A NOTE ON EXACT TESTS FOR SERIAL CORRELATION *

by

Christopher A. Sims

Discussion Paper No. 74-39, May 1974

* This paper supersedes Discussion Paper No. 26. Errors in the earlier version are corrected and new material is added.

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

ABSTRACT

A transformation of the OLS residual vector to achieve a desired covariance structure, proposed earlier by Durbin, is shown to be capable of substantially changing the OLS residual vector even when that vector already has nearly the desired covariance structure. This may explain its substantially inferior performance in Monte Carlo comparisons with the transform proposed by Abrahamse and Koerts. A new transform, involving only small alterations in Durbin's procedure, is shown to avoid the defect of the Durbin transform.

A NOTE ON EXACT TESTS FOR SERIAL CORRELATION

by Christopher A. Sims*

Durbin [3], Durbin and Watson [5], Abrahamse and Koerts [1], and Abrahamse and Louter [2] have recently discussed general methods for transforming the residuals from least-squares regression in such a way that the transformed residuals will have a standard covariance matrix which does not depend on the matrix of regressors. The methods they propose will apply with any choice of the standard covariance matrix for the transformed regressors, so long as the standard covariance matrix can be expressed in the form

$$\Omega = \sigma^2(I - X_2(X_2'X_2)^{-1}X_2') ,$$

where X_2 is some matrix of the same rank (k) as the matrix of regressors, which we shall call X_1. It is easily seen that the theory in the articles cited has Theil's BLUS transformation as a special case in which X_2 is a matrix of columns each of which has all but one element zero. The methods may prove useful in generating frequency-domain procedures for residual analysis, since by proper choice of X_2 it is possible to give the Fourier transform of the estimated residuals at n-k of the harmonic frequencies the same distribution on the null hypothesis as the Fourier transform of the true residuals at those n-k points.

Abrahamse and Louter assert that tests based on a von Neuman ratio computed from their transformed residuals are, in certain examples, more powerful than tests based on the same statistic calculated from

Durbin's transformed residuals. On the other hand, Durbin's transform can be calculated using standard operations from a least-squares fitting algorithm plus the taking of triangular square roots of two $k \times k$ matrices; Abrahamse and Louter show that their transform (which is the same as that described in Abrahamse and Koerts [1]) can be computed by a procedure which involves only inverting $k \times k$ matrices and finding eigenvectors and eigenvalues of a $k \times k$ matrix. When k is one or two, the difference in computational requirements of the two methods is insignificant. But when k is of the order of, say, five or more, the additional computational requirements of the Abrahamse-Koerts-Louter (A-K-L) transform might be large enough to preclude its routine use in situations where the Durbin transform might be routinely implemented.

There are two reasons for expecting that the Durbin or A-K-L transform might be profitably implemented with the k columns of X_2 chosen to be cosine functions at low harmonic frequencies, as both Durbin and A-K-L propose. One reason, made explicit by both sets of authors, is that in this case the von Neumann ratio calculated from the transformed residuals has the distribution of the d_u statistic which has already been tabulated by Durbin and Watson [4]. The other, made explicit by A-K-L, is that with this choice of X_2 one might hope that the space spanned by X_2 and by the regressor matrices X_1 found in economic applications would be nearly the same, that therefore the transformation might have little effect on the original least-squares residuals, and that therefore the loss of power from using the transformation might be small.[1/]

In this paper we show that when X_1 and X_2 span nearly the same space, the Durbin transform does not in general leave the least-squares residuals nearly unchanged, whereas the A-K-L transform does have this desirable property. It seems likely that this defect of the Durbin procedure explains the poor power performance for the Durbin transform found by Abrahamse and Louter in their experiments with von Neumann ratios from economic data. This paper introduces a third transform which eliminates an element of arbitrariness in the Durbin transform, thereby achieving the property that for X_1 and X_2 spanning nearly the same space the transformed residuals are nearly the same. The proposed transform differs from that proposed by A-K-L in that it requires computations of the same order of magnitude as the original Durbin transform, regardless of k.

1. AN EXAMPLE

Suppose we have a regression equation of the form

$$y = \underset{n\times 1}{X_1}\ \underset{k\times 1}{\beta} + \underset{n\times 1}{u}\ , \tag{1.1}$$

with X_1 fixed and u distributed normally with scalar covariance matrix. To make matters simple initially, take $k=1$, $X_1 = aZ_1 + Z_2$, where Z_1 and Z_2 are mutually orthogonal vectors of unit length. Take $X_2 = Z_2$. Clearly the Ω matrix, and hence the nature of the problem, is the same whether we take $X_2 = -Z_2$ or instead take $X_2 = Z_2$. Accordingly, the A-K-L transform gives the same result for both choices of X_2. The Durbin transform gives different results for the two cases. As we shall now see, with $X_2 = Z_2$ the Durbin transform and

the A-K-L transform differ, whereas with the other choice of X_2 they are the same.

Let $\hat{u}_{12}$ be the vector of least-squares residuals from a regression of y on X_1 and X_2 jointly. Let c_2 be the estimated coefficient of X_2 in this joint regression. Then the Durbin transform yields the following estimated residual vector, as is easily verified from Durbin's [3] expression on p. 425

$$\hat{u}_D = \hat{u}_{12} + aZ_1P_1c_2/P_2 , \tag{1.2}$$

where P_1 and P_2 are the estimated standard errors of the coefficients of X_1 and X_2, respectively, in the joint regression. Here $P_1/P_2 = 1/(1+a^2)^{\frac{1}{2}}$, regardless of whether X_2 is plus or minus Z_2. The terms a and Z_1 also are unaffected by whether X_2 is plus or minus Z_2. But c_2's sign varies according to whether X_2 is plus or minus Z_2.

Now maintaining the $X_2 = Z_2$ assumption the OLS residual vector, $\hat{u}_{OLS}$, can be expressed in a form analogous to (1.2) above, yielding

$$\hat{u}_{OLS} = \hat{u}_{12} + (a^2Z_2 - aZ_1)c_2/(1+a^2) , \tag{1.3}$$

where in this case $c_2 = -(Z_1'u)/a - (Z_2'u)/(1+a^2)$. As a tends to zero, (1.2) and (1.3) yield

$$\hat{u}_D \to \hat{u}_{12} - Z_1Z_1'u , \text{ and} \tag{1.4}$$

$$\hat{u}_{OLS} \to \hat{u}_{12} + Z_1Z_1'u .$$

Thus with this particular choice for X_2, the Durbin transform fails to let $\hat{u}_D$ converge to $\hat{u}_{OLS}$ as X_1 converges to X_2. It is obvious that the problem here could be remedied simply by making sure to choose

$X_2 = -Z_2$ at the start. In this example, the sole difference between the A-K-L transform and the Durbin transform is that the A-K-L transform optimizes the sign of the vector added to $\hat{u}_{12}$ in (1.2). To be precise, A-K-L compares the vector multiplying c_2 in (1.2) with that multiplying c_2 in (1.3) Since, in this example, the two vectors show a negative cross-product, (1.2) is altered in the A-K-L transform by changing the sign of the vector added to $\hat{u}_{12}$.

3. RELATIONS AMONG OLS, A-K-L, AND DURBIN RESIDUALS

The transformed A-K-L residuals, $\hat{u}_{AKL}$, can in general be expressed as $\hat{u}_{12}$ plus a correction vector which is a multiple of c_2 , analogous to the Durbin expression for his transform presented on p. 425 of his [3] article. To see this, consider the expression given by Abrahamse and Louter [2] on p. 54 for their transform:

$$\hat{u}_{AKL} = K(K'MK)^{-\frac{1}{2}}K'Mu \ , \tag{2.1}$$

where K is an $n\times k$ matrix of orthonormal eigenvectors corresponding to unit eigenvalues of Ω, $M = I - X_1(X_1'X_1)^{-1}X_1'$, and the exponent "$-\frac{1}{2}$" indicates the inverse of the symmetric square root.[2] Let Q be some $n\times(n-2k)$ matrix of orthonormal vectors all of which are orthogonal to X_1 and to X_2 . Let $Q_{1.2}$ be an $n\times k$ matrix of orthonormal vectors orthogonal to X_2 but in the space spanned by X_2 and X_1 . Let $Q_{2.1}$ be defined analogously. Then we can take $K = (Q_{1.2}, Q)$, and $L = (Q_{2.1}, Q)$. L is then a matrix of orthonormal eigenvectors of $M = LL'$. With this definition of K and L , we can rewrite (2.1) as

$$\hat{u}_{AKL} = Q_{1.2}(Q_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2})^{-\frac{1}{2}}Q_{1.2}'Q_{2.1}Q_{2.1}'u + QQ'u \ . \tag{2.2}$$

But $QQ'u$ is $\hat{u}_{12}$ and $Q'_{2.1}u$ can be expressed as the product of a $k \times k$ matrix with c_2. Hence we have the desired correspondence with Durbin's expression.

Setting $H = (Q'_{1.2}Q_{2.1}Q'_{2.1}Q_{1.2})^{-\frac{1}{2}}Q'_{1.2}Q_{2.1}$, (2.2) becomes

$$\hat{u}_{AKL} = \hat{u}_{12} + Q_{1.2}HQ'_{2.1}u \text{ .}^{3/} \tag{2.3}$$

If, following Durbin's notation, we take $X_{1.2}$ to be that part of X_1 orthogonal to X_2 (i.e., the residual matrix from a least squares fit of X_1 to X_2) and define $X_{2.1}$ analogously, one possible choice of $Q_{1.2}$ and $Q_{2.1}$ is $Q_{1.2} = X_{1.2}((X'_{1.2}X_{1.2})^{ts})^{-1}$, with $Q_{2.1}$ chosen analogously. Here the "ts" exponent denotes the taking of the positive definite triangular square root; i.e., A^{ts} is the lower-triangular matrix with positive diagonal elements such that $A = A^{ts}A^{ts'}$. With this choice of Q's, Durbin's transform amounts to choosing H in (2.3) as the identity matrix. Abrahamse and Koerts show that any orthonormal matrix can replace H in (2.3) without altering the distribution of the resulting transformed residual vector. However, to minimize $E((u-\hat{u}_{AKL})'(u-\hat{u}_{AKL}))$ one must choose a particular H, that implicit in (2.2).

A general expression for the OLS residuals analogous to (2.2) and (2.3) is

$$\hat{u}_{OLS} = \hat{u}_{12} + Q_{2.1}Q'_{2.1}u \text{ .} \tag{2.4}$$

Thus the problem of keeping the change in the OLS residuals small can be seen as the problem of choosing H so that $Q_{1.2}H$ will be close to $Q_{2.1}$.

3. THE THIRD TRANSFORM

The A-K-L choice of H has the form $H = CQ_{1.2}'Q_{2.1}$, where C is a matrix satisfying $C^{-1}C^{-1\prime} = Q_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2}$, i.e. where C is the inverse of a square root of $Q_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2}$. So long as $Q_{2.1}'Q_{1.2}$ has full rank, any choice of H can be expressed in this form, so the choice of C determines the choice of H. Durbin's transform amounts to taking $C = (Q_{1.2}'Q_{2.1})^{-1}$. If we wish to minimize the A-K-L criterion, $E((u-\hat{u}_{AKL})'(u-\hat{u}_{AKL}))$, we can see directly from (2.3) and (2.4) that the problem reduces to minimizing

$$(3.1) \qquad \operatorname{tr}((Q_{2.1}H'Q_{1.2}' - Q_{2.1}Q_{2.1}')(Q_{1.2}HQ_{2.1}' - Q_{2.1}Q_{2.1}'))$$

$$= 2\operatorname{tr}(Q_{2.1}Q_{2.1}') - 2\operatorname{tr}(HQ_{2.1}'Q_{1.2})$$

$$= 2k - 2\operatorname{tr}(CQ_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2})$$

$$= 2(k - \operatorname{tr} C^{-1}) .$$

Now it should be clear why the Durbin transform is so far from optimal. If no account is taken of X_1 in the choice of X_2, the Durbin choice of C could easily yield a C^{-1} with all its diagonal elements negative, as indeed happens in the example of section 1. On the other hand, as the space spanned by X_1 converges to that spanned by X_2[4/], $Q_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2}$ converges to the identity matrix. Since C^{-1} is a square root of $Q_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2}$, it is clear from (3.1) that the transformed vector of residuals converges to the OLS vector in these circumstances so long as the square root taken to obtain C^{-1} is one which converges to the identity matrix as the matrix whose root is

sought converges to the identity. Certainly the symmetric square root which yields the A-K-L transform has this property. The "square root" which yields the Durbin transform does not have the property.

However the difficulty with the Durbin transform can be avoided by choosing X_2 properly in relation to X_1. In particular, if we start with a matrix Z_2 which spans the space of eigenvectors of Ω corresponding to zero eigenvalues, then we can choose X_2 to be the projection of X_1 on the Z_2-space, i.e. $X_2 = Z_2(Z_2'Z_2)^{-1}Z_2'X_1$. With this choice of X_2, not only does $Q_{1.2}'Q_{2.1}Q_{2.1}'Q_{1.2}$ converge to I as the spaces spanned by X_1 and Z_2 converge, but in addition $Q_{1.2}'Q_{2.1}$ itself converges to $-I$. Durbin's choice of $C^{-1} = Q_{1.2}'Q_{2.1}$ would not work here, but <u>minus</u> his choice is fine. Thus with this choice of X_2 and a reversal of the sign on the term Durbin adds to $\hat{u}_{12}$, the rest of Durbin's procedure can be applied without danger of meeting the bad behavior the original transform may display for X_1-space nearly the same as X_2-space.

4. CONVENIENT FORMULA

Durbin gives a formula for computing his proposed exact test on the basis of two auxiliary regressions, one of y on X_1 and X_2 jointly, and one of X_1 on X_2. The proposed modified version of Durbin's test can be computed on the basis of essentially the same information, and we make the procedure explicit here.

We are given initially the $n \times k$ matrices Z_2 and X_1. We will need to determine $X_{1.2} = X_1 - X_2 = X_1 - Z_2(Z_2'Z_2)^{-1}Z_2'X_1$. This amounts to a set of least-squares regressions of X_1 on Z_2, from which both

residuals $X_{1.2}$ and predicted values X_2 are saved. (The original Durbin procedure required $X_{1.2}$ also, but made no use of the "predicted values" here called X_2 .) We then estimate a least-squares regression of y on X_1 and X_2 jointly, saving the coefficient vector c_2 which applies to X_2 , the residuals $\hat{u}_{12}$, and the covariance matrices G_1 and G_2 of the coefficients corresponding to X_1 and X_2 .[5/] Then taking $P_1 = G_1^{ts}$ and $P_2 = G_2^{ts}$, the modified transform proposed is

$$\hat{u}_M = \hat{u}_{12} - X_{1.2}P_1P_2^{-1}c_2 \quad . \tag{4.1}$$

The general formula for Durbin's transform is nearly identical, being

$$\hat{u}_D = \hat{U}_{12} + X_{1.2}P_1P_2^{-1}c_2 \quad . \tag{4.2}$$

The difference between $\hat{u}_M$ and the Durbin residuals $\hat{u}_D$ is only in the sign of the second term on the right of (4.1) and in the initial "alignment" of X_2 with X_1 .

5. CONSTANT TERMS

In practice, X_1 and X_2 will often both contain a constant vector. In this case $X'_{1.2}X_{1.2}$ and $X'_{2.1}X_{2.1}$ are singular, invalidating the formulas of the preceding section. In this situation one can start with all variables in the form of deviations from means, omitting the constant vector from both X_i's , and the formulas again become applicable. This means that in (4.1), $\hat{u}_{12}$ is the residual vector from a regression of y on X_1 and X_2 jointly, as before, but with one of the constant vectors suppressed in the computations, and the G_{ij} are the appropriate blocks of the covariance matrix of estimated coefficients in the joint regression after the row and

column corresponding to the constant term have been deleted.

6. CONCLUSION

The $\hat{u}_M$ residuals do not in general achieve the A-K-L criterion of minimizing expected squared deviation from the OLS residuals, and their greater computational convenience must be weighed against their not meeting this criterion. On the other hand, there is as yet no analytical argument available as to why minimizing expected squared deviation from OLS residuals should improve statistical properties of tests based on the A-K-L residuals. It seems intuitively plausible that there might be some advantage to the A-K-L transform under some circumstances. However when it is possible to achieve the desired distribution of the transformed residuals by a very small modification of the OLS residuals, the transform proposed in this paper will make nearly as small a change in the OLS residuals as the A-K-L transform; and when a larger change in the OLS residuals is required, it is quite possible that minimizing distance from the OLS residuals is no longer so desirable. Further study of these issues is called for.

FOOTNOTES

*/ Associate Professor, Department of Economics, University of Minnesota, Minneapolis 55455. The author wishes to acknowledge useful comments from a referee and an editor on earlier drafts of the paper. This research was supported by NSF Grant GS36838.

1/ My own view is that, while "typical" X_1 matrices from economics may sometimes have power concentrated at low frequencies in the sense required here, it is not at all clear that this will often remain true when X_1 has more than, say, 3 columns or that it will often remain true after the data have been prefiltered to eliminate gross serial correlation in residuals.

2/ I.e., if A is a positive definite matrix with the representation $W'DW$, $WW' = W'W = I$, D diagonal, $A^{\frac{1}{2}} = W'D^{\frac{1}{2}}W$, where $D^{\frac{1}{2}}$ is the matrix with elements the positive square roots of corresponding elements of D.

3/ Note that the matrix H in (2.3) is not the same as the matrix H in Abrahamse and Koerts [1]. The latter matrix would, with this K and L, be $\begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix}$.

4/ To be precise, we are taking $X_1 = Z\,a + X_2 b$, where $Z'Z = I$, $Z'X_2 = 0$, and letting $a'a/b'b$ go to zero.

5/ In fact, there may be a small computational saving when k is large in finding $X_{2.1}$ directly by least squares regression of X_2 on X_1, finding c_2 by regression of y on $X_{2.1}$, and finding G_1 and G_2 directly as $(X'_{1.2}X_{1.2})^{-1}$ and $(X'_{2.1}X_{2.1})^{-1}$.

REFERENCES

[1] Abrahamse, A.P.J., and J. Koerts, "New Estimators of Disturbances in Regression Analysis", Journal of the American Statistical Association, 66 (March 1971), 71-74.

[2] Abrahamse, A.P.J., and A.S. Louter, "One New Test for Autocorrelation in Least Squares Regression", Biometrika, 58 (April 1971), 53-60.

[3] Durbin, J., "An Alternative to the Bounds Test for Serial Correlation in Least-Squares Regression", Econometrica, 38 (May 1970), 422-429.

[4] Durbin, J., and G.S. Watson, "Testing for Serial Correlation in Least Squares Regression II", Biometrika, 38 (April 1951), 159-78.

[5] Durbin, J., and G.S. Watson, "Testing for Serial Correlation in Least Squares Regression III", Biometrika, 58 (April 1971), 1-20.

DISCUSSION PAPER SERIES

71-1	A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices	Peter K. Clark
71-2	A Note on Approximate Regression Disturbances	Clifford Hildreth
71-3	Induced Innovation in Agricultural Development	Yujiro Hayami and Vernon W. Ruttan
71-4	Wage Structures in Latin America	Peter Gregory
71-5	Optimization and Scale Economies in Urban Bus Transportation	Herbert Mohring
71-6	An Approach to the Study of Money and Nonmoney Exchange Structures	Neil Wallace
71-7	Coalitions, Core, and Competition	Marcel K. Richter
71-8	Instantaneous and Non-Instantaneous Adjustment to Equilibrium in Two-Sector Growth Models	Antonio Bosch Andreu Mas-Colell Assaf Razin
71-9	A Static Nonstationary Analysis of the Interaction between Monetary and Fiscal Policy	Neil Wallace
71-10	Are There Exogenous Variables in Short-Run Production Relations?	Christopher A. Sims
71-11	An Adjusted Maximum Likelihood Estimator of Autocorrelation in Disturbances	Clifford Hildreth Warren T. Dent
71-12	Wage Fund, Technical Progress and Economic Growth	Hukukane Nikaido
71-13	The Economics of Malnourished Children: A Study of Disinvestment in Human Capital	Marcelo Selowsky Lance Taylor
72-14	Industrial Capacity Utilization in Colombia: Some Empirical Findings	Francisco E. Thoumi
72-15	Monopolistic Competition, Objective Demand Functions and the Marxian Labor Value in the Leontief System	Hukukane Nikaido
72-16	The Stability of Models of Money and Growth with Perfect Foresight	Thomas J. Sargent Neil Wallace
72-17	Consumerism: Origin and Research Implications	E. Scott Maynes
72-18	Income and Substitution Effects in Labor Force Participation and Hours of Work	H. Gregg Lewis

72-19	On Investment in Human Capital Under Uncertainty	Assaf Razin
72-20	Ventures, Bets and Initial Prospects (Revised - March 1973)	Clifford Hildreth
72-21	A Competitive Market Model for "Futures" Price Determination	Peter K. Clark
72-22	Rational Expectations and the Dynamics of Hyperinflation	Thomas J. Sargent Neil Wallace
72-23	Seasonality in Regression	Christopher A. Sims
72-24	The Impact of the Wage-Price Freeze on Relative Shares: A Test of Short-Run Market Expectations	Sol S. Shalit Uri Ben-Zion
73-25	A Model of the Eurodollar Market	Charles Freedman
73-26	A Note on Exact Tests for Serial Correlation	Christopher A. Sims
73-27	The Cost of Capital and the Demand for Money by Firms	Uri Ben-Zion
73-28	Distributed Lags	Christopher A. Sims
73-29	Economic Factors Affecting Population Growth: A Preliminary Survey of Economic Analyses of Fertility	T. Paul Schultz
73-30	A Note on the Elasticity of Derived Demand under Decreasing Returns	Assaf Razin
73-31	The Use of Operational Time to Correct for Sampling Interval Misspecification	Peter K. Clark
73-32	International Capital Flows, Interests Rates, and the Money Supply	Charles Freedman
73-33	A Model of Investment under Interest Rate Uncertainty	Elisha A. Pazner Assaf Razin
73-34	An Intergenerational Model of Population Growth	Assaf Razin Uri Ben-Zion
73-35	On Money, Votes and Policy in a Democratic Society	Uri Ben-Zion Zeev Eytan
73-36	Why Don't Unions Charge High Initiation Fees?	N. J. Simler
74-37	Stochastic Dominance and the Maximization of Expected Utility	Leigh Tesfatsion
74-38	A Model of Choice with Uncertain Initial Prospect	Clifford Hildreth Leigh Tesfatsion
74-39	A Note on Exact Tests for Serial Correlation (Revised - April 1974)	Christopher A. Sims