*Brazilian Journal of Probability and Statistics* (2000), 14, pp. 167–184.


# DIAGNOSTIC METHODS IN ELLIPTICAL LINEAR REGRESSION MODELS

**Manuel Galea** and **Marco Riquelme**
*Universidad de Valparaíso*
*Chile*
*Manuel.Galea@uv.cl*

**Gilberto A. Paula**
*Instituto de Matemática e Estatística*
*Universidade de São Paulo*
*Caixa Postal 66281*
*05315-970 - São Paulo - SP - Brasil*
*giapaula@ime.usp.br*

**Summary**
We discuss in this paper the development of various diagnostic methods in multivariate elliptical linear regression models. In particular, we show the invariance property of some usual standardized residuals in the elliptical class of distributions. This invariance is also verified for some influence measures of dropping observations, such as the Cook distance. We also discuss the computation of the likelihood displacement as well as the normal curvature in the local influence method. An example with real data is given for illustration.

**Key Words**: Elliptical distributions; Influence diagnostic; Likelihood displacement; Local influence; Multivariate elliptical distributions; Residuals.

## 1 Introduction

Diagnostic methods for the normal linear regression model have been largely investigated in the statistical literature (see, for instance, Belsley et al., 1980; Cook and Weisberg, 1982; Atkinson, 1985 and Chatterjee and Hadi, 1988). The majority of the works have given emphasis in studying the effect of eliminating observations on the results from the fitted model, particularly the parameter estimates. Alternatively, Cook (1986) has proposed the local influence method to assess the effect on the parameter estimates of small perturbations in the model. Several authors have extended the local influence method to various regression models. Some references are Beckman et al.(1987), Lawrence (1988), Thomas and Cook (1990), Tsai and Wu (1992), Paula (1993), Kim (1995), Galea at al. (1997), Fung and Kwan (1997) among others.

The aim of this paper is to discuss the development of some traditional diagnostic methods, such as the deletion of individual observations, residual analysis and local influence in multivariate elliptical linear regression models. When the interest is only on the regression coefficients the diagnostic graphics are in general invariant with the error distribution so that the well known graphics developed for the normal linear case can be applied in the elliptical class. However, when the interest is also on the dispersion parameter the diagnostic graphics depend on the error distribution. In the application we compare the behavior of some graphics for two particular models, Student-t and exponential power.

## 2 Elliptical linear regression model

The class of elliptical distributions has received an increasing attention in the statistical literature (see, for instance, Fang et al, 1990; Fang and Zhang, 1990; Fang and Anderson, 1990; Gupta and Varga, 1993; Arellano, 1994 and Leiva, 1998). We say that an $(n \times 1)$ random vector $\mathbf{Y}$ has an elliptical distribution with an $(n \times 1)$ location parameter $\boldsymbol{\mu}$ and an $(n \times n)$ scale matrix $\boldsymbol{\Sigma}$ if its density function is expressed as

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} g[(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})], \tag{2.1}$$

$\mathbf{y} \in \mathcal{R}^n$, where the function $g : \mathcal{R} \to [0, \infty)$ is such that $\int_0^\infty u^{n-1} g(u^2) du < \infty$. The function $g(\cdot)$ is typically known as the density generator. For a vector $\mathbf{Y}$ distributed according to the density (2.1) we use the notation $\mathbf{Y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, or, simply, $\mathbf{Y} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In the case where $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$, we obtain the spherical family of densities. This class of symmetric distributions includes the normal, Student-t, contaminated normal and logistic (both, univariate and multivariate), among others, as considered, for example, by Fang et al. (1990). Table 1 below, taken from Fang et al. (1990), reports examples of distributions in the elliptical family. The notation $c_1, c_2, c_3, c_4$ and $c_5$ is used to denote normalizing constants.

Some properties of the elliptical distributions may be found, for instance, in Fang et al.(1990). A particular property given below will be useful in some demonstrations.

Suppose $\mathbf{Y} \sim El_n(\mathbf{0}, \mathbf{I})$ and let $\mathbf{T}(\mathbf{Y})$ be a statistic. Then, from Theorem 2.22 of Fang et al. (1990, p.51), the distribution of $\mathbf{T}(\mathbf{Y})$ remains unchanged as long as $\mathbf{Y} \sim El_n(\mathbf{0}, \mathbf{I})$ provided that

$$\mathbf{T}(k\mathbf{Y}) \stackrel{d}{=} \mathbf{T}(\mathbf{Y}), \forall k > 0,$$

**Table 1**
*Multivariate elliptical distributions*

| Distribution | Notation | Generating function |
|---|---|---|
| Normal | $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | $g(u) = c_1 \mathrm{e}^{-u/2}, \quad u \geq 0$ |
| Student-$t$ | $t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ | $g(u) = c_2(1 + u/\nu)^{-(\nu+n)/2}, \quad u \geq 0$ |
| Contaminated Normal | $CN_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta, \tau)$ | $g(u) = c_1\{(1-\delta)\mathrm{e}^{-u/2} + \delta\tau^{-n/2}\mathrm{e}^{-u/(2\tau)}\}, \quad u \geq 0$ |
| Cauchy | $C_n(\mu, \boldsymbol{\Sigma})$ | $g(u) = c_3(1 + u)^{-(n+1)/2}, \quad u \geq 0$ |
| Logistic | $L_n(\mu, \boldsymbol{\Sigma})$ | $g(u) = c_4 \mathrm{e}^{-u}/(1 + \mathrm{e}^{-u})^2, \quad u \geq 0$ |
| Exponential Power | $PE_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha)$ | $g(u) = c_5 \mathrm{e}^{-u^\alpha/2}, \quad u \geq 0$ |

where the operator $\overset{d}{=}$ indicates the same distribution. In this case $\mathbf{T}(\mathbf{Y}) \overset{d}{=} \mathbf{T}(\mathbf{Z})$, where $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$.

Consider now the linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{Y}$ is a $(n \times 1)$ vector of responses, $\mathbf{X}$ is a known $(n \times p)$ matrix of rank $p$, $\boldsymbol{\beta}$ is a $p$-dimensional vector of parameters and $\boldsymbol{\epsilon}$ is a $p$-dimensional error vector with distribution $El_n(\mathbf{0}, \phi\mathbf{I})$, where $\phi$ is the scale parameter. Thus, it follows that $\mathbf{Y} \sim El_n(\mathbf{X}\boldsymbol{\beta}, \phi\mathbf{I})$. This is typically called the elliptical linear regression model. If $g(\cdot)$ is a continuous and decreasing function then the maximum likelihood estimators of $\boldsymbol{\beta}$ and $\phi$ are given, respectively, by (see Fang and Anderson, 1990)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad \text{and} \quad \hat{\phi} = Q(\hat{\boldsymbol{\beta}})/u_0,$$

where $Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ and $u_0$ maximizes the function

$$h_n(u) = u^{n/2}g(u), \quad u \geq 0. \tag{2.2}$$

Typically, if $g(\cdot)$ in (2.2) is continuous and decreasing then its maximum $u_0$ exists and it is finite and positive. Moreover, if $g(\cdot)$ is continuous and differentiable then $u_0$ is the solution to (Fang and Anderson, 1990)

$$g'(u) + \frac{n}{2u}g(u) = 0,$$

or, equivalently, the solution to the equation

$$\frac{n}{2u} + W_g(u) = 0, \tag{2.3}$$

where $W_g(u) = d \log g(u)/du = g'(u)/g(u)$. It is easy to see that for the normal and Student-t distributions $u_0 = n$, while for the exponential power $u_0 = (n/\alpha)^{1/\alpha}$. However, for the contaminated normal and logistic distributions, $u_0$ has to be obtained numerically. In the case of the logistic distribution, for example, equation (2.3) yields

$$\frac{n}{2u} = \tanh\left(\frac{u}{2}\right),$$

where $\tanh(\cdot)$ denotes the hyperbolic tangent.

Using properties of the elliptical distributions we may show that

$$\hat{\boldsymbol{\beta}} \sim El_n(\mathbf{X}\boldsymbol{\beta}, \phi(\mathbf{X}^T\mathbf{X})^{-1})$$

and

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{ps^2} \sim F_{p,(n-p)},$$

where $F_{p,(n-p)}$ denotes the $F$ distribution with $p$ and $(n-p)$ degrees of freedom and $s^2 = Q(\hat{\boldsymbol{\beta}})/(n-p) = u_0\hat{\phi}/(n-p)$. Then, an $100(1-\gamma)\%$ confidence region for $\boldsymbol{\beta}$, where $0 < \gamma < 1$, is given by

$$\mathbf{R} = \{\boldsymbol{\beta} \in I\!R^p : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq ps^2 F_{p,(n-p)}(1-\gamma)\}, \tag{2.4}$$

where $F_{p,(n-p)}(1-\gamma)$ denotes the $100(1-\gamma)th$ quantile of the $F_{p,(n-p)}$ distribution.

In addition, the likelihood ratio statistic for testing $H_0 : \mathbf{A}\boldsymbol{\beta} = \mathbf{C}$ against $H_1 : \mathbf{A}\boldsymbol{\beta} \neq \mathbf{C}$, where $\mathbf{A}$ is a $(q \times p)$ matrix of rank $q$ and $\mathbf{C}$ is a $(q \times 1)$ vector of constants, takes the form

$$\lambda = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{C})^T\{\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T\}^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{C})}{\mathbf{Y}^T(\mathbf{I}_n - \mathbf{P})\mathbf{Y}}, \tag{2.5}$$

where $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, $(n-p)\lambda/q \overset{H_0}{\sim} F_{q,(n-p)}$ and $\lambda$ is independent of $g(\cdot)$.

# 3 Effects of individual observations

## 3.1 Residuals

In this section we discuss some properties of two standardized forms for the ordinary residual in elliptical linear regression models. The vector of ordinary residuals is defined by $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$. Then, it follows that

$$\mathbf{e} \sim El_n(\mathbf{0}, \phi(\mathbf{I} - \mathbf{P})),$$

and in particular $e_i \sim El(0, \phi(1 - p_{ii}))$, $i = 1, \ldots, n$. We may define two standardized versions for the residual $e_i$, namely

$$r_i = \frac{e_i}{s\sqrt{1 - p_{ii}}}$$

and

$$t_i = \frac{e_i}{s_{(i)}\sqrt{1 - p_{ii}}},$$

$i = 1, \ldots, n$, where $s = \sqrt{u_0\hat{\phi}/(n-p)}$ and $s_{(i)} = \sqrt{u_0^*\hat{\phi}_{(i)}/(n-p-1)}$, with $u_0^*$ denoting the maximum of the function $h_{n-1}(u)$ and $\hat{\phi}_{(i)} = Q_{(i)}(\hat{\boldsymbol{\beta}}_{(i)})/u_0^*$ denotes the maximum likelihood estimator of $\phi$ by dropping the $i$th observation. Furthermore, we may note that $r_i(k\boldsymbol{\epsilon}) = r_i(\boldsymbol{\epsilon})$ and $t_i(k\boldsymbol{\epsilon}) = t_i(\boldsymbol{\epsilon})$, $\forall k > 0$. Thus, by assuming $\boldsymbol{\epsilon} \sim El_n(\mathbf{0}, \phi\mathbf{I})$, it follows from the property given in Section 2 that

$$t_i \sim t_{(n-p-1)}$$

and

$$b_i = \frac{r_i^2}{(n-p)} \sim Beta(1/2, (n-p-1)/2),$$

$i = 1, \ldots, n$, where $t_{(n-p-1)}$ denotes the Student-t distribution with $(n - p - 1)$ degrees of freedom and $t_i$ and $b_i$ are independent of $g(\cdot)$. These results show the invariance of $t_i$ and $b_i$ in the elliptical class.

Consider now the mean-shift perturbation in the elliptical model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{d}_i\tau + \boldsymbol{\epsilon},$$

where $\mathbf{d}_i$ denotes an $(n \times 1)$ vector of zeros with one at the $i$th position. Then, using (2.5), the $F$ statistic to assess if the $i$th observation is an outlier, that corresponds in testing $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$, is given by

$$F_{(i)} = t_i^2 \sim F_{1,(n-p-1)} \quad \text{under} \quad H_0.$$

The demonstration of this result is similar to the one for the normal linear case (see, for instance, Chatterjee and Hadi, 1988).

### 3.2 Cook distance

In order to assess the influence of the $i$th observation on $\hat{\boldsymbol{\beta}}$, Cook (1977) proposed a distance between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(i)}$ based on the confidence region for $\boldsymbol{\beta}$ in the normal linear case, where $\hat{\boldsymbol{\beta}}_{(i)}$ denotes the maximum likelihood estimator of $\boldsymbol{\beta}$ by dropping the $i$th observation. This distance takes here the form

$$\begin{aligned} D_i &= (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^T (\mathbf{X}^T\mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})/ps^2 \\ &= \left(\frac{p_{ii}}{1-p_{ii}}\right)\frac{r_i^2}{p}, \end{aligned}$$

$i = 1, \ldots, n$, and since it depends only on the invariant quantities $p_{ii}$ and $r_i$ it is also invariant in the elliptical class.

### 3.3 Scale ratio

Similarly to the normal linear case (see, for instance, Belsley et al., 1980) we can assess the influence of the $i$th observation on the scale matrix $\mathbf{D}(\hat{\boldsymbol{\beta}}) = \hat{\phi}(\mathbf{X}^T\mathbf{X})^{-1}$, by using the influence measure

$$SCR_i = \frac{\det\{\hat{\phi}_{(i)}(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)})^{-1}\}}{\det\{\hat{\phi}(\mathbf{X}^T\mathbf{X})^{-1}\}},$$

$i = 1, \ldots, n$, where $\mathbf{X}_{(i)}$ denotes the matrix $\mathbf{X}$ without the $i$th row. Since $\det(\mathbf{X}_{(i)}^T\mathbf{X}_{(i)}) = (1-p_{ii})\det(\mathbf{X}^T\mathbf{X})$ and

$$\frac{\hat{\phi}_{(i)}}{\hat{\phi}} = \left(\frac{u_0}{u_0^*}\right)\left\{1 - \frac{r_i^2}{(n-p)}\right\} = \left(\frac{u_0}{u_0^*}\right)(1-b_i),$$

we obtain

$$SCR_i = \left(\frac{u_0}{u_0^*}\right)^p \frac{(1-b_i)^p}{(1-p_{ii})},$$

$i = 1, \ldots, n$, that depends on the quantities $u_0$ and $u_0^*$ which are not invariant. Therefore $SCR_i$ is also not invariant in the elliptical class.

### 3.4 Andrews-Pregibon measure

Andrews and Pregibon (1978) proposed a particular influence measure to detect remote observations in the subspace explained by the response and explanatory variable vectors. This measure takes, in the elliptical class, the form

$$\begin{aligned} AP_{(i)} &= \frac{Q_{(i)}(\hat{\boldsymbol{\beta}}_{(i)})\det\{\mathbf{X}_{(i)}^T\mathbf{X}_{(i)}\}}{Q(\hat{\boldsymbol{\beta}})\det(\mathbf{X}^T\mathbf{X})} \\ &= (1-p_{ii})(1-b_i), \end{aligned}$$

$i = 1, \ldots, n$, and since it depends only on the invariant quantities $p_{ii}$ and $b_{ii}$ it is also invariant in the elliptical class.

# 4 Likelihood displacement

Let $L(\boldsymbol{\theta})$ denote the log-likelihood function for the elliptical linear model, where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$. The likelihood displacement (see, for instance, Cook and Weisberg, 1982 and Cook et al., 1988) is defined by

$$LD_i(\boldsymbol{\theta}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{(i)})\},$$

where $\hat{\boldsymbol{\theta}}_{(i)}$ denotes the maximum likelihood estimator of $\boldsymbol{\theta}$ by dropping the $i$th observation.

For the elliptical model we find

$$L(\boldsymbol{\theta}) = -\frac{n}{2}\log(\phi) + \log\left[g\left\{\frac{Q(\boldsymbol{\beta})}{\phi}\right\}\right], \tag{4.1}$$

which evaluated at $\hat{\boldsymbol{\theta}}$ leads to

$$L(\hat{\boldsymbol{\theta}}) = -\frac{n}{2}\log(\hat{\phi}) + \log\{g(u_0)\}.$$

Evaluating (4.1) at $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\boldsymbol{\beta}}_{(i)}^T, \hat{\phi}_{(i)})^T$ we obtain

$$L(\hat{\boldsymbol{\theta}}_{(i)}) = -\frac{n}{2}\log(\hat{\phi}_{(i)}) + \log\left[g\left\{\frac{Q(\hat{\boldsymbol{\beta}}_{(i)})}{\hat{\phi}_{(i)}}\right\}\right].$$

Since

$$\frac{Q(\hat{\boldsymbol{\beta}}_{(i)})}{\hat{\phi}_{(i)}} = u_0^*\left\{1 + \frac{b_i}{(1-p_{ii})(1-b_i)}\right\},$$

we get

$$L(\hat{\boldsymbol{\theta}}_{(i)}) = -\frac{n}{2}\log(\hat{\phi}_{(i)}) + \log\left[g\left\{u_0^*\left(1 + \frac{b_i}{(1-p_{ii})(1-b_i)}\right)\right\}\right],$$

and consequently the distance $LD_i(\theta)$ may be expressed in the form

$$\begin{aligned} LD_i(\boldsymbol{\theta}) &= n\log\left\{\left(\frac{u_0}{u_0^*}\right)(1-b_i)\right\} \\ &\quad +2\log\left[g(u_0)/g\left\{u_0^*\left(1+\frac{b_i}{(1-p_{ii})(1-b_i)}\right)\right\}\right], \end{aligned}$$

$i = 1, \ldots, n$. Note that $LD_i(\boldsymbol{\theta})$ is not invariant in the elliptical class.

In particular for the Student-t distribution we have $g(u) = c_2(1 + u/\nu)^{-(n+\nu)/2}$, $u_0 = n$ and $u_0^* = n-1$. Then,

$$\begin{aligned} LD_i(\boldsymbol{\theta}) &= n\log\left\{\left(\frac{n}{n-1}\right)(1-b_i)\right\} \\ &\quad +(n+\nu)\log\left[\left\{\nu+\frac{(n-1)(p_{ii}b_i+1-p_{ii})}{(1-p_{ii})(1-b_i)}\right\}/(n+\nu)\right]. \end{aligned} \quad (4.2)$$

When $\nu \to \infty$ expression (4.2) reduces to

$$LD_i(\boldsymbol{\theta}) = n\log\left\{\left(\frac{n}{n-1}\right)(1-b_i)\right\} + \left(\frac{n-1}{1-p_{ii}}\right)\left(\frac{b_i}{1-b_i}\right) - 1,$$

which corresponds to the normal linear case, as expected (see, for instance, Cook et al., 1988).

It may be also shown that $LD_i(\boldsymbol{\theta})$ takes, for the exponential power model, the form

$$\begin{aligned} LD_i(\boldsymbol{\theta}) &= n\log\left\{\left(\frac{n}{n-1}\right)^{1/\alpha}(1-b_i)\right\} \\ &\quad +\frac{1}{\alpha}\left\{(n-1)\left(1+\frac{b_i}{(1-p_{ii})(1-b_i)}\right)^{\alpha} - n\right\}. \end{aligned}$$

## 4.1 Parameter subsets

Suppose now we have interest on the parameter vector $\boldsymbol{\beta}$ with $\phi$ being considered as a nuisance parameter. The likelihood displacement is defined in this case as (see, for instance, Cook et al., 1988)

$$LD_i(\boldsymbol{\beta} \mid \phi) = 2\{L(\hat{\boldsymbol{\theta}}) - \max_{\phi} L(\hat{\boldsymbol{\beta}}_{(i)}, \phi)\}. \quad (4.3)$$

We may show that the value of $\phi$ which maximizes $L(\hat{\boldsymbol{\beta}}_{(i)}, \phi)$ is $\tilde{\phi} = Q(\hat{\boldsymbol{\beta}}_{(i)})/u_0$. Then, expression (4.3) reduces to

$$LD_i(\boldsymbol{\beta} \mid \phi) = n\log\left(1+\frac{pD_i}{(n-p)}\right), \quad (4.4)$$

$i = 1, \ldots, n$, that agrees with expression (4) given in Cook et al. (1988). It is interesting to note that (4.4) is invariant in the elliptical class.

Similarly, we may show that

$$LD_i(\phi \mid \boldsymbol{\beta}) = n\log(\hat{\phi}_{(i)}/\hat{\phi}) + 2\log[g(u_0)/g\{u_0\hat{\phi}/\hat{\phi}_{(i)}\}], \qquad (4.5)$$

$i = 1, \ldots, n$, which depends on the elliptical distribution. In the normal case $g(u) = c_1\exp(-u/2)$, $u_0 = n$ and $u_0^* = n - 1$ so that (4.5) reduces to

$$LD_i(\phi \mid \boldsymbol{\beta}) = n\log\left\{\left(\frac{n}{n-1}\right)(1 - b_i)\right\} + \left(\frac{nb_i - 1}{1 - b_i}\right), \qquad (4.6)$$

$i = 1, \ldots, n$. As expected, expression (4.6) agrees with expression (11) given in Cook et al. (1988) for the normal linear model.

# 5 Local influence

Let $L(\boldsymbol{\theta})$ denote the log-likelihood function from the postulated model (here $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$), and let $\boldsymbol{\omega}$ be a $(q \times 1)$ vector of perturbations restricted to some open subset $\boldsymbol{\Omega} \in I\!R^q$. The perturbations are made on the likelihood function, such that it takes the form $L(\boldsymbol{\theta}|\boldsymbol{\omega})$. Denoting the vector of no perturbation by $\boldsymbol{\omega}_0$, we assume $L(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$. To assess the influence of the perturbations on the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, one may consider the likelihood displacement

$$LD(\boldsymbol{\omega}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\},$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the maximum likelihood estimate under the model $L(\boldsymbol{\theta}|\boldsymbol{\omega})$.

In some situations, though, it may be of interest to assess the influence on a subset $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$. For example, one may have interest on $\boldsymbol{\theta}_1^T = (\beta_1, \ldots, \beta_p)^T$ or $\boldsymbol{\theta}_1 = \phi$. In these cases, the likelihood displacement is defined as

$$LD_1(\boldsymbol{\omega}) = 2[L(\hat{\boldsymbol{\theta}}) - L\{\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}, \hat{\boldsymbol{\theta}}_2(\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}})\}],$$

where $\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}$ is obtained from $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}} = (\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}^T, \hat{\boldsymbol{\theta}}_{2\boldsymbol{\omega}}^T)^T$ and $\hat{\boldsymbol{\theta}}_2(\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}})$ is the maximum likelihood estimate of $\boldsymbol{\theta}_2$ for $\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}$ fixed in the perturbed model.

The idea of local influence (Cook, 1986) is concerned in characterizing the behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. The procedure consists in selecting a unit direction $\boldsymbol{\ell}$, $\parallel \boldsymbol{\ell} \parallel = 1$, and then to consider the plot of $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ against $a$, where $a \in I\!R$. This plot is called *lifted line*. Note that, since $LD(\boldsymbol{\omega}_0) = 0$, $LD(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ has a local minimum at $a = 0$. Each lifted line can be characterized by considering the normal curvature $C_{\boldsymbol{\ell}}(\boldsymbol{\theta})$ around $a = 0$. This curvature is interpreted as the inverse radius of the best fitting circle at $a = 0$. The suggestion is to consider the direction $\boldsymbol{\ell}_{max}$ corresponding to the largest curvature $C_{\boldsymbol{\ell}_{max}}(\boldsymbol{\theta})$. The index plot of $\boldsymbol{\ell}_{max}$ may reveal those observations that under small perturbations exercise notable influence on $LD(\boldsymbol{\omega})$.

Cook(1986) showed that the normal curvature at the direction $\boldsymbol{\ell}$ takes the form

$$C_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta}\boldsymbol{\ell}|, \tag{5.1}$$

where $-\ddot{\mathbf{L}}$ is the observed Fisher information matrix for the postulated model ($\boldsymbol{\omega} = \boldsymbol{\omega}_0$) and $\boldsymbol{\Delta}$ is the $(p+1) \times q$ matrix with elements

$$\Delta_{ij} = \frac{\partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_i \partial\omega_j},$$

evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $i = 1, \ldots, p+1$ and $j = 1, \ldots, q$.

Therefore, the maximization of (5.1) is equivalent to finding the largest eigenvalue $C_{\boldsymbol{\ell}_{max}}$ of the matrix $\mathbf{B} = \boldsymbol{\Delta}^T(\ddot{\mathbf{L}})^{-1}\boldsymbol{\Delta}$, and $\boldsymbol{\ell}_{max}$ is the corresponding eigenvector.

For the subset $\boldsymbol{\theta}_1$, the curvature at the direction $\boldsymbol{\ell}$ is given by

$$C_{\boldsymbol{\ell}}(\boldsymbol{\theta}_1) = 2|\boldsymbol{\ell}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}\boldsymbol{\ell}|,$$

where $\mathbf{B}_{22}$ is defined as

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix},$$

and $\ddot{\mathbf{L}}_{22}$ is obtained from the partition of $\ddot{\mathbf{L}}$ according to the partition of $\boldsymbol{\theta}$. The eigenvector $\boldsymbol{\ell}_{max}$ corresponds to the largest eigenvalue of the matrix $\mathbf{B} = \boldsymbol{\Delta}^T(\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22})\boldsymbol{\Delta}$.

Recently, Fung and Kwan (1997) presented an interesting discussion on the application of the local influence method for various influence measures. They showed that an influence measure, namely $\hat{T}_{\boldsymbol{\omega}}$, is scale invariant if $\dot{\boldsymbol{\Gamma}} = \partial\hat{T}_{\boldsymbol{\omega}}/\partial\boldsymbol{\omega}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}$. When this derivative is non-zero the ordering among the components of $\boldsymbol{\ell}_{max}$ is not necessarily preserved under changes in the scale. In particular, for the likelihood displacement, we have $\dot{\boldsymbol{\Gamma}} = \partial L(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})/\partial\boldsymbol{\omega}|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}$. This property also follows, for instance, for the influence measures proposed by Thomas and Cook (1990) and Paula (1993). But it does not hold for other influence measures as pointed out by Fung and Kwan (1997).

## 5.1 Curvature derivation

We assume the perturbation scheme $\boldsymbol{\epsilon} \sim El_n(\mathbf{0}, \phi\mathbf{D}^{-1}(\boldsymbol{\omega}))$, where $\mathbf{D}(\boldsymbol{\omega}) = \mathrm{diag}\{\omega_1, \ldots, \omega_n\}$ with $\omega_i$ denoting the weight corresponding to the $i$th case and $\boldsymbol{\omega}_0 = \mathbf{1}$. The normal curvature in the direction $\boldsymbol{\ell}$ for the vector $\boldsymbol{\theta}$ is given by (see Galea et al., 1997)

$$C_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^T[\mathbf{B}_1 + \mathbf{B}_2]\boldsymbol{\ell}|,$$

where

$$\mathbf{B}_1 = (2/\hat{\phi})W_g(u_0)\mathbf{D}(\mathbf{e})\mathbf{P}\mathbf{D}(\mathbf{e})$$

and

$$\mathbf{B}_2 = \frac{1}{\hat{\phi}^2}\frac{[W_g(u_0) + u_0W_g'(u_0)]^2}{[\frac{n}{2} + u_0\{2W_g(u_0) + u_0W_g'(u_0)\}]}\mathbf{D}(\mathbf{e})\mathbf{e}\mathbf{e}^T\mathbf{D}(\mathbf{e}).$$

In particular, if we are interested in the vector $\boldsymbol{\beta}$, the normal curvature in the direction $\boldsymbol{\ell}$ yields

$$C_{\boldsymbol{\ell}}(\boldsymbol{\beta}) = \frac{4}{\hat{\phi}}|W_g(u_0)||\boldsymbol{\ell}^T\mathbf{D}(\mathbf{e})\mathbf{P}\mathbf{D}(\mathbf{e})\boldsymbol{\ell}|,$$

where $\mathbf{D}(\mathbf{e}) = \mathrm{diag}\{e_1, \ldots, e_n\}$. Then, the index plot of $\boldsymbol{\ell}_{max}$ obtained from the matrix $\mathbf{D}(\mathbf{e})\mathbf{P}\mathbf{D}(\mathbf{e})$ may show how to perturb $\mathbf{D}(\boldsymbol{\omega})$ to obtain larger changes in the regression coefficients.

Similarly, the normal curvature for $\phi$ in the direction $\boldsymbol{\ell}$ takes the form

$$C_{\boldsymbol{\ell}}(\phi) = \frac{2}{\hat{\phi}^2}|C_\omega||\boldsymbol{\ell}^T\mathbf{D}(\mathbf{e})\mathbf{e}\mathbf{e}^T\mathbf{D}(\mathbf{e})\boldsymbol{\ell}|,$$

where

$$C_\omega = [W_g(u_0) + u_0W_g'(u_0)]^2/[\frac{n}{2} + u_0\{2W_g(u_0) + u_0W_g'(u_0)\}].$$

In this case, for the largest curvature,

$$\boldsymbol{\ell}_{max} \propto \mathbf{D}(\mathbf{e})\mathbf{e},$$

which means that observations with large values for $e_i^2$ are most influential on $\hat{\phi}$.

Therefore, since $\mathbf{e}$ is invariant in the elliptical class the vector $\boldsymbol{\ell}_{max}$ is invariant when we are interested in the vector $\boldsymbol{\beta}$ or in the scale parameter $\phi$. However, if we are interested in both, $\boldsymbol{\beta}$ and $\phi$, the vector $\boldsymbol{\ell}_{max}$ depends on the elliptical distribution under consideration.

# 6 Application

As illustration consider the data set reported by Ruppert and Carroll (1980) on the salinity of water during the spring in Pamlico Sound, North Carolina. The response $Y$ was the biweekly salinity, and the explanatory variables were salinity lagged 2 weeks, $x_1$, a dummy variable, $x_2$, for the time period and the river discharge, $x_3$. Twenty-eight observations were considered. The value of $x_{1i}$ may differ from $y_{i-1}$, since the data are not a contiguous sequence. Several authors have analyzed this data set, particularly under the diagnostic viewpoint. The linear model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i,$$

where $\epsilon_i$ follows an appropriate symmetric distribution, has been adopted.

Atkinson (1985), for instance, assumed a normal distribution for $\epsilon_i$ and using deletion diagnostic methods found cases 16 and 5 as the most influential on the parameter estimates. Davison and Tsai (1992) assumed a univariate Student-t distribution with 3 degrees of freedom for $\epsilon_i$ and in their deletion diagnostic analysis cases 16, 5 and 3 appeared as the most influential. Galea et al. (1997) assumed that $\boldsymbol{\epsilon} \sim El_n(\mathbf{0}, \phi\mathbf{I})$ and applied the local influence method for assessing the influence of the observations on $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$ under some multivariate elliptical distributions. They found case 16 as the most influential on $\hat{\boldsymbol{\beta}}$ and cases 9, 15, 16 and 17 most influential on $\hat{\phi}$.

Our analysis will be restricted on the normal, Student-t and exponential power distributions, which are the most well known in the elliptical class. Figures 1 and 2 present the index plot of $LD_i(\boldsymbol{\theta})$ under the exponential power and Student-t distributions for $\alpha = 0.1, 0.5$ and 1.2 and $\nu = 3, 30$ and 100 degrees of freedom, respectively. We can notice from Figure 1 that case 16 appears with more accentuated influence for $\alpha = 1.2$ rather than for $\alpha = 0.5$ and 0.1. It may be due to the fact that the exponential power distribution has lighter tails than the normal distribution as $\alpha$ becomes greater than one. Similar tendency is observed in Figure 2. The influence of observation 16 becomes less accentuated for small degrees of freedom. Figures 3 and 4 present the index plot of $LD_i(\phi \mid \boldsymbol{\beta})$. In these figures we can notice similar tendencies to the ones observed in Figures 1 and 2, respectively. The index plot of $LD_i(\boldsymbol{\beta} \mid \phi)$, that is invariant under the elliptical distributions and is omitted here, points out observation 16 as the most influential.

Figures 5a and 5b present the index plot of $\mid \boldsymbol{\ell}_{max} \mid$ for $\boldsymbol{\theta}$ under the exponential power distribution with $\alpha = 0.1$ and 1.2. Note that case 16 appears most influential in both graphics. In Figures 5c and 5d one has the index plot of $\mid \boldsymbol{\ell}_{max} \mid$ for $\boldsymbol{\theta}$ under the Student-t distribution with $\nu = 3$ degrees of freedom and under the normal distribution. Similarly to the exponential power distribution case 16 is influential in both situations.

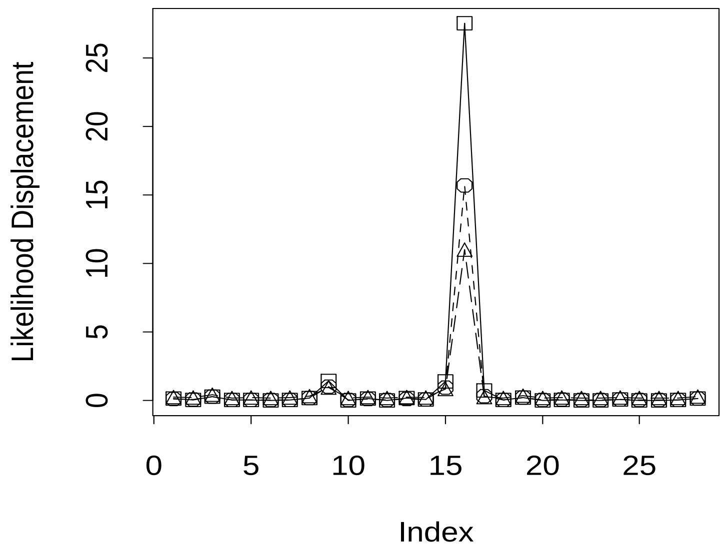


**Figure 1**
*Index plot of $LD_i(\boldsymbol{\theta})$ for the exponential power distribution with $\alpha = 0.1$ $(-\triangle-)$, $\alpha = 0.5$ $(- \circ -)$ and $\alpha = 1.2$ $(-\square-)$.*

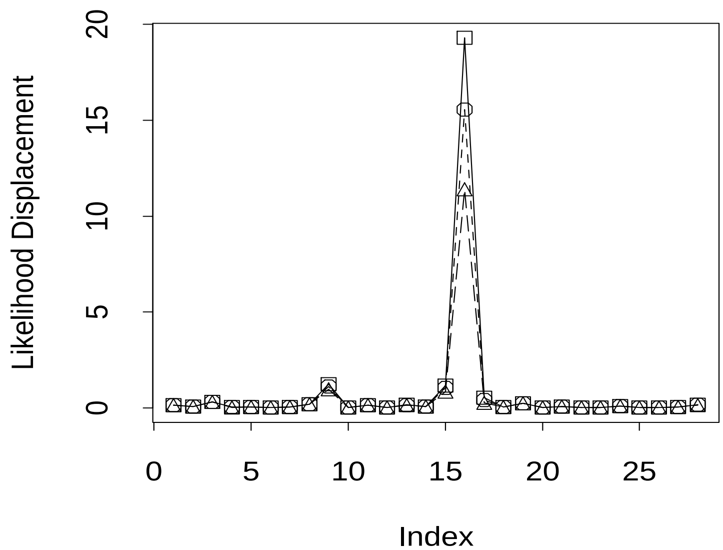


**Figure 2**
*Index plot of $LD_i(\boldsymbol{\theta})$ for the Student-t distribution with $\nu = 3$ $(-\triangle-)$, $\nu = 30$ $(- \circ -)$ and $\nu = 100$ $(-\square-)$ degrees of freedom.*

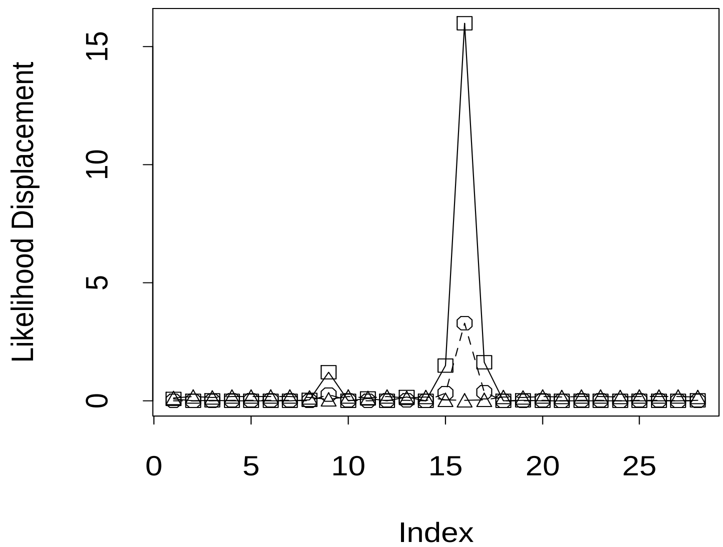


**Figure 3**
*Index plot of* $LD_i(\phi \mid \boldsymbol{\beta})$ *for the exponential power distribution with* $\alpha = 0.1$ $(-\triangle-)$, $\alpha = 1.0$ $(-\circ-)$ *and* $\alpha = 3.0$ $(-\square-)$.

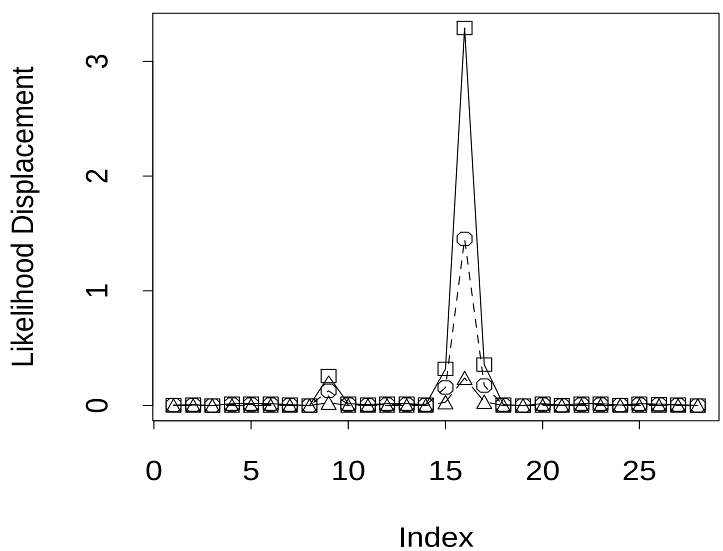


**Figure 4**
*Index plot of* $LD_i(\phi \mid \boldsymbol{\beta})$ *for the Student-t distribution with* $\nu = 3$ $(-\triangle-)$, $\nu = 30$ $(-\circ-)$ *and* $\nu = 100$ $(-\square-)$ *degrees of freedom.*

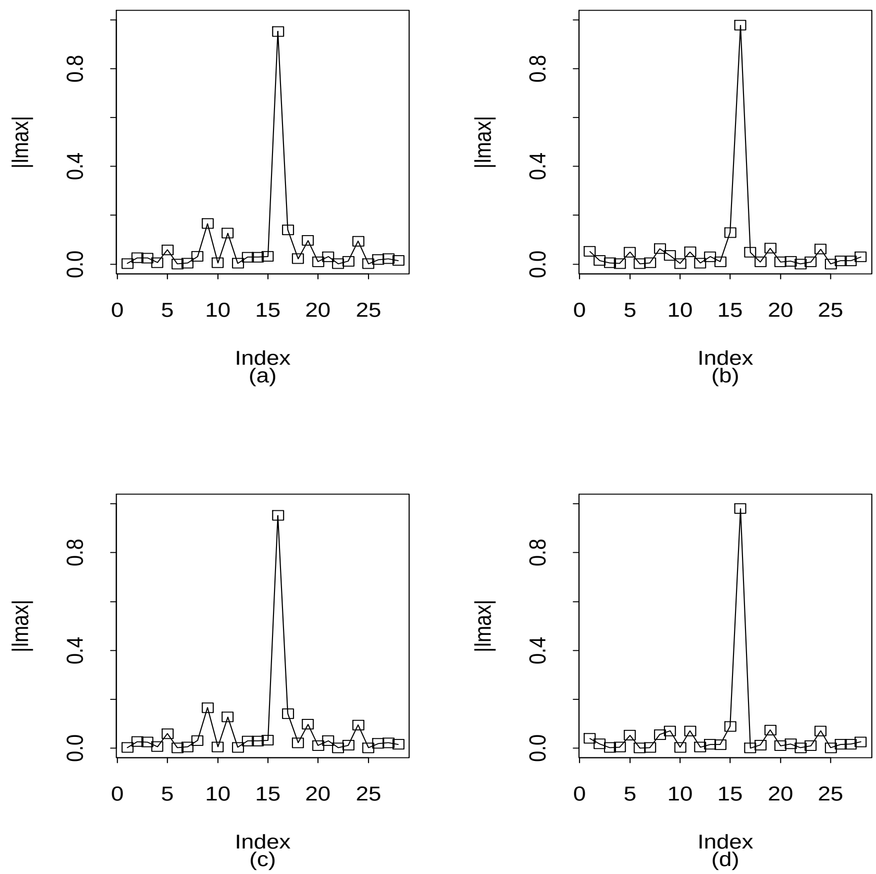


**Figure 5**
*Index plot of* $|\ell_{max}|$ *for* $\boldsymbol{\theta}$ *under the exponential power distribution with* $\alpha = 0.1$ *(a) and* $\alpha = 1.2$ *(b), Student-t distribution with* $\nu = 3$ *degrees of freedom (c) and normal distribution (d).*

# 7 Conclusions

Even though robustness is a more generic subject that involves aspects, which are not directly treated in this work, it is interesting to note from the example discussed in the last section the evident superiority of the deletion method over the local influence method in the sense of selecting the less sensitive elliptical model to the influential observations. For example, when the no invariant measure $LD_i(\boldsymbol{\theta})$ is used, the Student-t model with 3 degrees of freedom is less sensitive than the Student-t models with higher degrees of freedom to the influential observation 16. Similar tendency occurs with the exponential power model. In this case, the models with smaller values for the parameter $\alpha$ are less sensitive to the influential observations than the models with larger values of $\alpha$. Nevertheless, it is still open to discussion how to select between the "best" Student-t model and the "best" exponential power model.

Another important point that remains to be examined is on the definition of an appropriate residual capable of distinguishing among the elliptical distributions. As it was shown in Section 3.1 the residuals $t_i$ and $b_i$ are invariant in the elliptical class, so that they should not be used to select, among the elliptical distributions, one more appropriate to explain the responses. We think that defining residuals by using the log-likelihood function may be an interesting topic for future investigation. A possible suggestion would be the study of the deviance residual distribution (see, for instance, Davison and Gigli, 1989) in the elliptical class.

Further work on influence diagnostic in univariate elliptical linear models which assume, in contrast with the class considered in this work, independent distributions for the responses is being investigated by the authors and will be the subject of an incoming paper. However, such models do not enjoy the invariance property which is shared in the multivariate case. For more details on univariate elliptical models see, for instance, Fang and Anderson (1990).


# Acknowledgements

This work was supported by CNPq and FAPESP-Brazil, FONDECYT-1980451-Chile and Universidad del Valparaso, project DIUV 19/97. We wish to thank two referees and Dr. Heleno Bolfarine whose comments were valuable to improve the presentation of the paper.

*(Received December, 1999. Revised June, 2000.)*

## References


Andrews, D. F. and Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society B*, **40**, 80-93.

Arellano, R. (1994). *Elliptical Distributions: Properties, Inference and Applications in Regression Models*(in Portuguese). Unpublished Ph. D. Thesis, Department of Statistics, University of Sao Paulo, Brazil.

Atkinson, A. C. (1985). *Plots, Transformations and Regression*. Clarendon Press: Oxford.

Beckman, R. J., Nachtsheim, C. J. and Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413-426.

Belsley, D., Kuh, E. and Welsch, R. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. New York: Wiley.

Cook, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics*, **19**, 15-18.

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B*, **48**, 133-169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. London: Chapman and Hall.

Cook, R. D., Peña, D. and Weisberg S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics, Theory and Methods*, **17**, 623-640.

Davison, A. C. and Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika*, **76**, 211-221.

Davison, A.C. and Tsai, C-L. (1992). Regression model diagnostics. *International Statistical Review*, **60**, 337-353.

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

Fang, K. T. and Zhang, Y. T. (1990). *Generalized Multivariate Analysis*. New York: Springer-Verlag.

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

Fung, W. K. and Kwan, C. W. (1997). A note on local influence based on normal curvature.*Journal of the Royal Statistical Society B*, **59**, 839-843.

Galea, M.; Paula, G. A. and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, **46**, 71-79.

Gupta, A. K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics*. Kluwer Academic Publishers.

Kim, M. G. (1995). Local influence in multivariate regression. *Communications in Statistics, Theory and Methods*, **24**, 1271-1278.

Lawrance, A. J. (1988). Regression transformation diagnostics using local influence. *Journal of the American Statistical Association*, **84**, 125-141.

Leiva, V. (1998). *Inference on the variation coefficient in elliptical distributions*. Unpublished Ph.D. Thesis, University of Granada, Spain.

Paula, G. A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis*, **16**, 63-79.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, **75**, 828-838.

Thomas, W. and Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics*, **32**, 59-65.

Tsai, C. H. and Wu, X. (1992). Assessing local,influence in linear regression models with first-order autoregressive or heteroscedastic error structure. *Statistics and Probability Letters*, **14**, 247-252.