

Testing for multivariate normality of disturbances 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; Disturbances; Residuals; Multivariate normal distribution; Goodness-of-fit

Abstract. We suggest a characteristic test for testing the multivariate normal distribution of the disturbances in the multivariate linear regression model(MLRM). The test is based on the goodness-of-fit test for uniformity on the surface of a unit sphere. The asymptotic null distribution of the transformed residuals from the MLRM is obtained. An algorithm is given to approximate the critical values of the test by Monte Carlo simulation. The test possesses symmetry and can be easily computed for arbitrary dimension of the disturbance vectors.

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 R^m , following the model

$$y_j' = x_j' \beta + \varepsilon_j', j = 1, \dots, n, \quad (1)$$

$$E(\varepsilon_j) = 0, \text{Cov}(\varepsilon_j) = \Sigma, \quad (2)$$

where the prime “'” denotes transpose, the design vectors $x_j \in R^p$ are assumed to be nonrandom, β is an unknown $p \times m$ matrix of parameters called regression coefficients, $\varepsilon_j, j \leq n$ are the m -vectors of disturbances (or errors), and Σ is an unknown $m \times m$ positive definite matrix.

Classical theory on the multivariate linear model assume the disturbances ε_j in (1) are normally distributed. To avoid wrong conclusions in regression analysis, the distributional assumption on the disturbances should be checked. Let F be the unknown distribution of the disturbances ε_j and let F_0 be the $N(0, \Sigma)$ distribution. We want to test the hypothesis

$$F = F_0. \quad (3)$$

Based on a integral of the squared modulus of the difference between the empirical characteristic function of the residuals and the characteristic function under the null hypothesis, Gamero, García and Mejías(2005) proposed a goodness-of-fit test for any fixed distribution of disturbances in multivariate linear models[1]. Let Ω_m denote the surface of a unit sphere centered at the origin in R^m and let $U(\Omega_m)$ denote the uniform distribution on Ω_m . Based on the moment of inertia and the center of mass of the samples on Ω_m , Su and Yang (2009) proposed the goodness-of-fit tests for $U(\Omega_m)$, the power simulation has shown that the test has good power[2]. Su and Yang(2011) extended the test in [3], we presented a test for uniformity distribution on the surface of a unit sphere based on generalized inverse, the test possesses symmetry and has nice properties[3]. Hence, the test can increase the test power.

Let $\hat{\varepsilon}_j$ be the residuals of the multivariate linear regression model. The asymptotic null distribution of the transformed residuals is $U(\Omega_m)$. Therefore, the goodness-of-fit test for the multivariate normal distribution of the disturbances ε_j in (1) can be translated into the

goodness-of-fit test for $U(\Omega_m)$. Based on the goodness-of-fit test for $U(\Omega_m)$, this paper introduces a new test for the multivariate normal 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 parameter Σ of the $N(0, \Sigma)$ distribution. Thus, the critical values of the test statistic can be estimated by Monte Carlo methods with $\Sigma = I_m$, where I_m denotes the $m \times m$ identity matrix.

The paper is organized as follows. In Section 2, we introduce the multivariate linear regression model and some lemmas. In Section 3, the characterization-based test for multivariate normal disturbances is proposed. The asymptotic null distribution of the transformed residuals is obtained. In Section 4, the algorithm to estimate the critical values is given. Some discussions and further applications are given in Section 5.

The multivariate linear model and some lemmas

Let

$$Y = (y_1, \dots, y_n)', X = (x_1, \dots, x_n)', \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)' \quad (4)$$

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

$$Y = X\beta + \varepsilon, \quad (5)$$

$$E[\text{vec}(\varepsilon')] = 0, \text{Cov}[\text{vec}(\varepsilon')] = I_n \otimes \Sigma, \quad (6)$$

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 statement that the random matrix $\varepsilon \square N_{n \times m}(0, I_n \otimes \Sigma)$ is equivalent to the statement that the random vector $\text{vec}(\varepsilon) \square N_{nm}(0, I_n \otimes \Sigma)$. The multivariate linear model (5) -(6) 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^[4] Let the model $Y = X\beta + \varepsilon$ be defined in (5). Let $\text{rank}(X) = p$ and let

$$\varepsilon \square N_{n \times m}(0, I_n \otimes \Sigma), P_X = X(X'X)^{-1}X', \quad (7)$$

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

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)' = Y - X\hat{\beta}, \quad \hat{\Sigma} = \frac{1}{n-p} \hat{\varepsilon}'\hat{\varepsilon}, \quad \lim_{n \rightarrow \infty} \frac{1}{n} X'X = D, \quad (8)$$

where D is a positive definite matrix. Then

(a). $\hat{\beta}$ and $\hat{\Sigma}$ are independent. Moreover,

$$\hat{\beta} \xrightarrow{P} \beta, \hat{\Sigma} \xrightarrow{P} \Sigma, n \rightarrow \infty, \quad (9)$$

where \xrightarrow{P} denotes convergence in probability as $n \rightarrow \infty$.

(b). The residual matrix

$$\hat{\varepsilon} = (I_n - P_X)\varepsilon, \quad \hat{\varepsilon} \sim N(0, (I_n - P_X) \otimes \Sigma). \quad (10)$$

Moreover, the i th row of $\hat{\varepsilon}$, denoted as $\hat{\varepsilon}_i$, has an m -variate normal distribution, i.e.,

$$\hat{\varepsilon}_i \sim N(0, (1 - h_{ii})\Sigma), \quad i = 1, 2, \dots, n, \quad (11)$$

where h_{ii} indicates the i th element of P_X in (7).

Lemma2^[5] (The Cholesky decomposition). If A is an $m \times m$ positive definite matrix then there exists a unique $m \times m$ lower-triangular matrix L with positive diagonal elements such that $A = LL'$.

Lemma3^[5] Let $\varepsilon \square N_{n \times m}(0, I_n \otimes \Sigma)$ and let $\tilde{\Sigma} = \varepsilon'\varepsilon / (n-p)$. Let the Cholesky decomposition of $\tilde{\Sigma}$ be $\tilde{\Sigma} = [L(\tilde{\Sigma})][L(\tilde{\Sigma})']$ and let

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

where ε_i is defined in (4). Then the distribution of W does not depend on Σ .

Definition1 A matrix C^- such that $CC^-C = C$ is called a generalized inverse of C . The Moore-Penrose C^+ is a generalized inverse of a matrix C that satisfies the following requirements:

$$CC^+C = C, C^+CC^+ = C^+, (C^+C)' = C^+C, (CC^+)' = CC^+.$$

Definition2^[6] Let $U^{(m)} \square U(\Omega_m)$. An $m \times 1$ random vector ζ is said to have a spherical distribution if ζ has a stochastic representation $\zeta \stackrel{d}{=} \eta \cdot U^{(m)}$ for some random variable $\eta \geq 0$, which is independent of $U^{(m)}$. Here $\stackrel{d}{=}$ signifies that the two sides have the same distribution.

Lemma4^[6] If An $m \times 1$ random vector ζ has a spherical distribution then

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

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

Lemma5^[6] Let $U^{(m)} = (U_1, \dots, U_m)'$ $\square U(\Omega_m)$. Then

$$E(U^{(m)}) = 0, \text{Cov}(U^{(m)}) = \frac{1}{m} I_m.$$

Remark1 The covariance matrix $\text{Cov}(U^{(m)}) = (1/m)I_m$ is the main characterization of $U(\Omega_m)$. The $\text{Cov}(U^{(m)})$ of $U^{(m)}$ corresponds to the moment of inertia of $U^{(m)}$ (i.e., the second moment of the coordinate variable of $U^{(m)}$).

Lemma6^[2] Let $\tau^{(m)} = (\tau_1, \dots, \tau_m)'$ be an $m \times 1$ random vector. Let $\|\tau^{(m)}\| = 1$ and let $\text{Cov}(\tau^{(m)})$ exist. Then the expectation moment of inertia of $\tau^{(m)}$ about arbitrary direction $H^{(m)} = (h_1, \dots, h_m)'$ does not depend on $H^{(m)}$ if and only if

$$E(\tau_i^2) = \frac{1}{m}, i = 1, \dots, m, \quad E(\tau_i \tau_j) = 0, i \neq j, i, j = 1, \dots, m,$$

where $\|H^{(m)}\| = 1$.

Remark2 Lemma6 indicates that the expectation moment of inertia of $U^{(m)}$ about arbitrary direction is the same if $U^{(m)}$ is uniformly distributed on Ω_m .

Lemma7^[3] Let $U^{(m)} = (U_1, \dots, U_m)'$ $\square U(\Omega_m)$ 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. $\square U(\Omega_m)$ and let

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

Then

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

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

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

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

(c). $R_n = \sqrt{n}(V_n^{(m)} - \mu^{(m)}) \xrightarrow{d} N_m(0, \sigma^2 M), \quad R_n' \sigma^{-2} M^{-1} R_n \xrightarrow{d} \chi_{m-1}^2, \quad n \rightarrow \infty.$

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

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

Remark3 Since $\text{rank}(M) = m-1$, the covariance matrix of $G^{(m)} = (U_1^2, \dots, U_m^2)'$ is non-negative definite. Thus, the asymptotic chi-squared distribution of the statistic γ in (15) can be obtained by taking $M^+ = (1/m^2)M$.

Goodness of fit test for the multivariate normal distribution of disturbances

Let Σ and $\hat{\Sigma}$ be defined in (6) and (8), respectively. Let the Cholesky decomposition of Σ and $\hat{\Sigma}$ be

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

respectively. Let L^{-1} be the inverse of L and let $\hat{\varepsilon}_i$ be defined in (8). Let

$$z_i = [L(\hat{\Sigma})]^{-1} \hat{\varepsilon}_i, i = 1, \dots, n, \quad Z = (z_1, \dots, z_n)', \quad (17)$$

$$\xi_i^{(m)} = z_i / \|z_i\| = (\xi_{1i}, \dots, \xi_{mi})', i = 1, \dots, n, \quad \psi^{(m)} = (\xi_1^{(m)}, \dots, \xi_n^{(m)})'. \quad (18)$$

The z_i are known as the scaled residuals (or spherized data), $\xi_i^{(m)}$ are the projections of the z_i 's on the unit sphere.

Theorem1 Let the conditions of lemma1 hold. Let the $n \times m$ matrix Z and the m -vectors $\xi_i^{(m)}, i \leq n$ be defined in (17) and (18), respectively. Then

(a). The asymptotic distribution of z_i is $N_m(0, I_m)$ and z_1, \dots, z_n are asymptotically independent. The distribution of Z asymptotically does not depend on Σ in (7).

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

Proof By (7) and (8),

$$h_{ii} = x_i'(X'X)^{-1}x_i = \frac{1}{n} x_i' \left(\frac{1}{n} X'X \right)^{-1} x_i \rightarrow 0, n \rightarrow \infty.$$

Thus, we have by (11), the asymptotic distribution of $\hat{\varepsilon}_i$ is $N(0, \Sigma)$, which we write as $\hat{\varepsilon}_i \stackrel{a}{\sim} N(0, \Sigma)$, $i = 1, 2, \dots, n$. By Lemma1(a),

$$\hat{\varepsilon}_i \xrightarrow{P} \varepsilon_i, \quad L(\hat{\Sigma}) \xrightarrow{P} L(\Sigma), n \rightarrow \infty, \quad \hat{\varepsilon} \stackrel{a}{\sim} N(0, I_n \otimes \Sigma). \quad (19)$$

Thus,

$$z_i = [L(\hat{\Sigma})]^{-1} \hat{\varepsilon}_i \xrightarrow{P} \tilde{z}_i = [L(\Sigma)]^{-1} \varepsilon_i, n \rightarrow \infty, \quad (20)$$

where ε_i is defined in (4). Since $\varepsilon_i \square N_m(0, \Sigma)$, by (19) - (20), and Lemma3, we have

$$\tilde{z}_i \square N_m(0, I_m), \quad z_i \stackrel{a}{\square} N_m(0, I_m). \quad (21)$$

Thus, the desired results of (a) is proved. By (20) - (21), the desired result of (b) is obtained.

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

$$\tilde{Q}_{jn} = \frac{1}{n} \sum_{i=1}^n \xi_{ji}^2, j = 1, \dots, m, \quad \tilde{V}_n^{(m)} = (\tilde{Q}_{1n}, \dots, \tilde{Q}_{mn})', \quad (22)$$

$$\tilde{R}_n = \sqrt{n}(\tilde{V}_n^{(m)} - \mu^{(m)}), \quad \lambda = \lambda(\hat{\varepsilon}) = \tilde{R}_n'(\sigma m)^{-2} M \tilde{R}_n. \quad (23)$$

Remark4 Consider the null hypothesis (3), where F_0 denotes the $N(0, \Sigma)$ distribution with the parameter Σ unknown. By Theorem1, the goodness-of-fit test for F_0 can be translated into the goodness-of-fit test for $\xi_i^{(m)} \stackrel{a}{\square} U(\Omega_m), i = 1, \dots, n$. The multivariate normality is rejected for large values of $\lambda(\hat{\varepsilon})$ in (23)

Remark5 Theorem1 indicates that $\xi_1^{(m)}, \dots, \xi_n^{(m)}$ are asymptotically independent $U(\Omega_m)$ random vectors. Hence, the critical values of the test statistic $\lambda(\hat{\varepsilon})$ can be estimated by Monte Carlo simulation with $\Sigma = I_m$.

The algorithm to implement the test statistic

The algorithm to compute the test statistic

The algorithm to compute $\lambda(\hat{\varepsilon})$ in (23) consists of the following steps:

1. Compute the values of $\hat{\varepsilon}$ and $\hat{\Sigma}$ in (8), 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 (22).
5. Compute the values of \tilde{R}_n and $\lambda(\hat{\varepsilon})$ in (23), respectively.
6. The multivariate normality is rejected for large value of $\lambda(\hat{\varepsilon})$.

The algorithm to estimate the critical values

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

1. Generate $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)'$ from the multivariate normal distribution $N_{n \times m}(0, I_n \otimes I_m)$.
2. By Lemmal(b), compute

$$\hat{\varepsilon}^* = (\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*)' = (I_n - P_X)\varepsilon^*,$$

where P_X is defined in (7).

3. 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).

4. Compute $\xi_i^* = z_i^* / \|z_i^*\| = (\xi_{1i}^*, \dots, \xi_{mi}^*)', i = 1, \dots, n$.

5. Compute $\tilde{Q}_{jn}^* = \frac{1}{n} \sum_{i=1}^n [\xi_{ji}^*]^2, j = 1, \dots, m$, $\tilde{V}_n^* = (\tilde{Q}_{1n}^*, \dots, \tilde{Q}_{mn}^*)'$.

6. Compute $\tilde{R}_n^* = \sqrt{n}(\tilde{V}_n^* - \mu^{(m)})$,

where $\mu^{(m)} = (1/m, \dots, 1/m)'$.

7. Compute $\lambda^* = \lambda(\hat{\varepsilon}^*) = [\tilde{R}_n^*]'(\sigma m)^{-2} M \tilde{R}_n^*$,

where σ^2 and M are defined in (14), respectively.

Doing these N times gives a sample of replicates $\lambda_1^*, \dots, \lambda_N^*$. Let $\lambda_{(1)}^*, \dots, \lambda_{(N)}^*$ be the order statistics, the critical values for λ can be estimated from $\lambda_{(1)}^*, \dots, \lambda_{(N)}^*$.

Conclusions

When the distribution of the disturbances $\varepsilon_j, j \leq n$ in (1) enjoys multivariate normality, 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 $U(\Omega_m)$ in (14), the test statistic $\lambda(\hat{\varepsilon})$ in (23) is constructed which possesses symmetry. By Lemma7(b), the Moore-Penrose inverse $M^+ = (1/m^2)M$. Hence, the proposed test statistic $\lambda(\hat{\varepsilon})$ can be computed easily for any dimension of the disturbance vector.

The elliptically symmetric distribution is a natural extension of the multivariate normal distribution. The disturbance in a multivariate linear model can be assumed to have an elliptical distribution in robustness studies. Based on a property for the spherical symmetry and the modified EDF test, Su and Guo(2012) suggested the test procedures for testing the elliptical distribution[7]. The goodness-of-fit test for the multivariate normal distribution of the disturbances in the multivariate linear regression model(MLRM) can be extended to testing the elliptical distribution of the disturbances in MLRM.

Acknowledgements

This paper is supported by “ the Fundamental Research Funds for the Central Universities”.

References

- [1] Gamero M.D.J., García J.M., Mejías R.P., Testing goodness of fit for the distribution of errors in multivariate linear models. *Journal of multivariate analysis*, 2005, 95, 301~322.
- [2] Su Y., Yang Z.H., Goodness-of-fit analysis for uniformity on the surface of a unit sphere. *Acta Mathematicae Applicatae Sinica*, 2009, 32 (1), 93~105 (in Chinese).
- [3] Su Y., Yang Z.H., 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 Zhu K.L., Zhang H. Sydney: Aussino Academic Publishing House, 2011, Part 2, 1323~1327.
- [4] Greene W.H., *Econometric Analysis*, 4th ed. Prentice Hall, Inc., 2000.
- [5] Huffer F.W., Park C., A test for elliptical symmetry. *Journal of multivariate analysis*, 2007, 98, 256~281.
- [6] Fang K.T., Kotz S., Ng K.W., *Symmetric Multivariate and Related Distributions*. London, New York: Chapman & Hall, 1990.
- [7] Su Y., Guo L.H., A characterization-based test for the elliptically symmetric distribution. In: *5th International Institute of Statistics and Management Engineering Symposium 2012: Data-Driven Management Science under Developing*. Sydney: Aussino Academic Publishing House, 2012, 472-475.
- [8] Mardia K.V., Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 1970, 57, 519~530.
- [9] Liang J.J., Pan W.S.Y., Yang Z.H., Characterization-based Q-Q plots for testing multinormality. *Statistics & Probability Letters*, 2004, 70, 183~190.
- [10] Muirhead R.J., *Aspects of multivariate statistical theory*, New York - Chichester - Brisbane - Toronto - Singapore : John Wiley & Sons, Inc., 1982.
- [11] Zhu L.X., Zhu R.Q., Song S., Diagnostic checking for multivariate regression models. *Journal of Multivariate Analysis*, 2008, 99, 1841~1859.
- [12] Díaz-García J.A., Gutiérrez-Jáimez R., The distribution of the residual from a general elliptical multivariate linear model. *Journal of Multivariate Analysis*, 2006, 97, 1829~1841.
- [13] Ng V.M., Robust bayesian inference for seemingly unrelated regressions with elliptical errors. *Journal of Multivariate Analysis*, 2001, 83, 409~414.
- [14] Watson G.S., *Statistics on spheres*, New York : John Wiley & Sons, Inc., 1983.
- [15] Gilberto A. Paula, Marcio Medeiros, Filldor E. Vilca-Labra, Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics and probability letters*, 2009, 79, 339-346.
- [16] Pynnönen S., Distribution of an arbitrary linear transformation of internally studentized residuals of multivariate regression with elliptical errors. *Journal of Multivariate Analysis*, 2012, 107, 40~52.