Jackknife Estimation as a Robust Estimation in Linear Models Under Some Conditions

Lorenc EKONOMI[1], Osman HYSA[2], Eljona MILO[1], Lorena MARGO[1]
[1] *Department of Natural Sciences, "Fan S. Noli" University, Korça- ALBANIA*
lorencekonomi@yahoo.co.uk, eljonamilo@yahoo.com, lorena.margo@yahoo.com
[2] *Department of Mathematics, UAMD, Durrës- ALBANIA*
o.hysa@yahoo.com

ABSTRACT

One of the most important methods in statistics for estimating parameters is the resampling method. But estimations of unknown parameters with resampling methods demand a lot of calculations and it is very difficult to use them. These methods are found in lot of applications because of wide spread use of computers.

Tukey and Quenouoille [9] firstly introduced jackknife methods to estimate the bias and the variance during the estimation of unknown parameters. Afterwards, Efron [1-3] used jackknife methods to estimate the variance of least squares estimators in linear regression models. In many cases, jackknife estimators resulted successfully. But, there were some cases, when the results of jackknife estimators were far from the true values of the parameters [4]. This happened, because the jackknife estimators depended from various conditions that fulfill the model. These conditions were related with the values of the independent variables, the properties of the matrix of independent variables and the observation errors variances.

We have shown in our paper that ordinary jackknife estimations for variances of least squares estimators of unknown coefficients in linear regression models are not unbiased. Their accuracy depends on the variances of linear regression model errors and the nature of the matrix of the independent variables observations. We have found some conditions when the jackknife estimators are robust estimations (not influenced from the distribution of the sample elements) for the variances of least squares estimators of the linear regression model coefficients in the case when model errors are homoschedastic (the errors have equal variances) and in the case when model errors are heteroschedastic (the errors have not equal variances). We have analyzed the relationship between these conditions and have found that the relationship between them and the eigenvalues of the matrix $X^T X$, when X is the matrix of the observed independent variables.

INTRODUCTION

Jackknife estimation is a nonparametric method for the estimation of unknown parameters like the bias, the variance etc. Their beginners are of Quenouille (1949) [7] for the estimation of the bias and Tukey (1958) [8] for the estimation of the variance. Let we give the idea of the jackknife estimation.

Let suppose that $x_1,...,x_n$ are the sample observations of the random variable X and we have the statistic $\hat{\Theta}(X_1,...,X_n)$ to estimate the unknown parameter $^{-}$. Let we have $\hat{\Theta}_{(i)} = \hat{\Theta}(X_1,..., X_{i-1}, X_{i+1},..., X_n)$, if we have deleted the observation of i-s and let $\hat{\Theta}_{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Theta}_{(i)}$.

We assume that we want to estimate the variance of the estimator $\hat{\Theta}$. We know that the true variance of the estimator $\hat{\Theta}$ is $\mathrm{var}(\hat{\Theta}) = E(\hat{\Theta} - E(\hat{\Theta}))^2$.
Ordinary jackknife estimator for the variance of $\hat{\Theta}$ is [1-3]

$$\hat{\mathrm{var}}(\hat{\Theta}) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\Theta}_{(i)} - \hat{\Theta}_{(.)}\right)^2 . \qquad (1)$$

We can mention the fundamental work of Wu [9] about the jackknife in linear regression. We have studied the linear regression model in the case when the model errors have not with independent and identically distribution [4, 5] and in the case of the weighted regression model with unknown weight [6].

In our paper, we have analyzed the proprieties of the estimator (1) in linear regression models, where the unknown parameters are the coefficients of linear regression and they are estimated with OLS (ordinary least squares). Firstly we have taken the expression of the estimator (1) in linear models. Then, we have shown that this estimator in biased when the errors are homoscedastic and, under some conditions about the model, the estimator (1) is robust.

Jackknife estimation of linear regression coefficients Variance

It is given the linear regression model

$$y_i = x_i^T \quad + e_i \text{, for } i = 1,\dots,n , \tag{2}$$

where x_i a known vector kx1, the vector kx1 of unknown parameters, e_i the errors, that are uncorrelated and fulfill the conditions

$$E(e_i) = 0 \text{ and } var(e_i) = \pm_i^2 ,\ i = 1,\dots,n . \tag{3}$$

We denote $Y = (y_1,\dots,y_n)^T$; $e = (e_1,\dots,e_n)^T$ and $X = [x_1,\dots,x_n]^T$. We can write model in the form

$$Y = X \quad + e, \text{ where } E(e) = 0 \text{ and } \mathrm{var}(e) = diag[\pm_1^2,\dots,\pm_n^2]. \tag{4}$$

The OLS estimator for the unknown parameters is in the following form

$$\hat{} = \left(X^T X\right)^{-1} X^T Y . \tag{5}$$

We denote $\hat{}_{(i)}$ the OLS estimator for the parameters taken from (5), if we have delete the i-s observation (the vector x_i and y_i) and $\hat{}_{(.)} = \frac{1}{n}\sum_{i=1}^{n} \hat{}_{(i)}$. The expression (1) takes the form

$$\hat{\mathrm{var}}\left(\hat{}\right) = \frac{n-1}{n}\sum_{i=1}^{n}\left[\hat{}_{(i)} - \hat{}_{(.)}\right]\left[\hat{}_{(i)} - \hat{}_{(.)}\right]^T . \tag{6}$$

Let we see the following propositions.

Proposition 1.1 $\hat{}_{(i)} = \hat{} - \dfrac{\left(X^T X\right)^{-1} x_i r_i}{1 - x_i^T \left(X^T X\right)^{-1} x_i}$, where $r_i = y_i - x_i^T \hat{}$ for $i = 1,\dots,n$.

Proof. We denote $X_{(i)}$ and $Y_{(i)}$, respectively the matrices X dhe Y, after we have deleted the i-s observation (the vector x_i and the value y_i). Then

$$\hat{}_{(i)} = \left(X_{(i)}^T X_{(i)}\right)^{-1} X_{(i)}^T Y_{(i)} .$$

We see that $X^T X = X_{(i)}^T X_{(i)} + x_i x_i^T$ and $X^T Y = X_{(i)}^T Y_{(i)} + x_i y_i$ are true. We replace these expressions in the above expression for $\hat{\ }_{(i)}$ and we have

$$\hat{\ }_{(i)} = \left(X^T X - x_i x_i^T\right)^{-1}\left(X^T Y - x_i y_i\right).$$

In the first parenthesis we apply the matrix identity

$\left(X^T X - x_i x_i^T\right)^{-1} = \left(X^T X\right)^{-1} + \left(X^T X\right)^{-1} \dfrac{x_i x_i^T \left(X^T X\right)^{-1}}{1 - x_i^T \left(X^T X\right)^{-1} x_i}$. Then we do the following transformations

$$\hat{\ }_{(i)} = \left[\left(X^T X\right)^{-1} + \left(X^T X\right)^{-1} \frac{x_i x_i^T \left(X^T X\right)^{-1}}{1 - x_i^T \left(X^T X\right)^{-1} x_i}\right]\left(X^T Y - x_i y_i\right) =$$

$$= \hat{\ } + \left(X^T X\right)^{-1} x_i y_i - \frac{\left(X^T X\right)^{-1} x_i x_i^T \hat{\ } - \left(X^T X\right)^{-1} x_i x_i^T \left(X^T X\right)^{-1}}{1 - x_i^T \left(X^T X\right)^{-1} x_i} =$$

$$= \hat{\ } + \left(X^T X\right)^{-1} x_i y_i - \frac{\left(1 - x_i^T \left(X^T X\right)^{-1} x_i\right)\left(\left(X^T X\right)^{-1} x_i y_i + \left(X^T X\right)^{-1} x_i r_i\right)}{1 - x_i^T \left(X^T X\right)^{-1} x_i}$$

.

Although the expression $1 - x_i^T \left(X^T X\right)^{-1} x_i$ is a number, we take

$$\hat{\ }_{(i)} = \hat{\ } - \frac{\left(X^T X\right)^{-1} x_i r_i}{1 - x_i^T \left(X^T X\right)^{-1} x_i} . \square$$

Proposition 1.2 The jackknife estimator (6) for OLS estimation of linear regression coefficients has the form

$$\hat{\mathrm{var}}\left(\hat{\ }\right) = \frac{n-1}{n}\left(X^T X\right)^{-1}\left[X^T \Lambda X - \frac{1}{n} X^T q q^T X\right]\left(X^T X\right)^{-1}, \tag{7}$$

where $\Lambda = diag\left[\frac{r_1^2}{(1-w_1)^2},...,\frac{r_n^2}{(1-w_n)^2}\right]$, $q^T = \left(\frac{r_1}{1-w_1},...,\frac{r_n}{1-w_n}\right)$ and $w_i = x_i^T\left(X^TX\right)^{-1}x_i$.

Proof. From the Proposition 1.1, we have $\hat{}_{(.)} = \hat{} - \frac{1}{n}\left(X^TX\right)^{-1}\sum_{j=1}^{n}\frac{x_jr_j}{1-w_j}$. Then

$\hat{}_{(i)} - \hat{}_{(.)} = \left(X^TX\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j} - \frac{x_ir_i}{1-w_i}\right]$, and from here, we take

$$\left[\hat{}_{(i)} - \hat{}_{(.)}\right]\left[\hat{}_{(i)} - \hat{}_{(.)}\right]^T = \left(X^TX\right)^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j} - \frac{x_ir_i}{1-w_i}\right]\left[\frac{1}{n}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j} - \frac{x_ir_i}{1-w_i}\right]^T\left(X^TX\right)^{-1} =$$

$$= \left(X^TX\right)^{-1}\left[\frac{1}{n^2}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j}\sum_{i=1}^{n}\frac{x_j^Tr_j}{1-w_j} - \frac{2}{n}\left(\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j}\right)\frac{x_i^Tr_i}{1-w_i} + \frac{x_ir_i}{1-w_i}\frac{x_i^Tr_i}{1-w_i}\right]\left(X^TX\right)^{-1}$$

$$\sum_{i=1}^{n}\left[\hat{}_{(i)} - \hat{}_{(.)}\right]\left[\hat{}_{(i)} - \hat{}_{(.)}\right]^T = \left(X^TX\right)^{-1}\left[\sum_{i=1}^{n}\frac{x_ix_i^Tr_i^2}{(1-w_i)^2} - \frac{1}{n}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j}\right]\left(X^TX\right)^{-1}$$

or

Then

$$\hat{\mathrm{var}}\left(\hat{}\right) = \frac{n-1}{n}\sum_{i=1}^{n}\left[\hat{}_{(i)} - \hat{}_{(.)}\right]\left[\hat{}_{(i)} - \hat{}_{(.)}\right]^T =$$

$$= \frac{n-1}{n}\left(X^TX\right)^{-1}\left[\sum_{i=1}^{n}\frac{x_ix_i^Tr_i^2}{(1-w_i)^2} - \frac{1}{n}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j}\sum_{i=1}^{n}\frac{x_jr_j}{1-w_j}\right]\left(X^TX\right)^{-1}.$$

We denote $\Lambda = diag\left[\frac{r_1^2}{(1-w_1)^2},...,\frac{r_n^2}{(1-w_n)^2}\right]$, $q^T = \left(\frac{r_1}{1-w_1},...,\frac{r_n}{1-w_n}\right)$ and we have the expression (7). □

The errors of the linear regression model (4) are homoscedastic, if they have equal variances, so $var(e_i) = \pm_i^2 = \pm^2$, $i = 1,...,n$ or $\Sigma = \pm^2 I$.

Proposition 1.3 If in linear regression model (4) the errors are homoscedastic, then $E(r_i^2) = (1 - w_i)\pm^2$ and $E(r_i r_j) = -w_{ij}\pm^2$, where $w_{ij} = x_i^T (X^T X)^{-1} x_j$.

Proof. We have $r_i = e_i - x_i^T (X^T X)^{-1} Xe$ and $E(r_i) = 0$. From that we have

$$E(r_i^2) = E(e_i^2) - 2x_i^T (X^T X)^{-1} X^T E(e_i e) + x_i^T (X^T X)^{-1} X^T E(ee^T) X (X^T X)^{-1} x_i =$$

$$= \pm_i^2 - 2w_i \pm_i^2 + \sum_{j=1}^{n} w_{ij}^2 \pm_j^2 = (1 - w_i)\pm_i^2 + \sum_{j=1}^{n} w_{ij}^2 (\pm_j^2 - \pm_i^2),$$

We see that, when the errors are homoscedastic, we have $E(r_i^2) = (1 - w_i)\pm^2$. Let us find $E(r_i r_j)$, for $i \neq j$. We have

$$E(r_i r_j) = E(e_i e_j) - x_i^T (X^T X)^{-1} X^T E(e_i e) - x_j^T (X^T X)^{-1} X^T E(e_i e) +$$

$$+ x_i^T (X^T X)^{-1} X^T E(ee^T) X (X^T X)^{-1} x_j = -w_{ij}\pm_j^2 - w_{ji}\pm_i^2 +$$
$$x_i^T (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} x_j . \square$$

Let us see through the two following theorems the proprieties of jackknife estimators for the linear regression model coefficients variance estimated through OLS.

Theorem 1.1 If the errors variance in the linear regression model is equal, then the estimator (6) is biased.

Proof. From the Proposition 1.3, we have $E\left(r_i^2\right)=\left(1-w_i\right)\pm^2$ dhe $E\left(r_i r_j\right)=-w_{ij}\pm^2$. We take

$$E(\Lambda)=diag\left(\frac{\pm^2}{1-w_1},\ldots,\frac{\pm^2}{1-w_n}\right) \text{ and } E\left(qq^T\right)=\begin{cases}\dfrac{\pm^2}{1-w_i} & \text{for } i=j\\[2ex] \dfrac{w_{ij}\pm^2}{\left(1-w_i\right)\left(1-w_j\right)} & \text{for } i\neq j\end{cases}.$$

From that, we have $$E\left(\Lambda-\frac{1}{n}qq^T\right)=\begin{cases}\dfrac{n-1}{n}\dfrac{\pm^2}{1-w_i} & \text{for } i=j\\[2ex] -\dfrac{1}{n}\dfrac{w_{ij}\pm^2}{\left(1-w_i\right)\left(1-w_j\right)} & \text{for } i\neq j\end{cases}.$$

Thus,

$$E\left(\widehat{\mathrm{var}}\left(\hat{\ }\right)\right)=\left(X^TX\right)^{-1}X^TAX\left(X^TX\right)^{-1}, \tag{8}$$

where $$a_{ij}=\begin{cases}\left(\dfrac{n-1}{n}\right)^2\dfrac{\pm^2}{1-w_i} & \text{for } i=j\\[2ex] -\dfrac{n-1}{n^2}\dfrac{w_{ij}\pm^2}{\left(1-w_i\right)\left(1-w_j\right)} & \text{for } i\neq j\end{cases}.$$

On the other hand, the true variance of linear regression model coefficients OLS estimation is given by the expression $\mathrm{var}\left(\hat{\ }\right)=\pm^2\left(X^TX\right)^{-1}$. But, this expression is different with the expression (8), because, in general

$$\left(\frac{n-1}{n}\right)^2\frac{\pm^2}{1-w_i}\neq\pm^2. \square \tag{9}$$

We see, from the Theorem 1.1, than the estimator (6) is biased for the linear regression model coefficients variance estimated through OLS. From the expression (9) we see that the bias can be negligible or controlled if the linear regression model

(4) fulfill the conditions that the values $\left(\frac{n-1}{n}\right)^2 \frac{1}{1-w_i}$ are nearly 1. Let us see that via the following theorem.

Theorem 1.2 Let us suppose that in linear model (4) with homoscedastic errors the following condition is true

$$0 \le w_i \le \frac{c}{n}, \text{ for } i = 1,\dots,n, \tag{10}$$

where c is a constant not depended from n, and then the estimator (6) is a robust estimator for the linear regression coefficients variance estimated through OLS.

Proof. We have $\frac{n-c}{n} \le 1 - w_i \le 1$ or $\frac{1}{1-w_i} \le \frac{n}{n-c}$. From that we take

$$\left| \left(\frac{n-1}{n}\right)^2 \frac{\pm^2}{1-w_i} - \pm^2 \right| = \pm^2 \left| \frac{-2n+1}{n^2(1-w_i)} + \frac{w_i}{1-w_i} \right| \le \pm^2 \left(\left| \frac{-2n+1}{n^2(1-w_i)} \right| + \left| \frac{w_i}{1-w_i} \right| \right) \le$$

$$\le \pm^2 \left(\frac{3}{n}\frac{n-c}{n} + \frac{c}{n}\frac{n}{n-c} \right) \le \pm^2 \frac{c_1}{n} = O\left(\frac{1}{n}\right).$$

If we replace this expression in (8), we have

$$E\left(\hat{\text{var}}\left(\hat{\ }\right)\right) = \text{var}\left(\hat{\ }\right) + O\left(\frac{1}{n}\right), \tag{11}$$

or the estimator (6) is a robust estimation for the $\hat{\ }$ variance. □

2. Some conditions about linear regression model

We see from the Theorem 1.2 that, when the condition (10) is true, the estimator (6) is a robust estimator. In this section we will see some other conditions that induce the condition (10).

Let us have the following conditions:

A1. The eigenvalues of the matrix $\frac{1}{n}X_n^T X_n$ are uniformly bounded.

A2. The elements of the matrix X_n are uniformly bounded.

A3. The minimal eigenvalue of the matrix $\frac{1}{n}X_n^T X_n$ is bounded downwards and $\left|\frac{1}{n}X_n^T X_n\right| = O(1)$.

A4. The maximal eigenvalue of the matrix $\frac{1}{n}X_n^T X_n$ is bounded upwards and $\left|\frac{1}{n}X_n^T X_n\right| = O(1)$.

In the following, we show that some of the above conditions induce the condition (10).

Proposition 2.1 If the conditions (A1) and (A2) are true, then the condition (10) is true.

Proof. Let us have $\alpha_1^{(n)} \geq .... \geq \alpha_k^{(n)}$ the eigenvalues of the matrix $\frac{1}{n}X_n^T X_n$. From the condition (A1) we have $\alpha_1^{(n)} \leq c_1$ dhe $\alpha_k^{(n)} \geq c_2 > 0$. Then the eigenvalues of the matrix $\left(X_n^T X_n\right)^{-1}$ are $\frac{1}{n}\frac{1}{\alpha_1^{(n)}} \geq .... \geq \frac{1}{n}\frac{1}{\alpha_k^{(n)}}$ and we have $\frac{1}{n}\frac{1}{\alpha_1^{(n)}} \leq \frac{1}{nc_1}$.

From the condition (A2) we have $\left|x_{ij}\right| \leq c_3$ for $i = 1,..., n;\ \ j = 1,..., k$. We take $w_i = x_i^T\left(X^T X\right)^{-1} x_i \leq \frac{1}{n}\frac{1}{\alpha_1^{(n)}} x_i^T x_i \leq \frac{1}{nc_1} x_i^T x_i < \frac{kc_3}{nc_1} = \frac{c}{n}$, because the matrix $\frac{1}{n}X_n^T X_n$ is a symmetric matrix. □

Proposition 2.2 If the conditions (A2) and (A3) are true, then the condition (10) is true.

Proof. We can write

$$n\alpha_k^{(n)}\left(x_i^T x_i\right)x_i^T\left(X_n^T X_n\right)^{-1}x_i \le x_i^T\left(X_n^T X_n\right)x_i x_i^T\left(X_n^T X_n\right)^{-1}x_i \le \frac{\alpha_1^{(n)}}{\alpha_k^{(n)}}\left(x_i^T x_i\right)^2$$

. Then we have $\alpha_1^{(n)} = \dfrac{\left|\frac{1}{n}X_n^T X_n\right|}{\alpha_2^{(n)}\cdot\ldots\cdot\alpha_k^{(n)}} \le \dfrac{\left|\frac{1}{n}X_n^T X_n\right|}{\left(\alpha_k^{(n)}\right)^{k-1}}$ and the above expression takes the form $w_i = x_i^T\left(X_n^T X_n\right)^{-1}x_i \le \dfrac{1}{n}\dfrac{\left|\frac{1}{n}X_n^T X_n\right|}{\left(\alpha_k^{(n)}\right)^{k+1}} \le \dfrac{c}{n}$.□

Proposition 2.3 If the conditions (A2) and (A4) are true, then the condition (10) is true.

Proof. Reasoning in the some way with the Proposition 2.2, we have $\alpha_k^{(n)} = \dfrac{\left|\frac{1}{n}X_n^T X_n\right|}{\alpha_1^{(n)}\cdot\ldots\cdot\alpha_{k-1}^{(n)}} \ge \dfrac{\left|\frac{1}{n}X_n^T X_n\right|}{\left(\alpha_1^{(n)}\right)^{k-1}}$ or $\dfrac{1}{\alpha_k^{(n)}} \le \dfrac{\left(\alpha_1^{(n)}\right)^{k-1}}{\left|\frac{1}{n}X_n^T X_n\right|}$. Finally we have

$$w_i \le \frac{1}{n}\frac{\left(\alpha_1^{(n)}\right)^{2k-1}}{\left|\frac{1}{n}X_n^T X_n\right|} \le \frac{c}{n}.\square$$

Conclusion

From the above results we see that ordinary jackknife estimator is biased for the linear regression model coefficients estimated by OLS in the case when the model errors variances are homoscedastic. From the Theorem 1.2 we have arrived in the conclusions that ordinary jackknife estimator is a robust estimation, when the linear models fulfill some conditions. So, we can give the idea of ordinary jackknife estimator modification, to take into consideration the influence of the individual observations and the nature of the observations matrix.

REFERENCES

[26] Efron, B. (1979) Bootstrap methods: another look at the jackknife. *The Annals of Statistics*. **7**(1), 1-29

[27] Efron, B. (1982) The Jackknife, the Bootstrap and Other Resampling Plans. *SIAM*.

[28] Efron, B. and Gong, G. (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. *The American Statistician*, **37**(1), 36-47.

[29] Ekonomi L., Butka A. (2001) Jackknife estimations for linear regression coefficients variance in not i.i.d. situations. *Buletini i Universitetit “Fan S. Noli” Korce*, **3**, 121-128.

[30] Ekonomi L., Capollari Gj. (2004) A view about the Wu’s jackknife estimation terms and the design of an algorithm to calculate it. *Buletini shkencor UNIEL*. **1**, 19-32.

[31] Ekonomi L. (2004) Wu’s weighted jackknife estimation for the weighted linear regression coefficients variance with unknown weights. *Buletini Matematika dhe shkencat e natyres.***1**, 16-26.

[32] Quenouille, M. (1949) Approximate tests of correlation in time series. *J. Roy. Statist. Soc. Ser. B*, **11**, 18-84.

[33] Tukey, J. (1958) Bias and confidence in not quite large samples. *Ann. Math. Statist*., **29**, 614.

[34] Wu, C. F. J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, **14**(4), 1261-1295.