The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note

A. BUSE*

By means of simple diagrams this note gives an intuitive account of the likelihood ratio, the Lagrange multiplier, and Wald test procedures. It is also demonstrated that if the log-likelihood function is quadratic then the three test statistics are numerically identical and have χ^2 distributions for all sample sizes under the null hypothesis.

KEY WORDS: Likelihood ratio; Lagrange multiplier test; Wald test; Quadratic log-likelihood; Chi-squared distribution.

The level of rigor that is desirable or attainable in the teaching of an econometrics course will depend on the purpose of the course and the level of mathematical and statistical sophistication of the students taking it. In the case of undergraduate introductory econometrics courses the level of mathematical and statistical rigor will necessarily be limited. At the graduate level, on the other hand, there is always a need for a nonspecialist course in which the orientation is primarily to applications and rigor must give way to the need for broad coverage of various techniques and their limitations. In either case, but particularly in graduate courses, the teacher of such a course has a responsibility to incorporate recent developments and present them in an accessible form. Yet it is precisely the most recent developments that are available only in formal, rigorous presentations at the journal level. Thus, there is a short-term need to bridge the gap between the journal literature and the distilled versions of that literature that ultimately appear in the textbooks. The purpose of this note is to build such a bridge to the emerging literature on the use of the Wald (W) and Lagrange multiplier (LM) statistics in econometrics. It should be added here that the emphasis on econometrics in this note reflects the teaching experience of the author and is not meant to exclude readers from other fields. We feel that the exposition is applicable to any other field where these tests are used and we invite the reader to substitute his particular field whenever the term econometrics is used. Furthermore, since there are many levels of aspiration for an introductory course in econometrics, this note will assume that the students have the necessary mathematical and statistical background needed for a course taught at the level of Kmenta (1971).

*A. Buse is Professor, Department of Economics, University of Alberta, Edmonton, Alberta, Canada T6G 2H4. This is a revised version of a note written in May 1980 at the University of New South Wales while the author held a Leave Fellowship from the SSHRC. The author is indebted for advice and criticism to Ron Bewley, David Giles, Adrian Pagan, and Eric Sowey.

Within the framework of maximum likelihood methods the basic logic of these tests is developed by means of a simple diagram, and the connection of these tests to the standard likelihood ratio (LR) test is also established by the same diagrammatic device. Using the insights generated by our diagrammatic method we show that the distribution of these test statistics is easily derived if the log-likelihood is quadratic. Our exposition can be viewed as the geometric complement to the heuristic presentation of these tests that has been given by Silvey (1970, pp. 108–122). It can also be considered as complementary to the recent survey articles by Breusch and Pagan (1980) and Engle (1981).

To introduce our approach to these tests we apply it first to the LR test and give that test a visual format that we think is novel and pedagogically helpful. Consider a classroom situation in which the discussion of maximum likelihood estimation has been completed and the instructor has just explained the basic logic of the LR test. That is to say, he will have advanced the argument that taking the ratio of likelihoods with and without the restrictions of the null hypothesis imposed is a plausible basis for a test. Furthermore, the basic equation stating the asymptotic distribution of the test statistic,

$$\mathrm{LR} = 2(\log L(\hat{\theta}) - \log L(\tilde{\theta})) \sim \chi^2(g), \tag{1}$$

will have made an appearance, where $\hat{\theta}$ is the unrestricted estimate of the population vector, $\tilde{\theta}$ is the restricted estimate, and g is the number of restrictions imposed by the null hypothesis. Suppose that the vector θ consists of only one element, so that $\tilde{\theta} = \theta_0$, where θ_0 is the value of θ specified in a simple null hypothesis and the alternative hypothesis specifies that $\theta \neq \theta_0$. If we now plot the log-likelihood function, then the value of ½LR can be read directly from Figure 1 by noting the values of $\log L(\theta)$ at $\hat{\theta}$ and θ_0.

In Figure 1 we note that the distance, ½LR, will depend on the distance, $\hat{\theta} - \theta_0$, and the curvature[1] of the log-likelihood function, which we denote by $C(\hat{\theta})$ and which is defined by the absolute value of $(d^2 \log L)/d\theta^2$ evaluated at $\theta = \hat{\theta}$. Given $C(\hat{\theta})$, the larger the distance between $\hat{\theta}$ and θ_0, the further will $\log L(\theta_0)$ be from the maximum, $\log L(\hat{\theta})$, and the larger will be the distance ½LR. Conversely, for given distance $\hat{\theta} - \theta_0$, the greater the curvature $C(\hat{\theta})$, the larger will be the distance ½LR. It is these characteristics that provide the key to a diagrammatic derivation of the Wald test. Instead of considering the differences in log-likelihoods, this test takes the intuitively appealing approach of working with the squared distance between $\hat{\theta}$ and θ_0. Large deviations of

[1]Our use of the notion of curvature is intended to be informal and intuitive. However, since the first derivative is zero at $\theta = \hat{\theta}$, $C(\hat{\theta})$ does in fact measure curvature with respect to arc length.

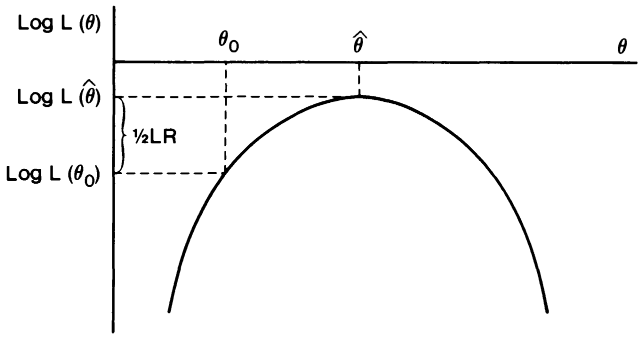

Figure 1. The Likelihood Ratio Test

$\hat{\theta}$ from θ_0 are taken as evidence that the data do not confirm the null hypothesis. However, given our earlier observation about curvature of the log-likelihood, the squared distance $(\hat{\theta} - \theta_0)^2$ must be weighted by $C(\hat{\theta})$ because two sets of data might produce the same value of $(\hat{\theta} - \theta_0)^2$, yet with one set less favorable to the hypothesis (from the perspective of the LR test) because the curvature of the log-likelihood is greater and the value of $\log L(\theta_0)$ is correspondingly much further away from the maximum. This point is illustrated in Figure 2, where both sets of data generate the same value of $(\hat{\theta} - \theta_0)^2$ but for case A the greater curvature produces a much smaller value of $\log L(\theta_0)$.

A Wald statistic can now be defined as

$$W = (\hat{\theta} - \theta_0)^2 C(\hat{\theta}), \tag{2}$$

which under H_0 is asymptotically distributed as χ^2 with one degree of freedom (as is the LR statistic). Large values of W lead to the rejection of the null hypothesis. Equation (2) is not the usual form of the Wald statistic, which is

$$W = (\hat{\theta} - \theta_0)^2 I(\hat{\theta}), \tag{3}$$

where $I(\theta) = E((d^2 \log L)/d\theta^2)$, the information matrix, so that the weighting of the squared distance is in terms of average curvature. The two statistics are, however, asymptotically equivalent because $C(\hat{\theta})$ is a consistent estimator of $I(\theta)$.

Having constructed the Wald statistic for the simple case, the student should now be less daunted by the general form given below. If $r(\theta) = 0$ is a vector of g

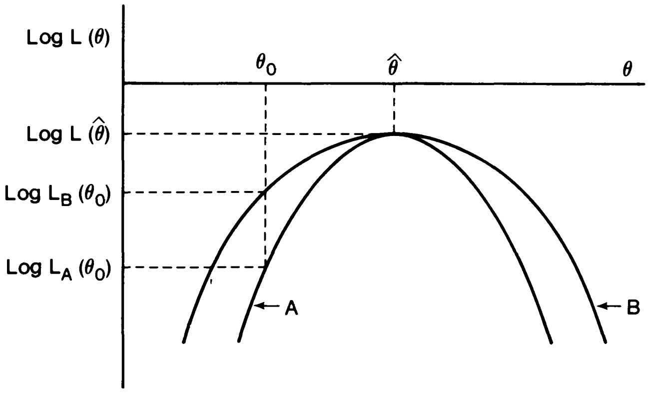

Figure 2. The Wald Test

functional restrictions imposed by H_0 on the k-vector θ $(k > g)$, then asymptotically

$$W = (r(\hat{\theta}))'[RI(\hat{\theta})^{-1}R']^{-1}[r(\hat{\theta})] \sim \chi^2(g), \tag{4}$$

where R is the $g \times k$ matrix of partial derivatives $\partial r(\theta)/\partial\theta$, evaluated at $\hat{\theta}$. Large values of W are generated by large deviations of $r(\hat{\theta})$ away from zero (the hypothesized value for the restrictions) and the deviations are weighted by a matrix involving the curvature of the log-likelihood. Equation (4) is a quadratic form in the vector $r(\hat{\theta})$ and it is clear that (3) is just a special case with $r(\theta)$ linear $(\theta - \theta_0 = 0)$ and R the identity matrix.

The Lagrange multiplier test also involves the curvature of the log-likelihood function but the basic idea behind the test focuses on the characteristics of the log-likelihood when the restrictions of the null hypothesis are imposed. Although the name of the test statistic derives from a restricted maximum likelihood estimation problem solved by the Lagrangian method, the basic idea of the test is seen most readily in the equivalent form of the "efficient score" statistic; see Breusch and Pagan (1980, p. 240). We will continue to use the term Lagrange multiplier test even though our account is based on the score formulation and despite the fact that the score form (Rao 1948) has historical precedent over the Lagrange form (Aitchison and Silvey 1958). The Lagrange terminology has become so firmly imbedded in the econometrics literature that to press for the score terminology would be futile. For those who work in fields in which the score terminology is more common we again invite them to make the appropriate substitution.

To proceed with the score approach we note that if the null hypothesis is true, the restricted maximum likelihood estimates will be near the unrestricted estimates. Since the unrestricted estimates maximize the log-likelihood, they satisfy the equation $S(\hat{\theta}) = 0$, where $S(\theta) = d\log L/d\theta$, and one can obtain a measure of the failure of the restricted estimates to reach the maximum by evaluating the extent of the departure of $S(\theta)$ from zero, where $\tilde{\theta}$ is again the restricted estimate. In our simple example $\tilde{\theta} = \theta_0$, and since we are indifferent to the sign of the slope, we can begin by suggesting $[S(\theta_0)]^2$ as a test statistic.[2] However, as in the case of the Wald statistic two data sets can generate the same slope of the log-likelihood, but one set has θ_0 closer to the maximum of the log-likelihood. To get around this difficulty we again weight the squared slope by the curvature, but in this case the greater the curvature the closer will θ_0 be to the maximum, so that if we weight by the inverse $C(\theta_0)^{-1}$, small values of the test statistic will be generated if θ_0 is close to the maximum. The point is illus-

[2]That the score and Lagrange test are equivalent can readily be seen by examining the constrained maximization problem associated with the Lagrange approach. Thus if $\phi = \log L(\theta) - \lambda(\theta - \theta_0)$, where λ is the Lagrange multiplier, then the first order conditions on the Lagrangian ϕ yield $S(\theta) = \lambda$ and $\tilde{\theta} = \theta_0$ and hence $S(\theta_0) = \lambda$. Thus large values of the slope are the same as large values of the Lagrange multiplier and both measure the cost of imposing the restriction.

trated in Figure 3, where the data set A has greater curvature at θ_0 and should therefore generate a smaller value of the test statistic because θ_0 is nearer the maximum of its log-likelihood.

$$\mathrm{LM} = [S(\theta_0)]^2 C(\theta_0)^{-1}, \tag{5}$$

which is distributed asymptotically as a χ^2 variable with one degree of freedom. As in the case of the W statistic the standard form for the LM statistic replaces the actual curvature by its expected value so that

$$\mathrm{LM} = [S(\theta_0)]^2 I(\theta_0)^{-1}. \tag{6}$$

The generalization to the vector version is a straightforward analog of (6) that specifies that asymptotically

$$\mathrm{LM} = S(\tilde{\theta})' I(\tilde{\theta})^{-1} S(\tilde{\theta}) \sim \chi^2(g) \tag{7}$$

It is clear that large values of LM leading to rejection of the null hypothesis will occur when the slope vector $S(\tilde{\theta})$ departs substantially from zero.

Looking back at Figures 1–3 we can note that the three tests differ in the kinds of information they require. The LR test requires both the restricted and unrestricted estimates of the parameter. Since the W test uses only the unrestricted estimates and the LM test uses only the restricted estimates, computationally the LR test is the most demanding. There is a direct physical analogy that can be made for each of these tests in terms of measuring the vertical distance between the top and a preassigned point on a hill. The LR test determines the distance by evaluating the height at both points. The W test works from the top of the hill and tries to establish the distance to a point lower down by looking at the horizontal distance between the top and the preassigned point and at the rate at which the hill curves away from the top. Finally, the LM test goes to the preassigned point and tries to determine how far it is to the top by considering the slope of the hill and the rate at which it is changing. While these physical analogies have limitations, they do provide a direct visual representation for students, and experience suggests that such analogies are an effective means of imparting the basic ideas. The careful teacher will, of course, point out these limitations and make suitable cautionary remarks. For example, one should note that implicit in

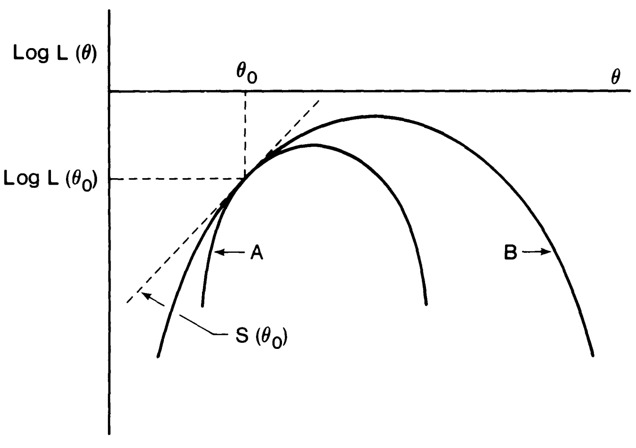

Figure 3. The Lagrange Multiplier Test

the diagrammatic presentation is the assumption that the argument is essentially a "local" one and in the case of log-likelihoods with multiple maxima one has to assume that the null is "sufficiently" close to the global maximum.

The hill-measuring analogy can be used to gain some additional insight into the distribution of the test statistics that, up to this point, have been asserted to be asymptotically χ^2. We now consider this question.

While we have drawn symmetric log-likelihoods as a matter of drafting convenience, symmetry is not essential to our visual argument. If in addition we assume that the log-likelihood is quadratic then we can exploit the hill-measuring analogy even further. In this case the W and LM methods must give the same numerical values as the LR method that measures the vertical distance directly. That this is so follows from the fact that three parameters are sufficient to determine a quadratic uniquely. In both the W and LM approaches we use three pieces of information to determine (implicitly) the parameters of the quadratic and hence the numerical value of the vertical distance. Thus in the W case we use the height of the hill at $\hat{\theta}$, $\log L(\hat{\theta})$, the value of the first derivative, $S(\hat{\theta}) = 0$, and second derivative, $C(\hat{\theta})$, whereas in the case of the LM we use the height of the hill at θ_0, $\log L(\theta_0)$, the value of the first derivative, $S(\theta_0)$, and second derivative, $C(\theta_0)$. Given that the numerical values of the statistics are the same, it follows that they have the same distribution. This does not, of course, say anything about the form of the distribution, but the quadratic likelihood can again be exploited for this purpose.

Consider the elementary textbook problem of testing the null hypothesis $\mu = \mu_0$ against $\mu \neq \mu_0$ for a sample size n drawn from a normal distribution with known variance, which for convenience we take as unity. Thus $X_i \sim N(\mu, 1)$ and

$$\log L(\mu) = -(n/2)(\log 2\pi) - \tfrac{1}{2}\Sigma(X_i - \mu)^2,$$

a quadratic in μ. Now

$$\frac{d\log L}{d\mu} = \Sigma(X_i - \mu) = n(\bar{X} - \mu) \tag{8}$$

$$\frac{d\log L}{d\mu^2} = -n \tag{9}$$

so that from (9) it is clear that the curvature is constant and this implies equality with the expected information. The maximum likelihood estimator is, of course, $\hat{\mu} = \bar{X}$, and a little algebra yields the LR statistic as

$$\begin{aligned} \mathrm{LR} &= 2[\log L(\hat{\mu}) - \log L(\mu_0)] \\ &= \Sigma(X_i - \mu_0)^2 - \Sigma(X_i - \bar{X})^2, \\ \mathrm{LR} &= n(\bar{X} - \mu_0)^2. \end{aligned} \tag{10}$$

The Wald statistic is given directly as

$$\mathrm{W} = (\hat{\mu} - \mu_0)^2 C(\hat{\mu}) = n(\bar{X} - \mu_0)^2, \tag{11}$$

and since from (8) $d\log L(\mu_0)/d\mu = n(\bar{X} - \mu_0)$ we have immediately that

$$LM = S(\mu_0)^2 C(\mu_0)^{-1} = n(\overline{X} - \mu_0)^2. \tag{12}$$

As predicted from our hill-measuring analogy, the values of all the test statistics are identical. Furthermore, since we know that $\overline{X} \sim N(\mu_0, n^{-1})$, each statistic is the square of a standardized normal variable and hence is distributed as a χ^2 variable with one degree of freedom. Thus in this particular example the test statistics are χ^2 for all sample sizes and therefore also asymptotically χ^2.

Our second example considers the multidimensional extension of the previous example by taking the standard econometric problem of testing for the validity of a set of linear restrictions. In the context of the linear model we assume that the variance-covariance matrix of the disturbances is known, and this reduces the problem to the quadratic case. Let $y = X\beta + u$ with $u \sim N(0, \Omega)$ where y is $n \times 1$, X is $n \times k$, β is $k \times 1$, and Ω is $n \times n$ and known. The null and alternative hypotheses are given by

$$\begin{aligned} H_0&: R\beta = r \\ H_1&: R\beta \neq r \end{aligned} \tag{13}$$

where R is $g \times k$ and r is $g \times 1$, both of which are known matrices.

The log-likelihood, a quadratic in the vector β, is given by

$$\begin{aligned} \log L(\beta) &= -(n/2)\log(2\pi) - ½\log|\Omega| \\ &\quad - ½(y - X\beta)'\Omega^{-1}(y - X\beta) \\ &= K - ½(y - X\beta)'\Omega^{-1}(y - X\beta), \end{aligned} \tag{14}$$

where the constant K is implicitly defined. From (14) we get

$$\frac{\partial \log L}{\partial \beta} = X'\Omega^{-1}y - X'\Omega^{-1}X\beta \tag{15}$$

and

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta'} = -(X'\Omega^{-1}X). \tag{16}$$

Given Ω, the Hessian (equation (16)) is constant and the "curvature" and the information matrix are again equal.

In order to construct the LR statistic we need both the unrestricted and the restricted estimates. The former is given directly from (15) as

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \tag{17}$$

whereas the latter can be obtained from the usual constrained maximization problem (see, e.g., Theil 1971, p. 285) as

$$\tilde{\beta} = \hat{\beta} - C^{-1}R'(RC^{-1}R^{-1})^{-1}(R\hat{\beta} - r) \tag{18}$$

with

$$C = (X'\Omega^{-1}X) \quad . \tag{19}$$

Defining $\hat{u} = y - X\hat{\beta}$ and $\tilde{u} = y - X\tilde{\beta}$ we can write the unrestricted log-likelihood as

$$\log L(\hat{\beta}) = K - ½\hat{u}'\Omega^{-1}\hat{u} \tag{20}$$

and the restricted log-likelihood as

$$\log L(\tilde{\beta}) = K - ½\tilde{u}'\Omega^{-1}\tilde{u} \quad . \tag{21}$$

The difference between (20) and (21) will be used to define the LR statistic but we first rewrite (21) by noting that

$$\tilde{u} = y - X\tilde{\beta} = \hat{u} - XC^{-1}R'(RC^{-1}R')^{-1}(R\hat{\beta} - r), \tag{22}$$

where we have substituted for $\tilde{\beta}$ from (18). We can therefore rewrite (21) as

$$\begin{aligned} \log L(\tilde{\beta}) &= K - ½\hat{u}'\Omega^{-1}\hat{u} \\ &\quad - ½(R\hat{\beta} - r)'(RC^{-1}R')^{-1}(R\hat{\beta} - r), \end{aligned} \tag{23}$$

where we have used the fact that $\hat{u}'\Omega^{-1}X = 0$. The LR statistic is now given as

$$\begin{aligned} LR &= 2[\log L(\hat{\beta}) - \log L(\tilde{\beta})] \\ &= (R\hat{\beta} - r)'(RC^{-1}R')^{-1}(R\hat{\beta} - r). \end{aligned} \tag{24}$$

The Wald statistic follows directly from (4) and is written as

$$W = (R\hat{\beta} - r)'(RC^{-1}R')^{-1}(R\hat{\beta} - r). \tag{25}$$

To obtain the LM statistic we note that from (15) we have

$$S(\tilde{\beta}) = R'(RC^{-1}R')^{-1}(R\hat{\beta} - r) \tag{26}$$

after we have substituted (18) for $\tilde{\beta}$. Thus,

$$\begin{aligned} LM &= S(\tilde{\beta})'C^{-1}S(\tilde{\beta}) \\ &= (R\hat{\beta} - r)'(RC^{-1}R^{1})^{-1}(R\hat{\beta} - r). \end{aligned} \tag{27}$$

We have therefore shown again that LR = W = LM. In fact, both our examples can be shown to be special cases of the general proposition that LR = W = LM for a test of linear restrictions if the log-likelihood is quadratic. The proof of this proposition is not included because it is straightforward and has the same structure as our second example. In fact, for the linear model the result on the equality of the statistics can be derived by a few judicious substitutions using the results in Breusch (1979). During the time that this paper was under review we received the paper by Engle (1981), who also proves this proposition and shows that it also holds when only a subset of the parameters are restricted by the null hypothesis. A generalization to non-linear restrictions on the vector β is possible (given Ω), but it requires that the restricted maximum likelihood estimate be obtained from the second iterate of the method of scoring. Furthermore, the first iterate must be the unrestricted estimator of β. This estimation method is equivalent to a minimum χ^2 approach in which the restrictions are linearized about the unrestricted estimates followed by restricted estimation using (18). (For more details, see Buse 1981).

The implications of the equality of the test statistics for tests of hypotheses are straightforward. Since $\hat{\beta} \sim N(\beta, C^{-1})$, each statistic is a quadratic form in the normal vector $R\hat{\beta} - r$ so that appeal to the relevant theorem gives us that each statistic is distributed as a

$\chi^2(g)$ variable, where g is the number of restrictions imposed by H_0. A less artificial case would require estimation of the variance-covariance matrix of the disturbances, which would destroy the equality between the statistics and the distributions would only be $\chi^2(g)$ asymptotically. In this case the linear model generates the systematic inequality $W \geq LR \geq LM$ first noted by Berndt and Savin (1977), which continues to attract attention in the econometric literature; see for example Evans and Savin (1980) and Fisher and McAleer (1980). Our results provide an interpretation of these inequalities in terms of the shape of the log-likelihood function. While the direction and order of the inequalities cannot be predicted from our geometrical arguments, it is clear the inequalities arise because of the departure of the log-likelihood from the quadratic shape. Whether one can formalize the extent of this departure and thus gain further insight into the genesis of the inequalities by taking higher-order terms in a Taylor series expansion of the log-likelihood is an open question to be pursued in future research.

Our approach to these test statistics leads to another research problem. The conventional format for the W and LM test (equations (3) and (6)) uses the information matrix rather than the Hessian of the log-likelihood (equations (2) and (5)). It is interesting to conjecture that the use of the information matrix rather than the Hessian could have a significant effect on the power of the LM and W tests in finite samples. Insofar as the LR test is based only on the data represented by the hill in Figure 1, the use of the information matrix in the W and LM tests is inappropriate because it does not represent the curvature of that particular hill of data. It is conceivable that the use of the information matrix rather than the Hessian is the cause of the very substantial underestimates of Type I error that have been observed in the use of the LM statistic for tests of heteroscedasticity; see, for example, Breusch and Pagan (1979) or Buse (1980). That using observed information (the Hessian) as opposed to expected information can have a substantial impact on the size of the Wald test is documented in Efron and Hinckley (1978). However, they dealt with the issue from the viewpoint of ancillary statistics and we plan a systematic investigation of this conjecture for the kinds of models typically used in econometrics.

Although these issues are not particularly appropriate for an expository note, it is gratifying to find that our relatively simple diagrammatic device leads to the research frontier. This, we believe, is in itself a good recommendation of its virtues and potential.

[*Received June 1981. Revised December 1981.*]

REFERENCES

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

BERNDT, E. R., and SAVIN, N. E. (1977), "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45, 1263–1278.

BREUSCH, T. S. (1979), "Conflict Among Criteria for Testing Hypotheses: Extension and Comment," *Econometrica*, 47, 203–208.

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

BREUSCH, T. S., and PAGAN, A. R. (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47, 239–254.

BUSE, A. (1980), "Tests for Additive Heteroskedasticity: Some Preliminary Monte Carlo Results," Paper presented to the Fourth World Congress of the Econometric Society, Aix-en-Provence.

——— (1981), "Pooling Time Series and Cross Section Data: A Minimum Chi-Squared Approach," *The Manchester School*, 49, 229–244.

EFRON, B., and HINKLEY, D. V., (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information," *Biometrika*, 65, 457–487.

ENGLE, R. F. (1981), "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics*, eds. Z. Griliches and M. Intriligator, Amsterdam: North Holland.

EVANS, G. B. A., and SAVIN, N. E. (1980), "Conflict Among the Criteria Revisited: The W, LR and LM Tests," Revised version of a paper presented to the Fourth World Congress of the Econometric Society, Aix-en-Provence.

FISHER, G., and McALEER, M. (1980), "Principles and Methods in the Testing of Alternative Models," Discussion Paper No. 400, Queen's University.

KMENTA, J. (1971), *Elements of Econometrics*, New York: Macmillan.

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, 50–57.

SILVEY, S. D., (1970), *Statistical Inference*, Harmondsworth: Penguin Books.

THEIL, H. (1971), *Principles of Econometrics*, New York: John Wiley.