

# Power analysis for linear models with spherical errors

Weijing Wang<sup>a</sup>, Martin T. Wells<sup>b,\*</sup>

<sup>a</sup>*Institute of Statistics, National Chiao-Tung University, Taiwan, ROC*

<sup>b</sup>*Department of Social Statistics, Cornell University, Ithaca, NY 14851-0952, USA*

---

## Abstract

In this article, we consider testing a general linear hypothesis for a regression model when the error distribution belongs to the class of spherical distributions. The distributional robustness of the  $F$ -statistics under a null hypothesis for spherically symmetric distributions is well understood. This invariance property, however, does not hold under the alternative hypothesis. Motivated by a simplified example, we study the relationship between power of the test and the error distribution's dispersion and kurtosis. We find that these two parameters are not sufficiently precise measures for determining the power behavior of a test.

© 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Dispersion; Elliptically and spherically symmetric distributions; Kurtosis; Linear models; Power; Tail behavior

---

## 1. Introduction

C.R. Rao's contributions to inference for linear models has had a far reaching impact on numerous branches of science. His book, *Linear Statistical Inference and its Applications*, has already been translated into six major languages, and is one of the most cited book in statistics. After his detailed study of the classical linear model and concise development of parametric inference, Prof. Rao (1965, Section 7e.1) pointed out, "in any problem we may have a set of observations and some partial information regarding the probability distribution of the observation". He then goes on to discuss the robustness properties of Student's  $t$ -test under a null hypothesis. This paper

---

\* Corresponding author. Department of Industrial and Labor Relations, Cornell University, 14850 Ithaca, USA.

*E-mail address:* [mtwl@cornell.edu](mailto:mtwl@cornell.edu) (M.T. Wells).

expands on Rao’s discussion of null robustness and examines the robustness and power properties of hypothesis tests for the general linear models with a spherically symmetric error term under an alternative hypothesis.

Consider the following regression model,  $Y = X\beta + \varepsilon$ , where  $Y$  is an  $n \times 1$  response vector,  $X$  is an  $n \times p$  nonrandom design matrix with rank  $p$ ,  $\beta$  is a  $p \times 1$  vector of unknown parameter of interest, and  $\varepsilon$  is an  $n \times 1$  vector of stochastic errors. To test the linear hypothesis,  $H_0 : R\beta = r$  vs.  $H_a : R\beta \neq r$ , where  $R$  is a  $q \times p$  matrix with rank  $q (q < p)$  and  $r$  is a  $q \times 1$  vector of constants, the following statistic is often used:

$$\lambda = \frac{(Y - X\hat{\beta}_0)^T(Y - X\hat{\beta}_0) - (Y - X\hat{\beta})^T(Y - X\hat{\beta})}{(Y - X\hat{\beta})^T(Y - X\hat{\beta})}, \tag{1}$$

where  $\hat{\beta} = (X^T X)^{-1} X^T Y$  is the least-squares estimator of  $\beta$  and

$$\hat{\beta}_0 = \hat{\beta} - (X^T X)^{-1} R^T [R(X^T X)^{-1} R^T]^{-1} (R\hat{\beta} - r)$$

is the least-squares estimator derived under  $H_0$ . The model above can be reexpressed as  $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ , where  $X_1 : n \times (p - q)$  and  $X_2 : n \times q$  are constructed by finding a  $(p - q) \times p$  matrix  $G$  complementary to  $R$  such that  $(X_1, X_2) = X \begin{pmatrix} G \\ R \end{pmatrix}^{-1}$ ,  $\beta_1 = G\beta : (p - q) \times 1$  and  $\beta_2 = R\beta : q \times 1$ . For details on the model re-expression, please refer to [Rao and Toutenburg \(1995, Section 3.7\)](#). Hence  $\lambda$  can be written as

$$\lambda = \frac{\tilde{\varepsilon}^T A_1 \tilde{\varepsilon}}{\varepsilon^T A_2 \varepsilon} = \frac{\varepsilon^T A_1 \varepsilon + 2\xi^T A_1 \varepsilon + \xi^T A_1 \xi}{\varepsilon^T A_2 \varepsilon}, \tag{2}$$

where  $\tilde{\varepsilon} = \varepsilon + X_2(\beta_2 - r)$ ,  $A_1 = M_1 X_2 D^{-1} X_2^T M_1$ ,  $M_1 = I - X_1(X_1^T X_1)^{-1} X_1^T$ ,  $D = X_2^T M_1 X_2$ ,  $\xi = X_2(\beta_2 - r)$ , and  $A_2 = I - X(X^T X)^{-1} X^T$ , where  $M_1 X_1 = 0$  and  $A_1 A_2 = 0$  and  $M_1$ ,  $A_1$  and  $A_2$  are all symmetric idempotent matrices with ranks equal to  $n - (p - q)$ ,  $q$ , and  $n - p$ , respectively. The expression of  $\lambda$  in (2) will be useful in further analysis.

Classical inference results are often derived assuming  $\varepsilon \sim N_n(0, \sigma^2 I_n)$ . Under the normality assumption,  $\lambda$  is a monotone function of the likelihood ratio (LR) statistic,  $\varepsilon^T A_2 \varepsilon \sim \sigma^2 X_{n-p}^2$  and  $\varepsilon^T A_1 \varepsilon + 2\xi^T A_1 \varepsilon + \xi^T A_1 \xi \sim \sigma^2 X_q^2(\delta^2)$  where  $\delta^2 = \xi^T A_1 \xi$  is the noncentrality parameter. Under  $H_0$  ( $\delta = 0$ ),  $(n - p)\lambda/q \sim F_{q, n-p}$  and in the noncentral case  $(n - p)\lambda/q \sim F_{q, n-p}(\delta^2)$ . As pointed out by [Rao \(1965, Section 7e.1\)](#), the spherical normality assumption may not be plausible in real applications. Many practitioners have encountered data or residual plots with heavier tails than those from a true normal population. Most often there is usually no scientifically sound reason to treat extreme observations as outliers and discard them from the data. Alternatively one may model the underlying population by some heavy tailed distributions. For example [Ullah and Zinde-Walsh \(1984\)](#) and [Zellner \(1976\)](#) studied regression models with multivariate

$t$  errors. [Anderson and Fang \(1982a\)](#) studied a regression model with spherically distributed errors and showed that if  $\varepsilon$  is spherically distributed, then for any scale invariant statistics,  $\psi(\varepsilon)$ , satisfying  $\psi(a\varepsilon) = \psi(\varepsilon)$  for any scalar  $a > 0$ , the distribution of  $\psi(\varepsilon)$  is the same as if  $\varepsilon \sim N_n(0, \sigma^2 I_n)$ . In the previous example under  $H_0$ ,  $\beta_2 = R\beta = r$ ,  $\tilde{\varepsilon} = \varepsilon$  and hence  $\lambda = (\varepsilon^T A_1 \varepsilon) / (\varepsilon^T A_2 \varepsilon)$  is scale invariant. Therefore, as long as  $\varepsilon$  has a spherical distribution,  $(n - p)\lambda/q \sim F_{q, n-p}$  as in the normal case ([Fang et al., 1990](#), pp. 54–55). This nice distributional invariance property, however, does not hold when

$H_0$  is not true. Specifically under  $H_a$ ,  $\tilde{\varepsilon}(\neq \varepsilon)$  is nonspherical and  $\lambda$  is no longer scale invariant. In general the distribution of  $\lambda$  and particularly the power of tests based on  $\lambda$  are both dependent on the underlying error distribution.

In this article, we study the regression model,  $Y = X\beta + \varepsilon$ , with  $\varepsilon$  being spherically distributed. We aim to find possible characteristics of the error distribution which may have direct effect on the power behavior of  $\lambda$ . In Section 2, previous related results on symmetric distributions are summarized and the distribution of  $\lambda$  under  $H_a$  is derived. In Section 3, a simplified example which motivates further analysis is discussed first. Then we derive the relationship between the power and the dispersion and kurtosis of the error distribution. Illustrative examples are provided in Section 4. Concluding remarks are given in Section 5.

## 2. Regression model with spherical errors

### 2.1. Spherical and elliptical distributions

Spherical distributions can be defined in various ways. Refer to Fang et al. (1990) for a thorough discussion. If the density of a spherical random vector  $Y = (Y_1, \dots, Y_n)^T$  exists, it must have the form of  $g(y^T y)$  where  $g(\cdot)$ , called the density generator, is some nonnegative scalar function. Specifically the univariate function  $C_n g(y^T y)$  defines a multivariate density of  $Y$  such that

$$C_n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g\left(\sum_{i=1}^n y_i^2\right) dy_1 \dots dy_n = C_n \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^{\infty} t^{n/2-1} g(t) dt = 1, \tag{3}$$

where

$$C_n = \frac{\Gamma(n/2)}{2\pi^{n/2} \int_0^{\infty} r^{n-1} g(r^2) dr} \tag{4}$$

is the normalizing constant. We will write  $Y \sim \sigma^2 S_n(g)$  to denote that  $Y$  belongs to the spherical family with the density generator  $g(\cdot)$  with  $E(Y) = 0$  and  $\text{Cov}(Y) = \sigma^2 I_n$ . The standard normal distribution is in the spherical family with  $C_n = (2\pi)^{-n/2}$  and  $g(t) = \exp(-t/2)$ . Table 3.1 in Fang et al. (1990) lists important subclasses of spherical distributions. Elliptical distributions can be derived from spherical distributions via an affine transformation as  $N_n(\mu, \Sigma)$  can be derived from  $N_k(0, I_k)$ . Specifically we say  $Y$  has an elliptical distribution, denoted by  $Y \sim EC_n(\mu, \Sigma, g)$ , if  $Y = \mu + A^T Z$  where  $Z \sim S_k(g)$ ,  $\mu : n \times 1$ ,  $A : k \times n$  and  $A^T A = \Sigma$  with  $\text{rank}(\Sigma) = k$  ( $k \leq n$ ). The density of  $Y \sim EC_n(\mu, \Sigma, g)$  (if it exists) is of the form

$$C_n |\Sigma|^{-1/2} g\{(y - \mu)^T \Sigma^{-1} (y - \mu)\}. \tag{5}$$

It should be mentioned that for members in the spherical family, the coordinates are uncorrelated but usually dependent unless it is normal (Fang et al., 1990, p. 106). The implication is that by extending the distributional assumption from normality to

spherical symmetry, the independence assumption is automatically dropped. Exchangeability is another way to extend the independence assumption. Spherical symmetry implies exchangeability but the converse is not true.

Several noncentral sampling distributions derived from elliptical distributions are discussed in Fang and Zhang (1990). For the present article the distribution of  $\lambda$  under the alternative hypothesis is related to the generalized  $F$ (GF) distribution which is defined as follows. If  $Y \sim EC_{m+n}(\mu, I_{m+n}, g)$  with  $\mu = (v^T, 0^T)^T: (m+n) \times 1$  and  $v: m \times 1$  is a vector of constants, then

$$U = \frac{U_1}{U_2} = \frac{Y_{(1)}^T Y_{(1)}/m}{Y_{(2)}^T Y_{(2)}/n} \sim GF_{m,n}(\delta^2, g), \tag{6}$$

where  $Y_{(1)} = (Y_1, \dots, Y_m)^T$ ,  $Y_{(2)} = (Y_{m+1}, \dots, Y_{m+n})^T$  and  $GF_{m,n}(\delta^2, g)$  is called the *generalized noncentral F* distribution with noncentrality parameter  $\delta^2 = v^T v$ . The p.d.f. of  $U \sim GF_{m,n}(\delta^2, g)$  has been derived by Fan (Theorem 2.9.5, Fang and Zhang, 1991) and is given by

$$f(u, g) = Cu^{(m-2)/2}(1+u)^{-(m+n)/2} \int_0^\infty \int_0^\pi \sin^{m-2}\theta y^{m+n-1} \times g(y^2 - 2\delta_1 y \cos \theta + \delta^2) d\theta dy, \tag{7}$$

where  $C = \frac{2mC_{m+n}(\delta^2)^{(m+n-1)/2}}{n\Gamma(m-1/2)\Gamma(n/2)}$ ,  $C_{m+n}$  can be computed using (4),  $u_1 = (m/n)u$ ,  $\delta_1 = \sqrt{u_1/(1+u_1)}\delta$ . Note that  $Y_{(1)}^T Y_{(1)}$  has the noncentral generalized chi-squared distribution with noncentrality parameter  $\delta^2$  and  $Y_{(2)}^T Y_{(2)}$  has the usual (central) chi-squared distribution (see Fang and Zhang, 1990, for a review). When  $g(t) = \exp(-t/2)$ , the GF distribution reduces to the usual noncentral  $F$  distribution. Note that unlike the ordinary  $F$  distribution, the GF distribution does not require stochastic independence between the numerator and denominator.

### 2.2. The distribution of $\lambda$

Consider the regression model  $Y = X\beta + \varepsilon$  where  $\varepsilon \sim \sigma^2 S_n(g)$ . From (5), the likelihood function of  $\beta$  can be written as

$$L(\beta) = C_n \sigma^{-n} g((Y - X\beta)^T (Y - X\beta) / \sigma^2),$$

where  $C_n$  is defined in (4). If  $g(t)$  is a monotone decreasing function of  $t (> 0)$ ,  $L(\beta)$  is maximized when  $(Y - X\beta)^T (Y - X\beta) / \sigma^2$  is minimized. Note that the monotonicity assumption corresponds to a unimodal distribution. In such a case the maximum likelihood estimator of  $\beta$  also equals the least-squares estimator. The distributions of  $\hat{\beta}$  and  $\hat{\varepsilon} = Y - X\hat{\beta}$  for  $\varepsilon \sim \sigma^2 S_n(g)$  have been derived by Anderson and Fang (1982a, b, c), specifically  $\hat{\beta} = (X^T X)^{-1} X^T Y \sim EC_p(\beta, \sigma^2 (X^T X)^{-1}, g)$ , and  $\hat{\varepsilon} = Y - X\hat{\beta} \sim EC_n(0, \sigma^2 [I_n + X(X^T X)^{-1} X^T], g)$ . When  $g(\cdot)$  is monotone decreasing, the LR statistic for testing  $H_0: R\beta = r$  is given by

$$l = \frac{\sup_{H_0} (\sigma^2)^{-n/2} g((Y - X\beta)^T (Y - X\beta) / \sigma^2)}{\sup (\sigma^2)^{-n/2} g((Y - X\beta)^T (Y - X\beta) / \sigma^2)} = \frac{L(\hat{\beta}_H, \hat{\sigma}_H^2)}{L(\hat{\beta}, \hat{\sigma}^2)} = \left[ \frac{\hat{\sigma}^2}{\hat{\sigma}_H^2} \right]^{n/2},$$

where  $\hat{\sigma}_H^2 = (Y - X\hat{\beta}_H)^\top(Y - X\hat{\beta}_H)/n$ ,  $\hat{\sigma}^2 = (Y - X\hat{\beta})^\top(Y - X\hat{\beta})/n$  and  $\hat{\sigma}_H^2 = (Y - X\hat{\beta}_H)^\top(Y - X\hat{\beta}_H)/n$ . Notice that  $\lambda = 1 - l^{-n/2}$ , the same as in the normal case. Therefore, the test based on  $\lambda$  is equivalent to the LR test and hence possesses the optimality property. As described earlier as  $\varepsilon \sim \sigma^2 S_n(g)$ ,  $(n - p)\lambda/q \sim F_{q,n-p}$  under  $H_0$ . This implies that when  $g(\cdot)$  is monotone decreasing, the LR statistic is also distributional invariant under the null. The statement is consistent with the result derived by Ullah and Zinde-Walsh (1984) who showed that the LR statistic is robust for multivariate  $t$  errors whose density generator is a monotone function.

Under  $H_a$  the distribution of  $\lambda$  depends on the density generator  $g(\cdot)$ . The following lemma, analogous to the normal sampling theory, states that the (noncentral) GF distribution can be derived from a ratio of two quadratic forms of spherical distributions.

**Lemma 1.** *Suppose that  $Y \sim EC_n(0, I_n, g)$  has the fourth moment and continuous positive density, then*

$$U = \frac{(Y^\top A_1 Y + 2b_1^\top A_1 Y + b_1^\top A_1 b_1)/n_1}{Y^\top A_2 Y/n_2} \sim GF_{n_1, n_2}(\delta^2, g) \tag{8}$$

if and only if  $\delta^2 = b_1^\top A_1 b_1$ ,  $A_i$  ( $i = 1, 2$ ) are symmetric matrices satisfying  $A_i^2 = A_i$ ,  $\text{rank}(A_i) = n_i$ ,  $A_i b_i = b_i$ ,  $A_1 A_2 = 0$ .

**Theorem 1.** *For the regression model,  $Y = X\beta + \varepsilon$ , with  $\varepsilon \sim \sigma^2 S_n(g)$ , then for the test statistics in (1),*

$$\frac{n - p}{q} \lambda = \frac{n - p}{q} \frac{\varepsilon^\top A_1 \varepsilon + 2\xi^\top A_1 \varepsilon + \xi^\top A_1 \xi}{\varepsilon^\top A_2 \varepsilon} \sim GF_{q, n-p}(\delta^2, g),$$

where  $A_1, A_2$  and  $\xi$  are defined in (2) and  $\delta^2 = \xi^\top A_1 \xi$  is the noncentrality parameter.

The proof of Lemma 1 is given in Appendix A. Theorem 1 can be proved by applying Lemma 1 to (2). Note that in the normal case with  $g(t) = \exp(-t/2)$ ,  $U \sim F_{n_1, n_2}(\delta^2)$ . Testing  $H_0 : R\beta = r$  is equivalent to testing  $H_0 : \delta = 0$  since  $A_1$  is positive definite this implies that  $\xi^\top A_1 \xi = 0$  if and only if  $\delta = 0$ . In fact when  $\delta = 0$ , the GF distribution reduces to the ordinary  $F$  distribution for all  $g(\cdot)$ . Under  $H_a$ ,  $(n - p)\lambda/q \sim GF_{q, n-p}(\delta^2, g)$  and the power function of the GF test statistic is given by

$$\mathcal{P}(\delta, g) = \int_{C_x}^{\infty} f(u, \delta, g) du = 1 - \int_0^{C_x} f(u, \delta, g) du, \tag{9}$$

where  $f(u, \delta, g)$  is the density of  $GF_{q, n-p}(\delta^2, g)$  in (7) and  $C_x$  is the critical value satisfying  $\int_{C_x}^{\infty} f(u, 0, g) du = \alpha$  where  $\alpha$  is the confidence level. Note that the value of  $C_x$  is the same for all  $g(\cdot)$  and can be found from the ordinary  $F$  table.

### 3. Power analysis

#### 3.1. A simple motivating example

To assess which characteristics of the error distribution affect the power behavior, consider a simplified model,  $Y_i = \tilde{\beta} + \varepsilon_i$  ( $i = 1, \dots, n$ ), where  $\tilde{\beta}$  is a scalar and  $(\varepsilon_1, \dots, \varepsilon_n) \sim \sigma^2 S_n(g)$ . To test  $H_0: \tilde{\beta} = 0$  vs.  $H_a: \tilde{\beta} \neq 0$ , the test statistic in (1) reduces to

$$\begin{aligned} \lambda(\tilde{\beta}) &= \frac{\sum_{i=1}^n Y_i^2 - \sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= A + B\Delta, \end{aligned} \tag{10}$$

where

$$\begin{aligned} A &= \frac{\sum_{i=1}^n \varepsilon_i^2 - \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2} + \frac{n\tilde{\beta}^2}{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2} > 0, \\ B &= \frac{\left| 2\tilde{\beta} \sum_{i=1}^n \varepsilon_i \right|}{\sqrt{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}} \frac{1}{\sqrt{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}} > 0 \end{aligned}$$

and  $\Delta = 1$  if  $\tilde{\beta}\bar{\varepsilon} > 0$  and  $\Delta = -1$  if  $\tilde{\beta}\bar{\varepsilon} < 0$ ,  $\bar{Y} = \sum_{i=1}^n Y_i/n$ ,  $\varepsilon_i = Y_i - \tilde{\beta}$ ,  $\bar{\varepsilon} = \sum_{i=1}^n \varepsilon_i/n$ . The power of  $\lambda(\tilde{\beta})$  equals  $\text{pr}(\lambda(\tilde{\beta}) > C_\alpha)$  where the critical value  $C_\alpha$  satisfies  $\text{pr}(\lambda(0) > C_\alpha) = \alpha$  and  $\alpha$  is the level of the test. The first component of  $A$  is scale invariant and the second term increases as the second sample moment decreases. The term  $B\Delta$ , on the average, is zero and asymptotically is of order  $O_p(n^{-1/2})$ . Therefore when  $n$  is large or  $\tilde{\beta}$  is far from zero, the effect of  $B\Delta$  is negligible and the value of  $\lambda(\tilde{\beta})$  is more affected by the second moment of  $\varepsilon$  in a monotonic way. Specifically when the second moment increases,  $\lambda(\tilde{\beta})$  tends to decrease which pulls down the power.

If  $\tilde{\beta}$  is close to zero and the sample size is moderate, the effect of  $B\Delta$  cannot be neglected. Notice that  $B$  is a product of a scale invariant ratio (the first component) and the reciprocal of the second moment of  $\varepsilon$  (the second component). Hence as the second moment increases,  $B$  tends to decrease. However the sign of  $\Delta$  can be positive or negative. Although by symmetry  $\text{pr}(\Delta = 1) = \text{pr}(\Delta = -1)$ , we find that since rejection region of the  $\lambda$  test is one sided, the sign of  $\bar{\varepsilon}$  or  $\Delta$  affects the power in an asymmetric way. To see this, without loss of generality suppose that  $A$  and  $B$  are fixed and  $\text{pr}(\Delta = 1) = \text{pr}(\Delta = -1) = 1/2$ . Notice that  $\text{pr}(\lambda(\tilde{\beta}) > C_\alpha) = \text{pr}(B > C_\alpha - A) \text{pr}(\Delta = 1) + \text{pr}(B < A - C_\alpha) \text{pr}(\Delta = -1)$ . When  $C_\alpha < A$ ,  $\text{pr}(B > C_\alpha - A) = 1$  but  $\text{pr}(B < A - C_\alpha)$  may be 0 or 1. When  $C_\alpha > A$ ,  $\text{pr}(B > C_\alpha - A) \geq 0$  but  $\text{pr}(B < A - C_\alpha) = 0$ . In both cases, the condition of  $\Delta = 1$ , corresponding to  $\bar{\varepsilon} > 0$ , has more contribution to the power. To sum up, in the simplified example the increase in the second moment tends to decrease the value of  $\lambda(\tilde{\beta})$  and hence decrease the power.

The above analysis can be extended to the general expression in (2). Specifically let  $A = (\varepsilon^T A_2 \varepsilon)^{-1} (\varepsilon^T A_1 \varepsilon + \delta^2)$ ,  $B = 2 \left| (\varepsilon^T A_2 \varepsilon)^{-1} (\xi^T A_1 \varepsilon) \right|$ ,  $\Delta = 1$  if  $\xi^T A_1 \varepsilon > 0$  and  $\Delta = -1$

if  $\xi^T A_1 \varepsilon < 0$ . It can be seen that the first component of  $A$  is scale invariant and  $\varepsilon^T A_2 \varepsilon$  also reveals a negative effect on the second component. According to the proof of Lemma 1,  $\xi^T A_1 \varepsilon$  can be written as  $\sum_{i=1}^g \xi_i^* X_i$  where  $X = (X_1, \dots, X_n)^T = \Gamma_1^T \varepsilon$ ,  $\xi^* = (\xi_1^*, \dots, \xi_n^*)^T = \Gamma_1^T \xi$  and  $\Gamma_1$  is an  $n \times n$  orthogonal matrix satisfying

$$\Gamma_1^T A_1 \Gamma_1 = \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that because  $X$  is also spherical,  $\text{pr}(\Delta = 1) = \text{pr}(\Delta = -1) = \frac{1}{2}$ . As in the simplified case, the increase in  $\varepsilon^T A_2 \varepsilon$  tends to decrease the value of  $\lambda$  and hence the power. The phenomenon seems to support the conjecture that the  $\lambda$  test would have bad power if the errors were generated from a heavy-tailed distribution with large second moment. The relationship between the power behavior and heaviness of tails is explored in more detail in the next subsection.

### 3.2. Dispersion, kurtosis, and power

Let  $Z \sim S_n(g)$ . The second moment, known as the dispersion parameter, of  $Z$  is defined as  $\rho = E(Z^T Z)/n$ . Kurtosis is usually used to measure tail's heaviness and according to Mardia (1970) it is defined as  $\kappa = E(Z^T Z)^2/n(n+2) - 1$ . Using the special property that all the marginal distributions of a spherical distribution are also spherical with the same generator up to a multiplicative constant, by simple algebra it follows that

$$\rho = \frac{1}{n} \sum_{i=1}^n E(Z_i^2) = C_1 \int_{-\infty}^{\infty} z^2 g(z^2) dz = C_1 \int_0^{\infty} t^{1/2} g(t) dt \tag{11}$$

and

$$\begin{aligned} \kappa &= \frac{1}{n(n+2)} \left\{ \sum_{i=1}^n E(Z_i^4) + \sum_{i \neq j} E(Z_i^2 Z_j^2) \right\} - 1 \\ &= \frac{C_1}{n+2} \int_{-\infty}^{\infty} z^4 g(z^2) dz + \frac{(n-1)C_2}{n+2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_1^2 z_2^2 g(z_1^2 + z_2^2) dz_1 dz_2 - 1 \\ &= \frac{C_1}{n+2} \int_0^{\infty} t^{3/2} g(t) dt + \frac{(n-1)C_2}{n+2} \int_0^{\infty} \int_0^{\infty} s^{1/2} t^{1/2} g(s+t) ds dt - 1, \tag{12} \end{aligned}$$

where  $C_1$  and  $C_2$  are the normalizing constants satisfying

$$C_1 \int_{-\infty}^{\infty} g(z^2) dz = 1, \quad C_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(z_1^2 + z_2^2) dz_1 dz_2 = 1.$$

The following theorem establishes the relationship between  $\rho, \kappa$  and  $\mathcal{P}(\delta, g)$  within a family which shares the same density generator indexed by a parameter, denoted as  $\tau$ .

**Theorem 2.** Consider the regression model,  $Y = X\beta + \varepsilon$ , where  $\varepsilon \sim \sigma^2 S_n(g)$ . Assume that the density generator  $g(t) \equiv g(t, \tau)$  is a continuous, differentiable and monotone function of some parameter  $\tau$  for all  $t > 0$ . Then

$$\text{sign}\left(\frac{\partial g(t, \tau)}{\partial \tau}\right) = \text{sign}\left(\frac{\partial \rho(\tau)}{\partial \tau}\right) = \text{sign}\left(\frac{\partial \kappa(\tau)}{\partial \tau}\right) = -\text{sign}\left(\frac{\partial \mathcal{P}(\delta, g, \tau)}{\partial \tau}\right),$$

where  $\mathcal{P}(\delta, g, \tau)$ ,  $\rho$  and  $\kappa$  are defined in (9), (11), (12), respectively.

Theorem 2 states that if  $\tau$  affects  $g(t, \tau)$  in a monotonic way, it also affects  $\rho$  and  $\kappa$  in the same direction and  $\mathcal{P}(\delta, g, \tau)$  in the opposite direction. It is worth noting to mention that the converse of Theorem 2 may not be true. That is if  $\mathcal{P}(\tau)$  or  $\kappa(\tau)$  is a monotonic function of  $\tau$ , there is no guarantee that  $\tau$  has a monotonic effect on  $g(t, \tau)$  and  $\mathcal{P}(\delta, g, \tau)$  if  $g(t, \tau)$  is not a monotone function of  $\tau$  for all  $t > 0$ . Theorem 2 does not allow one to compare  $\mathcal{P}(\delta, g)$  across different families of  $g(\cdot)$  simply based on the magnitude of  $\rho$  or  $\kappa$ . In other words, the dispersion and kurtosis are not sufficient to predict the power behavior of  $\lambda$ .

#### 4. Illustrative examples

We select some families of  $g(\cdot)$  from Fang et al. (1990, pp. 69–93) and compute the corresponding forms of  $\rho$  and  $\kappa$ . Figs. 1–3 show one-dimensional and two-dimensional density plots of these families. The power function of  $\text{GF}_{q,n-p}(\delta^2, g)$  is obtained using formula (9) where  $f(u, \delta, g)$  is given in (7). Fan (1984) derived analytic forms of the GF density for these families by expanding  $g(t)$  as the sum of an infinite series, if it converges everywhere, exchanging the summation and integration in (7) and then deriving the integration analytically. We found that Fan’s derivations had some typographical errors because the original formula of  $f(u, g)$  was in error. Here we present the corrected version which is then used to obtain analytic expressions of the corresponding power functions.

**Example 1.** The Kotz-type distribution,  $\text{Kotz}(a, \gamma)(a > -n/2, \gamma > 0)$ , is defined via the generating function  $g(t) = t^a \exp(-\gamma t)$  with  $C_n = \Gamma(n/2)\gamma^{n/2+a}/\pi^{n/2}\Gamma(n/2 + a)$ . It is easy to show that for a fixed  $a \leq 0$ ,  $g(t, \gamma)$  is decreasing in  $\gamma$  for  $t \geq 0$ .

Applying formula (11) and properties of the Gamma function, it can be shown that

$$\rho = C_1 \int_0^\infty t^{a+1/2} \exp(-\gamma t) dt = \frac{\Gamma(a + 3/2)}{\gamma \Gamma(a + 1/2)}.$$

Based on (12) it follows that for each  $i = 1, \dots, n$

$$E(Z_i^4) = C_1 \int_0^\infty t^{3/2+a} \exp(-\gamma t) dt = \frac{\Gamma(a + 5/2)}{\gamma^2 \Gamma(a + 1/2)}$$



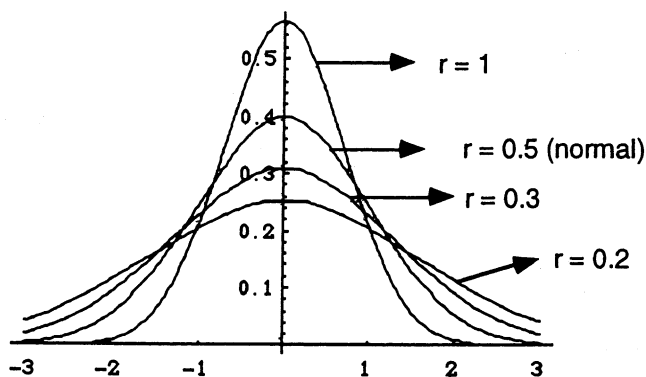


Fig. 1. One-dimensional density of Kotz( $r$ ).

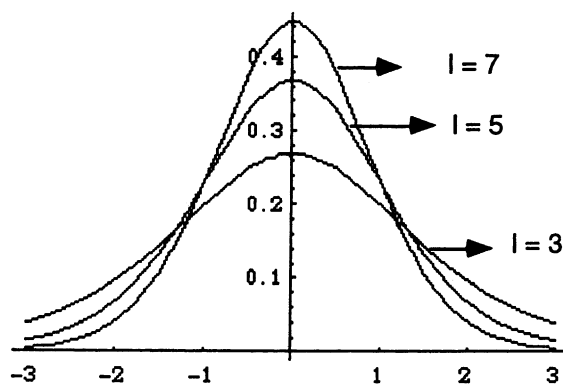


Fig. 2. One-dimensional density of MPVII( $l, h = 10$ ).

and by applying the formula of a Binomial series, it follows that for  $i \neq j$

$$\begin{aligned}
 E(Z_i^2 Z_j^2) &= C_2 \int_0^\infty \int_0^\infty s^{1/2} t^{1/2} (s+t)^a \exp\{-\gamma(s+t)\} ds dt \\
 &= 2C_2 \sum_{k=0}^\infty \binom{a}{k} \int_0^\infty t^{a+1/2-k} \exp(-\gamma t) \left[ \int_0^t s^{1/2+k} \exp(-\gamma s) ds \right] dt.
 \end{aligned}$$

The above expression, which involves integrating an incomplete Gamma function (i.e. the term in the bracket), does not reduce to a nice explicit formula. Nevertheless it still can be seen that  $g(t, \gamma)$ ,  $\rho(\gamma)$  and  $\kappa(\gamma)$  are all decreasing functions of  $\gamma$ . Also it can be shown that  $\partial \mathcal{P}(\delta, \gamma) / \partial \gamma$  equals

$$\begin{aligned}
 &\gamma \int_0^{C_z} \int_0^\infty \int_0^\pi C u_1^{(m-2)/2} (1+u_1)^{-(m+n)/2} \sin^{m-2} \theta y^{m+n-1} \\
 &\quad \times (y^2 - 2\delta_1 y \cos \theta + \delta^2)^{a+1} \exp\{-\gamma(y^2 - 2\delta_1 y \cos \theta + \delta^2)\} d\theta dy du. \quad (13)
 \end{aligned}$$

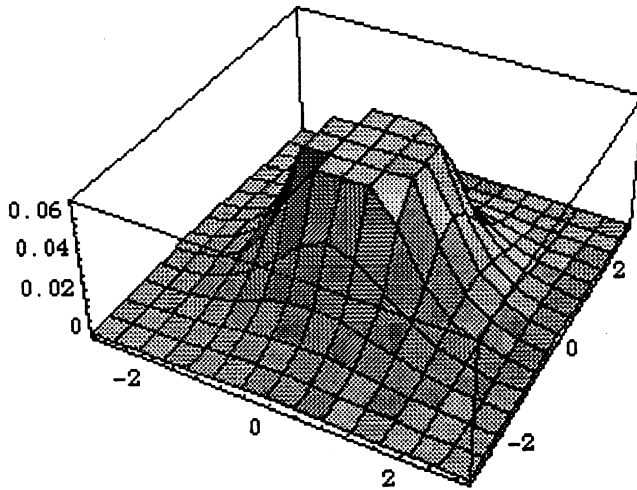


Fig. 3. Two-dimensional density of MPVII( $h = 1, l = 2$ ).

Because  $y^2 - 2\delta_1 y \cos \theta + \delta^2 > 0$ , the integrand of the above equation is positive and hence  $\partial \mathcal{P}(\delta, \gamma) / \partial \gamma > 0$ .

When  $a = 0$ , the Kotz family becomes a scale change of the normal distribution and more explicit results can be derived. It can be shown that  $\rho = 1/(2\gamma)$ ,  $E(Z_i^4) = 3/(4\gamma^2)$ ,

$$E(Z_i^2 Z_j^2) = C_2 \int_0^\infty \int_0^\infty s^{1/2} t^{1/2} \exp\{-\gamma(s+t)\} ds dt = C_2 \left( \frac{\Gamma(3/2)}{\gamma^{3/2}} \right)^2 = \frac{1}{4\gamma^2}$$

and hence  $\kappa = 3/4\gamma^2(n+2) + n - 1/4\gamma^2(n+2) - 1 = 1/4\gamma^2 - 1$ . The corrected formula of  $GF_{q,s}(\delta, \exp(-\gamma t))$ , based on Fan's original derivation, is given by

$$f(u, \delta) = \frac{q}{s} C_a u_1^{(q-2)/2} (1+u_1)^{-(q+s)/2} \exp(-\gamma\delta^2) {}_1F_1 \left( \frac{q+s}{2}, \frac{q}{2}, r\delta_1^2 \right),$$

where  $s = n - p$ ,  $C_x = \Gamma((q+s)/2) / \Gamma(q/2)\Gamma(s/2)$ ,  $u_1 = (q/s)u$ ,  $\delta_1 = \sqrt{[u_1/(1+u_1)]\delta}$ ,  ${}_1F_1(a, b, z) = \sum_{k=0}^\infty (a)_k z^k / (b)_k k!$  is a confluent hypergeometric function with  $(a)_k = a(a+1) \cdots (a+k-1) = \Gamma(a+k) / \Gamma(a)$ . Therefore one can write

$$\mathcal{P}(\delta, \gamma) = 1 - \sum_{k=0}^\infty C_a \frac{(\gamma\delta^2)^k ((q+s)/2)_k}{K!(q/2)_k} \exp(-\gamma\delta^2) S(k),$$

where

$$S(k) = \int_0^{C_x} \frac{q}{s} u_1^{(q-2)/2+k} (1+u_1)^{-[(q+s)/2]-k} du,$$

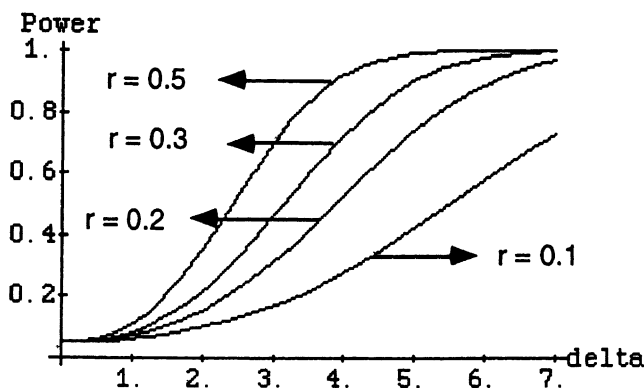


Fig. 4. Power of  $GF_{3,95}(\exp(-rt))$ .

which does not depend on  $\gamma$ . Note that unlike (13) directly checking the sign of  $\partial \mathcal{P}(\delta, \gamma) / \partial \gamma$  using the above explicit formula is very complex.

The power function of  $GF_{3,95}(\delta, \exp(-\gamma t))$  is depicted in Fig. 4 using the numerical integration function of *Mathematica*. The picture also shows that  $\mathcal{P}(\delta, \gamma)$  is an increasing function of  $\gamma$ . Note that using the formula of Binomial series again,  $S(k)$  can be expressed as the following analytic formula:

$$\begin{aligned}
 S(k) &= \sum_{i=0}^{\infty} \binom{-\frac{q+s}{2} - k}{i} \frac{1}{\frac{q-2}{2} + k + i + 1} \{C_{\alpha}^*\}^{\frac{q-2}{2} + k + i + 1} \quad \text{if } C_{\alpha}^* < 1 \\
 &= \sum_{i=0}^{\infty} \binom{-\frac{q+s}{2} - k}{i} \frac{1}{\frac{q-2}{2} + k + i + 1} \\
 &\quad + \sum_{i=0}^{\infty} \binom{-\frac{q+s}{2} - k}{i} \frac{1}{1 - (i + \frac{s+2}{2})} \left\{ \{C_{\alpha}^*\}^{-(i + \frac{s+2}{2}) + 1} - 1 \right\} \\
 &\quad \text{if } C_{\alpha}^* \geq 1,
 \end{aligned}$$

where  $C_{\alpha}^* = qC_{\alpha}/s$ .

**Example 2.** A Pearson-type VII distribution,  $MPVII(l, h)$ , is defined via the generating function  $g(t) = (1 + t/h)^{-l}$  and  $C_n = \Gamma(l)(\pi h)^{-n/2} / \Gamma(l - n/2) (t > 0, l > n/2, h > 0)$ . This family includes the multivariate  $t$  distribution with  $l = (n + h)/2$  and the multivariate Cauchy distribution with  $h = 1$  and  $l = (n + 1)/2$ . Notice that  $g(t)$  is also a monotone decreasing function of  $t$ .

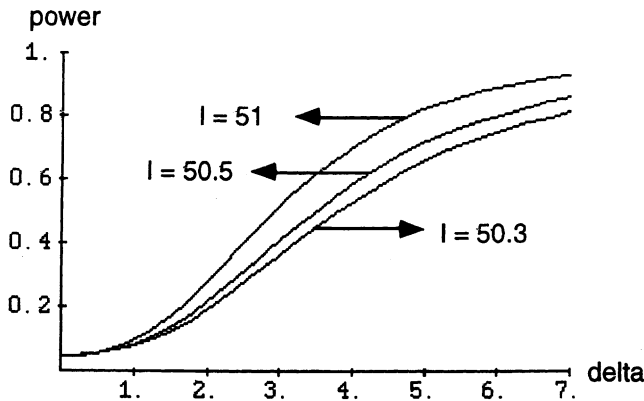


Fig. 5. Power of  $GF_{3,95}(MPVII(h=5, l))$ .

Using properties of the beta function, it follows that

$$\rho = E(Z_i^2) = C_1 h^{3/2} B(3/2, l - 3/2) = \frac{h}{2(l - 3/2)} \quad \text{if } l > 3/2$$

and

$$\begin{aligned} E(Z_i^4) &= C_1 \int_0^\infty t^{3/2} (1 + t/h)^{-l} dt = C_1 h^{5/2} \int_0^\infty s^{3/2} (1 + s)^{-l} ds \\ &= \frac{\Gamma(l)(\pi h)^{-1/2}}{\Gamma(l - 1/2)} h^{5/2} B\left(\frac{5}{2}, l - \frac{5}{2}\right) \\ &= \frac{3h^2}{4} \frac{\Gamma(l - 5/2)}{\Gamma(l - 1/2)} = \frac{3h^2}{4} \frac{1}{(l - 3/2)(l - 5/2)} \quad \text{if } l > 5/2. \end{aligned}$$

$$\begin{aligned} E(Z_i^2 Z_j^2) &= C_2 \int_0^\infty \int_0^\infty \sqrt{st} \left\{ 1 + \frac{s+t}{h} \right\}^{-l} ds dt \\ &= C_2 h^3 B\left(\frac{3}{2}, l - \frac{3}{2}\right) \int_0^\infty s^{1/2} (1 + s)^{-l+3/2} \\ &= \frac{h^2 \Gamma(l - 3)}{4\Gamma(l - 1)} = \frac{h^2}{4(l - 2)(l - 3)} \quad \text{for } l > 3. \end{aligned}$$

Hence the kurtosis is given by

$$\kappa = \frac{h^2}{4(n + 2)} \left[ \frac{3n}{(l - 3/2)(l - 5/2)} + \frac{n(n - 1)}{(l - 2)(l - 3)} \right] - 1 \quad (l > 5/2, h > 0).$$

Note that  $g(t)$ ,  $\rho$  and  $\kappa$  are all increasing in  $h$  and decreasing in  $l$ . Theorem 2 implies that  $\mathcal{P}(\delta, g)$  is increasing in  $l$  (as shown in Fig. 5) and decreasing in  $h$ .

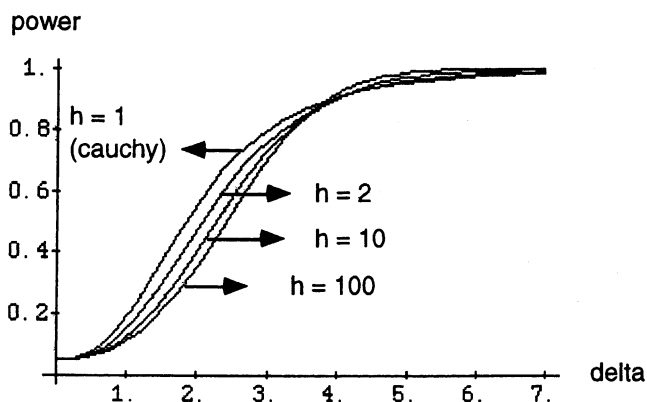


Fig. 6. Power of  $GF_{3,95}(MT)$ .

The corrected formula of  $GF_{q,s}(\delta, (1 + t/h)^{-l})$  is given by

$$f(u, \delta) = \frac{q}{s} C_a u_1^{(q-2)/2} (1 + u_1)^{-(q+s)/2} \left( \frac{h}{h + \delta^2} \right)^{l-(q+s)/2} \times {}_2F_1 \left( \frac{q+s}{2}, l - \frac{q+s}{2}, \frac{q}{2}, \frac{\delta_1^2}{h + \delta^2} \right).$$

It follows that  $\mathcal{P}(\delta, g) = 1 - C_a (h/(h + \delta^2))^{l-(q+s)/2} \sum_{k=0}^{\infty} [\delta^{2k} J(k)/(h + \delta^2)^k] S(k)$ , where  $C_a$  and  $S(k)$  are defined earlier,  ${}_2F_1(a, b, c; z) = \sum_{k=0}^{\infty} [(a)_k (b)_k / k! (c)_k] z^k$  is the Gauss-hypergeometric function and  $J(k) = ((q + s/2)_k (l - (q + s)/2)_k / [k! (q/2)_k])$ .

**Example 3.** The multivariate  $t$  distribution,  $MT(h)$ , is defined via the generating function  $g(t) = (1 + t/h)^{-(h+n)/2}$  where  $h$  is the degree-of-freedom parameter. Recall the multivariate  $t$  distribution is a special case of the Pearson type VII distribution with  $l = (h + n)/2$ . By simple algebra

$$E(Z_i^4) = \frac{3h^2}{4} \frac{\Gamma(h/2 - 2)}{\Gamma(h/2)} = \frac{3h^2}{(h - 2)(h - 4)}$$

and

$$E(Z_i^2 Z_j^2) = C_2 \int_0^{\infty} \int_0^{\infty} \sqrt{st} \left( 1 + \frac{s+t}{h} \right)^{-(h+2)/2} ds dt = \frac{h^2}{4} \frac{\Gamma(h/2 - 2)}{\Gamma(h/2)} = \frac{h^2}{(h - 2)(h - 4)}.$$

Therefore  $\rho = E(Z_i^2) = h/(h - 2)$  for  $h > 2$ , and  $\kappa = h^2/(h - 2)(h - 4) - 1$  for  $h > 4$ . Notice that both of  $\rho$  and  $\kappa$  are monotone decreasing functions of the degree-of-freedom

parameter,  $h$ . However  $g(t)$  is not a monotone function of  $h$  since the sign of the function,  $\partial g(t, h) / \partial h = g(t, h) \{-\frac{1}{2} \log(1 + t/h) + (h + n)/2t/(h + th)\}$ , depends on the value of  $t$ . Therefore Theorem 2 cannot be applied. In fact Fig. 6 shows that the power function of  $GF_{3,95}(MT(h))$  is not monotone in  $h$ . Surprisingly for small  $\delta (< 4)$ , the power is even higher for smaller value of  $h$  which correspond to heavier tails. In the case of  $h=1$  where the underlying distribution is Cauchy which has no finite moments, the power is the highest for  $\delta < 4$ .

## 5. Discussion

We have seen that  $\lambda$  is affected by  $\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2$  (in the simplified example) or by  $\varepsilon^T A_2 \varepsilon$  (in the general case), both of which are random variables and have the (central) generalized chi-squared distributions. When the noncentrality parameter is nonzero, the generalized chi-squared distribution always depends on  $g(\cdot)$ . However it has been shown that the density generator  $g(\cdot)$  affects the power of  $\lambda$  in an intricate way such that  $\rho$  and  $\kappa$  cannot completely determine the power behavior. We also found the interesting phenomenon for the multivariate  $t$  family that in the region of small noncentrality tail heaviness even has a positive effect on the power. Practically it is more difficult to detect the parameter value when it is close to the null hypothesis, thus the  $t$  family even has the advantage over the normal family in terms of the power consideration. We suspect that such a counter-intuitive result may be due to the effect of  $B$  in (10). For example for the Cauchy distribution, no moments exist so that even when  $\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2$  is large the value of  $B$  may still be large which pulls up the power. Recall that when  $g(\cdot)$  is a monotone function, the test statistic  $\lambda$  produces an equivalent test as the LR statistic which possesses some optimality properties. This implies that if  $\lambda$  has poor power, there is no much room to improve by constructing other tests. Note that the discussions in the article are based on finite samples. A referee pointed out that Fang and Yuan (1993) discussed asymptotic behavior of the power for some spherical distributions of the error.

## Acknowledgements

The support of NSC 87-2118 and National Science Foundation Grant DMS-9971586 is gratefully acknowledged. The authors thank Dr. Adam A. Ding and two anonymous referees for their helpful comments.

## Appendix A

### A.1. Proof of Lemma 1

There are two ways to prove the lemma. The first approach is to re-express  $U$  in (6) according to the definition of the GF distribution. Another alternative is to derive

the density of  $U$  from the joint density of  $(U_1, U_2)$  derived by Fan (1986) and show it as the same as the GF density in (7). Here we present the first approach and for the second method, please refer to Wang (1993). Specifically we want to show that

$$U = \frac{Y^T A_1 Y + 2b_1^T A_1 Y + b_1^T A_1 b_1}{Y^T A_2 Y} =^d \frac{W_{(1)}^T W_{(1)}/n_1}{W_{(2)}^T W_{(2)}/n_2},$$

where

$$\begin{pmatrix} W_{(1)} \\ W_{(2)} \end{pmatrix} \sim EC_{n_1+n_2} \left( \begin{pmatrix} \xi_{(1)} \\ I_{n_1+n_2} \\ \xi_{(2)} \end{pmatrix} \right),$$

$W_{(i)} : n_i \times 1 \xi_{(i)} : n_i \times 1 (i = 1, 2)$ ,  $\xi_{(1)}^T \xi_{(1)} = \delta^2 = c$ ,  $\xi_{(1)}^T \xi_{(2)} = 0$ , There exist orthogonal matrices,  $\Gamma_1$ , and  $\Gamma_2$ , such that

$$\Gamma_i^T A_i \Gamma_i = \begin{pmatrix} I_{n_i} & 0 \\ 0 & 0 \end{pmatrix} \quad (i = 1, 2).$$

Letting  $X = \Gamma_1^T Y$ , it is easy to show that

$$Y^T A_1 Y = Y^T (\Gamma_1^T)^{-1} \Gamma_1^T A_1 \Gamma_1 \Gamma_1^{-1} Y = X^T \begin{pmatrix} I_{n_1} & 0 \\ 0 & 0 \end{pmatrix} X = \sum_{i=1}^{n_1} X_i^2.$$

Letting  $b_1^* = \Gamma_1^T b_1$ , since  $\Gamma_1^T = \Gamma_1^{-1}$  it follows that

$$\begin{aligned} 2b_1^T Y &= 2b_1^T \Gamma_1 \Gamma_1^{-1} Y = 2b_1^T \Gamma_1 X = 2b_1^T A_1 \Gamma_1 X \\ &= 2b_1^T \Gamma \Gamma^T A_1 \Gamma X = 2b_1^{*T} \begin{pmatrix} I_{n_1} & 0 \\ 0 & 0 \end{pmatrix} X = 2 \sum_{i=1}^{n_1} b_{1i}^* X_i \end{aligned}$$

and  $b_1^T A_1 b_1 = b_1^T \Gamma_1 \Gamma_1^T A_1 \Gamma_1 \Gamma_1^{-1} b_1 = \sum_{i=1}^{n_1} b_{1i}^{*2} = \delta^2$ . Therefore one can write  $U_1 = \sum_{i=1}^{n_1} (X_i + b_{1i}^*)^2 = W_{(1)}^T W_{(1)}$ , where

$$W_{(1)} = \begin{bmatrix} X_{i1} + b_{1i}^* \\ \cdot \\ \cdot \\ \cdot \\ X_{in_1} + b_{1n_1}^* \end{bmatrix} \sim EC_{n_1} \left( \begin{bmatrix} b_{11}^* \\ \cdot \\ \cdot \\ \cdot \\ b_{1n_1}^* \end{bmatrix}, I_{n_1}, g \right).$$

Similarly letting  $Z = \Gamma_2^T Y$ , we get  $Y^T A_2 Y = \sum_{i=1}^{n_2} Z_i^2 = W_{(2)}^T W_{(2)}$  and  $W_{(2)} = (Z_1, \dots, Z_{n_2})^T \sim EC_{n_2}(0, I_{n_2}, g)$ . Then by the definition in (6), we have shown that  $U \sim GF_{n_1, n_2}(\delta^2)$ .  $\square$

*A.2. Proof of Theorem 2*

Since  $g(t, \tau)$  is a monotone differentiable function of  $\tau$ , then one can determine  $\Delta = \text{sign}(\partial g(t, \tau) / \partial \tau)$ . Since  $g$  is a density generator it follows from Lebesgue’s theorem, it follows that

$$\begin{aligned} \frac{\partial \rho(\tau)}{\partial \tau} &= C_1 \int_0^\infty t^{1/2} \frac{\partial g(t, \tau)}{\partial \tau} = \Delta C_1 \int_0^\infty \left| t^{1/2} \frac{\partial g(t, \tau)}{\partial \tau} \right| dt. \\ \frac{\partial \kappa(\tau)}{\partial \tau} &= \frac{C_1}{n+2} \int_0^\infty t^{3/2} \frac{\partial g(t, \tau)}{\partial \tau} dt + \frac{(n-1)C_2}{n+2} \int_0^\infty \int_0^\infty \sqrt{st} \frac{\partial g(s+t, \tau)}{\partial \tau} ds dt \\ &= \Delta \frac{C_1}{n+2} \int_0^\infty \left| t^{3/2} \frac{\partial g(t, \tau)}{\partial \tau} \right| dt \\ &\quad + \Delta \frac{(n-1)C_2}{n+2} \int_0^\infty \int_0^\infty \left| \sqrt{st} \frac{\partial g(s+t, \tau)}{\partial \tau} \right| ds dt. \end{aligned}$$

Since the integrands in the above expressions are positive functions and  $C_j > 0$  ( $j = 1, 2$ ), hence  $\text{sign}(\partial \rho(\tau) / \partial \tau) = \text{sign}(\partial \kappa(\tau) / \partial \tau) = \Delta$ . Similarly

$$\frac{\partial \mathcal{P}(\delta, g, \tau)}{\partial \tau} = - \int_0^{C_x} \frac{\partial f(u, \delta, g, \tau)}{\partial \tau} du$$

with  $f(u, \delta, g, \tau)$  given in (7). It is easy to show that the sign of  $\partial f(u, \delta, g, \tau) / \partial \tau$  is the same as  $\Delta$ . Hence  $\text{sign}(\frac{\partial \mathcal{P}(\delta, g, \tau)}{\partial \tau}) = -\Delta$ .

**References**

Anderson, T.W., Fang, K.T., 1982a. Maximum likelihood estimators and likelihood ratio criteria for multivariate elliptically contoured distributions. Stanford Technical Report, Vol. 1.  
 Anderson, T.W., Fang, K.T., 1982b. Distribution of quadratic forms and Cochran’s theorem for elliptically contoured distributions and their applications. Standard Technical Report, Vol. 53.  
 Anderson, T.W., Fang, K.T., 1982c. On the theory of multivariate elliptical contoured distributions and their applications. Standard Technical Report, Vol. 534.  
 Fan, J., 1984. Generalized non-central  $t$ -,  $F$ - and  $T^2$ - distributions. J. Graduate School 2, 134–148.  
 Fan, J., 1986. Distributions of quadratic forms and non-central Cochran’s theorem. Acta Math. Sinica 2, 185–198.  
 Fang, K.T., Yuan, K.H., 1993. The limiting distribution of some subclasses of the generalized non-central  $t$ -distributions. Acta Math. Appl. Sinica (English Ser.) 9, 71–81.  
 Fang, K.T., Zhang, Y.T., 1990. Generalized Multivariate Analysis. Science Press, Springer, Beijing, Berlin.  
 Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London.  
 Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.  
 Rao, C.R., 1965. Linear Statistical Inference and its Applications. Wiley, New York.



- Rao, C.R., Toutenburg, H., 1995. *Linear models: Least Squares and Alternatives*. Springer, New York.
- Ullah, A., Zinde-Walsh, V., 1984. On the robustness of LM, LR, and W tests in regression models. *Econometrica* 52, 1055–1066.
- Wang, W., 1993. *Linear models and covariance structures for elliptical distributions*. Cornell University Thesis.
- Zellner, A., 1976. Bayesian and non-Bayesian analysis of the regression model with multivariate student- $t$  error terms. *J. Amer. Statist. Assoc.* 71, 400–405.