

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: <http://www.nber.org/books/kmen80-1>

Publication Date: 1980

Chapter Title: Hypothesis Testing in Spectral Regression; the Lagrange Multiplier Test as a Regression Diagnostic

Chapter Author: Robert F. Engle

Chapter URL: <http://www.nber.org/chapters/c11707>

Chapter pages in book: (p. 309 - 321)

Hypothesis Testing in Spectral Regression; the Lagrange Multiplier Test as a Regression Diagnostic

ROBERT F. ENGLE

DEPARTMENT OF ECONOMICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA

1. Introduction

Many tests of the adequacy of the specification of a linear regression have been proposed and used by econometricians. One of the commonly used tests is to subdivide the sample period and test that the parameters remain constant between the subperiods. A closely related test has recently been formulated in the frequency domain by Engle (1974). In this test, the regression is estimated on subsets of the spectrum so that, for example, a regression could be estimated using only the low-frequency components of the data. An exactly analogous test is available to determine whether or not the parameters are the same for different frequency bands. Both of these tests provide useful regression diagnostics, and they complement each other in that they are powerful against different types of alternatives. For example, the time domain test has little power against dynamic misspecification while the frequency domain test finds it easy to detect such failures.

The frequency domain test has been applied to consumption functions to test whether or not the permanent and transitory propensities are the same (Engle, 1974) and to price equations to examine the timing of output price responses to input price changes (Engle, 1978). However, these applications rely upon the assumption that the disturbances under the null

hypothesis be white noise and that there be no lagged dependent or jointly endogenous variables. Engle (1980) has generalized the test for lagged dependent variables, and this paper will examine the hypothesis testing problem with a general stationary disturbance process.

This paper shows that the test of equality of parameters across frequency bands is a linear hypothesis test. Likelihood ratio (LR), Wald (W), and Lagrange multiplier (LM) tests are then developed for general linear hypotheses in this context. Following Savin (1976), Berndt & Savin (1977), Breusch (1979), the test criteria are shown to satisfy a numerical inequality even though they have the same asymptotic distribution. It is suggested that the Lagrange multiplier statistic may be a particularly useful formulation for testing for model misspecification. Examples are presented which show the simplicity of this test.

2. Serial Correlation

Consider the model

$$y = x\beta + u, \quad E(u|x) = 0, \quad u \sim N(0, \sigma^2\Omega), \quad \text{tr}\Omega = T, \quad (1)$$

where x is a $T \times K$ matrix of observations, β is a $K \times 1$ vector of unknown parameters, u is a $T \times 1$ vector of disturbances, and y is a $T \times 1$ vector of observations on the dependent variable. Letting the matrix of Fourier coefficients be defined by

$$W_{jk} = (1/\sqrt{T})e^{i(2\pi j/T)k}, \quad j, k = 1, \dots, T,$$

the amplitudes of frequency components can be written as

$$\tilde{y} = Wy.$$

The square of the absolute value of the j th component of \tilde{y} is defined to be the periodogram of y , $I_y(\theta_j)$, at frequency θ_j , and the cross product $\tilde{x}_j^\dagger \tilde{y}_j$ is defined to be the cross periodogram at frequency θ_j between x and y where the dagger represents the complex conjugate of the transpose and $\theta_j = 2\pi j/T$. The W matrix is unitary and has the useful property that it will approximately diagonalize any covariance matrix. This approximation becomes exact if either Ω is defined as a circulant¹ or the sample size is

¹ A stationary stochastic process is circular if the correlation between u_t and u_{t+s} is the same as between u_{T+j} and u_{s+j} for any j where T is the end of the sample period. That is, the observations after the sample period are the same as those at the beginning.

very large. These relations can be written for the exact case as

$$W^{\dagger}W = WW^{\dagger} = I, \quad W\Omega W^{\dagger} = D = \text{diag} \{f_u(\theta_j)\},$$

$$\theta_j = \frac{2\pi j}{T}, \quad j = 1, \dots, T, \quad (2)$$

where D is the diagonal matrix with the spectrum of the disturbances at the harmonic frequencies on the diagonal. Throughout this paper (2) will be taken to be exactly true.

The log likelihood function for (1) can be written

$$L(u) = -\frac{T}{2} \log 2\pi\sigma^2 - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma^2} u^{\dagger} \Omega^{-1} u. \quad (3)$$

Using relations (2) it can be rewritten

$$L(\tilde{u}) = -\frac{T}{2} \log 2\pi\sigma^2 - \frac{1}{2} \sum_j \log f_u(\theta_j) - \frac{1}{2\sigma^2} \tilde{u}^{\dagger} D^{-1} \tilde{u}$$

$$= -\frac{T}{2} \log 2\pi\sigma^2 - \frac{1}{2} \sum_j \log f_u(\theta_j) - \frac{1}{2\sigma^2} \sum_j f_u^{-1}(\theta_j) I_u(\theta_j). \quad (4)$$

If the spectrum of the disturbances is known, then (4) can be maximized directly with respect to σ^2 and β noting from (1) that

$$I_u(\theta_j) = I_y(\theta_j) + \beta' I_{xx}(\theta_j) \beta - (\beta' I_{xy}(\theta_j) + I_{yx}(\theta_j) \beta). \quad (5)$$

This gives exactly the familiar generalized least squares (GLS) estimator

$$\hat{\beta}_{\text{GLS}} = (x' \Omega^{-1} x)^{-1} x' \Omega^{-1} y.$$

In general, the spectrum of the disturbance must also be estimated assuming only that it is locally smooth. A convenient assumption which turned out to be quite satisfactory in a Monte Carlo experiment (Engle & Gardner, 1976) is to assume that f_u is constant within narrow mutually exclusive bands B_h , $h = 1, \dots, H$ with m_h elements in each band so that $\sum_h m_h = T$. For asymptotic results H must grow with T but more slowly than T . Hannan (1970, p. 442) and Espasa & Sargan (1975) assume $m_h = m$ and $\lim T/H^2 = 0$ with x satisfying Grenander's conditions which essentially require that x be stationary or only mildly explosive. These will be assumed to hold for all asymptotic results but are not necessary for the finite sample properties.

Differentiating (4) with respect to $f_u(\theta_j)$ and using $T = \text{tr} \hat{\Omega} = \sum_j \hat{f}_u(\theta_j)$, the first-order conditions, which define maximum likelihood estimators, become

$$\hat{f}_u(\theta_j) = \ddot{I}_u(\theta_j) / \hat{\sigma}^2, \quad (6a)$$

$$\hat{\sigma}^2 = \frac{1}{T} u' \hat{\Omega}^{-1} u = \frac{1}{T} \sum_j \frac{I_u(\theta_j)}{\hat{f}_u(\theta_j)} = \frac{1}{T} \sum_j I_u(\theta_j) = \frac{u'u}{T}, \quad (6b)$$

$$\ddot{I}_u(\theta_j) = \frac{1}{m_h} \sum_{j \in B_h} I_u(\theta_j) \quad \text{for } j \in B_h, \quad h = 1, \dots, H, \quad (6c)$$

where u is now interpreted as a vector of residuals. The second line of (6b) is a consequence of the simple assumption made on the smoothness of the spectrum.

Substituting (6) into (4), the concentrated log likelihood is

$$L_c = -\frac{T}{2} (1 + \log 2\pi) - \frac{1}{2} \sum_j \log \ddot{I}_u(\theta_j). \quad (7)$$

Maximizing (7) with respect to β gives the first order conditions

$$\begin{aligned} \frac{\partial L_c}{\partial \beta} &= -\frac{1}{2} \sum_j \ddot{I}_u^{-1}(\theta_j) \frac{\partial \ddot{I}_u(\theta_j)}{\partial \beta} = 0 \\ &= -\frac{1}{2} \frac{\partial}{\partial \beta} (\tilde{y} - \tilde{x}\beta)' \hat{D}^{-1} (\tilde{y} - \tilde{x}\beta) = 0, \end{aligned}$$

where \hat{D} is a $T \times T$ matrix with diagonal elements $\ddot{I}_u(\theta_j)$.

The result is clearly

$$\hat{\beta} = (\tilde{x}' \hat{D}^{-1} \tilde{x})^{-1} \tilde{x}' \hat{D}^{-1} \tilde{y} = (x' \hat{\Omega}^{-1} x)^{-1} x' \hat{\Omega}^{-1} y, \quad (8)$$

where $\hat{\Omega}^{-1} = W' \hat{D}^{-1} W$. The maximum likelihood estimator must satisfy (6) and (8), and therefore an iterative procedure is usually followed. A check of the second-order conditions would be recommended to insure that the solution represents a maximum of the likelihood function.

This estimator is an approximate GLS estimator which will be asymptotically efficient under the conditions mentioned above. It is Hannan's (1963) estimator in a slightly different form as pointed out by Engle & Gardner (1976), and it is exactly maximum likelihood for finite samples. The variance matrix of $\hat{\beta}$ will be consistently estimated by

$$V(\hat{\beta}) = \hat{\sigma}^2 (x' \hat{\Omega}^{-1} x)^{-1} = \hat{\sigma}^2 \left[\sum_j \ddot{I}_u^{-1}(\theta_j) I_{xx}(\theta_j) \right]^{-1}. \quad (9)$$

3. Band Spectrum Regression

The derivation of the maximum likelihood estimator makes it easy to consider testing the hypothesis that the regression coefficients are constant for all frequencies. Let A be a diagonal $T \times T$ matrix with ones for all frequencies to be included and zeroes for those without ambiguity, let it also represent the set of included frequencies. These frequency bands will be assumed to be collections of B_h s. Instead of maximizing the likelihood function of u or \tilde{u} in (4), the likelihood of $A\tilde{u}$ is to be maximized to obtain an estimate of β_A for this subset of frequencies. The first-order conditions for this maximum are found in exactly the same fashion as for $A = I$ and are

$$\begin{aligned} \hat{\beta}_A &= (\tilde{x}^\dagger \hat{D}_A^{-1} \tilde{x})^{-1} (\tilde{x}^\dagger \hat{D}_A^{-1} \tilde{y}) \\ &= \left(\sum_{j \in A} \ddot{I}_{u_A}^{-1}(\theta_j) I_{xx}(\theta_j) \right)^{-1} \sum_{j \in A} \ddot{I}_{u_A}^{-1}(\theta_j) I_{xy}(\theta_j) \quad (10) \\ \hat{D}_A^{-1} &= \begin{cases} \text{diag } \ddot{I}_{u_A}^{-1}(\theta_j), & \text{for } j \text{ included by } A \\ 0 & \text{otherwise,} \end{cases} \\ \ddot{I}_{u_A}(\theta_j) &= \frac{1}{m_h} \sum_{j \in B_h} I_{u_A}(\theta_j), \quad \text{for } j \text{ included by } A, \\ \hat{\sigma}_A^2 &= \frac{1}{\text{rank } A} \sum_{j \in A} I_{u_A}(\theta_j). \end{aligned}$$

Versions of this estimator are easy to calculate and the procedure is available in the TROLL system. The variance matrix is consistently estimated by

$$V(\hat{\beta}_A) = \hat{\sigma}_A^2 (\tilde{x}^\dagger \hat{D}_A^{-1} \tilde{x})^{-1} = \hat{\sigma}_A^2 \left[\sum_{j \in A} \ddot{I}_u^{-1}(\theta_j) I_{xx}(\theta_j) \right]^{-1}.$$

Similar relations apply for β_{I-A} , the coefficients corresponding to frequencies excluded by A . For further details and an extension to lagged dependent variables see Engle (1980).

4. General Linear Hypothesis Test

The test of particular interest from the point of view of a general specification test is the test that $\beta_A = \beta_{I-A}$. That this is a general linear hypothesis can be seen by examining the regression in the frequency domain. Multiplying (1) by W gives

$$\tilde{y} = \tilde{x}\beta + \tilde{u},$$

where each observation is on a different frequency. Under the alternative that the coefficients are different, the model can be rewritten as

$$\begin{bmatrix} \tilde{y}_A \\ \tilde{y}_{I-A} \end{bmatrix} = \begin{bmatrix} \tilde{x}_A & 0 \\ 0 & \tilde{x}_{I-A} \end{bmatrix} \begin{bmatrix} \beta_A \\ \beta_{I-A} \end{bmatrix} + \begin{bmatrix} \tilde{u}_A \\ \tilde{u}_{I-A} \end{bmatrix}. \quad (11)$$

where \tilde{y}_A is the vector of observations on \tilde{y} included by A . Thus the null hypothesis is just a linear restriction on this regression model. Premultiplying now by W^\dagger , the inverse Fourier transform, will convert the data back to the time domain. Equation (11) then becomes

$$y = x_A \beta_A + x_{I-A} \beta_{I-A} + u,$$

where x_A and x_{I-A} are time domain data series after eliminating some components. Because they correspond to mutually exclusive and collectively exhaustive frequency bands, $x_A' x_{I-A} = 0$ and $x_A + x_{I-A} = x$. The test in the time domain is thus just a test of whether or not two sets of variables have the same coefficients. Because the variables are orthogonal, there are many equivalent ways to formulate the test.

One formulation of the model is

$$y = x\beta + x_A \beta_A + u \quad (12)$$

so that under the null hypothesis x_A does not belong in the regression. Clearly, if there are variables whose coefficients are not assumed to vary under the alternative, these can be included directly into (12).

This formulation makes clear the type of alternative models for which this test would have substantial power. The omitted variable x_A might be the high-frequency component of x and therefore would be highly correlated with series such as the first different in x . Therefore, the test should be powerful against misspecification of the dynamics of a model. These are exactly the types of alternatives that the time domain partition or Chow test would be ineffective at detecting if the relation between included and omitted variables remains stable between the two periods.

To develop tests of these hypotheses it is sufficient to consider the general linear restriction $R\beta = r$. While the theory and practice of testing linear restrictions is well known when the variance-covariance matrix of the disturbances is known, there are several possible asymptotically equivalent tests when the disturbance covariances must also be estimated. The difference between the procedures lies in the choice of an estimated covariance or distance matrix. Wald (1943) proposed estimating the disturbance matrix Ω jointly with the parameters under the alternative hypothesis and then comparing the parameter estimates with the linear restriction. The likelihood ratio method estimates the disturbance process under both the null and the

alternative hypothesis and forms the ratio of the likelihoods. The less well known Lagrange multiplier test, originally suggested by Rao (1947) and more recently proposed by Silvey (1959) and Aitchison & Silvey (1958), estimates Ω only under the null hypothesis. This procedure is closest to the spirit of a regression diagnostic. Only one regression need be run, and from manipulation of the output, it is possible to test whether or not the specification should be rejected in favor of a more general model.

5. A Comparison of Test Criteria

Three criteria will be developed for testing the hypothesis that $R\beta = r$ against the alternative $R\beta \neq r$. Under the alternative hypothesis, the maximum likelihood estimator will be given by the solution to equations (6) and (8) which will generally require an iterative procedure. This solution will be denoted

$$\hat{\beta}, \hat{\Omega}, \hat{\sigma}^2,$$

corresponding to log likelihood value L_c from equation (7) and residual periodogram \hat{I}_u .

Under the null hypothesis, the constraint $R\beta = r$ is assumed to hold. It can always be rewritten as $\beta = c + C\gamma$ where γ is a vector of unknown parameters with dimension $K - \text{rank}(R)$ when β has dimension K . The model (1) can therefore be rewritten as

$$\begin{aligned} y &= x\beta + u = xc + xC\gamma + u, \\ y_0 &= x_0\gamma + u, \end{aligned} \tag{13}$$

where $y_0 = y - xc$ and $x_0 = xC$.

The maximum likelihood estimators of (13) can be found in exactly the same fashion and will be denoted

$$\hat{\beta}_0, \hat{\Omega}_0, \hat{\sigma}_0^2,$$

where $\hat{\beta}_0 = c + C\hat{\gamma}$, and the value of the log likelihood and the smoothed periodogram of the residuals are L_{c_0} and \hat{I}_{u_0} , respectively.

The likelihood ratio λ is therefore given by

$$\log(\lambda) = L_{c_0} - L_c.$$

The asymptotic distribution of $\zeta_{LR} = -2 \log(\lambda)$ is well known to be $\chi^2_{\text{rank}(R)}$ under the null hypothesis, and therefore the likelihood ratio criterion will

be defined as

$$\xi_{LR} = \sum_j \log \dot{I}_{u_0}(\theta_j) / \dot{I}_u(\theta_j). \quad (14)$$

The Wald test criterion can be easily formulated. From a direct unrestricted regression, the parameter distribution is known to be asymptotically normal with variance given by (9). The distribution does not depend upon the method for estimating Ω , as long as a consistent method is used (see, for example, Maddala (1971)); however, in deriving the inequalities between criteria, it will be assumed that the maximum likelihood procedure is used. The test statistic

$$\xi_w = (R\hat{\beta} - r)' [R(x'\hat{\Omega}^{-1}x)^{-1}R']^{-1} (R\hat{\beta} - r) / \hat{\sigma}^2 \quad (15)$$

is asymptotically distributed as $\chi_{\text{rank}(R)}^2$ when H_0 is true.

In practice this statistic is commonly calculated by running both a restricted and unrestricted regression using the Ω matrix estimated under the alternative. Letting ${}_1\text{SSR}_0$ and ${}_1\text{SSR}_1$ be the sums of squared residuals under the null and alternative regressions, respectively, each evaluated using the distance metric from the alternative (e.g., ${}_1\text{SSR}_1 = (y - x\hat{\beta})\hat{\Omega}^{-1}(y - x\hat{\beta})$), the statistic can be written as

$$\xi_w = ({}_1\text{SSR}_0 - {}_1\text{SSR}_1) / ({}_1\text{SSR}_1 / T). \quad (16)$$

This relation is true for any distance metric and is shown in Maddala (1977).

The Lagrange multiplier procedure tests the hypothesis by maximizing the log likelihood subject to restrictions. The Lagrange multipliers associated with the restrictions indicate the cost or shadow price of the constraints, and one would therefore reject the null if the multipliers were too large.

Maximizing (4) subject to $R\beta = r$ will give the estimate of the restricted maximum likelihood estimator as well as the Lagrange multipliers. Savin (1976, p. 1307) gives the expression for the multipliers

$$\hat{\mu} = [R(x'\hat{\Omega}_0^{-1}x)^{-1}R']^{-1}(r - Rb), \quad (17)$$

where b is the estimate of the full, unrestricted parameter vector, β using the disturbance process estimated under the null hypothesis $\hat{\Omega}_0$. The test criterion is then shown to be

$$\xi_{LM} = \hat{\mu}' [R(x'\hat{\Omega}_0^{-1}x)^{-1}R']^{-1} \hat{\mu} / \hat{\sigma}_0^2, \quad (18)$$

which is asymptotically distributed as $\chi_{\text{rank}(R)}^2$ under the null hypothesis.

Substituting (17) into (18) gives

$$\xi_{LM} = (Rb - r)' [R(x'\hat{\Omega}_0^{-1}x)^{-1}R']^{-1} (Rb - r) / \hat{\sigma}_0^2. \quad (19)$$

This can therefore also be rewritten in terms of sums of squared residuals.

Letting ${}_0\text{SSR}_1$ and ${}_0\text{SSR}_0$ be the sums of squared residuals under the alternative and the null hypotheses, each using $\hat{\Omega}_0$ as the distance metric, the statistic can be written as

$$\xi_{\text{LM}} = ({}_0\text{SSR}_0 - {}_0\text{SSR}_1)/({}_0\text{SSR}_0/T). \tag{20}$$

One useful form of the statistic is obtained by rewriting ${}_0\text{SSR}_1$ in terms of the residuals from the fit under the null hypothesis, $e_0 = y_0 - x_0\gamma$. In this form the test appears to be a way of analyzing the residuals for particular types of nonrandomness. In this sense the test is in the same spirit as those of Ramsey (1969).

Let $M_0 = (I - x_0(x_0'\hat{\Omega}_0^{-1}x_0)^{-1}x_0'\hat{\Omega}_0^{-1})$ and $M = (I - x(x'\hat{\Omega}_0^{-1}x)^{-1}x'\hat{\Omega}_0^{-1})$. Because x_0 is a subspace of x , $M_0M = M$. Thus

$$\begin{aligned} {}_0\text{SSR}_1 &= (y - xb)'\hat{\Omega}_0^{-1}(y - xb) = y'M'\hat{\Omega}_0^{-1}My = y_0'M'\hat{\Omega}_0^{-1}My_0 \\ &= y_0'M_0M\hat{\Omega}_0^{-1}MM_0y_0 = e_0'M'\hat{\Omega}_0^{-1}Me_0 \end{aligned}$$

since $y_0 = y - xc$ and $Mx = 0$. This is therefore also the sum of squared residuals from a regression of e_0 on x using $\hat{\Omega}_0$ as the distance metric. Letting \hat{e}_0 be the fitted values of this regression the LM test is therefore simply

$$\xi_{\text{LM}} = T\mathcal{R}^2 = T(\hat{e}_0'\hat{\Omega}_0^{-1}\hat{e}_0/e_0'\hat{\Omega}_0^{-1}e_0) \tag{21}$$

where \mathcal{R}^2 is the multiple correlation coefficient of this regression.² The procedure therefore is to estimate the model and Ω under the null hypothesis and save the (untransformed) residuals. These are then regressed on the full set of x s under the alternative being considered, using the estimated Ω matrix. The \mathcal{R}^2 of this regression times the sample size is the LM statistic.

Based upon this model, Breusch (1979) extends Savin's (1976) Theorem 1, proving that the three test statistics developed for the general linear hypothesis test satisfy a numerical inequality

$$\xi_{\text{LM}} \leq \xi_{\text{LR}} \leq \xi_{\text{W}}. \tag{22}$$

Since all three have asymptotically the same distribution under the null hypothesis, all three will have the same critical values, and thus the tests may give conflicting inference. Statistical grounds on which to choose among the tests would be finite sample size and power calculations, not asymptotic properties. As these are not yet known, the main implication is that the Lagrange multiplier test is the least stringent so that rejection of this criterion would imply rejection by all three, and conversely, the Wald test is the most stringent.

² This assumes that the computer reports the \mathcal{R}^2 of the transformed regression. See Buse 1973 for a discussion of this point.

6. A Time Domain Example

As an example, consider an equation presented in Engle (1979) which explains employment in Boston's textile industry as a function of the U.S. demand and prices, the stock of fixed factors in Boston, and the Boston wage rate. The equation is a reduced form derived from a simple production model with capital as a fixed factor and a constant price elasticity of demand. The variables are specific combinations of logarithms of the original data. Denote the dependent variable by y , and the independent variables by x_1 , x_2 , and include an intercept. The hypothesis to be tested is whether or not a time trend should also be introduced to allow technical progress in the sector. There is substantial serial correlation in the disturbance and several methods of parameterizing it are given in the original paper; however, it will here be assumed to follow a first order autoregressive process. There are 22 annual observations.

The basic estimate of the relation is

$$\hat{y} = 4.4 + .165x_1 + .669x_2, \quad \rho = .901, \quad \mathcal{R}^2 = .339.$$

(1.92) (2.45) (3.11)

The estimate is not particularly good, but it has the right signs and significant t statistics (shown in parentheses). Rho was estimated by searching over the unit interval, and the estimate is generalized least squares.

The residuals from this estimate were then regressed upon the expanded set of regressors to obtain

$$\hat{\epsilon} = 49.2 - .185x_1 - .045x_2 - .025 \text{ time}; \quad \rho = .901, \quad \mathcal{R}^2 = .171.$$

(1.90) (-1.61) (-.22) (1.93)

The same value of rho was imposed upon this estimate. The Lagrange multiplier statistic is $(22)(.171) = 3.76$ which is slightly below the 95% level for χ_1^2 but above the 90% level so it rejects at 90% but not at 95%.

For comparison, the full regression was estimated, including a reoptimization of rho. The results were

$$\hat{y} = 59.9 - .05x_1 + .611x_2 - .028 \text{ time}; \quad \rho = .970, \quad \mathcal{R}^2 = .480$$

(2.26) (-.45) (3.18) (2.13)

The Wald test involves merely looking at the t -statistic on time; however, the asymptotic formulation would estimate the standard error using 22 degrees of freedom rather than 18. In this case the t statistic is -2.35 so the test rejects at 95% but not 99%.

In this example the two test statistics give conflicting inference at the 95% level with the Wald statistic rejecting the null hypothesis and the La-

grange multiplier statistic accepting. However, at both the 90% and the 99% level, they agree. The numerical results illustrate the algebraic inequality given above. The benefits from using the Lagrange multiplier test lie primarily in the avoidance of a recalculation of ρ . This may appear a rather minimal saving for the first-order autoregressive case; for models postulated to have autoregressive, moving-average (ARMA) disturbance processes or general stationary error processes, the iterative procedures may be substantially more expensive and may require model identification for each alternative. In the next section, more complicated error processes are considered.

7. A Frequency Domain Example

In this section frequency domain and time domain partitions are used to test for the stability of the model. In both cases the LM technique proves to be highly valuable in simplifying computational procedures.

For the time domain partition the same value of the first-order autoregression is used in estimating the model for the 1950s and for the 1960s. The expression (20) is used in evaluating the outcome of the test. The statistic is 6.33 which is a χ^2 with 3 degrees of freedom and is not significant at the 95% level since the critical value is 7.81.

For the band spectrum regression test the procedures described in Section 3 are followed. The spectrum of the residuals is estimated under the null hypothesis using the procedures of Engle & Gardner (1976). The spectrum is generally downward sloping but turns upward for higher frequencies.

This spectrum is then used to estimate the model for each frequency band. The bands are defined to be from 0 to .25 and from .25 to .5 so that oscillations with periods longer than 4 years are considered low frequency and tested against shorter periodicities.

The model was estimated in the form of (12) with the high-frequency parts of x_1 and x_2 included as additional variables (the high-frequency part of the constant is zero). Using the same spectrum the results are

$$\hat{y} = -27.6 + .51x_1 + 2.12x_2 - .31x_1^{\text{high}} - 1.57x_2^{\text{high}}, \quad R^2 = .58.$$

$$\begin{array}{cccccc} (-2.11) & (3.81) & (3.59) & (-2.01) & (-2.37) & \end{array}$$

The LM test statistic is 6.01 which is very slightly above the 95% point for a chi square with two degrees of freedom (5.99). It appears from the statistics that the dynamics of both variables could be improved. Since the LM test rejects, one can be certain that the likelihood ratio and Wald tests will reject by (21) and there is no need to reestimate the spectrum of the residuals of the model under the alternative hypothesis.

This example illustrates the computational saving involved in the use of the LM test when there are substantial costs to reestimating the disturbance covariance matrix. The result is particularly clear here since one would be very unlikely to cease the investigation by accepting the alternative model that the frequency bands differ. Thus it is unlikely that the residual spectrum for this situation would ever be of interest. The implication of the test is that the dynamics are possibly faulty. This is attractive from an economic point of view since one might expect short-run and long-run price elasticities to differ. In Engle (1979) a dynamic model is therefore estimated to improve this relationship and allow a difference in the price elasticities.

ACKNOWLEDGMENTS

The research for this paper was undertaken while the author was a Senior Research Fellow at the London School of Economics and was completed at the University of California at San Diego with the support of NSF SOC77-07166. The author is indebted to Ken Wallis, Clive Granger, and a referee for this volume for helpful suggestions.

REFERENCES

- Aitchison, J., & Silvey, S. D. Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, 1958, **29**, 813–828.
- Berndt, E. R., & Savin, N. E. Conflict among criteria for testing hypotheses in the multivariate Linear regression model. *Econometrica*, 1977, **45** (July), 1263–1278.
- Breusch, T. S. (1979), Conflict among criteria for testing hypotheses: Extensions and comments. *Econometrica*, 1979, **47** (January) 203–207.
- Buse, A. Goodness of fit in generalized least squares estimation. *The American Statistician*, 1973, **27**, 106–108.
- Engle, R. F. Band spectrum regression. *International Economics Review*, 1974, **15**, 1–11. (a)
- Engle, R. F. Testing price equations for stability across frequencies. *Econometrica*, 1978, **46** (July), 869–881.
- Engle, R. F. Estimation of the price elasticity of demand facing metropolitan producers. *Journal of Urban Economics*, 1979, **6**, 42–64.
- Engle, R. F., Exact maximum likelihood methods for dynamic regressions and band spectrum regressions. *International Economic Review*, June, 1980, in press.
- Engle, R. F., & Gardner, R. E. Some finite sample properties of spectral estimators of a linear regression. *Econometrica*, 1976, **44**, 149–165.
- Espasa, A. & Sargan, D. The spectral estimation of simultaneous equation systems with lagged endogenous variables. *International Economic Review*, 1975, **18**, 583–607.
- Hannan, E. J. Regression for Time Series. In M. Rosenblatt (Ed.), *Time series analysis*. New York: Wiley, 1963. Pp. 17–37.
- Hannan, E. J. *Multiple time series*, New York: Wiley, 1970.
- Maddala, G. S. Generalized least squares with an estimated variance covariance matrix. *Econometrica*, 1971, **39**, 23–33.
- Maddala, G. S. *Econometrics*, New York: McGraw-Hill, 1977.

- Ramsey, J. B. Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, Series B*, 1969, 31(2), 350-371.
- Rao, C. R. Large sample tests of the statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 1947, 44, 50-57.
- Savin, N. E., Conflict among testing procedures in a linear regression model with autoregressive disturbances. *Econometrica*, 1976, 44, 1303-1313.
- Silvey, S. D. The Lagrangian multiplier test. *Annals of Mathematical Statistics*, 1959, 30, 389-407.
- Wald, A. Tests of hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 1943, 54, 426-482.