A GENERAL APPROACH TO THE CONSTRUCTION OF MODEL DIAGNOSTICS BASED UPON THE LAGRANGE MULTIPLIER PRINCIPLE

by

Robert F. Engle

NUMBER 156

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

A GENERAL APPROACH TO THE CONSTRUCTION OF MODEL DIAGNOSTICS BASED UPON THE LAGRANGE MULTIPLIER PRINCIPLE

by

Robert F. Engle

NUMBER 156

August 1979

This paper is a substantial revision of UCSD Discussion Paper 78-6. The writing and much discussion took place at the 1979 Warwick Summer Workshop, with support from the Social Science Research Council, and at the London School of Economics. The research was supported by NSF SOC 78-09476 and the International Centre for Economics and Related Disciplines. The author is indebted to a great many people for suggestions and discussions of this material. Trevor Breusch, Les Godfrey, Robert Hall, Andrew Harvey, David Hendry, Grayham Mizon, Adrian Pagan, Tom Rothenberg, Denis Sargan, Ken Wallis have all made helpful comments on the previous draft or the new material. All errors remain the author's responsibility.

Author's address:

Department of Economics B-003,
University of California, San Diego,
La Jolla CA. 92093, U.S.A.

1. Introduction

In order to assess the validity of the specification of an econometric model, it is useful to have a variety of diagnostic statistics which might provide evidence on the existence and possibly the type of misspecification involved. One source of diagnostics is hypothesis tests where the model under consideration is taken to be the null and the alternative is some generali ation. A particularly attractive approach is to construct optimal test statistics against a variety of specific alternatives. In this way is is possible to have reasonable power against a collection of interesting alternatives, although when looking at sets of non-independent statistics, one must be cautious about interpretations of the overall size of the test.

Strong rejection of any of these tests suggests some degree of misspecification as the data and model are apparently incompatible. The rejection does not however require acceptance of the alternative, as this test may have power against a collection of alternatives: comparisons of the test statistics with the underlying economic theory should help in formulating a strategy for respecifying the model.

In order to calculate such test statistics easily and inexpensively it is frequently desirable to avoid estimation of the alternative models. This is especially true when the alternative is more complicated than the null. The Lagrange Multiplier or Score test is ideal for this situation since it examines the fit of the model under the null for evidence of departures in the direction of interest. All of the tests formulated in this paper are calculated in terms of the residuals from estimation of the model under the null hypothesis. These LM tests can be thought of as ways of examining the residuals of a model for specific types of non-randomness.

The Lagrange Multiplier test has only recently become familiar to econometricaians through papers by Savin (1976), Berndt and Savin (1977) and Breusch (1979) who examine the numerical inequalities between LM Likelihood Ratio and Wald tests, Godfrey (1978) and Breusch and Pagan (1979) who develop tests for heteroscedasticity, Godfrey (1978a)(1978b) and Breusch and Pagan (1979a) who present tests for serial correlation non-linearities, variance components and non-nested hypotheses and Engle (1977) (1979) who derives tests for spectral regression problems and for autoregressive conditional heteroscedasticity. The basic principle was originally suggested by Rao (1948) and subsequently by Aitchison and Silvey (1958) and Silvey (1959) and examples are familiar in the Durbin Watson (1950) statistic, and Durbin's m test and approximately in Durbin's (1970) h test.

In this paper a different general approach is taken to the derivation of the tests, which eliminates the need to construct the information matrix of the full parameter set. In this way many apparently unrelated testing problems are easily seen to have the same solution. There are additional advantages for non-regular or non-normal problems although these are not exploited here. The approach is described in section 2 and the general theorems giving the distribution of the score are in section 3.

As an application of the approach, several groups of testing problems are analysed which have traditionally been found difficult. Section 4 gives simple examples while 5, 6 and 7 present results for non-linearities, common factor dynamics and simultaneous equation systems with particular attention to exogeneity testing.

2. The Approach

For most hypothesis testing problems, the Wald, Likelihood Ratio and Lagrange Multiplier tests share the optimality criterion of being asymptotically locally most powerful. This is most easily established for the LM procedure. Let $f(y;\theta)$ be the joint density of the data y as a function of a set of parameters θ. Under the null hypothesis the first p parameters θ_1 are restricted to take on the value θ_1^o. Under the alternative, $\theta_1 = \theta_1^o + \delta$ where for local alternatives δ will shrink to the zero vector usually as $1/\sqrt{T}$. The log of the likelihood ratio will be

$$\log f(y;\theta_1^o + \delta,\theta_2) - \log f(y;\theta_1^o,\theta_2)$$

which approaches the gradient or score times δ

$$\delta' \partial \log f(y;\theta_1^o,\theta_2)/\partial\theta_1^o$$

and therefore locally optimal tests will reject for large values of the score m in the relevant direction. The maximum likelihood estimates under the null of the nuisance parameters θ_2 would generally be used to evaluate the score statistic however other estimates with the same asymptotic distribution can be used with no change in large sample properties of the procedure. If θ_2 were known, then the problem would be a multidimensional Neyman-Pearson testing problem for which the critical region defined by the score would be locally most powerful invariant where the invariance comes from equal interest in all departures δ. If in addition, θ_1 were one dimensional and one sided tests were desired, then the score test would be locally most powerful, and, in the exponential family, uniformly most powerful. Even if θ_2 were estimated, frequently the critical region based on the score would be locally most powerful similar or unbiased. Asymptotically the parameters θ_2 can be considered known if they can be estimated consistently and therefore the asymptotic local optimality of the LM test follows directly. For more details see for example Cox and Hinckley (1974 , section 9.3).

This statistical argument suggests the asymptotic optimality of a critical region based upon the score. For economists another argument is perhaps more convincing. Estimation subject to a restriction can be considered as maximization of an objective function, the log likelihood, subject to a restriction. Associated with such a problem is a vector of Lagrange Multipliers which give the shadow price of the constraint. If the null hypothesis is false then the shadow price should be large and thus a critical region would be constructed for large values of the multipliers. For the set-up in the preceeding paragraph, the multipliers are simply the derivatives of the log likelihood evaluated at the null.

The standard approach to obtaining the critical region from the score is to use the information matrix. With L as the log likelihood, the information matrix and score are defined as

(1) $$d(y;\theta_1^o, \theta_2) = \partial L(y;\theta_1,\theta_2)/\partial\theta_1 \Big|_{\theta_1=\theta_1^o}$$

(2) $$\mathcal{J}(\theta_1,\theta_2) = -E(\partial^2 L/\partial\theta\,\partial\theta').$$

Letting $\hat{\theta}_2$ be the maximum likelihood estimate of θ_2 under the null, and $\mathcal{J}^{11}$ the partioned inverse of $\mathcal{J}$, the test is based on the statistic

(3) $$\xi = d(y;\theta_1^o, \hat{\theta}_2)' \mathcal{J}^{11}(\theta_1^o, \hat{\theta}_2)\, d(y;\theta_1^o, \hat{\theta}_2)$$

which has a limiting distribution which is χ^2_p when the null hypothesis is true. For the statistic ξ to have this limiting distribution, the regularity conditions of maximum likelihood theory are required. In addition, many familiar problems present non-standard situations with dependent sampling, non-identical distributions and boundary solutions. Hence the assertion of the limiting distribution frequently includes several unmentioned assumptions.

The approach used in this paper explicitly lists the assumptions required for a central limit theorem to hold for the score when the null is true.

Thus (4) $d(y;\theta_1^o, \hat{\theta}_2)/\sqrt{T} \xrightarrow{D} N(0,V_o)$

and the test is based upon

(5) $$\xi = d(y;\theta_1^o, \hat{\theta}_2)' V_o^{-1} d(y;\theta_1^o, \hat{\theta}_2)$$

which will by construction, have a limiting chi square distribution. Usually (4) will be satisfied without distributional assumptions on the data.

These approaches generally give the same test statistic, however there are several advantages to the formulation in (5). Most important for this paper, it is easier to derive a variety of tests from (5) than from (3) because the score vector frequently has the same form even though the information matrix may appear very different. Of theoretical importance, version (5) will often be available for non-standard problems such as when θ_1^o is on a boundary of the parameter space or when one tailed tests are required, or when the regularity conditions are not satisfied. Furthermore, (5) provides a certain degree of robustness to incorrect distributional assumptions. If an incorrect distribution is assumed for the data, the test in (5) will no longer be asymptotically optimal but at least it will have the appropriate size for large samples. The test in (3) will have unknown properties. Finally, the explicit listing of assumptions required for (4) is useful and holds out the possibility of better approximations to the asymptotic distribution than (5) itself. It also allows construction of a test when $\hat{\theta}_2$ is inefficient.

3. Limiting Distributions of the Score

All of the testing problems in this paper can be reparameterized to be omitted variable problems. In the model

$$(6) \qquad y = x\beta + z\theta + \varepsilon$$

the null hypothesis is simply $\theta=0$. If ε is normally distributed, the score for these problems will have the form

$$(7) \qquad d = z'\hat{\Omega}^{-1} u$$

where u are the residuals under the null and $\hat{\Omega}$ is the estimated covariance matrix under the null. The theorems below give sufficient conditions for normal limiting distribution of the score when the model is estimated with non-linearities, serial correlation, instrumental variables and several other complications. These conditions do not require normality of the data. The theorems also provide a unified approach to calculation of the test statistic.

To establish notation, let y_t, x_t and z_t be 1,k and p dimensional row vectors and let $y = (y_1,\ldots y_T)'$, $x = (x_1',\ldots,x_T')'$, $z = (z_1',\ldots,z_T')'$ be data matrices. All probability limits and limit statements are taken as T goes to infinity. The R^2 is taken as the (uncentered) sum of squares of the fitted values divided by the (uncentred) sum of squares of the dependent variable, all after any transformations. For models with an intercept and mean preserving transformations, the centered definition will give the same statistic.

Theorem 1

If

(i) $E(y_t | x_t) = x_t\beta$ and $E\,\varepsilon\varepsilon' = \sigma^2 I$ where $\varepsilon = y - x\beta$.

(ii) $\text{plim}\ w'w/T = Q$ which is non-singular. $Q = \begin{bmatrix} Q_{xx} & Q_{xz} \\ Q_{zx} & Q_{zz} \end{bmatrix}$

and $w = (x, z)$.

(iii) $w'\varepsilon / \sqrt{T} \xrightarrow{D} N(0, \sigma^2 Q)$

Then with $u = y - x(x'x)^{-1}x'y$

(a) $u'z / \sqrt{T} \xrightarrow{D} N(0, \sigma^2 (Q_{zz} - Q_{zx} Q_{xx}^{-1} Q_{xz}))$

(b) $\xi = u'z(z'z - z'x(x'x)^{-1} x'z)^{-1} z'u/\hat{\sigma}^2 \xrightarrow{D} \chi^2_p$ where $\hat{\sigma}^a = u'u/T$

(c) $\xi = TR^2$ of the regression of u on w.

Proof:

$$u'z/\sqrt{T} = \varepsilon'(I-x(x'x)^{-1}x')z/\sqrt{T}$$

$$= \varepsilon'(z-x(x'x)^{-1}x'z)/\sqrt{T}$$

$$= \frac{\varepsilon'w}{\sqrt{T}} \begin{pmatrix} -(x'x)^{-1}x'z \\ I \end{pmatrix}$$

$$= \varepsilon'w\, A_T/\sqrt{T}$$

By assumption (ii) plim A_T exists and let it be A. Then by (iii)

$$u'z/\sqrt{T} \xrightarrow{D} N(0,\sigma^2 A'QA) = N(0,\sigma^2(Q_{zz} - Q_{zx}Q_{xx}^{-1}Q_{xz}))$$

establishing (a). Result (b) is established by noting that the expression in the brackets in the statement, converges in probability to the variance of the limiting normal.

To establish (c)

$$TR^2 = Tu'w(w'w)^{-1}w'u/u'u$$

$$= (0,u'z)(w'w)^{-1}(0,u'z)'/u'u/T$$

because $u'x=0$. Taking the partitioned inverse of w'w yields the expression (b). Notice that if x includes an intercept, the centered and uncentered definitions of R^2 coincide.

Theorem 1 applies to the stochastic regressor problem in which least squares may be consistent but biased. Conditions (ii) and (iii) merely assert that the z's satisfy the same conditions as the x's themselves. More generally however, the z's may be smooth functions of estimated parameters such as $\hat{\beta}$, because in the limit $z(\hat{\beta})'\varepsilon/\sqrt{T}$ and $z(\hat{\beta})'w/T$ will have the same distributions for β as for $\hat{\beta}$.

Conditions (ii)(iii) can be established by appeal to more primitive assumptions. For example Schönfeld (1971) gives sufficient conditions for a dynamic simultaneous equation system to generate data satisfying these conditions. His theorem assumes that the system is stable, that the exogenous variables are bounded and have second moments and autocovariances at all lags and that the disturbances are gaussian. Weaker conditions are surely possible.

Several alternative computational approaches are available as is generally the case for linear hypothesis tests. Letting SSR_o be the sum of squared residuals from (6) under the null using the covariance matrix under the null and SSR_1 be the sum of squared residuals under the alternative using the covariance matrix from the null, Engle (1977) showed that $\xi = T(SSR_o - SSR_1)/SSR_o$. An even simpler procedure is to use the t or F test from a direct estimation of (6). These test statistics will differ from the pure LM test because σ^2 is estimated under the alternative and because it is defined as $u'u/T-K-P$ rather than by $u'u/T$. For finite samples it is not clear which is preferable and asymptotically they are equivalent tests. The use of the centered definition of R^2 may have a similar effect. If the mean of the residuals is not exactly zero then the test statistics will differ; however, the expected value will always be zero and therefore the statistics will not differ asymptotically.

These alternative computational approaches apply equally well to the results of theorems 1 and 2. For theorems 3 and 4 which deal with instrumental variables estimation, the t and F versions are available; however, the difference in the sums of squares residuals will no longer give the right statistic.

When there is only a single omitted variable, a one tailed test may be desirable. The square root of the chi square statistic will be normally distributed with the sign of the coefficient of z in the auxilliary regression.

To develop tests when the null hypothesis may be non-linear and there may be non-spherical disturbances, theorem 2 is required.

Theorem 2

If

(i) $E(y_t|x_t) = g(x_t,\beta)$, $E\,\varepsilon\varepsilon' = \sigma^2\Omega$, $\varepsilon = y-g$

(ii) $\Omega = \Omega(\omega)$ and $\hat{\omega}$ is a consistent estimator of ω. Let $\hat{\Omega} = \Omega\ (\hat{\omega})$.

(iii) $g(x_t,\beta)$ has second derivatives with respect to β which are uniformly bounded over t in a neighbourhood of the true β. $g(x_t,\tilde{\beta}) = g(x_t,\beta)$ for all t only if $\tilde{\beta}=\beta$.

(iv) $\text{Plim}\ w'\hat{\Omega}^{-1}\ w/T = Q$ non-singular with $w=(G,z)$ and $G= \left.\frac{\partial g}{\partial \beta}\right|_{\beta=\beta}$, a TxK matrix.

(v) $w'\hat{\Omega}^{-1}\varepsilon / \sqrt{T} \overset{D}{\to} N(0, {}^2Q)$.

(vi) $\hat{\beta}$ is the interior solution to $\min(y-g)'\hat{\Omega}^{-1}(y-g)$ and any one of the following conditions is true

(1) $g(x,\beta) = x\,\beta\,(\alpha)$ and $\beta\,(\alpha)$ has continuous first derivatives everywhere in the parameter space

(2) $\Omega = I$ and g has continuous first derivatives with respect to β everywhere in the parameter space

(3) $\hat{\beta}$ is consistent.

Then with $\hat{G} = \left.\frac{\partial g}{\partial \beta}\right|_{\beta=\hat{\beta}}$ and $u=y-\hat{g}$

(a) $z'\hat{\Omega}^{-1} u / \sqrt{T} \overset{D}{\to} N(0,\sigma^2 (Q_{zz} - Q_{zG} Q_{GG}^{-1} Q_{Gz})$

(b) $\xi = u'\hat{\Omega}^{-1} z(z'\hat{\Omega}^{-1} z - z'\hat{\Omega}^{-1} \hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})\hat{G}'\hat{\Omega}^{-1} z)^{-1} z'\hat{\Omega}^{-1} u/\hat{\sigma}^2 \overset{D}{\to} \chi^2_p$

$\hat{\sigma}^2 = u'\hat{\Omega}^{-1} u/T$.

(c) $\xi = T R^2$ of the regression of u on $(z,\hat{G})$ taking $\hat{\Omega}^{-1}$ as the covariance matrix.

(d) $\sqrt{T}\,(\hat{\beta} - \beta) \overset{D}{\to} N(0,\sigma^2 Q_{GG}^{-1})$

Proof : See Appendix.

Theorem 2 is a direct generalization of Theorem 1 to allow both non-linearities and non-spherical disturbances, in addition to stochastic regressors. If g is a linear function, then G=x.
Condition (v) excludes models with lagged dependent variables and serial correlation. In fact it excludes all cases where the information matrix is not block diagonal between the parameters ω and the coefficients of w.

The case of lagged dependent variables and serial correlation can generally be reformulated as a non-linear regression with white noise disturbances. In this case $\Omega = 1$, condition (v) is satisfied and Theorem 2 can be applied.

It is often difficult to verify (v) or even (vi) in the general case. Malinvaud (1970) Jennrich (1969) and Gallant and Holly (1978) give some sufficient conditions for the consistency and efficiency of non-linear least squares, which directly imply conditions (i)-(vi) when x is exogenous and $\Omega = I$. Even these more primitive assumptions are frequently extremely difficult to verify and often they are substantially stronger than would be necessary for the theorem. In practice, assumptions (iv) and (v) are quite plausible even in complicated models, and, as they are implicitly assumed for most econometric work, the more primitive assumptions are not explored.

If the model estimated under the null hypothesis has simultaneous equations bias or other stochastic problem which leads to correlation between x and ε, then theorem 3 is necessary. It deals with instrumental variables estimation but by careful choice of instruments, it also represents maximum likelihood estimators.

Theorem 3

Let $\hat{x}$ be a TxK matrix of instrumental variables which without loss of generality are assumed to satisfy $\hat{x}'x = \hat{x}'\hat{x}$.[1/]

If

(i) $y = x\beta + \varepsilon, \; E\varepsilon\varepsilon' = \sigma^2 I$

(ii) $\text{Plim}\; s's/T = Q$ where $s = (x, \hat{w}),\; \hat{w} = (\hat{x}, z)$ and $Q_{\hat{w}\hat{w}}$ is non-singular

(iii) $\hat{w}'\varepsilon/\sqrt{T} \xrightarrow{D} N(0, \sigma^2 Q_{\hat{w}\hat{w}})$

Then with $u = y - x(\hat{x}'\hat{x})^{-1}\hat{x}'y$ and $v = x - \hat{x}$

(a) $z'u/\sqrt{T} \xrightarrow{D} N(0, \sigma^2 P) \quad P = Q_{zz} - Q_{z\hat{x}} Q_{xx}^{-1} Q_{xz} - Q_{zx} Q_{\hat{x}\hat{x}}^{-1} Q_{\hat{x}z} + Q_{zx} Q_{\hat{x}\hat{x}}^{-1} Q_{xz}$

$$= Q_{zz} - Q_{z\hat{x}} Q_{\hat{x}\hat{x}}^{-1} Q_{\hat{x}z} + Q_{zv} Q_{\hat{x}\hat{x}}^{-1} Q_{vz}$$

(b) $\xi = u'z(\hat{P})^{-1} z'u/\hat{\sigma}^2 \xrightarrow{D} \chi^2_p$ where $\hat{\sigma}^2 = u'u/T$ and $\hat{P}$ is the sample value of P.

If in addition

(iv) $\text{Plim}\; z'(x - \hat{x})/T = 0$

then

1/ If $\hat{x}'x \neq \hat{x}'\hat{x}$, then let $x^* = x(\hat{x}'\hat{x})^{-1}\hat{x}'x$ to give $x^{*\prime}x^* = x'x^*$

(c) an asymptotically equivalent statistic is $T R^2$ of the regression of u on $\hat{w}$.

Proof : See Appendix.

Theorem 3 establishes the framework for calculating LM tests in simultaneous systems. The conditions (ii) and (iii) are sufficient for $\hat{x}$ to be a legitimate instrument, and imply the asymptotic normality of instrumental variable estimators. They are standard assumptions for instrumental variable problems. Assumption (iv) is somewhat stronger as it asserts that the omitted variable is asymptotically uncorrelated with the difference between x and its instrument. In a linear simultaneous equations set-up z might already be included with the predetermined variables and therefore (iv) would be satisfied exactly for two stage least squares. If the full set of candidate instruments is used in the first stage regression, then $x-\hat{x}$ is simply the reduced form disturbance which could reasonably be assumed uncorrelated with a new variable z. However, if less than the full instrument list is used, which might often be the case in non-linear simultaneous equation systems, then z could be correlated with the omitted candidate instruments and (iv) would fail. If an asymptotically efficient estimator is used then (iv) will be satisfied.

To generalize this theorem to system estimation and parameter restrictions within or between equations theorem 4 is required. Stacking the equations, the problem again becomes one of an instrumental variable estimator where there is a covariance matrix $\Omega=\Sigma \otimes I$ and the coefficient vector β may be a function of a smaller number of unrestricted parameters α.

Theorem 4

Let $\hat{x}$ be a matrix of instrumental variables which without loss of generality satisfy $\hat{x}'\hat{\Omega}^{-1}x = \hat{x}'\hat{\Omega}^{-1}\hat{x}$. If

(i) $y=x\beta + \varepsilon$, $E\varepsilon\varepsilon' = \sigma^2\Omega$,

(ii) $\Omega = \Omega(\omega)$ and $\hat{\omega}$ is a consistent estimator of ω. Let $\hat{\Omega} = \Omega(\hat{\omega})$.

(iii) $\beta = \beta(\alpha)$ where α is Jx1 J≤K. β has bounded derivatives to the second order in the neighbourhood of the true α. $\beta(\tilde{\alpha}) = \beta(\alpha)$ only if $\tilde{\alpha} = \alpha$.

(iv) Plim $s'\hat{\Omega}^{-1} s/T = Q$ where $s=(x,\hat{w})$, $\hat{w}=(\hat{x}, z)$ and $Q_{\hat{w}\hat{w}}$ is non-singular.

(v) $\hat{w}'\hat{\Omega}^{-1} \varepsilon / \sqrt{T} \overset{D}{\to} N(0, \sigma^2 Q_{\hat{w}\hat{w}})$

(vi) Let $\hat{\alpha}$ be the interior solution to min $(y-\hat{x}\beta)(\hat{\Omega}^{-1})(y-\hat{x}\beta)$ and let β have continuous first derivatives in the parameter space.

Then with $u=y-x\beta(\hat{\alpha})$, $H = \left.\frac{\partial\beta}{\partial\alpha}\right|_{\alpha=\alpha}$, $\hat{H} = \left.\frac{\partial\beta}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}$ each KxJ matrices and $v=x-\hat{x}$

(a) $z'\hat{\Omega}^{-1} u/ \sqrt{T} \overset{D}{\to} N(0,\sigma^2 P)$, $P = Q_{zz}-Q_{z\hat{x}} H(H'Q_{\hat{x}\hat{x}}H)^{-1} H'Q_{xz}$

$-Q_{zx}H(H'Q_{\hat{x}\hat{x}}H)^{-1} H'Q_{\hat{x}z} + Q_{zx}H(H'Q_{\hat{x}\hat{x}}H)^{-1}H'Q_{xz}$

$= Q_{zz}-Q_{z\hat{x}} H(H'Q_{\hat{x}\hat{x}}H)^{-1} H'Q_{\hat{x}z} + Q_{zv}H(H'Q_{\hat{x}\hat{x}}H)^{-1} Q_{vz}$

(b) $\xi = u'\hat{\Omega}^{-1}z\ (\hat{P})^{-1}\ z'\hat{\Omega}^{-1}\ u/\hat{\sigma}^2 \xrightarrow{D} \chi^2_p$ where $\hat{\sigma}^2 = u'\hat{\Omega}^{-1}\ u/T$ and $\hat{P}$ is the same as P with H replaced by $\hat{H}$ and all Q matrices replaced by sample estimates.

If in addition

(vii) $\text{Plim}\ z'\hat{\Omega}^{-1}\ (x-\hat{x})/T=0$

Then

(c) an asymptotically equivalent expression for ξ is $T\ R^2$ of the regression of u on $(\hat{x}\hat{H},\ z)$ using $\hat{\Omega}^{-1}$ as covariance matrix.

(d) $\sqrt{T}(\hat{\alpha}-\alpha) \xrightarrow{D} N(0,\ \sigma^2\ (H'Q_{\hat{x}\hat{x}}H)^{-1})$ under assumptions (i) - (vi).

Proof : See Appendix.

4. A Simple Example

In order to see the simplicity of these results consider the omitted variable problem in (6) with normal errors. If the model is a dynamic regression with white noise disturbances, then, the score is proportional to u'z. The test for θ=0 would be found by theorem 1 from TR^2 of the regression of the least squares residuals, u, on x and z.

If the model is a static model with non-white noise disturbances, then the test, by theorem 2, would be TR^2 of the regression of u on x and z using the estimated covariance matrix under the null. An alternative formulation is sometimes easier. If the data are transformed to make the disturbances white, let $\tilde{y}$ and $\tilde{x}$ be the transformed variables.

Then TR^2 can be calculated from the regression of $\tilde{u}$ on $\tilde{x}$, $\tilde{z}$.

If this model includes lagged dependent variables in either x or z, as well as serial correlation then theorem 2 must be applied as a non-linear regression problem with white noise errors. For the first order autoregression, this becomes TR^2 of the regression of $\tilde{u}$ on $\tilde{x}$, $\tilde{z}$, u_{-1}, where $\tilde{w} = w-\hat{\rho}w_{-1}$ for the estimate of ρ obtained under the null and $u = y-x\hat{\beta}$. See section 6 for more details.

If some of the x's are endogenous variables and are estimated with instruments $\hat{x}$, then theorem 3 must be invoked. Suppose the difference between x and $\hat{x}$ is purely a reduced form error term which is reasonably assumed to be uncorrelated with z. The test is calculated by TR^2 of the regression of u on $\hat{x}$, z, where u is the vector of residuals from instrumental variable estimation under the null. There is no need to recalcualte instruments.

If in addition this equation has first order autoregressive errors and there are no cross equation lags, then the auxillary regression would be $u-\hat{\rho}u_{-1}$ on $\hat{x}-\hat{\rho}x_{-1}$, $z-\hat{\rho}z_{-1}$, u_{-1} from which TR^2 would be calculated. This is an application of theorem 4.

If the equation is a stacked system of simulataneous equations, then theorem 4 is required where the covariance matrix has the form $\Omega \otimes I$. Let $\hat{x}$ be the instruments calculated under the null which may be the efficient instruments which give FIML as in Hendry (1976) or Hausman (1975) or simply reduced form regressions and let $\hat{\Omega}$ be the estimated contemporary covariance matrix. The test is based upon TR^2 of the regression of u on $\hat{x}$ and z using $\hat{\Omega} \otimes I$ as the covariance matrix.

In all of these cases the test statistic is asymptotically distributed as chi square with degrees of freedom equal to the rank of z. If this rank is one, then a one tailed test is available by taking the square root of the statistic and giving it the sign of the coefficient of z in the auxilliary regression.

5. Testing for Non-Linearities

Frequently an empirical relationship derived from economic theory is highly non-linear. This is typically approximated by a linear regression without any test of the validity of the approximation. The LM test generally provides a simple test of such restrictions because it uses estimates only under the null hypothesis. While it is ideal for the case where the model is linear under the null and non-linear under the alternative, the procedures also greatly simplify the calculation when the null is non-linear. Three examples will be presented which show the usefulness of this set of procedures.

Several studies have examined the demand for money to test for the existence of a liquidity trap. Pifer (1969), White (1972) and Eisner (1971) test for a liquidity trap in logarithmic or Box-Cox functional forms while Konstas and Khouja (1969) (K-K) use a linear specification. Most studies find maximum likelihood estimates of the interest rate floor to be about 2% but they differ on whether this figure is significantly different from zero. Pifer says it is not significant, Eisner corrects his likelihood ratio test and says it is, White generalizes the form using a Box-Cox transformation and concludes that it is not different from zero. Recently Breusch and

Pagan (1977a) have reexamined the Konstas and Khouja form and using a Lagrange Multiplier test, conclude that there is a liquidity trap.

Except for minor footnotes in some of the studies, there is no mention of the serial correlation which exists in the models. In re-estimating the Konstas-Khouja model, the Durbin-Watson statistic was found to be .3 which is evidence of a severe problem with the specification and that the distribution of all the test statistics may be highly misleading.

The model estimated by K-K is

$$M = \gamma Y + \beta(r-\alpha)^{-1} + \varepsilon \tag{8}$$

where M is real money demand, Y is real GNP and r is the interest rate. Perhaps their best results are when M1 is used for M and the long-term government bond rate is used for r. The null hypothesis to be tested is $\alpha = 0$. The normal score is proportional to $u'z$ where z is the derivative of the right-hand side with respect to α evaluated under the null:

$$z = \left.\frac{\partial g}{\partial \alpha}\right|_0 = \frac{\beta}{r^2}$$

Therefore the LM test is a test of whether $\frac{1}{r^2}$ belongs in the regression along with Y and 1/r.

Using the procedure from theorem 2, Breusch and Pagan obtain the statistic $\xi_{LM} = 11.47$ therefore rejecting $\alpha = 0$. Including a constant term this becomes 5.92 which is still very significant in the χ^2 table. However, correcting for serial correlation in the model under the null changes the results dramatically. A second-order autoregressive model was required to whiten the residuals with parameters 1.5295 and -.5597. These parameters are used in an auxiliary regression of the untransformed residuals on the three right-hand side variables and a constant, to obtain an R^2 = .01096. Thus, the LM statistic is $\xi_{LM} = .515$ which is distributed as χ^2 if the null is true. As can be seen it is very small suggesting that the liquidity trap is not significantly different from zero.

As a second example consider testing the hypothesis that the elasticity of substitution of a production function is equal to 1 against the alternative that it is constant but not unity. If y is output and X_1 and X_2 are factors of production, the model under the alternative can be written as

$$\text{(9)} \qquad \log y = -\frac{\alpha}{\rho} \log (\delta X_1^{-\rho} + (1-\delta) X_2^{-\rho}) + u.$$

If $\rho = 0$, the elasticity of substitution is one and the model becomes

$$\log y = \alpha \delta \log X_1 + \alpha (1-\delta) \log X_2 + u.$$

To test the hypothesis $\rho = 0$, it is sufficient to calculate $\left.\frac{\partial g}{\partial \rho}\right|_{\rho=0}$ and test whether this variable belongs in the regression. In this case $\left.\frac{\partial g}{\partial \rho}\right|_{\rho=0} = -\frac{\alpha}{2} \delta(1-\delta) (\log X_1/X_2)^2$ which is simply the Kmenta (1967) approximation. Thus the Cobb-Douglas form can be estimated with appropriate heteroscedasticity or serial correlation corrections and the unit elasticity assumption tested with power equal to a likelihood ratio test without ever doing a non-linear regression.

As a third example, Davidson, Hendry, Srba, and Yeo (1978) estimate a consumption function for the U.K. which pays particular attention to the model dynamics. The equation finally chosen can be expressed as

$$(10) \qquad \Delta_4 c_t = \beta_1 \Delta_4 y_t + \beta_2 \Delta_1 \Delta_4 y_t + \beta_3 (c_{t-4} - y_{t-4}) + \beta_4 \Delta_4 D_t + \beta_5 \dot{p}_t + \beta_6 \Delta_1 \dot{p}_t$$

where c y p are the logs of real consumption, real personal disposable income and the price level, and Δ_i is the i^{th} difference. In a subsequent paper Hendry and von Ungern-Sternberg (1979) argue that the income series is mis-measured in periods of inflation. The income which accrues from the holdings of financial assets should be measured by the real rate of interest rather than the nominal as is now done. There is a capital loss of $\dot{p}$ times the asset which should be netted out of income. The appropriate income measure is $y_t^* = \log (Y_t - \alpha \dot{p} L_{t-1})$ where L is liquid assets of the personal sector and α is a scale parameter to reflect the fact that L is not all financial assets.

The previous model corresponds to $\alpha=0$ and the argument for the respecification of the model rests on the presumption that $\alpha \neq 0$. The LM test can be easily calculated whereas the likelihood ratio and Wald tests require non-linear estimation. The derivative of y* with respect to α evaluated under the null is simply $-\dot{p} L_{t-1}/Y_t$. Denote this by x_t. The score is however u'z where $z = \hat{\beta}_1 \Delta_4 x_t + \hat{\beta}_2 \Delta_1 \Delta_4 x_t - \hat{\beta}_3 x_{t-4}$, and the betas are replaced by their estimates under the null. This is now a one degree of freedom test and can be simply performed using theorem 1. The test is significant with a chi squared value of 5. As a one tailed test it is significant at the 2.5% level.

6. Testing for Common Factor Dynamics

In formulating dynamic single equation models, it is common to consider some of the dynamics as due to the structure and some to the error term. In the model

(11) $\alpha^*(L)y_t = \gamma^*(L)x_t^{*\prime} + \varepsilon_t$

where $\alpha^*(L)$ is a general lag polynomial and $\gamma^*(L)$ is a matrix lag polynomial, Sargan (1964, 1975) and Hendry and Mizon (1978) suggest testing for a common factor such as that $\alpha^*(L) = \rho(L)\alpha(L)$ and $\gamma^*(L) = \rho(L)\gamma(L)$. In this case the restricted model can be rewritten as

(12) $\alpha(L)y_t = \gamma(L)\ x_t^{*\prime} + e_t,\ \rho(L)\ e_t = \varepsilon_t$

Putting the lagged y's on the right hand side and redefining the x* matrix and coefficient vector, this becomes

(13) $y_t = x_t\beta + e_t,\ \rho(L)\ e_t = \varepsilon_t$

Both Wald and likelihood ratio tests have been developed and employed by these authors. The Wald test has the advantage that the unrestricted model in (11) can be estimated by ordinary least squares and the test statistics computed by rearranging the output. Both likelihood ratio and LM tests require estimation by correcting for serial correlation.

However, once the investigator has chosen his model, he may want some final check on the validity of his dynamics. In this case, an LM test is ideal because he has already estimated the null and wants to examine it against a slightly less restricted model. In this spirit Godfrey (1979a) and Sargan (1969) have tested for omitted serial correlation.

Suppose the researcher conjectures that there are a set of lagged y, lagged x, and perhaps lagged u's which could enter the equation but are not already present. Call this data set z. The regression model is now

$$\text{(14)} \qquad \tilde{y} = \tilde{x}\beta + z\theta + \varepsilon, \qquad \tilde{x} = \rho(L)\, x$$

which is a non-linear least squares regression and the null hypothesis is $\theta=0$. The normal score is proportional to $\tilde{u}'z$ where $u=y-x\hat{\beta}$. To apply theorem 2 requires calculating the G matrix which is the derivative of ε with respect to the parameters β and ρ. The result is $G = (\tilde{x}, u_{-1}, \ldots, u_{-r})$ where r is the order of the serial correlation process, $\rho(L)$. The test is therefore easily computed by regressing $\tilde{u}$, the whitened residuals, on G and z and testing $T R^2$ as chi square with rank (z) degrees of freedom.

Frequently in the literature there is interest in testing the hypothesis that the order of the serial correlation process is r against the alternative that it is r-1. Consider the case where (13) is static. When r=1, this implies that $z=x_{-1}$. However, when $r > 1$ the formulation of the alternative must be carefully considered. Let $\rho(L) = (1-\phi L)\,\psi(L)$ where $\psi(L)$ is of order r-1 and ϕ is therefore one of the r possible roots of $\rho(L)$. Letting $\bar{x} = \psi(L)\, x$ the alternative can be written

$$\bar{y}_t = \phi\, \bar{y}_{t-1} + \bar{x}_t\beta + \bar{x}_{t-1}\gamma + \varepsilon_t$$

which can be expressed as

$$\text{(15)} \qquad \tilde{y}_t = \tilde{x}_t\beta + \bar{x}_{t-1}\,\theta + \varepsilon_t.$$

Thus the omitted variable is clearly $\bar{x}_{t-1}$. The LM test is easily calculated.

The difficulty with this test, however, is that it will give different answers depending upon which root ϕ is extracted. Furthermore, the full asymptotic power will not be achieved unless the appropriate choice is made. Unfortunately, when the null is true, the correct root is not identified, although there are only r possible values for ϕ. Several partial solutions are available. One might arbitrarily pick one of the $\bar{x}$ and sacrifice the optimal asymptotic power. One might include all of the $\bar{x}$'s and jointly test them which would give a test with rK degrees of freedom rather than K; however, this would in general be equivalent to testing against r=0. Finally one might follow Davies (1976) and develop a test based on $\max_i (\xi_i)$ but the distribution of r dependent chi square statistics would be difficult to find and thus the critical value would be unclear. In short, one must decide whether the alternative of serial correlation of order r-1 is of sufficiently great interest to solve this problem. Lots of very similar and sensible alternatives can easily be tested as described above.

7. Testing for Exogeneity

Tests for exogeneity are a source of controversy partly because of the variety of definitions of exogeneity implicit in the formulation of the hypotheses. In this paper the notions of weak and strong exogeneity as formulated by Richard et al (1979) will be used in the context of linear simultaneous equation systems. In this case weak exogeneity is essentially that the variables can be considered as instruments and the equations defining them can be ignored without a loss of information. Strong exogeneity implies, in addition, that the variables in question cannot be forecast by past values of endogenous variables which is the definition implicit in Granger (1969) "non-causality."

Consider a complete simultaneous equation system with G equations and K predetermined variables so that Y ε and V are TxG, X is TxK and the coefficient matrices are conformable. The structural and reduced forms are:

(16) $$YB = X\Gamma + \varepsilon \quad ; \quad E\varepsilon_t'\varepsilon_t = \Omega$$

(17) $$Y = X\Pi + V$$

where ε_t are rows of ε which are independent and the X are predetermined. Partitioning this set of equations into the first and the remaining G-1 the structure becomes

(18) $$y_1 - Y_2\beta = X_1\gamma + \varepsilon_1$$

(19) $$-y_1\alpha' + Y_2B_2 = X_2\Gamma_2 + \varepsilon_2$$

where X_2 may be the same as X and

(20) $$B = \begin{pmatrix} 1 & -\alpha' \\ -\beta & B_2 \end{pmatrix} \qquad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

The hypothesis that Y_2 is weakly exogeneous to the first equation in this full information context is simply the condition for a recursive structure:

(21) $$H_o : \alpha = 0, \Omega_{12} = 0$$

which is a restriction of 2G-2 parameters.

Several variations on this basic test are implicit in the structure. If the coefficient matrix is known to be triangular, then $\alpha = 0$ is part of the maintained hypothesis and the test becomes simply a test for $\Omega_{12} = 0$. This test is also constructed below: Holly (1979) generalizes the result to let the entire B matrix be assumed upper triangular and obtains a test of the diagonality of Ω. If some of the elements of β are known to be zero, then the testing problem remains the same except for the special case where B_2 is upper triangular between the included and excluded variables of Y_2 and the disturbances are uncorrelated with those of y_1 and the included y_2. Then it is only necessary to test that the α's and Ω's of the included elements of y_2 are zero. In effect, the excluded y_2 now form a higher level block of a recursive system and the problem can be defined a priori to exclude them also from Y_2. Thus without loss of generality the test in (21) can be used when some components of β take on known values.

To test (21) with (16) maintained, first construct the normal log likelihood L, apart from some arbitrary constants.

(22) $$L = T \log|B| - \frac{T}{2} \log |\Omega| - \tfrac{1}{2} \sum_{t=1,T} \varepsilon_t \, \Omega^{-1} \, \varepsilon_t$$

Partitioning this as in (20) using the identity $|\Omega| = |\Omega_{22}|\,|\Omega_{11} - \Omega_{12}\,\Omega_{22}^{-1}\,\Omega_{21}|$ gives

(23) $$L = T \log|B_2| + T \log |1-\alpha' B_2^{-1}\beta| - T/2 \log |\Omega_{22}| - T/2 \log |\Omega_{11} - \Omega_{12}\,\Omega_{22}^{-1}\,\Omega_{21}| - \tfrac{1}{2} \sum_t \varepsilon_{1t}\,\Omega^{11}\,\varepsilon_{1t}' - \tfrac{1}{2} \sum_t \mathcal{E}_{2t}\,\Omega^{22}\,\mathcal{E}_{2t}' - \sum_t \varepsilon_{1t}\,\Omega^{12}\,\mathcal{E}_{2t}$$

where the superscripts on Ω indicate the partitioned inverse and upper case characters represent matrices. Differentiating with respect to α and setting parameters to their values under the null gives the score:

$$(24) \qquad \left.\frac{\partial L}{\partial \alpha}\right|_0 = -T\,\hat{B}_2^{-1}\,\hat{\beta} + \sum_t \hat{\Omega}^{22}\,U_{2t}'\,y_{1t}$$

where hats represent estimates under the null and U_{2t} is the row vector of residuals under the null. Recognizing that $\sum_t \hat{\Omega}^{22}\,U_{2t}'\,U_{2t}/T = I$ this can be rewritten as

$$(25) \qquad \left.\frac{\partial L}{\partial \alpha}\right|_0 = \sum_t \hat{\Omega}^{22}\,U_{2t}'\,(y_{1t} - U_{2t}\,\hat{B}_2^{-1}\,\hat{\beta}) \equiv \sum_t \hat{\Omega}^{22}\,U_{2t}'\,(\bar{y}_{1t} + u_{1t})$$

where $\bar{y}_1$ is the reduced form prediction of y_1 which is given in this case as $X_1\hat{\gamma} + X_2\,\hat{\Gamma}_2\,\hat{B}_2^{-1}\,\hat{\beta}$. Clearly under the null hypothesis, the score will have expected value zero as it should. Using tensor notation this can be expressed as

$$(26) \qquad d_\alpha = (I \otimes (\bar{y}_1 + u_1))'(\hat{\Omega}_{22}^{-1} \otimes I)\ \text{vec}\ (U_2)$$

which is in the form of omitted variables from a stacked set of regressions with covariance matrix $\hat{\Omega}_{22}^{-1} \otimes I$. Theorem 4 applies directly and allows calculation of a test for $\alpha = 0$ under the maintained hypothesis that $\Omega_{12} = 0$.

The other part of the test in (21) is obtained by differentiating (23) with respect to Ω_{12} and evaluating under the null. It is not hard to show that all terms in the derivative vanish except the last. Because $\left.\partial\Omega^{12}/\partial\Omega_{12}\right|_0 = -\Omega_{11}^{-1}\ \Omega_{22}^{-1}$ the score can be written as

$$(27) \qquad d_{\Omega_{12}} = \sum_t u_{1t}\,\hat{\Omega}_{11}^{-1}\,\hat{\Omega}_{22}^{-1}\,U'_{2t}$$

which can be written in two equivalent forms

$$(28) \qquad d_{\Omega_{12}} = \hat{\Omega}_{11}^{-1} \; \hat{\Omega}_{22}^{-1} \; U'_2 \, u_1 .$$

$$(29) \qquad = \hat{\Omega}_{11}^{-1} \; (I \otimes u_1)' \; (\hat{\Omega}_{22}^{-1} \otimes I) \text{ vec } (U_2).$$

Either would be appropriate for testing $\Omega_{12} = 0$ when $\alpha = 0$ is part of the maintained hypothesis. In (28) the test would be performed in the first equation by considering U_2 as a set of G-1 omitted variables. In (29) the test would be performed in the other equations by stacking them and then considering $I \otimes u_1$ as the omitted set of variables. Clearly the former is easier in this case.

To perform the joint test, the two scores must be jointly tested against zero. Here (26) and (29) can easily be combined as they have just the same form. The test becomes a test for two omitted variables, $\bar{y}_1 + u_1$ and u_1, in each of the remaining G-1 equations. Equivalently, $\bar{y}_1$ and u_1 can be considered as omitted from these equations.

This test would be computed as an application of Theorem 4. Notice that the assumption that $Q_{\hat{w}\hat{w}}$ is non-singular is violated if the model is not identified under the alternative. Surely in this case the data would not be able to reject the hypothesis and indeed the likelihood would not increase at all by relaxing it. Thus a test on an unidentified model would give a zero test statistic (assuming the compute is able to take generalized inverses) and if the model is very weakly identified, the test would be likely to have very low power.

In the special case where G=2, the test is especially easy to calculate because both equations can be estimated by least squares under the null. Therefore theorem 1 can be applied directly.

As an example, the Michigan model of the monetary sector was examined. The equations are reported in Gardner and Hymans (1978). In this model, as in most models of the money market it is assumed that a short term interest rate can be taken as weakly exogenous in an equation for a long term rate. However, most portfolio theories would argue that all rates are set at the same time as economic agents shift from one asset to another to clear the market.

In this example a test is constructed for the weak exogeneity of the prime rate, RAAA , in the 35 year government bond rate equation, RG35 . The model can be written as

$$\text{(30)} \qquad RG35 = \beta \Delta RAAA + X_1\gamma + \varepsilon_1$$

$$\Delta RAAA = \alpha RG35 + X_2\gamma + \varepsilon_2$$

where the estimates assume $\alpha = \sigma_{12} = 0$, and the X's include a variety of presumably predetermined variables including lagged interest rates. Testing the hypothesis that $\alpha = 0$ by considering RG35 as an omitted variable is not legitimate as it will be correlated with ε_2 . If one does the test anyway, a chi squared value of 35 is obtained.

The appropriate test of the weak exogeneity of RG35 is done by testing u_1 and $RG35 - \hat{\beta}u_2$ as omitted from the second equation where $u_2 = \Delta RAAA - X_2\hat{\gamma}_2$. This test was calculated by regressing u_2 on X_2 , u_1 and $RG35 - \hat{\beta}u_2$. The resulting $TR^2 = 1.25$ which is quite small, indicating that the data does not contain evidence against the hypothesis. careful examination of X_1 and X_2 in this case shows that the identification of the model under the alternative is rather flimsy and therefore the best probably has very little power.

A second class of weak exogeneity tests can be formulated using the same analysis. These might be called limited information tests because it is assumed that there are no overidentifying restrictions available from the second block of equations. In this case equation (19) can be replaced by

$$(31) \qquad Y_2 = X\Pi_2 + \varepsilon_2 .$$

Now the definition of weak exogeneity is simply that $\Omega_{12} = 0$ because $\alpha = 0$ imposes no restrictions on the model. This situation would be expected to occur when the second equations are only very roughly specified.

A very similar situation occurs in the case where Y_2 is possibly measured with error. Suppose Y_2^* is the true unobserved value of Y_2 but one observes $Y_2 = Y_2^* + \eta$. If the equation defining Y_2^* is

$$Y_2^* = X_2\Gamma_2 + \varepsilon_2$$

where the assumption that Y_2^* belongs in the first equation implies $E\ \varepsilon_1'\varepsilon_2 = 0$. The observable equations become

$$(32) \qquad y_1 = Y_2\beta + X_1\gamma + \varepsilon_1 - \eta\beta$$

$$Y_2 = X_2\Gamma_2 + \varepsilon_2 + \eta$$

If there is no measurement error, then the covariance matrix of η will be zero, and $\Omega_{12} = 0$. This set up is now just the same as that used by Wu (1973) to test for weak exogeneity of Y_2 when it is know that $\alpha = 0$.

The procedure for this test has already been developed. The two forms of the score are given in (28) and (29) and these can be used with Theorem to test for the presence of U_2 in the first equation. This test is Wu's test and it is also the test derived by Hausman (1979) for this problem. By showing that these are Lagrange Multiplier tests, the asymptotic optimality of the procedures is established when the full set of X_2 is used. Neither Hausman nor Wu could establish this property.

Finally, tests for strong exogeneity can easily be performed. By definition, strong exogeneity requires weak exogeneity plus the non-predictability of Y_2 from past values of y_1. Partitioning X_2 in (19) into (y_1^0, X_3) where y_1^0 is a matrix with all the relevant lags of y_1, and similarly letting $\Gamma_2 = (\Gamma_{20}, \Gamma_{23})$ the hypothesis of strong exogeneity is

$$(33) \qquad H_0 \quad : \quad \alpha = 0, \Omega_{12} = 0, \Gamma_{20} = 0$$

This can clearly be jointly tested by letting u_1, $\bar{y}_1$ and y_1^0 be the omitted variables from each of the equations. Clearly the weak exogeneity and the Granger non-causality are very separate parts of the hypothesis and can be tested separately. Most often however when Granger causality is being tested on its own, the appropriate model is (31) as overidentifying restrictions are rarely available.

VII. Conclusions

Lagrange Multiplier tests have been derived and applied to a wide variety of situations from omitted variables to non-linearities, common factor dynamics, errors of measurement and exogeneity. In general they are inexpensive to compute and can be reported by a single statistic.

A recommendation of this paper is that a series of such tests be performed for each equation published and probably for each equation estimated by a researcher. The range of alternatives as well as the values of the statistics considered would provide a measure of confidence which a reader or researcher could have in the particular result. The approach has the advantage that each diagnostic test requires only a single number rather than a fully tabulated regression and thus leads to economies of reading, printing and digesting as well as computing. It is possible that wider use of these techniques would make a small step forward in improving the reliability of economic models. If so, the costs seem small.

APPENDIX

Proof of Theorem 2

First the limiting distribution of $\hat{\beta}$ must be established, by showing consistency (if vi(3) has not been assumed) and then the limiting distribution.

From condition (vi),

$$(A1) \qquad (y-\hat{g})'\hat{\Omega}^{-1}(y-\hat{g}) \leq \varepsilon'\hat{\Omega}^{-1}\varepsilon \ .$$

Applying the mean value theorem

$$(A2) \qquad \hat{g} = g + G^*(\hat{\beta} - \beta)$$

where G^* is G evaluated at β^* which lies between $\hat{\beta}$ and β .

Substituting $y = g + \varepsilon$ and (A2) into (A1) gives

$$(A3) \qquad (\hat{\beta} - \beta)'G^{*\prime}\hat{\Omega}^{-1}G^*(\hat{\beta} - \beta) \leq 2\ \varepsilon'\hat{\Omega}^{-1}G^*\ (\hat{\beta} - \beta)$$

If (vi(1)) is assumed then $G^* = x\ H^*$ where $H^* = \left.\frac{\partial\beta}{\partial\alpha}\right|_{\alpha = \alpha^*}$ so that the right hand side of (A3) becomes $\varepsilon'\ \hat{\Omega}^{-1}xH^*$. Dividing both sides by T and taking probability limits, establishes by (v) that the quadratic form

$$\text{plim}\ (\hat{\beta} - \beta)'H^{*\prime}\ Q_{xx}H^*(\hat{\beta} - \beta) \leq 0\ .$$

Because the interior matrix is positive definite, $\text{plim}(\hat{\beta} - \beta) = 0$. If on the other hand, (vi(2)) has been assumed, then the right hand side of (A3) is proportional to $\sum_t \varepsilon_t \left.\frac{\partial g(X_t, \beta)}{\partial \beta}\right|_{\beta = \beta^*}$. Because the conditional expectation of ε_t given x_t is zero, each term in the sum has expectation zero and the law of large numbers establishes that

$$\text{plim}(\hat{\beta} - \beta)' Q^{**}_{GG} (\hat{\beta} - \beta) \leqslant 0$$

and therefore $\hat{\beta}$ is consistent.

The estimate $\hat{\beta}$ must satisfy the first order conditions

$$\text{(A4)} \qquad (y - g)' \hat{\Omega}^{-1} \hat{G} = 0 \; .$$

From (A2) this becomes

$$\text{(A5)} \qquad \hat{\beta} - \beta = (\hat{G}' \hat{\Omega}^{-1} G^*)^{-1} \hat{G}' \hat{\Omega}^{-1} \varepsilon$$

which will be shown to have the same limiting distribution as

$$\text{(A6)} \qquad (G' \hat{\Omega}^{-1} G)^{-1} G' \hat{\Omega}^{-1} \varepsilon \; ,$$

which by (iv) and (v) establishes

$$\sqrt{T}\, (\hat{\beta} - \beta) \overset{D}{\rightarrow} N(0, \sigma^2 Q_{GG})$$

Now to show that (A5) and (A6) have the same limiting distribution.

$$\text{Plim} \left| \frac{w' \hat{\Omega}^{-1} (\hat{G} - G)}{T} \right| \leqslant \text{plim} \left| \frac{w' \hat{\Omega}^{-1} B}{T} \right| |b|$$

where B is a TxK matrix with the uniform upper bound of $\partial G/d\beta$ in the neighbourhood of β as guaranteed by (iii), and b is the maximum $|\hat{\beta}_j - \beta_j|$. From assumption (v) the first term has a finite probability limit while the second goes to zero because $\hat{\beta}$ is consistent. Hence $\hat{G}'\hat{\Omega}^{-1}G^*/T$ has the same limit as $\hat{G}'\hat{\Omega}^{-1}G/T$ which is the same as $G'\hat{\Omega}^{-1}G/T$. Similarly $\varepsilon'\hat{\Omega}^{-1}\hat{G}/\sqrt{T}$ has the same limiting distribution as $\varepsilon'\hat{\Omega}^{-1}G/\sqrt{T}$. This establishes (d).

Now to find the limiting distribution of the score.

$$z'\hat{\Omega}^{-1}u/\sqrt{T} = z'\hat{\Omega}^{-1}(\varepsilon + g - \hat{g})/\sqrt{T}$$

$$= z'\hat{\Omega}^{-1}(\varepsilon - G^*(\hat{\beta} - \beta))/\sqrt{T}$$

$$= z'\hat{\Omega}^{-1}(\varepsilon - G^*(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\ \hat{G}'\hat{\Omega}^{-1}\varepsilon)/\sqrt{T}$$

which has the same limiting distribution as

$$(z'\hat{\Omega}^{-1}\varepsilon - z'\hat{\Omega}^{-1}G(G'\hat{\Omega}^{-1}G)^{-1}\ G'\hat{\Omega}^{-1}\varepsilon)/\sqrt{T} = A_T\ w'\hat{\Omega}^{-1}\varepsilon/\sqrt{T}$$

where $A_T = (-z'\hat{\Omega}^{-1}G(G'\hat{\Omega}^{-1}G)^{-1}\ , I)$. The probability limit of A_T is $A = (-Q_{zG}Q_{GG}^{-1}, I)$ and therefore by (iv) and (v)

$$z'\hat{\Omega}^{-1}u/\sqrt{T} \xrightarrow{D} N(0, \sigma^2\ A\ Q\ A') = N(0, \sigma^2(Q_{zz} - Q_{zG}Q_{GG}^{-1}Q_{Gz}))$$

establishing a.

Propositions (b) and (c) follow directly from the same argument as follows theorem 1. When G includes an intercept and preserves means, the estimates based on the centered and uncentered R^2 will be numerically equal. However, when the mean only has expected value zero, the statistics will only be the same asymptotically.

Proof of Theorem 3

To establish the limiting distribution of the score

$$z'u/\sqrt{T} = z'(y - x(\hat{x}'x)^{-1}\hat{x}'y)/\sqrt{T}$$

$$= z'(\varepsilon - x(\hat{x}'\hat{x})^{-1}\hat{x}'\varepsilon)/\sqrt{T}$$

$$= A_T \hat{w}'\varepsilon/\sqrt{T}$$

where $A_T = (-z'x(\hat{x}'\hat{x})^{-1}, I)$. The probability limit of $A_T = A$ by assumption (ii) and therefore

$$z'u/\sqrt{T} \to N(0, \sigma^2 AQ_{\hat{w}\hat{w}}A') = N(0, \sigma^2 P)$$

where $P = Q_{zz} - Q_{z\hat{x}}Q_{\hat{x}\hat{x}}^{-1}Q_{xz} - Q_{zx}Q_{\hat{x}\hat{x}}^{-1}Q_{\hat{x}z} + Q_{zx}Q_{\hat{x}\hat{x}}^{-1}Q_{xz}$. This rather surprising expression is symmetric and positive definite. Letting $x - \hat{x} = v$ and defining Q_{zv} in the obvious fashion

$$P = Q_{zz} - Q_{z\hat{x}}Q_{\hat{x}\hat{x}}^{-1}Q_{\hat{x}z} + Q_{zv}Q_{\hat{x}\hat{x}}^{-1}Q_{vz} .$$

This expression differs from the previous cases because of the third term.

Under (iv) $Q_{zv} = 0$. Then noting that $u'\hat{w} = (0,u'z)$ the statistic can be calculated as $T R^2$ just as in theorem 1.

Proof of Theorem 4

First establish the limiting distribution of $\hat{\alpha}$ and $\hat{\beta}$. From the mean value theorem

$$\text{(A7)} \qquad \hat{\beta} = \beta + H^*(\hat{\alpha} - \alpha)$$

where $H^* = \left.\frac{\partial\beta}{\partial\alpha}\right|_{\alpha = \alpha^*}$ and α^* lies between $\hat{\alpha}$ and α .

The value $\hat{\alpha}$ must satisfy the first order conditions for a minimum;

$$0 = (y - \hat{x}\hat{\beta})'\hat{\Omega}^{-1} \hat{x} \hat{H} =$$

$$= (\varepsilon + x\beta - \hat{x}\hat{\beta})'\hat{\Omega}^{-1} \hat{x} \hat{H}$$

$$= (\varepsilon - \hat{x}(\hat{\beta} - \beta))'\hat{\Omega}^{-1} \hat{x} \hat{H} \quad \text{from the construction of } \hat{x}$$

$$= (\varepsilon - \hat{x}H^*(\hat{\alpha} - \alpha))'\hat{\Omega}^{-1} \hat{x} \hat{H} \quad \text{by (A7)}$$

$$\text{So} \quad \hat{\alpha} - \alpha = (\hat{H}'\hat{x}'\hat{\Omega}^{-1}\hat{x}H^*)^{-1} \hat{H}'\hat{x}\hat{\Omega}^{-1}\varepsilon$$

which has the same limiting distribution, by the argument of theorem 2, as

$$(H'x'\hat{\Omega}^{-1} x H)^{-1}H'x'\hat{\Omega}^{-1}\varepsilon$$

and therefore

$$\sqrt{T}(\hat{\alpha} - \alpha) \overset{D}{\to} N(0, \sigma^2 (H' Q_{\hat{x}\hat{x}} H)^{-1}) .$$

The score is given by

$$z'\hat{\Omega}^{-1}u/\sqrt{T} = z'\hat{\Omega}^{-1}(y - x\hat{\beta})/\sqrt{T}$$

$$= z'\hat{\Omega}^{-1}(\varepsilon - x(\hat{\beta} - \beta))/\sqrt{T}$$

$$= z'\hat{\Omega}^{-1}(\varepsilon - xH^*(\hat{\alpha} - \alpha))/\sqrt{T}$$

$$= z'\hat{\Omega}^{-1}(\varepsilon - xH^*(\hat{H}'\hat{x}'\hat{\Omega}^{-1}\hat{x}H^*)^{-1}\hat{H}'\hat{x}\hat{\Omega}^{-1}\varepsilon)/\sqrt{T}$$

$$= A_T\, w'\, \varepsilon/\sqrt{T}$$

where $A_T = (-z'\hat{\Omega}^{-1} xH^*(\hat{H}'\hat{x}'\hat{\Omega}^{-1}\hat{x}H^*)^{-1}\hat{H}', I)$. By (iv) A_T has a limiting distribution $A = (-Q_{zx}H(H'Q_{xx}H)^{-1}H', I)$ and therefore

$$z'\hat{\Omega}^{-1}u/T \overset{D}{\to} N(0, \sigma^2 AQ_{\hat{w}\hat{w}}A) = N(0, \sigma^2 P)$$

where P is given in the text.

When $Q_{zx} = Q_{z\hat{x}}$ then the expression for P simplifies to give the lower right hand corner of the partitioned inverse of $H'Q_{\hat{w}\hat{w}}H$ and therefore $T\,R^2$ of the regression of u on $\hat{w}$ using $\hat{\Omega}$ as covariance matrix will calculate the statistic.

BIBLIOGRAPHY

Aitchison, J. and S.D. Silvey (1958), "Maximum Likelihood Estimation of Parameters Subject to Restraints", Annals of Mathematical Statistics, 29, pp.813-828.

Berndt, E.R. and N.E. Savin (1977), "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", Econometrica, 45, pp. 1263- 1278.

Breusch, T.S. (1979), "Conflict Among Criteria for Testing Hypotheses", Econometrica, 47.

Breusch, T.S. and A.R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation", forthcoming Econometrica.

Breusch, T.S. and A.R. Pagan (1979a), "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics", Review of Economic Studies, forthcoming.

Cox, D.R. and D.V. Hinkley (1974), Theoretical Statistics, Chapman and Hall, London.

Davidson, J.E.H., D.F. Hendry, F. Srba and S. Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom", Economic Journal, 88, pp.661-692.

Davies, R.B. (1977), "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative", Biometrika, 64, pp.247-254.

Durbin, J. and G.S. Watson (1950), "Testing for Serial Correlation in Least Squares Regression", Biometrika, 37, pp.409-428; II, Biometrika, 38, 1951, pp.159-177.

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

Eisner, R. (1971), "Non-Linear Estimates of the Liquidity Trap", Econometrica, 39, pp.861-864.

Engle, R.F. (1977), "Hypothesis Testing in Spectral Regression: The Lagrange Multiplier as a Regression Diagnostic", in Criteria for Evaluation of Econometric Models, Ed. J. Kmenta and J. Ramsey, NBER Conference Proceedings, Ann Arbor, May.

Engle, R.F. (1979), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of Inflationary Expectations" manuscript presented to European Meetings of the Econometric Society, Athens, September.

Gallant, A.R. and A. Holly (1978), "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation", manuscript.

Godfrey, L.G. (1978), "Testing for Multiplicative Heteroskedasticity", Journal of Econometrics, 8, pp.227-236.

Godfrey, L.G. (1978a), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors include Lagged Dependent Variables", Econometrica, 46, pp.1293-1302.

Godfrey, L.G. (1978b), "Testing for Higher Order Serial Correlation in Regression Equations when the Regressors include Lagged Dependent Variables", Econometrica, 46, pp.1303-1310.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", Econometrica, 37, pp.424-438.

Hausman, J.A. (1975), "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models", Econometrica, 43, pp.727-738.

Hausman, J.A. (1979), "Specification Tests in Econometrics", Econometrica.

Hendry, D.F. (1976), "The Structure of Simultaneous Equations Estimators" Journal of Econometrics, 4, pp.51-88.

Hendry, D.F. and G.E. Mizon (1978), "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England", Economic Journal, 88, pp.549-563.

Hendry, D.F. and T. von Ungern-Sternberg (1979), "Liquidity and Inflation Effects on Consumer's Expenditure" forthcoming in Festschrift for Richard Stone, ed. Angus Deaton.

Holly, A. (1979), "Testing Recursiveness in a Triangular Simultaneous Equation Model", Warwick Economic Research Paper, No.154.

Jennrich, R.I. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimators", Annals of Mathematical Statistics, 40, pp.633-643.

Kmenta, J. (1967), "On Estimation of the CES Production Function", International Economic Review, 8, pp.180-189.

Konstas, Panos and Mohamad Khouja (1969), "The Keynesian Demand-for-Money Function: Another Look and Some Additional Evidence", Journal of Money, Credit and Banking, 1, pp.765-777.

Malinvaud, E. (1970), "The Consistency of Non-Linear Regressions", Annals of Mathematical Statistics, 41, pp.956-969

Pifer, Howard W. (1969), "A Nonlinear Maximum Likelihood Estimate of the Liquidity Trap", Econometrica, 37, pp.324-332.

Rao, C.R. (1948), "Large Sample Tests of Statistical Hypotheses concerning Several Parameters with Applications to Problems of Estimation", Proceedings of the Cambridge Philosophical Society, 44, pp.50-57.

Richard, J.F., R.F. Engle and D.F. Hendry (1979), "Exogeneity", manuscript.

Sargan, J.D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in Econometric Analysis for National Economic Planning (ed. P.E. Hart, G. Mills, and J.K. Whittaker) London: Butterworths Scientific Publications.

Sargan, J.D. (1975), "A Suggested Technique for Computing Approximations to Wald Criteria with Application to Testing Dynamic Specifications", unpublished manuscript, London School of Economics.

Savin, N.Eugene (1976), "Conflict Among Testing Procedures in a Linear Regression Model with Autoregressive Disturbances", Econometrica, 44, 1303-1313.

Schönfeld, P. (1971), "A Useful Central Limit Theorem for m-Dependent Random Variables", Metrika, 17, pp.116-128.

Silvey, D.S. (1959), "The Lagrangean Multiplier Test", Annals of Mathematical Statistics, 30, pp.389-407.

White, Kenneth J. (1972), "Estimation of the Liquidity Trap with a Generalized Function Form", Econometrica, 40, pp.193-199.

Wu, Demin (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances", Econometrica, 41, pp.733-750.