Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computational Statistics & Data Analysis 49 (2005) 689–708

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Restricted methods in symmetrical linear regression models

Francisco José A. Cysneiros[a], Gilberto A. Paula[b,1,*]

[a]*Universidade Federal de Pernambuco, Brazil*
[b]*Department of Statistics, Universidade de São Paulo, Rua do Matao 1010, 05508090 Sao Paulo SP, Brazil*

Received 28 May 2004; accepted 2 June 2004
Available online 25 June 2004

Abstract

In this paper we discuss the problem of testing equality and inequality constraints in symmetrical linear regression models. This class of models includes all symmetric continuous distributions, such as normal, Student-*t*, Pearson VII, power exponential and logistic, among others. It is commonly used for the analysis of data containing influential or outlying observations with responses supposedly normal. Iterative processes for evaluating the parameters under equality and inequality constraints are presented. The asymptotic null distribution of three asymptotically equivalent one-sided tests is showed to be invariant with the symmetrical error. A sensitivity study to investigate the robustness of the maximum likelihood estimates from some symmetrical models against high leverage and influential observations is presented. An illustrative example with presence of influential observations on the decisions from the statistical tests of different symmetrical models is given. The robustness aspects of such models are also discussed.

Keywords: Hypothesis testing; Symmetric distributions; One-sided tests; Order restrictions; Restricted estimation; Robustness

* Corresponding author. Tel.: +55-11-3091-6239; fax: +55-11-3091-6130.
E-mail address: giapaula@ime.usp.br (G.A. Paula).giapaula@ime.usp.br (G.A. Paula).

[1] Address for correspondence: Instituto de Matemática e Estatística, USP - Caixa Postal 66281 (Ag.Cidade de São Paulo), 05311-970 São Paulo - SP - Brazil.

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.06.001

1. Introduction

In this paper we discuss two situations of testing restricted hypotheses in symmetrical linear regression models. First, the problem of testing linear equality hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta}=\mathbf{d}$ against the linear inequality hypothesis $H_2 : \mathbf{C}\boldsymbol{\beta}\geqslant\mathbf{d}$, with at least one strict inequality in H_2 (case 1) is treated, and then, $H_2 : \mathbf{C}\boldsymbol{\beta}\geqslant\mathbf{d}$ against $\mathbb{R}^p$ - H_2 (case 2) is discussed. The problem of testing one-sided alternatives was originally treated by Bartholomew (1959a, b) for independent normal models and extended by Kudo (1963) for multivariate normal models. Nuesch (1966) also investigates this problem in normal models while Perlman (1969) extends the results for a more general class of multivariate normal models. Gourieroux et al. (1982) discuss the asymptotic null distribution of three asymptotically equivalent one-sided tests in multivariate normal models when the variance-covariance matrix may depend on a finite number of unknown parameters. Wolak (1987) proposes exact one-sided tests for multivariate normal models and Wolak (1989) extends the results from Gourieroux et al. (1982) for more general restricted hypotheses. Moving away from the normal case, Kodde and Palm (1986) and Silvapulle and Silvapulle (1995), for instance, present Wald and score type tests that may be applied for testing equality and inequality restrictions in general multivariate regression models. In case 2, the main difficulty is when the information matrix depends on the parameter $\boldsymbol{\beta}$. A consequence of this fact is that we should search through the set of null parameters for least favorable points. Wolak (1991) proposes a lemma in which a methodology to find a least favorable region is presented. An excellent review on this subject may be found in the book by Robertson et al. (1988) (see also, Sen and Silvapulle, 2002).

The paper is organized as follows. In Section 2 we discuss the unrestricted parameter estimation in symmetrical linear models. Iterative processes for evaluating the maximum likelihood restricted estimates under equality and inequality constraints are given in Section 3. Section 4 contains the expressions as well as the asymptotic null distribution of three asymptotically equivalent one-sided tests. In Section 5 a sensitivity study to investigate the robustness of the maximum likelihood estimates from some symmetrical models against high leverage and influential observations is given. An illustrative example in which influential observations change the decisions from the statistical tests of different symmetrical models is presented in Section 6. The robustness aspects of such models are discussed. The last section deals with some concluding remarks.

2. Symmetrical linear models

Suppose $Y_1, \ldots, Y_n$ independent random variables with density function of Y_i given by

$$f_{y_i}(y) = \frac{1}{\sqrt{\phi}} g\{(y-\mu_i)^2/\phi\}, \tag{1}$$

$y \in \mathbb{R}$, where the function $g : \mathbb{R} \to [0, \infty)$ is such that $\int_0^\infty g(u)\,\mathrm{d}u < \infty$. The function $g(\cdot)$ is typically known as the density generator. We will denote $Y_i \sim \mathrm{S}(\mu_i, \phi)$. The symmetrical linear regression models are defined as

$$Y_i = \mu_i(\boldsymbol{\beta}) + \varepsilon_i, \tag{2}$$

where $\mu_i(\boldsymbol{\beta}) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$, $\mathbf{x}_i$ is a $p \times 1$ vector of explanatory variable values and $\varepsilon_i \sim S(0, \phi)$. We have, when they exist, that $E(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \xi\phi$, where $\xi > 0$ is a constant that may be obtained from the expected value of the radial variable or from the derivative of the characteristic function (see, for instance, Fang et al., 1990). For example, for the Student-t distribution with v degrees of freedom one has $\xi = v/(v-2)$ $(v > 2)$.

The symmetrical family of distributions allows an extension of the normal distribution for statistical modeling of real data involving distributions with heavier and lighter tails than the ones of the normal distribution. Many authors such as Muirhead (1980, 1982), Berkane and Bentler (1986), Rao (1990) and Fang and Anderson (1990) have investigated these distributions. A review of different areas in which symmetrical distributions are applied is given by Chmielewski (1981). Influence diagnostic methods can be found, for instance, in Galea et al. (1997) and Galea et al. (2003).

The log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \phi)^{\mathrm{T}}$ takes the form

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\log\phi + \sum_{i=1}^{n}\log\{g(u_i)\}.$$

The function $L(\boldsymbol{\theta})$ is assumed to be regular (Cox and Hinkley, 1974, Chapter 9) with respect to $\boldsymbol{\beta}$ and ϕ. Regular conditions are also stated in Serfling (1980, p. 144). To obtain the score function and the Fisher information matrix, we need to derive $L(\boldsymbol{\theta})$ with respect to unknown parameters and then computing some moments of such derivatives. We suppose that such derivatives exist. However, some symmetric distributions do not satisfy the regularity conditions, for example, double exponential, Kotz and generalized Kotz. These cases will not be considered here.

We find the score functions

$$\mathbf{U}_{\beta}(\boldsymbol{\theta}) = \phi^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{3}$$

and

$$U_{\phi}(\boldsymbol{\theta}) = (2\phi)^{-1}\{\phi^{-1}Q_V(\boldsymbol{\beta}) - n\}, \tag{4}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, $\mathbf{X}$ is a $n \times p$ matrix with rows $\mathbf{x}_i^{\mathrm{T}}$, $\mathbf{D}(\mathbf{v}) = \mathrm{diag}\{v_1, \ldots, v_n\}$ with $v_i = -2\mathrm{W}_g(u_i)$, $\mathrm{W}_g(u_i) = g'(u_i)/g(u_i)$ with $g'(u_i) = \mathrm{d}g(u_i)/\mathrm{d}u_i$ and $Q_V(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{D}(\mathbf{v})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Let $\ddot{\mathbf{L}}_{\theta\theta}$ be the $(p+1) \times (p+1)$ matrix with ij-element $\partial^2 L(\boldsymbol{\theta})/\partial\theta_i\theta_j$. After some algebraic manipulations we find

$$\ddot{\mathbf{L}}_{\theta\theta} = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\phi} \\ \ddot{\mathbf{L}}_{\phi\beta} & \ddot{\mathrm{L}}_{\phi\phi} \end{bmatrix},$$

Table 1
Values of $g(u)$, $\mathrm{W}_g(u)$ and $\mathrm{W}'_g(u)$ for some symmetrical distributions

Distribution	g(u)	$\mathrm{W}_g(u)$	$\mathrm{W}'_g(u)$
Normal	$\frac{1}{\sqrt{2\pi}}\exp(-u/2)$	$-\frac{1}{2}$	0
Student-t	$\frac{v^{v/2}}{\mathrm{Beta}(1/2,v/2)}(v+u)^{-\frac{v+1}{2}}$	$-\frac{v+1}{2(v+u)}$	$\frac{(v+1)}{2(v+u)^2}$
Generalized Student-t	$\frac{s^{r/2}}{\mathrm{Beta}(1/2,r/2)}(s+u)^{-\frac{r+1}{2}}$	$-\frac{(r+1)}{2(s+u)}$	$\frac{(r+1)}{2(s+u)^2}$
Logistic-I	$c\frac{\exp(-u)}{[1+\exp(u)]^2}$	$-\tanh(\frac{u}{2})$	$-\mathrm{sech}(\frac{u}{2})/2$
Logistic-II	$\frac{\exp(-\sqrt{u})}{[1+\exp(-\sqrt{u})]^2}$	$-\frac{\exp(-\sqrt{u})-1}{(-2\sqrt{u})[1+\exp(-\sqrt{u})]}$	$\frac{2\exp(-\sqrt{u})\sqrt{u}+\exp(-2\sqrt{u})-1}{-4u^{3/2}[1+\exp(-\sqrt{u})]^2}$
Generalized logistic	$\frac{\alpha}{\mathrm{Beta}(m,m)}\left[\frac{\exp\{-\alpha\sqrt{u}\}}{(1+\exp\{-\alpha\sqrt{u}\})^2}\right]^m$	$\frac{-\alpha m[\exp(-\alpha\sqrt{u})-1]}{(-2\sqrt{u})[1+\exp(-\alpha\sqrt{u})]}$	$-\frac{\alpha m}{4}\frac{2\alpha\exp(-\alpha\sqrt{u})\sqrt{u}+\exp(-2\alpha\sqrt{u})-1}{u^{3/2}[1+\exp(-\alpha\sqrt{u})]^2}$
Power exponential	$\frac{\exp\{-\frac{1}{2}u^{1/(1+k)}\}}{\Gamma(1+\frac{1+k}{2})2^{1+(1+k)/2}}$	$-\frac{1}{2(1+k)(u)^{k/(k+1)}}$	$\frac{k}{(1+k)^2 2u^{(2k+1)/(1+k)}}$

Note: $\Gamma(\cdot)$ and $\mathrm{Beta}(\cdot)$ are gamma and beta functions respectively and c is a normalized constant, $c \approx 1.48430029$.

where

$$\ddot{\mathbf{L}}_{\beta\beta} = -\phi^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{a})\mathbf{X},$$
$$\ddot{\mathbf{L}}_{\beta\phi} = 2\phi^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{b} \quad \text{and}$$
$$\ddot{\mathrm{L}}_{\phi\phi} = \phi^{-2}\left\{\frac{n}{2} + \mathbf{u}^{\mathrm{T}}\mathbf{D}(\mathbf{c})\mathbf{u} - \phi^{-1}Q_V(\boldsymbol{\beta})\right\}$$

with $\mathbf{D}(\mathbf{a}) = \mathrm{diag}\{a_1, \ldots, a_n\}$, $\mathbf{D}(\mathbf{c}) = \mathrm{diag}\{c_1, \ldots, c_n\}$, $\mathbf{b} = (b_1, \ldots, b_n)^{\mathrm{T}}$, $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$, $a_i = -2\{\mathrm{W}_g(u_i) + 2u_i\mathrm{W}'_g(u_i)\}$, $c_i = \mathrm{W}'_g(u_i)$ and $b_i = \{\mathrm{W}_g(u_i) + u_i\mathrm{W}'_g(u_i)\}\varepsilon_i$.

The Fisher information matrix for $\boldsymbol{\beta}$ and ϕ can be expressed, respectively, as

$$\mathbf{K}_{\beta\beta} = \frac{4d_g}{\phi}\mathbf{X}^{\mathrm{T}}\mathbf{X} \quad \text{and} \quad K_{\phi\phi} = \frac{n}{4\phi^2}(4f_g - 1),$$

where $d_g = E\{\mathrm{W}_g^2(Z^2)Z^2\}$ and $f_g = E\{\mathrm{W}_g^2(Z^2)Z^4\}$ with $Z \sim S(0, 1)$. For some values of $g(u)$, $\mathrm{W}_g(u)$, $\mathrm{W}'_g(u)$, d_g, f_g and ξ for several symmetrical distributions, see Tables 1 and 2 It may be showed that $\mathbf{K}_{\beta\phi} = \mathbf{0}$, that is, $\boldsymbol{\beta}$ and ϕ are orthogonal parameters. A joint iterative process to solve $\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{0}$ and $U_\phi(\boldsymbol{\theta}) = 0$ is given by

$$\boldsymbol{\beta}^{(r+1)} = \{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{y} \tag{5}$$

Table 2
Values of d_g, f_g and ξ for some elliptical distributions

Distribution	d_g	f_g	ξ
Normal	$\frac{1}{4}$	$\frac{3}{4}$	1
Student-*t*	$\frac{(v+1)}{4(v+3)}$	$\frac{3(v+1)}{4(v+3)}$	$\frac{v}{v-2}$, $v > 2$
Generalized Student-*t*	$\frac{r(r+1)}{4s(r+3)}$	$\frac{3(r+1)}{4(r+3)}$	$\frac{s}{r-2}$, $s > 0, r > 2$
Logistic-I	0.369310044	1.003445984	0.79569
Logistic-II	$\frac{1}{12}$	0.60749	$\pi^2/3$
Generalized logistic	$\frac{\alpha^2 m^2}{4(2m+1)}$	$\frac{2m(2+m^2\psi'(m))}{4(2m+1)}$	$2\psi'(m)$
Power exponential	$\frac{\Gamma\{(3-k)/2\}}{4(2^{k-1})(1+k)^2\Gamma\{(k+1)/2\}}$	$\frac{(k+3)}{4(k+1)}$	$2^{(1+k)}\frac{\Gamma\{3(k+1)/2\}}{\Gamma\{(k+1)/2\}}$

Note: $\psi'(\cdot)$ is the trigamma function.

and

$$\phi^{(r+1)} = \frac{1}{n} Q_V(\boldsymbol{\beta}^{(r)}), \tag{6}$$

$r = 0, 1, \ldots$. The iterative process (6) guarantees positive solution for the maximum likelihood estimate of ϕ. We should start the iterative process (5)–(6) with initial values $\boldsymbol{\beta}^{(0)}$ and $\phi^{(0)}$. The procedure described above is the same described by Gourieroux and Monfort (Vol. I, p. 170, 1995). In the first step (5), we maximize the log-likelihood function with regard to $\boldsymbol{\beta}$ given ϕ, known as the concentrated log-likelihood, $\mathbf{L}_c(\boldsymbol{\beta} \mid \phi)$. In the second step (6), we maximize the log-likelihood function with regard to ϕ given $\boldsymbol{\beta}$, namely $\mathbf{L}_c(\phi \mid \boldsymbol{\beta})$. This procedure leads to the maximum likelihood estimate of $\boldsymbol{\theta}$.

Since $g(\cdot)$ is a nonincreasing function the concavity of $\mathbf{L}_c(\boldsymbol{\beta} \mid \phi)$ is guaranteed by Theorem 6.6 of Lehmann (1983, p. 53) if the matrix $\ddot{\mathbf{L}}_{\beta\beta} = -\frac{1}{\phi}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{a})\mathbf{X}$ is negative definite, that occurs when $a_i > 0, \ \forall i$. For example, $a_i = 1$ for the normal case and $a_i = (v+1)(v-u_i)/(v+u_i)^2$ for the Student-*t* distribution with v degrees of freedom. Therefore, we will have a_i positive whenever $\sqrt{v} > |y_i - \mu_i|/\sqrt{\phi}, \forall i$. Thus, it will be more difficult to find concavity for small than for large degrees of freedom. This is in agreement with the results of Pratt (1981). Nevertheless, we may attain concavity even for some a_i negative. The concavity of $\mathbf{L}_c(\phi \mid \boldsymbol{\beta})$ is guaranteed when $\ddot{\mathrm{L}}_{\phi\phi}$ is negative.

For the normal distribution the maximum likelihood estimates take closed-form expressions, because $v_i = 1, \forall i$. For the Student-*t* distribution with v degrees of freedom, we have $g(u) = c(1 + u/v)^{-(v+1)/2}$, $v > 0$ and $u > 0$ so that $\mathrm{W}_g(u_i) = -(v+1)/2(v+u_i)$

and $v_i = (v+1)/(v+u_i), \forall i$. In this case the current weight $v_i^{(r)}$ from (5) is inversely proportional to the distance between the observed value y_i and its current predicted value $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(r)}$, so that outlying observations tend to have small weights in the estimation process (see discussion, for instance, in Lange et al., 1989). For the power exponential distribution with shape parameter $\gamma = 1/(1+k)$ fixed, $g(u) = c\mathrm{e}^{-0.5u^{\gamma-1}}$, $u > 0$ and $\gamma \geqslant \frac{1}{2}$, then $\mathrm{W}_g(u_i) = -\frac{1}{2}\gamma u_i^{\gamma-1}$ and $v_i = \gamma u_i^{\gamma-1}$.

We further assume that $\boldsymbol{\beta} \in \Omega_\beta \subset \mathbb{R}^p$, where Ω_β is open with interior points. We have that $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, and

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}, \mathbf{J}_{\beta\beta}^{-1}), \text{ where } \mathbf{J}_{\beta\beta} = \lim_{n\to\infty} \frac{1}{n}\mathbf{K}_{\beta\beta}.$$

Then, $\hat{\mathbf{K}}_{\beta\beta}^{-1} = \frac{\hat{\phi}}{4d_g}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ is a consistent estimator of the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}$. Also $\hat{\phi}$ is a consistent estimator of ϕ, and

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{\mathrm{d}} \mathrm{N}(0, J_{\phi\phi}^{-1}), \text{ where } J_{\phi\phi} = \lim_{n\to\infty} \frac{1}{n}K_{\phi\phi}.$$

Then, $\hat{K}_{\phi\phi}^{-1} = 4\hat{\phi}^2/n(4f_g - 1)$ is a consistent estimator of the asymptotic variance of $\hat{\phi}$.

3. Restricted estimation

3.1. Equality constraints

Suppose first we are interested in estimating the parameter vector $\boldsymbol{\beta}$ under k linearly independent restrictions $\mathbf{C}_j^{\mathrm{T}}\boldsymbol{\beta} - d_j = 0$, where $\mathbf{C}_j$, $j = 1, \ldots, k$, are $p \times 1$ vectors and d_j, $j = 1, \ldots, k$, are scalars, both known fixed numbers. The problem here is to maximize the log-likelihood function $L(\boldsymbol{\theta})$ subject to the linear constraints $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$, where $\mathbf{C} = (\mathbf{C}_1, \ldots, \mathbf{C}_k)^{\mathrm{T}}$ and $\mathbf{d} = (d_1, \ldots, d_k)^{\mathrm{T}}$. Similarly to Nyquist (1991), that investigated this problem in generalized linear models, we will apply the methodology of penalty functions by considering the quadratic penalty function

$$P(\boldsymbol{\theta}, \boldsymbol{\tau}) = L(\boldsymbol{\theta}) - \frac{1}{2}\sum_{j=1}^{k} \tau_j (d_j - \mathbf{C}_j^{\mathrm{T}}\boldsymbol{\beta})^2.$$

The procedure consists in finding the solution of max $P_\theta(\boldsymbol{\theta}, \boldsymbol{\tau})$ for positive and fixed values of τ_j, $j = 1, \ldots, k$. The solution for $\boldsymbol{\beta}$ will be denoted by $\boldsymbol{\beta}(\boldsymbol{\tau})$. The equality restricted estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}^0 = \lim_{\tau_1, \ldots, \tau_k \to \infty} \boldsymbol{\beta}(\boldsymbol{\tau}).$$

Using similar approach of that given in Nyquist (1991) one may show that $\hat{\boldsymbol{\beta}}^0$ is the solution of the following iterative process:

$$\begin{aligned}\boldsymbol{\beta}^{0(r+1)} = & \{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{y} + \{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{C}^{\mathrm{T}} \\ & \times[\mathbf{C}\{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{C}^{\mathrm{T}}]^{-1}[\mathbf{d} - \mathbf{C}\{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{y}], \end{aligned} \tag{7}$$

for $r = 0, 1, \ldots$, where $\phi^{(r)}$ is obtained from (6). The iterative process (7) may be, alternatively, expressed as

$$\begin{aligned}\boldsymbol{\beta}^{0(r+1)} = & \mathbf{b}^{(r+1)} + \{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{C}^{\mathrm{T}}[\mathbf{C}\{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{v}^{(r)})\mathbf{X}\}^{-1}\mathbf{C}^{\mathrm{T}}]^{-1} \\ & \times(\mathbf{d} - \mathbf{C}\mathbf{b}^{(r+1)}), \end{aligned} \tag{8}$$

for $r = 0, 1, \ldots$, where $\mathbf{b}^{(r+1)}$ denotes $\boldsymbol{\beta}^{(r+1)}$ evaluated at the current restricted estimate. The authors have developed a library in S-Plus and R to fit symmetrical linear models based in some distributions and the iterative process (5–8) and more, some diagnostic graphics. This library is available in the web page http://www.de.ufpe.br/~cysneiros/elliptical/elliptical.html.

It may be showed under suitable regularity conditions (see, for instance, Gourieroux and Monford, 1995, Section 10.3) that $\hat{\boldsymbol{\beta}}^0$ is a consistent estimator of $\boldsymbol{\beta}$, and

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} N_p(\mathbf{0}, (\mathbf{J}^0_{\beta\beta})^{-1}),$$

where

$$\mathbf{J}^0_{\beta\beta} = \lim_{\tau_1,\ldots,\tau_k\to\infty}\left[\lim_{n\to\infty}\frac{1}{n}E\left\{-\frac{\partial \mathrm{P}(\boldsymbol{\theta}, \boldsymbol{\tau})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}\right]$$

and

$$E\left\{-\frac{\partial \mathrm{P}(\boldsymbol{\theta}, \boldsymbol{\tau})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\} = \frac{4d_g}{\phi}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{C}^{\mathrm{T}}\mathbf{D}(\boldsymbol{\tau})\mathbf{C},$$

with $\mathbf{D}(\boldsymbol{\tau}) = \mathrm{diag}\{\tau_1, \ldots, \tau_k\}$. Then, a consistent estimator of the asymptotic variance–covariance matrix of $\hat{\boldsymbol{\beta}}^0$ is given by

$$\lim_{\tau_1,\ldots,\tau_k\to\infty}\left\{\frac{4d_g}{\phi}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{C}^{\mathrm{T}}\mathbf{D}(\boldsymbol{\tau})\mathbf{C}\right\}^{-1} = \mathbf{K}^{-1}_{\beta\beta}\{\mathbf{I}_p - \mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{K}^{-1}_{\beta\beta}\mathbf{C}^{\mathrm{T}})^{-1}\mathbf{C}\mathbf{K}^{-1}_{\beta\beta}\},$$

which may be evaluated at some consistent estimator of $\boldsymbol{\beta}$, such as $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^0$.

Suppose now we have interest in testing the hypotheses $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ against $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$. The most usual methods for testing these linear hypotheses are the likelihood ratio, Wald and score (Rao) tests. The statistics become here, respectively, given by

$$\begin{aligned}
\xi^*_{LR} &= 2\{L(\hat{\boldsymbol\beta}, \hat\phi) - L(\hat{\boldsymbol\beta}^0, \hat\phi_0)\} \\
&= 2\left[\frac{n}{2}\log\left(\frac{\hat\phi_0}{\hat\phi}\right) + \sum_{i=1}^{n}\log\left\{\frac{g\{(y_i - \mathbf{x}_i^{\mathrm T}\hat{\boldsymbol\beta})^2/\hat\phi\}}{g\{(y_i - \mathbf{x}_i^{\mathrm T}\hat{\boldsymbol\beta}^0)^2/\hat\phi_0\}}\right\}\right], \\
\xi^*_{W} &= (\mathbf{C}\hat{\boldsymbol\beta} - \mathbf{d})^{\mathrm T}\hat{\mathrm{V}}\mathrm{ar}^{-1}(\mathbf{C}\hat{\boldsymbol\beta})(\mathbf{C}\hat{\boldsymbol\beta} - \mathbf{d}) \\
&= (\mathbf{C}\hat{\boldsymbol\beta} - \mathbf{d})^{\mathrm T}(\mathbf{C}\hat{\mathbf{K}}^{-1}_{\beta\beta}\mathbf{C}^{\mathrm T})^{-1}(\mathbf{C}\hat{\boldsymbol\beta} - \mathbf{d}) \\
&= \frac{4d_g}{\hat\phi}(\mathbf{C}\hat{\boldsymbol\beta} - \mathbf{d})^{\mathrm T}\{\mathbf{C}(\mathbf{X}^{\mathrm T}\mathbf{X})^{-1}\mathbf{C}^{\mathrm T}\}^{-1}(\mathbf{C}\hat{\boldsymbol\beta} - \mathbf{d}) \quad \text{and} \\
\xi^*_{SR} &= \{\mathbf{U}_\beta(\hat{\boldsymbol\beta}^0, \hat\phi_0) - \mathbf{U}_\beta(\hat{\boldsymbol\beta}, \hat\phi)\}^{\mathrm T}\hat{\mathrm{V}}\mathrm{ar}_0(\hat{\boldsymbol\beta})\{\mathbf{U}_\beta(\hat{\boldsymbol\beta}^0, \hat\phi_0) - \mathbf{U}_\beta(\hat{\boldsymbol\beta}, \hat\phi)\} \\
&= \mathbf{U}_\beta(\hat{\boldsymbol\beta}^0, \hat\phi_0)^{\mathrm T}(\hat{\mathbf{K}}^0_{\beta\beta})^{-1}\mathbf{U}_\beta(\hat{\boldsymbol\beta}^0, \hat\phi^2_0) \\
&= \frac{\hat\phi_0}{4d_g}\mathbf{U}_\beta(\hat{\boldsymbol\beta}^0, \hat\phi_0)^{\mathrm T}(\mathbf{X}^{\mathrm T}\mathbf{X})^{-1}\mathbf{U}_\beta(\hat{\boldsymbol\beta}^0, \hat\phi_0),
\end{aligned}$$

where $\hat{\mathbf{K}}_{\beta\beta}$ and $\hat{\mathbf{K}}^0_{\beta\beta}$ are evaluated at $(\hat{\boldsymbol\beta}, \hat\phi)$ and $(\hat{\boldsymbol\beta}^0, \hat\phi_0)$, respectively. It follows under H_0 and for large n that ξ^*_{LR}, ξ^*_W and ξ^*_{SR} have a central chi-squared distribution with k degrees of freedom.

3.2. Inequality constraints

The problem of maximizing log-likelihood functions restricted to linear inequality parameter constraints $\mathbf{C}\boldsymbol\beta - \mathbf{d} \geqslant \mathbf{0}$ have been investigated by various authors (see, for instance, Robertson et al., 1988; McDonald and Diamond, 1990; Fahrmeir and Klinger, 1994; Paula and Sen, 1995). Our primary interest is to obtain the maximum likelihood estimate of $\boldsymbol\beta$ in model (1) subject to the constraints $\mathbf{C}\boldsymbol\beta - \mathbf{d} \geqslant \mathbf{0}$; that is, we want to solve the problem $\max_{\{C\beta - d \geqslant 0\}} L(\boldsymbol\beta, \phi)$. We can apply the Kuhn-Tucker conditions to attain the restricted global maximum. Consider then the Lagrangian function

$$\mathscr{L}(\boldsymbol\beta, \phi) = L(\boldsymbol\beta, \phi) + \sum_{i=1}^{k}\lambda_j(\mathbf{C}_j^{\mathrm T}\boldsymbol\beta - d_j),$$

where $\boldsymbol\lambda = (\lambda_1, \ldots, \lambda_k)^{\mathrm T} \geqslant \mathbf{0}$ denotes the Lagrange multiplier vector. The sufficient conditions to guarantee that $\tilde{\boldsymbol\beta}$ corresponds to the inequality restricted estimate (see, for instance, Fahrmeir and Klinger, 1994) are given by

(i) $\mathbf{C}_j^{\mathrm T}\tilde{\boldsymbol\beta} - d_j = 0$ for $j \in \mathbf{I} \subseteq \{1, \ldots, k\}$ and $\mathbf{C}_j^{\mathrm T}\tilde{\boldsymbol\beta} > d_j$ for all $j \notin \mathbf{I}$; that is, $\tilde{\boldsymbol\beta}$ is an admissible point;

(ii) there exist $\tilde{\boldsymbol\lambda} = (\tilde\lambda_1, \ldots, \tilde\lambda_k)^{\mathrm T} \geqslant \mathbf{0}$ such that $\partial\mathscr{L}(\boldsymbol\beta, \phi)/\partial\boldsymbol\beta|_{(\tilde\beta^{\mathrm T}, \tilde\phi)^{\mathrm T}} = \mathbf{0}$; that is, $(\tilde{\boldsymbol\beta}^{\mathrm T}, \tilde\phi)^{\mathrm T}$ is a stationary point;

(iii) $\mathbf{s}^{\mathrm{T}}\ddot{\mathbf{L}}_{\theta\theta}\mathbf{s}|_{(\tilde{\beta}^{\mathrm{T}},\tilde{\phi})^{\mathrm{T}}} < \mathbf{0}$ for all $\mathbf{s} \neq \mathbf{0}$ and $\mathbf{s} \in \{\mathbf{s}|\mathbf{C}_j^{\mathrm{T}}\mathbf{s} - d_j = 0,\ j \in \mathbf{I},\ \tilde{\lambda}_j > 0$ and $\mathbf{C}_j^{\mathrm{T}}\mathbf{s} - d_j > 0,\ i \notin \mathbf{I},\ \tilde{\lambda}_j = 0\}$.

These conditions are equivalent to finding $\tilde{\boldsymbol{\beta}}$ from a searching procedure which consists in maximizing $L(\boldsymbol{\beta}, \phi)$ subject to $\mathbf{C}_j^{\mathrm{T}}\boldsymbol{\beta} - d_j = 0,\ j \in \mathbf{I}$, for each $\mathbf{I} \subseteq \{1, \ldots, k\}$. The inequality restricted estimate $\tilde{\boldsymbol{\beta}}$ is obtained from the maximization problem that fulfils conditions (i), (ii) and (iii). Thus, the inequality restricted problem reduces to a equality restricted problem that may be solved by the procedures given in Section 3.1.

The asymptotic distribution of $\tilde{\boldsymbol{\beta}}$ is not necessarily normal. It depends whether the true parameter value satisfies $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} > \mathbf{0}$ or $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$. For the first case the inequality restricted estimator coincides asymptotically with the unrestricted estimator and therefore $\tilde{\boldsymbol{\beta}}$ has the same asymptotic distribution as $\hat{\boldsymbol{\beta}}$. However, if the true value belongs to the boundary of the set of inequality parameter constraints, the asymptotic distribution of $\tilde{\boldsymbol{\beta}}$ has the form of a truncated normal distribution at $\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \mathbf{0}$ (see discussion, for instance, in Gourieroux and Monford, 1995, Section 21.1).

4. One-sided tests

4.1. Case 1

In this section we will consider the problem of testing the hypotheses $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ against $H_2 : \mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{d}$, with at least one strict inequality in H_2. The usual statistics likelihood ratio, Wald and score take, in this case, the forms

$$\begin{aligned}
\xi_{LR} &= 2[L(\tilde{\boldsymbol{\beta}}, \tilde{\phi}) - L(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0)] \\
&= 2\left[\frac{n}{2}\log\left(\frac{\hat{\phi}_0}{\tilde{\phi}}\right) + \sum_{i=1}^{n}\log\left\{\frac{g\{(y_i - \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}})^2/\tilde{\phi}\}}{g\{(y_i - \mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}^0)^2/\hat{\phi}_0\}}\right\}\right], \\
\xi_W &= (\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{d})^{\mathrm{T}}\{\mathbf{C}\tilde{\mathbf{K}}_{\beta\beta}^{-1}\mathbf{C}^{\mathrm{T}}\}^{-1}(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{d}) \\
&= \frac{4d_g}{\tilde{\phi}}(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{d})^{\mathrm{T}}\{\mathbf{C}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}^{\mathrm{T}}\}^{-1}(\mathbf{C}\tilde{\boldsymbol{\beta}} - \mathbf{d}) \quad \text{and} \\
\xi_{SR} &= \{\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0) - \mathbf{U}_\beta(\tilde{\boldsymbol{\beta}}, \tilde{\phi})\}^{\mathrm{T}}(\hat{\mathbf{K}}_{\beta\beta}^0)^{-1}\{\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0) - \mathbf{U}_\beta(\tilde{\boldsymbol{\beta}}, \tilde{\phi})\} \\
&= \frac{\hat{\phi}_0}{4d_g}\{\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0) - \mathbf{U}_\beta(\tilde{\boldsymbol{\beta}}, \tilde{\phi})\}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\{\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}^0, \hat{\phi}_0) - \mathbf{U}_\beta(\tilde{\boldsymbol{\beta}}, \tilde{\phi})\},
\end{aligned}$$

respectively. In addition, suppose the parameter space of $\boldsymbol{\beta}$ is open. Under the regular conditions given in Gourieroux and Monford (1995, Section 21.3), it follows that the statistics ξ_{LR}, ξ_W and ξ_{SR} are asymptotically equivalent as a mixture of chi-square distributions, namely

$$Pr\{\xi_{LR} \geqslant c\} = \sum_{\ell=0}^{k}\omega(k, \ell; \boldsymbol{\Delta})Pr\{\chi_\ell^2 \geqslant c\} + \mathrm{o}(1), \tag{9}$$

where $c \geqslant 0$, $\boldsymbol{\Delta} = \mathbf{C}\mathbf{K}_{\beta\beta}^{-1}\mathbf{C}^{\mathrm{T}}$, χ_0^2 denotes the degenerate distribution at the origin and $\omega(k, \ell; \boldsymbol{\Delta})$'s are known as level probabilities (see definition and expressions, for instance, in Shapiro, 1985) which are expressed as functions of correlation coefficients associated with the matrix $\boldsymbol{\Delta}$. These correlation coefficients are the minimum information necessary to compute the asymptotic null distribution given in (9) because $\omega(k, \ell; \boldsymbol{\Delta})$ depends on $\boldsymbol{\Delta}$ only through its correlation matrix. Due to the difficulty of computing the level probabilities for five or more constraints, several approximations have been proposed (see, for instance, Robertson et al., 1988, Chapter 3). Nevertheless, computational procedures for computing $w(k, \ell; \boldsymbol{\Delta})$ are available (see, for example, Bohrer and Chow, 1978 and Sun, 1988a, b). If the weights $\omega(k, \ell; \boldsymbol{\Delta})$'s do not depend on $\boldsymbol{\beta}$ through the correlation coefficients associated with the matrix $\boldsymbol{\Delta}$, then the distribution given in (9) is unique. Examining the expression of $\mathbf{K}_{\beta\beta}$ given in Section 2 we can conclude that $\boldsymbol{\Delta}$ does not depend on $\boldsymbol{\beta}$ in the class of symmetrical linear models. This properly does not follow in general. For instance, in generalized linear models it only occurs in some particular cases (see, for instance, Paula and Sen, 1995).

4.2. Case 2

Now, we will consider the hypotheses $H_2 : \mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{d}$ against $\mathbb{R}^p - H_2$. In this case, the usual statistics likelihood ratio, Wald and score take the forms

$$
\begin{aligned}
\xi_{LR}^c &= 2\{L(\hat{\boldsymbol{\beta}}, \hat{\phi}) - L(\tilde{\boldsymbol{\beta}}, \tilde{\phi})\} \\
&= 2\left[\frac{n}{2}\log\left(\frac{\tilde{\phi}}{\hat{\phi}}\right) + \sum_{i=1}^{n}\log\left\{\frac{g\{(y_i - \mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})^2/\hat{\phi}\}}{g\{(y_i - \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}})^2/\tilde{\phi}\}}\right\}\right], \\
\xi_{W}^c &= (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\tilde{\boldsymbol{\beta}})^{\mathrm{T}}\{\mathbf{C}\hat{\mathbf{K}}_{\beta\beta}^{-1}\mathbf{C}^{\mathrm{T}}\}^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\tilde{\boldsymbol{\beta}}) \\
&= \frac{4d_g}{\hat{\phi}}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\tilde{\boldsymbol{\beta}})^{\mathrm{T}}\{\mathbf{C}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C}^{\mathrm{T}}\}^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{C}\tilde{\boldsymbol{\beta}}) \quad \text{and} \\
\xi_{SR}^c &= \mathbf{U}_{\beta}(\tilde{\boldsymbol{\beta}}, \tilde{\phi})^{\mathrm{T}}(\tilde{\mathbf{K}}_{\beta\beta})^{-1}\mathbf{U}_{\beta}(\tilde{\boldsymbol{\beta}}, \tilde{\phi}) \\
&= \frac{\tilde{\phi}}{4d_g}\mathbf{U}_{\beta}(\tilde{\boldsymbol{\beta}}, \tilde{\phi})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{U}_{\beta}(\tilde{\boldsymbol{\beta}}, \tilde{\phi})^{\mathrm{T}}.
\end{aligned}
$$

Here we should search for the least favorable distribution $\sup_{\{C\beta \geqslant d\}} Pr\{\xi_{LR}^c \geqslant c\}$. However, due to the lack of functional dependence of $\boldsymbol{\Delta} = \mathbf{C}\mathbf{K}_{\beta\beta}^{-1}\mathbf{C}^{\mathrm{T}}$ on $\boldsymbol{\beta}$ the least favorable asymptotic null distribution of ξ_{LR}^c, ξ_W^c and ξ_{SR}^c for the purpose of testing H_2 against $\mathbb{R}^p - H_2$ is attained at $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ (see, for instance, Wolak, 1991). This distribution is uniquely determined and given by

$$
Pr\{\xi_{LR}^c \geqslant c\} = \sum_{\ell=0}^{k}\omega(k, k-\ell; \boldsymbol{\Delta})Pr\{\chi_\ell^2 \geqslant c\} + \mathrm{o}(1), \tag{10}
$$

where $c \geqslant 0$. When $\boldsymbol{\Delta}$ depends on $\boldsymbol{\beta}$ the asymptotic null distribution of ξ_{LR}^c is much more complicated than (10). A search algorithm should be required in these cases to find the least favorable situation under the null hypothesis.

5. Sensitivity study

It is well known that maximum likelihood estimates from symmetrical models with error distributions presenting heavier tails than the normal ones tend to be less sensitive to outlying observations. However, few has been investigated on the robustness of such estimates against influential or high leverage observations in these cases. In order to clarify this point for some symmetrical models we will present in the sequel a small sensitivity study in which a particular observation is perturbed in the sense of becoming high leverage in the simple linear regression $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $\varepsilon_i \sim \mathrm{S}(0, \phi)$. Then, the behavior of an appropriate influence measure is studied.

For some particular settings for β_0, β_1 and ϕ and $n = 50$, x_i was obtained from an exponential distribution of mean 1. We consider for illustration ε_i following normal, Student-t with 12 and 3 degrees of freedom, power exponential with $k=0.3$ and $k=0.6$ and logistic-II distributions. For the power exponential distribution with $k > 0$ the kurtosis coefficient γ_2 is positive and increases as k increases. For the logistic-II distribution one has $\gamma_2 = 1.2$ while for normal error $\gamma_2 = 0$. After generating the explanatory variable x_i we made a perturbation scheme in the largest explanatory variable value, $x_{\max}$, in order to make it a high leverage point, namely $x_{\max} \leftarrow x_{\max} + \delta\sigma_x$ for $\delta \in [0, 3]$. Then, we calculated the measure $\mathrm{W}(\delta) = \{\beta_1 - \hat{\beta}_1(\delta)\}^2 / \mathrm{Var}\{\hat{\beta}_1(\delta)\}$, where β_1 denotes the true value, $\hat{\beta}_1(\delta)$ denotes the estimate of β_1 under the perturbation δ and $\mathrm{Var}\{\hat{\beta}_1(\delta)\}$ is the approximate variance of $\hat{\beta}_1(\delta)$. Although $\mathrm{W}(\delta)$ assumes a similar form of the Wald statistic we are indeed assessing the distance between β_1 and $\hat{\beta}_1(\delta)$ under the metric $1/\mathrm{Var}\{\hat{\beta}_1(\delta)\}$. In order to confirm the high leverage of $x_{\max}$ under $\delta = 3$, we calculate the principal diagonal elements of the leverage matrix $\mathbf{H} = (\partial\hat{\mathbf{y}}/\partial\mathbf{y}^{\mathrm{T}})$. From Wei et al. (1998) we find that $\mathbf{H} = \mathbf{X}\{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{a})\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}(\mathbf{a})$. The index plot of $\hat{h}_{ii} = \hat{a}_i \mathbf{x}_i^{\mathrm{T}}\{\mathbf{X}^{\mathrm{T}}\mathbf{D}(\hat{\mathbf{a}})\mathbf{X}\}^{-1}\mathbf{x}_i$ for $\delta = 3$ is presented in Fig. 1 and the behavior of $\mathrm{W}(\delta)$ is described in Fig. 2 for the particular setting of $\beta_0 = 1$, $\beta_1 = 2$ and $\phi = 2$. As we can see from this last figure the maximum likelihood estimates of $\hat{\beta}_1$ from the logistic II, power exponential with $k = 0.6$ and $k = 0.3$ and Student-t with 3 and 12 degrees of freedom, respectively, seem to be more robust against the perturbation scheme and consequently against high leverage points. Similar tendencies were observed for other configurations of β_0, β_1 and ϕ.

6. Example

In this section we will reanalyze the example discussed by Ramanathan (1993) on a study in which seven variables were observed in 40 metropolitan areas (see Table 10.1 Ramanathan, 1993). The main interest is on regressing the number (in thousands) of subscribers with cable TV (Y) against the number (in thousands) of homes in the area (X_1), the per capita income for each television market with cable (X_2), the installation fee (X_3), the monthly service charge (X_4), the number of television signals carried by each cable system (X_5) and the number of television signals received with good quality without cable (X_6). Because Y corresponds to count data we will use a square root transformation in order to

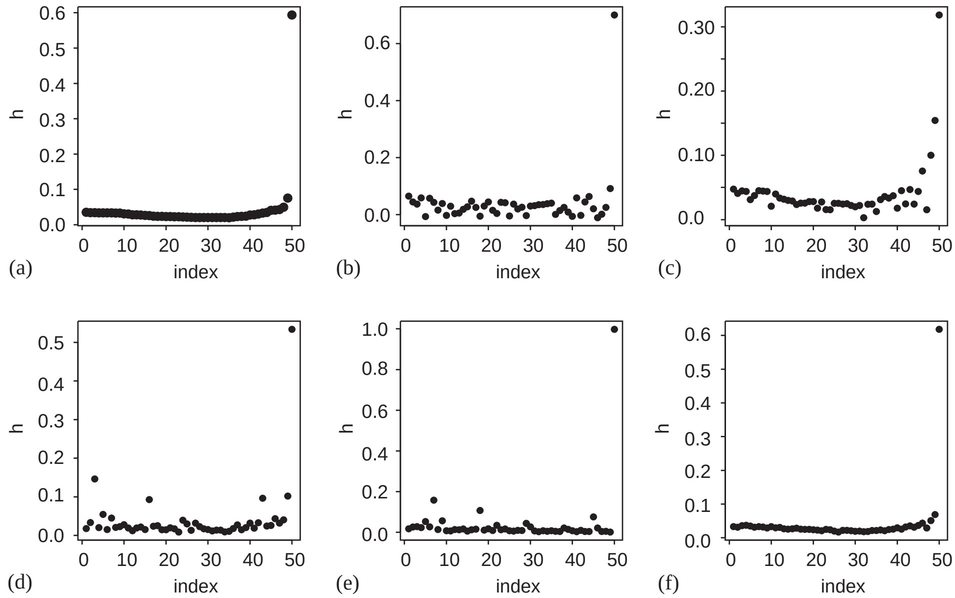

Fig. 1. Index plot of $\hat{h}_{ii}$ for the parameter estimates of the symmetrical perturbed models ($\delta = 3$) under errors (a) normal, (b) Student-t with 3 d.f., (c) Student-t with 6 d.f., (d) PE(0.3), (e) PE(0.6) and (f) Logistic-II.

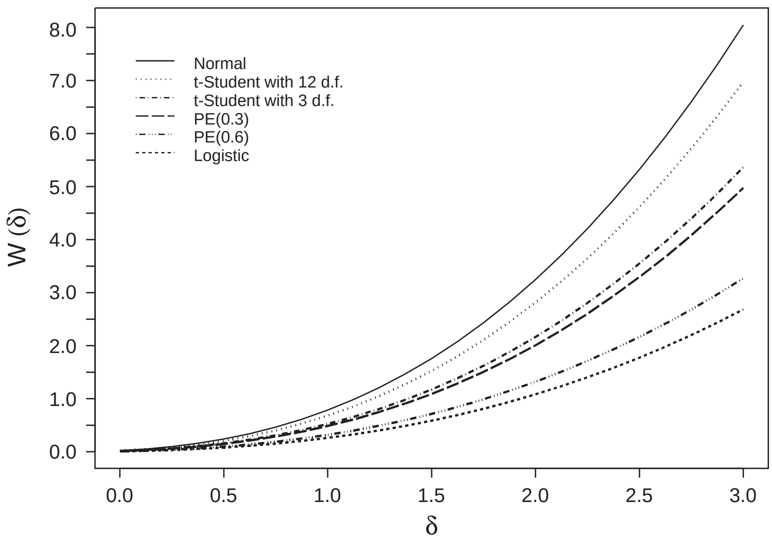

Fig. 2. Behavior of the distance $W(\delta)$ under perturbations in the largest explanatory variable value.

Table 3
Unrestricted maximum likelihood estimates (standard errors in parenthesis)

Parameter	Normal	t_6	PE(0.3)	Logistic − II
β_0	2.319	3.335	2.635	3.122
	(2.233)	(1.866)	(1.939)	(1.907)
β_1	0.034	0.035	0.034	0.034
	(0.002)	(0.002)	(0.002)	(0.002)
β_2	0.0002	0.0001	0.0002	0.0001
	(0.0003)	(0.0002)	(0.0002)	(0.0002)
β_3	0.035	0.010	0.023	0.014
	(0.040)	(0.033)	(0.034)	(0.034)
β_4	−0.245	−0.318	−0.268	−0.301
	(0.182)	(0.152)	(0.158)	(0.155)
β_5	0.134	0.118	0.122	0.119
	(0.059)	(0.049)	(0.052)	(0.051)
β_6	−0.361	−0.319	−0.335	−0.327
	(0.134)	(0.111)	(0.116)	(0.114)
ϕ	1.015	0.665	0.573	0.298
	(0.227)	(0.182)	(0.146)	(0.078)

try to stabilize the variance. Then, we propose the model

$$\sqrt{y_i} = \beta_0 + \sum_{j=1}^{6} \beta_j x_{ij} + \varepsilon_i, \qquad i = 1, \ldots, 40, \tag{11}$$

where $\varepsilon_i \sim S(0, \phi)$ are mutually independent errors. In addition, it is reasonable to expect in this example that the effect of each coefficient is unidirectional, although the opposite direction is not theoretically impossible. For instance, one may have interest in assessing if the number of subscribers decreases as the monthly service charge increases, that is, to assess $H_0 : \beta_4 = 0$ against $H_2 : \beta_4 < 0$. Following the same idea for the remaining variables one may have interest in assessing the directions $\beta_1 \geqslant 0$, $\beta_2 \geqslant 0$, $\beta_3 \leqslant 0$, $\beta_5 \geqslant 0$ and $\beta_6 \leqslant 0$.

6.1. Analysis under normal error

We first fitted the model (11) by assuming normal errors. The unrestricted estimates are given in the first column of Table 3 . Applying one-sided t tests we do not reject the hypotheses of each coefficient β_2, β_3 and β_4 be equal to zero, at the significance level of 5%, while some doubt appears for the coefficient β_5 for which the p-value is about 3%. The remaining coefficients β_1 and β_6 are highly significant in the one-sided directions. The only estimated coefficient with opposite sign is $\hat{\beta}_3$, but it is not due to multicollinearity that is negligible in this example. Thus, in order to assess if at least one of the coefficients β_2, β_3, β_4 and β_5 is in the one-sided direction, we apply the statistical tests defined in Section 4.1 to assess, the hypotheses $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ against $H_2 : \mathbf{C}\boldsymbol{\beta} \geqslant \mathbf{0}$, with at least one strict inequality

Table 4
Statistical test values and the *p*-values (in parenthesis)

Statistics	Normal	t_6	PE(0.3)	Logistic − II
ξ^*_{SR}	6.61	8.54	9.11	8.50
	(0.16)	(0.07)	(0.06)	(0.07)
ξ^*_{LR}	7.23	8.65	8.33	8.32
	(0.12)	(0.07)	(0.08)	(0.08)
ξ^*_{W}	7.92	8.87	7.7	8.21
	(0.09)	(0.06)	(0.10)	(0.08)
ξ_{SR}	5.87	8.39	8.66	8.26
	(0.06)	(0.02)	(0.02)	(0.02)
ξ_{LR}	6.33	8.56	7.94	8.14
	(0.05)	(0.02)	(0.02)	(0.02)
ξ_{W}	6.85	8.90	7.54	8.14
	(0.04)	(0.01)	(0.03)	(0.02)

in H_2, where

$$\mathbf{C}=\begin{bmatrix}0&0&1&0&0&0&0\\0&0&0&-1&0&0&0\\0&0&0&0&-1&0&0\\0&0&0&0&0&1&0\end{bmatrix}.$$

The results from the statistical tests (*p*-values in parenthesis) are available in the first column of Table 4 . It may be showed that the asymptotic null distribution of the statistics ξ_{LR}, ξ_{SR} and ξ_W is a mixture of chi-squared distributions with weights $\omega(0,4;\mathbf{\Delta})=0.0714$, $\omega(1,4;\mathbf{\Delta})=0.2610$, $\omega(2,4;\mathbf{\Delta})=0.3728$, $\omega(3,4;\mathbf{\Delta})=0.2389$ and $\omega(4,4;\mathbf{\Delta})=0.0556$. The results indicate that the null hypothesis is not in general rejected, at the level of 10% for the two-sided tests and at the level of 5% for the one-sided tests.

However, due to the lack of robustness of the least-squared estimates against outlying observations we performed some residual analysis. Fig. 3a presents the plot of $r_{s_i}=(y_i-\hat{y}_i)/\sqrt{\hat{\phi}\xi}$, $i=1,\ldots,n$, against the fitted values. The graphic does not give indication of any systematic tendency suggesting that $\sqrt{Y}$ should stabilize the variance of the errors. Nevertheless, area 14 appears with a large residual value (greater than 3) indicating the possible influence of this observation on the decisions from the statistical tests. The generated envelope, as proposed by Atkinson (1981), is presented in Fig. 4a noticing that the assumption of normal distribution for the errors does not seem to be inappropriate, even though area 14 appears outside the envelope. Elimination of this area reduces drastically the *p*-values of the statistical tests as we can see from the first column of Table 5 . Indeed, area 14 has a high installation fee and a relatively high monthly service charge which are in disagreement with the high proportion of homes with cable TV. The index plot of the local total influence C_i (see, for instance, Lesaffre and Verbeke, 1998; Galea et al., 2003) is given in Fig. 5a for the normal case. As we can see from this figure areas 14 and 1 appear as the most influential observations. Elimination of area 1 makes all the statistical tests non-significant (see the first column of Table 6). This area has a small number of homes with cable TV but a large number of television signals received with cable.

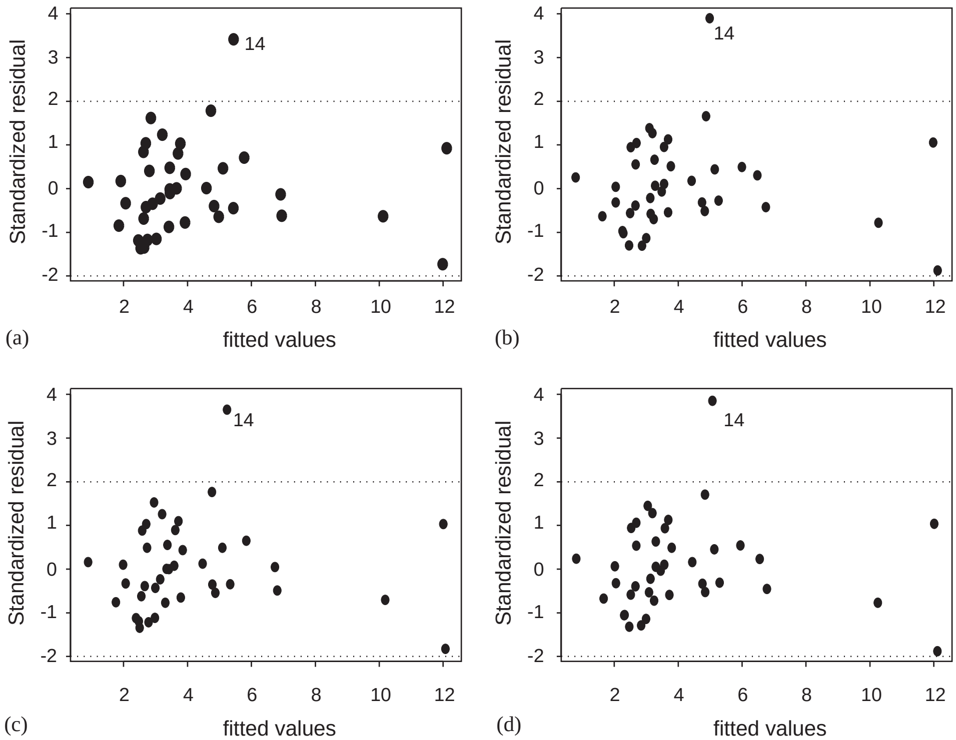

Fig. 3. Plots of r_{s_i} against the fitted values for the symmetrical model (11) under errors (a) normal, (b) Student-t with 6 d.f., (c) PE(0.3) and (d) Logistic-II.

Table 5
Statistical test values and the p-values (in parenthesis) dropping area 14

Statistics	Normal	t_6	PE(0.3)	Logistic − II
ξ^*_{SR}	10.90	11.14	11.73	11.38
	(0.03)	(0.02)	(0.02)	(0.02)
ξ^*_{LR}	12.79	11.42	12.42	11.67
	(0.01)	(0.02)	(0.01)	(0.02)
ξ^*_{W}	15.13	11.86	13.97	12.21
	(0.00)	(0.02)	(0.01)	(0.01)
ξ_{SR}	10.90	11.14	11.73	11.38
	(0.01)	(0.00)	(0.00)	(0.00)
ξ_{LR}	12.79	11.42	12.42	11.67
	(0.00)	(0.00)	(0.00)	(0.00)
ξ_{W}	15.13	11.86	13.97	12.21
	(0.00)	(0.00)	(0.00)	(0.00)

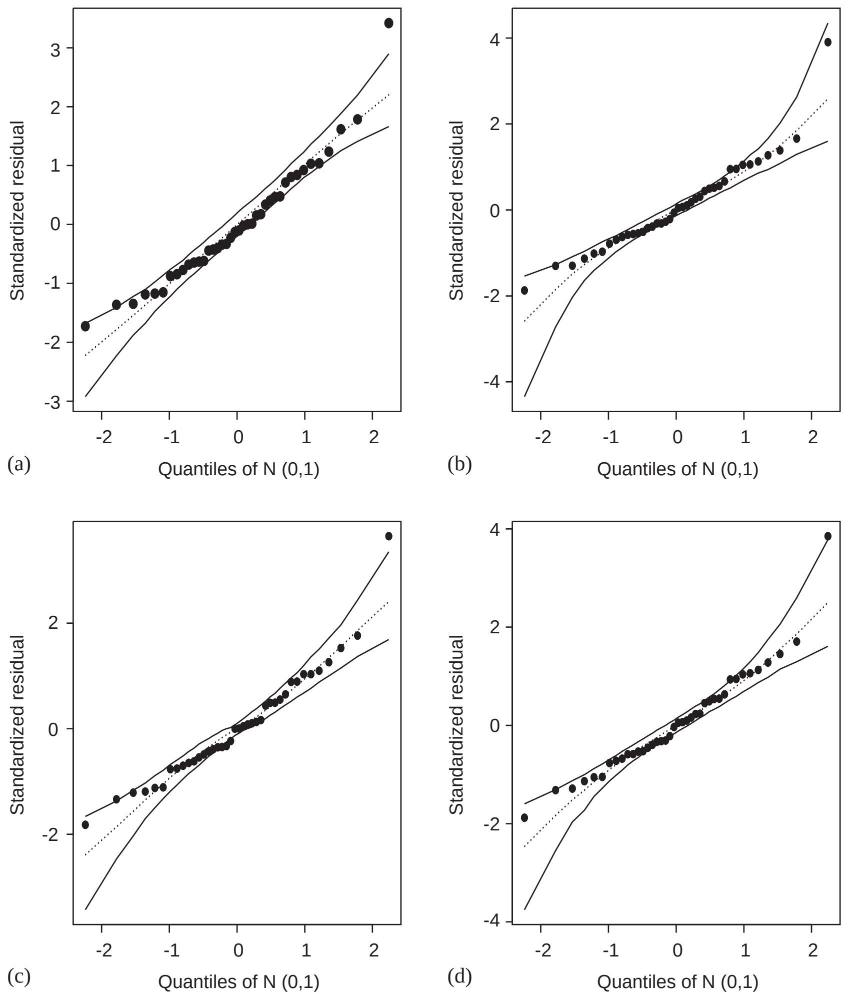

Fig. 4. Normal probability plots for the residual r_{s_i} for the symmetrical model (11) under errors (a) normal, (b) Student-t with 6 d.f., (c) PE(0.3) and (d) Logistic-II.

6.2. Analysis under other symmetrical errors

In order to accommodate areas 1 and 14 or at least to reduce their influence on the results from the statistical tests we refitted model (11) by assuming distributions for the errors with heavier tails than the ones of the normal distribution. First, we fitted a model with errors following a Student-t distribution with v degrees of freedom. If we assume $v > 4$, then a consistent estimate for v can be obtained from the residuals $r_i = y_i - \hat{y}_i$, $i = 1, \ldots, n$. This estimate is given by $\hat{v} = (4\hat{m}_{2,1} - 6)/(4\hat{m}_{2,1} - 3)$, where $\hat{m}_{2,1} = (1/n\sum_{i=1}^{n} r_i^4)/(1/n\sum_{i=1}^{n} r_i^2)^2$ (Arellano-Valle, 1994). For the data set of the example above we find $\hat{v} \approx 6$. The unrestricted

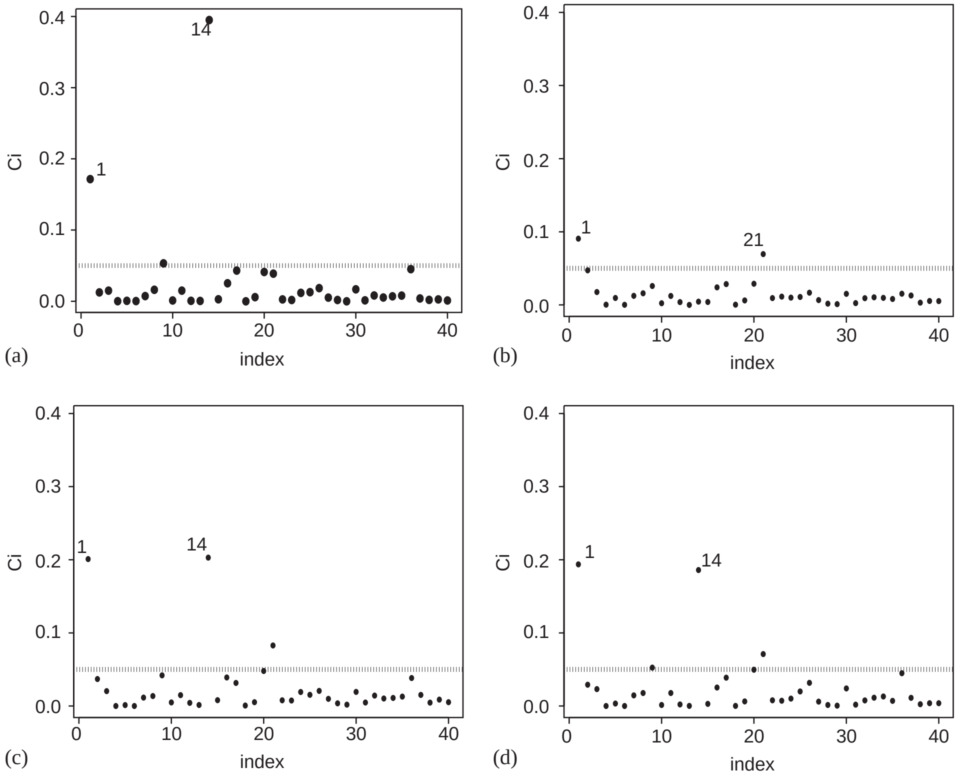

Fig. 5. Index plot of C_i for the parameter estimates of the symmetrical model (11) under errors (a) normal, (b) Student-t with 6 d.f., (c) PE(0.3) and (d) Logistic-II.

Table 6
Statistical test values and the p-values (in parenthesis) dropping area 1.

Statistics	Normal	t_6	PE(0.3)	Logistic − II
ξ^*_{SR}	3.10	6.39	5.52	5.89
	(0.54)	(0.17)	(0.24)	(0.21)
ξ^*_{LR}	3.23	6.40	5.02	5.66
	(0.52)	(0.17)	(0.28)	(0.22)
ξ^*_{W}	3.36	6.78	4.39	5.56
	(0.50)	(0.15)	(0.35)	(0.23)
ξ_{SR}	2.84	6.39	5.48	5.89
	(0.24)	(0.05)	(0.07)	(0.06)
ξ_{LR}	2.95	6.40	5.00	5.66
	(0.23)	(0.05)	(0.09)	(0.07)
ξ_{W}	3.06	6.80	4.45	5.56
	(0.22)	(0.04)	(0.12)	(0.07)

maximum likelihood estimates for the parameters under the Student-t distribution with $v=6$ degrees of freedom for the errors are given in the second column of Table 3. The values of the statistical tests for both two-sided and one-sided tests, as described in the second column of Table 4, indicate for the rejection of the null hypothesis at the significance levels of 10% and 5%, respectively. The residual analyses under the t model are described in Figs. 3b and 4b, indicating that area 14 which appears with a large residual as in the normal case is here accommodated into the envelope. The generated envelope for the Student-t model does not present any unusual feature. If we eliminate area 14 the values of the statistical tests, as described in the second column of Table 5, do not change as in the normal case confirming the robustness of the Student-t distribution against outlying observations. Fig. 5b presents the index plot of C_i for the Student-t model and as we can observe from this figure areas 1 and 21 appear with large influence. Elimination of area 1 (see Table 6) changes the decision based on the two-sided tests but does not change much the p-values from the one-sided tests.

Two other error distributions with heavier tails than the normal were also assumed, power exponential with $k=0.3$ and logistic-II models. We take arbitrarily $k=0.3$ in order to try accommodating the outlying observation 14. The unrestricted maximum likelihood estimates for the parameters of the power exponential and logistic-II models are, respectively, given in the third and fourth columns of Table 3. The values of the statistical tests are presented in the third and fourth columns of Table 4. We can notice a similarity among the results for these two models and the Student-t model with $v = 6$ degrees of freedom. Looking at Figs. 3c, 3d, 4c and 4d we can observe that area 14 also appears as an outlying observation under these two models, but the generated envelopes have similar behaviors to the one of the Student-t model. The index plot of C_i given in Figs. 5c and 5d confirm the influence of areas 1 and 14. Elimination of area 1 changes more the p-values from the statistical tests for the power exponential model, in the sense of non rejecting the null hypothesis, than under the Student-t and logistic-II models.

Our main conclusion for this example is that the transformation $\sqrt{Y}$ seems to stabilize the variance of the response, but the Student-t, power exponential and logistic-II models are less influenced by the outlying observation 14. The one-sided tests based on these three fitted models indicate for the rejection of the null hypothesis at the significance level of 5% while under the normal model the rejection becomes more evident after dropping the outlying observation 14. However, the Student-t model seems to be more robust against the influential observation 1 than the other three models. If we continue the selection process with this model only the coefficients β_2 and β_3 are removed from the model. Thus, the final model becomes given by $\sqrt{y_i} = \mu_i + \varepsilon_i$, where $\varepsilon_i \sim t_6(\mu_i, \phi)$ with $\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6}$ and $\hat{\beta}_0 = 4.727(0.993)$, $\hat{\beta}_1 = 0.035(0.002)$, $\hat{\beta}_4 = -0.293(0.147)$, $\hat{\beta}_5 = 0.111(0.049)$, $\hat{\beta}_6 = -0.263(0.092)$ and $\hat{\phi} = 0.672(0.184)$.

7. Concluding remarks

In this paper we develop iterative processes relatively simple to be implemented for evaluating restricted maximum likelihood estimates for the parameters in symmetrical linear regression models. We have developed codes in S-Plus and R to fit restricted and unrestricted symmetrical linear regression models, which are available in the web page http://www.de.ufpe.br/∼cysneiros/elliptical/elliptical.html. We also verified, under some regularity conditions, that the asymptotic null distribution of the one-sided statistical tests is a mixture of chi-square distributions. The asymptotic null distribution is unique for the cases 1 and 2.

Acknowledgements

The first author received financial support from CAPES and the second author was supported by FAPESP and CNPq, Brazil. The authors are grateful to one Referee for helpful comments and suggestions.

References

Arellano-Valle, R.B., 1994. Elliptical Distribution: Properties and Applications in Regression Models. Unpublished Ph.D.Thesis, Department of Statistics, University of São Paulo, Brazil.

Atkinson, A.C., 1981. Two graphical display for outlying and influential observations in regression. Biometrika 68, 13–20.

Bartholomew, D.J., 1959a. A test of homogeneity for ordered alternatives. I. Biometrika 46, 36–48.

Bartholomew, D.J., 1959b. A test of homogeneity for ordered alternatives. II. Biometrika 46, 328–335.

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

Bohrer, R., Chow, W., 1978. Algorithm AS122. Weights for one-sided multivariate inference. Applied Statistics 27, 100–104.

Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman & Hall, London.

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

Fahrmeir, L., Klinger, J., 1994. Estimating and testing generalized linear models under inequality restrictions. Statistical Papers 35, 211–229.

Fang, K.T., Anderson, T.W., 1990. Statistical Inference in Elliptical Contoured and Related Distributions. Allerton Press, New York.

Fang, K.T., Kotz, S., Ng, K.W, 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London.

Galea, M., Paula, G.A., Bolfarine, H., 1997. Local influence in elliptical linear regression models. The Statistician 46, 71–79.

Galea, M., Paula, G.A., Uribe-Opazo, M., 2003. On influence diagnostic in symmetrical linear regression models. Statistical Papers 44, 23–45.

Gourieroux, C., Holly, A., Monford, A., 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. Econometrica 50, 63–80.

Gourieroux, G., Monford, A., 1995. Statistics and Econometric Models, Vols. 1 and 2. Cambridge University Press, Cambridge.

Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restrictions. Econometrics 54, 1243–1248.

Kudo, N.M., 1963. A multivariate analogue of the one-sided test. Biometrika 50, 403–418.

Lange, K.L., Litte, R.J.A., Taylor, J.M.G., 1989. Robust statistical modeling using the *t* distribution. Journal of the American Statistical Association 84, 881–896.

Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York.

Lesaffre, F., Verbeke, G., 1998. Local influence in linear mixed models. Biometrics 57, 1166–1172.

McDonald, J.M., Diamond, I., 1990. On the fitting of generalized linear models with non-negativity parameter constraints. Biometrics 46, 201–206.

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

Muirhead, R., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Nyquist, H., 1991. Restricted estimation of generalized linear models. Applied Statistics 40, 133–141.

Nuesch, P.E., 1966. On the problem of testing location in multivariate populations for restricted alternatives. Annals of Mathematical Statistics 37, 113–119.

Paula, G.A., Sen, P.K., 1995. One-sided tests in generalized linear models with parallel regression lines. Biometrics 51, 1494–1501.

Perlman, M.D., 1969. One-sided problems in multivariate analysis. Annals of Mathematical Statistics 40, 549–567.

Pratt, J.W., 1981. Concavity of the log likelihood. Journal of the American Statistical Association 76, 103–106.

Ramanathan, R., 1993. Statistical Methods in Econometrics. Wiley, New York.

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

Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley, New York.

Sen, P.K., Silvapulle, M.J., 2002. An appraisal of some aspects of statistical inference under inequality constraints. Journal of Statistical Planning and Inference 107, 3–43.

Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Shapiro, A., 1985. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika 72, 133–144.

Silvapulle, M.J., Silvapulle, P., 1995. A score test against one-sided alternatives. Journal of the American Statistical Association 90, 342–345.

Sun, H.J., 1988a. A general reduction method for n-variate normal orthant probability. Communications in Statistics Theory and Methods 17, 3913–3921.

Sun, H.J., 1988b. A Fortran subroutine for computing normal orthant probabilities. Communications in Statistics, Simula 17, 1097–1111.

Wei, B.C., Hu, Y.Q., Fung, W.K., 1998. Generalized leverage and its applications. Scandinavian Journal of Statistics 25, 25–37.

Wolak, F.A., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association 82, 782–793.

Wolak, F.A., 1989. Testing inequality constraints in linear econometric models. Journal of Econometrics 41, 205–235.

Wolak, F.A., 1991. The local and global nature of hypothesis test involving inequality constraints in nonlinear models. Econometrika 59, 981–995.