A NOTE ON THE COOK'S DISTANCE

José A. Díaz-García and Graciela González-Farías

Comunicación Técnica No I-01-14/25-07-2001
(PE/CIMAT)

A Note on the Cook's Distance

José A. Díaz-García*
Departamento de Estadística y Cálculo
Universidad Autónoma Agraria Antonio Narro
25350 Buenavista, Saltillo, Coahuila, MÉXICO.
jadiaz@narro.uaaan.mx
and
Graciela González-Farías
Centro de Investigación en Matemáticas A. C.
Callejón de Jalisco s/n
36240 Guanajuato, MÉXICO
farias@cimat.mx

KEY WORD AND PHRASES: Diagnostic tools, Generalized Mahalanobis Distance, Influential data.

ABSTRACT

A modification of the classical Cook's distance is proposed, providing us with a generalized Mahalanobis distance in the context of multivariate elliptical linear regression models. We establish the exact distribution of a pivotal type statistics based on this generalized Mahalanobis distance, which provides critical points for the identification of outlier data points. We illustrate the procedure with an example, in the context of multiple and multivariate linear regression.

1. INTRODUCTION

The identification problem of outliers or influential data, in the univariate or multivariate linear regression setting and under the assumption of Gaussian errors has been studied by several authors like Cook (1977), Besley et al. (1980), Cook and Weisberg (1982) and Chatterjee and Hadi (1988), just to mention a few. Most of these results has been extended to the case of elliptical contour distributions, see for example Galea et al. (1997), Liu (2000) and Díaz-García et al. (2001), among others. One way or another, in all those works, the original idea of the so called Cook's distance is mentioned as a tool for identifying one influential point or sets of influential observations. However when we use this criteria we have only critical points

*This article was written while the first author was a Visiting Professor at the Department of Mathematics of ITESM, Monterrey, México and during a research visit at the Department of Probability and Statistics of CIMAT, Guanajuato, México

provided by an approximate central $\mathcal{F}$ distribution used as it was proposed by Cook (1977). Our purpose here is to modify this distance and derive its exact distribution.
Suppose $Y \in I\!R^{n\times p}$ has an elliptical distribution with location parameter $\mu \in I\!R^{n\times p}$ and scale matrix $\Sigma \otimes \Theta \in I\!R^{np\times np}$ with $\Sigma \in I\!R^{p\times p}$, $\Sigma > 0$ and $\Theta \in I\!R^{n\times n}$, $\Theta > 0$ then the density function is given by

$$f_{_Y}(Y) = |\Sigma|^{-p/2}|\Theta|^{-n/2} g[\mathrm{tr}(\Sigma^{-1}(Y-\mu)^T\Theta^{-1}(Y-\mu))]$$

where $g : I\!R \mapsto [0,\infty)$ is such that $\int_0^\infty u^{(np-2)/2} g(u)du < \infty$ being a density kernel. Let us denote this fact as $Y \sim \mathcal{E}l_{n\times p}(\mu, \Sigma\otimes\Theta, g)$.
This distribution family has been studied by different authors, see for example, Fang and Zhang (1990), Fang and Anderson (1990), Gupta and Varga (1993), among others. The elliptical distribution family includes subfamily distribution functions such as Gaussian, Pearson type VII and Logistic distributions, just to mention some.
Consider the multivariate linear regression model:

$$Y = X\beta + \boldsymbol{\epsilon}, \tag{1}$$

where $Y \in I\!R^{n\times p}$ is the response matrix, $X \in I\!R^{n\times q}$, with $r(X) = q$, $\beta \in I\!R^{q\times p}$ the matrix of the unknown parameters and $\epsilon \in I\!R^{n\times p}$ is an error matrix, such that $\boldsymbol{\epsilon} \sim \mathcal{E}l_{n\times p}(0, \Sigma\otimes I_n, g)$. This model is known as multivariate linear elliptical regression model. If g is a continuous and decreasing function, the maximum likelihood estimators for β and Σ are given by , see Fang and Zhang [pp. 129, 1990],

$$\widehat{\beta} = (X^TX)^{-1}X^TY = X^-Y \quad \text{and} \quad \widehat{\Sigma} = u_0(Y - X\widehat{\beta})^T(Y - X\widehat{\beta}),$$

where X^- is the Moore-Penrose inverse of X, and u_0 maximize the function

$$h(u) = u^{-np}g(p/u), u \geq 0$$

We consider a multivariate linear elliptical regression and propose an extension and modification of the Cook's distance. It will allow us to derive the exact distribution for the new distance providing a critical point to decide if a particular observation (or set of observations) behaves as an outlier.

2. MODIFIED DISTANCE : ONE OBSERVATION

Consider the modified multivariate linear elliptical regression model,

$$Y_{(i)} = X_{(i)}\beta^* + \boldsymbol{\epsilon}_{(i)}, \qquad \boldsymbol{\epsilon}_{(i)} \sim \mathcal{E}l_{(n-1)\times p}(0, \Sigma^*\otimes I_{n-1}, g_{(i)}), \tag{2}$$

we get this model from (1) deleting the ith row from Y, X and $\boldsymbol{\epsilon}$, that is, deleting the ith observation.
For the modified model we have:

$$\widehat{\beta}_{(i)} = (X_{(i)}^TX_{(i)})^{-1}X_{(i)}^TY_{(i)} = X_{(i)}^-Y_{(i)} \quad \text{and} \quad \widehat{\Sigma}_{(i)} = u_{0_{(i)}}(Y_{(i)} - X_{(i)}\widehat{\beta}_{(i)})^T(Y_{(i)} - X_{(i)}\widehat{\beta}_{(i)}),$$

First of all we need to work out a simple representation for $\widehat{\beta} - \widehat{\beta}_{(i)}$. For that, consider the following partition matrices:

$$Y = \begin{pmatrix} Y_1^T \\ Y_2^T \\ \vdots \\ Y_n^T \end{pmatrix}, \quad Y_i \in I\!R^p \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1^T \\ \boldsymbol{\epsilon}_2^T \\ \vdots \\ \boldsymbol{\epsilon}_n^T \end{pmatrix}, \quad \boldsymbol{\epsilon}_i \in I\!R^p \qquad X = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix}, \quad X_i \in I\!R^q.$$

therefore

$$X^T X = (X_1 X_2 \cdots X_n) \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix} = \sum_{k=1}^{n} X_k X_k^T = \sum_{k \neq i}^{n} X_k X_k^T + X_i X_i^T = X_{(i)}^T X_{(i)} + X_i X_i^T,$$

and

$$X^T Y = (X_1 X_2 \cdots X_n) \begin{pmatrix} Y_1^T \\ Y_2^T \\ \vdots \\ Y_n^T \end{pmatrix} = \sum_{k=1}^{n} X_k Y_k^T = \sum_{k \neq i}^{n} X_k Y_k^T + X_i Y_i^T = X_{(i)}^T Y_{(i)} + X_i Y_i^T.$$

Note that if e_i^n is the *ith* vector of the canonical base in $I\!R^n$, that is, the unit vector, $e_i^n = (0 \cdots 0\ 1\ 0 \cdots 0)^T$, then: ${e_i^n}^T Y = Y_i^T$, ${e_i^n}^T X = X_i^T$ and ${e_i^n}^T \boldsymbol{\epsilon} = \boldsymbol{\epsilon}_i^T$.
By Rao [pp. 33, 1973], if A is nonsingular, v and u are two arbitrary vector, then

$$(A - uv^T)^{-1} = A^{-1} + \frac{A^{-1} u v^T A^{-1}}{1 - v^T A^{-1} u}$$

therefore, if we defined $A = X^T X$ and $u = v = X_i$, we get

$$(X^T X - X_i X_i^T)^{-1} = (X_{(i)}^T X_{(i)})^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} X_i X_i^T (X^T X)^{-1}}{(1 - h_i)}, \tag{3}$$

with $h_i = X_i^T (X^T X)^{-1} X_i$.
By (3), we get,

$$\begin{aligned} \widehat{\beta} - \widehat{\beta}_{(i)} &= (X^T X)^{-1} X^T Y - (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T Y_{(i)} \\ &= \left((X_{(i)}^T X_{(i)})^{-1} - \frac{(X^T X)^{-1} X_i X_i^T (X^T X)^{-1}}{(1 - h_i)} \right) X^T Y - (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T Y_{(i)} \\ &= (X_{(i)}^T X_{(i)})^{-1} (X^T Y - X_{(i)}^T Y_{(i)}) - \frac{(X^T X)^{-1} X_i X_i^T (X^T X)^{-1} X^T Y}{(1 - h_i)} \\ &= (X_{(i)}^T X_{(i)})^{-1} X_i Y_i^T - \frac{(X^T X)^{-1} X_i X_i^T (X^T X)^{-1} X^T Y}{(1 - h_i)}. \end{aligned} \tag{4}$$

Using (3) on, the first part of (4) we have,

$$\begin{aligned}(X_{(i)}^T X_{(i)})^{-1} X_i Y_i^T &= \left((X^T X)^{-1} + \frac{(X^T X)^{-1} X_i X_i^T (X^T X)^{-1}}{(1-h_i)}\right) X_i Y_i^T \\ &= (X^T X)^{-1} X_i Y_i^T + \frac{h_i (X^T X)^{-1} X_i Y_i^T}{(1-h_i)} \\ &= \frac{(X^T X)^{-1} X_i Y_i^T}{(1-h_i)}. \end{aligned} \tag{5}$$

Substituting (5) in (4)

$$\begin{aligned}\widehat{\beta} - \widehat{\beta}_{(i)} &= \frac{(X^T X)^{-1} X_i Y_i^T - (X^T X)^{-1} X_i X_i^T (X^T X)^{-1} X^T Y}{(1-h_i)} \\ &= \frac{(X^T X)^{-1} X_i}{(1-h_i)} (Y_i^T - X_i^T (X^T X)^{-1} X^T Y) \end{aligned} \tag{6}$$

Now, since $\widehat{\boldsymbol{\epsilon}} = (Y - X\widehat{\beta}) = (I - XX^-)Y = (I-P)Y$, where P is the orthogonal projector over the image of X. Then $e_i^{n^T} \widehat{\boldsymbol{\epsilon}} = \widehat{\boldsymbol{\epsilon}}_i^T = e_i^{n^T}(Y - X\widehat{\beta}) = Y_i^T - X_i^T (X^T X)^{-1} X^T Y$, so, we obtain:

$$\widehat{\beta} - \widehat{\beta}_{(i)} = \frac{(X^T X)^{-1} X_i \widehat{\boldsymbol{\epsilon}}_i^T}{(1-h_i)} \tag{7}$$

Under the assumption of the elliptical distribution having moments, we propose the following modification to the Cook's distance, called $\mathcal{D}_m$:

$$\mathcal{D}_m = \operatorname{vec}(\widehat{\beta} - \widehat{\beta}_{(i)})^T \widehat{\operatorname{Cov}}(\operatorname{vec}(\widehat{\beta} - \widehat{\beta}_{(i)}))^- \operatorname{vec}(\widehat{\beta} - \widehat{\beta}_{(i)}) \tag{8}$$

Note 1. Expression (8) is no more than an extension for the squared Mahalanobis generalized distance, see Rao and Mitra [203-206, 1971].
The second step is to find a simple expression for the variance covariance matrix $\operatorname{Cov}(\operatorname{vec}(\widehat{\beta} - \widehat{\beta}_{(i)}))$, under the assumption that the error distribution has moments.
Recall that $Y \sim \mathcal{E}l_{n\times p}(\mu, \Sigma \otimes \Theta, g)$, then its characteristic function is given by

$$\Psi_{_Y}(T) = \operatorname{etr}(i\mu T^T)\phi(\operatorname{tr}(\Sigma T^T \Theta T))$$

and, $E(Y) = \mu$, $\operatorname{Cov}(\operatorname{vec}(Y)) = c_0(\Sigma \otimes \Theta)$, with $c_0 = -2\phi'(0)$, see Gupta and Varga [pp. 33, 1993].
Since $\widehat{\boldsymbol{\epsilon}}_i^T = e_i^{n^T}(Y - X\widehat{\beta}) = e_i^{n^T}(I-P)Y$, it is clear that $\operatorname{vec} \widehat{\boldsymbol{\epsilon}}_i^T = (I_p \otimes e_i^{n^T}(I-P)) \operatorname{vec} Y$, therefore

$$\begin{aligned}\operatorname{vec}(\widehat{\beta} - \widehat{\beta}_{(i)}) &= \frac{(I_p \otimes (X^T X)^{-1} X_i)}{(1-h_i)} (I_p \otimes e_i^{n^T}(I-P)) \operatorname{vec} Y \\ &= \frac{(I_p \otimes (X^T X)^{-1} X_i e_i^{n^T}(I-P))}{(1-h_i)} \operatorname{vec} Y \\ &= \frac{(I_p \otimes (X^T X)^{-1} X_i P_i^T)}{(1-h_i)} \operatorname{vec} Y \end{aligned} \tag{9}$$

where $P_i^T = {e_i^n}^T(I-P)$ is the ith row of the matrix $(I-P)$. Then

$$\begin{aligned}
\mathrm{Cov}(\mathrm{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})) &= \frac{(I_p\otimes(X^TX)^{-1}X_iP_i^T)}{(1-h_i)}\,\mathrm{Cov}(\mathrm{vec}\,Y)\frac{(I_p\otimes(X^TX)^{-1}X_iP_i^T)^T}{(1-h_i)} \\
&= \frac{(I_p\otimes(X^TX)^{-1}X_iP_i^T)}{(1-h_i)^2}(\Sigma^*\otimes I)(I_p\otimes P_iX_i^T(X^TX)^{-1}) \\
&= \frac{||P_i||^2(\Sigma^*\otimes(X^TX)^{-1}X_iX_i^T(X^TX)^{-1})}{(1-h_i)^2} \qquad (10)
\end{aligned}$$

where $\Sigma^* = c_0\Sigma$.
Note that,

$$\begin{aligned}
||P_i||^2 &= {e_i^n}^T(I-P)(I-P)e_i^n \\
&= {e_i^n}^T(I-P)e_i^n \\
&= {e_i^n}^Te_i^n - {e_i^n}^TX(X^TX)^{-1}X^Te_i^n \\
&= 1-X_i(X^TX)^{-1}X_i^T \\
&= 1-h_i \qquad (11)
\end{aligned}$$

Substituting (11) in (10) we get,

$$\mathrm{Cov}(\mathrm{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})) = \frac{(\Sigma^*\otimes(X^TX)^{-1}X_iX_i^T(X^TX)^{-1})}{(1-h_i)} \qquad (12)$$

Let $S_1 = \widehat{\Sigma}/(u_0(n-q))$ and observe that $E(S_1) = \Sigma^*$, see Fang and Zhang [pp. 138, 1990] , then

$$\widehat{\mathrm{Cov}}(\mathrm{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})) = \frac{(S_1\otimes(X^TX)^{-1}X_iX_i^T(X^TX)^{-1})}{(1-h_i)} \qquad (13)$$

Let $r_i = (X^TX)^{-1}X_i$. Based on the following standard results:

1. For $a\in\mathbb{R}^n$, $a^- = a^T/||a||^2$,
2. Given $A\in\mathbb{R}^{p\times q}$, $(AA^T)^- = A^{T-}A^-$ with $A^{-1}=A^-$ if A is non singular,
3. Given the matrices A and B, $(A\otimes B)^- = A^-\otimes B^-$,

   we get,

$$\begin{aligned}
(\widehat{\mathrm{Cov}}(\mathrm{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})))^- &= \widehat{\mathrm{Cov}}(\mathrm{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}))^- \\
&= \left(\frac{(S_1\otimes r_ir_i^T}{(1-h_i)})\right)^- \\
&= \frac{(1-h_i)}{||r_i||^4}(S_1^{-1}\otimes r_ir_i^T)
\end{aligned}$$

Therefore the modified Cook's distance can be rewritten as:

$$\begin{aligned}
\mathcal{D}_m &= \operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})^T \widehat{\operatorname{Cov}}(\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}))^- \operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}) \\
&= \left(\frac{(I_p\otimes(X^TX)^{-1}X_iP_i^T)\operatorname{vec}Y}{(1-h_i)}\right)^T \frac{(1-h_i)(S_1^{-1}\otimes r_ir_i^T)}{||r_i||^4}\left(\frac{(I_p\otimes(X^TX)^{-1}X_iP_i^T)\operatorname{vec}Y}{(1-h_i)}\right) \\
&= \frac{(1-h_i)^{-1}}{||r_i||^4}\operatorname{vec}^T Y(S_1^{-1}\otimes P_ir_i^Tr_ir_i^Tr_iP_i^T)\operatorname{vec}Y \\
&= (1-h_i)^{-1}\operatorname{vec}^T Y(S_1^{-1}\otimes P_iP_i^T)\operatorname{vec}Y. \qquad (14)
\end{aligned}$$

Alternatively, since $\operatorname{tr}BX^TCXD = \operatorname{vec}^T X(B^TD^T\otimes C)\operatorname{vec}X = \operatorname{vec}^T X(DB\otimes C^T)\operatorname{vec}X$, for matrices of the correct sizes, we can write $\mathcal{D}_m$ as

$$\mathcal{D}_m = (1-h_i)^{-1}\operatorname{tr}S_1^{-1}Y^TP_iP_i^TY.$$

On the other hand, since $\widehat{\boldsymbol{\epsilon}}_i^T = e_i^{nT}(Y-X\widehat{\beta}) = P_iY$, then,

$$\begin{aligned}
\mathcal{D}_m &= (1-h_i)^{-1}\operatorname{tr}S_1^{-1}\widehat{\boldsymbol{\epsilon}}_i\widehat{\boldsymbol{\epsilon}}_i^T \\
&= (1-h_i)^{-1}\operatorname{tr}\widehat{\boldsymbol{\epsilon}}_i^TS_1^{-1}\widehat{\boldsymbol{\epsilon}}_i \\
&= (1-h_i)^{-1}\widehat{\boldsymbol{\epsilon}}_i^TS_1^{-1}\widehat{\boldsymbol{\epsilon}}_i
\end{aligned}$$

In this way we have the following alternative expressions for the square of the modified Cook's distance:

$$\mathcal{D}_m = \begin{cases}
\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})^T\widehat{\operatorname{Cov}}(\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}))^-\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}) \\
(1-h_i)^{-1}\operatorname{vec}^TY(S_1^{-1}\otimes P_iP_i^T)\operatorname{vec}Y \\
(1-h_i)^{-1}\operatorname{tr}S_1^{-1}Y^TP_iP_i^TY \\
(1-h_i)^{-1}\widehat{\boldsymbol{\epsilon}}_i^TS_1^{-1}\widehat{\boldsymbol{\epsilon}}_i
\end{cases} \qquad (15)$$

Note 2. According with Chatterjee and Hadi [pp. 124, 1988], we could replace the matrix S_1 by one obtained using the reduced sample $(n-1)$, denoted by S_{1_i}.

Note 3.. Cook (1977), Chatterjee and Hadi [pp. 117, 1988] , Díaz- García et. al (2001), and many others use the variance covariance matrix of $vec(\widehat{\beta})$ to construct the distance measure. The reformulation we proposed is based on the replacement of that variance-covariance matrix by the variance covariance matrix of $\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})$. We can also find this idea on Chatterjee and Hadi [pp. 150, 1988], for the univariate case, but for the evaluation of influential data on a particular regression coefficient, only the variance of one coefficient is used instead of the variance of the difference. The problem when this idea is extended to the multivariate case is that such matrix is singular, so we need to consider the Moore-Penrose inverse for the variance covariance matrix of $\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})$.

Now we will consider the case when the errors from the multivariate regression model has an elliptical distribution without moments, (for example, when the errors have a matrix Cauchy distribution, see Gupta and Varga [pp. 76, 1993]), then the proposed distance maybe be defined by (15), without taking in account that,

$$\frac{(S_1\otimes(X^TX)^{-1}X_iX_i^T(X^TX)^{-1})}{(1-h_i)}$$

is the variance covariance matrix of $(\operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}))$. Let us called the distance matrix.

Theorem 1. *Consider the elliptical regression multivariate model given by (1), were the error distribution may or may not have moments. Then a modified squared Cook's distance to detect an outlier data, it can be written as:*

$$\mathcal{D}_m = \begin{cases} \operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)})^T \left(\dfrac{(S_1 \otimes (X^TX)^{-1}X_iX_i^T(X^TX)^{-1})}{(1-h_i)} \right)^{-} \operatorname{vec}(\widehat{\beta}-\widehat{\beta}_{(i)}) \\ (1-h_i)^{-1} \operatorname{vec}^T Y(S_1^{-1} \otimes P_iP_i^T) \operatorname{vec} Y \\ (1-h_i)^{-1} \operatorname{tr} S_1^{-1}Y^TP_iP_i^TY \\ (1-h_i)^{-1}\widehat{\boldsymbol{\epsilon}}_i^T S_1^{-1}\widehat{\boldsymbol{\epsilon}}_i \end{cases} \tag{16}$$

Note 4. From *(16)* is easy to see that if we want to implement this measure for all the data points, it is enough to fit the model once and from the usual output, we can construct the modified distance for each point. Note that the expression $\mathcal{D}_m$, on the univariate Normal case, coincides with the analysis of studentized residuals, see Besley et al. (p. 201, 1980) and Chatterjee (p. 78, 1988).

3. MODIFIED DISTANCE : MULTIPLE OBSERVATIONS

Let $I = \{i_1, i_2, \ldots, i_k\}$ a subset of size k from $\{1, 2, \ldots, n\}$, such that $(n-k) \geq q$. Now, under model (1), denote by $X_{(I)}$, $Y_{(I)}$ and $\widehat{\boldsymbol{\epsilon}}_{(I)}$, the regression, the data and the error matrices respectively, after deleting the corresponding observations according with the subindexes on I. Let $\widehat{\beta}_{(I)}$, and $\widehat{\Sigma}_{(I)}$ the corresponding maximum likelihood estimator in the model

$$Y_{(I)} = X_{(I)}\beta^* + \boldsymbol{\epsilon}_{(I)}, \qquad \boldsymbol{\epsilon}_{(I)} \sim \mathcal{E}l_{(n-k)\times p}(0, \Sigma^* \otimes I_{n-k}, g_{(I)}).$$

Based on the equality

$$(A - BCD^T)^{-1} = A^{-1} + A^{-1}B(D^{-1} - C^TA^{-1}B)^{-1}C^TA^{-1}$$

where A and D are non singular matrices of order s and m respectively, B and C matrices of order $s \times m$ and using similar procedures as those from section 2, it is easy to verify that,

$$\widehat{\beta} - \widehat{\beta}_{(I)} \;=\; (X^TX)^{-1}X_I(I-H_I)^{-1}\widehat{\boldsymbol{\epsilon}}_I,$$

with $(I - H_I) = (I_k - X_I^T(X^TX)^{-1}X_I)$ and X_I the matrix with the corresponding rows of X according with I. Observe that, $\widehat{\boldsymbol{\epsilon}}_I = U_I^T\widehat{\boldsymbol{\epsilon}} = U_I^T(I-P)Y$, where

$$U_I^T = \begin{pmatrix} e_{i_1}^{n^T} \\ e_{i_2}^{n^T} \\ \vdots \\ e_{i_k}^{n^T} \end{pmatrix}$$

We get,

$$\operatorname{vec}(\widehat{\beta} - \widehat{\beta}_{(I)}) = (I_p \otimes (X^TX)^{-1}X_I(I-H_I)^{-1}P_I) \operatorname{vec} Y,$$

with $P_I = U_I^T(I - P)$.
If we consider the case of existing moments for the distribution of errors,

$$\mathrm{Cov}(\mathrm{vec}(\widehat{\beta} - \widehat{\beta}_{(I)})) = (\Sigma^* \otimes (X^T X)^{-1} X_I (I - H_I)^{-1} X_I^T (X^T X)^{-1})$$

and

$$\widehat{\mathrm{Cov}}(\mathrm{vec}(\widehat{\beta} - \widehat{\beta}_{(I)})) = (S_1 \otimes (X^T X)^{-1} X_I (I - H_I)^{-1} X_I^T (X^T X)^{-1})$$

Under the same arguments for the non existing moments given in section 2, we have,

Theorem 2. *Consider the elliptical regression multivariate model given by (1), were the* $\boldsymbol{\epsilon}$ *may or may not have moments. Then a modified squared Cook's distance to detect k influential data points , it can be written as*

$$\mathcal{D}_{m_I} = \begin{cases} \mathrm{vec}(\widehat{\beta} - \widehat{\beta}_{(I)})^T (S_1 \otimes (X^T X)^{-1} X_I (I - H_I)^{-1} X_I^T (X^T X)^{-1})^- \mathrm{vec}(\widehat{\beta} - \widehat{\beta}_{(I)}) \\ \mathrm{vec}^T Y (S_1^{-1} \otimes P_I^T (I - H_I)^{-1} P_I) \,\mathrm{vec}\, Y \\ \mathrm{tr}\, S_1^{-1} Y^T P_I^T (I - H_I)^{-1} P_I Y \\ \mathrm{tr}\, S_1^{-1} \widehat{\boldsymbol{\epsilon}}_I^T (I - H_I)^{-1} \widehat{\boldsymbol{\epsilon}}_I \end{cases} \tag{17}$$

4. DISTRIBUTION FUNCTIONS ASSOCIATED WITH THE MODIFIED DISTANCES

The main reason to explore the modifications given in section 2 and 3 for the squared Cook's distance is that, instead of using an approximate $\mathcal{F}$ distribution we will be able to derived an exact distribution for $\mathcal{D}_m$, which is invariant under the family of elliptical distributions. We propose a pivotal type estimator with an exact distribution for $\mathcal{D}_{m_I}$, the case of detection of several influential data simultaneously.
We will derive the distribution under the pivotal statistics for both cases: one influential observation at that time and multiple observations.

Theorem 3. *Under the assumptions of Theorem 1, we have,*

$$\frac{(n-q-p)}{p(n-q-1)} \mathcal{D}_m^* \sim \mathcal{F}_{p,(n-q-p)} \tag{18}$$

where $\mathcal{F}_{p,(n-q-p)}$ *denote a central* $\mathcal{F}$ *distribution with* p *and* $(n-q-p)$ *degrees of freedom (df) and* $\mathcal{D}_m^*$ *is given by (15) substituting* S_1 *by* S_{1_i}*, see Note 2.*

Proof **:** It follows immediately from Theorem 5.2.2 in Anderson [p. 163, 1984] and Theorem 5.1.1 from Fang and Zhang [pp. 154, 158, 1990]. ∎

From Theorem 3, given a significance level α, we may write the following decision rule:
Y_i, $i = 1, 2, \ldots, n$, is an outlier observation if

$$\frac{(n-q-p)}{p(n-q-1)} \mathcal{D}_m^* \geq \mathcal{F}_{\alpha:p,(n-q-p)} \tag{19}$$

where $\mathcal{F}_{\alpha:p,(n-q-p)}$ the corresponding upper $\alpha - percentil$ from a $\mathcal{F}$ distribution with p and $(n-q-p)$ df .

Note 5. For the univariate case, $p = 1$, the decision rule becomes: Y_i, $i = 1, 2, \ldots, n$, is an outlier if

$$\mathcal{D}_m^* \geq \mathcal{F}_{\alpha:1,(n-q-1)} \tag{20}$$

where $\mathcal{F}_{\alpha:1,(n-q-1)}$ is the $\alpha - percentil$ from a $\mathcal{F}$ distribution with 1 and $(n-q-1)$ df, or equivalently:

$$\mathcal{D}_m^{*\,1/2} \geq \boldsymbol{t}_{\alpha/2:(n-q-1)} \tag{21}$$

where $\boldsymbol{t}_{\alpha/2:(n-q-1)}$ is the upper $\alpha/2 - percentil$ of a $\boldsymbol{t}$ distribution with $(n-q-1)$ df.

In a similar way, when we deal with multiple observations:

Theorem 4. *Under the assumptions of Theorem 2, we have,*

$$\frac{\mathcal{D}_{m_I}^*}{n-q-k} \sim \mathcal{LH}_{s,m,h} \tag{22}$$

where $\mathcal{LH}_{s,m,h}$ denote the central distribution for the Lawley-Hotelling statistics with parameters $s = \min(p,k)$,$m = (|p-k|-1)/2$ and $h = (n-q-p-1)/2$ and $\mathcal{D}_{m_I}^*$ is given by (17) substituting S_1 by S_{1_I}.

Proof: It follows immediately from Theorem 5.3.1 from Gupta A.K and Varga T. [pp. 182, 301, 1993] and Theorem 10.6.2 , Corolary 10.6.3, in Muirhead [pp. 468-471 and p. 471, 1982]. ■

5. AN APPLICATION

We illustrate the use of the exact test given in Section 4 under two scenarios : simple regression and multivariate multiple regression.
The first data set was presented by Cook and Weisberg (pp. 204-207, 1994). This is a small data set with observations on 21 children, giving their AGE in months at first spoken word, and a $SCORE$, which is a measure of the development of the child. A plot $(AGE, SCORE)$ is given in Figure *1 a).* It is clear that are three observations that have a distinguishable behavior: 18, 1 and 17. If we think on an ordinary least squares (OLS) linear fit, $SCORE$ seems to decrease with AGE . Case 18 appears to be poorly fitted by the linear trend, relative to the other data. Cases 1 and 17, have relatively large values of AGE. Figure *1 b),* shows the Q-Q plot of the residuals from the OLS linear fit. It is clear that only observation 18 seems to

be a candidate for an *outlier* as it is defined in Chatterjee (pp.94-95, 1988).

Figure 1. Original data for the adaptive score measure and a Q-Q plot for the residual on the simple regression of *SCORE* on *AGE*.

Figure 2 shows the identification and detection of influential and outlier points, using different techniques. This analysis emphasize the fact mentioned before: studentized residuals coincide with the modified distance for the univariate case. Moreover studentized residuals are the base for the John and Draper distance as it is discussed in Draper and Smith (pp. 169-175, 1981). Figure 2 a) shows how the Cook's distance detects observation 17th, which is a leverage point as it is described in Cook and Weisberg (1994). Taking into account Figure 1 b) and according with Figures 2 b) and 2 c), observation 18 was a candidate for an outlier and the tests are in favor to declare observation 18th as an outlier observation. Note that the critical value in Figure 2 c) is the approximate value of an $\mathcal{F}$-distribution multiplied by $s^2(i)$ for $i = 18$ in order to plot the original distance of Draper and John.
For the multivariate multiple regression example, we generated 19 observations from a model $y = X\beta + \epsilon$, with normal errors, $p = 2$, and $q = 4$.

Figure 2. Identification of influence and outlier points based on a) Cook's Distance, b) The Modified Cook's Distance according with (20), c) Draper & John Distance. $n = 21$, $q = 2$, $p = 1$ and $s^2(i)$, the residual variance without the i-th observation was use in all the cases according with Note 2.

We fitted the model and constructed E, the residual sum of square matrix, then applied the Mahalanobis squared distance as it is shown in Seber (pp. 152-153, 1984). Figure 3 a) shows the Mahalanobis squared distance and suggests observations 10 and 11 as possible outliers. We applied *(22)*, with $k = 2$, $Dm = 4.77$, compared with a critical value of 3.015; the percentile is approximated by an $\mathcal{F}$ distribution as suggested by Seber (pp. 38-39, 563-564, 1984). The test is in favor of considering observations 10 and 11 as outliers.
Figure 3 b) shows the same analysis taken one observation at a time. We use an $\mathcal{F}$ test based on $(1 - \alpha/n)$ instead of $(1 - \alpha)$ to get a simultaneous test with a nominal level at least α. In this case we get the same conclusion as with the test based on k observations. We recommend the use of the test based on k observations as given in (22), and selecting the k point using a graphical method as the one given in Figure 3 a).

It is important to recall this test is valid under elliptical distributions and not only for the normal error case.

Figure 3. Identification of outliers based on the Mahalanobis Distance on the residual matrix and detection of outliers based on the Modified Cook's Distance as given in Theorem 3.

6. ACKNOWLEDGEMENT

We are grateful to J. Ramón Domínguez for helping us with the figures. An S-plus program for all the calculations is available under request to jrdguez@cimat.mx

REFERENCES

[1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.

[2] Besley, D., Kuh, E., and Welsch, R. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York.

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

[4] Cook, R. D., and Weisberg, S. (1982). *Residual and Influence in Regression*, Chapman and Hall, London.

[5] Cook, R. D., and Weisberg, S. (1994). *An Introduction to Regression Graphics*, John Wiley & Sons. New York.

[6] Chatterjee, S., and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*, John Wiley & Sons. New York.

[7] Díaz-García, J.A., Galea, M., and Leiva- Sánchez, V. (2001). "Influence diagnostics for elliptical regression linear models", Submitted for publication.

[8] Draper, N., and Smith, H. (1981). *Applied Regression Analysis*, (2nd ed.), John Wiley & Sons, New York.

[9] Fang, K. T., and Anderson T. W. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*, Allerton Press Inc., New York.

[10] Fang, K. T., and Zhang, Y. T. (1990). *Generalized Multivariate Analysis*, Science Press, Beijing, Springer-Verlang.

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

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

[13] Liu, S. Z. (2000). "On local influence for elliptical linear models", *Statistical Papers* 41, 211-224.

[14] Muirhead, R. J. (1982). *Aspects of Multivariated Statistical Theory*, John Wiley & Sons, New York.

[15] C. R. Rao,(1973). *Linear Statistical Inference and its Applications* (2nd ed.), John Wiley & Sons, New York.

[16] Rao, C. R. and Mitra, S. K.(1971). *Generalized Inverse of Matrices and its Applications* (2nd ed.), John Wiley & Sons, New York.

[17] Seber, G.A.F.(1984). *Multivariate Observations* , John Wiley & Sons, New York.