King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 420

Jan 2011

Robustness of Correlation Coefficient and Variance Ratio in the Class of Bivariate Elliptical Distributions

Anwar H. Joarder

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: c-math@kfupm.edu.sa

Robustness of Correlation Coefficient and Variance Ratio in the Class of Bivariate Elliptical Distributions

Anwar H. Joarder
Department of Mathematics and Statistics
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia

Email: anwarj@kfupm.edu.sa

Keywords and Phrases: bivariate distribution; correlation coefficient; variance ratio; robustness; the class of elliptical distributions

AMS Mathematics Subject Classification (2000): 62E15, 60E05, 60E10

Parametric robustness of a statistic in a class of distributions implies that the distribution of the statistic is the same for any member of the class of distributions. The bivariate Wishart distribution involves three essential statistics, namely, two sample variances and the product moment correlation coefficient. The product moment correlation coefiificent is known to be robust in the class of elliptical distributions. In this paper, we prove that the variance ratio is also robust for the class of elliptical distributions.

1. Introduction

Consider a random variable following the class of bivariate elliptical distributions with mean θ (column vector order 2 components) and scale matrix Σ (a 2×2 matrix) where $\theta' = (\theta_1, \theta_2)$ and $\Sigma = (\sigma_{ik})$, $i = 1,2;\ k = 1,2$. Let $\sigma_{11} = \sigma_1^2,\ \sigma_{22} = \sigma_2^2,\ \sigma_{12} = \rho\sigma_1\sigma_2$ with $\sigma_1 > 0,\ \sigma_2 > 0$ and the quantity ρ $(-1 < \rho < 1)$ is the product moment correlation coefficient between X_1 and X_2. Let each of the sample observation X_j, $(j = 1, 2, \cdots, N)$, follow the class of elliptical distributions. The sample mean vector is $\overline{X}$, where $\overline{X}' = (\overline{X}_1, \overline{X}_2)$, so that the sums of squares and cross product matrix is given by $A = (a_{ik})$, where $a_{ik} = \sum_{J=1}^{N} (X_{ij} - \overline{X}_i)(X_{kj} - \overline{X}_k),\ i = 1,2; k = 1,2.$ Obviously, $a_{ii} = \sum_{j=1}^{N} (X_{ij} - \overline{X}_i)^2$, $(i = 1,2)$, and $a_{12} = \sum_{j=1}^{N} (X_{1j} - \overline{X}_1)(X_{2j} - \overline{X}_2).$ The sample correlation coefficient is then given by $r = a_{12} / (ms_1s_2),$ where $s_i^2 = a_{ii} / m\ (i = 1,2)$.

Fisher (1915) derived the distribution of the matrix A to study the distribution of correlation coefficient for a bivariate normal sample. The distribution of A is also known for the class of elliptical distributions. See, for example, Sutradhar and Ali (1989, 157).

A recent interest among the applied scientists is the use of fat tailed distribution for modeling business data such as stock returns. Since the bivariate *t*-distribution has fatter tails, it has been increasingly applied for modeling business data. Interested readers may go through

Sutradhar and Ali (1986), Lange, Little and Taylor (1989), Billah and Saleh (2000) and Kibria and Saleh (2000). A wider class of distributions accommodating bivarite *t*-distribution or bivariate normal distribution is the class of Compound Normal Distributions. Bivariate elliptical distributions accommodates all these distributions. We assume that in the case of bivariate elliptical distribution, the observations in the sample are uncorrelated but not necessarily independent.

It is proved by Ali and Joarder (1991) that the distribution of R is robust in the class of bivariate elliptical distributions. Thus the null $(H_0 : \rho = 0)$ or non-null $(H_1 : \rho \neq 0)$ distribution of the test statistic R is robust. However, if the parent population is bivariate normal, the test statistic $(m-1)^{1/2}(R-\rho)(1-R^2)^{-1/2}$ is known to have a *t*-distribution with m degrees of freedom under the null hypothesis. In view of Ali and Joarder (1991), the test statistic is robust in the class of bivariate elliptical distributions.

Consider the scaled variances $U = mS_1^2 / \sigma_1^2$ and $V = mS_2^2 / \sigma_2^2$. Assuming that the observations are from a bivariate normal population, Bose (1935) proved that the distribution of variance ratio $H = U / V$ has a correlated $F(m, m; \rho)$ distribution (See equation 5.1) which specializes to usual *F*-distribution $F(m, m)$ if $\rho = 0$. The random variables U and V have a bivariate chi-square distribution (Gunst and Webster, 1973) with correlation coefficient ρ^2 and found application in signal processing (Gerkmann and Martin, 2009).

Can we relax the assumption of bivariate normality to a broader class of distributions and study the behaviour of the variance ratio? In this paper, we prove that if the sample observations are governed by the class of bivariate elliptical distributions, the distribution of the variance ratio remains the same.

In Section 4, we report the joint distribution of scaled variances and correlation coefficient whenever the sample observations follow the class of bivariate elliptical distributions. It is pointed out that the distribution of the correlation coefficient is robust in the class of bivariate elliptical distributions. In Section 5, we prove that the distribution of variance ratio $(H = U / V)$ has a correlated $F(m, m; \rho)$ distribution even if the observations are governed by the class of bivariate elliptical distributions. Thus we say that the variance ratio is robust or its distribution is invariant in the class of bivariate elliptical distributions.

2. The Bivariate Normal Distribution

Let X be a bivariate normal random vector with density function

$$\xi_1(x) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp\left[\frac{-1}{2}\left((x-\theta)'\Sigma^{-1}(x-\theta)\right)\right], \qquad (2.1)$$

where $\Sigma > 0$. Now consider a sample $X_1, X_2, \cdots X_N$ $(N > 2)$ having the joint probability density function

$$\xi_2(x_1, x_2, \cdots, x_N) = (2\pi)^{-N} |\Sigma|^{-N/2} \exp\left(\frac{-1}{2}\sum_{j=1}^{N}(x_j-\theta)'\Sigma^{-1}(x_j-\theta)\right), \qquad (2.2)$$

where $\Sigma > 0.$ Each observation $X_j (j = 1, 2, \cdots, N)$ in (2.2) follows (2.1). Since the observations in (2.2) are uncorrelated, by virtue of normality, they are also independent. Fisher (1915) derived the distribution of A in order to study the distribution of correlation coefficient from a normal sample. The distribution of A is given by

$$\xi(A) = 4\pi\Gamma(m-1)\,|\,\Sigma\,|^{-m/2}|\,A\,|^{(m-3)/2} \exp\left(-\frac{1}{2} tr\Sigma^{-1}A\right), \quad A > 0, \tag{2.3}$$

where $m > 2$ and $\Sigma > 0,$ (See e.g. Anderson, 2003, 252).

3. The Class of Bivariate Elliptical Distributions

In this section, we will present the class of bivariate elliptical distributions which includes the class of compound bivariate normal distribution. The latter includes bivariate *t* and bivariate normal distributions as special cases.

Let X be a random variable following the class of bivariate elliptical distributions so that its density function is is given by

$$f_1(x) \propto |\,\Sigma\,|^{-1/2}\ g_{N,2}\left((x-\theta)'\Sigma^{-1}(x-\theta)\right), \tag{3.1}$$

where $\Sigma > 0$ and the normalizing constant is determined by the form of g (Sutradhar and Ali, 1989). Johnson (1987) describes the generation of random samples from the class of bivariate elliptical distributions. Now consider a sample $X_1, X_2, \cdots X_N\ (N > 2)$ having the joint probability density function

$$f_2(x_1, x_2, \cdots, x_N) \propto |\,\Sigma\,|^{-N/2}\ g_{N,2}\left(\sum_{j=1}^{N}(x_j-\theta)'\Sigma^{-1}(x_j-\theta)\right), \tag{3.2}$$

where $\Sigma > 0$ and the normalizing constant is determined by the form of $g.$ Each observation $X_j (j = 1, 2, \cdots, N)$ in (3.2) follows (3.1). Since the observations are uncorrelated but not necessarily independent, (3.2) is called Uncorrelated and Identical Bivariate Elliptical (UIBE) model for the sample. It can be checked that the coefficient of correlation between components X_{1j} and X_{2j} of $X_j (j = 1, 2, \cdots, N)$ is ρ (Ali and Joarder, 1991). Note that the sample observations in (3.2) are independent if $g_{N,2}(u) = e^{-u/2}$ in which case (3.2) defines the joint density function of N independent observations from a bivariate normal distribution. It is worth mentioning that if the sample observations are independent and Identical bivariate Elliptical distributions (3.1), then $f_2(x_1, x_2, \cdots, x_N) = \prod_{j=1}^{N} f_1(x_j),$ which is a Independent and Identical Bivariate Elliptical Model (IIBEM) for the sample.

Theorem 3.1 (Sutradhar and Ali, 1989, 158) Let A be the mean centered sum of squares and product matrix based on UIBE model (3.2). Then the density function of the A is given by

$$f(A) \propto |A|^{(m-3)/2} g_{m,2}(tr\Sigma^{-1}A), \quad A > 0, \tag{3.3}$$

where $m > 2$ and $\Sigma > 0$.

4. Robustness of Some Tests on Correlation Coefficient

In the following theorem, we report the joint density function of scaled variances and correlation coefficient.

Theorem 4.1 Let S_1^2, S_2^2 and R be variances and correlation coefficient based on Uncorrelated and Identical Bivariate Elliptical model (3.2). Then the joint density function of $U = mS_1^2/\sigma_1^2, V = mS_2^2/\sigma_2^2$ and R is given by

$$f_{U,V,R}(u,v,r) \propto (uv)^{(m/2)-1}(1-r^2)^{(m-3)/2} g_{m,2}\left(\frac{u+v}{1-\rho^2} - \frac{2\rho r\sqrt{uv}}{1-\rho^2}\right), \tag{4.1}$$

where $m > 2$ and $-1 < \rho < 1$.

The density function depends on the particular form of $g_{m,2}(.)$ implying that the joint distribution of scaled variances and correlation coefficient is not robust. Note that in case of sampling from a bivariate normal population, the joint density function of scaled variances and correlation coefficient is given by

$$f_{U,V,R}(u,v,r) = \frac{(1-\rho^2)^{-m/2}(uv)^{(m-2)/2}(1-r^2)^{(m-3)/2}}{2^m\sqrt{\pi}\;\Gamma(m/2)\Gamma\left((m-1)/2\right)} \exp\left(-\frac{u+v}{2(1-\rho^2)} + \frac{\rho r\sqrt{uv}}{1-\rho^2}\right), \tag{4.2}$$

where $m > 2$ and $-1 < \rho < 1$, (Joarder, 2009).

The density function of R was derived originally by Fisher (1915) assuming that the sample is from a bivariate normal distribution. The following theorem is due to Ali and Joarder (1991) who proved that the distribution of R remains the same as that obtained by Fisher (1915) if the sample observations follow UIBE model (3.2).

Theorem 4.2 The density function of sample correlation coefficient R based on a sample following Uncorrelated and Identical Bivariate Elliptical model (3.2) is given by

$$f_R(r) = \frac{2^{m-2}\left(1-\rho^2\right)^{m/2}}{\pi\;\Gamma(m-1)} \left(1-r^2\right)^{(m-3)/2} \sum_{k=0}^{\infty} \frac{(2\rho r)^k}{k!}\;\Gamma^2\left(\frac{m+k}{2}\right), \quad -1 < r < 1, \tag{4.3}$$

where $m > 2$ and $-1 < \rho < 1$.

Theorem 4.2 indicates robustness of the correlation coefficient in the class of bivariate elliptical populations. Thus the assumption of bivariate normality under which tests on correlation coefficient are developed can be relaxed to a broader class of bivariate elliptical distributions.

If $\rho = 0$, $R^2 \sim Beta(1/2, (m-1)/2)$ and $\sqrt{m-1}\, R\left(1-R^2\right)^{-1/2}$ has a Student t-distribution with $(m-1)$ degrees of freedom. The likelihood ratio test of the hypothesis $H_0 : \rho = 0$ against the alternative $H_1 : \rho \neq 0$ is done by the above statistic. Acceptance of the null hypothesis does not mean independence unless the sample is from bivariate normal distribution. In view of Ali and Joarder (1991), the test is robust in the class of bivariate elliptical distributions in which case the acceptance of $H_0 : \rho = 0$ implies uncorrelation but not necessarily independence.

In view of Ali and Joarder (1990), the tests presented in Muddapur (1988) are robust in the class of bivariate elliptical distributions.We warn that the distribution of R is not necessarily robust for independent observations from elliptical population. The models for samples considered in (3.2) imply that the observations X_j $(j = 1, 2, \cdots, N)$ are uncorrelated but not necessarily independent. The asymptotic distribution of R for independent observations from bivariate elliptical population was obtained by Muirhead (1982, 157).

5. The Distribution of Variance Ratio

In this section we prove that the distribution of $H = U/V$ is robust in the class of bivariate elliptical distributions.

Theorem 5.1 Let S_1^2, S_2^2 and R be variances and correlation coefficient based on Uncorrelated and Identical Bivariate Elliptical model (3.2). Also let $U = mS_1^2/\sigma_1^2$ and $V = mS_2^2/\sigma_2^2$ be scaled sample variances. Then the density function of $H = U/V$ is given by

$$f_H(h) = \frac{(1-\rho^2)^{m/2}}{B\left(m/2, m/2\right)} \frac{h^{(m/2)-1}}{(1+h)^m} \left(1 - \frac{4\rho^2 h}{(1+h)^2}\right)^{-(m+1)/2}, \; h > 0. \tag{5.1}$$

where $m > 2$ and $-1 < \rho < 1$.

Proof. It follows from (4.1) that

$$f_H(h) \propto h^{(m-2)/2} \int_{r=-1}^{1} \left(1-r^2\right)^{(m-3)/2} \int_{v=0}^{\infty} v^{m-1}\, g_{m,2}\left(\frac{v(1+h)}{1-\rho^2} - \frac{2\rho r v\sqrt{h}}{1-\rho^2}\right) dv dr,$$

which can be simplified to be

$$f_H(h) \propto \frac{h^{(m-2)/2}}{(1+h)^m} \int_{r=-1}^{1} \left(1 - \frac{2\rho r\sqrt{h}}{1+h}\right)^{-m} \left(1-r^2\right)^{(m-3)/2} dr.$$

Having completed the integral, we have

$$f_H(h) \propto \frac{h^{(m-2)/2}}{(1+h)^m}\left(1-\frac{4\rho^2 h}{(1+h)^2}\right)^{(m+1)/2},$$

which, apart from the normalizing constant, is the same as (5.1).

Equation (5.1) is well known for bivariate normal distribution (Bose, 1935). This proves the robustness of variance ratio in the class of bivariate elliptical distributions. The distribution of test statistic $H = U/V$ given by (5.1) will be denoted by $F(m, m; \rho)$.

Acknowledgements The author gratefully acknowledges the excellent research facility provided by King Fahd University of Petroleum & Minerals.

References

Ali, M.M. and Joarder, A.H. (1991). Distribution of the correlation coefficient for the class of bivariate elliptical models. *Canadian Journal of Statistics,* 19, 447-452.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons. New York.

Billah, M.B. and Saleh, A.K.M.E. (2000). Performance of the large sample tests in the actual forecasting of the pretest estimators for regression model with *t*-error. *Journal of Applied Statistical Science,* 9(3), 237-251.

Bose, S.S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population, *Sankhya*, 2, 65-72.

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507-521.

Gerkmann, T. And Martin, R. (2009). On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling. *IEEE Transactions on Signal Processing,* 57(11), 4165-4174.

Gunst, R.F. And Webster, J.T. (1973). Density functions of the bivariate chi-square distribution. *Journal of statistical computation and simulation*, 2, 275-288.

Joarder, A.H. (2009). Moments of the product and ratio of two correlated chi-square random variables. *Statistical Papers,* 50(3), 581-592.

Johnson, M.E. (1987). *Multivariate Statistical Simulation*. John Wiley and Sons, New York.

Kibria, B.M.G. and Saleh, A.K.M.E. (2000). Conflict among the shrinkage estimators induced by W, LR and LM tests under a Student's t regression model. *Journal of the Korean Statistical Society*, 33(4), 411-433.

Lange, K.L.; Little, R.J.A. and Taylor, J.M.G. (1989) Robust statistical modeling using the *t* distribution. *Journal of American Statistical Association*, 84, 881-896.

McGraw, D.K. and Wagner, J.F. (1968). Elliptically symmetric distributions. *IEEE Transactions on Information Theory*, 14, 110-120.

Muddapur, M.V. (1988). A simple test for correlation coefficient in a bivariate normal distribution. *Sankhya*, Series B, 50, 50-68.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley, New York.

Sutradhar, B.C. and Ali, M.M (1986). Estimation of the parameters of regression model with a multivariate *t* error variable. *Communications in Statistics: Theory and Methods*, 15, 429-450.

Sutradhar, B.C. and Ali, M.M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivatiate *t* model. *Journal of Multivariate Analysis*, 29(1), 155-162.

RESMSA \ BED \ RandFTR