UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

An Application of the Local Influence Diagnostics to Ridge Regression Under Elliptical Model

Filidor V. Labra
Reiko Aoki
Francisco Rojas

Nº 78

São Carlos - SP

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2577

An Application of the Local Influence Diagnostics to Ridge Regression Under Elliptical Model

Filidor V. Labra
Reiko Aoki
Francisco Rojas

Nº 78

NOTAS

Série Estatística

São Carlos – SP
Jul./2005

AN APPLICATION OF THE LOCAL INFLUENCE DIAGNOSTICS TO RIDGE REGRESSION UNDER ELLIPTICAL MODEL

Filidor V. Labra[1], Reiko Aoki[2] and Francisco Rojas[3]

(1) Departamento de Estatística, IMECC/UNICAMP, Caixa Postal 6065

13083-970 Campinas, SP, Brazil

fily@ime.unicamp.br

(2) Departamento de Ciências de Computação e Estatística,

ICMC/USP-São Carlos, Caixa Postal 668,

13560-970, São Carlos, SP, Brazil

reiko@icmc.usp.br

(3) Departamento de Estatística, Ufscar, Caixa Postal 676,

13565-905, São Carlos, SP, Brazil

Key Words: ridge regression; ill conditioned matrix; multicollinearity; local influence; elliptical distributions.

RESUMO

Neste trabalho, vamos estudar o efeito de pequenas perturbações nos estimadores "ridge" considerando a classe de distribuições elípticas para os erros. As matrizes necessárias para a análise de influência local considerando as perturbações nas variáveis explanatórias e na matriz escala são obtidas. Os dados de Longley são analizadas como uma ilustração.

ABSTRACT

In this paper, we study the effect of a minor perturbation on the ridge estimator considering the class of elliptical distribution for the errors. The necessary matrices for assessing the local influence under the perturbation of the explanatory variables and the scale matrix is derived. The Longley data is analyzed for illustration.

1. INTRODUCTION

Consider the usual multiple linear regression model,

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

where $\boldsymbol{Y}$ is an n vector of observable responses, $\boldsymbol{X}$ is an $n \times p$, centered and standardized, matrix of known constants of rank p, $\boldsymbol{\beta}$ is an $p \times 1$ vector of unknown parameters and $\boldsymbol{\epsilon}$ is an n vector of random errors, with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \sigma^2\mathbf{I_n}$. It is well known that in the presence of multicollinearity the ordinary least-squares (OLS) estimators in the linear regression model is not appropriate as the average distance from $\hat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}$ will tend to be large. Also, the OLS estimators is very sensitive to the presence of a few extreme observations in the response and explanatory variables, which can seriously affect the inferences. It is not unusual to have multicollinearity and influential observations simultaneously in a data set (Lawrence and Marsh, 1984). To overcome the difficulty of multicollinearity, Hoerl and Kennard (1970a, b) proposed the use of $(\mathbf{X}^\top\mathbf{X} + k\mathbf{I}_p)$, $k \geq 0$, rather than $\mathbf{X}^\top\mathbf{X}$ in the estimation of $\boldsymbol{\beta}$, which is known in the literature as ridge regression estimator. In the presence of collinearity, the ridge regression estimator shrinks the OLS estimator toward the origin and yields a biased estimator with smaller mean square error. Since then, a large number of papers have been written on the subject. Walker and Birch (1988) showed that when ridge regression is considered, the influence of each case changes as a function of the shrinkage parameter k, which means that once the value of k is determined, influence measures should be computed for that specific value of k. More recently Billor and Loynes (1999) considered the use of the local influence diagnostics proposed by Cook (1986) in ridge regression. They considered the normal distribution for the errors. However, the multicollinearity may occur in many applications of the multiple linear regression models with the errors having other distributions than the normal distribution. Several authors have considered the Student_t distribution as an alternative to the normal distribution as it can naturally accommodate outliers present in the data. Lange et. al. (1989) discussed the use of the Student_t distribution in regression models, as well as in problems related to multivariate analysis.

In this paper, we extend the results obtained in Billor and Loyness (1993, 1999), considering the class of elliptical distributions for the errors and also the perturbation in the explanatory variables, as the matrix of the predictor variables are supposed to be ill conditioned. The class of elliptical distributions has as its particular cases, some well known distributions, as normal, Student_t, contaminated normal and logistic, among other distributions (see Fang and Anderson (1990) and Fang and Zhang (1990), for example). In Section 2, the ridge regression model under elliptical distribution is described. In Section 3, the concept of the local influence approach is reviewed. In Section 4, some perturbation schemes for the analysis of the local influence is considered and the necessary matrices for the application of the results are given in closed form expressions. Finally, in Section 5 we apply and discuss the results obtained in earlier sections to the Longley data.

2. RIDGE REGRESSION UNDER ELLIPTICAL MODEL

In this section we consider the ridge regression under the class of the elliptical distributions. If the density of a n-dimensional random vector $\boldsymbol{Y}$ with the n-dimensional location vector $\boldsymbol{\mu}$ and the $n \times n$ scale matrix $\boldsymbol{\Lambda}$ is given by

$$f_{\mathbf{Y}}(\boldsymbol{y}) = |\boldsymbol{\Lambda}|^{-1/2} f((\boldsymbol{y}-\boldsymbol{\mu})^{\top}\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})), \quad \boldsymbol{y} \in \mathbb{R}^n,$$

for some non negative function $f(u)$, with $u \geq 0$ satisfying

$$\int_0^{\infty} u^{n-1} f(u^2) du < \infty,$$

then $\boldsymbol{Y}$ is distributed according to a elliptical distribution. The function f is known as the density generating function. We shall denote $\boldsymbol{Y} \sim El(\boldsymbol{\mu}, \boldsymbol{\Lambda}; f)$ or $\boldsymbol{Y} \sim El(\boldsymbol{\mu}, \boldsymbol{\Lambda})$.

This class of distributions has as its particular cases the normal distributions, student_t distributions, contaminated normal distributions, logistic distributions and exponential power distributions among others, see for example Fang et al. (1990) and Galea et al. (2000). In the case of a normal distribution, $f(u) = ce^{-u/2}$, where c is a normalizing constant.

Next, we introduce the ridge regression under the class of the elliptical distributions. Consider the linear regression model given by

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

where $\boldsymbol{Y}$ is a centered n-dimensional random vector, $\boldsymbol{X}$ is a fixed $n \times p$ matrix of covariates, with full rank p, $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters and $\boldsymbol{\epsilon}$ is a n-dimensional vector of unobservable errors. The errors are supposed to be uncorrelated and the matrix $\boldsymbol{X}$ is considered to be centered and scaled. Under the normal model the ridge estimator, $\widehat{\boldsymbol{\beta}}_*$, is obtained as a solution to the $(\boldsymbol{X}^\top\boldsymbol{X} + k\boldsymbol{I}_p)\widehat{\boldsymbol{\beta}}_* = \boldsymbol{X}^\top\boldsymbol{Y}$, which is given by

$$\widehat{\boldsymbol{\beta}}_* = (\boldsymbol{X}^\top\boldsymbol{X} + k\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{Y}, \quad k > 0.$$

As the method of local influence is based on the likelihood function, to study the effect of a minor perturbation in the data or in the model, it is necessary to obtain the maximum likelihood estimator of the parameter vector $\boldsymbol{\beta}$. In the normal model, Billor and Loynes (1999) considered the $(n+p) \times p$ matriz

$$\boldsymbol{X}_a = \begin{pmatrix} \boldsymbol{X} \\ k^{1/2}\boldsymbol{I}_p \end{pmatrix},$$

and the $(n+p)$-dimensional vector of responses $\boldsymbol{Y}_a = (\boldsymbol{Y}^\top, \boldsymbol{0}^\top)^\top$ satisfying the following relationship

$$\boldsymbol{Y}_a = \boldsymbol{X}_a\boldsymbol{\beta} + \boldsymbol{\epsilon}_a. \tag{2.1}$$

The elliptical model can be defined assuming that $\boldsymbol{\epsilon}_a \sim El_{n+p}(\boldsymbol{0}, \phi\boldsymbol{I}_{n+p}; f)$ and from the properties of the elliptical distributions $\boldsymbol{Y}_a \sim El_{n+p}(\boldsymbol{X}_a\boldsymbol{\beta}, \phi\boldsymbol{I}_{n+p}; f)$. The density function of $\boldsymbol{Y}_a$, treated as though all the components are ordinary observations, will be called a pseudo-density function. Thus, the pseudo-likelihood can be written as

$$L_p(\boldsymbol{\theta}) = -\frac{n+p}{2}\log\phi + \log f(u_a), \tag{2.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ and $u_a = \frac{1}{\phi}Q_a(\boldsymbol{\beta})$, with

$$Q_a(\boldsymbol{\beta}) = (\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta})^\top(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^\top\boldsymbol{\beta}.$$

If f is a continuous and decreasing function, then the maximum likelihood estimators of $\boldsymbol{\beta}$ and ϕ are given by (see, Fang and Anderson, 1990)

$$\widehat{\boldsymbol{\beta}}_* = (\boldsymbol{X}_a^\top \boldsymbol{X}_a)^{-1} \boldsymbol{X}_a^\top \boldsymbol{Y}_a = (\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_p)^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \quad \text{and} \quad \widehat{\phi} = Q_a(\widehat{\boldsymbol{\beta}}_*)/u_f,$$

where u_f maximizes the function $h(u) = u^{\frac{n+p}{2}} f(u)$, $u \geq 0$. That is, u_f is the solution of the equation

$$f'(u) + \frac{n+p}{2u} f(u) = 0$$

or equivalently, solution of the equation

$$W_f(u) + \frac{n+p}{2u} = 0, \text{ where } W_f(u) = \frac{f'(u)}{f(u)}. \tag{2.3}$$

If we consider, for example, the normal distribution, we obtain $u_f = n + p$ and $W_f(\hat{u}_a) = 1/2$.

3. LOCAL INFLUENCE DIAGNOSTICS

The local influence diagnostics is a method to assess the local influence of minor perturbations of a statistical model or data.

Case deletion is a popular way to asses the individual impact of cases on the estimation process. This approach can be regarded as a global measure of influence. An alternative methodology for the identification of groups of cases which may require some concern is local influence which is based on differential geometry instead of complete deletion. It employs a differential comparison of parameter estimates before and after perturbation to data values or model assumptions. As considered in Cook (1986), the likelihood displacement is used as the metric to assess the local influence.

Let $L(\boldsymbol{\theta})$ denote the log-likelihood function given in (2.2), $\boldsymbol{\omega}$ the vector of perturbation introduced in the model, where $\boldsymbol{\omega} \in \boldsymbol{\Omega} \subseteq \mathbb{R}^h$, $\boldsymbol{\Omega}$ an open subset and $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ the log-likelihood function corresponding to the perturbed data or model. Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_\omega$ denote the maximum likelihood estimates under the model defined by $L(\boldsymbol{\theta})$ and $L(\boldsymbol{\theta}|\boldsymbol{\omega})$, respectively, and assume that there is an $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ representing no perturbation, such that $L(\boldsymbol{\theta})=L(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$ for all $\boldsymbol{\theta}$. The influence of $\boldsymbol{\omega}$ can be assessed by the log-likelihood displacement

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

where $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\omega_0}$. Because evaluation of $LD(\boldsymbol{\omega})$ for all $\boldsymbol{\omega}$ is practically unfeasible, Cook (1986) proposed to study the local behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$, which can be performed by evaluating the normal curvature C_l of $LD(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\boldsymbol{l}$.

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

$$C_l = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \, \boldsymbol{I}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|,$$

where $\|\boldsymbol{l}\| = 1$, $\boldsymbol{I} = -\dfrac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$ is a $(p+1)\times(p+1)$ observed information matrix, and

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta}/\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}$$

are both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

Let $\boldsymbol{l}_{max}$ be the direction of the maximum normal curvature (C_{max}), which is the perturbation that produces the greatest local change in $\widehat{\boldsymbol{\theta}}$. The most influential elements of the data may be identified by looking the components of the vector $\boldsymbol{l}_{\max}$, which is relatively large. Furthermore, $\boldsymbol{l}_{\max}$ is just the eigenvector corresponding to the largest eigenvalue, (C_{max}), of $\boldsymbol{\Delta}^\top \, \boldsymbol{I}^{-1}\boldsymbol{\Delta}$. Other important direction is $\boldsymbol{l} = \boldsymbol{e}_j$, where $\boldsymbol{e}_j$ denotes a vector of zeros, with the element of the jth position assuming the value one. In that case, the normal curvature, called the total local influence of individual j, is given by $C_j = 2\boldsymbol{\Delta}_j^\top \boldsymbol{I}^{-1}\boldsymbol{\Delta}_j$, where $\boldsymbol{\Delta}_j$ is the jth column of $\boldsymbol{\Delta}$, $j = 1,\ldots,n$. From (2.1), it follows that $\boldsymbol{I}$ takes the form

$$\boldsymbol{I} = -\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top} \right],$$

where, $\boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\beta}, \phi$.

When a subset $\boldsymbol{\theta}_1$ from the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is of interest, influence diagnostics can be based on the matrix (Cook, 1986)

$$\boldsymbol{\Delta}^\top(\boldsymbol{I}^{-1} - \boldsymbol{B}_{22})\boldsymbol{\Delta},$$

with

$$\boldsymbol{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{22}^{-1} \end{pmatrix}$$

and $\boldsymbol{I}_{22}$ is determined by the partition of $\boldsymbol{I}$ accordingly with the partition of $\boldsymbol{\theta}$.

Billor and Loynes (1993) pointed out some practical and theoretical difficulties which arise from Cook's approach. For example, computability of the maximum curvature; lack of invariance of the curvature under reparametrization of the perturbation scheme; and lack of definition of the parameters. To overcome the difficulty of lack of definition they suggested the following measure proposed by Tsai (1986)

$$LD^*(\boldsymbol{\omega}) = -2[L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_\omega|\boldsymbol{w})],$$

where $L(\boldsymbol{\theta}|\boldsymbol{w})$ is the log-likelihood function of the perturbed model. They suggested the use of the first derivative of LD^* as it provides valuable informations about the local influence behavior of LD^*. In particular, the use of the direction which produces the maximum increment in LD^* is of interest and in this case the maximum slope is given by

$$d_{max} = \|\nabla LD^*(\boldsymbol{\omega}_0)\| = 2\|\nabla L(\widehat{\boldsymbol{\theta}}|\boldsymbol{w})\|.$$

If we take the perturbed model given by

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

where $Var(\boldsymbol{\epsilon}) = \phi\boldsymbol{W}$, with ϕ known and $\boldsymbol{W} = \text{diag}(1+w_1, \ldots, 1+w_n)$, with $\text{diag}(a_1, \ldots, a_n)$ denoting a square diagonal matrix with the elements of the diagonal given by $a_1, \ldots, a_n$, then Billor and Loynes (1999) obtained under ridge model that

$$\|\nabla LD^*(\boldsymbol{\omega}_0)\|^2 = \sum_{i=1}^{n}(1 - \frac{e_i^{*2}}{\phi})^2, \tag{3.1}$$

where $\Delta_f(u) = \frac{dW_f(u)}{du}$, $u \geq 0$. Note that, evaluating the derivatives above at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, we obtain that $(\boldsymbol{Y}_a - \boldsymbol{X}_a\widehat{\boldsymbol{\beta}}_*)^\top \boldsymbol{X}_a = 0$ and $Q_a(\widehat{\boldsymbol{\beta}}_*)/\widehat{\phi} = u_f$, so that

$$\boldsymbol{I}\mid_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\begin{pmatrix} \frac{2}{\widehat{\phi}} W_f(\widehat{u}_a)(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_p) & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n+p}{2\widehat{\phi}^2} + \frac{u_f}{\widehat{\phi}^2} A_f(\widehat{u}_a) \end{pmatrix}, \tag{3.2}$$

where $\widehat{u}_a = Q_a(\widehat{\boldsymbol{\beta}}_*)/\widehat{\phi} = u_f$ and $A_f(\widehat{u}_a) = 2W_f(\widehat{u}_a) + u_f\Delta_f(\widehat{u}_a)$.

Expressions for $W_f(u)$ and $\Delta_f(u)$ for some elliptical distributions, can be found in Galea et al. (2000). In particular under normal distribution, $u_f = n+p$, $W_f(\widehat{u}_a) = -1/2$ and $\Delta_f(\widehat{u}_a) = 0$, so that the matrix $\boldsymbol{I}\mid_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ reduces to

$$\boldsymbol{I}\mid_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \begin{pmatrix} (1/\widehat{\phi})(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_p) & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n+p}{2\widehat{\phi}^2} \end{pmatrix}.$$

4.2. PERTURBATION IN EXPLANATORY VARIABLES

First, we are going to consider the perturbation in the explanatory variables. We are assuming that the matrix $\boldsymbol{X}$ is ill-conditioned, which means that a small perturbation in the $\boldsymbol{X}$ matrix, may result in a great change in the estimated parameters. In this section we consider the influence that the perturbation in the explanatory variables may result on the ridge estimator.

Let s_i, $i = 1, \ldots, p$, denote the scale factors to account for the different measurement units associated with the columns of $\boldsymbol{X}$ and let

$$\boldsymbol{X}_w = \boldsymbol{X} + \boldsymbol{W}\boldsymbol{S},$$

where $\boldsymbol{W} = (w_{ij})$ is an $n \times p$ matrix of perturbations and $\boldsymbol{S} = diag(s_1, \ldots, s_p)$. In this case, we consider $\boldsymbol{w} = (\boldsymbol{w}_1^\top, \ldots, \boldsymbol{w}_p^\top)^\top = vec(\boldsymbol{W})$ that is a $np \times 1$ vector. Thus, the no perturbation case follows by taking $\boldsymbol{w} = \boldsymbol{0}$. Under this perturbation scheme, it follows that the $(p+1) \times np$ matrix $\boldsymbol{\Delta}$ is given by $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_p)$, where

$\boldsymbol{\Delta}_i = (\boldsymbol{\Delta}_{\beta i}^\top, \boldsymbol{\Delta}_{\phi i}^\top)^\top$, $i = 1, .., p$, with

$$\boldsymbol{\Delta}_{\beta i} = -\frac{2}{\widehat{\phi}} W_f(\widehat{u}_a) s_i(\boldsymbol{e}_*^\top \otimes e_p(i) - \widehat{\beta}_{i*}\boldsymbol{X}^\top) \text{ and } \boldsymbol{\Delta}_{\phi i} = -\frac{2}{\widehat{\phi}^2} W_f(\widehat{u}_a) s_i \widehat{\beta}_{i*} \boldsymbol{e}_*^\top,$$

where $e_p(i)$ is a $p \times 1$ vector with 1 in the ith position and zeros elsewhere, $\otimes$ denotes the Kronecker product, $\boldsymbol{e}_* = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_*$ and $\widehat{u}_a$ as in (3.2).

Concatenating $\boldsymbol{\Delta}_i$, $i = 1, \ldots, p$ horizontally, we have

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_\beta \\ \boldsymbol{\Delta}_\phi \end{pmatrix} = -\frac{2}{\phi} W_f(\widehat{u}_a) \begin{pmatrix} \boldsymbol{S} \otimes \boldsymbol{e}_*{}^\top - \widehat{\boldsymbol{\beta}}_*^\top \boldsymbol{S} \otimes \boldsymbol{X}^\top \\ \frac{1}{\phi} \widehat{\boldsymbol{\beta}}_*^\top \boldsymbol{S} \otimes \boldsymbol{e}_*^\top \end{pmatrix}.$$

In this case, the matrix $\boldsymbol{\Delta}^\top \boldsymbol{I}^{-1} \boldsymbol{\Delta}$, can be written as $\boldsymbol{\Delta}^\top \boldsymbol{I}^{-1} \boldsymbol{\Delta} = \boldsymbol{B}_1 + \boldsymbol{B}_2$, where $\boldsymbol{B}_1 = \frac{2}{\phi} W_f(\widehat{u}_a)(\boldsymbol{S} \otimes \boldsymbol{e}_* - \boldsymbol{S}\widehat{\boldsymbol{\beta}}_* \otimes \boldsymbol{X})(\boldsymbol{X}^\top \boldsymbol{X} + k\boldsymbol{I}_p)^{-1}(\boldsymbol{S} \otimes \boldsymbol{e}_*{}^\top - \widehat{\boldsymbol{\beta}}_*^\top \boldsymbol{S} \otimes \boldsymbol{X}^\top)$ and $\boldsymbol{B}_2 = \frac{4}{\phi^2} d_f(\boldsymbol{S}\widehat{\boldsymbol{\beta}}_* \widehat{\boldsymbol{\beta}}_*^\top \boldsymbol{S} \otimes \boldsymbol{e}_* \boldsymbol{e}_*^\top)$, with $d_f = W_f^2(\widehat{u}_a)/(\frac{n+p}{2} + u_f A_f(\widehat{u}_a))$. Thus, the normal curvature in the direction $\boldsymbol{l}$, when both $\boldsymbol{\beta}$ and ϕ are of interest is given by $C_l = 2|\boldsymbol{l}^\top(\boldsymbol{B}_1 + \boldsymbol{B}_2)\boldsymbol{l}|$.

4.3. PERTURBATION IN THE SCALE MATRIX

In this subsection, we extend the results given in Billor and Loynes (1993, 1999) for the normal model to the elliptical model. The model defined in Section 2 considers that the errors are homocedastic. We consider that the vector of errors in the perturbad model assumes the following form $\boldsymbol{\epsilon}_a \sim El_{n+p}(\boldsymbol{0}, \phi\boldsymbol{\Psi}; f)$, where

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{W}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p \end{pmatrix},$$

with $\boldsymbol{W} = diag(1+w_1, 1+w_2, \ldots, 1+w_n)$. If $w_i = 0$, $i = 1, \ldots, n$, the perturbed model is the same as the postulated model. Under the perturbad model $\boldsymbol{Y}_a \sim El_{n+p}(\boldsymbol{X}_a\boldsymbol{\beta}, \phi\boldsymbol{\Psi}; f)$ and the perturbed pseudo-likelihood function is given by

$$L_p(\boldsymbol{\theta}|\boldsymbol{w}) = -\frac{n+p}{2}\log\phi + \log f(u_{a,w}) + \frac{1}{2}\sum_{j=1}^{n}\log(1+w_j),$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ and $u_{a,w} = \frac{1}{\phi}((\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^\top\boldsymbol{\beta})$. In this case, $\boldsymbol{w}_0 = (0, \ldots, 0)^\top$ is such that $L_p(\boldsymbol{\theta}|\boldsymbol{w}_0) = L_p(\boldsymbol{\theta})$.

4.3.1. COOK'S APPROACH

Now, by setting $Q_{a,w}(\boldsymbol{\beta}) = (\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta})^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{\beta})$, we have that $u_{a,w} = Q_{a,w}(\boldsymbol{\beta})/\phi$,

$$\frac{\partial L_p(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\boldsymbol{\beta}} = -\frac{2}{\phi}W_f(u_{a,w})(\boldsymbol{X}^\top\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) - k\boldsymbol{\beta})$$

and

$$\frac{\partial L_p(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\phi} = -\frac{n+p}{2\phi} - \frac{1}{\phi^2}W_f(u_{a,w})Q_{a,w}(\boldsymbol{\beta}).$$

So that the elements of the $\boldsymbol{\Delta}$ matrix is given by

$$\frac{\partial^2 L_p(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\boldsymbol{\beta}\partial\boldsymbol{w}^\top} = -\frac{2}{\phi}(W_f(u_{a,w})\boldsymbol{X}^\top D(\boldsymbol{\epsilon}) + \frac{1}{\phi}\Delta_f(u_{a,w})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{\epsilon} - k\boldsymbol{\beta})\boldsymbol{\epsilon}^\top D(\boldsymbol{\epsilon}))$$

and

$$\frac{\partial^2 L_p(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\phi\partial\boldsymbol{w}^\top} = -\frac{1}{\phi^2}\{W_f(u_{a,w}) + \frac{1}{\phi}\Delta_f(u_{a,w})Q_{a,w}(\boldsymbol{\beta})\}\boldsymbol{\epsilon}^\top D(\boldsymbol{\epsilon}),$$

where $\boldsymbol{\epsilon} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$. Evaluating the matrix $\boldsymbol{\Delta}$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0$, it follows that

$$\boldsymbol{\Delta} = \begin{pmatrix} -(2/\widehat{\phi})W_f(\widehat{u}_a)\boldsymbol{X}^\top D(\boldsymbol{e}_*) \\ -\frac{1}{\widehat{\phi}^2}(W_f(\widehat{u}_a) + u_f\Delta_f(\widehat{u}_a))\boldsymbol{e}_*^\top D(\boldsymbol{e}_*) \end{pmatrix},$$

where $\boldsymbol{e}_* = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_*$. Note that, if $k = 0$, the expression of the $\boldsymbol{\Delta}$ coincides with the corresponding result given in Galea et al. (1997). Moreover, under the normal ridge regression models, that is, $W_f(u) = -1/2$, for all $u > 0$, it follows that

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{X}^\top D(\boldsymbol{e}_*)/\widehat{\phi} \\ \boldsymbol{e}_*^\top D(\boldsymbol{e}_*)/(2\widehat{\phi}^2) \end{pmatrix},$$

which have the same form as in Galea et. al. (1997) and Cook (1986), but in this case $\boldsymbol{e}_*$ is the vector of ridge residuals, which depends on the ridge estimator.

Under elliptical Model the matrix $\boldsymbol{\Delta}^\top\boldsymbol{I}^{-1}\boldsymbol{\Delta}$, can be written as

$$\boldsymbol{\Delta}^\top\boldsymbol{I}^{-1}\boldsymbol{\Delta} = \boldsymbol{B}_1 + \boldsymbol{B}_2,$$

where

$$\boldsymbol{B}_1 = \frac{2}{\widehat{\phi}}W_f(\widehat{u})D(\boldsymbol{e}_*)\boldsymbol{P}_aD(\boldsymbol{e}_*) \text{ and } \boldsymbol{B}_2 = \frac{1}{\widehat{\phi}^2}C_fD(\boldsymbol{e})\boldsymbol{e}_*\boldsymbol{e}_*^\top D(\boldsymbol{e}_*),$$

with

$$\boldsymbol{P}_a = \boldsymbol{X}^\top(\boldsymbol{X}^\top\boldsymbol{X} + k\boldsymbol{I}_p)^{-1}\boldsymbol{X} \text{ and } C_f = \frac{(W_f(\widehat{u}_a) + u_f\Delta_f(\widehat{u}_a))^2}{\frac{n+p}{2} + u_f A_f(\widehat{u}_a)}. \tag{3.3}$$

Thus, the normal curvature in the direction $\boldsymbol{l}$, when both $\boldsymbol{\beta}$ and ϕ are of interest is

$$C_l = 2|\boldsymbol{l}^\top(\boldsymbol{B}_1 + \boldsymbol{B}_2)\boldsymbol{l}|.$$

In the special case when we are interested in the vector $\boldsymbol{\beta}$, the normal curvature in the direction $\boldsymbol{l}$, is given by

$$C_l(\boldsymbol{\beta}) = 2|\boldsymbol{l}^\top\boldsymbol{B}_1\boldsymbol{l}| = \frac{4}{\widehat{\phi}}|W_f(\widehat{u}_a)|\boldsymbol{l}^\top D(\boldsymbol{e}_*)\boldsymbol{P}_a D(\boldsymbol{e}_*)\boldsymbol{l}.$$

Similarly, the normal curvature for the scale parameter ϕ in the direction $\boldsymbol{l}$, can be written as

$$C_l(\phi) = 2|\boldsymbol{l}^\top\boldsymbol{B}_2\boldsymbol{l}| = \frac{2}{\widehat{\phi}^2}|C_f||\boldsymbol{l}^\top D(\boldsymbol{e}_*)\boldsymbol{e}_*\boldsymbol{e}_*^\top D(\boldsymbol{e}_*)\boldsymbol{l}|,$$

where C_f is as in (3.3).

4.3.2. BILLOR AND LOYNES'S APPROACH

In this case, the rate of change of the function LD^* is measured by the gradient vector as in the normal case and is given by

$$\nabla LD_R^*(\boldsymbol{w}_0) = \frac{\partial LD^*(\boldsymbol{\theta}|\boldsymbol{w})}{\partial \boldsymbol{w}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{w}=\boldsymbol{w}_0} = 2\frac{\partial L_p(\boldsymbol{\theta}|\boldsymbol{w})}{\partial \boldsymbol{w}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{w}=\boldsymbol{w}_0}.$$

After algebraic manipulations, it can be shown that

$$\nabla LD_R^*(\boldsymbol{w}_0) = \frac{2}{\widehat{\phi}}W_f(\widehat{u}_a)D(\boldsymbol{e}_*)\boldsymbol{e}_* + \boldsymbol{1}_n,$$

with $\widehat{u}_a$ as given in (3.2) and $\boldsymbol{e}_*$ is the vector of the ridge residuals. The maximum increment in the LD^* is in the direction of the gradient $\nabla LD_R^*(\boldsymbol{w}_0)$ which is given by

$$\|\nabla LD_R^*\| = (\sum_{j=1}^{n}(1 + \frac{2}{\widehat{\phi}}W_f(\widehat{u}_a)e_{*j}^2)^2)^{1/2}, \tag{3.4}$$

where e_{*j} , $j = 1,..,n$ are the elements of the vector of the ridge residuals. Observe that under normality $\|\nabla LD_R^*\| = (\sum_{j=1}^{n}(1 - \frac{e_{*j}^2}{\widehat{\phi}})^2)^{1/2}$.

5. NUMERICAL ILLUSTRATION

In this Section we are going to apply the results obtained in the earlier sections to the Longley data (Longley, 1967). The data set consists of six explanatory variables with 16 observations. It was previously analyzed by Cook (1977) who has noted that the two observations (observation 5 and 16) with the largest values of D_i (Cooks Statistic) were those observations with great influence in the estimation of $\boldsymbol{\beta}$, while the effect of the observation with the largest studentized residual (observation 10) is not so important relative to the observations 5 and 16. The most influential observations in order were given by the observations 5, 16, 4, 10 and 15. The effect is relative to complete deletion. Later, Walker and Birch (1988) proposed approximate deletion formulas for the detection of influential observations in ridge regression and found that observations 16, 10, 4, 15 and 1 (in that order) were the most influential. More recently, Shi and Wang (1999) studied the effect of local influence proposed in Shi (1997) in ridge regression. Considering the variance perturbation they found that the cases 10, 4, 15, 16 and 1 in that order were the most influential. In the class of elliptical distributions, we are going to apply the methods described in earlier sections considering the Student_t distribution and the exponential power distribution. The density generating function for the Student_t distribution and exponential power distribution is, respectively, given by

$$f(u) = c_1(1+\frac{u}{\nu})^{-\frac{(\nu+n+p)}{2}}, u \ \geq 0 \text{ and } f(u) = c_2 e^{-\frac{u^\alpha}{2}}, \ u \geq 0,$$

with c_1 and c_2 denoting the normalizing constants. As recommended in Lange et al. (1989) and Berkane et al. (1994), we are going to consider 4 degrees os freedom for the Student_t distribution. The exponential power distribution with $\alpha = 1$ is the normal distribution, so we are going to consider $\alpha = 0.8$ and $\alpha = 1.2$. Considering the generating function given above, we have

$$W_f(u) = -\frac{(\nu+n+p)}{2(\nu+u)}, \ u_f = \hat{u}_a = n+p \text{ and } \Delta_f(u) = \frac{(\nu+n+p)}{2(\nu+u)^2},$$

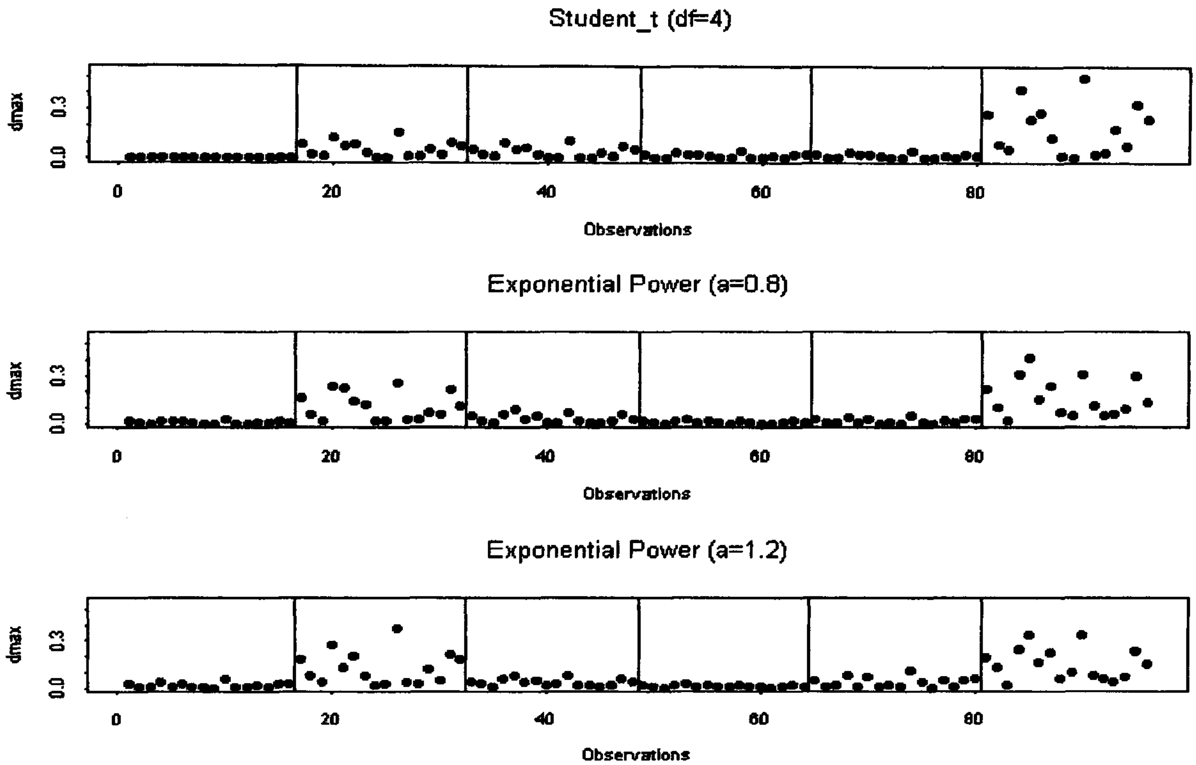

Figure 1: Perturbation in the Explanatory Variables.

for the Student_t distribution and

$$W_f(u) = -\frac{\alpha u^{\alpha-1}}{2},\ u_f = \hat{u}_a = \left(\frac{n+p}{\alpha}\right)^{\frac{1}{\alpha}} \text{ and } \Delta_f(u) = -\frac{\alpha(\alpha-1)u^{\alpha-2}}{2},$$

for the exponential power distribution.

In ridge regression it is well known that the influence of the observation depends on the value of the k. So we should define this value first and for that value of k we are going to determine the observations which are influential for each perturbation schemes. Considering the Hoerl and Kennard estimate for the shrinkage parameter k, given by

$$k = \frac{p\,\hat{\phi}}{\hat{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\beta}}},$$

where p is the number of parameters in the model not counting β_0, $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$ are the maximum likelihood estimator of $\boldsymbol{\beta}$ and ϕ, respectively, we notice that the value of k is the same for the Student_t distribution, whatever the value of the degrees of freedom and it is given by 0.00016. In the exponential power distribution, the value of k depends on the value of α and for $\alpha = 0.8$ ($\alpha = 1.2$) the value of k is given by 0.00006 (0.00032).

Considering the perturbation in the explanatory variables described in subsection 4.2, it was computed the values of d_{max} which are graphed against the observations one to sixteen for each of the covariates $\mathbf{x}_1, \ldots, \mathbf{x}_6$ in Figure 1. The effect of the small perturbation in the covariate six and two relative to the rest of the covariates are apparent.

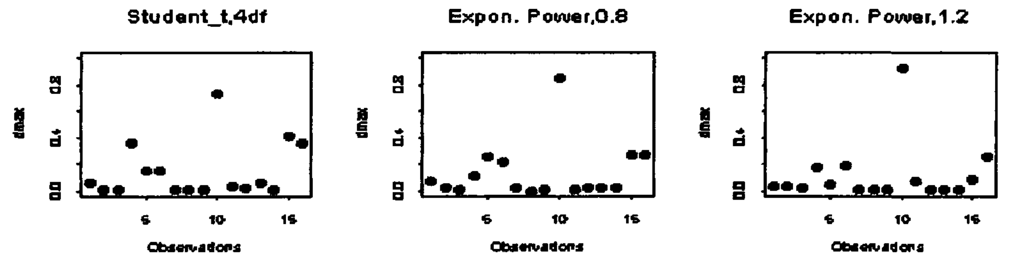

Figure 2: Perturbation in the Scale Matrix (Cook's approach).

The index plot of d_{max} when we perturb the scale matrix are shown in Figure 2. The case 10 stands out in the three graphs and it is followed by the cases 15, 4 and 16 in the Student_t model. Considering the exponential power model with α=0.8 (α=1.2) the most influential case is followed by the observations 16, 15, 5 and 6 (16, 6 and 4).

The individual values l_i of the equation (3.4) may reveal which observations are influential. In Figure 3 the index plot of l_i is shown. Clearly the observations 10 and 4 stands out.

In Figure 4 the total local influence of individual j, given by C_j is graphed. Considering the total local influence the most influential observations is given also by the observations 10 and 4. As can be seen, the influence of the observations change when we consider the ridge regression estimator instead of the ordinary least square estimator and also depends on the distribution of the errors.

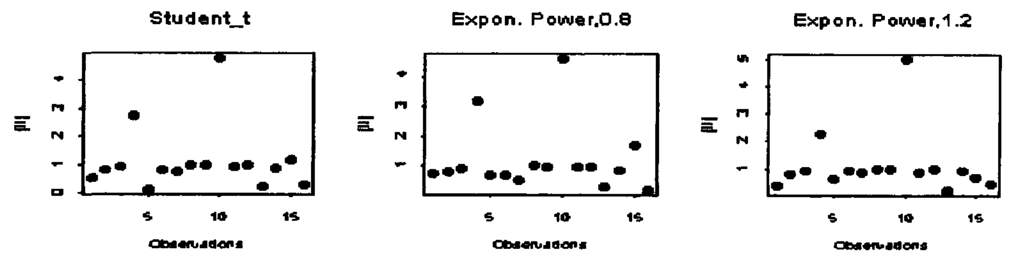

Figure 3: Perturbation in the Scale Matrix (Billor and Loynes's approach).

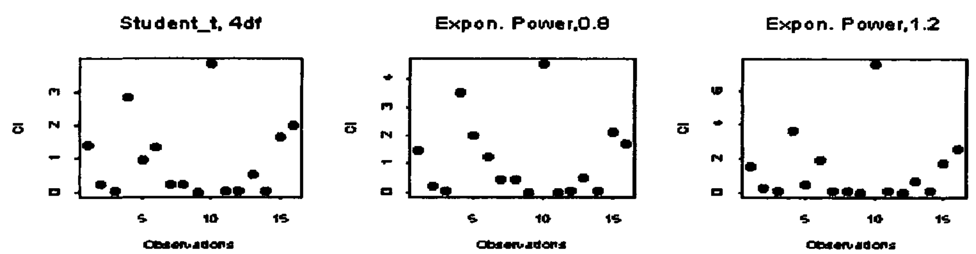

Figure 4: Perturbation in the Scale Matrix (Cj).

Acknowledgements

This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

[1] Berkane, M., Kano, Y. and Bentler, P. M. (1994). Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification. *Computational Statistics & Data Analysis*, **18**, 255-267.

[2] Billor, N. and Loynes, R. M. (1993). Local influence: a new approach. *Communications in Statistics: Theory and Methods*, **22**, 1595-1611.

[3] Billor, N. and Loynes, R. M. (1999). An application of the local influence approach to ridge regression. *Journal of Applied Statistics*, **2**, 177-183.

[4] Cook, D. (1977) - Detection of influential observations in linear regression, *Technometrics*, **19**, 15-18.

[5] Cook, D. (1986) - Assessment of Local Influence, *Journal of Royal Statistical Society*, **48**, B (2), 133-169.

[6] Fang, K. T. and Anderson T.W. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions.* London: Chapman and Hall.

[7] Fang, K. T. and Zhang, Y. T. (1990). *Generalized Multivariate Analysis.* London: Springer.

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

[9] Galea M., Riquelme M. and Paula G. A. (2000). Diagnostic Methods in Elliptical Linear Regression Models. *Brazilian Journal of Probability and Statistics*, **14** (2), 167-184.

[10] Hoerl, A. E. and Kennard, R. W. (1970a). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55-67.

[11] Hoerl, A. E. and Kennard, R. W. (1970b). Ridge regression: application to nonorthogonal problems. *Technometrics*, **12**, 69-82.

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

[13] Lawrence, K.D. and March, L.C. (1984). Robust Ridge Regression Methods for Predicting U.S. Coal Mining Fatalities. *Communication in Statistics-Theory and Methods*, **13**, 139-149.

[14] Longley, J.W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *Journal of the American Statistical Association*, **62** (319), 819-841.

[15] Shi, L. (1997). Local influence in principal components analysis. *Biometrika*, **84** (1), 175-186.

[16] Shi, L. and Wang X. (1999). Local influence in ridge regression. *Computational Statistics & Data Analysis*, **31**, 341-353.

[17] Tsai, C.L. (1986). Discussion of Assessment of Local Influence by R.D. Cook. *Journal of the Royal Statistical Society, Ser. B*, **48**, 165.

[18] Walker, E. and Birch, J. B. (1988). Influence measure in ridge regression. *Technometrics*, **30**, 221-227.

NOTAS DO ICMC

SÉRIE ESTATÍSTICA

077/2005 ANDRADE, M.G. – Estimação bayesiana para processos PAR (*pm*) com a transformação de Box-Cox.

076/2004 COSTA, E.F.; VAL, J.B.R. – Generalized LQ problem for infinite jump linear systems and the minimal solution of algebraic Riccati equations.

075/2004 LOIBEL, S.; VAL, J.B.R.; ANDRADE, M.G. – Inference for the Richards Growth model using Box and Cox transformation and Bootstrap techniques.

074/2003 AOKI, R.; BOLFARINE, H.; ACHCAR, J.A.; LEÃO P. JR., D. – Bayesian analysis of a multivariate null intercept errors-in-variables regression model.

073/2002 LEÃO P. JR., D.; AOKI, R.; SILVA, G.F.- Statistical analysis of proficiency testing results.

072/2002 ANDRADE, MARINHO G.; FERREIRA, VALÉRIA A. MARTINS; OLIVEIRA, SANDRA C. – Abordagem bayesiana com MCMC versus máxima verossimilhança para modelos ARCH(p)

071/2002 LOUZADA-NETO, F.; TOMAZELLA, V.L.D.; ANDRADE, M.G. – Bayesian analysis for reccurent lifetime data with a non homogeneous Poisson process with a frailty term.

070/2002 LEÃO JR., D.; FRAGOS, M. D.; VAL, J. B. R. DO; ANDRADE, M. G. – Separable Hansdorff measurable Radon spaces

069/2002 AOKI, R.; ACHCAR, J.A.; BOLFARINR, H.; SINGER, J.M. – Bayesian analysis of paired data in mixture of normal distributions.

068/2001 LOIBEL, S.; VAL, J.B.R.; ANDRADE, M.G. – Comparação entre as abordagens Bayesiana e clássica do problema de inferência com modelo de crescimento populacional de Richards.