Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables
Author(s): L. G. Godfrey
Source: *Econometrica*, Vol. 46, No. 6 (Nov., 1978), pp. 1293-1301
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/1913829
Accessed: 23/09/2014 15:35

Econometrica, Vol. 46, No. 6 (November, 1978)

TESTING AGAINST GENERAL AUTOREGRESSIVE AND MOVING AVERAGE ERROR MODELS WHEN THE REGRESSORS INCLUDE LAGGED DEPENDENT VARIABLES

BY L. G. GODFREY[1]

Since dynamic regression equations are often obtained from rational distributed lag models and include several lagged values of the dependent variable as regressors, high order serial correlation in the disturbances is frequently a more plausible alternative to the assumption of serial independence than the usual first order autoregressive error model. The purpose of this paper is to examine the problem of testing against general autoregressive and moving average error processes. The Lagrange multiplier approach is adopted and it is shown that the test against the nth order autoregressive error model is exactly the same as the test against the nth order moving average alternative. Some comments are made on the treatment of serial correlation.

1. INTRODUCTION

THE THEORY OF TESTING for serial correlation when the regressors of a regression equation include lagged dependent variables was greatly advanced by Durbin's [**1**] work. Durbin [**1**] developed a general procedure for testing for specification error and applied it to the problem of testing the null hypothesis of serial independence against the alternative that the errors of the regression model were generated by a stable first order autoregressive (AR(1)) process. His procedure yielded a very simple test, known as the h test, which was asymptotically equivalent to the appropriate likelihood ratio (LR) test.[2] This test is now widely used in empirical work.

However, given that regression relationships with lagged dependent variables can often be regarded as transformed versions of some sort of rational distributed lag model, it is clear that the moving average error model of order n (MA(n)) is frequently a more plausible alternative hypothesis than the AR(1) scheme (see Nicholls et al. [**7**, Section 2] and Sims [**11**, Section 6]). Unfortunately, Durbin's general procedure does not provide simple non-iterative tests against MA error processes (see Fitts [**2**, Section 4] for a discussion of the iterative procedure required for the MA(1) case).

The power of the h test against MA(n) errors is open to question (see Sims [**11**, pp. 320–321]), and Wallis [**12**, Sections 3.2 and 3.4] has provided some numerical examples which illustrate the inability of the h test to detect fourth order autocorrelation. The purpose of this paper is, therefore, to propose large sample tests of the serial independence assumption appropriate for the general alternative hypotheses of AR(n) and MA(n) errors. These tests are based upon Silvey's [**10**] Lagrange multiplier approach and have the properties that they are asymptotically equivalent to the corresponding LR tests and so share their

[1] The author is grateful to Christine Godfrey, D. S. Poskitt, A. R. Tremayne, and the referees for helpful comments. The author is also indebted to Dr. D. F. Hendry for his permission to refer to unpublished work.

[2] By "asymptotically equivalent", we mean that the two test statistics have the same asymptotic distribution under the null hypothesis, and the asymptotic relative efficiency of one compared to the other is unity.

desirable large sample properties, and they can be easily added to regression programs and do not require iterative calculations.

The contents of this paper are as follows. Section 2 contains details of notation and assumptions, and some general remarks on the use of the Lagrange multiplier technique. The tests for alternative hypotheses of AR(n) and MA(n) models are considered in Section 3, and the problems associated with testing against a mixed-autoregressive-moving-average (ARMA) process are discussed in Section 4. It is shown that the tests derived below depend only upon the order of the error process under the alternative hypothesis and not upon whether its form is AR or MA. It is also shown that test procedures break down when the ARMA model is employed as the alternative hypothesis. Section 5 contains some concluding remarks.

2. THE MODEL AND SOME PRELIMINARY RESULTS

Suppose that

(1) $$\underline{y} = X\underline{\beta} + \underline{u}$$

where $\underline{y}$ is a T by 1 vector of observations on the dependent variable, X is a T by k matrix of observations on the regressors which include lagged dependent variables, $\underline{\beta}$ is a k by 1 vector of unknown coefficients, and $\underline{u}$ is a T by 1 vector of error terms. The null hypothesis is that the $\{u_t\}$ are independently and normally distributed with zero mean and variance σ^2, and the two alternatives considered in Section 3 are H_A: the $\{u_t\}$ are generated by the AR(n) process

(2) $$u_t + \alpha_1 u_{t-1} + \ldots + \alpha_n u_{t-n} = \varepsilon_t, \quad \varepsilon_t \text{NID}(0, \sigma_\varepsilon^2),$$

and H_B: the $\{u_t\}$ are generated by the MA(n) process

(3) $$u_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_n \varepsilon_{t-n}, \quad \varepsilon_t \text{NID}(0, \sigma_\varepsilon^2).$$

It is assumed that the models consisting of (1) and (2), and (1) and (3) both satisfy the conditions set out by Durbin [**1**, pp. 412–413] with the α coefficients of (3) satisfying the invertibility condition.

The error models of (2) and (3) can be compactly written in the form

(4) $$M(\underline{\alpha})\underline{u} = \underline{\varepsilon} \ldots \text{AR}(n)$$

and

(5) $$\underline{u} = M(\underline{\alpha})\underline{\varepsilon} \ldots \text{MA}(n),$$

where $\underline{\alpha}' = (\alpha_1 \ldots \alpha_n)$, $\underline{u}' = (u_1 \ldots u_T)$, $\underline{\varepsilon}' = (\varepsilon_1 \ldots \varepsilon_T)$, and

$$M(\underline{\alpha}) = \begin{bmatrix} 1 & & & & & 0 \\ \alpha_1 & \ddots & & & & \\ \vdots & \ddots & \ddots & & & \\ \alpha_n & & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & \alpha_n & \cdots & \alpha_1 & 1 \end{bmatrix} \quad \text{(see Pagan [**8**]).}$$

Note that, for simplicity, the pre-sample values $(u_0, \ldots, u_{1-n})$ and $(\varepsilon_0, \ldots, \varepsilon_{1-n})$ have been set equal to zero and not treated as random variables. This assumption does not affect the asymptotic properties of the tests developed below.

Estimators of $\underline{\alpha}$, $\underline{\beta}$, and σ_ε^2 which are asymptotically equivalent to the exact MLE can then be obtained by maximizing

$$(6) \qquad L_1(\underline{\alpha}, \underline{\beta}, \sigma_\varepsilon^2) = -\tfrac{1}{2} \ln \sigma_\varepsilon^2 - T^{-1}(\underline{\varepsilon}'\underline{\varepsilon}/2\sigma_\varepsilon^2)$$

where $\underline{\varepsilon} = M(\underline{\alpha})(\underline{y} - X\underline{\beta})$ for the AR(n) case and $\underline{\varepsilon} = M(\underline{\alpha})^{-1}(\underline{y} - X\underline{\beta})$ for the MA(n) case.[3] The parameter σ_ε^2 can be eliminated by concentrating the log likelihood function $L_1(\cdot)$ to obtain

$$(7) \qquad L_2(\underline{\alpha}, \underline{\beta}) = -\tfrac{1}{2} \ln (\underline{\varepsilon}'\underline{\varepsilon}/T),$$

where $\underline{\varepsilon} = M(\underline{\alpha})(\underline{y} - X\underline{\beta})$ for the AR(n) case and $\underline{\varepsilon} = M(\underline{\alpha})^{-1}(\underline{y} - X\underline{\beta})$ for the MA(n) case. The dimensionality of the parameter space can be further reduced by replacing $\underline{\beta}$ in $L_2(\cdot)$ by

$$(8) \qquad \underline{\beta} = (X'S(\underline{\alpha})^{-1}X)^{-1}(X'S(\underline{\alpha})^{-1}\underline{y}),$$

where $S(\underline{\alpha}) = M(\underline{\alpha})^{-1}(M(\underline{\alpha})^{-1})'$ for the AR(n) case and $S(\underline{\alpha}) = M(\underline{\alpha})M(\underline{\alpha})'$ for the MA(n) case, in order to obtain a concentrated log likelihood function $L_3(\underline{\alpha})$.

The null hypothesis of serial independence will be tested by testing $H_0: \underline{\alpha} = \underline{0}$ and, in order to derive a suitable test statistic based only upon results from estimation under H_0, we adopt Silvey's [**10**] Lagrange multiplier approach. The Lagrangian will be written as

$$(9) \qquad L_2^*(\underline{\alpha}, \underline{\beta}, \underline{\lambda}) = L_2(\underline{\alpha}, \underline{\beta}) - \underline{\lambda}'\underline{\alpha}$$

where $\underline{\lambda}' = (\lambda_1 \ldots \lambda_n)$ is the vector of multipliers. It will be useful to denote the maximizers of $L_2^*(\cdot)$ by caret and the maximizers of $L_2(\cdot)$ by tilde, and to introduce the vector $\underline{\theta}' = (\underline{\alpha}' \; \underline{\beta}')$.

The first order conditions for a maximum of $L_2^*(\cdot)$ with respect to $\underline{\lambda}$ imply that $\underline{\hat{\alpha}} = \underline{0}$, so that $M(\underline{\hat{\alpha}}) = I$, $\hat{\beta} = (X'X)^{-1}X'y$ (the OLS estimate), and $\underline{\hat{\varepsilon}} = \underline{\hat{u}} = y - X\hat{\beta}$ (the OLS residual vector). The vector $\underline{\hat{\lambda}}$ is given by

$$(10) \qquad \underline{\hat{\lambda}} = (\partial L_2(\underline{\theta})/\partial \underline{\alpha})_{\underline{\theta} = \underline{\hat{\theta}}}$$

and Silvey's approach to testing H_0 is to test the joint significance of the elements of $\hat{\lambda}$. Now, the results obtained by Silvey [**10**] imply that, under H_0, $\sqrt{T}\underline{\hat{\lambda}} \to N(\underline{0}, \Psi(\underline{\alpha}))$ where

$$\Psi(\underline{\alpha}) = \text{plim}\, (\partial^2 L_3(\underline{\alpha})/\partial \underline{\alpha} \partial \underline{\alpha}')_{\underline{\alpha} = \underline{\hat{\alpha}}}.$$

The matrix $\Psi(\underline{\alpha})$ is, however, simply the inverse of the asymptotic variance-covariance matrix of $\sqrt{T}(\underline{\tilde{\alpha}} - \underline{\alpha})$ when H_0 is true, so that if W is a consistent estimator of this variance-covariance matrix, then a valid large sample test can be based upon

$$(11) \qquad l = (\sqrt{T}\underline{\hat{\lambda}})' W (\sqrt{T}\underline{\hat{\lambda}})$$

[3] The constant terms of log likelihood functions will be omitted.

which is asymptotically distributed as χ_n^2 under H_0, with significantly large values of l indicating that the null hypothesis is not consistent with the sample data.

The l test is asymptotically equivalent to the LR test, but requires only estimation under H_0. In the context of testing for serial correlation, this implies that the Lagrange multiplier approach provides simple non-iterative tests based upon OLS results which have good asymptotic power properties. The exact forms of these tests for the alternatives H_A and H_B are obtained in the next section, and are compared to those derived from Durbin's [**1**] general method.

3. LAGRANGE MULTIPLIER TESTS AGAINST AR(n) AND MA(n) ALTERNATIVES

Consider first the alternative hypothesis H_A that the $\{u_t\}$ are generated by an AR(n) process, so that

$$
\begin{aligned}
(12) \qquad \hat{\underline{\lambda}} &= (\partial L_2(\underline{\theta})/\partial \underline{\alpha})_{\underline{\theta}=\hat{\underline{\theta}}} \\
&= -\tfrac{1}{2}[(\underline{\varepsilon}'\underline{\varepsilon}/T)^{-1}T^{-1}\partial(\underline{\varepsilon}'\underline{\varepsilon})/\partial\underline{\alpha}]_{\underline{\theta}=\hat{\underline{\theta}}} \\
&= -[(\underline{\varepsilon}'\underline{\varepsilon}/T)^{-1}T^{-1}(\partial\underline{\varepsilon}/\partial\underline{\alpha})'\underline{\varepsilon}]_{\underline{\theta}=\hat{\underline{\theta}}},
\end{aligned}
$$

with $M(\underline{\alpha})\underline{u}=\underline{\varepsilon}$, so that

$$
\begin{aligned}
\partial\underline{\varepsilon}/\partial\alpha_1 &= (\partial M(\underline{\alpha})/\partial\alpha_1)\underline{u} \\
&= L\underline{u},
\end{aligned}
$$

where L is a T by T matrix with elements on the diagonal immediately below the main diagonal being equal to unity and all other elements equal to zero. (Fitts [**2**, p. 365] points out that L can be thought of as the lag operator.) In general,

$$
(13) \qquad \partial\underline{\varepsilon}/\partial\alpha_i = L^i\underline{u} = \underline{u}_i \qquad (i=1,\ldots,n)
$$

where $\underline{u}_i'=(0\ldots 0\; u_1 \ldots u_{T-i})$. Combining (12) and (13), it is easy to see that

$$
\begin{aligned}
\hat{\lambda}_i &= -(\hat{\underline{u}}'\hat{\underline{u}}/T)^{-1}T^{-1}\hat{\underline{u}}_i'\hat{\underline{u}} \\
(14) \qquad &= -(\hat{\underline{u}}_i'\hat{\underline{u}})/(\hat{\underline{u}}'\hat{\underline{u}}) \\
&= -r_i \qquad (i=1,\ldots,n),
\end{aligned}
$$

where $\hat{\underline{u}}_i'=(0\ldots 0\;\hat{u}_1\ldots\hat{u}_{T-i})$ and r_i is an estimate of the ith autocorrelation coefficient of the $\{u_t\}$ calculated from the OLS residuals. It follows that

$$
(15) \qquad \hat{\underline{\lambda}} = -(T\hat{\sigma}^2)^{-1}\hat{U}_n'\hat{\underline{u}}
$$

where $\hat{\sigma}^2=(\hat{\underline{u}}'\hat{\underline{u}}/T)$ and $\hat{U}_n=(\hat{\underline{u}}_1\ldots\hat{\underline{u}}_n)$.

Sargan's [**9**] results imply that, under H_0, a consistent estimator of the asymptotic variance-covariance matrix of $\sqrt{T}(\tilde{\underline{\alpha}}-\underline{\alpha})$ is $W = T\hat{\sigma}^2[\hat{U}_n'\hat{U}_n-\hat{U}_n'X(X'X)^{-1}X'\hat{U}_n]^{-1}$.[4] The statistic for the Lagrange multiplier

[4] The autoregressive least squares estimator can be regarded as a special case of the nonlinear instrumental variable estimator with the instruments being the regressors of the unrestricted transformed equation, and the results of Sargan [**9**, Section 5] are, therefore, applicable.

test of H_0 against H_A is, therefore,

$$\begin{aligned} l_{AR} &= T^{-1}\hat{\sigma}^{-4}(\underline{\hat{u}}'\hat{U}_n)W(\hat{U}_n'\underline{\hat{u}}) \\ &= \underline{\hat{u}}'\hat{U}_n[\hat{U}_n'\hat{U}_n - \hat{U}_n'X(X'X)^{-1}X'\hat{U}_n]^{-1}\hat{U}_n'\underline{\hat{u}}/\hat{\sigma}^2. \end{aligned} \tag{16}$$

Note that $\hat{\sigma}^2$ can be replaced in (16) by the asymptotically equivalent estimator $s^2 = (\underline{\hat{u}}'\underline{\hat{u}})/(T-k)$ without affecting the asymptotic properties of the l_{AR} test, so that an asymptotically equivalent test statistic is

$$l_{AR}^* = \underline{\hat{u}}'\hat{U}_n[s^4T(I_n) - \hat{U}_n'XVX'\hat{U}_n]^{-1}\hat{U}_n'\underline{\hat{u}} \tag{17}$$

where V is the usual estimator of the asymptotic variance-covariance matrix of the least squares estimator, i.e. $V = s^2(X'X)^{-1}$, and we have used the result that plim $T^{-1}(\hat{U}_n'\hat{U}_n) = \sigma^2 I_n$ when H_0 is true.

Consider now the case in which the alternative hypothesis is that the errors are generated by the MA(n) scheme of equation (3). The vector $\hat{\underline{\lambda}}$ is still given by (12), but $\underline{u} = M(\underline{\alpha})\underline{\varepsilon}$, so that

$$\begin{aligned} \partial\underline{\varepsilon}/\partial\alpha_i &= -M(\underline{\alpha})^{-1}(\partial M(\underline{\alpha})/\partial\alpha_i)M(\underline{\alpha})^{-1}\underline{u} \\ &= -M(\underline{\alpha})^{-1}L^iM(\underline{\alpha})^{-1}\underline{u}. \end{aligned} \tag{18}$$

Combining (12) and (18) yields

$$\begin{aligned} \hat{\lambda}_i &= (\underline{\hat{u}}'\underline{\hat{u}}/T)^{-1}T^{-1}(\underline{\hat{u}}_i'\underline{\hat{u}}) \\ &= r_i \qquad\qquad (i = 1, \ldots, n). \end{aligned} \tag{19}$$

It can be seen from (14) and (19) that the only effect of using the MA alternative H_B, rather than the AR alternative H_A, is to change the sign of every element of $\hat{\underline{\lambda}}$. The multipliers, therefore, have the same asymptotic variance-covariance matrix under H_0, and the test statistic for the MA alternative is exactly the same as the test statistic for the AR alternative since $(-\sqrt{T}\hat{\underline{\lambda}})'W(-\sqrt{T}\hat{\underline{\lambda}}) = (\sqrt{T}\hat{\underline{\lambda}})'W(\sqrt{T}\hat{\underline{\lambda}})$.

This equivalence does not hold for the tests derived from Durbin's [**1**] procedure. In the case of testing against AR(n) errors, Durbin's test would be based upon the value of $\underline{\alpha}$ maximizing $L_2(\underline{\alpha}, \hat{\underline{\beta}})$, i.e. upon $\underline{\alpha}^* = (\hat{U}_n'\hat{U}_n)^{-1}\hat{U}_n'\underline{\hat{u}}$. But $\underline{\alpha}^*$ is obviously a nonsingular transformation of $\hat{\underline{\lambda}}$ of (15) and so the two approaches lead to test statistics which differ only in the treatment of asymptotically negligible terms. When the alternative is H_B, i.e. the MA(n) case, then Durbin's test would again be based upon the value of $\underline{\alpha}$ maximizing $L_2(\underline{\alpha}, \hat{\underline{\beta}})$ which is, of course, the minimizer of $\underline{\hat{u}}'(M(\underline{\alpha})^{-1})'M(\underline{\alpha})^{-1}\underline{\hat{u}}$. The minimization of $\underline{\hat{u}}'(M(\underline{\alpha})^{-1})'M(\underline{\alpha})^{-1}\underline{\hat{u}}$ is, however, a nonlinear optimization problem and its solution involves iterative calculations (see Fitts [**2**, Section 4] for details of the generalization of Durbin's test for a MA(1) alternative). The test based upon Durbin's approach would be asymptotically equivalent to the Lagrange multiplier test since both are asymptotically equivalent to the LR test, but its computational cost is unlikely to make it attractive to applied workers.

The fact that the Lagrange multiplier test statistic does not depend upon whether the alternative is H_A or H_B does raise the question of how to proceed if a significant value of the test statistic is obtained. Assuming that re-estimation, rather than re-specification, of (1) is the right course of action, the researcher will have to choose between the error models (2) and (3), and the consequences of an incorrect choice are obviously of interest. Some evidence on this matter is available in the Monte Carlo work of Hendry and Trivedi [**6**, Part 2] who found that, if the bias and mean square error of the estimator of β were used as criteria, then the quality of an estimated AR(1) or MA(1) approximation did not depend upon whether the true model was AR(1) or MA(1), thus implying that the correct order of the error model was more important than its form. Analytic support for this result is to be found in a recent paper by Hendry [**5**] who also shows that, in terms of goodness of fit and final residual autocorrelations, there is little to choose between an ARMA(2, 0) approximation and an ARMA(1, 1) scheme when the latter model is appropriate. If the results obtained by Hendry and Trivedi [**6**], and by Hendry [**5**] can be applied in more general situations, then presumably researchers will employ the AR(n) model if serial correlation is detected since it implies a smaller computational burden than the MA(n) error model.[5]

To sum up, the test procedure of this paper for testing H_0 against either H_A or H_B is as follows: (a) estimate the regression equation (1) by OLS and obtain the OLS residuals $\{\hat{u}_t\}$; (b) construct $\hat{U}_n^* = (\underline{\hat{u}}_1 \ldots \underline{\hat{u}}_n)$, calculate

$$l = \underline{\hat{u}}'\hat{U}_n[\hat{U}_n'\hat{U}_n - \hat{U}_n'X(X'X)^{-1}X'\hat{U}_n]^{-1}\hat{U}_n'\underline{\hat{u}}/\hat{\sigma}^2,$$

and compare l to the chosen critical value of a χ_n^2 variate. Significantly large values of l imply that the assumption that the errors are serially independent is not consistent with the sample data.[6]

The analysis above is easily generalized to the case in which the regressors include endogenous variables and the equation is estimated by the method of instrumental variables. If X includes only exogenous and lagged exogenous variables, then, when H_0 is true, the right-hand side of (16) is asymptotically equivalent to $\underline{\hat{u}}'\hat{U}_n(\hat{U}_n'\hat{U}_n)^{-1}\hat{U}_n'\underline{\hat{u}}/\hat{\sigma}^2$ which is in turn asymptotically equivalent to $T(r_1^2 + \ldots + r_n^2)$.

4. TESTING AGAINST ARMA ALTERNATIVES

The alternative hypothesis considered in this section is H_c: the $\{u_t\}$ are generated by the ARMA(p, q) model

$$(20) \qquad u_t + \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} = \varepsilon_t + \mu_1\varepsilon_{t-1} + \ldots + \mu_q\varepsilon_{t-q}, \quad \varepsilon_t \mathrm{NID}(0, \sigma_\varepsilon^2).$$

Assuming that the pre-sample values $(u_0, \ldots, u_{1-p}: \varepsilon_0, \ldots, \varepsilon_{1-q})$ are equal to

[5] The results of [**5** and **6**] are, of course, somewhat stronger than might be expected from the well known result that a stationary and invertible ARMA process can be approximated arbitrarily closely by an AR process of sufficiently high order.

[6] Subroutines to compute l and the appropriate generalization of Godfrey's [**3**] π test will be available from the author on request.

zero, the error model under H_c can be written as

$$M(\underline{\rho})\underline{u} = N(\underline{\mu})\underline{\varepsilon} \tag{21}$$

where $\underline{\rho}' = (\rho_1 \dots \rho_p)$, $\underline{\mu}' = (\mu_1 \dots \mu_q)$,

$$M(\underline{\rho}) = \begin{bmatrix} 1 & & & & & 0 \\ \rho_1 & \ddots & & & & \\ \vdots & \ddots & \ddots & & & \\ \rho_p & & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & \rho_p & \dots & \rho_1 & 1 \end{bmatrix},$$

and

$$N(\underline{\mu}) = \begin{bmatrix} 1 & & & & & 0 \\ \mu_1 & \ddots & & & & \\ \vdots & \ddots & \ddots & & & \\ \mu_q & & \ddots & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & \mu_q & \dots & \mu_1 & 1 \end{bmatrix}.$$

If $\underline{\alpha}$ is defined by $\underline{\alpha}' = (\underline{\rho}' \ \underline{\mu}')$, then the concentrated log likelihood function is given by

$$L_4(\underline{\alpha}, \underline{\beta}) = -\tfrac{1}{2} \ln (\underline{\varepsilon}'\underline{\varepsilon}/T) \tag{22}$$

where $\underline{\varepsilon} = N(\underline{\mu})^{-1} M(\underline{\rho})(\underline{y} - X\underline{\beta})$. The Lagrangian used in testing H_0: $\underline{\alpha} = \underline{0}$ is then

$$L_4^*(\underline{\alpha}, \underline{\beta}, \underline{\lambda}) = L_4(\underline{\alpha}, \underline{\beta}) - \underline{\lambda}'\underline{\alpha} \tag{23}$$

where $\underline{\lambda}$ is a $(p+q)$ by 1 vector of undetermined multipliers. If caret is used to denote a maximizer of $L_4^*(\cdot)$ and $\underline{\theta}' = (\underline{\alpha}' \ \underline{\beta}')$, then, as before, $\hat{\underline{\alpha}} = \underline{0}$, so that $M(\hat{\underline{\rho}}) = N(\hat{\underline{\mu}}) = I$, $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$ (the OLS estimate) and $\hat{\underline{\varepsilon}} = \hat{\underline{u}} = \underline{y} - X\hat{\underline{\beta}}$ (the OLS residual vector). The estimated multipliers are given by

$$\hat{\underline{\lambda}} = (\partial L_4(\underline{\theta})/\partial \underline{\alpha})_{\underline{\theta} = \hat{\underline{\theta}}}$$

and it is easy to show that

$$\hat{\underline{\lambda}}' = (-r_1 \ -r_2 \dots \ -r_p : r_1 \ r_2 \dots r_q) \tag{24}$$

where the $\{r_i\}$ are as defined in the previous section. The vector $\hat{\underline{\lambda}}$ obviously has a singular multivariate normal distribution asymptotically and the test procedure described above cannot be applied.[7]

[7] If the results obtained by Hendry [**5**, Section 3] on the adequacy of an AR(2) approximation to a true ARMA(1, 1) model generalize to the ARMA(p, q) case (and this matter clearly merits investigation), then it may be useful to test H_0 against H_A with $n = p + q$ if there is reason to believe that H_C should be employed as the alternative hypothesis.

This phenomenon is, of course, not peculiar to the Lagrange multiplier approach. Fitts [**2**, p. 370] points out that, in the case of the ARMA(1, 1) model, the MLE of μ_1 and ρ_1 are perfectly correlated when both parameters equal zero with the likelihood function being maximized anywhere along the line $\mu_1 = \rho_1$. This perfect correlation between the MLE of the parameters of the error model implies the singularity of their asymptotic variance–covariance matrix and this singularity also occurs when H_0 is true and the ARMA(p, q) model is used as the alternative. It follows that the usual asymptotic tests based upon the unrestricted MLE are not applicable for testing H_0 against H_C. Similar results in the time series analysis literature are to be found in Hannan's [**4**] book in which he shows that the MLE of the parameters of an ARMA(p, q) process have a singular asymptotic variance-covariance matrix if $\rho_p = \mu_q = 0$, and points out that this implies that it is not possible to carry out a valid test if the alternative is H_C and the null hypothesis is obtained by decreasing both p and q (see [**4**, pp. 388–389 and pp. 409–414].

5. CONCLUDING REMARKS

This paper has explored the application of Silvey's [**10**] Lagrange multiplier approach to the problem of testing for serial correlation in the errors of dynamic regression equations. It has been shown that the tests against the alternatives of AR(n) and MA(n) errors are identical, both being simply tests of the joint significance of the first n autocorrelations of the OLS residuals. (It might be thought unusual that the partial autocorrelations play no role in the analysis, but, under H_0, the first n sample autocorrelations will be asymptotically equivalent to the corresponding estimated partial autocorrelations.)

It is clear that if one is willing to estimate under the alternative hypothesis using some nonlinear least squares estimator, then the likelihood ratio and Wald approaches can be used to obtain tests which are asymptotically equivalent to the Lagrange multiplier test. In practice, however, many researchers prefer to estimate under the null hypothesis using the OLS estimator, and to consider more complicated estimators only if the OLS residuals provide significant evidence of serial correlation. It could be argued that this approach is short sighted since if the null hypothesis is rejected then the model must be reestimated to obtain the unrestricted maximum likelihood estimates which are calculated for the likelihood ratio and Wald tests. It is, however, not always the case that the appropriate treatment of serial correlation is the reestimation of the original model and it is reasonable to suppose that serial correlation will sometimes be caused by misspecification of the regression model, e.g. misspecified dynamic structure, incorrect functional form. It would, of course, be of interest to examine the behavior of the Lagrange multiplier, likelihood ratio, and Wald tests in small samples, e.g. by Monte Carlo methods. (A referee has suggested that the finite sample performance of the Lagrange multiplier test might be improved by omitting asymptotically negligible terms, e.g., covariances between the lagged values of disturbances and exogenous regressors.)

University of York

Manuscript received May, 1977.

REFERENCES

[1] DURBIN, J.: "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38 (1970), 410–421.

[2] FITTS, J.: "Testing for Autocorrelation in the Autoregressive Moving Average Error Model," *Journal of Econometrics*, 1 (1973), 363–376.

[3] GODFREY, L. G: "Testing for Serial Correlation in Dynamic Simultaneous Equation Models," *Econometrica*, 44 (1976), 1077–1084.

[4] HANNAN, E. J.: *Multiple Time Series.* New York: Wiley, 1970.

[5] HENDRY, D. F.: "On the Time Series Approach to Econometric Model Building," Discussion Paper No. A8, London School of Economics, 1976.

[6] HENDRY, D. F., AND P. K. TRIVEDI: "Maximum Likelihood Estimation of Difference Equations with Moving Average Errors: A Simultation Study," *The Review of Economic Studies*, 39 (1972), 117–145.

[7] NICHOLLS, D. F., A. R. PAGAN, AND R. D. TERREL: "The Estimation and Use of Models with Moving Average Disturbance Terms: A Survey," *International Economic Review*, 16 (1975), 113–134.

[8] PAGAN, A. R.: "A Generalized Approach to the Treatment of Autocorrelation," *Australian Economic Papers*, 13 (1974), 267–280.

[9] SARGAN, J. D.: "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society*, Series B, 21 (1959), 91–105.

[10] SILVEY, S. D.: "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30 (1959), 389–407.

[11] SIMS, C. A.: "Distributed Lags," in *Frontiers of Quantitive Economics*, Vol. II, ed. by M. D. Intriligator and D. A. Kendrick. Amsterdam: North-Holland, 1974, pp. 289–332.

[12] WALLIS, K. F.: "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations," *Econometrica*, 40 (1972), 617–636.