Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 6920

ON A NEW TEST FOR AUTOCORRELATION IN LEAST-SQUARES REGRESSION

by A.P.J. Abrahamse and A.S. Louter

August 5, 1969

Preliminary and Confidential

ON A NEW TEST FOR AUTOCORRELATION IN LEAST-SQUARES REGRESSION

by A.P.J. Abrahamse and A.S. Louter

Contents

1. INTRODUCTION

Recently, a new kind of estimators was derived for the disturbances in the general linear model, see [2]. They are best in the class of all linear unbiased estimators having some a priori specified idempotent covariance matrix of order n and rank n - k, where n is the number of observations and k the number of unknown parameters in the general linear model. It was shown that, in particular cases, they reduce to the BLUS estimators and the BLU estimators (the least-squares residuals), respectively. The new estimators have the convenient property that their distribution does not depend on the values taken by the explanatory variables which makes them particularly useful for a test on serial correlation of the disturbances. The BLUS residuals also possess this property but their covariance matrix must be chosen to be scalar which is an unnecessary restriction in view of testing for serial correlation; they are merely particular cases of the new estimators.

In [2] it was suggested that, in the relevant econometric applications, a test on serial correlation based on the new estimators, given an appropriate covariance matrix, might have a higher power than one based on BLUS estimators. This suggestion is supported by observations made by Hannan [4] and Theil-Nagar [7]. They pointed out that the regression vectors often behave in a particular way. The new estimators can probably be constructed so as to approximate this behaviour more accurately than the BLUS residuals. An additional advantage to the BLUS estimators is that, in using the proposed estimators, there is no ambiguity as to the choice of the basis since no disturbances must be put equal to zero (or something else) in advance.

The purpose of this paper is threefold. Firstly, in section 2, the authors recommend a particular procedure to compute the estimates. Secondly, the connection between the new estimators and the least-squares residuals is examined in section 3. Finally, in section 4, powers of a test against serial correlation are computed for some examples and compared with the corresponding quantities of the BLUS test and the Durbin-Watson test.

2. COMPUTATIONAL ASPECTS

We consider the general linear model

$$y = X\beta + u \tag{2.1}$$

where y is a vector of n values taken by the dependent variable, X an $n \times k$ matrix of values taken by the k non-stochastic explanatory variables, one of them being a constant term, β a vector of k unknown parameters and u a vector of unknown random variables (the disturbances) with mean zero and variance σ^2.

Let K be some $n \times (n - k)$ matrix satisfying $K'K = I_{(n-k)}$ and, let P be an $n \times (n - k)$ matrix whose columns are scaled eigenvectors corresponding to the unit roots of $M = I_{(n)} - X(X'X)^{-1}X'$. Then the vector of new estimators is given by

$$v = K(K'MK)^{\frac{1}{2}}(P'K)^{-1}P'y \tag{2.2}$$

where $(K'MK)^{\frac{1}{2}}$ is defined as QDQ', D being a diagonal matrix with the square roots of the eigenvalues of $K'MK$ on its main diagonal and Q an orthogonal matrix of corresponding eigenvectors.

It will now be shown that the matrix $(K'MK)^{\frac{1}{2}}$ can be computed in a simple way requiring only little time. The matrix $K'MK$ has $n - 2k$ eigenvalues equal to 1 if we assume $K'X$ to have rank k. For then $K'X(X'X)^{-1}X'K$ has rank k and $K'MK = I_{(n-k)} - K'X(X'X)^{-1}X'K$ has $n - 2k$ unit roots. We can thus write for $K'MK$

(2.3)
$$K'MK = \sum_{1}^{k} \delta_i^2 q_i q_i' + \sum_{k+1}^{n-k} q_i q_i'$$

where δ_i^2 $i = 1, \ldots, k$ are the eigenvalues of K'MK differing from one, q_i for $i = 1, \ldots, k$ corresponding eigenvectors, and q_i for $i = k + 1, \ldots, n - k$ eigenvectors corresponding to the remaining eigenvalues of K'MK. The δ_i^2 are all non-negative because K'MK is positive semi-definite.

The positive square root of K'MK is then

(2.4)
$$(K'MK)^{\frac{1}{2}} = \sum_{1}^{k} \delta_i q_i q_i' + \sum_{k+1}^{n-k} q_i q_i'$$

Making use of $\sum_{1}^{n-k} q_i q_i' = I_{(n-k)}$, (2.4) can be rewritten as

(2.5)
$$(K'MK)^{\frac{1}{2}} = I_{(n-k)} - \sum_{1}^{k} (1 - \delta_i) q_i q_i'$$

Writing L for the orthogonal $n \times k$ matrix orthogonal to K satisfying $LL' + KK' = I_{(n)}$, we shall prove the following lemma.

LEMMA 1. Let the eigenvalues of the $k \times k$ matrix $(X'L)^{-1}X'KK'X(L'X)^{-1}$ be denoted by λ_i and corresponding eigenvectors by ℓ_i $i = 1, \ldots, k$. Then $\delta_i^2 = 1/(1 + \lambda_i)$ and $q_i = K'X(L'X)^{-1}\ell_i/\surd\lambda_i$ $i = 1, \ldots, k$.

<u>Proof</u>: By multiplication and making use of $KK' + LL' = I_{(n)}$, it is easily verified that

(2.6)
$$I_{(n-k)} + K'X(X'LL'X)^{-1}X'K = (K'MK)^{-1}$$

Let ℓ_i be a scaled eigenvector of $(X'L)^{-1}X'KK'X(L'X)^{-1}$ corresponding to the eigenvalue λ_i. Then we have

(2.7)
$$(X'L)^{-1}X'KK'X(L'X)^{-1}\ell_i = \lambda_i \ell_i$$

Premultiplication of (2.7) by $K'X(L'X)^{-1}$ gives

(2.8)
$$K'X(X'LL'X)^{-1}X'K\{K'X(L'X)^{-1}\ell_i\} = \lambda_i\{K'X(L'X)^{-1}\ell_i\}$$

from which we conclude that

(2.9)
$$q_i = K'X(L'X)^{-1}\ell_i$$

is an (unscaled) eigenvector of $K'X(X'LL'K)^{-1}X'K = (K'MK)^{-1} - I_{(n-k)}$ (and thus also of K'MK) corresponding to the eigenvalue λ_i for $i = 1, \ldots, k$. It is also seen that

$$(2.10) \qquad \delta_i^2 = 1/(1 + \lambda_i) \qquad i = 1, \ldots, k$$

are the k eigenvalues of K'MK differing from one. Finally, the q_i are scaled by dividing them by

$$(2.11) \qquad \{\ell_i'(X'L)^{-1}X'KK'X(L'X)^{-1}\ell_i\}^{\frac{1}{2}} = (\lambda_i \ell_i'\ell_i)^{\frac{1}{2}} = \lambda_i^{\frac{1}{2}} \qquad i = 1, \ldots, k$$

which completes the proof.

Writing

$$Z = K'X(L'X)^{-1}$$

we obtain for the square root of K'MK

$$(2.12) \qquad (K'MK)^{\frac{1}{2}} = I_{(n-k)} - \sum_{1}^{k} \{\frac{1 - 1/(1 + \lambda_i)^{\frac{1}{2}}}{\lambda_i})(Z\ell_i)(Z\ell_i)'\}$$

Lemma 1 clearly simplifies the determination of the eigenvalues and eigenvectors of K'MK considerably. They can be obtained from those of a matrix of order k only by means of simple operations. Since k is often not more than 2 or 3, the process usually takes very little time.

For the matrix P such a simple procedure has not been found and its columns which are eigenvectors corresponding to the unit roots of $M = I_{(n)} - X(X'X)^{-1}X'$ must be computed by means of a standard method for the determination of eigenvalues and eigenvectors of a symmetric matrix. It was proved in [2] that the columns of P need not necessarily be eigenvectors of M. It is sufficient that they form an orthonormal basis for the space spanned by M.

The remaining operations required to obtain the estimators are elementary and simple.[1]

---

[1] A computer program is available at the Econometric Institute and can be obtained on request.

3. SOME PROPERTIES

In [2] it was shown that

(3.1) $$E(v) = E(u) = 0, \quad E(vv') = KK'\sigma \quad \text{and} \quad v'v = u'Mu$$

We shall now derive the average inaccuracy of the estimators v, defined as the ratio of the expected sum of squares of the estimation errors to the expected sum of squares of the disturbances to be estimated:

(3.2) $$I_v = \frac{E\{(v - u)'(v - u)\}}{E(u'u)}$$

Writing

(3.3) $$B' = K(K'MK)^{\frac{1}{2}}(P'K)^{-1}P'$$

we represent the error vector by

(3.4) $$v - u = B'u - u = (B' - I_{(n)})u$$

Its expected inner product is

(3.5) $$\begin{aligned} E\{(v - u)'(v - u)\} &= E\{u'(B - I_{(n)})(B' - I_{(n)})u\} \\ &= \sigma^2 \operatorname{tr} (BB' - B - B' + I_{(n)}) \\ &= \sigma^2(n - k + n - 2 \operatorname{tr} B) \end{aligned}$$

where use is made of

(3.6) $$\operatorname{tr} (BB') = \operatorname{tr} (B'B) = \operatorname{tr} (K'K) = \operatorname{tr} (I_{(n-k)}) = n - k$$

according to (3.3) and $K'K = I_{(n-k)}$.
The trace of the matrix B can be expressed as follows.

(3.7)
$$\text{tr } B = \text{tr } K(K'MK)^{\frac{1}{2}}(P'K)^{-1}P'$$
$$= \text{tr } (K'MK)^{\frac{1}{2}}(P'K)^{-1}(P'K)$$
$$= \text{tr } QDQ'$$
$$= \text{tr } D$$
$$= \sum_{1}^{n-k} \delta_i$$

which is equal to

(3.8)
$$\text{tr } B = n - 2k + \sum_{1}^{k} \delta_i$$

since δ_i $i = k + 1, \ldots, n - k$ are equal to 1.

Combining (3.5) and (3.8), we obtain

(3.9)
$$E\{(v - u)'(v - u)\} = \sigma^2(3k - 2\sum_{1}^{k} \delta_i)$$

Dividing (3.9) by $E(u'u) = n\sigma^2$, we obtain

(3.10)
$$I_v = \frac{k}{n} - \frac{2}{n}\sum_{1}^{k}(1 - \delta_i)$$

The expected sum of squares of the estimation errors of the least-squares estimators is

(3.11)
$$E\{(\hat{u} - u)'(\hat{u} - u)\} = E\{u'(M - I_{(n)})(M - I_{(n)})'u\}$$
$$= \sigma^2 \text{ tr } (M - 2M + I_{(n)})$$
$$= \sigma^2(n - k - 2n + 2k + n) = \sigma^2 k$$

It follows that the average inaccuracy of $\hat{u}$ equals

(3.12)
$$I_{\hat{u}} = \frac{E\{(\hat{u} - u)'(\hat{u} - u)\}}{E(u'u)} = \frac{\sigma^2 k}{\sigma^2 n} = \frac{k}{n}$$

We shall call the ratio of (3.12) to (3.10) "the efficiency of v with respect to $\hat{u}$:[2]

[2] This definition is analogous to one given by Koerts for the BLUS estimators in [5].

$$E = \frac{I_{\hat{u}}}{I_v} = \frac{k}{k + 2 \sum_1^k (1 - \delta_i)} \qquad 0 < E \leq 1 \tag{3.13}$$

Further, we shall determine the expected sum of squares of the differences between the estimators v and the least-squares residuals $\hat{u} = Mu$ as follows.

$$\begin{aligned} E\{(v - \hat{u})'(v - \hat{u})\} &= E\{u'(B - M)(B - M)'u\} \\ &= \sigma^2 \text{ tr } (BB' - BM - MB' + M) \\ &= \sigma^2\{2(n - k) - 2 \text{ tr } B\} \\ &= 2\sigma^2(k - \sum_1^k \delta_i) = 2\sigma^2 \sum_1^k (1 - \delta_i) \end{aligned} \tag{3.14}$$

where use is made of (3.6), (3.8), and $B'X = 0$ which follows from $E(v) = 0$. The efficiency is equal to 1 if the new estimators coincide with the least-squares residuals, which occurs if the X-matrix happens to be orthogonal to the a priori chosen matrix K. In other cases it is smaller than one; this is the price to paid for the additional restriction of a certain specified covariance matrix $KK'\sigma^2$. The more M differs from KK', the lower the efficiency. E can be useful for analyzing X-matrices as indicated at the end of the next section.

An interesting property is that (3.9) is equal to the sum of (3.11) and (3.14). Thus the expected sum of squares of estimation errors of the new estimators equals the sum of the expected sum of squares of estimation errors of the least-squares estimators.

4. TESTING FOR SERIAL CORRELATION

It is known to be important that the covariance matrix of estimators on which a test for serial correlation is based be independent of the values taken by the explanatory variables. Only in that case tabulation of the test statistic's significance points makes sense. The BLUS estimators do have such a covariance matrix and lead to Hart's table with significance points of the Von Neumann ratio. However, it has been shown in [1] that the "BLUS power" may be considerably lower than the "Durbin-Watson" power. One additional disadvantage of the BLUS estimators is that a number of disturbance estimators (equal to the number of explanatory variables including the constant term) must a priori be chosen equal to zero which gives the BLUS estimators a more or less ambiguous character.

The difference in power between "BLUS" and "Durbin-Watson" justifies an attempt to find better estimators than the BLUS residuals whose covariance matrix remains independent of the values taken by the explanatory variables. For this purpose we can use the new estimators based on an appropriate covariance matrix $\sigma^2 KK'$. In the case of the BLUS estimators KK' is required to be of the form $I_{(n-k)}$ filled up with zeros so that the class of admitted covariance matrices is unnecessarily restricted.

The matrix KK' is completely determined by K, which should be chosen so that, on the average, $\sigma^2 KK'$ approximates the covariance matrix $\sigma^2 M$ of the (best linear unbiased) least-squares estimators as well as possible. Of course, this makes only sense if, in the relevant applications, M and hence X show some regularity. In other words, the X-matrices of distinct models should fluctuate around some "mean X-matrix" with not too large variation. K should then be based on this "mean X-matrix".

Does such an X-matrix exist? It seems so. On the basis of remarks by Hannan and Theil-Nagar it was pointed out in [2] that such a matrix may be given by

$$(4.1) \qquad L = [h_1, \ldots, h_k]$$

where

$$(4.2) \qquad h_i = \frac{1}{C_i} [\cos \frac{\pi(i-1)}{2n} \cos \frac{3\pi(i-1)}{2n} \ldots \cos \frac{(2n-1)\pi(i-1)}{2n}]' \quad i = 1, \ldots, n$$

where $C_1 = n^{\frac{1}{2}}$ and $C_i = (n/2)^{\frac{1}{2}}$ for $i = 2, \ldots, n$.

If, on the average L approximates the X-matrix sufficiently well, the same holds for $KK' = I_{(n)} - LL'$ and $M = I_{(n)} - X(X'X)^{-1}X'$. So we should choose K as follows.

$$(4.3) \qquad K = [h_{k+1}, \ldots, h_n]$$

In [2] it was shown that the distribution of the Durbin-Watson statistic, based on the new estimators with K as specified in (4.3), is that of the Durbin-Watson upper bound and, hence, tables with significance points are available.

For illustrative purposes we shall compute the powers of a test against positive serial correlation based on v with K as specified in (4.3) and L as specified in (4.1) for some examples, and compare them with the corresponding powers of "BLUS" and "Durbin-Watson".

We use the test statistic

$$Q = \frac{\sum_{2}^{n} (v_t - v_{t-1})^2}{\sum_{1}^{n} v_t^2} \qquad (4.4)$$

where $v = \{v_t\} = K(K'MK)^{\frac{1}{2}}(P'K)^{-1}P'y$ with $(K'MK)^{\frac{1}{2}}$ computed as indicated in (2.12). Appropriate significance points are taken from the Durbin-Watson upper-bound table.

The two examples are the "Textile example" as described in [6] which has 15 observations on 2 explanatory variables (k = 3), and the "spirit example" consisting of the first 15 observations of the example presented by Durbin-Watson in [3], here also k = 3. In both examples the probability of a Type-I error is taken equal to 0.05 and the alternative hypothesis is a first-order Markov scheme $u_t = \rho u_{t-1} + \varepsilon_t$ with the ε_t independent and normally distributed with mean zero and variance σ^2. For three different values of ρ the powers are computed by means of a method described in [6]. They are presented in the table below.

Powers in two examples

Textile example n = 15, k = 3, P(I) = 0.05

ρ	Durbin-Watson	New test	BLUS[3]
.3	.19	.19	.14
.6	.40	.39	.30
.8	.51	.50	.41

Spirit example n = 15, k = 3, P(I) = 0.05

ρ	Durbin-Watson	New test	BLUS[3]
.3	.20	.19	.16
.6	.46	.43	.37
.8	.61	.56	.52

[3] In both examples the last three disturbance estimates are a priori put equal to zero.

The results show that in both examples the powers of the new test lie between those of "Durbin-Watson" and "BLUS", most of the difference between the latter being regained by using the new estimators. Hence, to a certain extent we can say that the X-matrices in both examples "behave like L".

The authors are conscious of the fact that the value of the procedure depends on the question whether or not the X-matrices in economic models where testing for serial correlation is necessary, generally behave in this way. As said before remarks by Hannan and Theil-Nagar point to an affirmative answer to this question. Moreover, the Textile example and the Spirit example are both arbitrary models containing time series and thus cases where testing for serial correlation is needed.

A thorough examination of the X-matrices which can arise in models capable for testing against serial correlation can provide a more definite answer. This can be done by determining for distinct relevant X-matrices (either generated or taken from the literature) the efficiency E, defined in (3.14) and see whether or not it shows large fluctuations.

It goes without saying that K may be taken different from that which was adopted here e.g. by choosing some of the other columns h_i. In particular h_1 corresponds to the constant term in the regression so that for a regression without a constant term this vector may perhaps better be replaced by a different one. However, the distribution of Q will then no longer be that of the Durbin-Watson upper bound and a new table of significance points is needed. Such a table can be constructed by means of a procedure described in [1, 6].

REFERENCES

[1] Abrahamse, A.P.J., and J. Koerts (1969)."A Comparison between the Power of the Durbin-Watson Test and the Power of the BLUS Test".Journal of the American Statistical Association, 64.

[2] Abrahamse, A.P.J., and J. Koerts (1969)."New Estimators of Disturbances in Regression Analysis". Report 6906 of the Econometric Institute of the Netherlands School of Economics. Rotterdam.

[3] Durbin, J. and G.S. Watson (1951)."Testing for Serial Correlation in Least-Squares Regression, II".Biometrika, 38, 159-178.

[4] Hannan, E.J. (1960). _Time Series Analysis_. London, New York.

[5] Koerts, J. (1967)."Some Further Notes on Disturbance Estimates in Regression Analysis". _Journal of the American Statistical Association_, 62, 169-183.

[6] Koerts, J. and A.P.J. Abrahamse (1968). "On the Power of the BLUS Procedure". Journal of the American Statistical Association, 63, 1227-1236.

[7] Theil, H., and A.L. Nagar (1961)."Testing the Independence of Regression Disturbances". _Journal of the American Statistical Association_, 56, 793-806.