Brazilian Journal of Probability and Statistics (2001), 15, pp. 49–67.

CORRECTED LIKELIHOOD RATIO TESTS IN CLASS OF SYMMETRIC LINEAR REGRESSION MODELS

Silvia L. P. Ferrari[1] and **Miguel A. Uribe-Opazo**[2]

[1] *Departamento de Estatística, Universidade de São Paulo, Brazil. E-mail: sferrari@ime.usp.br*

[2] *Centro de Ciências Exatas e Tecnológicas, Universidade Estadual do Oeste do Paraná, Brazil*

Summary
In this paper we derive general formulae for Bartlett corrections to likelihood ratio statistics in a class of symmetric linear regression models. This is a wide class of models that has the normal linear regression model as a special case and covers other models with heavy tails. For instance, the t linear regression model, which is commonly used as an alternative to the usual normal regression model when the data contain extreme or outlying observations is another important special case of this class. Simulation results show that the corrected tests perform much better than their uncorrected counterparts in samples of small to moderate sizes. A real data example is presented.

Key Words: Bartlett correction; likelihood ratio test; linear regression; symmetric distribution; t distribution.

1 Introduction

In regression analyses it is a common practice to assume that the observations follow a normal distribution. However, it is well known that the normal distribution is not always suitable for data containing extreme or outlying observations. The t distribution provides a useful extension of the normal distribution when a previous analysis of the data indicates the presence of errors that follow a symmetric distribution with longer–than–normal tails. Here, we consider a wide class of symmetric distributions that covers the normal and the t distributions and has some other important distributions as special cases.

In this paper we present Bartlett corrections to likelihood ratio statistics (Lawley, 1956) in linear regression models with errors that follow a symmetric distribution. We generalize the results obtained by Ferrari and Arellano–Valle (1996) who considered a t distribution for the errors. Bartlett corrections for other important classes of regression models are found in Cordeiro (1983, 1987), Cordeiro, Paula and Botter (1994) among others.

In Section 2, we define the class of linear regression models with a symmetric distribution for the errors and show how the maximum likelihood estimates are obtained. It becomes clear that inferences for a number of models in this class are robust in the sense that outlying observations have less weight in the estimation process than the others. In Section 3, we develop Bartlett corrections to likelihood ratio statistics and present some special cases. In Section 4, simulation results are presented and, finally, in Section 5, a real data example is discussed. Technical details are left for an appendix.

2 Symmetric models and maximum likelihood estimation

Let $y_1, \ldots, y_n$ be n independent random variables, each y_l having a continuous symmetric distribution with location parameter $\mu \in \mathbb{R}$ and scale parameter $\phi > 0$, and density function

$$\pi(y_l; \mu_l, \phi) = \frac{1}{\phi} h\left(\left(\frac{y_l - \mu_l}{\phi}\right)^2\right), \quad y \in \mathbb{R}, \tag{2.1}$$

for some positive function $h(.)$ (named generating density function), defined on $\mathbb{R}^+$, and $\int_0^\infty u^{-1/2} h(u) \mathrm{d}u = 1$. This condition guarantees that $\pi(\cdot; \mu_l, \phi)$ is a density function (Fang, Kotz and Ng, 1990). A number of important distributions have density function (2.1) as it is shown in Table 1. This table also shows the coefficient of kurtosis γ_2 for each distribution. It can be seen that the class of distributions considered here covers distributions with larger as well as smaller coefficients of kurtosis than the normal distribution[1]. The class of symmetric distributions defined in (2.1) has been considered by several authors (Kelker, 1970; Chu, 1973; Cambanis, Huang and Simons, 1981). The properties of these distributions have been explored by Muirhead (1980, 1982), Berkane and Bentler (1986), Rao (1990) and Fang, Kotz and Ng (1990). It is easy to find many properties of the symmetric distributions parallel to those of the normal distribution. A review

[1]Note that the coefficient of kurtosis of the generalized logistic distribution does not depend on the parameter α. However, its variance equals $2\psi'(m)\phi^2/\alpha$ which reduces to ϕ^2 if $\alpha = \sqrt{2\psi'(m)}$.

of different areas in which symmetric distributions are applied is given by Chmielewski (1981).

Table 1
Generating density functions and coefficients of kurtosis for some symmetric distributions

DISTRIBUTION	$h(u)$	γ_2
Normal	$h(u)=\frac{1}{\sqrt{2\pi}}\exp\{-u/2\}, u>0$	3
Cauchy	$h(u)=\frac{1}{\pi}(1+u)^{-1}, u>0$	-
t	$h(u)=\frac{\nu^{\nu/2}}{B(1/2,\nu/2)}[\nu+u]^{-\frac{\nu+1}{2}}, \nu>0,$ $u>0,\ \nu>4$	$3+\frac{6}{(\nu-4)},$
Generalized t	$h(u)=\frac{s^{r/2}}{B(1/2,r/2)}(s+u)^{-\frac{r+1}{2}},\ s,r>0,$ $u>0,\ r>4$	$3+\frac{6}{(r-4)},$
Type I Logistic	$h(u)=c\frac{e^{-u}}{(1+e^{-u})^2}, u>0, c\approx 1.484300029$	2.385165
Type II Logistic	$h(u)=\frac{e^{-u^{1/2}}}{(1+e^{-u^{1/2}})^2}, u>0$	4.2
Generalized Logistic	$h(u)=\frac{\alpha}{B(m,m)}\left[\frac{e^{-\alpha\sqrt{u}}}{(1+e^{-\alpha\sqrt{u}})^2}\right]^m,$ $m>0, u>0, \alpha>0$	$3+\frac{\psi'''(m)}{(2\psi(m)^2)}$
Contaminated Normal	$h(u)=(1-\varepsilon)\frac{1}{\sqrt{2\pi}}\exp\{-u/2\}$ $+\varepsilon\frac{1}{\sqrt{2\pi}\sigma}\exp\{-u/(2\sigma^2)\},$ $u>0, \sigma>0, 0\le\varepsilon\le 1$	$3\frac{(1+\varepsilon(\sigma^4-1))}{(1+\varepsilon(\sigma^2-1))^2}$
Power Exponential	$h(u)=C(k)\exp\left\{-\frac{1}{2}u^{1/(1+k)}\right\},$ $-1<k\le 1, u>0$ where $C(k)^{-1}=\Gamma(1+\frac{1+k}{2})2^{1+(1+k)/2}$	$\frac{\Gamma\left(\frac{5(1+k)}{2}\right)\Gamma\left(\frac{1+k}{2}\right)}{\Gamma\left(\frac{3(1+k)}{2}\right)^2}$

Note: $B(.,.)$, $\Gamma(\cdot)$ and $\psi(\cdot)$ are the beta, gamma and digamma functions respectively.

We assume that the parameter vector $\mu=(\mu_1,\ldots,\mu_n)^\top$ follows the linear structure

$$\mu = X\beta, \tag{2.2}$$

where X is an $n\times p$ matrix of known constants with rank$(p)<n$ and $\beta=(\beta_1,\ldots,\beta_p)^\top$ is a set of unknown regression parameters.

Let $L(\theta)$ be the total log–likelihood function for the regression model defined by (2.1) and (2.2) given $y_1, \ldots, y_n$, where $\theta = (\beta^\top, \phi)^\top$. We have

$$L(\theta) = -n \log \phi + \sum_{l=1}^{n} t(z_l), \tag{2.3}$$

where

$$t(z_l) = \log h(z_l^2),$$

with

$$z_l = \frac{y_l - x_l^\top \beta}{\phi} \tag{2.4}$$

representing the standardized error for the l-th observation.

For obtaining the likelihood equations, the information matrix and the Bartlett corrections, one needs to derive the log–likelihood function with respect to the unknown parameters and compute some moments of such derivatives. In the following, we assume that such derivatives and moments exist. Therefore, the symmetric distributions that do not satisfy this condition will be excluded. For instance, for the double exponential distribution, the first derivative of $L(\theta)$ with respect to β does not exist for all $\beta \in \mathbb{R}^p$ and, hence, it will not be considered here. In other cases, the derivatives of $L(\theta)$ exist for all the values of the unknown parameters in the parameter space only if the extra known parameter (for example, k in the power exponential distribution) belongs to a certain interval that will be indicated whenever it is necessary.

The first derivative $\partial L(\theta)/\partial\theta$ of the log–likelihood function given in (2.3) is obtained from

$$\frac{\partial L(\theta)}{\partial \beta_r} = -\frac{1}{\phi} \sum_{l=1}^{n} t_l^{(1)} x_{lr} \quad \text{and} \quad \frac{\partial L(\theta)}{\partial \phi} = -\frac{n}{\phi} - \frac{1}{\phi} \sum_{l=1}^{n} t_l^{(1)} z_l,$$

where z_l is defined in (2.4) and $t_l^{(1)} = \mathrm{d}t(z_l)/\mathrm{d}z_l$, for $l = 1, \ldots, n$, assuming that such derivative exists for all $z_l \in \mathbb{R}$. We have $t_l^{(1)} = -z_l w_l$, where

$$w_l = w_l(\theta) = -2 \frac{\mathrm{d}}{\mathrm{d}u} \log h(u)\big|_{u=z_l^2}. \tag{2.5}$$

The maximum likelihood estimates of β and ϕ are obtained from the likelihood equations

$$\widehat{\beta} = (X^\top \widehat{W} X)^{-1} X^\top \widehat{W} y$$

and

$$\widehat{\phi}^2 = \frac{1}{n} \widehat{e}^\top \widehat{W} \widehat{e},$$

where $\widehat{W} = W(\widehat{\theta}) = \mathrm{Diag}\{\widehat{w}_1, \ldots, \widehat{w}_n\}$ with w_l defined in (2.5) and $\widehat{e} = (y - X\widehat{\beta})$ is the vector of ordinary residuals. The likelihood equations are non–linear in $\widehat{\beta}$ and $\widehat{\phi}^2$ except for the normal model ($w_l = 1$, for $l = 1, \ldots, n$) and hence an iterative algorithm must be used. A simple iterative procedure replaces the above equations by $\widehat{\beta}_{(r+1)} = (X^\top \widehat{W}_{(r)} X)^{-1} X^\top \widehat{W}_{(r)} y$ and $\widehat{\phi}^2_{(r+1)} = n^{-1} \widehat{e}^\top_{(r+1)} \widehat{W}_{(r+1)} \widehat{e}_{(r+1)}$, for $r = 0, 1, 2, \ldots$, where $\widehat{W}_{(r)} = W(\widehat{\theta}_{(r)})$ and $\widehat{e}_{(r)} = (y - X\widehat{\beta}_{(r)})$ with $\widehat{\theta}_{(r)} = (\widehat{\beta}_{(r)}, \widehat{\phi}_{(r)})$ representing the estimate of θ in the r-th step. The procedure may be initialized by taking $\widehat{\beta}_{(0)} = (X^\top X)^{-1} X^\top y$, the ordinary least squares estimate of β, and $\widehat{\phi}^2 = n^{-1}(y - X\widehat{\beta}_{(0)})^\top (y - X\widehat{\beta}_{(0)})$. Standard iterative algorithms, such as the Newton-Raphson, Fisher scoring or EM algorithms, may also be used. The EM algorithm is particularly useful for the t and generalized t distributions (see Lange, Little and Taylor, 1989). The maximum likelihood estimate of β are unbiased up to an error of order n^{-2} (see Cordeiro, Ferrari, Uribe–Opazo and Vasconcellos, 2000).

Note that w_l may be regarded as the contribution of the l-th observation for the estimation of the parameters. Table 2 shows the values of w for some of the distributions given in Table 1. For the normal model all the observations have the same weight in the estimation of β and ϕ. For the Cauchy, t, generalized t, type II logistic, generalized logistic, contaminated normal and power exponential distributions, w_l is a decreasing function of $|z_l| = |y_l - x_l^\top \beta|/\phi$. Hence, the maximum likelihood estimates of β and ϕ are robust in the sense that the observations with large $|z_l|$ have small weight w_l. For the type I logistic distribution, the weights w_l are increasing functions of $|z_l|$. This was expected since this distribution has shorter–than–normal tails.

The information matrix for $(\beta^\top, \phi)$ (see Appendix) is block–diagonal and is given by $K = K(\theta) = \mathrm{Diag}\{K_{\beta,\beta}, K_{\phi,\phi}\}$ with

$$K_{\beta,\beta} = \frac{\delta_{(2,0,0,0,0)}}{\phi^2} X^\top X$$

and

$$K_{\phi,\phi} = \kappa_{\phi,\phi} = \frac{n}{\phi^2}(\delta_{(2,0,0,0,2)} - 1),$$

where

$$\delta_{(a,b,c,d,e)} = E[t^{(1)^a} t^{(2)^b} t^{(3)^c} t^{(4)^d} z^e], \tag{2.6}$$

for a, b, c, d, $e = 0, 1, 2, 3, 4$, with $t^{(r)} = \mathrm{d}^r t(z)/\mathrm{d}z^r$. Therefore, β and ϕ are globally orthogonal parameters and their maximum likelihood estimates are asymptotically uncorrelated.

Table 2
w for some symmetric distributions

DISTRIBUTION	w
Normal	1
Cauchy	$2/(1+z^2)$
t	$(\nu+1)/(\nu+z^2)$
generalized t	$(r+1)/(s+z^2)$
Type I Logistic	$\dfrac{2(1-e^{-z^2})}{1+e^{-z^2}} = 2\tanh(z^2/2)$
Type II logistic	$(e^{\|z\|}-1)/(\|z\|(1+e^{\|z\|}))$
Generalized logistic	$\alpha m(e^{\alpha\|z\|}-1)/(\|z\|(1+e^{\alpha\|z\|}))$
Contaminated normal	$f_1(z^2)/f_0(z^2)$, where $f_i(z^2) = (1-\varepsilon)e^{-z^2/2} + \varepsilon(\sigma^2)^{-1/2-i}e^{-z^2/2\sigma^2}$, $i = 0, 1$.
Power exponential	$1/[(1+k)(z^2)^{k/(1+k)}]$, for $-1 < k < 1$.

3 Likelihood ratio tests and Bartlett corrections

Consider the null hypothesis $H_0 : \beta_1 = \beta_1^{(0)}$ to be tested against the alternative hypothesis $H_1 : \beta_1 \neq \beta_1^{(0)}$, where $\beta_1 = (\beta_1, \ldots \beta_q)^\top$, $q \leq p$, with $\beta_2 = (\beta_{q+1}, \ldots, \beta_p)^\top$ and ϕ representing nuisance parameters. The likelihood ratio statistic for testing H_0 against H_1 is

$$LR = 2(L(\widehat{\theta}) - L(\widetilde{\theta})),$$

where $L(\theta)$ is the log–likelihood function given in (2.3), $\widehat{\theta} = (\widehat{\beta}^\top, \widehat{\phi})^\top$ and $\widetilde{\theta} = (\beta_1^{(0)^\top}, \widetilde{\beta}_2^\top, \widetilde{\phi})^\top$, with $\widetilde{\beta}_2$ and $\widetilde{\phi}$ denoting the restricted maximum likelihood estimates for β_2 and ϕ respectively.

It is well known that, under suitable regularity conditions, LR has a χ^2_q distribution asymptotically under the null hypothesis. However, for small to moderate samples, the asymptotic distribution may deliver poor approximations to the true sizes of the tests. This becomes clear in the simulation results that will be presented later. For this reason, it is convenient to define a modified likelihood ratio statistic whose null distribution is better approximated by the χ_q^2 distribution. This may be achieved by using the Bartlett corrected likelihood ratio statistic (see Lawley, 1956)

$$LR^* = LR\left(1 - \frac{d}{n}\right), \tag{3.1}$$

where d is given in the Appendix and is a function of some cumulants of log-likelihood derivatives. For the class of symmetric linear regression models considered here we obtain (see Appendix),

$$d = \frac{d_0}{q} h_0(X_1, X_2) + d_1 + \left(\frac{2p - q}{2}\right) d_2, \tag{3.2}$$

where

$$d_0 = \frac{\delta_{(0,0,0,1,0)}}{4\delta^2_{(2,0,0,0,0)}}, \quad d_1 = -\frac{m_2 m_3}{2m_1^2} - \frac{2m_3 + m_3^2 + m_4}{2m_1}, \quad d_2 = -\frac{m_3^2}{2m_1}, \tag{3.3}$$

with

$$\begin{aligned} m_1 &= \delta_{(0,1,0,0,2)} - 1, \quad m_2 = 4 - \delta_{(0,0,1,0,3)} - 6\ \delta_{(0,1,0,0,2)}, \\ m_3 &= (\delta_{(0,0,1,0,1)} + 2\ \delta_{(0,1,0,0,0)})/\delta_{(2,0,0,0,0)}, \\ m_4 &= (\delta_{(0,0,0,1,2)} - 6\ \delta_{(1,1,0,0,1)})/\delta_{(2,0,0,0,0)}, \end{aligned}$$

and

$$h_0(X_1, X_2) = \rho_{ZZ} - \rho_{Z_2 Z_2}, \tag{3.4}$$

with $\rho_{Z_2Z_2} = n\text{tr}(Z_{2d}Z_{2d})$, $\rho_{ZZ} = n\text{tr}(Z_d Z_d)$. Here,

$$Z = \{z_{lm}\} = X(X^\top X)^{-1}X^\top, \quad Z_2 = \{z_{2lm}\} = X_2(X_2^\top X_2)^{-1}X_2^\top,$$

where the model matrix X is partitioned as $X = (X_1, X_2)$ with X_1 corresponding to the first q columns of X and X_2 corresponding to the remaining columns. Then Z and Z_2 are $n \times n$ matrices of ranks p and $(p - q)$, respectively. If $p = q$ we set $Z_2 = 0$, where 0 denotes an $n \times n$ null matrix. Note that $Z\phi^2\delta^{-1}_{(2,0,0,0,0)}$ and $Z_2\phi^2\delta^{-1}_{(2,0,0,0,0)}$ are the asymptotic covariance matrices of $X\widehat{\beta}$ and $X_2\widetilde{\beta}_2$, respectively. Also,

$$Z_d = \text{Diag}\{z_{11}, \ldots, z_{nn}\} \quad \text{and} \quad Z_{2d} = \text{Diag}\{z_{211}, \ldots, z_{2nn}\}$$

represent the diagonal matrices obtained from the diagonal elements of Z and Z_2, respectively.

Note that d is a linear combination of d_0, d_1 and d_2, which depend on the distribution assumed for the data through the δ's, with coefficients that depend on the model matrix X (through $h_0(X_1, X_2)$), the number of regression parameters (p), the number of parameters of interest (q) and the sample size (n). A possible motivation for developing closed form formulae for Bartlett corrections to the likelihood ratio statistics is that the formulae usually reveal which aspects of the model contribute to the possible poor reliability of the first order χ^2 approximation. For example, the formula of the Bartlett correction given in (3.2) reveals that the number of nuisance parameters and the difference between the total sums of the squares of the diagonal elements of the projection matrices Z and Z_2 have influence on the goodness of the χ^2 approximation to the likelihood ratio statistic.

We now examine some special linear structures that lead to simplifications in the formula for $h_0(X_1, X_2)$ given in (3.4). If the null hypothesis is $H_0 : \beta = \beta^{(0)}$ we have $p = q$ ($Z_2 = 0$) and hence $h_0(X_1, X_2) = \rho_{ZZ}$. In particular, if β is a scalar parameter ($p = q = 1$) then $h_0(X_1, X_2) = \overline{s}_4/\overline{s}_2^2$, where $\overline{s}_a = \sum_{i=1}^{n} x_i^a/n$, for $a = 2, 4$ and for the i.i.d. case we have $h_0(X_1, X_2) = 1$. For the likelihood ratio test of homogeneity of the location parameters in a one–way classification model, $h_0(X_1, X_2) = n\sum_{i=1}^{p}(1/n_i) - 1$, where $n_1, \ldots, n_p$ are the number of observations in the p independent random samples and $n = \sum_{i=1}^{p} n_i$. If $n_1 = n_2 = \cdots = n_p$, then $h_0(X_1, X_2) = p^2 - 1$. For the simple linear regression model, $h_0(X_1, X_2) = \overline{S}_4/\overline{S}_2^2 + 2$, where $\overline{S}_a = \sum_{i=1}^{n} (x_i - \overline{x})^a/n$, for $a = 2, 4$. Hence the approximation of the null distribution of the likelihood ratio statistic by the χ_1^2 distribution is sensitive to changes in the sample kurtosis $\overline{S}_4/\overline{S}_2^2$ of the covariate.

Suppose now that the null hypothesis $H_0 : \phi = \phi^{(0)}$ is to be tested against $H_1 : \phi \neq \phi^{(0)}$, where β is the nuisance parameter and $\phi^{(0)}$ is a specified value for the dispersion parameter ϕ. The likelihood ratio statistic is

$$LR = 2\{L(\widehat{\beta}, \widehat{\phi}) - L(\widetilde{\beta}, \phi^{(0)})\}.$$

Under H_0, LR has a χ^2 distribution with one degree of freedom. In the Appendix we show that the Bartlett corrected likelihood ratio statistics of H_0 against H_1 is given by (3.1) with

$$d = d_3 + d_1 p + \frac{d_2 p^2}{2}, \tag{3.5}$$

where

$$\begin{aligned} d_3 \;=\; & \frac{1}{4(\delta_{(0,1,0,0,2)} - 1)^3}\{(\delta_{(0,1,0,0,2)} - 1)(\delta_{(0,0,0,1,4)} + 16\ \delta_{(0,0,1,0,3)} \\ & +26(2\delta_{(0,1,0,0,2)} - 1)) - \frac{5}{3}(6\delta_{(0,1,0,0,2)} + \delta_{(0,0,1,0,3)} - 4)^2\}. \end{aligned} \tag{3.6}$$

We note that the Bartlett correction does not depend on the model matrix X and is a second degree polynomial in the number of regression parameters (p).

We now apply formulae (3.3) and (3.6) to some symmetric distributions.

(i) Normal: $d_0 = 0, \quad d_1 = 1, \quad d_2 = 1, \quad d_3 = 1/3.$

(ii) Cauchy: $d_0 = 3/4, \quad d_1 = 1/2, \quad d_2 = 1, \quad d_3 = 1/4.$

(iii) t ($\nu > 0$ known):

$$\begin{aligned} d_0 &= \frac{3(\nu+2)(\nu+3)^2}{2\nu(\nu+1)(\nu+5)(\nu+7)}, \; d_1 = \frac{(\nu+2)(\nu+3)(\nu^2+9\nu+2)}{\nu(\nu+5)^2(\nu+7)}, \\ d_2 &= \frac{(\nu+3)(\nu+2)^2}{\nu(\nu+5)^2}, \quad d_3 = \frac{(\nu+3)(\nu^3+14\nu^2+14\nu+25)}{3\nu(\nu+5)^2(\nu+7)}. \end{aligned}$$

(iv) Generalized t ($s, r > 0$ known):

$$\begin{aligned} d_0 &= \frac{3(r+2)(r+3)^2}{2r(r+1)(r+5)(r+7)}, \quad d_1 = \frac{(r+2)(r+3)(r^2+9r+2)}{r(r+5)^2(r+7)}, \\ d_2 &= \frac{(r+3)(r+2)^2}{r(r+5)^2}, \quad d_3 = \frac{(r+3)(r^3+14r^2+14r+25)}{3r(r+5)^2(r+7)}. \end{aligned}$$

(v) Type I logistic: $d_0 \approx 0.4879224$, $d_1 \approx 1.470555027$, $d_2 \approx 1.362652631$, $d_3 \approx 0.481968387$.

(vi) Type II logistic: $d_0 = 3/20, \quad d_1 \approx 0.985182, \quad d_2 \approx 0.786756, \quad d_3 \approx 0.22795.$

(vii) Generalized logistic ($m > 0$ known);

$$\begin{aligned} d_0 &= \frac{(2m+1)}{4m(2m+3)}, \\ d_1 &= \frac{(2m+1)(4m^2(m+1)\psi'(m) + 8m^2 - 2m - 1)}{2(m+1)^2(2m^2\psi'(m) + 2m - 1)^2} \\ &\quad + \frac{(2m+1)(m^2\psi'(m) - 6m^2 - 3m - 1)}{m(2m^2\psi'(m) + 2m - 1)}, \\ d_2 &= \frac{(2m+1)^3}{2(2m^2\psi'(m) + 2m - 1)(m+1)^2}, \\ d_3 &= (2m+1)\{6m^3(m+1)\psi'(m)^2 + m^3(m+1)\psi'''(m) \\ &\quad -4m(26m^3 + 41m^2 + 15m + 3)\psi'(m) - 52m^2(m+1) + 21m \\ &\quad +9\}/\{2(2m+3)(m+1)(2m^2\psi'(m) + 2m - 1)^2\} \\ &\quad +5(2m+1)^3(3m^2\psi'(m) + 2m - 1)^2/\{3(m+1)^2(2m^2\psi'(m) \\ &\quad +2m - 1)^3\}. \end{aligned}$$

(viii) Power exponential ($-1 < k < -1/2$, known);

$$d_0 = k(1-k)\frac{\Gamma\left(\frac{1-3k}{2}\right)\Gamma\left(\frac{1+k}{2}\right)}{8\Gamma\left(\frac{3-k}{2}\right)^2}, \quad d_1 = d_2 = \frac{1}{k+1}, \quad d_3 = \frac{1}{3(k+1)}.$$

Note that the results of the t distribution depend only on the (known) parameter ν and agree with the results obtained by Ferrari and Arellano–Valle (1996). When $\nu \to \infty$, d_1 and d_2 equal the corresponding d's of the normal distribution. The results for the generalized t distribution depend only on the known parameter (r) and are equal to the corresponding results obtained for the t distribution with ν replaced by r. The d's for the power exponential distribution depend on the kurtosis parameters k which is assumed to be known. For the normal, Cauchy, type I and type II logistic distributions the d's are fixed constants. The d's for the generalized logistic distribution depend only on the parameter m and converge to the corresponding d's for the normal distribution when $m \to \infty$.

4 Simulation results

In this section we use Monte Carlo experiments to compare the size performance of the likelihood ratio (LR) and Bartlett corrected likelihood ratio (LR^*) statistics for testing $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ in the linear regression model (2.1) and (2.2) with $p = 7$ (i.e. 7 components in the regression parameter). We considered the following distributions for the errors: Cauchy and t with $\nu = 2$ and 4 degrees of freedom. For the t distribution ν is assumed to be known. We also consider the case where ν is (wrongly) assumed to be equal to 4 but the data were generated from a t distribution with 6 degrees of freedom.

The values for the covariates were selected as follows: x_1 is a vector of ones; x_2, x_3, x_4, x_5 and x_6 are random samples of the standard normal, uniform in the (0,1) interval, Cauchy, $F(8,6)$ and exponential (with mean equal to 8) distributions, respectively, and x_7 has its first $n/2$ components equal to zero and the remaining components equal to 1. All results are based on 10,000 Monte Carlo replications and are displayed in Table 3, where all entries are percentages.

The figures in Table 3 reveal that the likelihood ratio test tends to be quite oversized at least for the cases considered here. The size performance of the corrected test is much better than that of its uncorrected counterpart even for the Cauchy distribution and the t distribution with the number of degrees of freedom wrongly assumed to be equal to 4 while the correct value was 6 (a discussion of the use $\nu = 4$ when the true value of ν is unknown may be found in Lange, Little and Taylor, 1989, p.892). The effect of the Bartlett correction to the likelihood ratio statistic is remarkable. For example, for the t distribution with 2 degrees of freedom, $n = 30$ and a

10% nominal size, the simulated size of the likelihood ratio test drops from 19.7% (almost two times the nominal size) to 10.4% when the correction is used.

Table 3

Simulated sizes of the LR and corrected LR tests

Sample	Nominal	Cauchy		t_2		t_4		t_6 ($\nu = 4$)(*)	
size	size	*LR*	*LR**	*LR*	*LR**	*LR*	*LR**	*LR*	*LR**
20	5.0	25.0	5.7	14.9	3.8	11.7	4.3	12.5	4.7
	10.0	33.8	21.2	22.3	8.2	18.6	8.5	19.8	9.4
30	5.0	17.5	3.9	11.9	5.3	10.0	4.7	10.8	5.8
	10.0	26.2	8.7	19.7	10.4	16.0	10.0	17.8	11.3
40	5.0	12.8	4.8	8.9	4.8	8.4	5.3	8.8	5.4
	10.0	20.6	9.9	15.7	9.6	15.0	10.3	15.3	10.9
50	5.0	10.3	4.9	8.0	4.8	7.4	5.0	8.0	5.5
	10.0	17.5	9.7	14.6	10.0	13.4	10.2	13.9	10.7
60	5.0	9.1	4.7	7.2	4.8	6.6	4.9	7.5	5.5
	10.0	16.0	9.7	13.2	10.0	12.4	9.8	13.6	10.7

(*) ν is assumed to be equal to 4 while its correct value is 6.

5 A numerical example

As a numerical application of our results we consider a data set presented by Draper and Stoneman (1966) (see Table 4). For these data we assume that

$$y_l = \beta_0 + \beta_1 x_{1l} + \beta_2 x_{2l} + \phi z_l, \;\; l = 1, 2, \ldots, 10,$$

where the errors z_l are i.i.d. random variables. We first fitted the model assuming a normal distribution for the errors. In Figure 1 we present a plot of the usual Studentized deleted residuals against the fitted values. These residuals have a t distribution with $n - p - 2$ degrees of freedom if the errors follow a standard normal distribution. The plot reveals that the first observation is an outlying case (its Studentized deleted residual equals -3.25). This suggests that a distribution with longer-than-normal tails may be more appropriate for modelling these data. In Table 5 we present the maximum likelihood estimates of the parameters assuming the following distributions for the errors: standard normal, t with $\nu = 4$, power exponential with $k = 0.5$ and generalized logistic with $m = 0.4$ and $\alpha = \sqrt{2\psi'(m)}$.

Table 4
Wood beam data

strenght (y)	specific gravity (x_1)	moisture content (x_2)
11.14	0.499	11.1
12.74	0.558	8.9
13.13	0.604	8.8
11.51	0.441	8.9
12.38	0.550	8.8
12.60	0.528	9.9
11.13	0.418	10.7
11.70	0.480	10.5
11.02	0.406	10.5
11.41	0.467	10.7

Table 5
Parameter estimates

Distribution	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\phi}$
$N(0,1)$	10.302	8.495	-0.266	0.230
	(1.583)	(1.491)	(0.103)	(0.051)
t_4	9.068	9.231	-0.175	0.159
	(1.295)	(1.219)	(0.084)	(0.047)
PE($k = 0.5$)	9.368	9.095	-0.198	0.138
	(1.682)	(1.449)	(0.113)	(0.037)
GL($m = 0.4$)	9.104	9.226	-0.178	0.217
	(1.345)	(1.266)	(0.087)	(0.060)

Note: Standard errors are given in parentheses.

We now delete the first observation (outlying case) and re-fit the same models. Table 6 show the new parameter estimates. We note that under the normal assumption, the deletion of a single observation has a remarkable impact on the estimates while, under the assumption of longer-than-normal-tailed distributions, the impact is much less pronounced. For example, when the first observation is excluded, $\widehat{\beta}_1$ becomes 21%, 7%, 8% and 7% bigger under the normal, t_4, PE($k = 0.5$) and GL($m = 0.4$) distributions respectively. This suggests that the heavy-tailed distributions are, in fact, more appropriate for this data set than the normal distribution.

Table 6
Parameter estimates excluding the first observation

Distribution	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\phi}$
$N(0,1)$	7.592	10.267	-0.073	0.140
	(1.163)	(1.002)	(0.078)	(0.032)
t_4	8.148	9.833	-0.109	0.106
	(1.042)	(0.898)	(0.070)	(0.037)
PE$(k = 0.5)$	8.028	9.862	-0.098	0.087
	(1.275)	(1.098)	(0.085)	(0.025)
GL$(m = 0.4)$	8.139	9.844	-0.108	0.135
	(0.986)	(0.849)	(0.066)	(0.041)

Note: Standard errors are given in parentheses.

Table 7
LR and corrected LR statistics

Distribution	LR	LR^*
N(0,1)	5.068	3.295
	(0.024)	(0.047)
t_4	3.205	1.955
	(0.073)	(0.162)
GL$(m = 0.4)$	3.666	2.139
	(0.056)	(0.144)

Note: p-values are given in parentheses.

Table 8
LR and corrected LR statistics excluding the first observation

Distribution	LR	LR^*
N(0,1)	0.769	0.470
	(0.381)	(0.493)
t_4	2.074	1.162
	(0.149)	(0.281)
GL$(\nu = 0.4)$	2.526	1.336
	(0.112)	(0.247)

Note: p-values are given in parentheses.

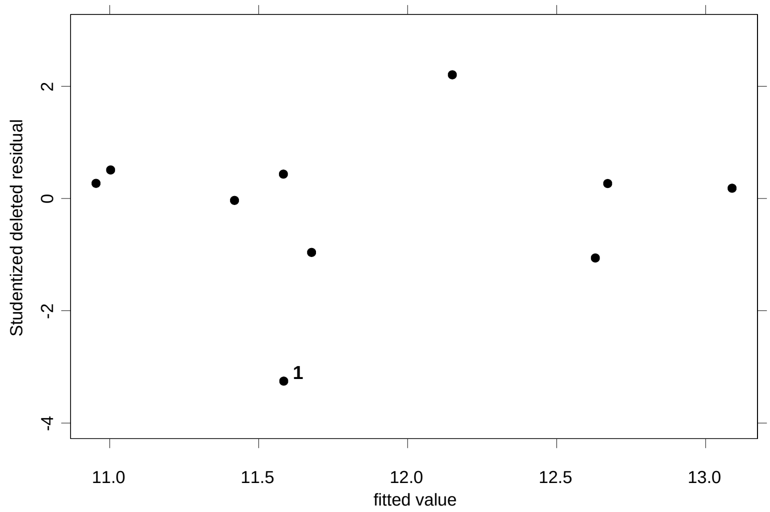

Figure 1
Plot of Studentized deleted residuals against fitted values

Tables 7 and 8 show the values of the likelihood ratio statistics and those of their corrected versions for the test of $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ calculated with all the observations and without the first one, respectively[2]. The figures in these tables show that deletion of the outlying observation has an enormous impact on the p-values of all the tests under the normality assumption. On the other hand, the changes in the p-values are much smaller under the alternative models. In particular, for a 10% nominal level, the corrected likelihood ratio test under longer-than-normal-tailed distributions do not lead to the rejection of the null hypothesis whether or not the first observation is eliminated. If the normal assumption is considered, H_0 is not rejected if the first observation is deleted and rejected otherwise. This example suggests that the class of symmetric models studied in this paper is robust in the presence of outlying observations.

6 Concluding remarks

In this paper we derived Bartlett corrections to likelihood ratio statistics in a class of symmetric linear regression models that has the normal and

[2]The power exponential ($k = 0.5$) is not considered here because the Bartlett correction is valid only for $-1 < k < -1/2$.

the t linear regression models as special cases and covers other models with heavy tails. The Bartlett factor relative to tests concerning the regression parameters depends on the distribution assumed for the data, in particular on the coefficient of kurtosis of such distribution, and also on the model matrix, the number of regression parameters, the number of nuisance parameters and the sample size. Besides, the Bartlett factor relative to tests on the dispersion parameter depends on the same quantities except for the model matrix which is irrelevant for determining the correction. Our simulation results reveal that the likelihood ratio test tends to be quite oversized and that the size performance of the corrected test is much better than that of its uncorrected counterpart.

Appendix

Let $L = L(\theta)$ be the total log-likelihood function given in (2.3). We shall use the following notation for cumulants of log-likelihood derivatives, where all suffices, except ϕ, range from 1 to p: $\kappa_{rs} = E(\partial^2 L/\partial\beta_r\partial\beta_s)$, $\kappa_{rst} = E(\partial^3 L/\partial\beta_r\partial\beta_s\partial\beta_t)$, $\kappa_{\phi\phi} = E(\partial^2 L/\partial\phi^2)$, $\kappa_{r\phi} = E(\partial^2 L/\partial\beta_r\partial\phi)$, $\kappa_{r,s} = E(\partial L/\partial\beta_r\, \partial L/\partial\beta_s)$ and so on. Derivatives of such cumulants are denoted by $\kappa_{rs}^{(t)} = \partial E(\partial^2 L/\partial\beta_r\beta_s)/\partial\beta_t$, $\kappa_{rs}^{(\phi)} = \partial E(\partial L/\partial\beta_r\beta_s)/\partial\phi$, $\kappa_{rs}^{(tu)} = \partial^2 E(\partial L/\partial\beta_r\beta_s)/\partial\beta_t\partial\beta_u$, etc. Note that $\kappa_{r,s}$ and $\kappa_{\phi,\phi}$ are elements of the information matrix for $(\beta^\top,\phi)^\top$. The corresponding elements of its inverse are denoted by $\kappa^{r,s} = -\kappa^{rs}$ and $\kappa^{\phi,\phi} = -\kappa^{\phi\phi}$. From the first derivative of the log–likelihood function given in Section 2, we have

$$\begin{aligned}\frac{\partial^2 L}{\partial\beta_r\partial\beta_s} &= -\frac{1}{\phi}\sum\frac{\partial t_l^{(1)}}{\partial\beta_s}x_{lr} = -\frac{1}{\phi}\sum\frac{\mathrm{d}t_l^{(1)}}{\mathrm{d}z_l}\frac{\partial z_l}{\partial\mu_l}\frac{\partial\mu_l}{\partial\beta_s}x_{lr} = \frac{1}{\phi^2}\sum t_l^{(2)}x_{lr}x_{ls},\\ \frac{\partial^2 L}{\partial\phi^2} &= \frac{1}{\phi^2}\sum_{l=1}^{n}(1+2t_l^{(1)}z_l+t_l^{(2)}z_l^2),\ \frac{\partial^2 L(\theta)}{\partial\beta_r\partial\phi} = \frac{1}{\phi^2}\sum_{l=1}^{n}(t_l^{(2)}z_l+t_l^{(1)})x_{lr}.\end{aligned}$$

Assuming standard regularity conditions and taking expectations we get

$$\kappa_{rs} = -\frac{\delta_{(2,0,0,0,0)}}{\phi^2}\sum_{l=1}^{n}x_{lr}x_{ls},\quad \kappa_{\phi\phi} = -\frac{n}{\phi^2}(\delta_{(2,0,0,0,2)}-1),\quad \kappa_{r\phi} = 0,$$

where the δ's come from (2.6). From the above equations, it follows the information matrix for $(\beta^\top,\phi)^\top$ given in Section 3. Note that, from Bartlett identities (Lawley, 1956), the δ's sastify some relations such as $\delta_{(0,1,0,0,1)} = \delta_{(1,0,0,0,0)} = 0$, $\delta_{(0,0,0,1,0)} = -\delta_{(1,0,1,0,0)}$, $\delta_{(1,1,0,0,1)} + \delta_{(0,0,1,0,1)} = -\delta_{(0,1,0,0,0)}$, $\delta_{(0,1,0,0,2)} = 2-\delta_{(2,0,0,0,2)}$, $\delta_{(4,0,0,0,0)} = -3\delta_{(2,1,0,0,0)}$, $\delta_{(1,1,0,0,3)} = -3\delta_{(0,1,0,0,2)} -\delta_{(0,0,1,0,3)}$, $\delta_{(0,1,0,0,0)} = -\delta_{(2,0,0,0,0)}$ and $2\delta_{(0,0,1,0,1)} = -\delta_{(0,0,0,1,2)} - \delta_{(1,0,1,0,2)}$.

Other cumulants of log-likelihood derivatives that will be needed for obtaining the Bartlett correction to the likelihood ratio statistic are given by

$$\begin{aligned}
\kappa_{rst} &= -\frac{1}{\phi^3}\sum E(t_l^{(3)})x_{lr}x_{ls}x_{lt} = 0,\\
\kappa_{rstu} &= \frac{1}{\phi^4}\sum E(t_l^{(4)}))x_{lr}x_{ls}x_{lt}x_{lu} = \frac{1}{\phi^4}\delta_{(0,0,0,1,0)}\sum x_{lr}x_{ls}x_{lt}x_{lu},\\
\kappa_{\phi rs} &= -\frac{1}{\phi^3}(\delta_{(0,0,1,0,1)} + 2\delta_{(0,1,0,0,0)})\Sigma x_{lr}x_{ls}, \quad \kappa_{\phi\phi r} = 0,\\
\kappa_{\phi rst} &= 0, \quad \kappa_{\phi\phi rs} = \frac{1}{\phi^4}(6\delta_{(0,1,0,0,0)} + 6\delta_{(0,0,1,0,1)} + \delta_{(0,0,0,1,2)})\Sigma x_{lr}x_{ls},\\
\kappa_{\phi\phi\phi r} &= 0, \quad \kappa_{\phi\phi\phi} = -\frac{n}{\phi^3}(6\delta_{(0,1,0,0,2)} + \delta_{(0,0,1,0,3)} - 4),\\
\kappa_{\phi\phi\phi\phi} &= \frac{n}{\phi^4}(\delta_{(0,0,0,1,4)} + 12\delta_{(0,0,1,0,3)} + 36\delta_{(0,1,0,0,2)} - 18).
\end{aligned}$$

Suppose now that ϕ is known. The general expression for the coefficient d (see eq. (3.1)) of the Bartlett correction to the likelihood ratio statistic for testing $H_0 : \beta_1 = \beta_1^{(0)}$ against $H : \beta_1 \neq \beta_1^{(0)}$ is

$$d = \frac{n}{q}\left(\sum{}'(\ell_{rstu} - \ell_{rstuvw}) - \sum{}''(\ell_{rstu} - \ell_{rstuvw})\right) \tag{A.1}$$

where

$$\ell_{rstu} = \kappa^{rs}\kappa^{tu}\left(\frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)}\right), \tag{A.2}$$

$$\begin{aligned}
\ell_{rstuvw} &= \kappa^{rs}\kappa^{tu}\kappa^{vw}\left\{\kappa_{rtv}\left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)}\right) + \kappa_{rtu}\left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)}\right)\right.\\
&\left.+\kappa_{rt}^{(v)}\kappa_{sw}^{(u)} + \kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right\}
\end{aligned} \tag{A.3}$$

with $\sum'$ and $\sum''$ representing the summation over all the subscripts ranging in $\{1,\ldots,p\}$ and $\{q+1,\ldots,p\}$ respectively (see Lawley, 1956). Plugging the cumulants given above in (A.2) we get

$$\begin{aligned}
\sum{}'\ell_{rstu} &= \frac{\delta_{(0,0,0,1,0)}}{4\phi^4}\sum{}'\kappa^{rs}\kappa^{tu}\sum x_{lr}x_{ls}x_{lt}x_{lu}\\
&= \frac{\delta_{(0,0,0,1,0)}}{4\phi^4}\sum(-\sum x_{lr}x_{ls}\kappa^{rs})(-\sum x_{lt}x_{lu}\kappa^{tu})\\
&= \frac{\delta_{(0,0,0,1,0)}}{4\delta_{(2,0,0,0,0)}^2}\sum z_{ll}^2 = \frac{\delta_{(0,0,0,1,0)}}{4\delta_{(2,0,0,0,0)}^2}\mathrm{tr}(Z_dZ_d).
\end{aligned}$$

Note that $\sum' \ell_{rstuvw}$ vanishes because all the cumulants that appear in ℓ_{rstuvw} are equal to zero. Now, if ϕ is known the coefficient d comes from (A.1) as

$$d = \frac{d_0}{q} h_0(X_1, X_2) \qquad \text{(A.4)},$$

where d_0 and $h_0(X_1, X_2)$ are given in (3.3) and (3.4) respectively.

Now, suppose that ϕ is unknown. It is convenient to write d as

$$d = d_\beta + d_{\beta\phi},$$

where d_β is given by (A.4) and $d_\beta\phi$ is given by (A.1) with the summations replaced by $\sum'_{\beta\phi}$ and $\sum''_{\beta\phi}$. These summations indicate that the subscripts range in $\{1, \ldots, p, \phi\}$ and $\{q+1, \ldots, p, \phi\}$ respectively with at least one subscript equal to ϕ. From (A.2) and (A.3) and the cumulants given above, we have

$$\sum\nolimits'_{\beta\phi} \ell_{rstu} = \ell_{\phi\phi\phi\phi} + \left(\frac{6m_3 - m_4}{2nm_1}\right) p \qquad \text{(A.5)}$$

and

$$\sum\nolimits'_{\beta\phi} \ell_{rstuvw} = \ell_{\phi\phi\phi\phi\phi\phi} + \frac{1}{2n}\frac{m_2 m_3}{m_1^2} p + \frac{4}{n}\frac{m_3}{m_1} p + \frac{1}{2n}\frac{m_3^2}{m_1} p + \frac{1}{2}\frac{m_3^2}{m_1}\frac{p^2}{2}, \quad \text{(A.6)}$$

where m_1, m_2 and m_3 are given in Section 3,

$$\ell_{\phi\phi\phi\phi} = \frac{1}{4n(\delta_{(2,0,0,0,2)} - 1)^2}\{\delta_{(0,0,0,1,4)} - 12\delta_{(0,1,0,0,2)} + 6\}$$

and

$$\begin{aligned} \ell_{\phi\phi\phi\phi\phi\phi} \;=\; & \frac{1}{4n(\delta_{(0,1,0,0,2)} - 1)^3}\Big\{\frac{5}{3}(6\delta_{(0,1,0,0,2)} + \delta_{(0,0,1,0,3)} - 4)^2 \\ & -16(\delta_{(0,1,0,0,2)} - 1)(4\delta_{(0,1,0,0,2)} + \delta_{(0,0,1,0,3)} - 2)\Big\}. \end{aligned}$$

The summations $\sum''_{\beta\phi}$ and $\sum''_{\beta\phi}$ are given respectively by (A.5) and (A.6) with p replaced by q. Now, it is easy to get equation (3.2).

If the hypothesis $H_0 : \phi = \phi^{(0)}$ is to be tested against the alternative hypothesis $H_0 : \phi = \phi^{(0)}$, the coefficient d of the Bartlett correction is given by (A.1) with q replaced by 1 and with $\sum'$ and $\sum''$ representing the summation over all the subscripts ranging in $\{1, \ldots, p, \phi\}$ and $\{1, \ldots, p\}$ respectively. It is now clear that, in this case,

$$d = n \sum\nolimits'_{\beta\phi} (\ell_{rstu} - \ell_{rstuvw}).$$

From (A.5) and (A.6) we get (3.5) after some algebra.

Acknowledgements

We gratefully acknowledge partial financial support from CNPq, FAPESP and FINEP.

(Received June, 2000. Revised August, 2001.)

References

Berkane, M., Bentler, P.M. (1986). Moments of elliptically distributed random variates. *Statistics and Probability Letters*, **4**, 333–335.

Cambanis, S., Huang, S., Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**, 368–385.

Chmielewski, M.A. (1981). Elliptically simmetric distributions: a review and bibliography. *International Statistical Review*, **49**, 67–74.

Chu, K. (1973). Estimation and decision for linear systems with elliptical random processes. *I.E.E.E. Trans. Auto. Control*, **18**, 499–505.

Cordeiro, G.M., (1983). Impoved likelihood ratio tests generalized linear models. *Journal of the Royal Statistical Society* B, **45**, 404–413.

Cordeiro, G.M., (1987). Impoved likelihood ratio tests generalized linear models. *Biometrika,* **74**, 265–274.

Cordeiro, G.M., Paula, G.A., Botter, D.A. (1994). Impoved likelihood ratio tests for dispersion models. *International Statistical Review*, **62**, 257–276.

Cordeiro, G.M., Ferrari, S.L.P., Uribe–Opazo, M.A., Vasconcellos, K.L.P. (2000). Corrected maximum likelhood estimation in a class of symmetric nonlinear regression models. *Statistical and Probability Letters*, **46**, 317–328.

Draper, N.R., Stoneman, D.M. (1966). Testing for the inclusion of variables in linear regression by a randomization technique. *Technometrics*, **8**, 695–699.

Fang, K.T., Kotz, S., Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions.* London: Chapman and Hall.

Ferrari, S.L.P., Arellano–Vale, R.B. (1996). Bartlett corrected tests for regression models with Student–t independent errors. *Brazilian Journal of Probability and Statistics*, **10**, 15–33.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā* A, **32**, 419–430.

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

Lawley, D.N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, **71**, 233–244.

Muirhead, R. (1980). The effects of symmetric distributions on some standard procedures involving correlation coefficients. In *Multivariate Statistical Analysis* (ed. R.P. Gupta), North-Holland, 143–159.

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

Rao, B.L.S.P. (1990). Remarks on univariate symmetric distributions. *Statistics and Probability Letters*, **10**, 307–315.