat{\Sigma})]^{-1} \hat{\varepsilon}_{i}, i = 1, \cdots, n, \quad Z = (z_{1}, \cdots, z_{n})^{'},$$
(17)

$$\xi_{i}^{(m)} = z_{i} / \left\| z_{i} \right\| = (\xi_{1i}, \cdots, \xi_{mi}), i = 1, \cdots, n, \quad \psi^{(m)} = (\xi_{1}^{(m)}, \cdots, \xi_{n}^{(m)}).$$
(18)

Theorem1 Let the conditions of Lemma2 hold. Let the $n \times m$ matrix Z and the *m*-vectors $\xi_i^{(m)}, i \le n$ be defined in (17) and (18), respectively. Let $\alpha^2 = m^{-1}E(\kappa)$ with $\alpha > 0$, where κ is defined in (4). Then

(a). The asymptotic distribution of z_i is $EC_m(0, \alpha^{-2}I_m)$, which we write as $z_i \stackrel{a}{\sim} EC_m(0, \alpha^{-2}I_m)$, $i = 1, \dots, n$.

(b). z_1, \dots, z_n are asymptotically independent and the distribution of Z asymptotically does not depend on the parameter Σ_m of $EC_m(0, \Sigma_m)$.

(c). The asymptotic distribution of $\xi_i^{(m)}$ is $U(\Omega_m)$ and $\xi_1^{(m)}, \dots, \xi_n^{(m)}$ are asymptotically independent.

Let $\xi_i^{(m)} = (\xi_{1i}, \dots, \xi_{mi})$ be defined in (18) and let σ^2 and *M* are defined in (14), respectively. Let

$$\tilde{Q}_{jn} = \frac{1}{n} \sum_{i=1}^{n} \xi_{ji}^{2}, j = 1, \cdots, m, \quad \tilde{V}_{n}^{(m)} = (\tilde{Q}_{1n}, \cdots, \tilde{Q}_{mn})^{'},$$
(19)

$$\tilde{R}_{n} = \sqrt{n} (\tilde{V}_{n}^{(m)} - \mu^{(m)}), \quad T = T(\hat{\varepsilon}) = \tilde{R}_{n} (\sigma m)^{-2} M \tilde{R}_{n},$$
(20)

where $\mu^{(m)}$ is defined in (15).

Remark2 Consider the null hypothesis (3), where F_0 denotes the $EC_m(0, \Sigma_m)$ distribution with the parameter Σ_m unknown. By Theorem1, the goodness-of-fit test for F_0 can be translated into the goodness-of-fit test for $\xi_i^{(m)} \sim U(\Omega_m)$, $i = 1, \dots, n$. By Lemma6(c), the elliptical symmetry is rejected for large values of $T(\hat{\varepsilon})$ in (20).

The algorithm to implement the test statistic

The algorithm to compute the test statistic.

The algorithm to compute $T(\hat{\varepsilon})$ in (20) consists of the following steps:

- 1. Compute the values of $\hat{\varepsilon}$ and $\hat{\Sigma}$ in (9), respectively.
- 2. Compute the value of Z in (17).
- 3. Compute the value of $\psi^{(m)}$ in (18).
- 4. Compute the value of $\tilde{V}_n^{(m)}$ in (19).
- 5. Compute the values of \tilde{R}_n and $T(\hat{\varepsilon})$ in (20), respectively.

The elliptical symmetry is rejected for large value of $T(\hat{\varepsilon})$.

The algorithm to estimate the critical values.

Let z_i be defined in (17) and let

$$\tau_i = \|z_i\|, i = 1, \cdots, n, \tag{21}$$

where $\|\cdot\|$ denotes the Euclidean norm. The bootstrap method uses the empirical distribution of τ_i , $i = 1, \dots, n$ to approximate the distribution of κ in (4).

The algorithm to estimate the critical values of $T(\hat{\varepsilon})$ consists of the following steps:

- 1. Sample $\tau_1^*, \dots, \tau_n^*$ with replacement from the values τ_1, \dots, τ_n .
- 2. Generate U_1^*, \dots, U_n^* which are i.i.d. uniform on Ω_m . Compute $\varepsilon_i^* = \tau_i^* U_i^*, i = 1, \dots, n$.
- 3. Let $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)'$. Compute $\hat{\varepsilon}^* = (\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*)' = (I_n P_X)\varepsilon^*$, where P_X is defined in (8).
- 4. Compute $\hat{\Sigma}^* = [\hat{\varepsilon}^*] \hat{\varepsilon}^* / (n-p)$, $z_i^* = [L(\hat{\Sigma}^*)]^{-1} \hat{\varepsilon}_i^*, i = 1, \dots, n$, where $\hat{\Sigma}^* = [L(\hat{\Sigma}^*)][L(\hat{\Sigma}^*)]'$ (the Cholesky decomposition).
- 5. Compute $\xi_i^* = z_i^* / ||z_i^*|| = (\xi_{1i}^*, \dots, \xi_{mi}^*), i = 1, \dots, n$.

6. Compute
$$\tilde{Q}_{j_n}^* = \frac{1}{n} \sum_{i=1}^n [\xi_{j_i}^*]^2, j = 1, \cdots, m, \quad \tilde{V}_n^* = (\tilde{Q}_{j_n}^*, \cdots, \tilde{Q}_{j_n}^*)^{'}.$$

7. Compute $\tilde{R}_{n}^{*} = \sqrt{n} (\tilde{V}_{n}^{*} - \mu^{(m)})$, where $\mu^{(m)} = (1/m, \dots, 1/m)'$.

8. Compute $T^* = T(\hat{\varepsilon}^*) = [\tilde{R}_n^*](\sigma m)^{-2} M \tilde{R}_n^*$, where σ^2 and M are defined in (14), respectively.

Doing these *N* times gives a sample of replicates T_1^*, \dots, T_N^* . Let $T_{(1)}^*, \dots, T_{(N)}^*$ be the order statistics, the critical values for $T(\hat{\varepsilon})$ can be estimated from $T_{(1)}^*, \dots, T_{(N)}^*$.

Conclusions

Elliptical distribution plays an important role in generalized multivariate analysis. When the distribution of the errors ε_j , $j = 1, \dots, n$ in (1) enjoys elliptical symmetry, the direction vectors $\xi_i^{(m)}$ in (18) should be, approximately, uniformly distributed on the surface of the unit sphere Ω_m . Based on the generalized inverse of the covariance matrix $\sigma^2 M$ of $G^{(m)}$ in Lemma6, $T(\hat{\varepsilon})$ in (20) is constructed which possesses symmetry. Hence, the proposed test statistic $T(\hat{\varepsilon})$ will have good power for testing goodness of fit to the elliptical distribution of errors in the multivariate linear model.

The multivariate times series ζ_t follows a vector autoregressive(VAR) model, if $\frac{p}{r}$

$$\zeta_t = \phi_0 + \sum_{i=1}^{p} \phi_i \zeta_{t-i} + a_t,$$

where a_t is a sequence of i.i.d. random vectors with mean zero and covariance matrix Σ_a . The goodness-of-fit test for the elliptical distribution of the errors ε_t in the multivariate linear regression model can be extended to testing the elliptical distribution of the innovations a_t in the VAR model.

Appendix

Proof of Lemma2(b). By (a), we have

$$\hat{\Sigma} = \frac{1}{n-p} \hat{\varepsilon}' \hat{\varepsilon} = \frac{1}{n-p} \varepsilon' (I_n - X(X'X)^{-1}X') \varepsilon$$
$$= \frac{n}{n-p} \left[\frac{1}{n} \varepsilon' \varepsilon - \left(\frac{1}{n} \varepsilon' X\right) \left(\frac{1}{n} X'X\right)^{-1} \left(\frac{1}{n} X'\varepsilon\right) \right].$$
(22)

By Lemma1 and the law of large numbers,

$$\frac{1}{n}\varepsilon \varepsilon \xrightarrow{p} \frac{E(\kappa^2)}{m} \cdot \Sigma_m = \Sigma = (\sigma_{ij})_{m \times m}, \quad n \to \infty.$$
(23)

Let $X = (X_{(1)}, \dots, X_{(p)})$, $\varepsilon = (\varepsilon_{(1)}, \dots, \varepsilon_{(m)})$ and $X' \varepsilon = (\lambda_{ij})_{p \times m}$. Then $\lambda_{ij} = X_{(i)} \varepsilon_{(j)}, i = 1, \dots, p, j = 1, \dots, m.$

By Chebyshev's inequality for any $\Delta > 0$,

$$P(\left|n^{-1}\lambda_{ij}\right| > \Delta) \le \frac{1}{n^{2}\Delta^{2}}Cov(X_{(i)}\varepsilon_{(j)}) = \frac{X_{(i)}X_{(i)}}{n\Delta^{2}} \cdot \frac{\sigma_{jj}}{n},$$
(24)

where $\sigma_{jj} > 0$ is defined in (23). By (9), we have

$$n^{-1}\lambda_{ij} \xrightarrow{p} 0, \quad n^{-1} \cdot X \stackrel{p}{\varepsilon} \xrightarrow{p} 0, \quad n \to \infty,$$
 (25)

By (23) and (25), $\hat{\Sigma} \xrightarrow{p} \Sigma$, $n \to \infty$. **Proof of Theorem1**. By (9) and Lemma2(a),

$$\hat{\varepsilon}_{i} = y_{i} - x_{i} \hat{\beta} \xrightarrow{p} \varepsilon_{i}, n \to \infty.$$
(26)

Thus, the asymptotic distribution of $\hat{\varepsilon}_i$ is $EC_m(0, \Sigma_m)$, i.e., $\hat{\varepsilon}_i \stackrel{a}{\sim} EC_m(0, \Sigma_m)$, $i = 1, 2, \dots, n$.

By Lemma1 and (16), we have $\Sigma = \alpha^2 \Sigma_m$. By Lemma2(b),

$$L(\hat{\Sigma}) \xrightarrow{r} L(\Sigma) = \alpha L(\Sigma_m), n \to \infty, \qquad (27)$$

Thus,

$$z_i = [L(\hat{\Sigma})]^{-1} \hat{\varepsilon}_i \xrightarrow{P} \tilde{z}_i = [\alpha L(\Sigma_m)]^{-1} \varepsilon_i, n \to \infty, \qquad (28)$$

where ε_i is defined in (5). Since $\varepsilon_i \sim E_m(0, \Sigma_m)$, by (27) - (28) and Lemma3, we have

$$\tilde{z}_i \sim EC_m(0, \alpha^{-2}I_m), \quad z_i \sim EC_m(0, \alpha^{-2}I_m), \tag{29}$$

Thus, the desired results of (a) and (b) are proved. Since \tilde{z}_i in (29) has a spherical distribution, we have by Lemma5

$$\xi_i^{(\mathrm{m})} \xrightarrow{d} \tilde{z}_i / \|\tilde{z}_i\| \sim U(\Omega_m), n \to \infty$$

The desired result of (c) is obtained.

References

- [1] M.D.J. Gamero, J.M. García and R.P. Mejías: Testing goodness of fit for the distribution of errors in multivariate linear models. Journal of multivariate analysis, 95(2005), p. 301~322.
- [2] Y. Su, Z.H. Yang: Goodness-of-fit test for uniformity on the surface of a unit sphere based on generalized inverse. In : Recent Advance in Statistics Application and Related Areas (Conference Proceedings of The 4th International Institute of Statistics & Management Engineering Symposium, Dalian, China), edited by K.L. Zhu, H. Zhang, Aussino Academic Publishing House, Part 2, Sydney, (2011) p.1323~1327.
- [3] K.T. Fang, S. Kotz and K.W. Ng: Symmetric Multivariate and Related Distributions. Chapman & Hall, London, New York (1990).
- [4] W.H. Greene: Economrtric Analysis, 4th ed. Prentice Hall, Inc.(2000).
- [5] F.W. Huffer, C. Park: A test for elliptical symmetry. Journal of multivariate analysis, 98 (2007), p. 256~281.
- [6] R.J. Muirhead: Aspects of multivariate statistical theory. John Wiley & Sons, Inc., New York -Chichester - Brisbane - Toronto - Singapore (1982).
- [7] J.A. Díaz-García, R. Gutiérrez-Jáimez: The distribution of the residual from a general elliptical multivariate linear model. Journal of Multivariate Analysis, 97 (2006), p.1829~1841.
- [8] B.M.G. Kibria, Haq M.S. Haq: Predictive inference for the elliptical linear model. Journal of Multivariate Analysis, 68 (1999), p. 235-249.
- [9] L.X. Zhu, R.Q. Zhu and S.Song: Diagnostic checking for multivariate regression models. Journal of Multivariate Analysis, 99(2008), p.1841~1859.
- [10] J.D. Hamilton: Time series analysis, Princeton University Press(1994).
- [11] S.Pynnönen: Distribution of an arbitrary linear transformation of internally studentized residuals of multivariate regression with elliptical errors. Journal of Multivariate Analysis, 107 (2012), p. 40~52.
- [12] R.S. Tsay: Multivariate time series analysis with R and financial applications. John Wiley & Sons, Inc. Hoboken, New Jersey (2014).