TESTING FOR AUTOCORRELATION IN DYNAMIC LINEAR MODELS*

T. S. BREUSCH

Australian National University

I. Introduction

If the disturbances of a linear model are autocorrelated, ordinary least squares (OLS) estimates of the coefficient parameters are inefficient but unbiased. However, in a dynamic equation where lagged values of the dependent variable appear as regressors, least squares estimates are biased and generally inconsistent. For this reason it is important to have available tests against autocorrelation, particularly when it is a dynamic model which is proposed to be estimated by OLS.

But the standard tests based on OLS residuals, notably that of Durbin and Watson [4] and the extensions by Schmidt [10] and Wallis [12], are invalid when some of the regressors are lagged values of the dependent variable. These tests are attractive because they avoid the obvious approach of estimating the model with the disturbance process incorporated explicitly and making inference from estimates of parameters in the disturbance process. This latter approach is available in the case of dynamic models but is not widely practised because of the relatively heavy computational requirement to obtain maximum likelihood (ML) estimates of the parameters.

Thus the importance of the seminal contribution of Durbin [5] who showed that it was possible to construct (asymptotically) valid tests against autoregressive disturbances using the estimates obtained by OLS under the assumption that no autocorrelation is present. When appropriate account is taken of the interaction between the dynamics of the model and the dynamic nature of autoregressive disturbances, a modification to the standard statistic is obtained without the computational difficulties of estimating the full model.

In section II, the general philosophy of Durbin's derivation is discussed. The argument is, in places, simplified and some points are amplified if they are important for the developments which follow. Durbin's requirement was to obtain a test statistic using only those estimates obtained with the null hypothesis imposed. Therefore, it might be expected to be closely related to the Lagrange multiplier (LM) approach of Aitchison and Silvey [2].

*I would like to thank the participants in the ANU Econometrics Seminar in March 1977 for comments on an earlier version of this paper.

This relationship is explored in sections III and IV. Section III has a discussion of the LM test in the framework adopted by Durbin for his general discussion and in section IV the LM statistic for testing against autoregressive disturbances in a dynamic model is derived. The conclusion is that the LM and Durbin statistics are asymptotically equivalent but differ in that Durbin's statistic uses estimates of autoregressive parameters while the LM statistic uses simple autocorrelations of the OLS residuals.

In section V some special cases are considered. Apart from the familiar Durbin h-statistic for first order autocorrelation, the simplifications of the general form of the statistic for simple kth order and the important case of joint first and fourth are obtained. The relationship between Durbin's test and that of Box and Pierce [3] is not immediately apparent but it is shown how the latter may be obtained from the LM statistic by additional approximations.

Durbin developed his test for the case where the alternative hypothesis is that the disturbances follow an autoregressive (AR) process. Fitts [6] has attempted to use the same general method to obtain a test against the hypothesis that the disturbances are generated by a moving average (MA) but his procedure is not very practicable. In section VI the LM test is obtained for this situation with the surprising result that the statistic is exactly the same as in the AR case.

The possibility of applying the LM approach to testing for composite (ARMA) disturbances is discussed briefly in section VII.

In his original paper, Durbin recognized that his statistic sometimes may not be defined and gave an alternative procedure which avoids this difficulty. In section VIII, the problem is linked to the ways that different elements of the asymptotic information matrix are estimated and the alternative Durbin statistic is shown to be the LM statistic with a particular choice of estimate of the information matrix.

II. Durbin's General Testing Procedure

In this section a somewhat simplified exposition of the general method used by Durbin to derive the test statistics is given. The general framework is a family of models which, for a sample size n, give a likelihood $L(\alpha,\beta)$ depending on two sets α and β of p and q parameters respectively. Denote $\ell(\alpha,\beta)=\log L(\alpha,\beta)$. The likelihood is assumed to be sufficiently regular to give the familiar asymptotic result:

$$\begin{bmatrix} n^{-\frac{1}{2}}\dfrac{\partial \ell(\alpha,\beta)}{\partial \alpha} \\ \\ n^{-\frac{1}{2}}\dfrac{\partial \ell(\alpha,\beta)}{\partial \beta} \end{bmatrix} \equiv \begin{bmatrix} v \\ \\ u \end{bmatrix} \underset{\sim}{\alpha} N(0,\mathcal{I})$$

where $\underset{\sim}{\alpha}$ means "is asymptotically distributed as" and $\mathcal{I}$ is the asymptotic information matrix,

$$\mathcal{I} = plim \left\{ -\frac{1}{n} \begin{bmatrix} \frac{\partial^2 \ell(\alpha,\beta)}{\partial\alpha\partial\alpha'} & \frac{\partial^2 \ell(\alpha,\beta)}{\partial\alpha\partial\beta'} \\ \frac{\partial^2 \ell(\alpha,\beta)}{\partial\beta\partial\alpha'} & \frac{\partial^2 \ell(\alpha,\beta)}{\partial\beta\partial\beta'} \end{bmatrix} \right\}$$

$$\equiv \begin{bmatrix} A & C \\ C' & B \end{bmatrix}$$

It is further assumed that the nature of the likelihood permits solution of the first order conditions for a maximum of the likelihood to give unique ML estimates which are consistent. That is, the MLEs $\hat{\alpha}$ and $\hat{\beta}$ which are given by the joint solution of

$$\frac{\partial \ell(\hat{\alpha},\hat{\beta})}{\partial\alpha} = 0 \quad \text{and} \quad \frac{\partial \ell(\hat{\alpha},\hat{\beta})}{\partial\beta} = 0 \tag{1}$$

are consistent estimators of the parameters α and β.

If the first and second derivatives of $\ell(\alpha,\beta)$ with respect to α and are continuous everywhere in the neighbourhood of the true values α and β, the asymptotic distribution of $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by expanding (1) by the mean value theorem.

$$\frac{\partial \ell(\hat{\alpha},\hat{\beta})}{\partial\alpha} = 0 = \frac{\partial \ell(\alpha,\beta)}{\partial\alpha} + \frac{\partial^2 \ell(\alpha^*,\beta^*)}{\partial\alpha\partial\alpha'}(\hat{\alpha}-\alpha) + \frac{\partial^2 \ell(\alpha^*,\beta^*)}{\partial\alpha\partial\beta'}(\hat{\beta}-\beta) \tag{2a}$$

$$\frac{\partial \ell(\hat{\alpha}\hat{\beta})}{\partial\beta} = 0 = \frac{\partial \ell(\alpha,\beta)}{\partial\beta} + \frac{\partial^2 \ell(\alpha^{**},\beta^{**})}{\partial\beta\partial\alpha'}(\hat{\alpha}-\alpha) + \frac{\partial^2 \ell(\alpha^{**},\beta^{**})}{\partial\beta\partial\beta'}(\hat{\beta}-\beta) \tag{2b}$$

where (α^*,β^*) and (α^{**},β^{**}) are between $(\hat{\alpha},\hat{\beta})$ and (α,β) so that $plim\ \alpha^* = plim\ \alpha^{**} = \alpha$ and $plim\ \beta^* = plim\ \beta^{**} = \beta$. Multiplying through (2) by $n^{-1/2}$ and letting $n \to \infty$,

$$\begin{bmatrix} A & C \\ C' & B \end{bmatrix} \begin{bmatrix} \sqrt{n}(\hat{\alpha}-\alpha) \\ \sqrt{n}(\hat{\beta}-\beta) \end{bmatrix} = \begin{bmatrix} v \\ u \end{bmatrix} \underset{\sim}{a} N(0,\mathcal{I})$$

so that

$$\begin{bmatrix} \sqrt{n}(\hat{\alpha}-\alpha) \\ \sqrt{n}(\hat{\beta}-\beta) \end{bmatrix} \underset{\sim}{a} N(0,\mathcal{I}^{-1}). \tag{3}$$

The hypothesis to be tested specifies that the parameter vector α is equal to some given vector,

$$H_o\!: \ \alpha = \alpha_0 \ .$$

By partitioned inversion of $\mathcal{I}$ in (3) it can be seen that

$$\sqrt{n}(\hat{\alpha}-\alpha) \ \underset{\sim}{a} N\left(0, (A-CB^{-1}C')^{-1}\right)$$

so that an asymptotic test of H_o is provided by the quantity

$$n(\hat{\alpha}-\alpha_0)'(A-CB^{-1}C')(\hat{\alpha}-\alpha_0) \tag{4}$$

When H_o is correct this will be asymptotically distributed as χ^2 (p). But the elements in the covariance matrix in (4) are unknown and a practical test would require that they be estimated. Replacing A, B and C by consistent estimates (by replacing α and β by $\hat{\alpha}$ and $\hat{\beta}$ and/or replacing population moments by sample moments) gives a statistic with the same distribution as (4). This is often called the Wald test or, notably in the time series literature, the χ^2 test by "overfitting". The hypothesis H_o is rejected when the calculated value of the statistic exceeds the appropriate upper point of the χ^2 (p) distribution.

An alternative test statistic which is asymptotically equivalent to (4) is provided by the likelihood ratio procedure. For this, ML estimates of the parameter vector β under the assumption that H_o is correct are also required. This restricted estimator b of β is given by the solution of

$$\frac{\partial \ell\,(\alpha_0, b)}{\partial \beta} = 0 \tag{5}$$

The likelihood ratio is defined as

$$\lambda = \frac{\sup_{\beta} L\,(\alpha_0, \beta)}{\sup_{\alpha,\beta} L\,(\alpha,\beta)} = \frac{L\,(\alpha_0, b)}{L\,(\hat{\alpha},\hat{\beta})}.$$

It is a well known result that in regular problems the likelihood ratio statistic $-2 \log\lambda$ is distributed as χ^2 (p) and provides an equivalent test of H_o with the same asymptotic properties as (4).

In many situations, joint solution of equations (1) is difficult requiring iterative methods while it may be possible to solve for b in (5) relatively easily. Correspondingly, if the true value of β were known, obtaining the ML estimator $\tilde{\alpha}$ of α by solving

$$\frac{\partial \ell\,(\tilde{\alpha},\beta)}{\partial \alpha} = 0 \tag{6}$$

may be a relatively simple operation. Of interest would be a test statistic which uses expressions like (5) and (6) and avoids the need to solve for the "full" ML estimates $\hat{\alpha}$ and $\hat{\beta}$.

With the true value of β known, $\tilde{\alpha}$ is a consistent estimator of α and its asymptotic distribution can be obtained by applying the mean value theorem to (6).

$$\frac{\partial \ell\,(\tilde{\alpha},\beta)}{\partial \alpha} = 0 = \frac{\partial \ell\,(\alpha,\beta)}{\partial \alpha} + \frac{\partial^2 \ell\,(\alpha^*,\beta)}{\partial \alpha \partial \alpha'}\,(\tilde{\alpha}-\alpha) \tag{7}$$

where α^* is between $\tilde{\alpha}$ and α, so that $plim\ \alpha^* = \alpha$. Multiplying through (7) by $n^{-1/2}$ and letting $n \to \infty$,

$$A\sqrt{n}(\tilde{\alpha}-\alpha) = v \underset{\sim}{a} N(0,A)$$

so that

$$\sqrt{n}(\tilde{\alpha}-\alpha) \underset{\sim}{a} N(0,A^{-1})$$

But the true value of β is not known so Durbin constructs his test by considering the ML estimator a of α assuming $\beta{=}b$, that is the solution of

$$\frac{\partial \ell\,(a,b)}{\partial \alpha} = 0 \tag{8}$$

Note that a has been obtained in two steps: firstly assume that $\alpha=\alpha_0$ is correct and find the estimate b of β, then take $\beta=b$ and find the estimate a of α. In the context of testing for autocorrelation in a linear regression model, β would be the vector of regression coefficients and the disturbance variance, α would be the vector of parameters in the disturbance process. When the null hypothesis is $\alpha=0$, the first step is to assume that the disturbances are not autocorrelated and estimate the regression by ordinary least squares. In the second step, the parameters of the disturbance process are estimated from the least squares residuals.

To construct a test of H_0, we need to find the asymptotic distribution of the estimates a. It should be obvious that, *when H_0 is correct*, b is a consistent estimator of β so that a is a consistent estimator of $\alpha(=\alpha_0)$. The additional information needed to form the test statistic is the asymptotic covariance matrix of a.

Durbin calls the "naive" test the one that assumes that a has the same covariance matrix as $\tilde{\alpha}$, *i.e.*, it takes

$$n(a-\alpha_0)'\tilde{A}\,(a-\alpha_0) \tag{9}$$

to be distributed as $\chi^2\,(p)$. However a is calculated using b and applying the mean value theorem to (5) and (8) which define b and a gives when H_0 is correct:

$$\frac{\partial \ell\,(a,b)}{\partial\alpha} = 0 = \frac{\partial \ell\,(\alpha,\beta)}{\partial\alpha} + \frac{\partial^2 \ell\,(\alpha^*,\beta^*)}{\partial\alpha\partial\alpha'}\,(a-\alpha) + \frac{\partial^2 \ell\,(\alpha^*,\beta^*)}{\partial\alpha\partial\beta'}\,(b-\beta) \tag{10a}$$

$$\frac{\partial \ell\,(\alpha_0,b)}{\partial\beta} = 0 = \frac{\partial \ell\,(\alpha,\beta)}{\partial\beta} + \frac{\partial^2 \ell\,(\alpha^{**},\beta^{**})}{\partial\beta\partial\beta'}\,(b-\beta) \tag{10b}$$

where (α^*,β^*) is between (a,b) and (α,β) and where (α^*,β^*) is between (α_0,b) and (α,β). When H_0 is correct, $\alpha=\alpha_0$, so that *plim* $\alpha^*=$*plim* $\alpha^{**}=\alpha$ and *plim* $\beta^*=$*plim* $\beta^{**}=\beta$. Multiplying through (10) by $n^{-1/2}$ and letting $n\to\infty$,

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}\begin{bmatrix} \sqrt{n}\,(a-\alpha) \\ \sqrt{n}\,(b-\beta) \end{bmatrix} = \begin{bmatrix} v \\ u \end{bmatrix} \underset{\sim}{a} N(0,\mathcal{I})$$

so that

$$\begin{bmatrix} \sqrt{n}\,(a-\alpha) \\ \sqrt{n}\,(b-\beta) \end{bmatrix} \underset{\sim}{a} N\left(0,\begin{bmatrix} A^{-1}(A-CB^{-1}C')A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}\right)$$

Since

$$\sqrt{n}\,(a-\alpha) \underset{\sim}{a} N(0,A^{-1}\,(A-CB^{-1}C')A^{-1})$$

the test statistic which is asymptotically $\chi^2\,(p)$ when H_0 is correct is

$$n\,(a-\alpha_0)'\tilde{A}(\tilde{A}-\tilde{C}\tilde{B}^{-1}\tilde{C}')^{-1}\tilde{A}\,(a-\alpha_0) \tag{11}$$

where A, B and C have been replaced by estimates which are consistent under H_0. These may be obtained by replacing α by a or α_0 and β by b and/or replacing population moments by sample moments.

By comparing the covariance matrices in (9) and (11) the properties of the naive test can be established. Unless $C=0$, it would tend to underestimate the significance of an observed value of a because the difference between the covariance matrices,

$$A\,(A - CB^{-1}C')^{-1}A - A,$$

is a non-negative definite matrix. Only when the MLEs of α and β in the full model are asymptotically uncorrelated (so that $C=0$) will the naive test give the correct probability of Type I error.

When testing for autocorrelation in the linear regression model it will be seen in section IV that MLEs of coefficients of exogenous regressors are asymptotically uncorrelated with those of the parameters in the disturbance process. If all regressors are exogenous the naive test will be asymptotically valid, but if some of the regressors are lagged values of the dependent variable then this is no longer true. Estimates a computed from least squares residuals will have smaller asymptotic variance than the corresponding quantities $\tilde{\alpha}$ based on the true disturbances. Durbin's static can be viewed as an adjustment of the naive one so that the test has correct Type I error.

The distribution of the statistics has been discussed only under the assumption that the null hypothesis is true. The question of relative powers of the tests is of considerable interest. By considering alternative hypotheses of the form $\alpha=\alpha_0+\gamma/\sqrt{n}$ with γ fixed, Durbin showed that the statistic he derived has, asymptotically, the same local power as the Wald test and hence as the likelihood ratio test.

III. The Lagrange Multiplier Test and Durbin's Procedure

The test developed by Durbin is of interest when it is relatively easy to estimate the parameters assuming that the null hypothesis is correct compared with the difficulty of estimating the parameters in the "full" model. But there is a considerable body of literature covering the problem of testing using only the estimates obtained with the null hypothesis imposed as a restriction on the parameter space. The work that is probably most familiar to econometricians is the so-called Lagrange multiplier (LM) test exposited in a series of papers by Aitchison and Silvey [1, 2, 11].

The general framework of this procedure is as follows. Suppose there is a sample of size n from some distribution which is known apart from a finite number s of unknown parameters $\theta=(\theta_1, \ldots, \theta_s)$ giving a likelihood $L(\theta)$. The hypothesis to be tested is specified as $p<s$ restrictions on θ,

$$H_O : h_j(\theta) = 0 \qquad j = 1, \ldots, p$$

Aitchison and Silvey approach the problem of testing H_O by considering the $\tilde{\theta}$ and λ which emerge as solutions to the first order conditions for a restricted maximum of the log likelihood, $\ell(\theta)$, by the method of Lagrange multipliers.

$$n^{-1}\frac{\partial}{\partial\theta}\,\ell(\tilde{\theta}) + H\lambda = 0$$

$$h_j(\tilde{\theta}) = 0 \qquad j = 1, \ldots, p$$

where

H is the $(s\mathrm{x}p)$ matrix $\left\{ \dfrac{\partial h_j(\tilde{\theta})}{\partial \theta_i} \right\}$

λ is a $(p\mathrm{x}1)$ vector of Lagrange multipliers.

The idea underlying the test is that if the null hypothesis is correct then the restricted estimator $\tilde{\theta}$ of θ will tend to be very near to the unrestricted MLE, at least in large samples, so that the first derivatives of the log likelihood when evaluated at $\theta = \tilde{\theta}$,

$$\left[\frac{\partial}{\partial \theta} \ell(\tilde{\theta})\right] \tag{12}$$

will be close to zero. But if the hypothesis is false there is no reason for this quantity to be small. The Lagrange multiplier test is based on the statistic[1]

$$n\tilde{\lambda}'H'\tilde{\mathcal{I}}^{-1}H\tilde{\lambda} = \frac{1}{n}\left[\frac{\partial}{\partial \theta} \ell(\tilde{\theta})\right]'\tilde{\mathcal{I}}^{-1}\left[\frac{\partial}{\partial \theta} \ell(\tilde{\theta})\right] \tag{13}$$

where $\tilde{\mathcal{I}}$ is the asymptotic information matrix evaluated under H_O using the restricted estimator $\tilde{\theta}$.

Silvey [11] has shown that, in most practical situations, the LM test is asymptotically equivalent to the likelihood ratio test. That is, when H_O is true, the LM statistic (13) is asymptotically distributed as $\chi^2\,(p)$ and the test which rejects H_O when the statistic is greater than the appropriate upper point of the $\chi^2\,(p)$ distribution has (locally) the same power as the likelihood ratio test.

The LM test has been developed in a very general framework and encompasses the situation considered by Durbin as a special case. The specialization is that the parameter set θ is partitioned into two subsets α and β and the restriction under test is that one of the subsets of parameters equals particular values, *i.e.*

$$\theta = \begin{bmatrix}\alpha\\ \beta\end{bmatrix} \quad \text{where } H_O\colon\ h(\theta) = (I_p : 0)\begin{bmatrix}\alpha\\ \beta\end{bmatrix} - \alpha_0 = 0.$$

In this special case the statistic can be reduced further. The vector in (13) is

$$\left[\frac{\partial}{\partial \theta} \ell(\tilde{\theta})\right] = \begin{bmatrix}\dfrac{\partial \ell(\alpha_0, b)}{\partial \alpha}\\[2ex] \dfrac{\partial \ell(\alpha_0, b)}{\partial \beta}\end{bmatrix} = \begin{bmatrix}\dfrac{\partial \ell(\alpha_0, b)}{\partial \alpha}\\[2ex] 0\end{bmatrix}$$

where the zero in the second equality follows from the definition of the restricted estimator b as in (5). Partitioning the asymptotic information matrix as in section II, the LM

[1] Note that the form of the statistic on the right side of (13) does not require explicitly that estimates of the Lagrange multipliers be computed. It is identical to the "efficient score test" proposed by Rao [see 9, p. 418].

statistic for the general framework considered by Durbin becomes

$$\frac{1}{n}\left[\frac{\partial \ell(\alpha_0,b)}{\partial \alpha}\right]'(\tilde{A}-\tilde{C}\tilde{B}^{-1}\tilde{C}')^{-1}\left[\frac{\partial \ell(\alpha_0,b)}{\partial \alpha}\right] \tag{14}$$

Comparing the LM statistic (14) with Durbin's statistic (11), it can be seen that the present approach does not involve estimation of the α parameters which are given values by the hypothesis. While the general theory discussed above indicates that the LM and Durbin statistics will have the same asymptotic properties, the relationship between them is even stronger. There will be some situations where

$$n^{-1/2}\left[\frac{\partial \ell(\alpha_0,b)}{\partial \alpha}\right] = n^{1/2}\tilde{A}\,(a-\alpha_0) \tag{15}$$

so that (11) and (14) will be identical.[2] Interestingly the situations in which this might occur are closely related to those in which Durbin's approach is most practicable.

Suppose that the log likelihood is quadratic in the α parameters. Then $[\partial\ell(\alpha,\beta)/\partial\alpha]$ will be linear in α making solution of equations such as (6) or (8) a simple operation and this is one of the conditions for the Durbin test to be attractive. Also $[\partial^2\ell(\alpha,\beta)/\partial\alpha\partial\alpha']$ will not be a function of the α parameters; we will write it as $[\partial^2\ell(\cdot,\beta)/\partial\alpha\partial\alpha']$ to indicate the lack of dependence on the value of α· The following expansion of (8) will be exact:

$$\frac{\partial \ell(a,b)}{\partial \alpha} = 0 = \frac{\partial \ell(\alpha_0,b)}{\partial \alpha} + \frac{\partial^2 \ell(\cdot,b)}{\partial\alpha\partial\alpha'}(a-\alpha_0) \tag{16}$$

Thus, if the log likelihood is quadratic in α and if the component A of the information matrix is estimated as

$$\tilde{A} = \left[-\frac{1}{n}\,\frac{\partial^2 \ell(\cdot,b)}{\partial\alpha\partial\alpha'}\right] \tag{17}$$

then (15) follows. The statistics given alternatively by following Durbin's prescription or by the LM method will be exactly the same (subject to the *caveat* in footnote 2).

It will be seen in section IV when the methods are applied to the problem of testing for autoregressive disturbances in a linear regression model with lagged dependent variables that (16) holds. But the two methods do not give exactly the same test statistic; the difference arises because of the estimate used for the component A of the covariance matrix. Unless A is estimated as in (17)—Durbin estimates A not by the right side of (17) but by its limiting value—the two methods will lead to different numerical values for the test statistic. This matter is taken up again in the latter part of section IV.

[2] This assumes that the same estimate of the information matrix is used in both cases.

IV. Testing for Autoregressive Disturbances

In this section the Lagrange multiplier method is applied to the problem of testing for autoregressive disturbances in a linear regression model where some of the regressors are lagged dependent variables. The model is

$$\gamma(L)y_t = \sum_{j=1}^{s} \beta_j x_{jt} + u_t \tag{18}$$

$$\alpha(L)u_t = e_t \qquad t = 1, \ldots, n \tag{19}$$

where $\gamma(L)$ and $\alpha(L)$ are polynomials in the lag operator L:

$$\gamma(L) = 1 - \sum_{j=1}^{m} \gamma_j L^j$$

$$\alpha(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j$$

Assumptions:
- i) x_{jt} are exogenous with $lim \left\{\frac{1}{n} \Sigma\ x_{jt}x_{kt}\right\}$ finite.
- ii) e_t are *i.i.d.* $N(0, \sigma^2)$.
- iii) $\gamma(Z)=0$ and $\alpha(Z)=0$ both have all their roots outside the unit circle.
- iv) $y_0, y_{-1}, \ldots, y_{-m}$ and $u_0, u_{-1}, \ldots, u_{-p}$ are constants.

The hypothesis to be tested is that the disturbances in (18) are not autocorrelated, that is

$$H_o:\ \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0.$$

Estimation of the parameters of this model when H_o is assumed to be true is simple: ML estimates $(\hat{\gamma}, \hat{\beta})$ of the coefficients are given by ordinary least squares applied to (18). Denoting the least squares residuals by $\hat{u}_t$, the ML estimate of the disturbance variance is $\hat{\sigma}^2 = n^{-1}\Sigma\hat{u}_t^2$.

To form the LM statistic (14) we need the first derivatives and the limits of the second derivatives of the log likelihood in the full model. These are to be evaluated under H_o at the restricted (OLS) estimates. For convenience, we use the notation '$\mathfrak{E}$' for this operation. From the above assumptions, the log likelihood is

$$\ell = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\ \Sigma\ e_t^2$$

where e_t is to be considered as a function of the data and the unknown parameters.[3] The

[3] All summations are over t to n unless indicated otherwise.

first derivatives and their values under H_o are as follows:

$$\alpha_j: \quad \frac{1}{\sigma^2} \; \Sigma \; e_t u_{t-j}$$

$$\overset{e}{\to} \frac{1}{\hat{\sigma}^2} \; \Sigma \; \hat{u}_t \hat{u}_{t-j} = n \hat{r}_j; \text{ for } j = 1, \ldots, p$$

where $\hat{r}_j = \Sigma \hat{u}_t \hat{u}_{t-j} / \Sigma \hat{u}_t^2$.

$$\gamma_j \; : \; \frac{1}{\sigma^2} \; \Sigma \, e_t . \, \alpha(L) y_{t-j}$$

$$\overset{e}{\to} \frac{1}{\hat{\sigma}^2} \; \Sigma \, \hat{u}_t y_{t-j} = 0; \text{ for } j = 1, \ldots, m \tag{20a}$$

$$\beta_j \; : \; \frac{1}{\sigma^2} \; \Sigma \, e_t . \, \alpha(L) x_{jt}$$

$$\overset{e}{\to} \frac{1}{\hat{\sigma}^2} \; \Sigma \, \hat{u}_t x_{jt} = 0; \text{ for } j = 1, \ldots, s \tag{20b}$$

$$\sigma^2 \; : \; -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \; \Sigma \, e_t^2$$

$$\overset{e}{\to} -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \; \Sigma \, \hat{u}_t^2 = 0 \tag{20c}$$

Since all of the off-diagonal blocks of $\mathcal{J}$ connecting σ^2 with other parameters are zero, we can ignore the parameter σ^2 without affecting the statistic. Collecting together all of above estimates of elements of the asymptotic information matrix gives

$$\tilde{\mathcal{J}} = \begin{bmatrix} I_p & \hat{G} & 0 \\ \hat{G}' & \multicolumn{2}{c}{\begin{bmatrix} (n\hat{V})^{-1} \end{bmatrix}} \\ 0 & & \end{bmatrix} \begin{array}{l} \} \; p \; \alpha's \\ \} \; m \; \gamma's \\ \} \; s \; \beta's \end{array} \tag{22}$$

where G is the matrix formed from the first m columns of

$$\begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \psi_1 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \psi_{p-1} & \cdots & \psi_1 & 1 & 0 \cdots & 0 \end{bmatrix}$$

Partitioned inversion of (22) gives the test statistic (14) in this application to be

$$n \, \hat{r}' \, [I_p - n \, \hat{G} \hat{V}_{11} \hat{G}']^{-1} \hat{r} \tag{23}$$

where $\hat{V}_{11}$ is the top left $(m \times m)$ block of $\hat{V}$, *i.e.* the estimated covariance matrix of the $\hat{\gamma}$ estimates in OLS estimation of (18). The LM test takes (23) as $\chi^2 \; (p)$, rejecting the null hypothesis of no autocorrelation when the calculated value of the statistic exceeds the appropriate upper point.

It will be obvious that the log likelihood is quadratic in the α parameters so that (21) is not a function of α. But the LM statistic (23) differs from that proposed by Durbin in one significant feature. While the matrix in the quadratic form that defines the statistic is the same in both cases (and we have estimated it in the same way), the vector is different. In the LM statistic it is the vector of autocorrelations

$$\hat{r} = (\hat{r}_1, \ldots, \hat{r}_p)' \quad \text{where} \quad \hat{r}_j = \Sigma \hat{u}_t \hat{u}_{t-j} / \Sigma \hat{u}_t^2$$

while Durbin [5, p. 420] would use the vector of estimates a of α obtained by the least squares regression of $\hat{u}_t$ on $\hat{u}_{t-1}, \ldots, \hat{u}_{t-p}$. These estimates a are given by the solution of

$$\begin{bmatrix} \Sigma \hat{u}_{t-1} \hat{u}_t \\ \vdots \\ \Sigma \hat{u}_{t-p} \hat{u}_t \end{bmatrix} = \begin{bmatrix} \Sigma \hat{u}_{t-1}^2 & \ldots & \Sigma \hat{u}_{t-1} \hat{u}_{t-p} \\ \vdots & & \vdots \\ \Sigma \hat{u}_{t-p} u_{t-1} & \ldots & \Sigma \hat{u}_{t-p}^2 \end{bmatrix} a$$

Comparing this with equation (15), it can be seen that for the two statistics to be the same, not only does the information matrix have to be estimated in the same way in both cases, but the submatrix A corresponding to the α parameters in the disturbance process would have to be estimated as

$$\tilde{A} = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} \Sigma \hat{u}_{t-1}^2 & \ldots & \Sigma \hat{u}_{t-1} \hat{u}_{t-p} \\ \vdots & & \vdots \\ \Sigma \hat{u}_{t-p} \hat{u}_{t-1} & \ldots & \Sigma \hat{u}_{t-p}^2 \end{bmatrix}$$

This agrees with the observation made in the discussion in section III; only when A is estimated by simply evaluating second derivatives at the restricted estimates will the two statistics be the same. In the above derivation of the LM statistic we followed Durbin and estimated A by its limiting value under H_o, *i.e.* $\tilde{A} = I_p$, so the two statistics generally will be different.

There is one simple case, however, where the two statistics coincide. If $p=1$ then $a_1 = \hat{r}_1$ (neglecting end effects possibly arising from summation over different ranges of t) and the statistics given by the LM approach and the Durbin method will be the same.

Note that equations (20) are zero for the same reason that (5) was zero; they are the normal equations for restricted estimation of $(\gamma, \beta, \sigma^2)$.

The vector required to form the LM statistic (14) is therefore

$$n\,\hat{r} \quad \text{where} \quad \hat{r} = (\hat{r}_1, \ldots, \hat{r}_p)'.$$

The other component required for the statistic is an estimate of the asymptotic information matrix

$$\mathcal{I} = plim \left[-\frac{1}{n} \frac{\partial^2 \ell}{\partial\theta \partial\theta'} \right]$$

with $\theta = (\alpha, \gamma, \beta, \sigma^2)$, evaluated under H_o at the restricted estimates. We follow Durbin in estimating the submatrices of $\mathcal{I}$ corresponding to γ and β parameters by replacing population moments with sample moments. The second derivatives of the log likelihood, multiplied by $(-n^{-1})$, and their estimates are as follows:

$$\gamma_j \text{ with } \gamma_k: \quad \frac{1}{n\sigma^2} \Sigma\, [\alpha(L)y_{t-j}.\alpha(L)y_{t-k}]$$

$$\overset{e}{\rightarrow} \quad \frac{1}{n\hat{\sigma}^2} \Sigma\, y_{t-j}y_{t-k}$$

$$\gamma_j \text{ with } \beta_k: \quad \frac{1}{n\sigma^2} \Sigma\, [\alpha(L)y_{t-j}.\alpha(L)x_{kt}]$$

$$\overset{e}{\rightarrow} \quad \frac{1}{n\hat{\sigma}^2} \Sigma\, y_{t-j}x_{kt}$$

$$\beta_j \text{ with } \beta_k: \quad \frac{1}{n\sigma^2} \Sigma\, [\alpha(L)x_{jt}.\alpha(L)x_{kt}]$$

$$\overset{e}{\rightarrow} \quad \frac{1}{n\hat{\sigma}^2} \Sigma\, x_{jt}x_{kt}$$

The submatrix of $\tilde{\mathcal{I}}$ corresponding to γ and β parameters can be written as $(n\hat{V})^{-1}$ where $\hat{V}$ is the estimated covariance matrix of the regression coefficients when OLS is applied to (18). This provides a convenient way of obtaining consistent estimates of these elements of the information matrix.

For the remaining elements of $\tilde{\mathcal{I}}$, we take limits under H_0 and then replace parameters by estimates. The notation '$\overset{p}{\rightarrow}$' is used to indicate probability limit under the null hypothesis.

$$\alpha_j \text{ with } \alpha_k: \quad \frac{1}{n\sigma^2} \Sigma\, u_{t-j}u_{t-k} \tag{21}$$

$$\overset{p}{\rightarrow} \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

$$\alpha_j \text{ with } \gamma_k: \quad \frac{1}{n\sigma^2} \Sigma\, [u_{t-j}.\alpha(L)y_{t-k} + e_t y_{t-k}]$$

$$\overset{p}{\rightarrow} \begin{cases} 0 & k > j \\ \psi_{j-k} & j \geqslant k \end{cases}$$

where the ψs are coefficients of the power series expansion of the inverse of lag operator polynomial on y_t, that is,

$$\gamma(L)^{-1} = \sum_{l=0}^{\infty} \psi_l L^l \quad \text{with } \psi_0 = 1.$$

$$\alpha_j \text{ with } \beta_k: \quad \frac{1}{n\sigma^2} \Sigma\, [u_{t-j}.\alpha(L)x_{kt} + e_t x_{kt}] \overset{p}{\rightarrow} 0$$

$$\alpha_j \text{ with } \sigma^2: \quad \frac{1}{n\sigma^4} \Sigma\, u_{t-j}e_t \overset{p}{\rightarrow} 0$$

γ_j with σ^2 : $\quad \dfrac{1}{n\sigma^4} \; \Sigma\alpha(L) y_{t-j} . e_t \overset{p}{\to} 0$

β_j with σ^2 : $\quad \dfrac{1}{n\sigma^4} \; \Sigma\alpha(L) x_{jt} . e_t \overset{p}{\to} 0$

σ^2 with σ^2 : $\quad \dfrac{1}{2\sigma^4} + \dfrac{1}{n\sigma^6} \; \Sigma \; e_t^2 \overset{p}{\to} \dfrac{1}{2\sigma^4}.$

V. Special Cases

Some interesting simplifications to the LM test statistic (23) are possible in certain cases.

The first is the well known Durbin h-statistic for first order autocorrelation ($p=1$). Then $G=(1 \;\; 0 \; . \; . \; . \; 0)$ and $\hat{r}=\hat{r}_1$ so that the test statistic which is taken as χ^2 (1) is $n \, \hat{r}_1{}^2 \; [1-n \, \hat{V}(\hat{\gamma})]^{-1}$ or using the square root of the statistic,

$$h = \hat{r}_1 \sqrt{\frac{n}{1-n\hat{V}(\hat{\gamma}_1)}} \tag{24}$$

is taken as a standard normal deviate. As was pointed out at the end of the last section, the difference between the LM and Durbin statistics does not arise when the methods are applied to testing for a first order autoregression.

There is one case (apart from the obvious one of no lagged dependent variables as regressors) in which the "naive" test is valid. If the only lags in y_t are those of higher order than p then $G=(0 \; . \; . \; . \; 0)$ and the statistic is $n\hat{r}'\hat{r}$.

Another sort of simplification is possible when only the first lag of the dependent variable appears as an explanatory variable. Then $m=1$ and $\psi_j=\gamma^j$ for $j=0,1,2, \; . \; . \; . \;$. If we are interested in testing just one particular parameter, *i.e.*, $H_o : \alpha_k=0$ where $\alpha_j=0$ for $j \neq k$ under both null and alternative hypotheses, then $G=\psi_{k-1}=\hat{\gamma}^{k-1}$. The LM and Durbin statistics coincide having the simple form $n\hat{r}_k^2 \, [1-n\hat{\gamma}^{2(k-1)} \; \hat{V}(\hat{\gamma})]^{-1}$ which is distributed as χ^2 (1). Alternatively its square root, $\hat{h}_k=\hat{r}_k \; \{n/[1-n\hat{\gamma}^{2(k-1)} \, \hat{V}(\hat{\gamma})]\}^{1/2}$, can be taken as N (0,1). This form of the statistic gives a convenient test against seasonal autocorrelation (*e.g.*, $k=4$ in a quarterly model).

Of particular interest in many econometric applications would be a test against joint first and fourth autocorrelation when the basic model has one lagged dependent variable ($m=1$). The LM statistic becomes

$$n \; [\hat{r}_1 \; \hat{r}_4] \; [I_2 - n \begin{bmatrix} 1 \\ \hat{\psi}_3 \end{bmatrix} \hat{V}(\hat{\gamma}) \; (1 : \hat{\psi}_3)]^{-1} \begin{bmatrix} \hat{r}_1 \\ \hat{r}_4 \end{bmatrix}$$

$$= \quad n \frac{\hat{r}_1^2 + \hat{r}_4^2 - n\hat{V}(\hat{\gamma}) \, (\hat{r}_1 \, \hat{\gamma}^3 - \hat{r}_4)^2}{1 - n\hat{V}(\hat{\gamma}) \, (1 + \gamma^6)}$$

which would be taken as χ^2 (2).

There is one other important class of models where a substantial simplification may be made, but only with an additional approximation which may in certain circumstances be reasonable. In the last sentence of his paper Durbin says:

> "The result for the case where the x's are absent . . . and where $p > 1$ was obtained by Box and Pierce by a different method."

Box and Pierce [3] consider a model with no exogenous regressors, that is the straight autoregression

$$\gamma(L) y_t = u_t \tag{25}$$

where $\gamma(L)$ is defined in section IV, and they find the approximate asymptotic distribution of the first p (assumed $p > m$) autocorrelations of the least squares residuals which we have called the vector $\hat{r}$. Under the hypothesis that the disturbances u_t are not autocorrelated, the Box-Pierce test takes

$$Q = n \hat{r}'\hat{r} = n \sum_{j=1}^{p} \hat{r}_j^2$$

and treats this as $\chi^2\,(p-m)$. Now this statistic is different from the Durbin one in that it uses residual autocorrelations rather than estimates of autoregressive parameters. The LM statistic developed in section IV uses the autocorrelations $\hat{r}$ and it would be instructive to compare the LM statistic with that developed by Box and Pierce. For testing the hypothesis $H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$ when the disturbances in (25) are p th order autoregressive,[4] *i.e.* $\alpha(L) u_t = e_t$ where $\alpha(L)$ is defined in section IV, the LM statistic would be

$$n \hat{r}' [I_p - \hat{G}\hat{W}^{-1}\hat{G}']^{-1} \hat{r} \tag{26}$$

where

$$W_{jk} = plim \left\{ \frac{1}{n\sigma^2} \Sigma \alpha(L) y_{t-j} \cdot \alpha(L) y_{t-k} \right\}$$

evaluated under H_0. The W matrix was previously estimated by the covariance matrix of the (OLS) estimates of the γ parameters with H_0 imposed. An alternative estimate could be made after taking the limit under the null hypothesis:

$$\begin{aligned} W_{jk} &= \frac{1}{\sigma^2} plim \left\{ \frac{1}{n} \Sigma \gamma(L)^{-1} e_{t-j} \cdot \gamma(L)^{-1} e_{t-k} \right\} \\ &= \sum_{l=0}^{\infty} \psi_l \psi_{l+|j-k|} \end{aligned} \tag{27}$$

[4] In the next section it is shown that exactly the same statistic arises if, under the alternative hypothesis, the disturbances follow a pth order moving average rather than the autoregression specified here.

From the definition of the matrix G in the previous section, the (j,k)th element of $G'G$ would be

$$\sum_{l=0}^{p-g} \psi_l \psi_{l+|j-k|} \tag{28}$$

where $g=\max(j,k)$. If (27) is approximated by (28) then $W=G'G$ and the matrix in (26) becomes $[I_p-\hat{G}(\hat{G}'\hat{G})^{-1}\hat{G}']$ which is singular (in fact it is idempotent to rank $p-m$), therefore the inverse which appears in (26) does not exist. This difficulty arises because, with the approximation that is made, the joint asymptotic distribution of the elements of the vector $\hat{r}$ has an idempotent covariance matrix of rank $p-m$. Thus, by an extension of theorem (4.8) of Graybill [7], $n\,\hat{r}'\hat{r}$ will be asymptotically distributed as $\chi^2\ (p-m)$ when H_0 is correct.

Comparing (27) and (28) it can be seen that the approximation to arrive at the Box-Pierce statistic involves truncation of terms in ψ_l, the coefficients in the expansion of the inverse of the polynomial $\gamma(L)$. The worst approximation occurs when j or k is equal to m when the summation is truncated after $p-m$ terms. For the approximation to be a good one it is required that p be much larger than m and that the roots of $\gamma(Z)=0$ be well outside the unit circle so that the ψ_l go quickly to zero as l increases.

The LM or Durbin statistics do not involve this approximation and are applicable in more general circumstances such as $p<m$ or when the model includes exogenous regressors. But, except in very simple cases, the Box-Pierce statistic (if applicable) provides a test which is much simpler computationally.

This connection between the LM statistic and the one proposed by Box and Pierce sheds some light on what appears to be an arbitrary choice of the number of autocorrelations to use in forming the latter statistic. In effect the Box-Pierce test is against an alternative that the disturbances follow a pth order autoregression (or a pth order moving average as will be shown in the next section).

VI. Testing for Moving Average Autocorrelation

The statistics we have been considering are appropriate for testing the null hypothesis of no autocorrelation in the disturbances against the alternative that the disturbances are generated by an autoregressive process (AR) as in (19). In recent years considerable attention has been given to other processes which generate autocorrelated disturbances. An important alternative model is the moving average (MA)

$$u_t = \alpha(L)e_t \tag{29}$$

where the e_t are *i.i.d.* $N(0,\sigma^2)$ and

$$\alpha(L) = 1 + \sum_{j=1}^{p} \alpha_j L^j$$

While little work has been done on the problem of testing for MA autocorrelation, the methods considered in sections II and III could provide comparable tests for this type of autocorrelation.

The advantage of the Durbin or LM approaches to testing for AR disturbances in a dynamic model was that the relatively complicated estimation problem was avoided. However the method of Hatanaka [8] provides a simple two-step method of estimating the full model without iteration but yields estimates with the same asymptotic properties as maximum likelihood. In situations where iterative solutions are not feasible it is still possible to apply the Wald test since the two-step estimates can be used in forming the test statistic.

It is probably more important to have tests against MA disturbances which do not require estimation of the full model for several reasons. One is the inherent difficulty of maximum likelihood estimation of models with MA disturbances. This problem has been given a lot of attention and while satisfactory methods are available they still require iterative solutions. In the case of an MA no efficient two-step estimator akin to the Hatanaka estimator is available so a test which does not require estimates of parameters in the full model would be of considerable interest.

The possibility of using the Durbin approach as outlined in section II for testing the null hypothesis that the α parameters in the disturbance process (29) for the dynamic regression model (18) are all zero has been considered by Fitts [6]. He takes the simple case of one lagged dependent variable and one exogenous regressor and a MA (1) disturbance, $p=m=s=1$. Fitts approached the problem in exactly the same way as Durbin exposited his test for AR autocorrelation. That is, assume $\alpha=0$ and estimate γ, β and σ^2 parameters by OLS then estimate the α parameters in (29) using the OLS residuals $\hat{u}_t$ instead of the true disturbances u_t. The problem here is that, unlike estimating an autoregression in the residuals, estimating the parameters of even a simple moving average is not a one-step operation. Iterative methods are still required to estimate α from the OLS residuals. In addition, the expression obtained by Fitts for the asymptotic variance [6, p. 372] is very complicated relative to that given by Durbin for the AR disturbances case. This approach is of little practical importance since it offers little, if any, computational advantage over the Wald test or likelihood ratio test because its complexity rivals that of estimation under the alternative hypothesis.

A relatively simple test which requires estimation only under the null hypothesis of no autocorrelation is available. Although the Durbin and LM methods generally do not coincide in the MA case (the log likelihood is not quadratic in the parameters under test) and the Durbin approach does not avoid the complexity of estimation of the full model, the LM statistic is readily derived. The remarkable result established here is that the LM statistic for testing against autocorrelation of the MA type is exactly the same as that for testing against an AR of the same order. Surprisingly, the well known Durbin h-statistic for first order AR in the disturbances would also be appropriate for testing against a first order MA.

The assumptions, derivation and notation parallel those used in section IV for the AR case and will not be restated. Again the null hypothesis is $H_O: \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$. The only difference here is that the disturbances u_t are now assumed to follow the MA process (29) instead of the AR in (19). The first derivatives of the log likelihood with respect to α and their values when evaluated at the estimates under H_0 are:

$$\alpha_j : \frac{1}{\sigma^2} \Sigma\, e_t.\alpha(L)^{-2}u_{t-j}$$

$$\overset{e}{\to} \frac{1}{\hat{\sigma}^2} \Sigma \hat{u}_t \hat{u}_{t-j} = n\hat{r}_j$$

As before, the other first derivatives under H_0 are zero giving the normal equations for OLS estimates of the regression parameters. The second derivatives with respect to γ and β parameters multiplied by $(-n^{-1})$ and their estimates under H_0 are as follows:

$$\gamma_j \text{ with } \gamma_k \; : \; \frac{1}{n\sigma^2} \Sigma\,[\alpha(L)^{-1}y_{t-j}.\alpha(L)^{-1}y_{t-k})]$$

$$\overset{e}{\to} \frac{1}{n\hat{\sigma}^2} \Sigma\, y_{t-j}y_{t-k}$$

$$\gamma_j \text{ with } \beta_k \; : \; \frac{1}{n\sigma^2} \Sigma\,[\alpha(L)^{-1}y_{t-j}.\alpha(L)^{-1}x_{kt}]$$

$$\overset{e}{\to} \frac{1}{n\hat{\sigma}^2} \Sigma\, y_{t-j}x_{kt}$$

$$\beta_j \text{ with } \beta_k \; : \; \frac{1}{n\sigma^2} \Sigma\,[\alpha(L)^{-1}x_{jt}.\alpha(L)^{-1}x_{kt}]$$

$$\overset{e}{\to} \frac{1}{n\hat{\sigma}^2} \Sigma\, x_{jt}x_{kt}$$

The estimates of the elements of the asymptotic information matrix corresponding to γ and β parameters are exactly the same as in the AR disturbance case of section IV.

For the other elements we take the limits as $n\to\infty$ of $(-n^{-1})$ times the second derivatives under H_0.

$$\alpha_j \text{ with } \alpha_k : \frac{1}{n\sigma^2} \Sigma\,[\alpha(L)^{-2}u_{t-j}.\alpha(L)^{-2}u_{t-k} + 2\, e_t.\alpha(L)^{-3}u_{t-j-k}]$$

$$\overset{p}{\to} \begin{cases} 1 & j=k \\ 0 & j\neq k \end{cases}$$

$$\alpha_j \text{ with } \gamma_k : \frac{1}{n\sigma^2} \Sigma\,[\alpha(L)^{-2}u_{t-j}.\alpha(L)^{-1}y_{t-k} + e_t.\alpha(L)^{-2}y_{t-j-k}]$$

$$\overset{p}{\to} \begin{cases} 0 & k>j \\ \psi_{j-k} & j\geqslant k \end{cases}$$

where, as before, $\gamma(L)^{-1} = \sum_{l=0}^{\infty} \psi_l L^l$ with $\psi_0 = 1$

$$\alpha_j \text{ with } \beta_k : \frac{1}{n\sigma^2} \Sigma\,[\alpha(L)^{-2}u_{t-j}.\alpha(L)^{-1}x_{kt} + e_t.\alpha(L)^{-2}x_{k,t-j}] \overset{p}{\to} 0$$

$$\alpha_j \text{ with } \sigma^2: \frac{1}{n\sigma^4} \Sigma e_t . \alpha(L)^{-2} u_{t-j} \overset{p}{\rightarrow} 0$$

$$\gamma_j \text{ with } \sigma^2: \frac{1}{n\sigma^4} \Sigma e_t . \alpha(L)^{-2} u_{t-j} \overset{p}{\rightarrow} 0$$

$$\beta_j \text{ with } \sigma^2: \frac{1}{n\sigma^4} \Sigma e_t \, \alpha(L)^{-2} x_{jt} \overset{p}{\rightarrow} 0$$

$$\sigma^2 \text{ with } \sigma^2: -\frac{1}{2\sigma^4} + \frac{1}{n\sigma^6} \Sigma e_t^2 \overset{p}{\rightarrow} \frac{1}{2\sigma^4}$$

Comparing the above quantities with the corresponding ones of section IV, it can be seen that they are exactly the same. The LM test statistic for the MA case is exactly the same as for the AR case. This probably says something about the power (or lack thereof) of procedures such as the LM statistic for testing against parametric autocorrelation processes. The same statistic arises whether the form of autocorrelations allowed under the alternative hypothesis is MA or AR. The LM test would seem not to exploit an important aspect of the prior information implied in hypothesizing a particular alternative process.

The discussion in previous sections showed that in the case of a first order AR disturbance, the Durbin h-statistic could be obtained as the LM statistic. Since the LM statistic is the same whether the disturbances are AR or MA the h-statistic would be appropriate when the alternative hypothesis is a first order MA.

VII. Composite Disturbances

We have shown that the Lagrange multiplier test provides the same statistic for testing against pth order autocorrelation whether generated by an AR or an MA process. It is interesting to consider what form the statistic would take if the disturbances are hypothesized to follow a composite ARMA process.

For simplicity we consider only the first order ARMA (1,1) case but extension to higher order processes would be relatively straightforward using the results of sections IV and VI. The model is the linear regression with lagged dependent variables as in (18) where the disturbances now follow

$$(1-\phi L)u_t = (1+\alpha L)e_t$$

and the e_t are *i.i.d.* $N(0,\sigma^2)$ as before. The hypothesis to be tested specifies that both the AR and MA parameters are zero, $H_o: \phi=\alpha=0$ so that the u_t disturbances are independent.

Under the null hypothesis, ML estimates are given by applying OLS to (18) and the test would be constructed as in section IV. However, the method used previously will break down because the information matrix of the "full" model will be singular when H_o is imposed. As we have seen, the statistic is exactly the same when either an AR or MA process is assumed. Allowing both to be present introduces into the information matrix identical rows (and columns) which makes the matrix singular. Another way of looking at the same problem is to note that, under H_o, ϕ and α assume the same numerical value, zero. Both sides of the ARMA process then have the same root so that ϕ and α are not separately identifiable and this problem manifests itself as a singular information matrix. The vector in the LM statistic would be $(\hat{r}_1\hat{r}_1)'$. Appropriate treatment, whether by taking a generalized inverse or by deleting one of the duplicate rows and columns and the corresponding element in the vector in the test statistic, would lead to the same statistic as in the simple AR or MA cases.

VIII. Alternative Forms of the Statistic

One difficulty with the Durbin statistic (shared by the LM statistic as developed in section IV) is that the information matrix as estimated may not be positive definite. In following Durbin, we have estimated some elements of the matrix by their limiting values (as $n \to \infty$) and others simply by evaluating second derivatives with population moments replaced by sample moments. This was done because it provided a convenient means of obtaining an estimate of the information matrix using the usual computational output from (OLS) estimation under the null hypothesis when the alternative hypothesis is first order autoregressive disturbances. The possible failure of the matrix to be positive definite means that the computed value of the statistic (a quadratic form) may be negative and hence meaningless as a χ^2 variable. In the simple case of testing for a first order autoregression, the h-statistic (24) will be undefined if $n\hat{V}(\hat{\gamma}_1) > 1$.

An alternative test which does not suffer from this difficulty was also proposed by Durbin. The alternative approach requires two OLS regressions and is almost as easy to apply in simple cases as the h-statistic. Generally it will be more convenient than direct extensions of the h-statistic when testing for higher orders of autocorrelation. He obtained the alternative procedure by a separate derivation in which a quadratic approximation is made to the log likelihood by expanding it (in the general framework of our section II) about α and b. We give another interpretation of the test as the LM statistic of section IV with the matrix in the quadratic form estimated in a way which ensures its positive definiteness.

We consider the regression model with lagged dependent variables of section IV in which we wish to test the null hypothesis of no autocorrelation in the disturbances against an alternative of a pth order autoregression. The model (18) and (19) can be written more conveniently for the present purpose as

$$y_t = Y_t\gamma + X_t\beta + u_t$$

$$u_t = U_t\alpha + e_t$$

where

$$Y_t = \{y_{t-1}, \ldots, y_{t-m}\}$$

$$X_t = \{x_{1t}, \ldots, x_{st}\}$$

$$U_t = \{u_{t-1}, \ldots, u_{t-p}\}$$

for $t=1, \ldots, n$. We wish to test $H_0 : \alpha=0$. All of the observations can be written in compact matrix form as follows:

$$y = Y\gamma + X\beta + u$$

$$u = U\alpha + e$$

Denoting $Z=(Y : X)$ and $\delta'=(\gamma' : \beta')$, the OLS estimates of the parameters in the model restricted by H_0 are $\hat{\delta}=(Z'Z)^{-1}Z'y$ with residuals $\hat{u}=y-Z\hat{\delta}$.

The principal distinction to be made in this section concerns the way in which the information matrix is estimated. Instead of estimating $\mathscr{I}$ as in (22) where limits under H_0 were taken for some components, we could have obtained consistent estimates of all components of $\mathscr{I}$ by evaluating $(-n^{-1})$ times the second derivatives at the restricted estimates of the parameters as was done for other components of $\mathscr{I}$. This would give

$$\text{for } \alpha_j \text{ with } \alpha_k: \quad \frac{1}{n\hat{\sigma}^2} \Sigma \hat{u}_{t-j}\hat{u}_{t-k}$$

$$\text{for } \alpha_j \text{ with } \gamma_k: \quad \frac{1}{n\hat{\sigma}^2} \Sigma \hat{u}_{t-j}y_{t-k}$$

$$\text{for } \alpha_j \text{ with } \beta_k: \quad \frac{1}{n\hat{\sigma}^2} \Sigma \hat{u}_{t-j}x_{kt}$$

where, in the notation introduced above, $\hat{\sigma}^2 = n^{-1} \hat{u}'\hat{u}$. The alternative estimate of $\mathscr{I}$ is then

$$\tilde{\mathscr{I}} = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} \hat{U}'\hat{U} & \hat{U}'Y & \hat{U}'X \\ Y'\hat{U} & Y'Y & Y'X \\ X'\hat{U} & X'Y & X'X \end{bmatrix} = \frac{1}{n\hat{\sigma}^2} \begin{bmatrix} \hat{U}'\hat{U} & \hat{U}'Z \\ Z'\hat{U} & Z'Z \end{bmatrix} \tag{30}$$

In similar notation, the vector in the test statistic is

$$n\,\hat{r} = n\,\frac{\hat{U}'\hat{u}}{\hat{u}'\hat{u}} = \frac{1}{\hat{\sigma}^2}\,\hat{U}'\hat{u}$$

Combining these into the quadratic form which constitutes the LM statistic gives

$$\hat{u}'\hat{U}\left\{\hat{U}'\hat{U} - \hat{U}'Z(Z'Z)^{-1}Z'\hat{U}\right\}^{-1}\hat{U}'\hat{u} \,/\, \hat{\sigma}^2 \tag{31}$$

Note that the matrix (30) can be written as $\tilde{\mathscr{I}}=P'P$ where P is a matrix of real elements. Unless the columns of P are linearly dependent, and there is no general reason why this should be so, the matrix $\tilde{\mathscr{I}}$ will be positive definite.

From the orthogonality of least squares regressors and residuals, $Z'\hat{u}=0$. Thus (31) can be interpreted as nR^2 where R^2 is the usual coefficient of determination in the regression of $\hat{u}$ on $\hat{U}$ and Z. The LM test is performed by taking nR^2 from this second application of OLS as $\chi^2\,(p)$.

Several variants of this procedure, using different aspects of the output from the second regression yield statistics with exactly the same asymptotic properties. Firstly, instead of taking R^2 which is the ratio of explained to total sums of squares, we could take the ratio of explained to residual sums of squares. Denote the vector of residuals from the second regression by $\hat{e}$ and let $S^2 = n^{-1}\hat{e}'\hat{e}$. Then the sum of squares decomposition is

$$\hat{u}'\hat{u} = \hat{e}'\hat{e} + \hat{u}'\hat{U}\{\hat{U}'\hat{U} - \hat{U}'Z(Z'Z)^{-1}Z'\hat{U}\}^{-1}\hat{U}'\hat{u}.$$

Dividing through by $n\hat{\sigma}^2$ gives

$$1 = S^2 / \hat{\sigma}^2 + \hat{u}'\hat{U}\{\hat{U}'\hat{U} - \hat{U}'Z(Z'Z)^{-1}Z'\hat{U}\}^{-1}\hat{U}'\hat{u} / n\hat{\sigma}^2 \tag{32}$$

Consider the second term on the right side of (32) which is (n^{-1}) times the LM statistic as given in (31). If (31) has an asymptotic distribution under H_0 and for local alternatives then (n^{-1}) times (31) converges in probability to zero. Hence $plim(S^2/\hat{\sigma}^2) = 1$ so that the asymptotic properties of the statistic are unaffected if S^2 is used in place of $\hat{\sigma}^2$ in (31). This form of the LM statistic

$$\hat{u}'\hat{U}\{\hat{U}'\hat{U} - \hat{U}'Z(Z'Z)^{-1}Z'\hat{U}\}^{-1}\hat{U}'\hat{u} / S^2 \tag{33}$$

taken as $\chi^2(p)$, will be recognized as the usual statistic for testing the overall significance of the second regression. (Usual, that is, when one is limited to tests based on asymptotic theory so that F-tests are not applicable). While (33) will be numerically larger than (31) it provides a test with exactly the same asymptotic properties.

In the regression of $\hat{u}$ on $\hat{U}$ and Z, the estimate of the vector of coefficients of $\hat{U}$ would be

$$d = \{\hat{U}'\hat{U} - \hat{U}'Z(Z'Z)^{-1}Z'\hat{U}\}^{-1}\hat{U}'\hat{u}$$

so (31) can also be written as

$$d'\{\hat{U}'\hat{U} - \hat{U}'Z(Z'Z)^{-1}Z'\hat{U}\}d / S^2 \tag{34}$$

This would be the quantity taken as $\chi^2(p)$ in the usual (asymptotic) test of the significance of the coefficients of $\hat{U}$ and is precisely Durbin's alternative test [5, p. 420].

Another variant of the statistic could use y instead of $\hat{u}$ as the dependent variable in the second application of OLS. From elementary regression theory, it will be known that the estimated coefficients of $\hat{U}$ and the estimated covariance matrix of coefficient estimates remain unchanged. Hence the test of significance of these coefficients with y as dependent variable provides the same statistic as (34) above. This formulation of the LM test is related to the Hatanaka [8] two-step efficient procedure for this problem. The Hatanaka estimator provides consistent and asymptotically efficient estimates of the parameters when the alternative hypothesis is assumed to be true but requires the first step estimates to be consistent under the alternative. In the present context, the LM statistic as in (34) can be viewed as a significance test based on a Hatanaka type estimator but with first step estimates given by OLS for regression coefficients and zeros for autocorrelation parameters. While these initial estimates do not allow desirable properties to be assigned to the second step estimates when the alternative hypothesis is true, the relationship with the LM statistic indicates that the test for significance of the coefficients of $\hat{U}$ does have appropriate asymptotic properties.

The forms of the LM statistic discussed in this section avoid the difficulties associated with the better known Durbin procedure. They have the additional attraction that they are simple to apply, requiring two OLS regressions to compute them.

REFERENCES

1. J. Aitchison and S. D. Silvey, "Maximum-likelihood estimation of parameters subject to restraints", *Annals of Mathematical Statistics*, vol. 29, 1958.
2. J. Aitchison and S. D. Silvey, "Maximum-likelihood estimation procedures and associated tests of significance", *Journal of the Royal Statistical Society (B)*, vol. 22, 1960.
3. G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models", *Journal of the American Statistical Association*, vol. 65, 1970.
4. J. Durbin and G. S. Watson, "Testing for serial correlation in least-squares regression, I and II", *Biometrika*, vol. 37, 1950 and vol. 38, 1951.
5. J. Durbin, "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables", *Econometrica*, vol. 38, 1970.
6. John Fitts, "Testing for autocorrelation in the autoregressive moving average error model", *Journal of Econometrics*, vol. 1, 1973.
7. F. A. Graybill, *An Introduction to Linear Statistical Models, Volume 1* (New York: McGraw-Hill, 1961).
8. M. Hatanaka, "An efficient two-step estimator for the dynamic adjustment model with autoregressive errors", *Journal of Econometrics*, vol. 2, 1974.
9. C. R. Rao, *Linear Statistical Inference and Its Applications* (New York: Wiley, 1973).
10. P. Schmidt, "A generalization of the Durbin-Watson test", *Australian Economic Papers*, vol. 11, 1972.
11. S. D. Silvey, "The Lagrangian multiplier test", *Annals of Mathematical Statistics*, vol. 30, 1959.
12. K. F. Wallis, "Testing for fourth order autocorrelation in quarterly regression equations", *Econometrica*, vol. 40, 1972.