A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 90, 1989

Zbigniew Wasilewski *

REMARKS ON BLUS RESIDUALS

1. INTRODUCTION

The BLUS residuals theory (Best Linear Unbiased residuals with Scalar covariance matrix) developed by Theil [7] and Koerts [2] is generaly related to problems of estimating the error term in linear regression models. This estimator possesses the same correlation structure as the unknown disturbances. This seems to be important for statistical inference about the stochastic structure of the regression model.

We have a regression model of the form

$$y = X\beta + u \tag{1}$$

where under common assumptions:

(a) X is an n:k nonstochastic matrix of rank k which contains the values taken by the k independent variables in n periods;

(b) $\lim_{n \to \infty} n^{-1}(X'X)$ is a finite nonsingular matrix;

(c) the vector of random disturbances, has uncorrelated elements with zero mean and constant variance, i.e. $E(u) = 0$, $E(uu') = \sigma^2 I$;

(d) in addition it is often assumed, that the disturbances are normally distributed.

* Lecturer at the Institute of Econometrics and Statistics, University od Łódź.

Due to lack of knowledge about real values of disturbances, inferences about the stochastic assumptions (c) and (d) must be based on some estimate of u having related properties. It seems straightforward to treat as such an estimate the vector of the approximation errors $y - Xb$ where b is given by minimization of the chosen approximation criterion, for instance $\Sigma(y_i - x'_{i.}b)^2$ or $\Sigma|y_i - x'_{i.}b|$. The minimization of the error sum of squares is so far the most popular criterion of estimation, in part due to its attractive analytical and numerical properties. Using this criterion we obtain the well-known least squares estimator of β

$$b = (X'X)^{-1}X'y \tag{2}$$

and the corresponding vector of least-squares residuals

$$e = y - Xb = My = Mu \tag{3}$$

where, $M = (I - X(X'X)^{-1} X')X$ is the idempotent $n:n$ projection matrix of rank $n - k$.

The estimator of u has the following desirable properties:

(a) it is linear in the dependent variable;

(b) it is unbiased;

(c) it has the smallest expected sum of squares of the estimation errors, within the set of all linear and unbiased estimators.

On the other hand, however, the covariance matrix of this estimator, assuming that $E(ee') = \sigma^2 I$, is given by

$$E(ee') = E(Muu'M) = \sigma^2 M \tag{4}$$

Thus the least squares residuals are correlated and their covariance matrix depends on the particular X matrix. This makes the least squares estimator of disturbances less useful for testing purposes. Clearly e can be transformed to have a different correlation structure, but due to the fact that l - s residuals are singulary distributed (the rank of the projection matrix M is $n - k$) we can obtain only $n - k$ transformed residuals that are uncorrelated. Moreover the obtained solution is not unique and the choice of the k residuals that are not estimated can be very important. In the paper we try to make some evidence, about the mutual relations between the basis of the transformation and the accuracy of the esimation, measured by mean square error (MSE), in the case of outliers among data. We propose to use residuals minimizing the sum of absolute deviations to obtain non-biased residuals with scalar covariance matrix, in such a case.

2. NOTES ON THE CONSTRUCTION OF THE BLUS RESIDUALS

Suppose that an $(n - k) : n$ matrix C defines a linear transformation $\hat{e} = C'y$. We will call $\hat{e}$ a vector of Linear Unbiased residuals with Scalar covariance matrix (LUS) if,

(i) $E(\hat{e}) = 0$ and (ii) $E(\hat{e}\hat{e}') = \sigma^2 I$

The conditions (i), (ii) require only that $C'X = 0$ and $C'C = I$.

There exist a few methods of derivation of the matrix C which fulfill these conditions. All of them require the choice of $n - k$ residuals to be estimated. The choice of these residuals is more or less arbitrary, so that the definition of the LUS residuals is not unique.

Suppose we partition $u' = (u'_o\ u'_1)$, $X' = (X'_o X'_1)$ and $C' = (C'_o C'_1)$ such that the subscript 0 corresponds to the k components of u which are not estimated and subscript 1 corresponds to the remaining $n - k$ cases. This is always possible by simple reordering of the rows of the matrix X.

We additionaly assume, that the $k : k$ matrix X_o is nonsingular. The conditions $C'X = 0$ and $C'C = I_{n-k}$ can now be written using the partition of the matrices C and X as follows,

$$C'X = (C'_o C'_1)\begin{pmatrix} X_o \\ X_1 \end{pmatrix} = C'_o X_o + C'_1 X_1 = 0 \tag{5}$$

$$C'C = (C'_o C'_1)\begin{pmatrix} C_o \\ C_1 \end{pmatrix} = C'_o C_o + C'_1 C_1 = I_{n-k} \tag{6}$$

C_o is thus determined uniqualy from (5) and equals,

$$C'_o = -C'_1 X_1 X_o^{-1} \qquad (5a)$$

where the assumption of nonsingularity of X_o is used.
Substituting (5a) in (6) obtain,

$$C'C = C'_1 X_1 X_o^{-1} X'_o{}^{-1} X'_1 C_1 + C'_1 C_1 = I_{n-k} \qquad (7)$$

Substituting $X'_o X_o$ by $X'X - X'_1 X_1$ we obtain[1],

$$C'_1 X_1 (X'X - X'_1 X_1)^{-1} X'_1 C_1 + C'_1 C_1 = C'_1 X_1 [(X'X)^{-1} +$$
$$- (X'X)^{-1} X'_1 (I - X_1 (X'X)^{-1} X'_1)^{-1} X_1 (X'X)^{-1}] X'_1 C_1 + C'_1 C_1 \qquad (8)$$
$$= C'_1 [X_1 (X'X)^{-1} X'_1 + X_1 (X'X)^{-1} X'_1 (I - X_1 (X'X)^{-1} X'_1)^{-1} \cdot$$
$$\cdot X_1 (X'X)^{-1} X'_1 + I] C_1 = I_{n-k}$$

Let us put now $A = X_1 (X'X)^{-1} X'_1$ then (8) can be written as

$$C'_1 (A + A(I - A)^{-1} A + I) C_1 =$$
$$= C'_1 (A(I + (I - A)^{-1} A) + I) C_1 =$$
$$= C'_1 [A((I - A)^{-1}(I - A) + (I - A)^{-1} A) +$$
$$+ (I - A)(I - A)^{-1}] C_1 = \qquad (8a)$$
$$= C'_1 [A (I - A)^{-1} (I - A + A) + (I - A)(I - A)^{-1}] C_1 =$$
$$= C'_1 (I - A)^{-1} C_1 = C'_1 (I - X_1 (X'X)^{-1} X'_1)^{-1} C_1$$

---

[1] In the derivation the inversion of the product $X'X - X_1 X_1$ was obtained using the following updating formulae due to Gauss.
Lema 1. Let A be p : p rank p symmetrix matrix, and suppose that X and Y are q : p rank q matrices. Then, provided that the inverses exist, $(A + X'Y)^{-1} = A^{-1} - A^{-1} X' (I + YA^{-1} X')^{-1} YA^{-1}$.

It follows from 8a that C_1 must satisfy,

$$C'_1(I - X_1(X'X)^{-1} X'_1)^{-1}C_1 = I_{n-k} \tag{9}$$

Thus C_1 must be chosen to be any factorization of the matrix $(I - X_1(X'X)^{-1}X'_1)^{-1}$ and C_o is then determined uniquely from $C'_o = -C'_1X_1X_o^{-1}$.

T h e i l [7], [8] showed that the use of spectral decomposition of the matrix $(I - X_1(X'X)^{-1}X'_1)^{-1}$ to find C_1 leads to LUS residuals with the smallest expected residual sum of squares. Due to this additional property these residuals are called the Best Linear Unbiased Scalar covariance matrix (BLUS) residuals.

It follows from direct multiplication (Theil [7]), that the inverse of the matrix $(I - X_1(X'X)^{-1}X'_1)$ exists and is of the form,

$$(I - X_1(X'X)^{-1}X'_1)^{-1} = I + X_1(X'_oX_o)^{-1}X'_1 = I + ZZ' \tag{10}$$

where $Z = X_1X_o^{-1}$.

The matrix $I - X_1(X'X)^{-1}X'_1$ is, as a nonsingular submatrix of the positive semi-definite matrix $M = I - X(X'X)^{-1}X'$, positive definite. Thus, there exists a square orthogonal matrix P, such that

$$\begin{aligned} P'(I - X_1(X'X)^{-1}X_1)P &= D \\ P'(I - X_1(X'X)^{-1}X_1)^{-1}P &= D^{-1} \end{aligned} \tag{11}$$

where $P'P = I$ and D, D^{-1} are diagonal matrices with the latent roots of $I - X_1(X'X)^{-1}X_1$ and $(I - X_1(X'X)^{-1}X_1)^{-1}$ on the main diagonal[2].

On premultiplying both sides of the second equation in (11) by P and postmultiplying by P' we obtain,

$$(I - X_1(X'X)^{-1}X_1)^{-1} = I + ZZ' = PD^{-1}P' \tag{12}$$

---

[2] Given the positive definitness of $I - X_1(X'X)^{-1}X_1$ all these latent roots are positive.

The condition (9) can be thus written in the form,

$$C'_1 P D^{-1} P' C_1 = I \tag{13}$$

which is fulfilled for the matrix C_1 given by,

$$C_1 = P D^{1/2} P' \tag{14}$$

Let us now consider the characteristic equation for the matrix $M_{11} = (I + ZZ')^{-1}$,

$$[(I + ZZ')^{-1} - d_i I] \, p_i = 0 \tag{15}$$

On premultiplying (15) by $I + ZZ'$ we have,

$$[I - d_i I - ZZ' d_i] p_i = 0 \tag{16}$$

and after dividing by $(-d_i)$ and substituting $X_1 X_o^{-1}$ for Z we finally obtain,

$$[X_1 X_o^{-1} (X_1 X_o^{-1})' - (1/d_i - 1)] p_i = 0 \tag{17}$$

From (17) we find that the characteristic vectors of $I + - X_1 (X'X)^{-1} X_1$ are the principal components of the matrix $X_1 X_o^{-1}$. This matrix can be treated as a matrix of "index transformed values" of explanatory variables with matrix X_o as a basis of this transformation.

Since the positive semi definite matrix $X_1 X_o^{-1} (X_1 X_o^{-1})'$ is of order $(n - k) : (n - k)$ and of rank k or less it has at least $n - 2k$ zero latent roots and at most k positive latent roots. Hence, at least $n - 2k$ of the d's are equal to 1 and at most k of them are less than 1.

Taking this into account we can now rewrite C_1 in the form,

$$C_1 = \sum_{i=1}^{k} d_i^{1/2} \, p_i p_i' + \sum_{i=1}^{n-k} p_i p'_i =$$

$$= M_{11} \sum_{i=1}^{k} d_i^{1/2} (1 - d_i^{1/2}) \, p_i P'_i \tag{18}$$

Thus the transformation matrix C_1 for BLUS residuals is obtained from submatrix M_{11} of the l-s transformation matrix M by adding k matrices $p_i p'_i$. of unit rank scaled by the factor $d_i^{1/2}(1 - d_i^{1/2})$.

3. THE PRICE OF THE SCALAR COVARIANCE CONDITION AND THE CHOICE OF THE BASIS OF THE TRANSFORMATION

The expected sum of squares of the BLUS residuals is, appart from the factor σ^2, equal to (T h e i l ([7])

$$E[(\hat{e} - u_1)'(\hat{e} - u_1)] = 2(n - k) - 2trC_1 =$$

$$= 2(n - k) - 2(n - 2k + \sum_{i=1}^{k} d_i^{1/2}) = 2 \sum_{i=1}^{k} (1 - d_i^{1/2}) \qquad (19)$$

Thus it depends on the choice of the basis X_o.

N e u d e c k e r [4] showed that the covariance matrix of the BLUS residuals equals the sum of the covariance matrices of the l - s residuals and the l-s - BLUS differences,

$$E[(\hat{e} - u)(\hat{e} - u)'] = E[(e - u)(e - u)'] +$$

$$+ E[(\hat{e} - e)(\hat{e} - e)'] \qquad (20)$$

where $\hat{e} = (0', \hat{e}')'$, $e = (e'_o, e'_1)'$ and $u = (u'_o, u'_1)'$ are the vectors of BLUS, l-s residuals and the error term, respectively.

The problem of the best choice of the basis X_o is, however, mainly related to the power of the corresponding test based on BLUS residuals. T h e i l [8] proposed to choose such a basis by the selection of, a so called, permitted set of bases with respect to the given testing problem, using a minimum expected residual sum of squares criterion. The "permitted set of bases" ought to be chosen in such a way, that the basis observations ought to have "less information value" with respect to the alternative hypothesis, than the remaining observations. In the case of testing for serial correlation, for instance, such a set of "permitted bases" consists of the bases that contain the first m and last k - m cases, where $0 \leqslant m \leqslant k$. Philips and Harvey [5] indicate, however, that there do not exist uniformely best bases for all alternative hypotheses. It is possible only, by proper choice of the basis, to succeed in avoiding tests which have relatively low powers.

In the same work a comparison of the exact tests for serial correlation based on BLUS and recursive residuals is made. The recursive residuals are another type of LUS residuals that can be obtained by using the Cholesky decomposition of the matrix $(I - X_1(X'X)^{-1}X'_1)^{-1}$ to find C_1 in (8a). This type of residuals seems to be specially attractive due to the simplicity of the recursive computations (Philips and Harvey [5]). The tests for serial correlation based on recursive residuals, are only a bit less powerful than the BLUS tests.

4. CHOICE OF THE BASIS IN THE PRESENCE OF OUTLIERS

The fact that the LUS residuals estimated only n - k components of the vector u blurs the relationship between residuals and cases much more than it is in the case of the l - s residuals. In certain cases, especially when there are outliers among sample data, this can have serious implications on the inference based on LUS residuals. In such cases also the choice of the proper basis seems to be more important.

In the case of l - s estimation the effect of the outliers is spread, by means of the projection matrix M, to all residuals. Denoting the ij - th element of the matrix $H = X(X'X)^{-1}X'$ by h_{ij} the i - th l - s residual can be written in the form,

$$e_i = (1 - h_{ii})y_i - \sum_{j \neq i} h_{ij}y_j \tag{21}$$

where due to the indempotency of H, $h_{ii} = \sum_{j=1}^{n} h^2_{ij}$ and $h_{ii} \geq 1/n$ provided the model contains a constant term.

Hence if h_{ii} is close to 1 a gross error in y_i will not necessarily show up in e_i, but it might show up elsewhere, say in e_k, if h_{ki} happens to be large and it affects all the residuals. The same effect is present in LUS residuals, but now it is strongly dependend not only on the structure of the matrix X and character of contamination but also on the choice of the basis. This is evident in the case of recursive residuals. The contamination of one of basis observations will manifest itself by increasing residual errors. This influence decreses with n. On the other hand the occurrence of contamination in the observation used at the end of the recurrent procedure will influence only the last residuals. From this point of view the recursive residuals are appropriate for examining assumptions that depend on the order of the cases. In the case of BLUS residuals the proper choice of the basis when there are outliers among the sample data is not so evident but also seems to be more important than in the normal case. Generally the basis of the LUS-transformation should not contain any outlying observations.

In order to give some evidence about the possible influence of outliers on BLUS residuals with respect to the choice of the basis we make some simple numerical experiments. Each experiment consisted of 500 replications. In each replication we generated 15-elements sample from the "mean shift outlier model" with given magnitude and configuration of a single outlier,

$$Y = X\beta + \Phi_i\ (\alpha\sigma) + u$$

where $\Phi_i\ (\alpha\beta)$ denotes the dummy variable with contamination constant $\alpha\sigma$ in i-th position and zeros elsewhere and $u \sim N(0, \sigma\ )$.

The contamination constant was equal 5σ and 10σ and was added to the first, central and last observation in sample respectively. For each sample the values of BLUS residuals with four different bases were calculated and the values of the mean square estimation errors (MSE) were compared. The bases for BLUS computations consist of the firs k (BLUSb), the central k (BLUSc) and the last k (BLUSe) observations as well as the k observations corresponding to zero values of least absolute deviation residuals (BLUSl). The first three criterions were often used in literature (see eg. T h e i l [8]), the last one is proposed here mainly because of its robustness to outliers.

The mean values of the MSE over 500 replications for each type of BLUS residuals are given in Table 1. This table contains also the fractions of the replications in which the corresponding BLUS residual had the smallest MSE value.

T a b l e 1

Mean values of the MSE

Index of contaminated observation and the value of contamination		Type of BLUS residuals			
		BLUSb	BLUSc	BLUSe	BLUSl
1	5σ	1.0179 (0.08)	0.5001 (0.412)	0.5929 (0.172)	0.5124 (0.336)
8	5σ	0.7014 (0.494)	0.9557 (0.054)	0.7660 (0.288)	0.7783 (0.164)
15	5σ	4.1716 (0.552)	5.0073 (0.002)	18.1478 (0.000)	4.3454 (0.446)
1	10σ	1.8990 (0.036)	0.6486 (0.490)	0.7509 (0.180)	0.7217 (0.294)
8	10σ	1.1726 (0.518)	1,7468 (0.022)	1.2292 (0.344)	1.3146 (0.116)

N o t e: Mean values of the MSE over 500 replications of the experiment for four types of BLUS residuals and five variants of contamination. The values in brackets denotes the fraction of the replications in which the corresponding MSE value was the smallest one.

The results of experiments gathered in Table 1 indicates a relatively high increase in MSE in the case of base-choice with outliers. This increase depends on the relative magnitude of the outlier in comparison with other observations. Thus the proper choice of the basis for LUS computations when there are outliers in the data seems to be of special importance. The identification of outliers can be difficult in some cases. Taking this into account, the choice of the basis corresponding to zero least absolute deviations (LAD) residuals seems to be a good choice. This type of BLUS residuals (in Table 1 noted as BLUSl) can be directly obtained from n - k nonzero LAD residuals by means of the C'_1 transformation. Thus there exists mutual correspondence between LAD and BLUS1 residuals assuming k non-estimated BLUS residuals are equal zero.

The system of equations for the solution of the least absolute deviation problem can be written as,

$$\begin{pmatrix} y_o \\ y_1 \end{pmatrix} = \begin{pmatrix} X_o & 0 \\ X_1 & I \end{pmatrix} \begin{pmatrix} b_{LAD} \\ e_{LAD} \end{pmatrix} \tag{22}$$

where the subscriptsm 0 and 1 refer to the observations with zero and nonzero residuals respectively.

This corresponds to reordering the observations so that the first k are those lying in the regression hyperplane. Accordingly X_o and X_1 are respectively k : k and (n - k) : k matrices. The vector e_{LAD} refers to the n - k nonzero residuals.

Assuming nonsingularity of X_o the vector b_{LAD} of estimated parameters can be obtained from (22) and is equal to,

$$b_{LAD} = X_o^{-1} y_o \tag{23}$$

The vector of LAD residuals is then given by,

$$e_{LAD} = y_1 - X_1 b_{LAD} = y_1 - X_1 X_o^{-1} y_o \tag{24}$$

Given the ordering of the cases corresponding to the solution (24) the respective BLUS estimator of residuals can be written as,

$$\hat{e} = C'y = (C'_o C'_1) \begin{pmatrix} y_o \\ y_1 \end{pmatrix} = C'_o y_o + C'_1 y_1 = \tag{25}$$

$$= -C'_1 X_1 X_o^{-1} y_o + C'_1 y_1 = C'_1 (y_1 - X_1 X_o^{-1} y_o) = C'_1 e_{LAD}$$

Thus choosing the basis corresponding to zero LAD residuals we can obtain the vector of BLUS residuals directly from n - k nonzero LAD residuals. Note that in this case, the information contained in k non-estimated residuals is not spread into the remaining n - k residuals, which is the caes for BLUS residuals obtained from 1 - s residuals.

REFERENCES

[1] C o o k D. R., W e i s b e r g S., (1982), *Residuals and Influence in Regression*, Chapman and Hall, New York.

[2] K o e r t s J. (1967), *Some Further Notes on Disturbance Estimates in Regression Analysis*, J. Amer. Statist. Assoc., 1067-1079.

[3] K o e r t s J., A b r a h a m s e A. P. J. (1969), *On the Theory and Application of the General Linear Model*, Rotterdam University Press.

[4] N e u d e č k e r H. (1969), *A Note on BLUS Estimation*, J. Amer. Statist. Assoc., 949-952.

[5] P h i l i p s G. D. A., H a r v e y A. C. (1974), *A Simple Test for Serial Correlation in Regression Analysis*, J. Amer. Statist. Assoc., 935-939.

[6] T a y l o r L. (1974), *Estimation by Minimizing the Sum of Absolute Errors*, [in:] *Frontiers in Econometrics*, ed. P. Zarembka, New York 1974.

[7] T h e i l H. (1965), *The Analysis of Disturbances in Regression Analysis*, J. Amer. Statist. Assoc.

[8] T h e i l H. (1971), *Principles of Econometrics*, North Holland Publishing Co., Amsterdam.

Zbigniew Wasilewski

UWAGI O RESZTACH BLUS

W pracy rozważa się problemy dotyczące wyboru bazy przekształcania, prowadzącego do otrzymania estymatora wektora reszt m.n.k. o skalarnej macierzy wariancji kowariancji, w przypadku występowania obserwacji nietypowych. Zaproponowano wykorzystanie zerowych reszt otrzymanych w wyniku estymacji minimalizującej sumę odchyleń bezwzględnych do określania bazy tej transformacji. Umożliwia to uniknięcie wyboru obserwacji, nietypowych dla bazy, co znacznie poprawia jakość estymacji.