The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics

Author(s): T. S. Breusch and A. R. Pagan

Source: *The Review of Economic Studies*, Jan., 1980, Vol. 47, No. 1, Econometrics Issue (Jan., 1980), pp. 239-253

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/2297111

Review of Economic Studies (1980) XLVII, 239-253
0034-6527/80/00110239$02.00

The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics

T. S. BREUSCH
University of Southampton

and

A. R. PAGAN
Australian National University

1. INTRODUCTION

Many econometric models are susceptible to analysis only by asymptotic techniques and there are three principles, based on asymptotic theory, for the construction of tests of parametric hypotheses. These are: (i) the Wald (W) test which relies on the asymptotic normality of parameter estimators, (ii) the maximum likelihood ratio (LR) procedure and (iii) the Lagrange multiplier (LM) method which tests the effect on the first order conditions for a maximum of the likelihood of imposing the hypothesis. In the econometric literature, most attention seems to have been centred on the first two principles. Familiar "t-tests" usually rely on the W principle for their validity while there have been a number of papers advocating and illustrating the use of the LR procedure. However, all three are equivalent in well-behaved problems in the sense that they give statistics with the same asymptotic distribution when the null hypothesis is true and have the same asymptotic power characteristics. Choice of any one principle must therefore be made by reference to other criteria such as small sample properties or computational convenience.

In many situations the W test is attractive for this latter reason because it is constructed from the unrestricted estimates of the parameters and their estimated covariance matrix. The LM test is based on estimation with the hypothesis imposed as parametric restrictions so it seems reasonable that a choice between W or LM be based on the relative ease of estimation under the null and alternative hypotheses. Whenever it is easier to estimate the restricted model, the LM test will generally be more useful. It then provides applied researchers with a simple technique for assessing the adequacy of their particular specification.

This paper has two aims. The first is to exposit the various forms of the LM statistic and to collect together some of the relevant research reported in the mathematical statistics literature. The second is to illustrate the construction of LM tests by considering a number of particular econometric specifications as examples. It will be found that in many instances the LM statistic can be computed by a regression using the residuals of the fitted model which, because of its simplicity, is itself estimated by OLS.

The paper contains five sections. In Section 2, the LM statistic is outlined and some alternative versions of it are discussed. Section 3 gives the derivation of the statistic for several econometric specifications. Applications in this section are the testing for a liquidity trap, autocorrelation, the error components model, diagonality of a covariance matrix in seemingly unrelated equation systems and choice between models generated by separate families of distributions. Section 4 considers the construction of a pseudo-LM statistic when estimation is difficult even under the null hypothesis and discusses the derivation of the exact distribution of the statistic. A concluding summary is given as Section 5.

2. THE FORMS OF THE LM STATISTIC

Consider a sample of size N from some distribution which is known apart from a finite number K of unknown parameters $\theta = (\theta_1, \ldots, \theta_K)'$ giving a log likelihood $L(\theta)$. The hypothesis to be tested is specified as $p < K$ restrictions on θ,

$$H_0\colon h_j(\theta) = 0 \qquad j = 1, \ldots, p. \qquad \ldots(1)$$

Aitchison and Silvey (1958), (1960) approach this problem by setting up the Lagrangean function

$$L(\theta) + \textstyle\sum_{j=1}^{p} \lambda_j h_j(\theta) \qquad \ldots(2)$$

and differentiating this with respect to the unknown parameters θ and the Lagrange multipliers λ_j to yield $\tilde{\theta}$ and $\tilde{\lambda}$ as the solution of the first order conditions

$$\tilde{D} + \tilde{H}\tilde{\lambda} = 0 \qquad \ldots(3)$$

$$h_j(\tilde{\theta}) = 0 \qquad j = 1, \ldots, p \qquad \ldots(4)$$

where

$$\tilde{D} \text{ is the } (K \times 1) \text{ vector } \left\{\frac{\partial L}{\partial \theta_i}(\tilde{\theta})\right\},$$

$$\tilde{H} \text{ is the } (K \times p) \text{ matrix } \left\{\frac{\partial h_j}{\partial \theta_i}(\tilde{\theta})\right\},$$

and

$$\tilde{\lambda} \text{ is a } (p \times 1) \text{ vector of Lagrange multipliers.}$$

The idea underlying the test is that when the null hypothesis is correct the restricted estimate $\tilde{\theta}$ will tend to be near the unrestricted maximum likelihood estimate so that $\tilde{D}$ will be close to the zero vector. Under the regularity condition that the order of differentiation and integration can be interchanged, Feigin (1976) shows that D will be a zero mean martingale, and under conditions set out in Crowder (1976) $c_N^{-\frac{1}{2}} D \xrightarrow{\mathscr{D}} \mathcal{N}(0, \lim_{N\to\infty} c_N^{-\frac{1}{2}} \mathscr{I} c_N^{-\frac{1}{2}})$ where $\mathscr{I} = E(-\partial^2 L/\partial\theta\partial\theta')$ is Fisher's information matrix and c_N is some suitably chosen norming matrix. This leads to the LM test based upon the statistic

$$\mathrm{LM} = \tilde{D}'\tilde{\mathscr{I}}^{-1}\tilde{D} = \tilde{\lambda}'\tilde{H}'\tilde{\mathscr{I}}^{-1}\tilde{H}\tilde{\lambda} \qquad \ldots(5)$$

where $\tilde{\mathscr{I}}$ is the information matrix when the null hypothesis is true, evaluated at the restricted estimates $\tilde{\theta}$. The term $\tilde{D}'\tilde{\mathscr{I}}^{-1}\tilde{D}$ is the "score" statistic (see Rao (1973)) while $\tilde{\lambda}'\tilde{H}'\tilde{\mathscr{I}}^{-1}\tilde{H}\tilde{\lambda}$ is the Lagrangean multiplier statistic (see Byron (1970)), making it clear that the two test statistics are identical so that the choice of which form to use is based on convenience. Throughout the remainder of this paper it is the score test form that we employ, although we will continue to refer to it as the LM test.

Under the usual maximum likelihood regularity conditions, the LM statistic in (5) is asymptotically equivalent to the W and LR statistics. That is, when H_0 is true it is asymptotically distributed as $\chi^2(p)$ and the test which rejects H_0 when the statistic is greater than the appropriate upper point of the $\chi^2(p)$ distribution has the same asymptotic power characteristics as the other tests, sharing the optimality criterion of maximum local power (Silvey (1959) and Cox and Hinkley (1974)).

For later reference it will be useful to establish the form of the LM statistic in the special case when θ is partitioned into two subsets θ_1 and θ_2 and the restriction under test is that one of the subsets of parameters equals particular values i.e. $\theta' = (\theta_1'\theta_2')$ and

$$H_0: \theta_1 = \theta_{10}$$

or

$$H_0: h(\theta) = [I_p : 0]\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} - \theta_{10} = 0$$

Partitioning D and $\mathscr{I}$ conformably,

$$\tilde{D} = \begin{bmatrix}\tilde{D}_1\\ \tilde{D}_2\end{bmatrix} = \begin{bmatrix}\tilde{D}_1\\ 0\end{bmatrix}$$

and

$$\mathscr{I} = \begin{bmatrix}\mathscr{I}_{11} & \mathscr{I}_{12}\\ \mathscr{I}_{21} & \mathscr{I}_{22}\end{bmatrix}$$

where $\tilde{D}_2 = 0$ from the first order conditions (3). In this case, the LM statistic will be

$$\tilde{D}_1'(\tilde{\mathscr{I}}_{11} - \tilde{\mathscr{I}}_{12}\tilde{\mathscr{I}}_{22}^{-1}\tilde{\mathscr{I}}_{21})^{-1}\tilde{D}_1 \equiv \tilde{D}_1'\tilde{\mathscr{I}}^{11}\tilde{D}_1. \qquad \ldots(6)$$

Furthermore, if $\tilde{\mathscr{I}}_{12} = \tilde{\mathscr{I}}_{21} = 0$ so that $\tilde{\mathscr{I}}$ is block diagonal, expression (6) reduces to

$$\tilde{D}_1'\tilde{\mathscr{I}}_{11}^{-1}\tilde{D}_1. \qquad \ldots(7)$$

Either (6) or (7) is generally needed in any particular application and both will be used in later sections.

It is worth noting that, whenever the regularity condition cited earlier holds, the information matrix can be obtained from the first derivatives of the log likelihood using the standard result from likelihood theory:

$$\mathscr{I} = E\left[-\frac{\partial^2 L}{\partial\theta\partial\theta'}\right] = E\left[\left(\frac{\partial L}{\partial\theta}\right)\left(\frac{\partial L}{\partial\theta}\right)'\right]$$

Alternative formulations of the statistic can also be obtained using other matrices in place of $\mathscr{I}$. In many instances, see Crowder (1976),

$$\text{plim}\left\{\tilde{\mathscr{I}}^{-1}\left[-\frac{\partial^2 L}{\partial\theta\partial\theta'}(\tilde{\theta})\right]\right\} = I_K$$

so that the asymptotic properties of the statistic will be unaffected if the negative of the Hessian (matrix of second derivatives) or its limiting form are used instead of the information matrix in constructing the test statistic.

Although it will emerge later that application of the formula in (6) or (7) frequently results in the LM statistic being computed by a regression, there are instances in which the analytic derivation of the required quantities would be so difficult as not to be worthwhile. For example, if one wished to construct hypothesis tests against the Goldfeld–Quandt (1973) switching regression model or the Rosenberg (1973) stochastically varying coefficient regression model, it is very difficult to construct the information matrix. In such cases, it is useful to derive an indirect method of computing the LM statistic which is probably available to most researchers estimating these models. The indirect method employs the fact that the LM test may be constructed by applying the Wald formula to the estimates from the first iteration of Fisher's scoring algorithm, when the starting values employed are the maximum likelihood values under the null hypothesis. To see this, observe that the method of scoring, starting from $\theta = \tilde{\theta}$, would yield as the first estimate of θ:

$$\hat{\theta} = \tilde{\theta} + \tilde{\mathscr{I}}^{-1}\tilde{D}.$$

Alternatively

$$\tilde{D} = \tilde{\mathscr{I}}(\hat{\theta} - \tilde{\theta})$$

so that

$$\begin{aligned} \text{LM} &= (\hat{\theta} - \tilde{\theta})'\tilde{\mathscr{I}}\tilde{\mathscr{I}}^{-1}\tilde{\mathscr{I}}(\hat{\theta} - \tilde{\theta}) \\ &= (\hat{\theta} - \tilde{\theta})'\tilde{\mathscr{I}}(\hat{\theta} - \tilde{\theta}). \end{aligned}$$

But this quantity is what would be obtained if a "Wald test" for the "hypothesis" $\theta = \tilde{\theta}$ was performed at the end of the first iteration of the scoring algorithm. Furthermore, under the null hypothesis

$$c_N^{-\frac{1}{2}}(\tilde{\mathscr{I}} - \hat{\mathscr{I}})c_N^{-\frac{1}{2}} \xrightarrow{p} 0$$

so that the statistic $(\hat{\theta} - \tilde{\theta})'\hat{\mathscr{I}}(\hat{\theta} - \tilde{\theta})$ will have the same asymptotic distribution as LM (under the null) and this would represent a "Wald test" of the "hypothesis" $\theta = \hat{\theta}$. Even if scoring has not been implemented for the problem, it will generally be possible to use the negative of the Hessian in place of $\tilde{\mathscr{I}}$, allowing the LM test to be based upon estimates after the first iteration of the Newton–Raphson algorithm with analytic or numerical derivatives.[1] Hence, if the direct approach is difficult to apply, this indirect approach enables the construction of an asymptotically valid test statistic to determine whether estimation under the alternative (i.e. continued iteration) is likely to be worthwhile.

3. EXAMPLES OF THE LM TEST

There is a class of problems for which the functional form may be written as a non-linear regression,

$$y_t = g(x_t;\, \theta) + e_t \equiv g_t + e_t \qquad t = 1, \ldots, N \qquad \ldots(8)$$

where the e_t are identically and independently distributed as $\mathcal{N}(0, \sigma^2)$ and g_t is independent of e_t. The K-dimension parameter set θ is partitioned into θ_1 and θ_2 of p and $K - p$ parameters respectively and H_0: $\theta_1 = \theta_{10}$ for some specific value θ_{10}. Two members of this class will be considered below.

Under the above assumptions the log likelihood (omitting constants) would be $-\frac{1}{2}N \log \sigma^2 - \frac{1}{2}\sigma^{-2}e'e$ where $e' = (e_1, \ldots, e_N)$.[2] It is easily seen that the information matrix will be block diagonal between θ and σ^2, therefore only derivatives with respect to θ are required allowing attention to be concentrated upon the exponent $-\frac{1}{2}\sigma^{-2}e'e$. The quantities needed to evaluate (5) will be

$$D = \frac{\partial L}{\partial \theta} = \sigma^{-2}G'e$$

$$\mathscr{I} = E\left[\left(\frac{\partial L}{\partial \theta}\right)\left(\frac{\partial L}{\partial \theta}\right)'\right] = \sigma^{-2}E(G'G)$$

where G is the $(N \times K)$ matrix containing $\partial g_t/\partial \theta_k$ in the (t, k)th position. Hence, the LM statistic is based upon the quantity

$$\tilde{\sigma}^{-2}\tilde{e}'\tilde{G}(\tilde{E}(G'G))^{-1}\tilde{G}'\tilde{e} \qquad \ldots(9)$$

where $\tilde{e}$ are the residuals from the model after fitting the $K-p$ estimates of θ_2 while restricting $\theta_1=\theta_{10}$, and $\tilde{E}(G'G)$ denotes that $E(G'G)$ is evaluated at the restricted estimates $\tilde{\theta}_1=\theta_{10}$ and $\tilde{\theta}_2$. If $E(G'G)$ in (9) is estimated as $\tilde{G}'\tilde{G}$ then (9) can be interpreted as NR^2 where R^2 is the usual coefficient of determination in the regression of $\tilde{e}$ on $\tilde{G}$. The statistic is taken as $\chi^2(p)$ under H_0.[3]

3.1. *Testing for a Liquidity Trap*

Konstas and Khouja (K–K) (1969) tested for a liquidity trap in the demand for money function but, as Spitzer (1976) observed, their algorithm failed to globally maximize the likelihood so that their LR test was misleading. Because such difficulties do not arise in the construction of an LM statistic, it is a good example of the utility of this approach.[4]

The equation estimated by K–K was

$$M_t=\gamma Y_t+\beta(R_t-\alpha)^{-1}+e_t\equiv g_t+e_t$$

with M_t, Y_t and R_t being money demand, income and the rate of interest respectively, and attention centred on whether $\alpha=0$ i.e. whether there was a liquidity trap. The LM test for this hypothesis can be constructed by setting $\theta_1=\alpha$, $\theta_2'=(\gamma,\beta)$ and testing $\theta_1=0$. This can be done in two steps:

(a) Regress M_t against Y_t and R_t^{-1} to get $\tilde{e}_t$ (this is the form under H_0 since $\alpha=0$).

(b) Form the derivatives $\partial g_t/\partial\gamma=Y_t$, $\partial g_t/\partial\beta=(R_t-\alpha)^{-1}$, $\partial g_t/\partial\alpha=\beta(R_t-\alpha)^{-2}$ and observe that, under H_0, these are Y_t, R_t^{-1} and $\tilde{\beta}R_t^{-2}$. Regress $\tilde{e}_t$ against Y_t, R_t^{-1} and R_t^{-2} and test NR^2 from this regression as $\chi^2(1)$. The value of LM $=11{\cdot}47$ indicates that the null hypothesis that there is no liquidity trap is rejected; a result which agrees with Spitzer's likelihood ratio test.

3.2. *Testing for Autocorrelation*

It has been observed by a number of authors–Aldrich (1978), Breusch (1978) and Godfrey (1978*c*)–that Durbin's h statistic can be derived from the general theory of the LM test. As there has been a recent trend in econometrics to utilize the autocorrelation function (a.c.f.) of residuals and transformations of it such as the Box–Pierce statistic for more comprehensive diagnostic tests for autocorrelation, it is interesting to demonstrate the connection between these procedures and the LM statistic, since this resolves some difficulties that have arisen in employing these quantities.

To begin, consider the regression model

$$y=X\beta+u$$

$$u=u_{-j}\rho_j+e$$

where $e\sim\mathcal{N}(0,\sigma^2I)$, u_{-j} is an $(N\times 1)$ vector containing u_{t-j} and X and is independent of u, i.e. it can contain lagged values of y. The transformed model is

$$y=\rho_j y_{-j}+(X-X_{-j}\rho_j)\beta+e\equiv g+e \qquad \ldots(10)$$

where the notation is obvious.

As (10) is in the format (8), the LM test for H_0: $\rho_j=0$ can be constructed by identifying θ_1 as ρ_j, θ_2 as β and applying (9). To construct the LM test $\tilde{e}$ and $\tilde{G}$ are needed. Because the restricted maximum likelihood estimates are $\rho_j=0$ and $\hat{\beta}_{\text{OLS}}$ from the regression of y against X, $\tilde{e}=\hat{u}$ (the least squares residuals from such a regression). To construct $\tilde{G}$ observe that $\partial g/\partial\rho_j=(y_{-j}-X_{-j}\beta)'=\mu'_{-j}$ and $\partial g/\partial\beta=(X-X_{-j}\rho_j)'$ so that $\tilde{G}=(\hat{u}_{-j}X)$. Since $\tilde{D}=\tilde{\sigma}^{-2}\tilde{G}'\tilde{e}$, $\tilde{D}_1=\hat{\sigma}^{-2}\hat{u}'_{-j}\hat{u}$. Now define the jth estimated autocorrelation coefficient as $r_j = (\hat{u}'\hat{u})^{-1}\hat{u}'_{-j}\hat{u}$ so that $\tilde{D}_1 = Nr_j$ because $\hat{\sigma}^2 = N^{-1}\hat{u}'\hat{u}$. Therefore it is clear that the LM test statistic for $\rho_j = 0$ involves testing if r_j differs significantly from zero.

To derive the LM test it is necessary to obtain $\tilde{\mathscr{I}}^{11}$. Because $\tilde{\mathscr{I}}$ can be estimated by $\tilde{\sigma}^{-2}\tilde{G}'\tilde{G}$,

$$\tilde{\mathscr{I}}^{11} = [\hat{\sigma}^{-2}(\hat{u}'_{-j}\hat{u}_{-j} - \hat{u}'_{-j}X(X'X)^{-1}X'\hat{u}_{-j})]^{-1}$$

and the LM statistic will be

$$\begin{aligned} &N^2 r_j^2[\hat{\sigma}^{-2}(\hat{u}'_{-j}\hat{u}_{-j} - \hat{u}'_{-j}X(X'X)^{-1}X'\hat{u}_{-j})]^{-1} \\ &\quad = Nr_j^2[N^{-1}\hat{\sigma}^{-2}\hat{u}'_{-j}\hat{u}_{-j} - \hat{\sigma}^{-2}(N^{-1}\hat{u}'_{-j}X)(N^{-1}X'X)^{-1}(N^{-1}X'\hat{u}_{-j})]^{-1}. \end{aligned}$$

If X is strictly exogenous and the null hypothesis is correct,

$$\hat{\sigma}^{-2}N^{-1}\hat{u}'_{-j}\hat{u}_{-j} \xrightarrow{p} 1, \quad \hat{\sigma}^{-2}N^{-1}\hat{u}'_{-j}X \xrightarrow{p} 0$$

so that the LM statistic becomes Nr_j^2 asymptotically, i.e. $N^{\frac{1}{2}}r_j$ could be treated as a standard normal deviate. But it is exactly this quantity which is central to most time series programmes based on the work of Box and Jenkins (1970), revealing that tests based on the a.c.f. are in fact LM statistics and so have asymptotic power equivalent to the LR and W statistics. It is also worthwhile noting that the same result would have been obtained if u_t followed an MA process as shown in Breusch (1978) and Godfrey (1978*b*).

What if X contains lagged dependent variables? Then $\hat{\sigma}^{-2}N^{-1}\hat{u}'_{-j}X$ does not generally tend in probability to zero and it is difficult to find its probability limit except in simple cases. One simple case is that given in Durbin (1970) where it tends to $(1, 0, \ldots, 0)'$ and the LM statistic will be Durbin's h statistic. In other cases it is simpler to obtain a test of the correct size by multiplying the usual a.c.f. output of Nr_j by the correction factor $[\hat{\sigma}^{-2}(\hat{u}'_{-j}\hat{u}_{-j} - \hat{u}'_{-j}X(X'X)^{-1}X'\hat{u}_{-j})]^{-\frac{1}{2}}$, and this is easily computed since $(X'X)^{-1}$ is already available in most computer programmes. Therefore there seems little reason for users of a.c.f. information to continue to assume that this correction factor is close to unity (as is done at present).

Now consider autocorrelation of the form

$$u_t = \rho_1 u_{t-1} + \ldots + \rho_m u_{t-m}$$

with $H_0: \rho_1 = \rho_2 = \ldots = \rho_m = 0$. Because $\partial g/\partial\rho_j = u'_{-j}$ it is easily seen that this hypothesis involves testing if $r_1, \ldots, r_m$ are jointly zero. But this is exactly what the Box–Pierce (1970) statistic $Q = N\sum_{k=1}^{m} r_k^2$ purports to do and it differs from the LM statistic in that it is distributed as $\chi^2(m-p)$ where p is the order of autoregression in y whereas LM would be $\chi^2(m)$. However, the Box–Pierce result follows by recognizing that the covariance matrix of $r_1, \ldots, r_m$ is approximately idempotent of order $(m-p)$, so that LM and Q should be close unless m is small (the nature of the approximation is spelled out in Breusch (1978).

Recognition of the fact that the Box–Pierce statistic is essentially the LM test for $H_0: \rho_1 = \ldots = \rho_m = 0$ emphasizes a number of points about it. Firstly, the implicit alternative hypothesis is that the disturbances follow an mth order linear process and that m parameters are being "estimated" to construct the test. This suggests that efforts to find better corrections to the distribution of Q under the null are only marginally useful for large m, as the power of the test is likely to be weak in such circumstances, and that the way to improve power is to ensure that m is kept small relative to N. In fact it seems doubtful that any investigator is really interested in testing such a general hypothesis as that embodied in the Q statistic (with m large), as evident from the few attempts made at estimating this alternative if the null is rejected. It is more likely that only a few of the ρ_j's would be estimated reflecting the fact that "some are more equal than others" for any particular type of data e.g. ρ_1, ρ_4, ρ_5 with quarterly economic data, and it would therefore seem more sensible for investigators to construct LM statistics for such alternatives than to rely upon "portmanteau" statistics of dubious power.

Secondly, the Q statistic has to be treated cautiously in regressions. When X contains only exogenous variables it is easily shown that Q should be referred to a $\chi^2(m)$ distribution but, when X contains both exogenous variables and lagged dependent variables, the covariance matrix of $r_1, \ldots, r_m$ will not generally be idempotent so that the Q statistic is entirely inappropriate. An LM statistic can be constructed by treating NR^2 from the regression of $\hat{u}_t$ against $\hat{u}_{t-1}, \ldots, \hat{u}_{t-m}, X_t$ as a $\chi^2(m)$ variable.

The previous examples were a non-linear regression model and a model that could be readily transformed into a non-linear regression. In the autocorrelation example, the Jacobian of the transformation could be neglected without affecting the asymptotic properties of the estimates or test statistic. There are other cases in econometrics where it is not possible to transform the model into an equivalent one with i.i.d. disturbances while retaining the same dependent variable. In many of these the analytic derivation of the LM test is so difficult that one would probably use the indirect method of solution. Nevertheless, there is another interesting class of problems in which a direct approach is feasible.

Consider the standard regression model with non-spherical disturbances

$$y = X\beta + u \qquad \ldots(11)$$

with $E(u) = 0$, $E(uu') = \Omega$ where $\Omega = \Omega(\theta)$ is a well-defined function of θ, a $(k \times 1)$ vector of parameters. With the regressors exogenous and the disturbances normally distributed the log likelihood (omitting constants) is

$$L = -\tfrac{1}{2}\log|\Omega| - \tfrac{1}{2}u'\Omega^{-1}u.$$

The information matrix will be block diagonal between θ and β parameters, and when the hypothesis to be tested involves only the θ parameters, the following quantities are required:[5]

$$\frac{\partial L}{\partial \theta} = -\tfrac{1}{2}A'(\Omega^{-1} \otimes \Omega^{-1})\operatorname{vec}(\Omega - uu') \qquad \ldots(12)$$

$$E\left[-\frac{\partial^2 L}{\partial\theta\partial\theta'}\right] = \tfrac{1}{2}A'(\Omega^{-1}\otimes\Omega^{-1})A \qquad \ldots(13)$$

where

$$A' = \partial \operatorname{vec} \Omega/\partial\theta.$$

Using $\tilde{\beta}$ and $\tilde{\theta}$ to denote ML estimates of the parameters when they are constrained by the null hypothesis, the LM statistic can be constructed from a regression of $\operatorname{vec}(\tilde{\Omega}^{-1} - \tilde{u}\tilde{u}')$ against $\tilde{A}$ in the metric of $(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1})$. In the special situation where $\Omega = \sigma^2 I$ under the null hypothesis (say $\theta_k = \sigma^2$ and H_0: $\theta_1 = \ldots = \theta_{k-1} = 0$) the statistic becomes:

$$\frac{1}{2\tilde{\sigma}^4}[\operatorname{vec}(\tilde{\sigma}^2 I - \tilde{u}\tilde{u}')]'\tilde{A}(\tilde{A}'\tilde{A})^{-1}\tilde{A}'\operatorname{vec}(\tilde{\sigma}^2 I - \tilde{u}\tilde{u}')$$

$$= \tfrac{1}{2}\left[\operatorname{vec}\left(I - \frac{\tilde{u}\tilde{u}'}{\tilde{\sigma}^2}\right)\right]'\tilde{A}(\tilde{A}'\tilde{A})^{-1}\tilde{A}'\operatorname{vec}\left(I - \frac{\tilde{u}\tilde{u}'}{\tilde{\sigma}^2}\right).$$

The statistic is one half of the explained sum of squares from the regression of $\operatorname{vec}(I - \tilde{u}\tilde{u}'/\tilde{\sigma}^2)$ against $\tilde{A}$ where the last column of $\tilde{A}$ will be $\operatorname{vec}(I)$ corresponding to $\theta_k = \sigma^2$.

Problems in this class include all of the familiar parametric models of non-spherical disturbances including heteroscedasticity and autocorrelation. As examples, we take the error components model of Balestra and Nerlove (1966) and testing diagonality of a covariance matrix in seemingly unrelated equation systems.

3.3. *The Error Components Model*

Balestra and Nerlove (1966) proposed the following model for N individuals observed over T time periods:

$$y_{it} = X_{it}\beta + u_{it} \qquad \dots(14)$$

$$u_{it} = \mu_i + \lambda_t + \nu_{it} \qquad \nu_{it} \sim \mathcal{N}(0, \sigma_\nu^2)$$
$$\mu_i \sim \mathcal{N}(0, \sigma_\mu^2) \qquad i = 1, \dots, N$$
$$\lambda_t \sim \mathcal{N}(0, \sigma_\lambda^2) \qquad t = 1, \dots, T \qquad \dots(15)$$

and a number of authors have provided estimators for it. (See Swamy (1973) for a survey). It seems fair to say that the model has not been extensively used, with most investigators preferring to allow for individual effects with dummy variables (taking μ_i, $i = 1, \dots, N$ as a set of constants) and omitting the λ_t time effects. To some extent, one suspects that the lack of enthusiasm stems from the fact that no test for the presence of individual and time effects, based on least squares residuals, has been available. In what follows, we demonstrate that the LM statistic provides such a test.

Following Nerlove (1971), the equations (14) and (15) may be written in matrix form as (11) with

$$E(uu') = \Omega = \sigma_\nu^2 I_{NT} + \sigma_\mu^2 (I_N \otimes e_T e'_T) + \sigma_\lambda^2 (e_N e'_N \otimes I_T)$$

where $e = (1, 1, \dots, 1)'$.

Setting $\theta = (\sigma_\mu^2, \sigma_\lambda^2, \sigma_\nu^2)$, the null hypothesis is that $\sigma_\mu^2 = \sigma_\lambda^2 = 0$ so that both individual and time effects are missing. Under the null hypothesis, $\Omega = \sigma_\nu^2 I_{NT}$ so ML estimates are obtained by OLS giving residuals $\tilde{u}$ and $\tilde{\sigma}_\nu^2 = \tilde{u}'\tilde{u}/NT$ as the estimate of σ_ν^2. It is interesting to observe that $\sigma_\mu^2 = \sigma_\lambda^2 = 0$ lie on the boundary of the parameter space and, as Chernoff (1954) pointed out, this creates some difficulties for the W and LR statistics (although these are fairly easily surmountable). Moran (1971) and Chant (1974) have analysed the properties of the LM statistic in such a "non-standard" situation and have shown that it is distributed as χ^2 under the null hypothesis.

The LM statistic will be given by one half of the explained sum of squares from the regression of vec $(I_{NT} - \tilde{u}\tilde{u}/\tilde{\sigma}^2)$ against vec $(I_N \otimes e_T e'_T)$, vec $(e_N e'_N \otimes I_T)$ and vec (I_{NT}). In some situations, this will be a convenient way to compute the statistic but, with only two parameters under test, it can be simplified considerably. In the notation introduced in Section 2:

$$\tilde{D} = -\frac{1}{2\tilde{\sigma}_\nu^4}\begin{bmatrix} \tilde{\sigma}^2 NT - \tilde{u}'(I_N \otimes e_T e'_T)\tilde{u} \\ \tilde{\sigma}^2 NT - \tilde{u}'(e_N e'_N \otimes I_T)\tilde{u} \\ \tilde{\sigma}^2 NT - \tilde{u}'\tilde{u} = 0 \end{bmatrix}$$

$$\tilde{\mathcal{J}} = \frac{NT}{2\tilde{\sigma}_\nu^4}\begin{bmatrix} T & 1 & 1 \\ 1 & N & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Hence the LM statistic would be

$$\mathrm{LM} = \frac{NT}{2}\left\{\frac{1}{T-1}\left[\frac{\tilde{u}'(I_N \otimes e_T e'_T)\tilde{u}}{\tilde{u}'\tilde{u}} - 1\right]^2 + \frac{1}{N-1}\left[\frac{\tilde{u}'(e_N e'_N \otimes I_T)\tilde{u}}{\tilde{u}'\tilde{u}} - 1\right]^2\right\}$$

taken as $\chi^2(2)$ under the null hypothesis. Note that

$$\tilde{u}'(I_N \otimes e_T e'_T)\tilde{u} = \textstyle\sum_{i=1}^N (\sum_{t=1}^T \tilde{u}_{it})^2 \quad \text{and} \quad \tilde{u}'(e_N e'_N \otimes I_T)\tilde{u} = \sum_{t=1}^T (\sum_{i=1}^N \tilde{u}_{it})^2.$$

If the time effects λ_t are assumed absent throughout, a common specification, the LM statistic for $\sigma_\mu^2 = 0$ is simply

$$\frac{NT}{2(T-1)}\left[\frac{\tilde{u}'(I_N \otimes e_T e'_T)\tilde{u}}{\tilde{u}'\tilde{u}} - 1\right]^2$$

to be taken as $\chi^2(1)$. This test statistic could then be viewed as an alternative to that described in Pudney (1977).

3.4. *Testing for Diagonality of the Error Covariance Matrix of a Seemingly Unrelated Equation* (SUE) *System*

It is sometimes of interest to test if the covariance matrix of disturbances in an SUE system is diagonal–see Albon and Valentine (1977)–and the LM statistic is a natural candidate for this, as estimation under the null hypothesis of diagonality involves OLS applied to each equation. Following the notation in Maddala (1977, pp. 465–466) the SUE system of m equations can be written as

$$y = X\beta + u$$

with the covariance matrix of u being $\sum \otimes I_N$. Excluding the diagonal elements there are $\frac{1}{2}m \times (m-1)$ unknown parameters in $\sum$ and these can be arranged in the vector θ, so that H_0: $\theta = 0$ and the model is in the general format to apply the LM test for non-spherical disturbances. Expression (13) gives the information matrix as $\frac{1}{2}A'(\Omega^{-1} \otimes \Omega^{-1})A$ and it is instructive to consider the element of this corresponding to the covariance between $\partial L/\partial\sigma_{ij}$ and $\partial L/\partial\sigma_{kl}$. Denoting A'_{ij} as

$$\frac{\partial \operatorname{vec} \Omega}{\partial \sigma_{ij}} = \left[\operatorname{vec}\left(\frac{\partial \sum}{\partial \sigma_{ij}} \otimes I_N\right)\right]',$$

under H_0: $\theta = 0$, $\tilde{\Omega}$ is diagonal and $A_{ij} = \operatorname{vec}(\imath_{ij} \otimes I_N)$ where $\imath_{ij}$ is an $(m \times m)$ matrix with unity in the (i, j)th and (j, i)th positions but zeros elsewhere. Therefore the restricted estimate of the information matrix has $\frac{1}{2}A'_{ij}(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1})A_{kl}$ as typical element, and

$$\begin{aligned}
(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1})A_{kl} &= (\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) \operatorname{vec}(\imath_{kl} \otimes I_N) \\
&= \operatorname{vec}(\tilde{\Omega}^{-1}(\imath_{kl} \otimes I_N)\tilde{\Omega}^{-1}) \\
&= \tilde{\sigma}_{kk}^{-1}\tilde{\sigma}_{ll}^{-1} \operatorname{vec}(\imath_{kl} \otimes I_N) \\
\therefore \quad A'_{ij}(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1})A_{kl} &= \tilde{\sigma}_{kk}^{-1}\tilde{\sigma}_{ll}^{-1} [\operatorname{vec}(\imath_{ij} \otimes I_N)]'[\operatorname{vec}(\imath_{kl} \otimes I_N)]
\end{aligned}$$

from which it is clear that the information matrix is diagonal with $N\tilde{\sigma}_{kk}^{-1}\tilde{\sigma}_{ll}^{-1}$ as diagonal element. This result implies that the LM statistic for H_0: $\theta = 0$ can be formed from

$$\textstyle\sum_{i=1}^{m} \sum_{j=1}^{i-1} \tilde{D}'_{ij}(\tilde{\sigma}_{ii}^{-1}\tilde{\sigma}_{jj}^{-1}N)^{-1}\tilde{D}_{ij}$$

where $D_{ij} = \partial L/\partial\sigma_{ij}$. But

$$\begin{aligned}
\tilde{D}_{ij} &= -\tfrac{1}{2}A'_{ij}(\tilde{\Omega}^{-1} \otimes \tilde{\Omega}^{-1}) \operatorname{vec}(\tilde{\Omega} - \tilde{u}\tilde{u}') \\
&= \tfrac{1}{2}\tilde{\sigma}_{ii}^{-1}\tilde{\sigma}_{jj}^{-1}\tilde{u}'(\imath_{ij} \otimes I_N)\tilde{u} \\
&= Nr_{ij}\tilde{\sigma}_{ii}^{-\frac{1}{2}}\tilde{\sigma}_{jj}^{-\frac{1}{2}}
\end{aligned}$$

where

$$r_{ij} = N^{-1}(\tilde{\sigma}_{ii}\tilde{\sigma}_{jj})^{-\frac{1}{2}}\tilde{u}'_i\tilde{u}_j.$$

Hence

$$\mathrm{LM} = N \textstyle\sum_{i=1}^{m} \sum_{j=1}^{i-1} r_{ij}^2 \qquad \ldots(16)$$

and this will be distributed as $\chi^2(\frac{1}{2}m \times (m-1))$ if the null hypothesis of diagonality is correct. Such a result is in accord with what would be expected and it seems likely to have been employed by many researchers over the years, and the present development demonstrates that it has asymptotic power equivalent to a likelihood ratio test.

There is one further example of the LM statistic which reveals, rather nicely, the appeal of basing inference upon the first derivatives of the likelihood rather than its level (LR) or a transformation of it (W). This arises in the situation when hypotheses are not nested. Pesaran (1974) and Pesaran and Deaton (1978) have exploited Cox's (1962) test for this, and the following paragraphs demonstrate that their statistic is in fact an LM statistic.

3.5. *Non-Nested Hypotheses*

The connection between LM and Cox's statistic was mentioned in Atkinson (1970, pp. 332–335) (although his equation (27) is the $C(\alpha)$ statistic referred to in the next section) and the connection can be seen as follows. Let $L_0^*(\theta_0)$ and $L_1^*(\theta_1)$ be the likelihoods under the two hypotheses H_0 and H_1. To make a choice between H_0 and H_1, one procedure would be to form a joint likelihood of the form $\mathrm{L}^* = (\int L_0^{*\lambda} L_1^{*1-\lambda})^{-1} L_0^{*\lambda} L_1^{*1-\lambda}$ and test if $\lambda = 0$ or 1 (notice that the factor $\int L_0^{*\lambda} L_1^{*1-\lambda}$ makes this a proper p.d.f.). The log likelihood is then $L = \lambda L_0 + (1-\lambda) L_1 - \log \int L_0^{*\lambda} L_1^{*1-\lambda}$, where $L_i = \log L_i^*$, and the LM test would be based upon $\partial L/\partial\lambda$.

Now, it is not hard to show that

$$\frac{\partial L}{\partial \lambda} = L_0 - L_1 - E(L_0 - L_1) = -L_{10} + E(L_{10})$$

where

$$L_{10} = L_1 - L_0.$$

The LM test for $\lambda = 1$ might therefore be based upon the score $\partial L/\partial\lambda$. To do so the joint distribution of the scores under H_0 is required and it is instructive here to observe that

$$\frac{\partial L}{\partial \theta_1} = (1-\lambda)\frac{\partial L_1}{\partial \theta_1} - \left(\int L_0^{*\lambda} L_1^{*1-\lambda}\right)^{-1} \int L_0^{*\lambda}(1-\lambda)\frac{\partial L_1^*}{\partial \theta_1} = 0$$

when $\lambda = 1$. Therefore the relevant parts of the information matrix are (when evaluated under H_0)

$$\mathscr{I}_{\lambda\lambda} = -E_0\left(\frac{\partial^2 L}{\partial\lambda\,\partial\lambda'}\right) = E_0(L_{10} - E(L_{10}))^2 = \mathrm{var}_0\,(L_{10}) = V_0(L_{10})$$

$$\mathscr{I}_{\lambda\theta_0} = -E_0\left(\frac{\partial^2 L}{\partial\lambda\,\partial\theta_0'}\right) = E_0\left(\frac{\partial L_{10}}{\partial\theta_0'}\right) = \eta'$$

$$\mathscr{I}_{\theta_0\theta_0} = -E\left(\frac{\partial^2 L}{\partial\theta_0\partial\theta_0'}\right) = -E_0\left(\frac{\partial^2 L_0}{\partial\theta_0\partial\theta_0'}\right) = Q$$

which follows by noting that $\partial L_1/\partial\theta_0 = 0$, $\lambda = 1$ and $E(\partial L_0/\partial\theta_0) = 0$. Thus the LM statistic becomes

$$[\hat{L}_{10} - E_0(\hat{L}_{10})]'[V_0(L_{10}) - \eta' Q \eta]^{-1}[\hat{L}_{10} - E_0(\hat{L}_{10})]$$

which is Cox's test in the type of notation in Pesaran and Deaton (1978, p. 681).

The above analysis demonstrates that Cox's test can be regarded as an LM test and this makes it easier to derive similar statistics for other ways of combining the likelihoods e.g. Quandt (1974) experimented with $\lambda L_0^* + (1-\lambda) L_1^*$. What is not answered in the above analysis is just what evaluating L_{10} under H_0 means. Notice that $\hat{L}_{10} = \hat{L}_1 - \hat{L}_0$ and $E_0(\hat{L}_{10}) = E_0(\hat{L}_1) - \hat{L}_0$ so that $\hat{L}_{10} - E_0(\hat{L}_{10}) = \hat{L}_1 - E_0(\hat{L}_1)$ i.e. the test is based on comparing the observed value of $\hat{L}_1$ with its expected value if H_0 were true. But L_1 is a function of θ_1 and θ_1 does not appear under H_0 so that it is not entirely clear how to evaluate $\hat{L}_1$. In fact two methods have been adopted. Cox forms $L_1(\hat{\theta}_1) - E_0(L(\hat{\theta}_{10}))$

where $\hat{\theta}_1$ maximizes L_1 and $\hat{\theta}_{10}$ is the plim of $\hat{\theta}_1$ under H_0, whereas Atkinson forms $L_1(\hat{\theta}_{10}) - E_0(L(\hat{\theta}_{10}))$. It would seem logical to adopt the latter from an LM viewpoint as the score is then unbiased, but de Pereira (1977) has recently claimed that the resulting test statistic is inconsistent, whereas Cox's is consistent. Clearly, this issue needs investigating as the differences in the statistics can be quite marked.

4. THE $C(\alpha)$ STATISTIC, THE SMALL SAMPLE PROPERTIES OF THE LM STATISTIC AND ITS USE AS A CONTROL VARIATE

Earlier sections have relied heavily upon the assumption that estimation under the null can be easily performed, generally by OLS, and have shown that many interesting econometric specifications fall in this class. Nevertheless, there do exist situations in which maximum likelihood estimation under the null is difficult, and yet it might be of interest to obtain an LM test. As an example, consider the heteroscedastic regression model

$$y_t = x_t'\beta + u_t \qquad u_t \sim \mathcal{N}(0, \sigma_t^2)$$
$$\sigma_t^2 = z_t'\alpha = \alpha_0 + \alpha_1 z_{1,t} + \dots + \alpha_q z_{q,t}.$$

Under suitable assumptions upon x_t and z_t, Breusch and Pagan (1979) and Godfrey (1978*a*) have shown that the LM test for $\alpha_1 = \dots = \alpha_q = 0$ (i.e. homoscedasticity) can be constructed from the OLS residuals. Now suppose one wished to test $\alpha_2 = \dots = \alpha_q = 0$. Then the maximum likelihood estimator of β, α_0 and α_1 under H_0 will involve the non-linear techniques discussed in Goldfeld and Quandt (1972) and the attractiveness of the LM test seems to disappear.

The difficulties raised for the LM test in such a case were investigated by Neyman (1959) and Buhler and Puri (1966), while Moran (1970) has written a good survey of this literature. Essentially, Neyman proposed what is called a $C(\alpha)$ test which involves the construction of a pseudo-LM test, with identical asymptotic properties to the true one, but requiring only root-N consistent estimators of the unknown parameters rather than ML ones. To appreciate how this modifies earlier theory, consider again the partitioning $\theta' = (\theta_1'\theta_2')$ where it is desired to test $\theta_1 = \theta_{10}$. Then if the information matrix is partitioned conformably as

$$\begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix},$$

the $C(\alpha)$ test is based upon

$$(\bar{D}_1 - \bar{\mathcal{I}}_{12}\bar{\mathcal{I}}_{22}^{-1}\bar{D}_2)'\bar{\mathcal{I}}^{11}(\bar{D}_1 - \bar{\mathcal{I}}_{12}\bar{\mathcal{I}}_{22}^{-1}\bar{D}_2) \qquad \dots(17)$$

where the estimates $\bar{\theta}_2$ used to evaluate the statistic now only need to be root-N consistent for θ_2 when H_0 is correct. With some algebra using the partitioned inverse of $\bar{\mathcal{I}}$, an alternative form of (17) is

$$C(\alpha) = \bar{D}'\bar{\mathcal{I}}^{-1}\bar{D} - \bar{D}_2'\bar{\mathcal{I}}_{22}^{-1}\bar{D}_2. \qquad \dots(18)$$

It seems reasonable to call this a pseudo-LM statistic although it is not directly related to the Lagrange multipliers in constrained maximization because it nevertheless is very similar to the form of the LM statistic we have used. When estimates which are only consistent under H_0 are used, $\bar{D}_2 \neq 0$; so to preserve the asymptotic properties of the statistic a correction needs to be made. This can be considered in terms of expression (17) to focus on the distribution of $\bar{D}_1$ conditional on the value of $\bar{D}_2$. Alternatively the pseudo-LM can be thought of as having the same form as (5) above, but reduced by an amount which accounts for the distribution of $\bar{D}_2$ when it is not set to zero as in maximum likelihood estimation. One interesting fact emerging from (18) is that the LM statistic evaluated at $\bar{\theta}_2$, $\bar{D}'\bar{\mathscr{I}}^{-1}\bar{D}$, will be larger than the appropriate $C(\alpha)$ statistic and this has application in testing for autocorrelation in systems of equations.

To illustrate this procedure, consider the pseudo-LM test for the non-linear regression specification of Section 3. Then the pseudo-LM statistic is calculated from three rather than two steps viz.

(i) Obtain root-N consistent estimates of the parameters under H_0, $\bar{\theta}$ and $\bar{\sigma}^2$, and denote the residuals by $\bar{e}$.

(ii) Regress $\bar{e}$ against $\bar{G}$ where all derivatives are included, and denote the coefficient of determination by R^*. This calculates $\bar{D}'\bar{\mathscr{I}}^{-1}\bar{D}$.

(iii) Regress $\bar{e}$ against $\bar{G}_2$ where G_2 are the derivatives with respect to θ_2 only and denote the coefficient of determination by R^{**}. This calculates $\bar{D}_2'\bar{\mathscr{I}}_{22}^{-1}\bar{D}_2$.

(iv) $N(R^*-R^{**})$ is $\chi^2(p)$ under H_0.[6]

There appear to be only a few studies of the small sample properties of the LM statistic. Peers (1971) investigated the local power to $O(N^{-\frac{1}{2}})$ of the LM test compared with the LR and W tests and found that there was no uniform ranking, the result depending upon the parameters of the model. Similar results were reported by Lee (1971) who examined the three tests in an application to multivariate analysis. Because the Durbin h-statistic can be interpreted as an LM statistic, the Monte Carlo studies by Maddala and Rao (1973) can be mentioned. Again, the relative rankings of LM and LR depend upon the model, but there was no hint that the LM statistic was in any sense inferior. Another study in which the performance of LM is comparable to other alternatives is that of Harvey and Phillips (1978) on heteroscedasticity in simultaneous equations.

This represents a very small range of comparative experience and suggests that much more detailed investigation of particular structures is required to provide any definitive answer. In the meantime, the computational advantages of the LM statistic constitute a weighty argument for its use. It would also seem valuable to apply Sargan's (1976*b*) small sample correction to any LM statistics (if these can be obtained).

Another interesting feature of the LM statistic is that its small sample distribution may sometimes be determined by numerical means. To illustrate this point, consider the LM statistic for the error components model

$$\mathrm{LM}=\frac{NT}{2}\left\{\frac{1}{(T-1)}\left[\frac{\tilde{u}'(I_N\otimes e_Te_T')\tilde{u}}{\tilde{u}'\tilde{u}}-1\right]^2+\frac{1}{(N-1)}\left[\frac{\tilde{u}'(e_Ne_N'\otimes I_T)\tilde{u}}{\tilde{u}'\tilde{u}}-1\right]^2\right\}$$

$$=\frac{NT}{2}\left\{\frac{1}{T-1}\left[\frac{u'Au}{u'Bu}-1\right]^2+\frac{1}{N-1}\left[\frac{u'Du}{u'Bu}-1\right]^2\right\}$$

where $B=(I-X(X'X)^{-1}X')$, $A=B(I_N\otimes e_Te_T')B$ and $D=B(e_Ne_N'\otimes I_T)B$. Now the distribution of either $(u'Bu)^{-1}u'Au$ or $(u'Bu)^{-1}u'Du$ is independent of σ_v^2, when $u\sim N(0,\sigma_v^2I_{NT})$, so that u can be regarded as a vector of standard normal deviates. Because A, B and D are functions only of the data, prob $(\mathrm{LM}>c)$ can then be determined by the following Monte Carlo method.

(i) Generate N observations on u.

(ii) Compute a value for LM, $\hat{LM}$, with (i) and compare to c.

(iii) Repeat (i) and (ii) R times and estimate prob $(\mathrm{LM}>c)$ by the ratio r/R where r is the number of times $\hat{LM}>c$.

The accuracy of this method improves as $R\to\infty$ but $R=5000$ was found to be satisfactory in Breusch and Pagan (1979). For some cases of the LM statistic e.g. that in Breusch and Pagan (1979), the exact probability of Type I error may alternatively be obtained by Imhof's method (1961) of numerically inverting the characteristic function, but the limited experiments carried out in Breusch and Pagan (1979) suggested that the Monte Carlo method was very much cheaper and of comparable accuracy.[7]

The fact that the small sample distribution of the LM statistic is available in some cases suggest an interesting use for it as a control variate in Monte Carlo studies of test statistics.

As Sargan (1976*a*, Appendix D) emphasizes, very large numbers of replications are needed to make a reasonable assessment of, say, the true Type I errors of test statistics corresponding to the large sample 1 per cent and 5 per cent critical values, making a control variate very useful for such investigations. Because the LM, W and LR statistics are asymptotically equivalent, it follows that LM may be a good control variate whenever its exact small sample distribution can be obtained e.g. as in the error components case above. Even when this is not possible however, it may be the case that a *statistic with the same form as* LM, *but evaluated with the true disturbances* rather than the residuals, has a small sample distribution that can be determined. One example of this would be for the h-statistic, i.e the control variate would be taken as

$$\frac{\sum u_t u_{t-1}}{\sum u_t^2}\sqrt{\left(\frac{N}{1-NV(\beta)}\right)}$$

where $V(\beta)$ would be the analytical expression for any particular model and u_t the true disturbance. This is the ratio of quadratic forms in normal variables and Imhof's method may be used to get the distribution.

5. CONCLUSION

The LM statistic has been exposited and derived for a number of interesting model specifications and it has been seen to reduce to some well known diagnostic tests for autocorrelation. In fact, the LM statistic appears very frequently in statistics and econometrics in a variety of guises–a partial listing would have to include the Wu–Hausman unobserved variable test and Kmenta's test for a Cobb Douglas Production function (see Engle (1978)), the Chow test, Andrews' functional form test (1971), the χ^2 contingency table test and the partial autocorrelation function. No doubt there are many more and it seems time to recognize that these diagnostic statistics all belong to one family, as extensions are thereby simplified. Apart from these general conceptual benefits the advantages of the LM statistic appear to be threefold: it normally requires least squares residuals; it is rarely difficult to compute; and its exact small sample distribution may be found in certain specific cases. The specifications examined in this paper were chosen to exhibit the different types of problem to which the LM statistic can be applied but they constitute only a small subset of those of interest to econometricians. However, it would not seem to be a difficult task for an applied researcher to construct an LM test for the particular specification in which he was interested.

We would like to thank Grayham Mizon and Stephen Yeo for their comments. The results on non-nested models were initially obtained with Angus Deaton during a visit to C.O.R.E.

NOTES

1. A similar result holds for the Gauss–Newton algorithm if this is applicable to the problem.
2. If x contains any lagged dependent variables there will be some added terms which disappear in large samples provided the autoregressive process is stable.
3. The test could be generalized to systems of "seemingly unrelated equations", in which case (9) would become

$$\tilde{e}'\tilde{V}^{-1}\tilde{G}(\tilde{E}(G'\tilde{V}^{-1}G))^{-1}\tilde{G}'\tilde{V}^{-1}\tilde{e}$$

where $V=\sum\otimes I$ is the covariance matrix of the system disturbances. The NR^2 version of this statistic has been also noted by Engle (1977) and Godfrey (1978*b*).

4. Spitzer mentions that the particular functional form chosen by K–K is probably not the best and suggests a use of the generalized Box–Cox transformation as in Zarembka (1968). It is worth noting that an LM test can be constructed for any particular functional form using this transformation, being of the same form as (9) if only transformations of X are made but probably needing to be calculated by the indirect method if the form of the transformation for y is unknown. We also point out that this example was selected to illustrate the steps leading to the construction of an LM statistic, rather than because we believe that the K–K specification admits of a reasonable test for a liquidity trap.

5. The following analysis utilizes a number of properties of vectors and the reader is referred to Magnus (1978) for a full account of these. In particular we use vec $(ABC) = C' \otimes A$ vec (B) and vec $(A)'$ vec $(B) =$ tr $A'B$.

6. The heteroscedasticity case this section was begun with is also a three regression one, as Amemiya (1977) has shown that estimates of α from the regression of the squares of the OLS residuals against z_t are root-N consistent.

7. There is one special case of equation (8) for which the exact small sample distribution of the LM statistic is F. This arises if (8) can be written as $y_t = z_t\beta + \phi(x_t; \alpha) + e_t$ with z_t and x_t strictly exogenous and, under $H_0: \alpha = \alpha_0$, the model reduces to a linear one. The LM statistic is then easily shown to involve testing if the variables $\Phi_t = \partial\phi(x_t; \alpha)/\partial\alpha$ evaluated at $\alpha = \alpha_0$ are significant in the regression of y_t against z_t and Φ_t.

REFERENCES

AITCHISON, J. and SILVEY, S. D. (1958), "Maximum-likelihood Estimation of Parameters Subject to Restraints", *Annals of Mathematical Statistics*, **29**, 813–828.

AITCHISON. J. and SILVEY, S. D. (1960), "Maximum-likelihood Estimation Procedures and Associated Tests of Significance", *Journal of the Royal Statistical Society*, Series B, **22**, 154–171.

ALDRICH, J. (1978), "An Alternative Derivation of Durbin's *h* Statistic", *Econometrica*, **46**, 1493–1494.

ALBON, R. P. and VALENTINE, T. J. (1977), "The Sectoral Demand for Bank Loans in Australia", *Economic Record*, **53**, 167–181.

AMEMIYA, T. (1977), "A Note on a Heteroscedastic Model", *Journal of Econometrics*, **6**, 365–370.

ANDREWS, D. F. (1971), "A Note on the Selection of Data Transformations", *Biometrika*, **58**, 249–254.

ATKINSON, A. C. (1970), "A Method for Discriminating Between Models", (with discussion), *Journal of the Royal Statistical Society*, Series B, **32**, 323–353.

BALESTRA, P. and NERLOVE, M. (1966), "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", *Econometrica*, **34**, 585–612.

BOX, G. E. P. and JENKINS, G. M. (1970) *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day).

BOX, G. E. P. and PIERCE, D. A. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, **65**, 1509–1526.

BREUSCH, T. S. (1978), "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economic Papers*, **17**, 334–355.

BREUSCH, T. S., AND PAGAN, A. R. (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica*, **47**, 1287–1294.

BUHLER, W. J. and PURI, P. S. (1966), "On Optimal Asymptotic Tests of Composite Hypotheses with Several Constraints", *Zeit. Wahrscheinlichkeitstheorie*, **5**, 71–88.

BYRON, R. P. (1970), "The Restricted Aitken Estimation of Sets of Demand Relations", *Econometrica*, **38**, 816–830.

CHANT, D. (1974), "On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions", *Biometrika*, **61**, 291–298.

CHERNOFF, H. (1954), "On the Distribution of the Likelihood Ratio", *Annals of Mathematical Statistics*, **25**, 573–578.

COX, D. R. (1962), "Further Results on Tests of Separate Families of Hypotheses", *Journal of the Royal Statistical Society*, Series B, **24**, 406–424.

COX, D. R. and HINKLEY, D. V. (1974) *Theoretical Statistics* (London: Chapman and Hall).

CROWDER, M. J. (1976), "Maximum Likelihood Estimation for Dependent Observations", *Journal of the Royal Statistical Society*, Series B, **38**, 45–53.

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

ENGLE, R. F. (1977), "Hypothesis Testing in Spectral Regression: The Lagrange Multiplier Test as a Regression Diagnostic", (paper presented to NBER Conference on Criteria for Evaluation of Econometric Models, Ann Arbor, 1977).

ENGLE, R. F. (1978), "The Lagrange Multiplier Test as a Regression Diagnostic", (Discussion Paper 78-6, University of California at San Diego).

FEIGIN, P. D. (1976), "Maximum Likelihood Estimation for Continuous-Time Stochastic Processes", *Advances in Applied Probability*, **8**, 712–736.

GODFREY, L. G. (1978*a*), "Testing for Multiplicative Heteroscedasticity", *Journal of Econometrics*, **8**, 227–236.

GODFREY, L. G. (1978*b*), "Testing against General Autoregressive and Moving Average Error Models when the Regressors include Lagged Dependent Variables", *Econometrica* **46**, 1293–1302.

GODFREY, L. G. (1978*c*), "Testing for Higher Order Serial Correlation in Regression Equations when the Regressors include Lagged Dependent Variables", *Econometrica* **46**, 1303–1310.

GOLDFELD, S. M. and QUANDT, R. E. (1972) *Nonlinear Methods in Econometrics* (Amsterdam: North-Holland).

GOLDFELD, S. M. and QUANDT, R. E. (1973), "The Estimation of Structural Shifts by Switching Regressions", *Annals of Economic and Social Measurement*, **2**, 475–486.

HARVEY, A. C. and PHILLIPS, G. D. A. (1978), "Testing for Heteroscedasticity in Simultaneous Equation Models", (Working Paper No. 8, University of Kent at Canterbury).

IMHOF, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables", *Biometrika*, **48**, 419–426.

KONSTAS, P. and KHOUJA, M. (1969), "The Keynesian Demand for Money Function", *Journal of Money, Credit and Banking*, **1**, 765–777.

LEE, YOONG-SIN (1971), "Asymptotic Formulae for the Distribution of a Multivariate Test Statistic: Power Comparisons of Certain Multivariate Tests", *Biometrika*, **58**, 647–651.

MADDALA, G. S. (1977) *Econometrics* (New York: McGraw Hill).

MADDALA, G. S. and RAO, A. S. (1973), "Tests for Serial Correlation in Regression Models with Lagged Dependent Variables and Serially Correlated Errors", *Econometrica*, **41**, 761–774.

MAGNUS, J. R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", *Journal of Econometrics*, **7**, 281–312.

MORAN, P. A. P. (1971), "Maximum Likelihood Estimation in Non-Standard Conditions", *Proceedings of the Cambridge Philosophical Society*, **70**, 441–450.

MORAN, P. A. P. (1970), "On Asymptotically Optimal Tests of Composite Hypotheses", *Biometrika*, **57**, 47–55.

NERLOVE, M. (1971), "A Note on Error Components Models", *Econometrica*, **39**, 383–396.

NEYMAN, J. (1959) "Optimal Asymptotic Tests of Composite Statistical Hypotheses", in *Probability and Statistics*, pp. 213–234 (U. Grenander (ed.)), (New York: Wiley).

PEERS, H. W. (1971), "Likelihood Ratio and Associated Test Criteria", *Biometrika*, **58**, 577–587.

PESARAN, M. H. (1974), "On the General Problem of Model Selection", *Review of Economic Studies*, **41**, 153–171.

PESARAN, M. H. and DEATON, A. S. (1978), "Testing Non-Tested Nonlinear Regression Models", *Econometrica*, **46**, 677–694.

DE PEREIRA, B. (1977), "A Note on the Consistency and on the Finite Sample Comparisons of Some Tests of Separate Families of Hypotheses", *Biometrika*, **64**, 109–114.

PUDNEY, S. (1977), "The Estimation of and Testing of Some Error Components Models" (Mimeo, London School of Economics).

QUANDT, R. E. (1974), "A Comparison of Methods for Testing Nonnested Hypotheses", *Review of Economics and Statistics*, **56**, 92–99.

RAO, C. R. (1973) *Linear Statistical Inference and Its Applications* 2nd ed., (New York: Wiley).

ROSENBERG, B. (1973), "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression", *Annals of Economic and Social Measurement*, **2**, 399–428.

SARGAN, J. D. (1976*a*), "Econometric Estimators and the Edgeworth Approximation", *Econometrica*, **44**, 421–448.

SARGAN. J. D. (1976*b*), "Some Approximations to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-Squared" (Mimeo, London School of Economics).

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

SPITZER, J. J. (1976), "The Keynesian Demand for Money Function Revisited", *Journal of Monetary Economics*, **2**, 381–387.

SWAMY, P. A. V. B. (1973), "Criteria, Constraints and Multicollinearity in Random Coefficient Regression Models", *Annals of Economic and Social Measurement*, **2**, 429–450.

ZAREMBKA, P. (1968), "Functional Form and the Demand for Money", *Journal of the American Statistical Association*, **43**, 502–511.