The Distribution of the Residual from a General Linear Model: Univariate and Multivariate Cases Under Normal and Elliptical Models

José A. Díaz-García and Ramón Gutiérrez-Jáimez

Comunicación Técnica No I-04-11/30-11-2004
(PE/CIMAT)

The Distribution of the Residual from a General Linear Model: Univariate and Multivariate Cases Under Normal and Elliptical Models

José A. Díaz-García* and Ramón Gutiérrez-Jáimez†

Given a general linear model of full or less than full rank, we find the distributions of normalised, standardised and studentised (internally and externally studentised) residuals, in univariate and multivariate cases, assuming normal and elliptical distributions. Also, we propose an alternative approach to the results by Ellenberg (1973) and Beckman and Trusell (1974).

KEY WORDS: Residual, normalised residual, studentised residual, Pearson type II distribution, matrix-variate t distribution, elliptical distribution, singular distribution, matric-variate t distribution.

1. INTRODUCTION

Consider the multivariate general linear model

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

where $\mathbf{Y}$ and $\boldsymbol{\epsilon}$ are $n \times p$ random matrices, $\mathbf{X}$ is a known $n \times q$ matrix, and $\boldsymbol{\beta}$ is an unknown $q \times p$ matrix of parameters called regression coefficients. We shall assume throughout this work that $\mathbf{X}$ has rank $\alpha \leq q$, $n \geq p+\alpha$. First, we shall assume that $\boldsymbol{\epsilon}$ has a matrix-variate normal distribution, that is $\boldsymbol{\epsilon} \sim \mathcal{N}_{n\times p}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$ such that $\mathbf{Y} \sim \mathcal{N}_{n\times p}(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n \otimes \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix, $\boldsymbol{\Sigma} > 0$. Thus the maximum likelihood estimates of $\mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are

$$\widetilde{\mathbf{X}\boldsymbol{\beta}} = \mathbf{X}\widetilde{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{Y} = \mathbf{X}\mathbf{X}^+Y \tag{2}$$

and

$$\widetilde{\boldsymbol{\Sigma}} = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}) \tag{3}$$

where $\mathbf{A}^-$ is the generalised inverse (also termed c-inverse) such that $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ and $\mathbf{A}^+$ denote the Moore-Penrose inverse of A. Thus, the estimator $\mathbf{X}\widetilde{\boldsymbol{\beta}}$ is invariant under any generalised inverse $(\mathbf{X}^T\mathbf{X})^-$ of

*José A. Díaz-García is Professor, Department of Statistics and Computation, Universidad Autónoma Agraria Antonio Narro, 25350 Buenavista, Saltillo, Coahuila, MÉXICO; E-mail: jadiaz@uaaan.mx. This work was supported partially by the research project 39017E of CONACYT-México.

†Ramón Gutiérrez-Jáimez is Head, Department of Statistics and O.R, University of Granada, Granada 18071, SPAIN; E-mail: rgjaimez@goliat.ugr.es

$\mathbf{X}^T\mathbf{X}$, see Graybill (1985), Srivastava and Khatri (1979, p. 171) and Muirhead (1982, p. 430). Moreover, $\mathbf{X}\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\Sigma}}$ are independently distributed; $\mathbf{X}\widetilde{\boldsymbol{\beta}} \sim \mathcal{N}_{n\times p}(\mathbf{X}\boldsymbol{\beta}, \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T \otimes \boldsymbol{\Sigma})$ and $n\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{W}_p(n-\alpha, \boldsymbol{\Sigma})$, see Srivastava and Khatri (1979, p. 171) and Muirhead (1982, p. 431). Finally, we shall denote by $\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}} = n\widetilde{\boldsymbol{\Sigma}}/(n-\alpha)$, the unbiased estimators of $\mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, respectively.

The residual matrix is defined as $\widehat{\boldsymbol{\epsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\widehat{\beta} = (\mathbf{I}_n - \mathbf{X}\mathbf{X}^+)\mathbf{Y} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}$, where $\mathbf{H} = \mathbf{X}\mathbf{X}^+$ is the orthogonal projector on the image of $\mathbf{X}$. Then $\widehat{\boldsymbol{\epsilon}}$ has a singular matrix-variate normal distribution of rank $p(n-\alpha)$, i.e. $\widehat{\boldsymbol{\epsilon}} \sim \mathcal{N}_{n\times p}^{(n-\alpha),p}(\mathbf{0}, (\mathbf{I}_n - \mathbf{H}) \otimes \boldsymbol{\Sigma})$, with $\operatorname{cov}(\operatorname{vec}(\widehat{\boldsymbol{\epsilon}}^{\,T})) = ((\mathbf{I}_n - \mathbf{H}) \otimes \boldsymbol{\Sigma})$, see Khatri (1968) and Díaz-García et al. (1997). Also, observe that the i-th row of $\widehat{\boldsymbol{\epsilon}}$, denoted as $\widehat{\boldsymbol{\epsilon}}_i$, has a nonsingular p-variate normal distribution, i.e., $\widehat{\boldsymbol{\epsilon}}_i \sim \mathcal{N}_p(0, (1-h_{ii})\boldsymbol{\Sigma})$, $\mathbf{H} = (h_{ij})$, for all $i = 1, \cdots, n$. Given that the $\widehat{\boldsymbol{\epsilon}}_i$ are linearly dependent, we define the index $I = \{i_1, \cdots, i_k\}$, with $i_s = 1, \cdots, n$; $s = 1, \cdots, k$ and $k \leq (n-\alpha)$, such that the vectors $\widehat{\boldsymbol{\epsilon}}_{i_1}, \cdots, \widehat{\boldsymbol{\epsilon}}_{i_k}$ are linearly independent. Thus we define the matrix

$$\widehat{\boldsymbol{\epsilon}}_I = \begin{pmatrix} \widehat{\boldsymbol{\epsilon}}_{i_1}^{\,T} \\ \vdots \\ \widehat{\boldsymbol{\epsilon}}_{i_k}^{\,T} \end{pmatrix} \tag{4}$$

and observe that $\widehat{\boldsymbol{\epsilon}}_I$ has a matrix-variate normal nonsingular distribution, moreover $\widehat{\boldsymbol{\epsilon}}_I \sim \mathcal{N}_{k\times p}(\mathbf{0}, (\mathbf{I}_k - \mathbf{H}_I) \otimes \boldsymbol{\Sigma})$. $\mathbf{H}_I$ is obtained from the matrix $\mathbf{H}$ by deleting the row and the column associated to the index I.

For a univariate model, i.e. when $p = 1$, different classes of residuals have been proposed, see Chatterjee and Hadi (1988):

$$\begin{aligned} a_i &= \frac{\widehat{\epsilon}_i}{\|\widehat{\boldsymbol{\epsilon}}\|} && \text{normalised residual} \\ b_i &= \frac{\widehat{\epsilon}_i}{\widehat{\sigma}} && \text{standardised residual} \\ r_i &= \frac{\widehat{\epsilon}_i}{\widehat{\sigma}\sqrt{1-h_{ii}}} && \text{internally studentised residual} \\ u_i &= \frac{\widehat{\epsilon}_i}{\widehat{\sigma}_{(i)}\sqrt{1-h_{ii}}} && \text{externally studentised residual} \end{aligned}$$

where $\|\mathbf{y}\|$ is the Euclidean norm of the vector $\mathbf{y}$; $\widehat{\sigma}^2 = \|\boldsymbol{\epsilon}\|^2/(n-\alpha)$; and $\widehat{\sigma}_{(i)}$ is the estimated standard deviation. Here, $\widehat{\sigma}_{(i)}$ is obtained by removing the i-th observation from the sample.

Analogously to the definition of $\widehat{\boldsymbol{\epsilon}}_I$, establish

$$\mathbf{r}_I^{\tau} = \begin{pmatrix} \mathbf{r}_{i_1} \\ \vdots \\ \mathbf{r}_{i_k} \end{pmatrix} = \begin{pmatrix} \dfrac{\widehat{\boldsymbol{\epsilon}}_{i_1}}{\widehat{\sigma}\sqrt{1-h_{i_1 i_1}}} \\ \vdots \\ \dfrac{\widehat{\boldsymbol{\epsilon}}_{i_k}}{\widehat{\sigma}\sqrt{1-h_{i_k i_k}}} \end{pmatrix} = \frac{1}{\widehat{\sigma}}\mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I, \tag{5}$$

where $\mathbf{D}^{-1/2}$ is a diagonal matrix with elements $(1-h_{i_1 i_1})^{-1/2}, \cdots, (1-h_{i_k i_k})^{-1/2}$.

Moreover, note that $\mathbf{r}_I^{\tau}$ can be defined as (see Chatterjee and Hadi (1988, p. 190)),

$$\mathbf{r}_I = \frac{1}{\widehat{\sigma}}(\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I \tag{6}$$

and when $I = \{i\}$

$$\mathbf{r}_I = \mathbf{r}_i = \mathbf{r}_I^\tau \tag{7}$$

The joint externally studentised residuals $\mathbf{u}_I^\tau$ and $\mathbf{u}_I$ can be defined similarly. But in this case,

$$\mathbf{u}_I^\tau = \begin{pmatrix} \mathbf{u}_{i_1} \\ \vdots \\ \mathbf{u}_{i_k} \end{pmatrix} = \begin{pmatrix} \dfrac{\widehat{\boldsymbol{\epsilon}}_{i_1}}{\widehat{\sigma}_{(i_1)}\sqrt{1-h_{i_1 i_1}}} \\ \vdots \\ \dfrac{\widehat{\boldsymbol{\epsilon}}_{i_k}}{\widehat{\sigma}_{(i_k)}\sqrt{1-h_{i_k i_k}}} \end{pmatrix} = \mathbf{D}_{\widehat{\sigma}}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I, \tag{8}$$

and the diagonal matrix $\mathbf{D}_{\widehat{\sigma}}^{-1/2}$ has dependent elements $\dfrac{(1-h_{i_1 i_1})^{-1/2}}{\widehat{\sigma}_{(i_1)}}, \dots, \dfrac{(1-h_{i_k i_k})^{-1/2}}{\widehat{\sigma}_{(i_k)}}$; $\widehat{\sigma}_{(i_s)}$, $s = 1, 2, \dots, k$. Then, problems arise when we try to find their distributions. A similar situation occurs when $\mathbf{u}_I$ is defined. Alternative definitions, as a way to avoid such problems, have been proposed for $\mathbf{u}_I$ and $\mathbf{u}_I^\tau$, as follows:

$$\mathbf{u}_I^\tau = \frac{1}{\widehat{\sigma}_{(I)}}\mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I \quad \text{and} \quad \mathbf{u}_I = \frac{1}{\widehat{\sigma}_{(I)}}(\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I, \tag{9}$$

where $\widehat{\sigma}_{(I)}$ is the standard deviation computed by removing, from the sample, the corresponding observations associated to the indexes in I.

Once again, note that under any of the possible definitions of the internally studentised residuals,

$$\mathbf{u}_I = u_i = \mathbf{u}_I^\tau, \quad I = \{i\}. \tag{10}$$

Analogously, the normalised and standardised residuals can be defined as $\mathbf{a}_I = \widehat{\boldsymbol{\epsilon}}_I/\|\widehat{\boldsymbol{\epsilon}}\|$ and $\mathbf{b}_I = \widehat{\boldsymbol{\epsilon}}_I/\widehat{\sigma}$, respectively.

Multivariate versions ($p > 1$) for the internally and externally studentised residuals are

$$\mathbf{r}_i = \frac{1}{\sqrt{1-h_{ii}}}\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i \quad \text{and} \quad \mathbf{u}_i = \frac{1}{\sqrt{1-h_{ii}}}\widehat{\mathbf{\Sigma}}_{(i)}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i,$$

respectively, where $\widehat{\boldsymbol{\epsilon}}_i : p \times 1$ and $\mathbf{A}^{1/2}$ is the definite non-negative squared root of $\mathbf{A}$, such that $(\mathbf{A}^{1/2})^2 = \mathbf{A}$. Given the index I, the following definitions are established

$$\begin{array}{rclcrcl} \mathbf{r}_I^\tau & = & \mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\widehat{\mathbf{\Sigma}}^{-1/2} & \qquad & \mathbf{u}_I^\tau & = & \mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\widehat{\mathbf{\Sigma}}_{(I)}^{-1/2} \\ \mathbf{r}_I & = & (\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\widehat{\mathbf{\Sigma}}^{-1/2} & & \mathbf{u}_I & = & (\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\widehat{\mathbf{\Sigma}}_{(I)}^{-1/2}. \end{array}$$

The multivariate versions of expressions (7) and (10) are also true in such cases.

The study of all kinds of residual distributions is very important in different fields of statistics, especially in sensibility analysis (or regression diagnostics) and in linear models. The effect of a variable on a regression model is usually studied by different kinds of graphic representations of residuals, Chatterjee and Hadi (1988, Section 3.8). Similarly, the effect of one or more observations on the parameters of a regression model is evaluated or measured by different measures or distances such as: Cook, Welsch or modified Cook distances, among many others. These measures can be expressed as functions of internally and externally studentised residuals. In the same way, other diagnostic measures based on volumes of ellipsoids of confidence or quotient of variances can also be expressed as a function of internally and externally studentised residuals, see Chatterjee and Hadi (1988, chapters 4 and 5) or Besley et al. (1980, Chapter 2). Unfortunately, the distributions of many of these measures are unknown, which means that decisions must be taken on the basis of a graphic representation and/or a list of values derived by computing the above-cited metrics.

Many researchers have avoided the problem of finding the joint distributions of different classes of residuals because they are singular, i.e. singular distributions do not exist with respect to the Lebesgue measure in $\mathbb{R}^n$. The problem is overcome by observing that singular distributions exist with respect to the Hausdorff measure defined over an affine subspace, see Díaz-García et al. (1997) and Díaz-García and González-Farías (2004). However, when other kinds of residuals are obtained under transformations of the singular distribution, then the Jacobians with respect to the Hausdorff measure shall be required; such problems are currently being investigated, Díaz-García and González-Farías (2004).An alternative approach was adopted by Ellenberg (1973), who proposed studying the distribution $\widehat{\boldsymbol{\epsilon}}_I$ defined by (4) and which already has a non-singular distribution, i. e. the distribution exists with respect to the Lebesgue measure in $\mathbb{R}^k$. Now it is possible to define the remaining classes of residuals, for the univariate and multivariate cases: we start with $\widehat{\boldsymbol{\epsilon}}_I$, and then determine their densities, which are non-singular under the hypothesis of the model (1).

The distribution of $\mathbf{r}_I^\tau$ was studied by Ellenberg (1973) (where the distribution of $\mathbf{r}_i$ is a particular case), and Beckman and Trusell (1974) studied the distribution of $\mathbf{u}_i$. These two results were found for the univariate general linear model of full rank ($\alpha = q$) and the three results are summarised in Chatterjee and Hadi (1988, see theorems 4.1 and 4.2, pp. 76-79).

The diagnostic problem for one observation in the multivariate case was studied by Caroni (1987), who determined the distributions of proportional amount to the Euclidean norm of $\mathbf{r}_i$ and $\mathbf{u}_i$. For more than one observation, the problem was addressed in Díaz-García and González-Farías (2004) by determining the distribution of matrices proportional to the $\mathbf{r}_I^T\mathbf{r}_I$ and $\mathbf{u}_I^T\mathbf{u}_I$ matrices.

The present paper starts by proposing a straightforward way for finding the distribution of r_i. The theory presented here avoids the usual methods: first, the general distribution given in Ellenberg (1973) is computed; second, a particular distribution is derived from the general one; third, the particular distribution is extended to a less than full rank model. Here, the distributions of r_i are used for deriving the densities of a_i and b_i. At the end of Section 2 we give a different proof to that of Beckman and Trusell (1974) for the density of $\mathbf{u}_i$ under models of full and less than full rank. The distributions of $\mathbf{r}_I^\tau$, $\mathbf{r}_I$, $\mathbf{u}_I^\tau$ and $\mathbf{u}_I$ are found for full and less than full rank models in the univariate case. The extensions to the multivariate case are treated in Section 4. An important application in econometric theory is presented in Section 5. In Section 6, we extend all the preceding results under the assumption that the distributions of the errors are elliptical. The paper ends with a list of conclusions in Section 7.

2. UNIVARIATE RESIDUAL

In the classical procedure for finding the distribution of $\mathbf{r}_i$, we first need to determine the distribution of $\mathbf{r}_I^\tau$, see Ellenberg (1973). In this section, we present a straightforward method for finding the distribution of $\mathbf{r}_i$ by considering a parallel approach to that of Cramér (1999, pp. 240-241). The result is established for the case of a less than full rank model; for the full rank model, the same result is easily obtained by taking $\alpha = q$ below. First, let us consider the following definition, Gupta and Varga (see 1993, p. 76), Dickey (1967) and Press (1982, pp. 138-141):

Definition 1. The $p \times n$ random matrix $\mathbf{X}$

i) is said to have a matrix variate symmetric Pearson type II distribution (also called inverted matrix variate $t-$distribution) with parameters $q \in \mathbb{R}$, $\mathbf{M} : p \times n$, $\mathbf{\Sigma} : p \times p$, $\Phi : n \times n$, with $q > -1$, $\mathbf{\Sigma} > 0$, and $\Phi > 0$ if its probability density function is

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{\Gamma\left[\dfrac{pn}{2} + q + 1\right]}{\pi^{pn/2}\Gamma[q+1]|\mathbf{\Sigma}|^{n/2}|\Phi|^{p/2}} \left(1 - \operatorname{tr}\left((\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\Phi^{-1}\right)\right)^q$$

where $\operatorname{tr}\left((\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\Phi^{-1}\right) \leq 1$, and it is denoted by $\mathbf{X} \sim \mathcal{P}II_{p\times n}(q, \mathbf{M}, \mathbf{\Sigma} \otimes \Phi)$.

ii) is said to have a matrix variate $t-$distribution with parameters $r \in \mathbb{R}$, $\mathbf{M} : p \times n$, $\mathbf{\Sigma} : p \times p$, $\Phi : n \times n$, with $r > 0$, $\mathbf{\Sigma} > 0$, and $\Phi > 0$ if its probability density function is

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{\Gamma[(pn+r)/2]}{(\pi r)^{pn/2}\Gamma[r/2]|\mathbf{\Sigma}|^{n/2}|\Phi|^{p/2}} \left(1 + \frac{\operatorname{tr}\left((\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\Phi^{-1}\right)}{r}\right)^{-(pn+r)/2}$$

and it is denoted by $\mathbf{X} \sim \mathcal{M}\mathbf{t}_{p\times n}(r, \mathbf{M}, \mathbf{\Sigma} \otimes \Phi)$ or by $\mathbf{X} \sim \mathbf{t}_p(r, \mathbf{M}, \mathbf{\Sigma})$ when $n = 1$.

iii) is said to have a matric-variate symmetric Pearson type II distribution (also called inverted matrix $T-$distribution) with parameters $q \in \mathbb{R}$, $\mathbf{M} : p \times n$, $\mathbf{\Sigma} : p \times p$, $\Phi : n \times n$, with $q > -1$, $\mathbf{\Sigma} > 0$, and $\Phi > 0$ if its probability density function is

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{\Gamma_n[q/2]}{\pi^{pn/2}\Gamma_n[(q-p)/2]|\mathbf{\Sigma}|^{n/2}|\Phi|^{p/2}} \left| \mathbf{I}_n - (\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\Phi^{-1}\right|^{-(q-p-n-1)/2}$$

where $\left(\mathbf{I}_n - (\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\Phi^{-1}\right) > 0$, and it is denoted by $\mathbf{X} \sim \mathcal{MP}II_{p\times n}(q, \mathbf{M}, \mathbf{\Sigma} \otimes \Phi)$.

iv) is said to have a matric-variate $T-$distribution with parameters $r \in \mathbb{R}$, $\mathbf{M} : p \times n$, $\mathbf{\Sigma} : p \times p$, $\Phi : n \times n$, with $r > 0$, $\mathbf{\Sigma} > 0$, and $\Phi > 0$ if its probability density function is

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{\Gamma_n[r/2]}{\pi^{pn/2}\Gamma_n[(r-p)/2]|\mathbf{\Sigma}|^{n/2}|\Phi|^{p/2}} \left| \mathbf{I}_n + (\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\Phi^{-1}\right|^{-r/2}$$

and it is denoted by $\mathbf{X} \sim \mathcal{MT}_{p\times n}(r, \mathbf{M}, \mathbf{\Sigma} \otimes \Phi)$.

where $\Gamma_n[a]$ denoted the multivariate gamma function, $\Gamma_n[a] = \pi^{n(n-1)/4}\prod_{i=1}^{n}\Gamma(a-(1-i)/2)$, see Muirhead (1982, p. 61).

Theorem 1 (Internally studentised residual). *Under the model (1) with $p = 1$ (univariate case), $\mathbf{r}_i$ has a Pearson Type II distribution, $r_i \sim \mathcal{P}II((n-\alpha-3)/2, 0, n-\alpha)$. Thus, its density function is given by*

$$g_{r_i}(r_i) = \frac{\Gamma[(n-\alpha)/2]}{\sqrt{\pi(n-\alpha)}\ \Gamma[(n-\alpha-1)/2]} \left(1 - \frac{r_i^2}{(n-\alpha)}\right)^{(n-\alpha-3)/2}, \quad |r_i| \leq \sqrt{(n-\alpha)}.$$

Proof: Define $\theta_i = \sqrt{1-h_{ii}}r_i$, note that

$$\theta_i = \sqrt{1-h_{ii}}\ \frac{\widehat{\epsilon}_i}{\widehat{\sigma}\sqrt{1-h_{ii}}} = \frac{\widehat{\epsilon}_i}{\widehat{\sigma}} = \sqrt{n-\alpha}\ \frac{\widehat{\epsilon}_i}{\|\widehat{\boldsymbol{\epsilon}}\|} \tag{11}$$

and observe that $\theta_i^2 = (n-\alpha)\widehat{\epsilon}_i{}^2/\|\widehat{\boldsymbol{\epsilon}}\|^2$ with $\widehat{\epsilon}_i{}^2 > 0$ and $\|\widehat{\boldsymbol{\epsilon}}\|^2 > 0$. Besides, $\|\widehat{\boldsymbol{\epsilon}}\|^2 \geq \widehat{\epsilon}_i{}^2$. Then

$$0 \leq \frac{\widehat{\epsilon}_i{}^2}{\|\widehat{\boldsymbol{\epsilon}}\|^2} \leq 1$$

Thus $\theta_i^2 \leq (n-\alpha)$ or equivalently $|\theta_i| \leq \sqrt{n-\alpha}$. This means that the density function of θ_i is zero outside the interval $[-\sqrt{n-\alpha}, \sqrt{n-\alpha}\ ]$. Now define

$$v_i = \sqrt{\frac{n-\alpha-1}{(n-\alpha)(1-h_{ii})}}\frac{\theta_i}{\sqrt{1-\dfrac{\theta_i^2}{n-\alpha}}}$$

so, by (11)

$$v_i = \frac{\widehat{\epsilon}_i}{\sqrt{\dfrac{(1-h_{ii})}{(n-\alpha-1)}\displaystyle\sum_{j\neq i}^{n}\widehat{\epsilon}_j{}^2}}. \tag{12}$$

Now, note that $\widehat{\epsilon}_i/\sqrt{\sigma^2(1-h_{ii})} \sim \mathcal{N}(0,1)$ is independent of $\displaystyle\sum_{j\neq i}^{n}\widehat{\epsilon}_j{}^2/\sigma^2 \sim \chi^2(n-\alpha-1)$, where $\chi^2(m)$ denotes the central chi-squared distribution with m degrees of freedom. Then

$$\frac{\dfrac{\widehat{\epsilon}_i}{\sigma\sqrt{(1-h_{ii})}}}{\sqrt{\dfrac{1}{\sigma^2(n-\alpha-1)}\displaystyle\sum_{j\neq i}^{n}\widehat{\epsilon}_j{}^2}} = \frac{\widehat{\epsilon}_i}{\sqrt{\dfrac{(1-h_{ii})}{(n-\alpha-1)}\displaystyle\sum_{j\neq i}^{n}\widehat{\epsilon}_j{}^2}} = v_i \sim t(n-\alpha-1) \tag{13}$$

Here, $t(m)$ denotes the one-dimensional central distribution t with m degrees of freedom. Also, note that if θ_i varies in the interval $[-\sqrt{n-\alpha}, \sqrt{n-\alpha}\ ]$, v_i then it takes values in the interval $(-\infty, \infty)$. Thus

$$v_i \leq \sqrt{\frac{n-\alpha-1}{(n-\alpha)(1-h_{ii})}}\frac{x}{\sqrt{1-\dfrac{x^2}{n-\alpha}}}$$

is equivalent to, $\theta_i \leq x$. So

$$P(\theta_i \leq x) = P\left(v_i \leq \sqrt{\frac{n-\alpha-1}{(n-\alpha)(1-h_{ii})}}\frac{x}{\sqrt{1-\dfrac{x^2}{n-\alpha}}}\right) = \int_{-\infty}^{\sqrt{\frac{n-\alpha-1}{(n-\alpha)(1-h_{ii})}}\frac{x}{\sqrt{1-\frac{x^2}{n-\alpha}}}} t(v_i; n-\alpha-1)dv_i$$

where $t(y;m)$ denotes the density function of a random variable x with t distribution and m degrees of freedom. But $\theta_i = \sqrt{1-h_{ii}}r_i$, thus

$$P\left(r_i \leq \frac{x}{\sqrt{1-h_{ii}}}\right) = P\left(v_i \leq \frac{1}{\sqrt{1-h_{ii}}}\sqrt{\frac{n-\alpha-1}{(n-\alpha)(1-h_{ii})}}\frac{x}{\sqrt{1-\dfrac{x^2}{n-\alpha}}}\right)$$

Taking the derivative with respect to x gives

$$r_i \sim \sqrt{\frac{n-\alpha-1}{n-\alpha}}\left(1-\frac{r_i^2}{n-\alpha}\right)^{-3/2} t\left(\sqrt{\frac{n-\alpha-1}{(n-\alpha)(1-h_{ii})}}\frac{r_i}{\sqrt{1-\dfrac{r_i^2}{n-\alpha}}}; n-\alpha-1\right), \quad |r_i| \leq \sqrt{n-\alpha}.$$

And then the desired result is obtained. ∎

Note that the distributions of the normalised and studentised residuals, a_i and b_i, respectively, are easily found; both residuals can be expressed as a function of $\mathbf{r}_i$, in the following way;

$$a_i = \sqrt{\frac{(1-h_{ii})}{(n-\alpha)}}\ r_i \quad \text{and} \quad b_i = \sqrt{1-h_{ii}}\ r_i.$$

Thus, from Gupta and Varga (1993, Theorem 2.1.2, p. 20);

Corollary 1 (Normalised and standardised residuals). *The distributions of the normalised and studentised residuals are given by*

i) $a_i \sim \mathcal{P}II((n-\alpha-3)/2, 0, (1-h_{ii})), \quad |a_i| \leq \sqrt{(1-h_{ii})}$

ii) $b_i \sim \mathcal{P}II((n-\alpha-3)/2, 0, (n-\alpha)(1-h_{ii})), \quad |b_i| \leq \sqrt{(n-\alpha)(1-h_{ii})},$

respectively.

Finally, an alternative proof to Beckman and Trusell (1974) and Chatterjee and Hadi (1988, p. 78) is given to determine the distribution of the externally studentised residuals under a full and less than full rank.

Theorem 2 (Externally studentised residual). *Under the model (1) with $p = 1$ univariate case),* $\mathbf{u}_i$ *has a t distribution with $(n-\alpha-1)$ degrees of freedom.*

Proof: The demonstration follows from (13), noting that $(n-\alpha-1)\widehat{\sigma}^2_{(i)} = \sum_{j\neq i}^{n} \widehat{\epsilon}_j^{\,2}$, thus

$$v_i = \frac{\widehat{\epsilon}_i}{\sqrt{\dfrac{(1-h_{ii})}{(n-\alpha-1)}\displaystyle\sum_{j\neq i}^{n}\widehat{\epsilon}_j^{\,2}}} = \frac{\widehat{\epsilon}_i}{\sqrt{(1-h_{ii})\widehat{\sigma}^2_{(i)}}} = u_i.$$

∎

3. JOINT UNIVARIATE RESIDUAL

In this section, the joint distributions of the externally and internally studentised residuals shall be studied under different definitions. The first theorem gives the less than full rank model version for the result of Ellenberg (1973).

Theorem 3 (Internally studentised residual, I). *If we consider the univariate model (1), then,* $\mathbf{r}_I^\tau \sim \mathcal{P}II_k\left(\frac{(n-\alpha-k)}{2}-1, \mathbf{0}, (n-\alpha)\mathbf{V}\right)$*, with* $\mathbf{V} = \mathbf{D}^{-1/2}(\mathbf{I}_k - \mathbf{H}_I)\mathbf{D}^{-1/2}$*, i.e.*

$$f_{\mathbf{r}_I^\tau}(\mathbf{r}_I^\tau) = \frac{\Gamma[(n-\alpha)/2]}{\Gamma[(n-\alpha-k)/2][\pi(n-\alpha)]^{k/2}|\mathbf{V}|^{1/2}}\left(1-\frac{\mathbf{r}_I^{\tau T}\mathbf{V}^{-1}\mathbf{r}_I^\tau}{(n-\alpha)}\right)^{(n-\alpha-k)/2-1}, \quad \mathbf{r}_I^{\tau T}\mathbf{V}^{-1}\mathbf{r}_I^\tau \leq (n-\alpha).$$

Proof: The demonstration reduces to that given in Ellenberg (1973), taking α instead of k (use Ellenberg's notation). ■

Theorem 4 (Internally studentised residual, II). *Consider model (1) with* $p = 1$*. Then,* $\mathbf{r}_I \sim \mathcal{P}II_k\left(\frac{(n-\alpha-k)}{2}-1, \mathbf{0}, (n-\alpha)\mathbf{I}_k\right)$*, i.e.*

$$f_{\mathbf{r}_I}(\mathbf{r}_I^\tau) = \frac{\Gamma[(n-\alpha)/2]}{\Gamma[(n-\alpha-k)/2][\pi(n-\alpha)]^{k/2}}\left(1-\frac{||\mathbf{r}_I||^2}{(n-\alpha)}\right)^{(n-\alpha-k)/2-1}, \quad ||\mathbf{r}_I||^2 \leq (n-\alpha).$$

Proof: The technique used is analogous to the given in 3. Alternatively, note that

$$\mathbf{r}_I^\tau = \frac{\mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I}{\widehat{\sigma}} = \frac{\mathbf{D}^{-1/2}(\mathbf{I}_k-\mathbf{H}_I)^{1/2}(\mathbf{I}_k-\mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I}{\widehat{\sigma}} = \mathbf{D}^{-1/2}(\mathbf{I}_k-\mathbf{H}_I)^{1/2}\mathbf{r}_I,$$

with $(d\mathbf{r}_I^\tau) = |\mathbf{D}^{-1}(\mathbf{I}_k - \mathbf{H}_I)|^{1/2}(d\mathbf{r}_I)$; where $(d\mathbf{X})$ denotes the exterior product of the elements of the vector (or matrix) of differentials $d\mathbf{X}$, Muirhead (1982, p. 57). Thus, by applying this transformation to the density in Theorem 3, the desired result is obtained. ■

With the aim of determining the joint distribution of the normalised and standardised residuals $\mathbf{a}_I$ and $\mathbf{b}_I$, respectively, note that for the first of these

$$\mathbf{a}_I = \frac{\widehat{\boldsymbol{\epsilon}}_I}{||\widehat{\boldsymbol{\epsilon}}||} = \begin{cases} \dfrac{\mathbf{D}^{1/2}}{\sqrt{n-\alpha}}\,\mathbf{r}_I^\tau & \text{with} \quad (d\mathbf{r}_I^\tau) = (n-\alpha)^{k/2}|\mathbf{D}|^{-1/2}(d\mathbf{a}_I) \\ \dfrac{(\mathbf{I}_k-\mathbf{H}_I)^{1/2}}{\sqrt{n-\alpha}}\,\mathbf{r}_I & \text{with} \quad (d\mathbf{r}_I) = (n-\alpha)^{k/2}|\mathbf{I}_k-\mathbf{H}_I|^{-1/2}(d\mathbf{a}_I). \end{cases} \tag{14}$$

Similarly, observe that

$$\mathbf{b}_I = \frac{\widehat{\boldsymbol{\epsilon}}_I}{\widehat{\sigma}} = \begin{cases} \mathbf{D}^{1/2}\,\mathbf{r}_I^\tau & \text{with} \quad (d\mathbf{r}_I^\tau) = |\mathbf{D}|^{-1/2}(d\mathbf{b}_I) \\ (\mathbf{I}_k-\mathbf{H}_I)^{1/2}\,\mathbf{r}_I & \text{with} \quad (d\mathbf{r}_I) = |\mathbf{I}_k-\mathbf{H}_I|^{-1/2}(d\mathbf{b}_I). \end{cases} \tag{15}$$

Then, considering transformations (14) and (15) and their corresponding Jacobians, the distributions of $\mathbf{a}_I$ and $\mathbf{b}_I$ are straightforwardly derived from Theorems 3 and 4. In summary:

Theorem 5 (Normalised and standardised residuals). *For the univariate model (1), we have,*

i). $\mathbf{a}_I \sim \mathcal{P}II_k\left(\frac{(n-\alpha-k)}{2}-1,\ \mathbf{0}, (\mathbf{I}_k-\mathbf{H}_I)\right)$

ii). $\mathbf{b}_I \sim \mathcal{P}II_k\left(\frac{(n-\alpha-k)}{2}-1,\ \mathbf{0}, (n-\alpha)(\mathbf{I}_k-\mathbf{H}_I)\right)$.

For the case of externally studentised residuals, observe that, see Ellenberg (1973, Lemma 2)

$$\widehat{\boldsymbol{\epsilon}}_I \sim \mathcal{N}_k(\mathbf{0}, \sigma^2(\mathbf{I}_k - \mathbf{H}_I)) \quad \text{independently of} \quad \frac{(n-\alpha-k)\sigma^2_{(I)}}{\sigma^2} \sim \chi^2(n-\alpha-k)$$

where $\sigma^2_{(I)}$ is given in (9). Then,

$$\frac{\dfrac{(I-\mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I}{\sigma}}{\sqrt{\dfrac{(n-\alpha-k)\widehat{\sigma}^2_{(I)}}{(n-\alpha-k)\sigma^2}}} = \frac{(I-\mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I}{\widehat{\sigma}_{(I)}} = \mathbf{u}_I \sim \mathbf{t}_k((n-\alpha-k), \mathbf{0}, \mathbf{I}_k).$$

by equation (1.2) in Kotz and Nadarajah (2004, p. 2).

Also, observe that

$$\mathbf{u}_I = \frac{(\mathbf{I}_k-\mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I}{\widehat{\sigma}_{(I)}} = \frac{(\mathbf{I}_k-\mathbf{H}_I)^{-1/2}\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I}{\widehat{\sigma}_{(I)}} = (I-\mathbf{H}_I)^{-1/2}\mathbf{D}^{1/2}\mathbf{u}^\tau_I,$$

from which, the following result can be established

Theorem 6 (Externally studentised residual). *Under the general univariate linear model (1) we have*

i) $\mathbf{u}_I \sim \mathbf{t}_k((n-\alpha-k), \mathbf{0}, \mathbf{I}_k)$

ii) $\mathbf{u}^\tau_I \sim \mathbf{t}_k((n-\alpha-k), \mathbf{0}, \mathbf{V})$, *with* $\mathbf{V} = \mathbf{D}^{-1/2}(I-\mathbf{H}_I)\mathbf{D}^{-1/2}$.

4. JOINT MULTIVARIATE RESIDUAL

In the multivariate case, the distributions of $\mathbf{r}_i$ and $\mathbf{u}_i$ are difficult to find. For example, for the externally studentised residual $\mathbf{u}_i = \dfrac{1}{\sqrt{(1-h_{ii})}}\widehat{\mathbf{\Sigma}}^{-1/2}_{(i)}\widehat{\epsilon}_i$ with $\dfrac{\widehat{\boldsymbol{\epsilon}}_i}{\sqrt{(1-h_{ii})}} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ independent $(n-\alpha-1)\widehat{\mathbf{\Sigma}}_{(i)} \sim \mathcal{W}_p(n-\alpha-1, \mathbf{\Sigma})$, note that the distributions of $\mathbf{u}_i$ cannot be found unless, as Ellenberg (1973) assumes, $\mathbf{\Sigma} = \mathbf{I}_p$ is taken in the two above distributions, by which we obtain $\mathbf{u}_i \sim \mathbf{t}_p\left((n-p-\alpha), \mathbf{0}, \dfrac{(n-p-\alpha)}{(n-\alpha-1)}\mathbf{I}_p\right)$. For applying the general definition of the multidimensional t distribution, it is required that $\mathbf{\Sigma}$ be proportional to $\mathbf{I}_p$ in the distribution of $\mathbf{u}_i$, see Kotz and Nadarajah (2004, p. 7).

Instead, we find the distributions of $\mathbf{r}_I$ and $\mathbf{u}_I$. Then, as corollaries, $I = \{i\}$, the distributions of $\mathbf{r}_i$ and $\mathbf{u}_i$ are found.

For extending Theorem 3 to its multivariate version, consider the following result.

Lemma 1. *If* $\Re e(a) > (p-1)/2$ *and* $\mathbf{\Theta}$ *is a symmetric* $p \times p$ *matrix with* $\Re e(\mathbf{\Theta}) > 0$ *then*

$$\int_{\mathbf{R}>0} \operatorname{etr}\left(\tfrac{1}{2}\mathbf{\Theta}^{-1}\mathbf{R}^2\right)|\mathbf{R}|^{2a-p}\prod_{i<j}(\lambda_i+\lambda_j)(d\mathbf{R}) = \Gamma_p[a]|\mathbf{\Theta}|^a 2^{p(a-1)}$$

where $\lambda_i, \quad i = 1, \dots, p$ *are the eigenvalues of the matrix* $\mathbf{R} : p \times p$.

Proof: From Theorem 2.1.11 in Muirhead (1982, p. 61) we know that for $\Re e(a) > (p-1)/2$ and $\mathbf{\Theta}$ is a symmetric $p \times p$ matrix with $\Re e(\mathbf{\Theta}) > 0$

$$\int_{\mathbf{B}>0} \operatorname{etr}\left(\tfrac{1}{2}\mathbf{\Theta}^{-1}\mathbf{B}\right) |\mathbf{B}|^{a-(p+1)/2}(d\mathbf{B}) = \Gamma_p[a]|\mathbf{\Theta}|^a 2^{ap}$$

the result follows taking $\mathbf{R}$ such that $(\mathbf{R})^2 = \mathbf{B}$ with $(d\mathbf{B}) = 2^p|\mathbf{R}|\prod_{i<j}(\lambda_i+\lambda_j)(d\mathbf{R})$, where $\lambda_i, \quad i = 1,\dots,p$ are the eigenvalues of the matrix $\mathbf{R}$, see Díaz-García and González-Farías (2004), Olkin and Rubin (1964) or Magnus (1988, p. 128). ∎

Theorem 7 (Internally studentised residual, II). *Under model (1),* $\mathbf{r}_I$ *has a matric-variate symmetric Pearson Type II distribution,* $\mathbf{r}_I \sim \mathcal{MPII}_{k\times p}((n-\alpha), \mathbf{0}, (n-\alpha)(\mathbf{I}_k \otimes \mathbf{I}_p))$*, moreover, its density function is*

$$g_{\mathbf{r}_I}(\mathbf{r}_I) = \frac{\Gamma_p[(n-\alpha)/2]}{(\pi(n-\alpha))^{kp/2}\ \Gamma_p[(n-\alpha-k)/2]}\left|\ \mathbf{I}_p - \frac{1}{(n-\alpha)}\mathbf{r}_I^T\mathbf{r}_I\right|^{(n-\alpha-k-p-1)/2}, \quad \left|\ \mathbf{I}_p - \frac{1}{(n-\alpha)}\mathbf{r}_I^T\mathbf{r}_I\right| > 0$$

Proof: Suppose $(n-\alpha)\widehat{\mathbf{\Sigma}} = \mathbf{A} = \widehat{\boldsymbol{\epsilon}}^T\widehat{\boldsymbol{\epsilon}}$, Then, generalising Lemma 1 in Ellenberg (1973), $(n-\alpha)\widehat{\mathbf{\Sigma}} = (n-\alpha-k)\widehat{\mathbf{\Sigma}}_{(I)} + \widehat{\boldsymbol{\epsilon}}_I(\mathbf{I}_k - \mathbf{H}_I)^{-1}\widehat{\boldsymbol{\epsilon}}_I$, where $\mathbf{A}_I = (n-\alpha-k)\widehat{\mathbf{\Sigma}}_{(I)} = \widehat{\boldsymbol{\epsilon}}_I\widehat{\boldsymbol{\epsilon}}_I \sim \mathcal{W}_p((n-\alpha-k), \mathbf{\Sigma})$ independently of $\widehat{\boldsymbol{\epsilon}}_I$, see Lemma 2 in Ellenberg (1973), then, denoting $m = n-\alpha-k$

$$f_{\widehat{\boldsymbol{\epsilon}}_I,\mathbf{A}_I}(\widehat{\boldsymbol{\epsilon}}_I, \mathbf{A}_I) = \frac{|\mathbf{A}_I|^{(m-p-1)/2}}{(2\pi)^{kp/2}2^{pm/2}\Gamma_p[m/2]|\mathbf{I}_k - \mathbf{H}_I|^{p/2}|\mathbf{\Sigma}|^{(n-\alpha)/2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Sigma}^{-1}(\mathbf{A}_I + \widehat{\boldsymbol{\epsilon}}_I(\mathbf{I}_k - \mathbf{H}_I)^{-1}\widehat{\boldsymbol{\epsilon}}_I)\right) \tag{16}$$

Now define $\mathbf{r}_I = (\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\widehat{\mathbf{\Sigma}}^{-1/2}$ and note that $(n-\alpha)\widehat{\mathbf{\Sigma}} = \mathbf{A}_I + \widehat{\boldsymbol{\epsilon}}_I(\mathbf{I}_k - \mathbf{H}_I)^{-1}\widehat{\boldsymbol{\epsilon}}_I$, so

$$\widehat{\boldsymbol{\epsilon}}_I = (\mathbf{I}_k - \mathbf{H}_I)^{1/2}\mathbf{r}_I\widehat{\mathbf{\Sigma}}^{1/2} \quad \text{and} \quad \mathbf{A}_I = (n-\alpha)\widehat{\mathbf{\Sigma}} + \widehat{\boldsymbol{\epsilon}}_I(\mathbf{I}_k - \mathbf{H}_I)^{-1}\widehat{\boldsymbol{\epsilon}}_I$$

which implies

$$(d\widehat{\boldsymbol{\epsilon}}_I)(d\mathbf{A}_I) = (n-\alpha)^{p(p+1)/2}|\mathbf{I}_k - \mathbf{H}_I|^{p/2}|\widehat{\mathbf{\Sigma}}|^{k/2}(d\mathbf{r}_I)(d\widehat{\mathbf{\Sigma}}).$$

Besides, observe that for $\mathbf{S} : p \times p$, such that $\mathbf{S} = \mathbf{R}^2 > 0$, $(d\mathbf{S}) = 2^p|\mathbf{R}|\prod_{i\neq j}^{p}(\lambda_i - \lambda_j)(d\mathbf{R})$, with λ_i the eigenvalues of $\mathbf{R}$. Thus

$$(d\widehat{\boldsymbol{\epsilon}}_I)(d\mathbf{A}_I) = 2^p(n-\alpha)^{p(p+1)/2}|\mathbf{I}_k - \mathbf{H}_I|^{p/2}\left|\widehat{\mathbf{\Sigma}}^{1/2}\right|^{k+1}\prod_{i\neq j}^{p}(\lambda_i - \lambda_j)(d\mathbf{r}_I)\left(d\widehat{\mathbf{\Sigma}}^{1/2}\right). \tag{17}$$

Substituting (17) in (16) and simplifying, we obtain

$$f_{\mathbf{r}_I,\widehat{\mathbf{\Sigma}}^{1/2}}(\mathbf{r}_I, \widehat{\mathbf{\Sigma}}^{1/2}) = \frac{\left|\ \mathbf{I}_p - \dfrac{\widehat{\mathbf{r}}_I^T\widehat{\mathbf{r}}_I}{(n-\alpha)}\right|^{(m-p-1)/2}\left|\ \widehat{\mathbf{\Sigma}}^{1/2}\right|^{n-\alpha-p}\operatorname{etr}\left(-\dfrac{(n-\alpha)}{2}\mathbf{\Sigma}^{-1}\left(\widehat{\mathbf{\Sigma}}^{1/2}\right)^2\right)\prod_{i\neq j}^{p}(\lambda_i - \lambda_j)}{(2\pi)^{kp/2}2^{pm/2-p}\Gamma_p[m/2]|\mathbf{\Sigma}|^{(n-\alpha)/2}(n-\alpha)^{-pm/2}}$$

Because

$$\begin{aligned}\mathbf{A}_I &= (n-\alpha)\widehat{\mathbf{\Sigma}} - \widehat{\boldsymbol{\epsilon}}_I(\mathbf{I}_k - \mathbf{H}_I)^{-1}\widehat{\boldsymbol{\epsilon}}_I^T \\ &= (n-\alpha)\widehat{\mathbf{\Sigma}} - \widehat{\mathbf{\Sigma}}^{1/2}\widehat{\mathbf{r}}_I(\mathbf{I}_k - \mathbf{H}_I)^{1/2}(\mathbf{I}_k - \mathbf{H}_I)^{-1}(\mathbf{I}_k - \mathbf{H}_I)^{1/2}\widehat{\mathbf{r}}_I \\ &= (n-\alpha)\widehat{\mathbf{\Sigma}}^{1/2}\left(\mathbf{I}_p - \frac{1}{(n-\alpha)}\widehat{\mathbf{r}}_I^T\widehat{\mathbf{r}}_I\right)\widehat{\mathbf{\Sigma}}^{1/2},\end{aligned}$$

with $|\mathbf{A}_I| = \left|\mathbf{I}_p - \frac{1}{(n-\alpha)}\widehat{\mathbf{r}}_I^T\widehat{\mathbf{r}}_I\right||(n-\alpha)\widehat{\mathbf{\Sigma}}|$. Integrating with respect to $\widehat{\mathbf{\Sigma}}^{1/2}$ using Lemma 1 the desired result is obtained. ■

We saw how Theorem 4 in the univariate case was derived from Theorem 7, just taking $p = 1$. Similarly, we straightforwardly obtain the marginal distribution of $\mathbf{r}_i$, starting from Theorem 7.

Corollary 2. *Setting $k = 1$ in Theorem 7, we have the marginal distribution of $\mathbf{r}_i$, and, the density of $\mathbf{r}_i$ is*

$$g_{\mathbf{r}_i}(r_I) = \frac{\Gamma[(n-\alpha)/2]}{(\pi(n-\alpha))^{p/2}\,\Gamma[(n-\alpha-p)/2]}\left(1 - \frac{1}{(n-\alpha)}||\mathbf{r}_i||^2\right)^{(n-\alpha-p-2)/2}, \quad ||\mathbf{r}_i||^2 < (n-\alpha)$$

this is, $\mathbf{r}_i \sim \mathcal{P}II_p\left(\frac{(n-\alpha-p-2)}{2}, \mathbf{0}, (n-\alpha)\mathbf{I}_p\right)$.

Proof: The demonstration is straightforward from Theorem 7, just noting that

$$\frac{\Gamma_p[(m+1)/2]}{\Gamma_p[m/2]} = \frac{\Gamma[(m+1)/2]}{\Gamma[(m+1-p)/2]} \tag{18}$$

with $m = n - \alpha - 1$. ■

Now, observing that

$$\mathbf{r}_I^\tau = \mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\widehat{\mathbf{\Sigma}}^{-1/2} = \mathbf{D}^{-1/2}(\mathbf{I}_k - \mathbf{H}_I)^{1/2}\mathbf{r}_I$$

with $(d\mathbf{r}_I) = |\mathbf{V}|^{p/2}(d\mathbf{r}_i^\tau)$ y $\mathbf{V} = \mathbf{D}^{-1/2}(\mathbf{I}_k - \mathbf{H}_I)\mathbf{D}^{-1/2}$, we get that

Theorem 8 (Internally studentised residual, I). *Under model (1), $\mathbf{r}_I^\tau$ has a matric-variate symmetric Pearson Type II distribution, $\mathbf{r}_I^\tau \sim \mathcal{MPII}_{k\times p}((n-\alpha), \mathbf{0}, (n-\alpha)(\mathbf{V}\otimes\mathbf{I}_p))$, con $\mathbf{V} = \mathbf{D}^{-1/2}(\mathbf{I}_k - \mathbf{H}_I)\mathbf{D}^{-1/2}$.*

Theorem 9 (Externally studentised residual). *Under the general multivariate linear model (1) we have that*

i) $\mathbf{u}_I \sim \mathcal{MT}_{k\times p}((n-\alpha), \mathbf{0}, (n-\alpha-k)(\mathbf{I}_k\otimes\mathbf{I}_p))$,

ii) $\mathbf{u}_I^\tau \sim \mathcal{MT}_{k\times p}((n-\alpha), \mathbf{0}, (n-\alpha-k)(\mathbf{V}\otimes\mathbf{I}_p))$, *with* $\mathbf{V} = \mathbf{D}^{-1/2}(\mathbf{I}_k - \mathbf{H}_I)\mathbf{D}^{-1/2}$.

Proof: The demonstration is parallel to the proof of Dickey (1967, Theorem 3.1), just note that $(\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I \sim \mathcal{N}_{k\times p}(0, I\otimes\mathbf{\Sigma})$ and that $(n-\alpha-k)\widehat{\mathbf{\Sigma}}_{(I)} \sim \mathcal{W}_p((n-\alpha-k), \mathbf{\Sigma})$. ■

By (10), and taking $k = 1$ in Theorem 9 we have:

Corollary 3. *Under the general multivariate linear model (1),*

$$\mathbf{u}_i \sim \mathbf{t}_p\left((n-p-\alpha), \mathbf{0}, \frac{(n-p-\alpha)}{(n-\alpha-1)}\mathbf{I}_p\right).$$

Remark 1. Generally, when the residuals are used for a sensibility analysis, it is traditional to take proportional amounts to $||\mathbf{r}_i||^2$ and $||\mathbf{u}_i||^2$ because their distributions are known, see Caroni (1987). In the multivariate case those results were extended: the distributions of proportional matrices to the matrices $\widehat{\mathbf{r}}_I{}^T\widehat{\mathbf{r}}_I$ $\widehat{\mathbf{u}}_I{}^T\widehat{\mathbf{u}}_I$ were found and several metrics associated to those matrices were determined, see Díaz-García and González-Farías (2004).

5. JOINT MULTIVARIATE RESIDUAL: SPECIAL CASE

In this section we consider the general multivariate linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{19}$$

but assuming that $\boldsymbol{\epsilon} \sim \mathcal{N}_{n\times p}(\mathbf{0}, \mathbf{I}_n \otimes \sigma^2\mathbf{W})$, with $\mathbf{W} > 0$ known. This model is very interesting for different fields of statistics, but especially in econometric methods, see Johnston (1972, Chapter 7).

For this model, the normal equations are given by $\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}\widetilde{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{W}^{-1}\mathbf{Y}$ so,

$$\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{Y} \quad \text{and} \quad \widehat{\boldsymbol{\epsilon}} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}$$

where in this case $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{W}^{-1}$. Moreover, $\widehat{\boldsymbol{\epsilon}} \sim \mathcal{N}_{n\times p}^{(n-\alpha),p}(\mathbf{0}, (\mathbf{I}_n - \mathbf{H}) \otimes \sigma^2\mathbf{W})$ and in particular, $\epsilon_i \sim \mathcal{N}_p(\mathbf{0}, \sigma^2(1-h_{ii})\mathbf{W})$. By vectorising model (1), see Muirhead (1982, p. 74), we find that

$$\begin{aligned} p(n-\alpha)\widehat{\sigma}^2 &= \left\|\left(\mathbf{W}^{-1/2} \otimes \mathbf{I}_n\right) \operatorname{vec} \widehat{\boldsymbol{\epsilon}}\, \right\|^2 \\ &= \operatorname{vec}^T \mathbf{Y} \left(\mathbf{W}^{-1} \otimes \mathbf{I}_n - \mathbf{W}^{-1} \otimes \mathbf{H}\right) \operatorname{vec} \mathbf{Y}, \end{aligned}$$

such that $\dfrac{p(n-\alpha)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{p(n-\alpha)}$. For this model we obtain that:

$$\begin{aligned} \mathbf{r}_I^\tau &= \frac{1}{\widehat{\sigma}}\mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\mathbf{W}^{-1/2} & \mathbf{u}_I^\tau &= \frac{1}{\widehat{\sigma}_{(I)}}\mathbf{D}^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\mathbf{W}^{-1/2} \\ \mathbf{r}_I &= \frac{1}{\widehat{\sigma}}(\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\mathbf{W}^{-1/2} & \mathbf{u}_I &= \frac{1}{\widehat{\sigma}_{(I)}}(\mathbf{I}_k - \mathbf{H}_I)^{-1/2}\widehat{\boldsymbol{\epsilon}}_I\mathbf{W}^{-1/2}. \end{aligned}$$

Theorem 10 (Internally studentised residual). *Under model (19) it is obtained*

i) $\mathbf{r}_I \sim \mathcal{P}II_{k\times p}\left(\dfrac{p(n-\alpha)-k}{2} - 1, \mathbf{0}, p(n-\alpha)(\mathbf{I}_k \otimes \mathbf{I}_p)\right)$

ii) $\mathbf{r}_I^\tau \sim \mathcal{P}II_{k\times p}\left(\dfrac{p(n-\alpha)}{2} - 1, \mathbf{0}, p(n-\alpha)(\mathbf{V} \otimes \mathbf{I}_p)\right)$, *con* $\mathbf{V} = \mathbf{D}^{-1/2}(I - \mathbf{H}_I)\mathbf{D}^{-1/2}$

Proof: i)The proof is similar to that of Theorem 3, but considering the distribution of $\operatorname{vec}\widehat{\boldsymbol{\epsilon}}_I$ instead of that of the distribution of $\widehat{\boldsymbol{\epsilon}}_I$.

ii) It is analogous to the proof of the Theorem 4. ∎

Now, observe that

$$\frac{\mathbf{W}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i}{\sigma\sqrt{(1-h_{ii})}} \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p) \quad \text{independent of} \quad \frac{(n-\alpha-1)\widehat{\sigma}^2_{(i)}}{\sigma^2} \sim \chi^2(p(n-\alpha)-1)$$

then

$$\frac{\dfrac{\mathbf{W}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i}{\sigma\sqrt{(1-h_{ii})}}}{\dfrac{(n-\alpha-1)\widehat{\sigma}^2_{(i)}}{\sigma^2}} = \frac{\mathbf{W}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i}{\widehat{\sigma}\sqrt{(1-h_{ii})}} = \mathbf{u}_i \sim \mathbf{t}_p(p(n-\alpha)-1,\mathbf{0},\mathbf{I}_p)$$

see (Kotz and Nadarajah, 2004, p. 2). Moreover

$$\left\|\frac{\mathbf{W}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i}{\sigma\sqrt{(1-h_{ii})}}\right\|^2 \sim \chi(p) \quad \text{independent of} \quad \frac{(n-\alpha-1)\widehat{\sigma}^2_{(i)}}{\sigma^2} \sim \chi^2(p(n-\alpha)-1)$$

thus

$$\frac{\dfrac{1}{p}\left\|\dfrac{\mathbf{W}^{-1/2}\widehat{\boldsymbol{\epsilon}}_i}{\sigma\sqrt{(1-h_{ii})}}\right\|^2}{\dfrac{(n-\alpha-1)\widehat{\sigma}^2_{(i)}}{(n-\alpha-1)\sigma^2}} = \frac{1}{p}\|\mathbf{u}_i\|^2 \sim \mathcal{F}(p,p(n-\alpha)-1),$$

where $\mathcal{F}(p,p(n-\alpha)-1)$ denotes the central $\mathcal{F}$ distribution with p and $p(n-\alpha)-1$ degrees of freedom.

Theorem 11 (Externally studentised residual). *Under model (19) we have that*

i) $\mathbf{u}_I \sim \mathcal{M}\mathbf{t}_{k\times p}\left(p(n-\alpha)-k,\mathbf{0},(\mathbf{I}_k\otimes\mathbf{I}_p)\right)$

ii) $\mathbf{u}_I^\tau \sim \mathcal{M}\mathbf{t}_{k\times p}\left(p(n-\alpha),\mathbf{0},(\mathbf{V}\otimes\mathbf{I}_p)\right)$, *con* $\mathbf{V}=\mathbf{D}^{-1/2}(\mathbf{I}_k-\mathbf{H}_I)\mathbf{D}^{-1/2}$

Proof: The demonstration is analogous to that given in Graybill (1985, Theorem 6.6.1, pp. 201-202), but using the distribution of $\operatorname{vec}\widehat{\boldsymbol{\epsilon}}_I$ instead of that of $\widehat{\boldsymbol{\epsilon}}_I$. ∎

Remark 2. The marginal distributions of $\mathbf{r}_i$ and $\mathbf{u}_i$ are obtained from Theorems 10 and 11 by taking $k=1$ and, for the univariate case, i.g. $\boldsymbol{\epsilon}\sim\mathcal{N}_n(\mathbf{0},\sigma^2\mathbf{W})$ by setting $p=1$ in the same theorems.

6. RESIDUAL UNDER MATRIX-VARIATE ELLIPTICAL DISTRIBUTION

In this section we consider models (1) and (19) but assume that $\boldsymbol{\epsilon}\sim\mathcal{E}_{n\times p}(\mathbf{0},\mathbf{I}_n\otimes\boldsymbol{\Sigma},h)$ y $\boldsymbol{\epsilon}\sim\mathcal{E}_{n\times p}(\mathbf{0},\mathbf{I}_n\otimes\sigma^2\mathbf{W},h)$, respectively, see Gupta and Varga (1993, p. 26) or Fang and Zhang (1990, p. 103). Note that if $K(\widehat{\boldsymbol{\epsilon}})$ denotes generically any kind of residual, then $K(\cdot)$ takes the form

$$K(\widehat{\boldsymbol{\epsilon}}) = \frac{g(\widehat{\boldsymbol{\epsilon}})}{||\widehat{\boldsymbol{\epsilon}}||},$$

and it is such that for $a > 0$ we get

$$K(a\widehat{\boldsymbol{\epsilon}}) = \frac{g(a\widehat{\boldsymbol{\epsilon}})}{||a\widehat{\boldsymbol{\epsilon}}||} = \frac{ag(\widehat{\boldsymbol{\epsilon}})}{a||\widehat{\boldsymbol{\epsilon}}||} = \frac{g(\widehat{\boldsymbol{\epsilon}})}{||\widehat{\boldsymbol{\epsilon}}||} = K(\widehat{\boldsymbol{\epsilon}})$$

and so, by Theorem 5.3.1 in Gupta and Varga (1993, p. 182), the distributions of all residual classes found in the above sections are invariant under the whole family of elliptical distributions. Moreover, they coincide with the distributions under the normality assumption. In summary, all the distributions found in this research are true, not only under normality, but also under an elliptical model.

7. CONCLUSIONS

We have shown that the distributions of different kinds of internally studentised residuals belong to a family of Pearson Type II distributions and that the externally studentised residuals belong to a family of t distributions. We remark that this provides a method for finding many numbers and/or random vectors with these distributions. Moreover, it is possible to determine the distributions of the $||\mathbf{r}_i||^2$ and $||\mathbf{u}_i||^2$ starting from the distributions of $\mathbf{r}_i$ and $\mathbf{u}_i$, respectively. Similarly, we can now easily find the distributions of the matrices $\widehat{\mathbf{r}}_I{}^T\widehat{\mathbf{r}}_I$ and $\widehat{\mathbf{u}}_I{}^T\widehat{\mathbf{u}}_I$ (or of the matrices proportional to these) starting from the distributions of $\widehat{\mathbf{r}}_I$ and $\widehat{\mathbf{u}}_I$, respectively; this goal can be reached, by just following the method described for finding the Wishart distribution.

The reader might expect the results of Section 5 to be particular cases of the result in Section 4. However, this is not so: from Kotz and Nadarajah (2004, p. 2, 4), we know that a random p-dimensional vector with distribution t can be defined in two ways; namely:

$$\mathbf{t} = \begin{cases} S^{-1}\mathbf{Y} + \boldsymbol{\mu}, & \text{with} \quad \dfrac{\nu S^2}{\sigma^2} \sim \chi^2(\nu) \quad \text{and} \quad \mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}) \\ \mathbf{W}^{-1/2}\mathbf{Y} + \boldsymbol{\mu}, & \text{with} \quad \mathbf{W} \sim \mathcal{W}_p(\nu + p - 1, \boldsymbol{\Sigma}) \quad \text{and} \quad \mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \nu\mathbf{I}_p) \end{cases}$$

with $(\mathbf{W}^{1/2})^2 = \mathbf{W}$ y $\mu : p \times 1$ a constant vector. Consider the sample $\mathbf{t}_1, \ldots, \mathbf{t}_n$ of a multivariate population with $\mathbf{t}$ distribution, arranged in the matrix $\mathbf{T} = (\mathbf{t}_1 \cdots \mathbf{t}_n) : p \times n$, then

$$\mathbf{T} = \begin{pmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_n^T \end{pmatrix}^T = \begin{cases} \begin{pmatrix} S^{-1}\mathbf{Y}_1^T + \boldsymbol{\mu}_1^T \\ \vdots \\ S^{-1}\mathbf{Y}_n^T + \boldsymbol{\mu}_n^T \end{pmatrix}^T = S^{-1}\mathbb{Y} + \mathbf{M} \quad \text{with} \quad \begin{cases} \dfrac{\nu S^2}{\sigma^2} \sim \chi^2(\nu) \\ \text{and} \\ \mathbb{Y} \sim \mathcal{N}_{p\times n}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n) \end{cases} \\ \text{or} \\ \begin{pmatrix} \mathbf{Y}_1^T\mathbf{W}^{-1/2} + \boldsymbol{\mu}_1^T \\ \vdots \\ \mathbf{Y}_n^T\mathbf{W}^{-1/2} + \boldsymbol{\mu}_n^T \end{pmatrix}^T = \mathbf{W}^{-1/2}\mathbb{Y} + \mathbf{M} \quad \text{with} \quad \begin{cases} \mathbf{W} \sim \mathcal{W}_p(\nu + p - 1, \boldsymbol{\Sigma}) \\ \text{and} \\ \mathbb{Y} \sim \mathcal{N}_{p\times n}(\mathbf{0}, \nu(\mathbf{I}_p \otimes \mathbf{I}_n)) \end{cases} \end{cases}$$

where $\mathbf{M} = (\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_n) : p \times n$, and $\mathbb{Y} = (\mathbf{Y}_1 \cdots \mathbf{Y}_n)$. But matrix $\mathbf{T}$ does not have the same distribution under the above two representations, even when their rows have the same distribution. In the representation, $\mathbf{T}$ has a matric-variate $t-$distribution and under the second one it has a matric-variate $T-$distribution, see Definition 1.

REFERENCES

Beckman, R. J., and Trussell, H. J. (1974), "The distribution of an arbitrary studentized residual and effects of updating in multiple regression," *Journal of the American Statistical Association*, 69, 199-201.

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

Caroni, C. (1987), "Residuals and influence in the multivariate linear model", *The Statistician*, 36, 365-370.

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

Cramér, H. (1999), *Mathematical Methods of Statistics*, Nineteenth printing, Princenton University Press, Princeton.

Dickey, J. M. (1967), "Matricvariate generalizations of the multivariate t- distribution and the inverted multivariate t-distribution," *Annals of Mathematical Statistics* 38, 511-518.

Díaz-García, J. A., Gutiérrez-Jáimez, R., and Mardia, K. V. (1997), "Wishart and Pseudo-Wishart distributions and some applications to shape theory," *Journal of Multivariate Analysis*, 63, 73-87.

Díaz-García, J.A., and González-Farías G. (2004), "A Note on the Cook's Distance," *Journal of Statistical Planning and Inference,* 120/1-2, 112-136.

Díaz-García, J. A. and González-Farías, G. (2004), "Singular Random Matrix decompositions: Jacobians," *Journal of Multivariate Analysis*, in press. See, http://www.scincedirect.com.

Díaz-García, J. A. and González-Farías, G. (2004), "Singular Random Matrix decompositions: Distributions," *Journal of Multivariate Analysis*, in press. See, http://www.scincedirect.com.

Ellenberg, J. H. (1973), "The joint distribution of the standardized least squares residual from general linear regression," *Journal of American Statistical Association*, 68, 941-943.

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

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

Graybill, F., (1985), *Theory and Application of the Linear Model*, Wadsworth Publishing Company, Inc., Belmont, California.

Johnston, J. (1972), *Econometric methods*, McGraw Hill, New York.

Khatri, C. G. (1968), "Some results for the singular normal multivariate regression models," *Sankhyā A* 30, 267-280.

Kotz, S., and Nadarajah, S. (2004), *Multivariate t Distributions and Their Applications*, Cambridge University Press, United Kingdom.

Magnus, J. R. (1988), *Linear Structures*, Charles Griffin & Company Ltd, London.

Muirhead, R. J. (1982), *Aspects of multivariate statistical theory*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Olkin, I., and Rubin, H. (1964), "Multivariate beta distribution and independence properties of the Wishart distribution," *Annals of Mathematical Statistics* 35, 261-269.

Press, S. J. (1982), *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, Second Edition, Robert E. Krieger Publishing Company, Malabar, Florida.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, Second Edition, John Wiley & Sons, Inc., New York.

Srivastava, M. S., and Khatri, C. G. (1979), *An Introduction to Multivariate Analysis*, North-Holland Publ., Amsterdam.