Erasmus University Rotterdam
Econometric Institute

Report 7604/S

On typical characteristics of economic time series and the relative qualities of five autocorrelation tests

by C. Dubbelman, A.P.J. Abrahamse, and A.S. Louter

february, 1976

1. Introduction

In 1971 an article was published presenting a new estimator for the disturbance vector in the linear model, see Abrahamse and Koerts (1971). This estimator was developed for the purpose of constructing a tabulable test on autocorrelation in the disturbances, u, of the general linear model $y = X\beta + u$.

In a following article, Abrahamse and Louter (1971), some properties of the new estimator were examined and a test on autocorrelation, based on the new estimator, was evaluated on some examples stemming from the field of economic time series.
The powers appeared to be high in comparison with those of some competing procedures.

The test statistic involved is the well-known Durbin-Watson (DW) statistic, with the least-squares residual vector replaced by the new estimator. This modification makes the distribution of the DW statistic, under the null hypothesis of no autocorrelation, independent of the regressors. This happens because the new estimator, besides being linear in y and unbiased, possesses a fixed, a priori chosen, covariance matrix. Different choices for this covariance matrix imply different estimators.

The estimator is best in the sense that the "distance" from the (best linear unbiased) least-squares residual vector is minimized. This distance will vary, however, if the same estimator (having the same given covariance matrix) is used in different sets of data.
In the two basic articles mentioned above, it was suggested that the quality of a test based on the new estimator would only be attractive if the distance from the least-squares residual vector were not too great. Then it is important to choose the covariance matrix such that,

for the sets of data where the new estimator is used, the distance between the least-squares residual vector and the new estimator is small, on the average.
Obviously, the use of the new estimator in a given category of sets of data would only be relevant if the distance within that category would show little variation. Of course, in the case of much variation, one could decompose the category into subcategories and use different covariance matrices and, thus, different new estimators for the different subcategories. But this would require the computation of a table of significance points for each subcategory. In fact there is no end; one could go so far as to define an estimator for each particular set of data, which would then equal the least-squares residual vector. And this is precisely what we do not want. For instance, when we want to test for autocorrelation in economic time series problems, we prefer to use a standard procedure with one single table of significance points. Evidently, there should be a sufficient amount of regularity in the behaviour of regressors ("X-matrices") in order to form a category in the above sense.

Now it was known that economic time series contain some regularity and in Abrahamse and Koerts (1971) a particular new estimator was given which should be close to the least-squares residual vector in problems involving economic time series. This estimator was specified by fixing its covariance matrix in such a way that the DW statistic based on this estimator has the distribution of the Durbin-Watson upper bound.

For some examples, the powers of an autocorrelation test based on the new estimator have been given in Abrahamse and Louter (1971), and they were compared with the corresponding powers of some customary test procedures. In these examples, the powers of the new test appear to be relatively high.

In the course of time, we have collected several X-matrices appearing in econometric literature and we have used them to compute powers of the new procedure and of competing procedures. The almost unanimous result is that the new procedure is superior to the BLUS procedure and also to "Durbin's exact alternative to the bounds test". These results are published in Dubbelman (1972).

Now l' Esperance and Taylor (1975) have written an article in which they come to quite different results. There are many inconsistencies between our results and those by l' Esperance and Taylor. Some of them, in particular those referring to Durbin's exact test, are probably due to an incorrect derivation of the formulae needed to compute powers. Most of them, however, can be explained by a basic difference in experimental design. Whereas we always took X-matrices from the literature, l' Esperance and Taylor generated X-matrices by simulating columns from a first-order autoregressive scheme. We believe that the particular scheme used by l' Esperance and Taylor is not capable of generating series having the basic characteristics of economic time series. The line of our article is as follows. In Section 2, we first make critical remarks on some propositions and derivations appearing in the article by l' Esperance and Taylor and, accordingly, we correct some errors. All five tests we are considering can be brought into a general framework. This is done in Section 3.1 by defining a general test statistic Q which serves as a parameter for selecting the particular test from the general framework. In Section 3.2 the distribution of Q is discussed, in particular the way in which power functions can be computed. In fact, this section contains a concise summary of known results and brings in only a few new elements. We feel that it is now time to give such a summary.

In Section 4, the qualities of the tests are considered from a theoretical point of view. In Section 5, we discuss the typical characteristics of sets of data consisting of economic time series and a design to generate such kind of data is proposed. Finally, in Section 6, the results of a number of power calculations for five tests, based on this design, are presented. The results appear to affirm the conclusions mentioned in Abrahamse and Louter (1971) and, accordingly, the conclusions drawn by l' Esperance and Taylor should be revised radically.

2. Comments and corrections on the article by l' Esperance and Taylor

(i) On page 2 of their article, l' Esperance and Taylor (to be called "the authors" henceforth) claim to review an alternative BLUE approach to derive the AL estimator. However, their approach is not really an alternative. The derivation they give is incomplete and contains a superfluous part, namely, that where the Lagrangian multipliers are used. The result in their formula (14),

$$B'(M-I) = -B'X(X'X)^{-1}X' = 0$$

follows immediately from the restriction $B'X = 0$ given in (8) (i). Thus, the interjacent derivation is redundant. The rest of the authors' proof, as far as it is presented, has already been given in Abrahamse and Koerts (1971). Formula (3) should read $KK' = \Omega$.

(ii) At the bottom of page 5, the authors make the proposition that any BLUS estimator entails a loss of degrees of freedom so that, for small samples, the testing procedure would become less powerful. This proposition rests on a misconception. Both a BLUS estimator and the least-squares residual vector possess (T-k)-dimensional distributions. That this is true for the least-squares residual vector can immediately be concluded by considering the rank of its covariance matrix $\sigma^2 [I-X(X'X)^{-1}X']$, which equals T-k. Thus, both estimators have the same number of degrees of freedom.

(iii) At the top of page 6, the authors say that, if K is specified as a subset of eigenvectors of A corresponding to the k' smallest non-zero eigenvalues of A, then the probability distribution of the DW statistic based on the AL estimator is the same as that of the DW upper bound. This is not true as can immediately be concluded by considering the rank (T-k) of K. In fact, it was shown in Abrahamse and Koerts (1971) that, for this purpose, K should contain eigenvectors corresponding to the T-k largest eigenvalues of A.

(iv) According to the authors, the eigenvalues of A would be $2[1-\cos(\pi j/T-1)]$, $j = 0, \ldots, T-1$. In this formula -1 should be omitted, see (2) in the present paper.

(v) In their Appendix, the authors say to derive the power functions of three test procedures. This suggests that they were not aware of the fact that the procedure outlined in Abrahamse and Koerts (1969) is completely general. It is obvious that the formula $(A - q^*E)C'\Omega C$, which is denoted by (3.4) in that article, and which refers to formula (2.4) in the same article:

$$Q^* = \frac{u'CAC'u}{u'CEC'u}, \qquad u \sim N(0, \Omega),$$

is valid for any matrix C, provided C is independent of y[1]. Q* is the DW statistic based on a disturbance estimator C'y. Thus to compute powers for the exact DW test, the only thing one has to do is to replace C by

$$M = I - X(X'X)^{-1}X'.$$

For the AL estimator one should choose

$$C = K(K'MK)^{-\frac{1}{2}}K'M,$$

and so on. If Ω is replaced by the identity matrix I, one obtains significance points.

This method has been used to compute powers and significance points for BLUS, DW's exact test, and other procedures in several of the articles mentioned in the reference list. Thus, the derivations by l' Esperance and Taylor do not add new results.

(vi) Because the procedure of approximating power functions is so general, the procedure for Durbin's exact alternative was not spelled out in Abrahamse and Louter (1971), though, in that article, it is mentioned that the power has been computed for certain cases. Perhaps this was an unwise decision because it may be a problem to write Durbin's

----

1) In the rest of this article, the symbols C and Ω are no longer used in this sense. They are replaced by the symbols B and Γ, respectively.

estimator as a linear form in y. Whatever it may be, the derivation given by l' Esperance and Taylor in their Appendix (c) is incorrect and, consequently, their power computations are wrong. They slip in assuming equivalent distributions for Durbin's transformed disturbances z and $\hat{w}$ as defined in their formula (48), while this is only true under the null hypothesis of no autocorrelation, as Durbin has shown. In particular, formula (56) is senseless, since it is independent of the regression matrix, which is impossible because, then, the power function would be independent of the regression matrix.

In this connection, the difference between the results for Durbin's exact test for $\lambda = 0.1$ and $\lambda = 0.7$ as presented by the authors in their tables is inexplicable.

In the appendix to the present article Durbin's implied estimator is written as a linear form in y.

(vii) In the summary of their article, the authors mention that the parameters of their regression model are presumed to be known. The sense of this is not clear since the power functions of all test procedures investigated are completely independent of the parameters. The fact that the authors always mention the β-parameters explicitly in the heads of their tables suggests that they influence the results, which should apparently be untrue, see our Section 3.2.

3. The form and the distribution of the test statistic Q

3.1. Definition of Q

It is assumed that y has a normal distribution with mean $X\beta$ and covariance matrix $\sigma^2\Gamma$, where X is an $n \times k$ matrix with rank k, specified by observation, where β is a k-element vector of unknown constants, σ^2 is an unknown scalar constant, and Γ is the $n \times n$ matrix

$$(1)\quad \Gamma = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \vdots & & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 \end{bmatrix},$$

$$\Gamma^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

We wish to test for H_0 (the null hypothesis): $\rho = 0$ against H_A (the alternative hypothesis): $\rho > 0$. We consider five tests, all of which have a critical region of the form [1)]

$$Q = \frac{(B'y)'A(B'y)}{(B'y)'E(B'y)} \leq q$$

for some matrix B, where q is determined by $\Pr[Q \leq q | H_0] = \alpha$, the significance level of the test. Here A and E are the matrices

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \quad \text{and } E = I - \iota(\iota'\iota)^{-1}\iota'$$

where $\iota = [1 \; 1 \; 1 \; \ldots \; 1]'$. The order of A and E is $n \times n$ in all tests but the BLUS test, where the order is $(n-k) \times (n-k)$. Usually, most of the tests under consideration are not defined with E in the denominator of Q. However, we restrict the present analysis to X-matrices including a constant term, and

---

1) When testing against H_A: $\rho < 0$, the critical region has the form $Q > q$ with q determined by $\Pr[Q \leq q \mid H_0] = 1 - \alpha$.

in this case E can be inserted, since the constant term implies EB' = B', as we shall see. The matrices A and E are related as follows:

$$A=H^*\begin{bmatrix} d_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & d_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & d_3 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & d_{n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & d_n \end{bmatrix}H^{*\prime} \qquad E=H^*\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}H^{*\prime}$$

where $H^{*\prime} = H^{*-1}$, D is the diagonal matrix with i-th diagonal element d_i,

(2) $$d_i = 2\,[1 - \cos\{\pi(i-1)/n\}]$$

so that we have $0 = d_1 < d_2 < \dots < d_n < 4$, and where the j-th element of h_i^*, the i-th column of H^*, is

$$h_i^*(j) = c\cos\{\pi(i-1)(j-\tfrac{1}{2})/n\} \qquad c = \sqrt{\tfrac{1}{n}} \text{ if } i = 1$$
$$c = \sqrt{\tfrac{2}{n}} \text{ if } i = 2,3,\dots,n$$

It is easily verified that AE = A.

As we just said, five tests will be considered, which are denoted by the abbreviations: test (D-W), test (BLUS), test (A-L), test (DURA), and test (sel). The matrices B' in these tests are as follows:

test (D-W) $B' = M = I_{(n)} - X(X'X)^{-1}X'$. This is the well-known exact Durbin-Watson test, based on $\hat{u} = M'y$.

test (BLUS) $B' = (J'MJ)^{-\frac{1}{2}}J'M$, where J is an $n \times (n-k)$ matrix obtained from $I_{(n)}$ by deleting k of its columns. This test is based on the BLUS estimator of J'u. The matrix J is chosen in accordance with the selection device given in Theil (1971), p. 218. See also Appendix A.

test (A-L) $B' = K(K'MK)^{-\frac{1}{2}}K'M$, where K is the $n \times (n-k)$ matrix

$[h^*_{k+1} \vdots h^*_{k+2} \vdots \ldots \vdots h^*_n]$. This test is based on the new estimator of u, which was introduced in Abrahamse and Koerts (1971).

test (DURA): $B' = KK'[I - \tilde{X}(\tilde{X}'KK'\tilde{X})^{-1}\tilde{X}'KK' + \tilde{X}P_1P_2'\tilde{P}']M$, where K is equal to K in test (A-L), $\tilde{X}$ is the $n \times (k-1)$ matrix obtained from X by deleting the constant term column, $\tilde{P}$ is the $n \times (k-1)$ matrix $[h^*_2 \vdots h^*_3 \vdots \ldots \vdots h^*_k]$, and P_1 and P_2 are lower triangular matrices such that $P_1P_1' = (\tilde{X}'KK'\tilde{X})^{-1}$ and $P_2P_2' = (\tilde{P}'M\tilde{P})^{-1}$. This is Durbin's exact alternative to the bounds test, introduced in Durbin (1970). The formula of B' has not been given by Durbin, but it can be derived from his formula for the calculation of z, the vector of regression residuals, see Appendix B. The above formula has been presented and used in Dubbelman (1972).

test (sel) B' is equal to B' in test (A-L) apart, incidentally, from the specification of the $n \times (n-k)$ matrix K. In this test K consists of n-k h*-vectors which are selected after the X-matrix has become known. The selection device has been introduced in Dubbelman (1972). It reads: compose K of those n-k vectors h^*_i for which $h^{*\prime}_i M h^*_i$, $i = 1,2,\ldots, n$, takes the greatest values. For instance, when X contains a constant term, then $h^{*\prime}_1 M h^*_1 = 0$ and K does certainly not include h^*_1. Quite often, test (sel) and test (A-L) coincide.

Test (BLUS) is defined with E in the denominator of Q. In all other tests E can be omitted, since $EB' = B'$. (Note that $\iota' M = 0$ since $X'M = 0$ and X contains ι, and that $\iota' K = 0$ since $\iota = \sqrt{n} h^*_1$ and $h^{*\prime}_1 h^*_i = 0$ for $i = 2,3, \ldots,n$. Note also that $B'X = 0$, in all five tests.)

3.2. The distribution of Q

The probability distribution function of Q is

$$F(q) = \Pr\,(Q \leq q) = \Pr\left[\frac{y'BAB'y}{y'BEB'y} \leq q\right]$$

$$= \Pr\,[y'BAB'y \leq q\, y'BEB'y] = \Pr\,[y'B(A-qE)B'y \leq 0]$$

where $y \sim N(X\beta, \sigma^2\Gamma)$ and B satisfies $B'X = 0$. Let us define F and Δ by $B'\Gamma B = F\Delta F'$ such that $F'F = I$ and Δ is diagonal and nonsingular. In other words, the diagonal elements of Δ are the nonzero (positive) eigenvalues of $B'\Gamma B$ and the columns of F are the corresponding eigenvectors. Denoting the space spanned by the columns of a matrix X by $M(X)$, we have $M(F) = M(F\Delta^{\frac{1}{2}}) = M(F\Delta F') = M(B'\Gamma B) = M(B')$, since both Δ and Γ are nonsingular. Hence, the vector $B'y$ lies in $M(B') = M(F\Delta^{\frac{1}{2}})$, while the columns of $F\Delta^{\frac{1}{2}}$ form a basis of this space. It follows that $B'y = F\Delta^{\frac{1}{2}}v$ for some unique vector v for given y. We obtain

$$F(q) = \Pr\,[v'\Delta^{\frac{1}{2}}F'(A-qE)F\Delta^{\frac{1}{2}}v \leq 0]$$

where $v = \Delta^{-\frac{1}{2}}F'B'y \sim N(0,\sigma^2 I)$. Writing $\Delta^{\frac{1}{2}}F'(A-qE)F\Delta^{\frac{1}{2}} = H\Lambda H'$ with $H' = H^{-1}$ and Λ diagonal, and defining $z = \frac{1}{\sigma}H'v$, we have

$$(3) \qquad F(q) = \Pr\,[z'\Lambda z \leq 0] = \Pr\,[\Sigma\lambda_i z_i^2 \leq 0]$$

where $z \sim N(0,I)$ and the λ_i are the eigenvalues of $\Delta^{\frac{1}{2}}F'(A-qE)F\Delta^{\frac{1}{2}}$. Below it is proved that the nonzero λ_i are the eigenvalues of $S'B(A-qE)B'S$, where S is a square matrix such that $SS' = \Gamma$. For instance, given Γ in (1), one may take

$$(4) \qquad S = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \rho & a & 0 & \dots & 0 & 0 \\ \rho^2 & a\rho & a & \dots & 0 & 0 \\ & & & & & \\ \rho^{n-1} & a\rho^{n-2} & a\rho^{n-3} & \dots & a\rho & a \end{bmatrix} \qquad a = \sqrt{1-\rho^2}$$

Under H_0: $\Gamma = I$, the calculations can be simplified, as we list below.

From $B'\Gamma B = B'SS'B = F\Delta F'$ it follows that $B'S = F\Delta^{\frac{1}{2}}V'$, since $M(B'S) = M(B'SS'B) = M(F\Delta F') = M(F\Delta^{\frac{1}{2}})$, and V satisfies $V'V = I$, since $B'SS'B = F\Delta^{\frac{1}{2}}V'V\Delta^{\frac{1}{2}}F' = F\Delta F'$ where $F'F = I$ and $\Delta^{-\frac{1}{2}}$ exists.
Then $S'B(A-qE)B'S = V\Delta^{\frac{1}{2}}F'(A-qE)F\Delta^{\frac{1}{2}}V' = (VH)\,\Lambda\,(VH)'$ where $(VH)'(VH) = I$, so that Λ contains all nonzero eigenvalues of $S'B(A-qE)B'S$. In all tests,

the calculation of $\mathcal{F}(q|H_A)$ requires the complete calculation of the matrix S'B(A-qE)B'S and the determination of its (at most n-k) nonzero eigenvalues. In test (D-W), the calculation of $\mathcal{F}(q|H_0)$ requires the determination of the eigenvalues of M(A-qE)M (here B' = M and S = I), so that the distribution depends on X by means of M. In test (BLUS), we have $B'\Gamma B = I_{(n-k)}$ under H_0, so that we may take $F = \Delta = I_{(n-k)}$ and hence $\mathcal{F}(q|H_0)$ depends on the eigenvalues of A-qE. In accordance with (2), these eigenvalues are

$$\lambda_1 = 0;\ \lambda_i = 2\,[1 - \cos\{\pi(i-1)/(n-k)\}] - q, \qquad i = 2,3,\ldots, n-k$$

In test (A-L), test (DURA), and test (sel), we have $B'\Gamma B = KK'$ under H_0, so that we may take F = K and $\Delta = I_{(n-k)}$ and hence $\mathcal{F}(q|H_0)$ depends on the eigenvalues of $K'(A-qE)K = K'AK-qI_{(n-k)}$ (since $\iota'K = 0$ in all cases). In test (A-L) and test (DURA) we have $K = [h^*_{k+1} \vdots h^*_{k+2} \vdots \ldots \vdots h^*_n]$, so that

$$\lambda_i = d_{k+i} - q, \qquad i = 1,2,\ldots, n-k$$

In test (sel) we have

$$0 \leq h^{*\prime}_{j_1} M h^*_{j_1} \leq h^{*\prime}_{j_2} M h^*_{j_2} \leq \ldots \leq h^{*\prime}_{j_n} M h^*_{j_n}$$

say, and hence

$$\lambda_i = d_{j_i} - q, \qquad i = k+1, k+2, \ldots, n$$

(A choise problem could arise when $h^{*\prime}_{j_k} M h^*_{j_k} = h^{*\prime}_{j_{k+1}} M h^*_{j_{k+1}}$, so that it is not clear whether $h^*_{j_k}$ or $h^*_{j_{k+1}}$ must be included into K. In this case we recommend to include $h^*_{j_{k+1}}$ if $j_{k+1} > j_k$, and to include $h^*_{j_k}$ otherwise.)

4. The quality of the tests theoretically seen

For arbitrary X, none of the five tests can claim theoretical superiority in a power sense. Test (D-W) has the best theoretical recommendations: for a restricted class of X-matrices, namely all X-matrices whose k columns are linear combinations of k h^*-vectors, this test is uniformly most powerful similar when testing for H_0: $\rho = 0$ against H_A: $\rho > 0$ in

$$\Gamma^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1+\rho^2-\rho & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1+\rho^2-\rho \end{bmatrix}$$

which matrix is close to Γ^{-1} in (1). Also, this test is uniformly most powerful invariant in the neighbourhood of $\rho = 0$. Berenblut and Webb (1973) constructed a uniformly most powerful invariant test in the neighbourhood of $\rho = 1$. The power of their test is, roughly speaking, 1.05 times the power of test (D-W) at $\rho = 0.9$, while the difference between the power functions is very small at $0 < \rho < 0.7$.

In this paper, the power function of test (D-W) is regarded as indicating the maximum power attainable. Therefore the powers of this test are included in Tables 3, 4 and 5. The other four tests claim some optimality in an estimation sense, i.e. they approximate test (D-W) either in the sense that B'y is close to (a subvector of) $\hat{u}$, like test (BLUS), or in the sense that the covariance matrix of B'y is close to the covariance matrix of $\hat{u}$, which is $\sigma^2 M$, like test (DURA), or both, like test (A-L) and, a fortiori, test (sel).

For the comparison of the powers of the tests, it is necessary to adopt several X-matrices and to calculate the powers for several values of ρ, for each of these X-matrices. The choice of the X-matrices is the subject of the following section. We emphasize that knowledge of β or a sample y within this context is redundant.

5. Typical characteristics of economic time series.

We wish to evaluate the five tests by comparing their powers in applications to economic time series. To this end one or more X-matrices must be adopted, which can be regarded as representative for economic time series. Unfortunately, no analytical rule can be given to decide whether a series does or does not belong to the class of economic time series. It must be learned from practice, which properties are typical of economic time series X-matrices. Before turning to the relevant remarks, made in the literature, we observe that in all tests but test (DURA) the matrix B' depends on X by means of M alone. Let R be an $n \times k$ orthogonal matrix such that $X = RG$, where G is some nonsingular $k \times k$ matrix, then

$$M = I - X(X'X)^{-1}X' = I - RG(G'R'RG)^{-1}G'R' = I - RR'$$

This means that the relevant feature of X is its space, which is spanned by the columns of R, while the location of the columns of X within this space, represented by G, is irrelevant. Test (DURA) is not independent of G, which, we feel, is a deficiency of this test. Accepting the space of X, which is often called the regression space, as the only relevant feature of X, our question becomes: which regression space is typical of economic time series?

5.1. Typical characteristics in the frequency domain

L'Esperance and Taylor simulated their series by means of a first-order autoregressive scheme $x_i = \lambda x_{i-1} + \eta_i$, $\eta_i \sim N(0, 1)$, where $\lambda = 0.1$ and 0.7. The power functions of the tests concerned and, thus, the results obtained by L'Esperance and Taylor are conditional on these series. Since the power functions are highly sensitive to changes in the character of the underlying series, the relevance of the results depends on whether the first-order autoregressive scheme adopted by l'Esperance and Taylor generates series having the essential characteristics of time series.

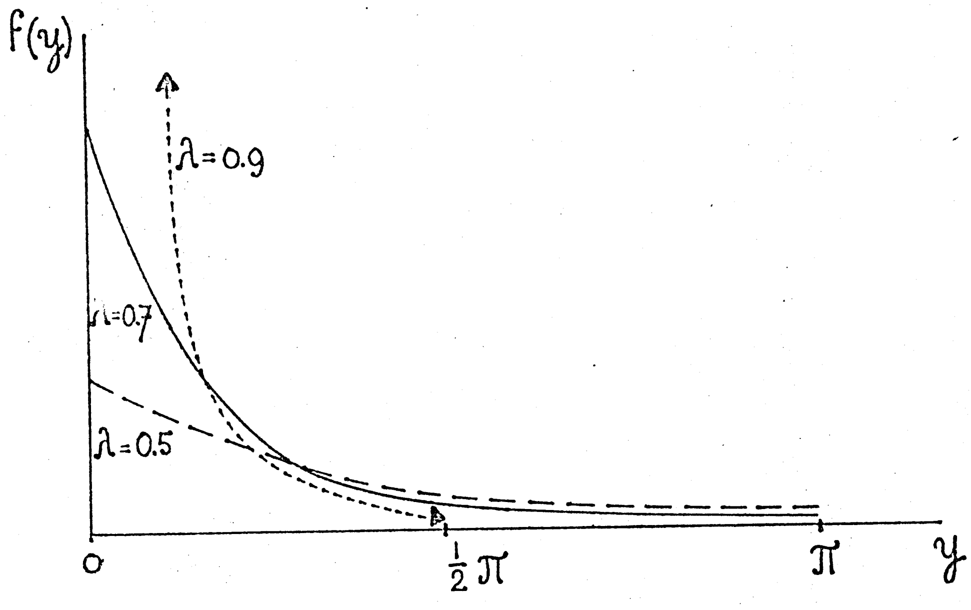
One possible way of examining the structure of a series is to consider its power spectrum. As is known, covariance stationary stochastic processes can be decomposed into a number of orthogonal components, each of which is associated with a given frequency. The power spectrum records the contribution of the components belonging to a given frequency band to the total variance of the process.

The power spectrum of the above mentioned autoregressive scheme is defined by

$$f(y) = \frac{1}{2\pi(1 + \lambda^2 - 2\lambda \cos y)}$$

The shape of f(y) is as follows

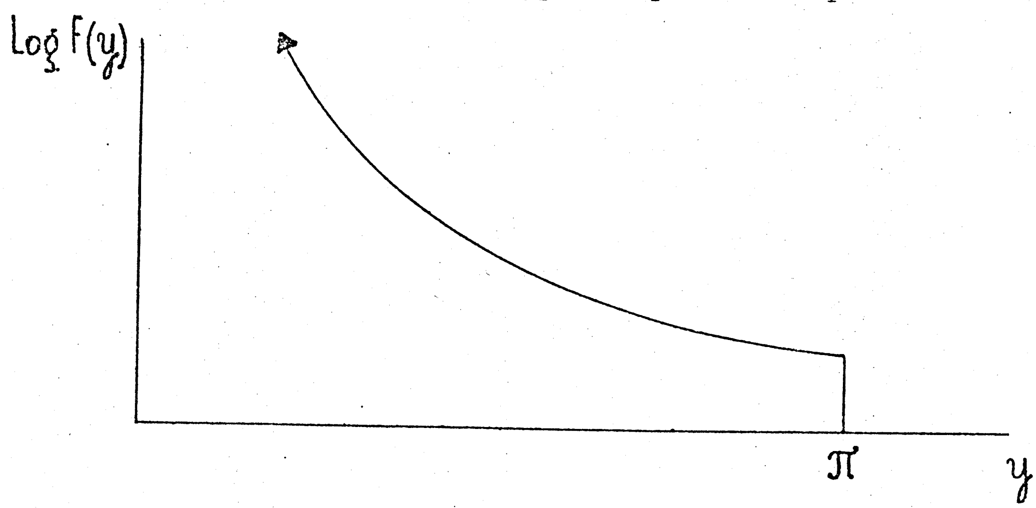
Figure 1. Spectral shape



For $\lambda = 0$ f(y) is a horizontal line. The larger λ becomes the higher f(y) is at frequency zero and the lower it is at frequency π.

In the course of time a fairly large number of power spectra have been estimated for economic series. The results obtained thus far were reviewed by Granger (1966). He concluded that there is regularity in the basic characteristics of economic variables. It appeared that the vast majority of economic variables investigated, after removal of trend and seasonal components, have similarly shaped power spectra, the typical shape being as in the following figure[1)]

Figure 2. Typical spectral shape



Thus the contribution of the low frequency components to the variance of an economic series is relatively high. One of the most obvious properties of a sample from a stochastic process having a spectrum of the typical shape is a visual long-term fluctuation which is not periodic.

Comparison of figures 1 and 2 suggests that a first-order autoregressive scheme might generate typical economic series. Empirical investigations seem to approve this conclusion provided the autocorrelation parameter is given a very high value e.g. 0.95 or higher, so that much weight is given to the low frequency components, see Granger. This agrees with the idea that economic processes are often almost unstable.

If λ = 0.95, e.g., the height of f(y) at frequency zero is $\frac{200}{\pi}$, whereas

----

1) This figure is the lower part of a scaled up version of a graph published in Granger (1966). Unlike Figure 1, this figure uses a logarithmic scale for measuring the power spectrum. This should be born in mind when comparing both figures.

for $\lambda = .7$ (the value used by l'Esperance and Taylor) it is $\frac{2}{\pi}$. Thus the power spectrum corresponding to the series simulated by l'Esperance and Taylor differs very much from the typical power spectrum of trendless economic time series. The simulation method used by l'Esperance and Taylor can of course be repaired by taking a much higher value for λ and adding a trend in the means of the series. This raises the power spectrum at low frequencies. The model then becomes x_i = trend in mean + e_i where $e_i = \lambda e_{i-1} + \eta_i$.

5.2. Typical characteristics in the time domain

In Dubbelman (1972) the space $M(P)$, where P is the $n \times k$ matrix $[h^*_1 \vdots h^*_2 \vdots \dots \vdots h^*_k]$, emerges as the mean of m regression spaces $M(X_i)$, where X_i is an empirical $n \times k$ matrix including a constant term, $i = 1,2, \dots, m$. This result has been found for $n = 15$, $k = 3$, and $m = 30$, and the assumed general validity of this result is verified for $n = 12$, $n = 10$, and for $k = 4$, and $k = 5$, both with and without a constant term.

When we would adopt only X-matrices with $R = P = [h^*_1 \vdots h^*_2 \vdots \dots \vdots h^*_k]$, which matrix is regarded as typical of economic time series, then test (D-W) = test (A-L) = test (sel) = test (DURA), and there is nothing left to be evaluated. Practically, it would be a matter of pure coincidence that the k columns of an X-matrix are linear combinations of h^*_1 through h^*_k. Every vector x, consisting of subsequent time series data, can be written as a linear combination of all h*-vectors.

$$x = c_1 h^*_1 + c_2 h^*_2 + \dots + c_k h^*_k + \dots + c_n h^*_n = H^* c$$

The point is that, on average, the values of c_i with i small are much greater than the values of c_i with i large. That is, the major components of x are, generally, a constant term (h^*_1), an almost linear trend (h^*_2), and waves

with low frequency (h_3^*, h_4^*), while the higher frequency vectors are relatively unimportant. (It is a matter of taste to call h_4^*, or even h_5^*, a low frequency vector for n in the neighbourhood of 15.) The observation, that the low frequency vectors are the major components of a vector of time series data, has been mentioned in literature by several authors. Several approximation procedures of significance points are based on this observation. In particular, Durbin (1970) says:
"Subsequently, Theil and Nagar [13] argued that in many applications the regressors are "slowly changing" in the sense that their successive differences of orders one to four are small in relation to the ranges of the variables themselves. More precisely, they assumed that quantities such as tr $\{X'A^rX(X'X)^{-1}\}$, $r = 1, \ldots, 4$, are small enough to be neglected in formulae for the first four moments of d, X being the matrix of observations of the regressors".

Below we use $\tau(s)$, defined by

$$\tau(s) = \text{tr } \{X'A^sX(X'X)^{-1}\} \qquad s = 1, 2, 3, 4$$

as indicators. It would be nice when we could say: a matrix X is not representative of economic time series if $\tau(s)$ exceeds some given number, which is possibly a function of s. However, at this moment we have not the faintest idea about such numbers. Therefore we must learn from empirical data what values of $\tau(s)$ can be expected.

To keep the analysis simple, we consider $k = 2$, and we write $X = [\iota \vdots x] = RG$, as above, where $R = [h_1^* \vdots r]$ and $r = (x'Ex)^{-\frac{1}{2}}Ex$. We obtain

$$\begin{aligned}\tau(s) &= \text{tr } \{G'R'A^sRG(G'R'RG)^{-1}\} \\ &= \text{tr } (R'A^sR) \\ &= r'A^sr \\ &= x'A^sx/x'Ex\end{aligned}$$

where we used $Ah_1^* = 0$. For several 15-element vectors x we calculated $\tau(s)$ for $s = 1, \ldots, 4$, see Table 1. It may be noted that

$\tau(s) = h_i^{*\prime} A^s h_i^* / h_i^{*\prime} E h_i^* = d_i^s$ when we take $x = h_i^*$, $i = 2, 3, \ldots, 15$.

Table 1. Values of $\tau(s)$ for several x-vectors

name of x-vector	$\tau(1)$	$\tau(2)$	$\tau(3)$	$\tau(4)$
h_2^*	0.04	0.00	0.00	0.00
h_3^*	0.17	0.03	0.01	0.00
h_4^*	0.38	0.15	0.06	0.02
h_5^*	0.66	0.44	0.29	0.19
h_6^*	1.00	1.00	1.00	1.00
h_7^*	1.38	1.91	2.64	3.65
h_8^*	1.79	3.21	5.74	10.29
h_9^*	2.21	4.88	10.78	23.81
h_{10}^*	2.62	6.85	17.94	46.98
h_{11}^*	3.00	9.00	27.00	81.00
h_{12}^*	3.34	11.14	37.20	124.19
h_{13}^*	3.62	13.09	47.36	171.35
h_{14}^*	3.83	14.65	56.05	214.52
h_{15}^*	3.96	15.65	61.93	244.99
$x_C(1)$	0.25	0.15	0.23	0.56
$x_C(2)$	0.52	1.12	3.09	9.39
$x_D(1)$	0.22	0.31	0.82	2.50
$x_D(2)$	0.30	0.46	1.06	2.82
$x_K(1)$	0.49	0.42	0.81	2.46
$x_K(2)$	0.43	0.59	1.53	4.96
$x_S(1)$	0.07	0.05	0.14	0.43
$x_S(2)$	0.88	2.13	5.88	17.43
$x_T(1)$	0.36	0.43	1.18	4.12
$x_T(2)$	0.16	0.31	0.90	2.86

The vectors $x_C(1)$, $x_C(2)$, $x_D(1)$, ..., $x_T(2)$ are 15-element regressor vectors, described in Appendix C. Expressing an n-element vector x as a linear combination of the n-element h*-vectors, $x = H^*c$, we have

$$\tau(s) = c'H^{*\prime}A^sH^*c / c'H^{*\prime}EH^*c =$$

$$= \sum_{i=2}^{n} c_i^2 d_i^s \Big/ \sum_{i=2}^{n} c_i^2$$

$$= \sum_{i=2}^{n} \frac{c_i^2}{\sum_{j=2}^{n} c_j^2} d_i^s$$

i.e. $\tau(s)$ is a weighted sum of the d_i^s, the weights being nonnegative and adding up to 1. Hence, each row in the lower half of Table 1 is such a weighted sum of the rows in the upper half of that table.

On the basis of our conclusions in the previous section and our findings in the present section, we propose to consider the following scheme

$$(5) \qquad \begin{aligned} x_i &= i\,t + e_i \\ e_i &= \lambda e_{i-1} + \eta_i \end{aligned} \qquad i = 1, 2, \ldots, n$$

where t is a constant, to be interpreted as a trend, λ is a first-order autocorrelation coefficient, $|\lambda| < 1$, and $e_0, \eta_1, \eta_2, \ldots, \eta_n$ are independent random drawings, e_0 from $N\left(0, \frac{1}{1-\lambda^2}\right)$ and $\eta_1, \eta_2, \ldots, \eta_n$ from $N(0, 1)$. For $t = 0$, this scheme is identical to the scheme adopted by l'Esperance and Taylor. In Table 2 the average values of $\tau(s)$ are presented for several values of t and λ. Every presented value of $\tau(s)$ in Table 2 is the average of 500 values of $\tau(s)$; for every row in Table 2 we simulated another set of 500 x-vectors. When comparing Table 2 and the lower half of Table 1, we conclude that, on average, the scheme (5) with $t = 0$ is not representative for our empirical data, apart, perhaps, from $x_S(2)$ when $\lambda = 0.9$. All other empirical data show values for $\tau(4)$ which

Table 2. Average values of $\tau(s)$ for 500 x-vectors simulated by (5)

t	λ	$\tau(1)$	$\tau(2)$	$\tau(3)$	$\tau(4)$
0.0	0.1	1.85	5.29	17.13	58.69
	0.3	1.57	4.21	13.17	44.20
	0.5	1.27	3.26	10.09	33.67
	0.7	1.00	2.42	7.37	24.53
	0.9	.70	1.59	4.74	15.54
0.2	0.1	1.09	3.10	10.09	34.69
	0.3	0.96	2.53	7.97	26.85
	0.5	0.82	2.00	6.09	20.15
	0.7	0.74	1.73	5.21	17.19
	0.9	0.59	1.28	3.77	12.31
0.4	0.1	0.50	1.30	4.16	14.11
	0.3	0.44	1.08	3.39	11.43
	0.5	0.40	0.92	2.81	9.36
	0.7	0.38	0.81	2.38	7.81
	0.9	0.39	0.83	2.49	8.20
0.6	0.1	0.27	0.65	2.05	7.15
	0.3	0.24	0.52	1.62	5.48
	0.5	0.22	0.45	1.37	4.54
	0.7	0.21	0.39	1.17	3.86
	0.9	0.21	0.37	1.10	3.60
0.8	0.1	0.18	0.39	1.23	4,17
	0.3	0.16	0.30	0.94	3.14
	0.5	0.15	0.26	0.77	2.55
	0.7	0.14	0.22	0.64	2.08
	0.9	0.13	0.20	0.57	1.83

correspond to $t \geq 0.4$ in Table 2. Considering $t > 0.4$, the trend becomes so dominant that $\tau(1)$ becomes very small while $\tau(4)$ is not that small. In fact, when looking at $\tau(1)$ in Table 2, we prefer $t = 0.4$; but when looking at $\tau(4)$, we prefer $t = 0.8$. Reconsidering the scheme (5), and writing $[e_1 \ e_2 \ \dots \ e_n]' = e$, the simulated x-vector is approximately equal to $c_2h_2^* + e$ for some value of c_2. Probably the scheme can be improved, in the sense of a better balance between $\tau(1)$ and $\tau(4)$, when we add another component, so that the simulated x-vector is approximately equal to $c_2h_2^* + c_3h_3^* + e$ for some values of c_2 and c_3. However, such an extension of the scheme would require a further study. In this paper we adopt the scheme (5) and we shall take $t = 0.4$ and $t = 0.8$, with $\lambda = 0.7$ in both cases.

6. The quality of the tests empirically seen

Table 3 contains average powers of test (D-W) for 15×2 X-matrices, all of which consist of a constant term column and a 15-element vector simulated according to the scheme (5). Given t and λ, five X-matrices are simulated. We calculated the powers of test (D-W) at significance level 0.05 against various alternative (to H_0: $\rho = 0$) values of ρ for these five X-matrices. The average power values are presented in the table. There we consider five different sets of values for t and λ, so that we simulated five sets of five X-matrices. Analogously, we calculated the average powers of the four tabulable tests for the same sets of X-matrices. The average losses of power, i.e. the average power of test (D-W) minus the average power of a tabulable test, are also presented in Table 3. Note that a negative loss of power is a positive gain of power. The average gains of power never exceeded 0.005.

A remark must be made with regard to test (DURA). In Appendix B it is proved that, when $k = 2$, test (DURA) and test (A-L) are identical if $x'h_2^* < 0$, where x is the column of X other than the constant term column. All tests but test (DURA) are invariant under multiplication by -1 of either x or h_2^* or both. On our computer, the h^*-vectors are generated by a subroutine such that the first (nonzero) element of each h^*-vector is positive. For the

Table 3. Average powers and losses of power for simulated 15 × 2 X-matrices

t	λ	ρ	Power of test (D-W)	Loss of power in the case of test (BLUS)	test (A-L)	test (DURA)	test (sel)
0.0	0.1	0.9	0.85	0.10	0.39	0.41	-0.00
		0.7	0.69	0.10	0.25	0.27	0.00
		0.5	0.46	0.07	0.13	0.15	0.00
		-0.5	0.48	0.08	0.02	0.07	0.00
		-0.7	0.72	0.12	0.01	0.09	0.00
		-0.9	0.90	0.11	-0.00	0.06	-0.00
0.0	0.7	0.9	0.78	0.10	0.12	0.20	0.06
		0.7	0.64	0.09	0.07	0.16	0.04
		0.5	0.43	0.07	0.04	0.11	0.02
		-0.5	0.47	0.08	0.00	0.08	0.00
		-0.7	0.72	0.13	0.00	0.10	0.00
		-0.9	0.90	0.15	-0.00	0.08	-0.00
0.0	0.95	0.9	0.76	0.09	0.13	0.24	0.07
		0.7	0.61	0.09	0.08	0.19	0.05
		0.5	0.41	0.08	0.04	0.12	0.02
		-0.5	0.46	0.10	0.01	0.08	0.00
		-0.7	0.71	0.16	0.01	0.11	-0.00
		-0.9	0.89	0.19	0.01	0.09	-0.00
0.4	0.7	0.9	0.72	0.16	0.09	0.27	0.02
		0.7	0.61	0.13	0.05	0.22	0.01
		0.5	0.42	0.10	0.02	0.14	0.01
		-0.5	0.47	0.10	0.01	0.10	0.00
		-0.7	0.73	0.15	0.01	0.13	-0.00
		-0.9	0.90	0.16	0.00	0.11	-0.00
0.8	0.7	0.9	0.73	0.11	0.00	0.19	0.00
		0.7	0.61	0.11	0.00	0.17	0.00
		0.5	0.42	0.09	0.00	0.12	0.00
		-0.5	0.48	0.10	-0.00	0.06	-0.00
		-0.7	0.73	0.15	0.00	0.07	0.00
		-0.9	0.91	0.16	0.00	0.04	0.00

sake of a useful comparison we changed the sign of x, where necessary, such that test (DURA) and test (A-L) differ. In our experiments with $k = 3$ we did not alter the signs of the simulated x-vectors. Sims (1975) proposed a small change in the calculation procedure of Durbin's alternative disturbance estimator z, see also Appendix B, which is theoretically advantageous in that it reduces the "distance" between z and $\hat{u}$.

Considering Table 3, we see that the first two subtables account for the cases considered by l'Esperance and Taylor. Test (sel) emerges as the best tabulable test, particularly when testing against negative autocorrelation, in which case test (A-L) is almost as good as test (sel). The third subtable accounts for the case which arose from a consideration of the typical power spectrum in time series. Also here test (sel) imparts the smallest loss of power for all values of ρ. In the last two subtables, the values of t and λ are chosen on the basis of the slowly changing character of time series regressors. Again, test (sel) is the best test, while test (sel) and test (A-L) coincide in all individual cases when $t = 0.8$. As stated in the preceding paragraph, test (DURA) and test (A-L) would coincide if the sign of every simulated vector x is such that $x'h_2^* < 0$. Summarizing, test (sel) is the best tabulable test against positive and negative autocorrelation, both when the X-matrices show some of the typical characteristics of economic time series and when this is not the case. Against negative autocorrelation, test (A-L) is a very good second best choice. Of course, these conclusions concern 15×2 X-matrices.

We next consider 15×3 X-matrices. In Table 4 the powers and the losses of power are presented for five empirical X-matrices, described in Appendix C. The average powers and losses of power for these five X-matrices are presented in Table 5, together with the average results for five simulated 15×3 X-matrices. All five simulated X-matrices consist of a constant term column and two vectors, both of which are simulated according to the scheme (5) with $t = 0.4$ and $\lambda = 0.7$.

Comparing the losses of power on each row in Table 4, we see that the smallest loss of power is found in the last column in all cases but one,

Table 4. Powers and losses of power for empirical 15 × 3 X-matrices

X	ρ	Power of	Loss of power in the case of			
		test (D-W)	test (BLUS)	test (A-L)	test (DURA)	test (sel)
X_C	0.9	0.57	0.11	0.10	0.06	0.10
	0.7	0.48	0.10	0.05	0.07	0.05
	0.5	0.35	0.08	0.02	0.06	0.02
	-0.5	0.44	0.09	-0.00	0.09	-0.00
	-0.7	0.70	0.15	-0.00	0.12	-0.00
	-0.9	0.90	0.17	-0.00	0.07	-0.00
X_D	0.9	0.65	0.13	0.05	0.12	0.02
	0.7	0.54	0.12	0.04	0.11	0.02
	0.5	0.37	0.10	0.03	0.08	0.01
	-0.5	0.44	0.12	0.01	0.02	-0.00
	-0.7	0.70	0.21	0.02	0.02	-0.00
	-0.9	0.89	0.24	0.01	0.01	-0.00
X_K	0.9	0.69	0.19	0.03	0.08	0.02
	0.7	0.55	0.18	0.02	0.07	0.01
	0.5	0.38	0.13	0.01	0.05	0.01
	- 0.5	0.44	0.14	0.00	0.03	0.00
	- 0.7	0.69	0.24	0.00	0.03	0.00
	- 0.9	0.89	0.28	0.00	0.02	0.00
X_S	0.9	0.68	0.19	0.32	0.32	0.00
	0.7	0.56	0.16	0.22	0.23	0.00
	0.5	0.38	0.11	0.12	0.12	0.00
	- 0.5	0.44	0.12	0.03	0.03	0.01
	- 0.7	0.70	0.19	0.03	0.03	0.01
	- 0.9	0.89	0.19	0.02	0.02	0.01
X_T	0.9	0.54	0.17	0.01	0.16	0.01
	0.7	0.46	0.17	0.01	0.15	0.01
	0.5	0.33	0.13	0.00	0.10	0.00
	- 0.5	0.42	0.15	-0.00	0.09	-0.00
	- 0.7	0.68	0.27	-0.00	0.12	-0.00
	- 0.9	0.88	0.33	-0.00	0.10	-0.00

Table 5. Average powers and losses of power for simulated 15 × 3 X-matrices ($t = 0.4$, $\lambda = 0.7$) and for the five empirical 15 × 3 X-matrices

Kind of the data	ρ	Power of test (D-W)	Loss of power in the case of test (BLUS)	test (A-L)	test (DURA)	test (sel)
simulated	0.9	0.63	0.14	0.15	0.21	0.04
	0.7	0.53	0.14	0.10	0.17	0.03
	0.5	0.37	0.11	0.06	0.11	0.01
	-0.5	0.43	0.12	0.02	0.06	0.02
	-0.7	0.68	0.21	0.02	0.08	0.03
	-0.9	0.88	0.24	0.01	0.07	0.03
empirical	0.9	0.63	0.16	0.10	0.15	0.03
	0.7	0.52	0.14	0.07	0.13	0.02
	0.5	0.36	0.11	0.04	0.08	0.01
	-0.5	0.44	0.12	0.01	0.05	0.00
	-0.7	0.69	0.21	0.01	0.06	0.00
	-0.9	0.89	0.24	0.00	0.04	0.00

namely for $X = X_c$ and $\rho = 0.9$. Test (A-L) and test (sel) coincide for $X = X_c$ and $X = X_T$, where test (DURA) imparts a considerable loss of power for $\rho < 0$. A dramatic loss of power is scored by test (A-L) and test (DURA) for $X = X_s$ and $\rho > 0$. Apart from the case $X = X_c$ and $\rho > 0$, test (BLUS) cannot stand the power comparison. Table 4 clearly suggests that test (sel) is the best tabulable test against autocorrelation in the case of empirical 15 × 3 X-matrices. Against negative autocorrelation, test (A-L) is a very good second best choice, see also the lower half of Table 5. The conclusion from the upper half of Table 5 is almost the same. Here test (A-L) is even slightly better than test (sel) when $\rho < 0$.

The unanimous result of all power calculations is that test (sel) is the best tabulable test against autocorrelation. The only real competitive procedure is test (A-L) against H_A: $\rho < 0$. Compared with test (A-L), test (sel) has the disadvantage that a much larger table of significance points is required. Tables of 5 and 10 per cent significance points for test (sel), ranging from 9 to 20 for n and from 1 to 4 for k, are available. A Fortran computer program for test (sel) is given in Louter and Dubbelman (1973).

We considered only 15×2 and 15×3 X-matrices in this paper. In the first place, we believe that a much larger number of observations rarely occurs in practice, and in the second place we believe that the losses of power of all tests diminish when n increases, see e.g. Koerts and Abrahamse (1969), p. 106. With respect to k we remark that it would be better to compare powers for empirical X-matrices rather than simulated X-matrices when $k \geq 3$. In our opinion the scheme (5) does not account for the typical characteristics of economic time series other than a trend component, see Section 5.2. As far as our experience goes, the powers of test (sel) for $n \times k$ X-matrices, with $10 \leq n \leq 20$ and $2 \leq k \leq 4$, are satisfactory, compared with the powers of test (D-W).

References

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, 938-948.

Abrahamse, A.P.J. and J. Koerts, 1971, New estimators of disturbances in regression analysis, Journal of the American Statistical Association 66, 71-74.

Abrahamse, A.P.J. and A.S. Louter, 1971, On a new test for autocorrelation in least squares regression, Biometrika 58, 53-60.

Berenblut, I.I. and G.I. Webb, 1973, A new test for autocorrelated errors in the linear regression model, Journal of the Royal Statistical Society B 35, 33-50.

Chow, G.C., 1957, Demand for automobiles in the United States (North-Holland Publ. Co., Amsterdam).

Dubbelman, C., 1972, A priori fixed covariance matrices of disturbance estimators, European Economic Review 3, 413-436.

Dubbelman, C., 1973, BLU disturbance estimators with fixed covariance matrices, Report 7302 of the Erasmus University Rotterdam.

Durbin, J., An alternative to the bounds test for testing for serial correlation in the least-squares regression, Econometrica 38, 422-429.

Durbin, J. and G.S. Watson, 1951, Testing for serial correlation in least-squares regression II, Biometrika 38, 159-178.

Esperance, W.L. l' and D. Taylor, 1975, The power of four tests of autocorrelation in the linear regression model, Journal of Econometrics 3, 1-21.

Granger, C.W.J., 1966, The typical spectral shape of an economic variable, Econometrica 34, 150-161.

Klein, L.R., 1950, Economic fluctuations in the United States 1921-1941 (Wiley, New York).

Koerts, J. and A.P.J. Abrahamse, 1969, On the theory and application of the general linear model (Rotterdam University Press, Rotterdam).

Louter, A.S. and C. Dubbelman, 1973, An exact autocorrelation test for small n and k, a computer program, and a table of significance points, Report 7304 of the Erasmus University Rotterdam.

Sato, R., 1970, The estimation of biased technical progress and the production function, International Economic Review 11, 179-208.

Sims, C.A. 1975, A note on exact tests for serial correlation, Journal of the American Statistical Association 70, 162-165.

Theil, H., Principles of Econometrics (Wiley, New York).

Appendix A

The BLU, BLUS and BLUF disturbance estimators

Abrahamse and Koerts (1971) presented their new estimators as follows. Given the linear model, $y = X\beta + u$, where $E(u) = 0$ and $E(uu') = \sigma^2 I_{(n)}$, the new estimator v of u is linear (L) in y, so that $v = B'y$ with B independent of y, and v is unbiased (U) in the sense that $E(v) = E(u) = 0$, so that $B'X = 0$, and the covariance matrix of v is $\sigma^2\Omega$, where Ω is a fixed (F) symmetric idempotent $n \times n$ matrix with rank $n - k$, so that $\Omega = KK'$ with $K'K = I_{(n-k)}$, and, finally, v is best (B) in the sense that it minimizes $E[(v-u)'(v-u)]$. The reasoning, used to derive v, can also be applied to a more general context, see Dubbelman (1973), as follows.

First, replace $E(uu') = \sigma^2 I_{(n)}$ by $E(uu') = \sigma^2\Gamma$, where Γ is a fixed symmetric positive definite $n \times n$ matrix. Second, consider a linear unbiased estimator w of $J'u$, where J is a fixed $n \times p$ matrix, so that w has p elements. Third, impose $E(ww') = \sigma^2\Omega$, where Ω is a fixed symmetric $p \times p$ matrix with rank r, not necessarily idempotent, so that a $p \times r$ matrix K exists such that $\Omega = KK'$. Fourth, take w best in the sense that it minimizes $E[(w-J'u)'Q(w-J'u)]$, where Q is a fixed symmetric nonnegative definite $p \times p$ matrix. Provided that $r \leq n-k$ and that $K'QJ'M^*\Gamma JQK$ is nonsingular, where

$$M^* = I_{(n)} - X(X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}$$

the BLUF estimator of $J'u$ is

$$w = K(K'QJ'M^*\Gamma JQK)^{-\frac{1}{2}}K'QJ'M^*y$$

It is seen that $w = v$ if $J = Q = \Gamma = I_{(n)}$ and Ω idempotent, and w is the least-squares (BLU) residual vector if besides $M = KK'$, and w is a BLUS vector if $\Gamma = I_{(n)}$, $K = Q = I_{(n-k)}$, and J consists of n-k columns from $I_{(n)}$.

Appendix B

Durbin's estimator z versus BLUF

Assuming $y = X\beta + u$ with $u \sim N(0,\sigma^2 I)$, Durbin (1970) constructed the vector z such that $z = B_z' y \sim N(0,\sigma^2\Omega)$, where $\Omega = KK' = I_{(n)} - PP'$ and $P = [h_1^* \vdots h_2^* \vdots \dots \vdots h_k^*]$. X and P are partitioned as $[h_1^* \vdots \tilde{X}]$ and $[h_1^* \vdots \tilde{P}]$, respectively. Durbin proposed the following computing procedure for z:

$$z = y - h_1^* a - \tilde{X} b_1 - \tilde{P} b_2 + \Omega\tilde{X} P_1 P_2^{-1} b_2 \qquad \text{(B.1)}$$

where a, b_1, and b_2 are the coefficients of h_1^*, $\tilde{X}$, and $\tilde{P}$ in the least-squares fit of the regression of y on h_1^*, $\tilde{X}$, and $\tilde{P}$, and where P_1 and P_2 are lower triangular matrices such that $\sigma^2 P_1 P_1'$ and $\sigma^2 P_2 P_2'$ are the covariance matrices of b_1 and b_2, respectively. For the calculation of the power of test (DURA) an explicit expression for B_z is required, which reads

$$B_z' = \Omega - \Omega\tilde{X}(\tilde{X}'\Omega\tilde{X})^{-1}\tilde{X}'\Omega + \Omega\tilde{X} P_1 P_2' \tilde{P}' M \qquad \text{(B.2)}$$

Below, (B.2) is derived from (B.1). Sims (1975) proposed another vector, which we denote by s,

$$s = y - h_1^* a - \tilde{X} b_1 - \tilde{P}\tilde{P}'\tilde{X} b_3 - \Omega\tilde{X} P_1 P_3^{-1} b_3 \qquad \text{(B.3)}$$

where $b_3 = (\tilde{P}'\tilde{X})^{-1} b_2$ and P_3 is lower triangular such that $\sigma^2 P_3 P_3'$ is the covariance matrix of b_3. Sims argued that his modifications of the computing procedure for z, namely reversion of the sign before Ω in (B.1) and use of $\tilde{P}\tilde{P}'\tilde{X}$ instead of P, improve that vector. Following the notation of Sims, we write

$$\text{(B.4)} \qquad \begin{aligned} \hat{u} &= [\Omega - Q_1 Q_1' + Q_2 Q_2']\, y \\ z &= [\Omega - Q_1 Q_1' + Q_1 Q_2']\, y \\ s &= [\Omega - Q_1 Q_1' - Q_1 H_s Q_2']\, y \\ w &= [\Omega - Q_1 Q_1' + Q_1 H_w Q_2']\, y \end{aligned}$$

where $Q_1 = \Omega\tilde{X}P_1$, $Q_2 = M\tilde{P}P_2$, $H_s = P_3^{-1}(\tilde{P}'\tilde{X})^{-1}P_2$, and $H_w = (Q_1'Q_2Q_2'Q_1)^{-\frac{1}{2}}Q_1'Q_2$. It is proved that every vector, which can be written as $[\Omega - Q_1Q_1' + Q_1HQ_2']y$ has zero mean and that such a vector has covariance matrix $\sigma^2\Omega$ if and only if $H' = H^{-1}$. Finally, we consider $k = 2$, in which case H and $\tilde{P}'\tilde{X}$ are scalars. In particular, $z = w$ if the scalar $\tilde{P}'\tilde{X}$ is negative, while $s = w$, regardless of the sign of the scalar $\tilde{P}'\tilde{X}$.

Consider

$$Z = [h \vdots \tilde{X} \vdots \tilde{P}] = [X \vdots \tilde{P}] \text{ and } \gamma = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}$$

Here h stands for h_1^*, for notational convenience. Then $\gamma = (Z'Z)^{-1}Z'y$. It can be verified that

$$Z'Z = \left[\begin{array}{c|c|c} 1 & h'\tilde{X} & 0 \\ \hline \tilde{X}'h & \tilde{X}'\tilde{X} & \tilde{X}'\tilde{P} \\ \hline 0 & \tilde{P}'\tilde{X} & I \end{array}\right] = \left[\begin{array}{c|c} X'X & X'\tilde{P} \\ \hline \tilde{P}'X & I \end{array}\right]$$

and that

$$(Z'Z)^{-1} = \left[\begin{array}{c|c|c} 1+h'\tilde{X}V\tilde{X}'h & -h'\tilde{X}V & h'\tilde{X}V\tilde{X}'\tilde{P} \\ \hline -V\tilde{X}'h & V & -V\tilde{X}'\tilde{P} \\ \hline \tilde{P}'\tilde{X}V\tilde{X}'h & -\tilde{P}'\tilde{X}V & I+\tilde{P}'\tilde{X}V\tilde{X}'\tilde{P} \end{array}\right] = \left[\begin{array}{c|c} (X'X)^{-1}+G'(\tilde{P}'M\tilde{P})^{-1}G & -G'(\tilde{P}'M\tilde{P})^{-1} \\ \hline -(\tilde{P}'M\tilde{P})^{-1}G & (\tilde{P}'M\tilde{P})^{-1} \end{array}\right]$$

with $V = (\tilde{X}'\Omega\tilde{X})^{-1}$ and $G = \tilde{P}'X(X'X)^{-1}$. We find $P_1P_1' = (\tilde{X}'\Omega\tilde{X})^{-1}$ and $P_2P_2' = (\tilde{P}'M\tilde{P})^{-1} = I + \tilde{P}'\tilde{X}(\tilde{X}'\Omega\tilde{X})^{-1}\tilde{X}'\tilde{P}$. Using the first partitioning of Z, we have

$$\gamma = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} h'(I+\tilde{X}V\tilde{X}'hh'-\tilde{X}V\tilde{X}'+\tilde{X}V\tilde{X}'\tilde{P}\tilde{P}') \\ -V\tilde{X}'hh'+V\tilde{X}'-V\tilde{X}'\tilde{P}\tilde{P}' \\ \tilde{P}'(\tilde{X}V\tilde{X}'hh'-\tilde{X}V\tilde{X}'+I+\tilde{X}V\tilde{X}'\tilde{P}\tilde{P}') \end{bmatrix} y = \begin{bmatrix} h'(I-\tilde{X}V\tilde{X}'\Omega) \\ V\tilde{X}'\Omega \\ \tilde{P}'(I-\tilde{X}V\tilde{X}'\Omega) \end{bmatrix} y$$

and from the second partitioning it follows that $b_2 = (\tilde{P}'M\tilde{P})^{-1}\tilde{P}'My$, so that $P_2^{-1}b_2 = P_2^{-1}P_2P_2'\tilde{P}'My = P_2'\tilde{P}'My$. Substitution of the expressions for a, b_1, b_2, and $P_2^{-1}b_2$ into (B.1) yields

$$B_z' = I - hh'(I - \tilde{X}V\tilde{X}'\Omega) - \tilde{X}V\tilde{X}'\Omega - \tilde{P}\tilde{P}'(I - \tilde{X}V\tilde{X}'\Omega) + \Omega\tilde{X}P_1P_2'\tilde{P}'M$$
$$= I - hh' - \tilde{P}\tilde{P}' - (I - hh' - \tilde{P}\tilde{P}')\tilde{X}V\tilde{X}'\Omega + \Omega\tilde{X}P_1P_2'\tilde{P}'M$$

and (B.2) follows. It may be noted that $[\Omega - \Omega\tilde{X}(\tilde{X}'\Omega\tilde{X})^{-1}\tilde{X}'\Omega]X = [0 \vdots \Omega\tilde{X}] - \Omega\tilde{X}(\tilde{X}'\Omega\tilde{X})^{-1}[0 \vdots \tilde{X}'\Omega\tilde{X}] = 0$, so that $B_z'M = B_z'$.

Let Q_1 and Q_2 be orthogonal $n \times (k-1)$ matrices, whose column vectors lie in $M(X \vdots P)$, such that $Q_1'P = 0$ and $Q_2'X = 0$, and let Q_0 be an orthogonal $n \times (n-2k+1)$ matrix whose column vectors span $M(X \vdots P)^{\perp} = M(h_1^* \vdots \tilde{X} \vdots \tilde{P})^{\perp}$. Then the column vectors of $K = [Q_0 \vdots Q_1]$ span $M(P)^{\perp}$, $K'K = I_{(n-k)}$, and the column vectors of $N = [Q_0 \vdots Q_2]$ span $M(X)^{\perp}$, $N'N = I_{(n-k)}$, and we have $\Omega = KK'$ and $M = NN'$. The matrices Q_1 and Q_2 can be made explicit as follows. We consider Q_1, whose column vectors lie in $M(X \vdots P) = M(\tilde{X} \vdots P)$, i.e. $Q_1 = \tilde{X}F_1 + PF_2$ for some matrices F_1 and F_2. From $P'Q_1 = 0$ it follows that $P'\tilde{X}F_1 = -F_2$, so that $Q_1 = \tilde{X}F_1 - PP'\tilde{X}F_1 = \Omega\tilde{X}F_1$. Taking $F_1 = P_1$, the lower triangular matrix such that $P_1P_1' = (\tilde{X}'\Omega\tilde{X})^{-1}$, we have $Q_1'Q_1 = I_{(k-1)}$. Of course, we could also take $F_1 = (\tilde{X}'\Omega\tilde{X})^{-\frac{1}{2}}$ in order to obtain $Q_1'Q_1 = I_{(k-1)}$. Analogously, $Q_2 = M\tilde{P}P_2'$. Both Q_1 and Q_2 can be postmultiplied by an arbitrary orthogonal $(k-1) \times (k-1)$ matrix. It is easily seen that $\hat{u} = My = NN'y = [Q_0Q_0' + Q_2Q_2']y$, where $Q_0Q_0' = KK' - Q_1Q_1'$. Further, substitution of Q_1 and Q_2 into (B.2) yields $z = [\Omega - Q_1Q_1' + Q_1Q_2']y$. Comparing (B.1) and (B.3), we see that $s = [\Omega - Q_1Q_1']y - Q_1P_3^{-1}b_3$. Using $b_3 = (\tilde{P}'\tilde{X})^{-1}b_2 = (\tilde{P}'\tilde{X})^{-1}P_2P_2'\tilde{P}'My$, we find $s = [\Omega - Q_1Q_1' - Q_1H_sQ_2']y$, where $H_s = P_3^{-1}(\tilde{P}'\tilde{X})^{-1}P_2$. Since the covariance matrix of b_2 is $\sigma^2P_2P_2'$, the covariance matrix of $b_3 = (\tilde{P}'\tilde{X})^{-1}b_2$ is $\sigma^2P_3P_3' = (\tilde{P}'\tilde{X})^{-1}(\sigma^2P_2P_2')(\tilde{X}'\tilde{P})^{-1}$. Hence, $H_sH_s' = I$. Considering $w = K(K'MK)^{-\frac{1}{2}}K'My$, we first observe that $Q_0'MQ_0 = I_{(n-2k+1)}$, $Q_0'MQ_1 = 0$, and $Q_0'M = Q_0'$. Then

$$w = [Q_0 \vdots Q_1]\begin{bmatrix} I_{(n-2k+1)} & 0 \\ 0 & Q_1'MQ_1 \end{bmatrix}^{-\frac{1}{2}}\begin{bmatrix} Q_0'M \\ Q_1'M \end{bmatrix} y$$

$$= [Q_0 \vdots Q_1]\begin{bmatrix} I_{(n-2k+1)} & 0 \\ 0 & (Q_1'MQ_1)^{-\frac{1}{2}} \end{bmatrix}\begin{bmatrix} Q_0' \\ Q_1'M \end{bmatrix} y$$

$$= [Q_0Q_0' + Q_1(Q_1'MQ_1)^{-\frac{1}{2}}Q_1'M]y$$

Substitution of $Q_0Q_0' = \Omega - Q_1Q_1'$ and of $Q_1'M = Q_1'Q_2Q_2'$ yields $w = [\Omega - Q_1Q_1' + Q_1H_wQ_2']y$, where $H_w = (Q_1'Q_2Q_2'Q_1)^{-\frac{1}{2}}Q_1'Q_2$ and H_w satisfies $H_w' = H_w^{-1}$.

In view of $Q_0'X = 0$ and $Q_2'X = 0$, we have $E[\Omega - Q_1Q_1' + Q_1HQ_2']y = E[\Omega - Q_1Q_1' + Q_1HQ_2']u = 0$ for H arbitrary. The covariance matrix of $[\Omega - Q_1Q_1' + Q_1HQ_2']y = [Q_0Q_0' + Q_1HQ_2']y$ is $E\{[Q_0Q_0' + Q_1HQ_2']uu'[Q_0Q_0' + Q_2H'Q_1']\}$ $= \sigma^2[Q_0Q_0' + Q_1HQ_2'Q_2H'Q_1'] = \sigma^2[Q_0Q_0' + Q_1HH'Q_1']$, where we used $Q_0'Q_2 = 0$ and $Q_2'Q_2 = I$. Now $Q_0Q_0' + Q_1HH'Q_1' = KK' = Q_0Q_0' + Q_1Q_1'$ if and only if $Q_1HH'Q_1' = Q_1Q_1'$, i.e. if and only if $HH' = I$, in view of $Q_1'Q_1 = I$.

Given the vector $[\Omega - Q_1Q_1' + Q_1HQ_2']y$ with $H' = H^{-1}$, the matrix H is a scalar if k = 2, so that H = 1 or H = -1. In the vector w we have $H = H_w = (Q_1'Q_2Q_2'Q_1)^{-\frac{1}{2}}Q_1'Q_2$, where $(Q_1'Q_2Q_2'Q_1)^{-\frac{1}{2}}$ is positive, by definition, so that the sign of H_w is equal to the sign of $Q_1'Q_2 = P_1'\tilde{X}'M\Omega\tilde{P}P_2$, where P_1 and P_2 are positive, by definition. Using $\Omega = I - h_1^*h_1^{*\prime} - \tilde{P}\tilde{P}'$, we have $\tilde{X}'M\Omega\tilde{P} = -\tilde{X}'\tilde{P}\tilde{P}'M\tilde{P}$, since $[h_1^* \vdots \tilde{X}]'M = X'M = 0$, and $\tilde{P}'M\tilde{P}$ is positive. Hence, the sign of H_w is opposite to the sign of $\tilde{X}'\tilde{P}$. Clearly, H = 1 yields the vector z, so that z = w if $\tilde{X}'P$ is negative. In the vector s we have $-H = H_s = P_3^{-1}(\tilde{P}'\tilde{X})^{-1}P_2$, where P_3 and P_2 are positive, by definition. Hence, the sign of H_s is equal to the sign of $\tilde{X}'\tilde{P}$, so that s = w, regardless of the sign of $\tilde{X}'P$, owing to the minus sign before $Q_1H_sQ_2'$ in (B.4). When k is greater than 2, z and s and w are all different, generally.

Appendix C

The empirical data

The 15-element x-vectors listed below consist of subsequent annual observations. The 15 × 3 matrices X_C, X_D, X_K, X_S, and X_T consist of two of these vectors and a constant term vector.

X_C Chow (1957), logarithms of Table 1; $x_C(1)$: log automobile stock per capita; $x_C(2)$: log personal money stock per capita; for the United States, 1921 - 1935.

X_D Durbin and Watson (1951), Table 1; $x_D(1)$: log real income per capita; $x_D(2)$: log relative price of spirits; for the United Kingdom, 1870 - 1884.

X_K Klein (1950), p. 135; $x_K(1)$: profits; $x_K(2)$: wages; for the United States, 1923 - 1937.

X_S Sato (1970), p. 203; $x_S(1)$: capital; $x_S(2)$: man hours; for the United States, 1946 - 1960.

X_T Theil (1971), Table 3.1; $x_T(1)$: log real income per capita; $x_T(2)$: log relative price of textiles; for the Netherlands, 1923 - 1937.