Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters

Author(s): Christian Gouriéroux, Alberto Holly and Alain Monfort

Source: *Econometrica*, Jan., 1982, Vol. 50, No. 1 (Jan., 1982), pp. 63-80

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/1912529

Econometrica, Vol. 50, No. 1 (January, 1982)

LIKELIHOOD RATIO TEST, WALD TEST, AND KUHN-TUCKER TEST IN LINEAR MODELS WITH INEQUALITY CONSTRAINTS ON THE REGRESSION PARAMETERS

BY CHRISTIAN GOURIÉROUX, ALBERTO HOLLY, AND ALAIN MONFORT

This paper considers the problem of testing statistical hypotheses in linear regression models with inequality constraints on the regression coefficients. The Kuhn-Tucker multiplier test statistic is defined and its relationships with the likelihood ratio test and the Wald test are examined. It is shown, in particular, that these relationships are the same as in the equality constrained case. It is emphasized, however, that their common asymptotic distribution is a mixture of chi-square distributions under the null hypothesis.

1. INTRODUCTION

THE PURPOSE OF THIS PAPER is to extend the usual asymptotically equivalent procedures for testing whether the regression coefficients of linear regression models satisfy equality constraints. These tests are the likelihood ratio test, the Wald test [**23**], and the Lagrange multiplier test [**1**, **22**], also known as Rao's efficient score test [**19**].

We consider in this paper the linear regression model

$$y = X\beta + u$$

where β is subject to the inequality constraint[1]

$$R\beta \geqq r$$

and we consider procedures for testing the null hypothesis $R\beta = r$. These test procedures rely on constrained and unconstrained estimators or on Lagrange multipliers and Kuhn-Tucker multipliers. These estimators and multipliers are defined in Section 2.

We consider in Section 3 the likelihood ratio test, the Wald test, and the Kuhn-Tucker multiplier test when the covariance matrix of the disturbances of the model is known. The main new development contained in Section 3 is the definition of the Kuhn-Tucker multiplier test which is a natural extension of the Lagrange multiplier test when inequality constraints are considered. It is shown that these three test statistics are equal. In Section 4 we show that the distribution of these test statistics under the null hypothesis is a mixture of chi-square distributions.

In Section 5 we consider the case where Ω depends on a finite number of parameters. In that case we prove that the result of Section 4 remains valid asymptotically. It is also shown that the inequalities between the likelihood ratio test, the Wald test, and the Lagrange multiplier obtained by Berndt and Savin [**6**]

[1] If x and y are vectors of the same dimension then the notation $x \geqq y$ means that $x_i \geqq y_i$ for every i.

(see also Savin **[20]** and Breusch **[7]**) still hold for the three test statistics considered in this paper. Finally we establish that, as in the equality constraint case (see Breusch and Pagan **[8]** and Engle **[11]**), we may evaluate the Kuhn-Tucker test statistic as a determination coefficient in the context of a two-step procedure.

2. THE DIFFERENT ESTIMATORS AND MULTIPLIERS

Consider the linear model

$$y = X\beta + u$$

where y is a $(n \times 1)$ vector, X a $(n \times K)$ matrix, and β a $(K \times 1)$ vector. We assume that u is an n-dimensional normal vector $\mathcal{N}(0, \Omega)$ where Ω is known and nonsingular.

Let R be a known $(p \times K)$ matrix of rank p and r a known $(p \times 1)$ vector. β may have to satisfy either the inequality constraint $R\beta \geqq r$ or the equality constraint $R\beta = r$. It also may be an unconstrained vector. In this section we define three estimators associated with these three situations.

In order to state the optimization problems in a standard form, we assume that the inequality constrained estimator $\tilde{\beta}$ is the solution of

$$\begin{aligned} &\max -(y - X\beta)'\Omega^{-1}(y - X\beta) \\ &\text{subject to} \qquad R\beta \geqq r. \end{aligned}$$

This is a quadratic programming problem which can be solved by several algorithms (see Judge and Takayama **[13]** and Liew **[15]**). The Kuhn and Tucker (K.T.) vector of multipliers associated with the constraint $R\beta \geqq r$ is denoted by $\tilde{\lambda}$.

The equality constrained estimator will be denoted by $\hat{\beta}_0$. The vector of Lagrange multipliers (L.M.) associated with the constraint $R\beta = r$ is denoted by $\hat{\lambda}_0$. The unconstrained estimator is denoted by $\hat{\beta}$. As a convention, we may associate with $\hat{\beta}$ a LM vector $\hat{\lambda}$ which is equal to zero.

Under these conditions we may verify that the three estimators of β satisfy the following condition:

$$\frac{\partial}{\partial \beta}\left[-(y - X\beta)'\Omega^{-1}(y - X\beta) + \lambda'(R\beta - r)\right] = 0.$$

This condition implies that

$$\tilde{\beta} = \hat{\beta} + (X'\Omega^{-1}X)^{-1}R'\tilde{\lambda}/2$$

and

$$\hat{\beta}_0 = \hat{\beta} + (X'\Omega^{-1}X)^{-1}R'\hat{\lambda}_0/2.$$

From these equations we have

$$(1) \qquad \tilde{\beta} - \hat{\beta}_0 = (X'\Omega^{-1}X)^{-1} R'(\tilde{\lambda} - \hat{\lambda}_0)/2$$

which will be extensively used in the following Section.

3. THREE EQUIVALENT TEST STATISTICS WHEN Ω IS A KNOWN MATRIX

In this section we assume that the regression model is specified as

$$y = X\beta + u,$$

$$R\beta \geqq r,$$

u is a $(K \times 1)$ random vector $\mathcal{N}(0, \Omega)$.

In order to test the null hypothesis,

$$R\beta = r,$$

we may use one of the following test statistics.

The likelihood ratio test statistic (LR) is defined, as usual, by

$$\xi_{\mathrm{LR}} = -2 \log \Lambda = 2(\tilde{L} - \hat{L}_0)$$

where $\tilde{L}$ and $\hat{L}_0$ are respectively the maximum value of the logarithm of the likelihood under the maintained hypothesis $H : \{R\beta \geqq r\}$ and the null hypothesis $H_0 : \{R\beta = r\}$. It may be verified that

$$\xi_{\mathrm{LR}} = -(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta}) + (y - X\hat{\beta}_0)'\Omega^{-1}(y - X\hat{\beta}_0).$$

It is important to note that ξ_{LR} is the optimum value of the objective function of the following maximization problem.

$$(\mathrm{P}) \qquad \begin{array}{l} \max -(y - X\beta)'\Omega^{-1}(y - X\beta) + (y - X\hat{\beta}_0)'\Omega^{-1}(y - X\hat{\beta}_0) \\ \text{subject to} \qquad R\beta \geqq r. \end{array}$$

(P) is a primal problem which is quadratic in β. From general results on quadratic optimization under inequality constraints (see Ekeland [**12**]) we may see that the dual problem of (P) is

$$(\mathrm{D}) \qquad \begin{array}{l} \min (\lambda - \hat{\lambda}_0)'R(X'\Omega^{-1}X)^{-1}R'(\lambda - \hat{\lambda}_0)/4 \\ \text{subject to} \qquad \lambda \leqq 0. \end{array}$$

We define the Kuhn-Tucker test statistic ξ_{KT} as the optimum value of (D), that is

$$\xi_{\mathrm{KT}} = (\tilde{\lambda} - \hat{\lambda}_0)'R(X'\Omega^{-1}X)^{-1}R'(\tilde{\lambda} - \hat{\lambda}_0)/4.$$

Since the objective functions of (P) and (D) evaluated at the optimum are equal, we have

$$\xi_{\mathrm{LR}} = \xi_{\mathrm{KT}}.$$

This last equality may be sufficient to justify the definition of the Kuhn-Tucker test statistic ξ_{KT}. This statistic may however be justified on intuitive grounds by using arguments similar to those in Silvey [**22**] for the Lagrange multiplier test.

In effect, let us assume that $(y - X\beta)'\Omega^{-1}(y - X\beta)/T$ converges in probability (or almost surely) to a function of β and β_* where β_* is the true value of β. Also assume that β_* is the unique value of β which maximizes this limit function under the equality constraint. Let $\tilde{\lambda}^\infty$ and $\hat{\lambda}_0^\infty$ be, respectively, the K-T and LM multipliers associated with the inequality constrained and equality constrained limit problem. Under the null hypothesis $H_0: \{R\beta_* = r\}$ we have $\tilde{\lambda}^\infty = \hat{\lambda}_0^\infty$. As a consequence ξ_{KT} may be interpreted as a distance between $\tilde{\lambda} - \hat{\lambda}_0$ and 0, which is the value of $\tilde{\lambda}^\infty - \hat{\lambda}_0^\infty$ under the null hypothesis; therefore this statistic is appropriate for testing H_0.

Finally the Wald test statistic is defined as usual by

$$\xi_{\mathrm{W}} = (R\tilde{\beta} - r)'\left[R(X'\Omega^{-1}X)^{-1}R'\right]^{-1}(R\tilde{\beta} - r).$$

Since $R\hat{\beta}_0 = r$, we have, according to (1),

$$R\tilde{\beta} - r = R(\tilde{\beta} - \hat{\beta}_0) = \left[R(X'\Omega^{-1}X)^{-1}R'\right](\tilde{\lambda} - \hat{\lambda}_0)/2$$

and hence, we also have

$$\xi_{\mathrm{W}} = \xi_{\mathrm{KT}}.$$

We have defined three test statistics which are equal when the disturbances have a normal distribution with a known covariance matrix Ω. Obviously they have the same distribution which will be described in detail in the next section.

4. THE DISTRIBUTION OF THE THREE TEST STATISTICS WHEN Ω IS KNOWN

Before we turn to the discussion of the distribution of the likelihood ratio test statistic, it is of some interest to observe that the problem of particular concern bears a strong analogy with a problem which has been extensively studied in the statistical literature. To this purpose we first observe that the likelihood ratio test statistic for the regression model as specified in Section 3 is identical with the likelihood ratio test statistic for the model specified as

$$\hat{\beta} = \beta + v,$$
$$R\beta \geqq r,$$

(2) v is a $K \times 1$ random vector $\mathcal{N}\left(0, (X'\Omega^{-1}X)^{-1}\right)$ where $\hat{\beta}$ is the unconstrained M.L. estimator of β, i.e., the G.L.S. estimator.

This may be proved by using sufficiency arguments. It is simple, however, to use a more direct approach. Writing

$$y - X\beta = (y - X\hat{\beta}) - X(\beta - \hat{\beta})$$

and using the normal equations gives

$$(y - X\beta)'\Omega^{-1}(y - X\beta) = (y - X\hat{\beta})'\Omega^{-1}(y - X\hat{\beta}) + (\beta - \hat{\beta})'(X'\Omega^{-1}X)(\beta - \hat{\beta}).$$

This result may be used to show that ξ_{LR} is the optimum value of the objective function of the maximization problem

$$\max -(\beta - \hat{\beta})'(X'\Omega^{-1}X)(\beta - \hat{\beta}) + (\hat{\beta}_0 - \hat{\beta})'(X'\Omega^{-1}X)(\hat{\beta}_0 - \hat{\beta})$$

$$\text{subject to} \quad R\beta \geqq r$$

which is precisely the maximization problem associated with the specification given in (2).

Having established this, we next observe that the problem considered in this paper is related to the following one-sided testing problem in multivariate analysis. Suppose one obtains independent observations from a p-dimensional normal distribution with mean vector μ and variance matrix Σ. The problem of testing $H_0: \mu = 0$ against the restricted alternative $H_1: \mu_i \geqq 0$, $i = 1, \ldots, p$ with at least one inequality strict, has been studied by several authors. In particular Bartholomew [**3**, **4**, **5**], Chacko [**9**], Kudô [**14**], Nüesch [**16**, **17**], Perlman [**18**], and Shorack [**21**] derived the likelihood ratio test statistic (see also Barlow, Bartholomew, Bremner and Brunk [**2**] and the references contained therein). They showed that, under the null hypothesis, this test statistic is distributed as a mixture of chi-squared distributions of the form

$$\sum_{i=0}^{p} w(p,i)\chi^2(i)$$

where $\chi^2(0)$ is the unit mass at the origin. An alternative proof, based on geometrical reasoning, is given in the Appendix. To illustrate the general theory, the special cases of one and two constraints are examined in the next section.

Before we turn to these particular cases a comment on the power of these test procedures is in order. In fact, intuition suggests that a test procedure derived by a method which takes into account the *a priori* constraints would have better power characteristics than test procedures which do not take them into account. This is particularly clear in the case of the single constraint as explained in the next section. In principle, the power function of the three procedures can be found by methods similar to those used to find their distribution under the null hypothesis. The key difficulty, however, is the fact that the weights will also depend on the alternative hypothesis. The complexity of the analysis is such that the results on power are limited in scope.

It is worth mentioning, however, that Bartholomew [**5**] and Nüesch [**16**] have studied the power function of the likelihood ratio test for the above mentioned problem of testing the mean of a multivariate normal distribution. (See also Barlow, Bartholomew, Bremner, and Brunk [**2**].) The computations of these authors show that this procedure has substantially higher power than the conventional L.R. test used for testing $\mu = 0$ against the unconstrained alternative $\mu \neq 0$. It seems reasonable to conjecture that similar results hold for the case we are concerned with here.

4.1 *One Inequality Constraint*

In this case R is a $(1 \times K)$ row vector and ξ_W is such that:

$$\xi_W = \frac{(R\tilde{\beta} - r)^2}{R(X'\Omega^{-1}X)^{-1}R'} = \frac{(R\tilde{\beta} - r)^2}{\operatorname{var}(R\hat{\beta})}.$$

Moreover

$$R\tilde{\beta} = \begin{cases} R\hat{\beta} & \text{if } R\hat{\beta} > r, \\ r & \text{otherwise;} \end{cases}$$

therefore, if we denote

$$\frac{R\hat{\beta} - r}{(\operatorname{var} R\hat{\beta})^{1/2}}$$

by t, we have

$$\xi_W = \begin{cases} t^2 & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Under the null hypothesis $H_0:(R\beta = r)$, the statistic t has a standard normal distribution and the distribution of ξ_W is $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$. The critical region at level α is of the form $\{\xi_W > c\}$, where c is defined by $P(\xi_W > c) = \alpha$; c is easily obtained by noting that $P(Z > c) = 2\alpha$, where Z has a $\chi^2(1)$ distribution. The critical region can also be written $\{t > \sqrt{c}\}$. This shows that the above test is simply the one sided U.M.P. test. Note that if the standard theory were used, the critical region would be $\{\xi_W > c'\}$ where c' is such that $P(Z > c') = \alpha$; since c' is greater than c this procedure would lead to acceptance of the null hypothesis too often.

4.2. *Two Inequality Constraints*

In order to easily derive the distribution of the statistics it is useful to transform the initial model. We first write it in the form given in (2). Next, by applying an appropriate invertible affine transformation, we may write the initial specification as

$$\hat{\gamma} = \gamma + w,$$

$$\gamma_1 \geqq 0,$$

$$\gamma_2 - \theta\gamma_1 \geqq 0,$$

w is a $K \times 1$ random vector $\mathcal{N}(0, I_K)$.

The null hypothesis is transformed into

$$H_0 = \{\gamma_1 = 0, \gamma_2 - \theta\gamma_1 = 0\}$$

which is equivalent to $\{\gamma_1 = 0, \gamma_2 = 0\}$. In addition, the statistic ξ_{LR} remains unchanged, and we can write

$$\xi_{LR} = \|\hat{\gamma} - \hat{\gamma}_0\|^2 - \|\hat{\gamma} - \tilde{\gamma}\|^2$$

where $\hat{\gamma}_0$ and $\tilde{\gamma}$ are the M.L. estimates of γ under H_0 and under the maintained hypothesis respectively. Note that $\hat{\gamma}_0$ is the orthogonal projection of $\tilde{\gamma}$ on the subspace corresponding to H_0, and, therefore:

$$\xi_{LR} = \|\tilde{\gamma} - \hat{\gamma}_0\|^2 = \tilde{\gamma}_1^2 + \tilde{\gamma}_2^2,$$

where $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are the first two coordinates of $\tilde{\gamma}$.

The point $(\tilde{\gamma}_1, \tilde{\gamma}_2)$ is the orthogonal projection of $(\hat{\gamma}_1, \hat{\gamma}_2)$ on the cone defined by $\gamma_1 \geqq 0, \gamma_2 - \theta\gamma_1 \geqq 0$, as Figure 1 illustrates.

$$\begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} \quad \text{if} \quad \hat{\gamma}_1 \geqq 0,\ \hat{\gamma}_2 - \theta\hat{\gamma}_1 \geqq 0;$$

$$\begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\gamma}_2 \end{pmatrix} \quad \text{if} \quad \hat{\gamma}_1 \leqq 0,\ \hat{\gamma}_2 \geqq 0;$$

$$\begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{if} \quad \hat{\gamma}_2 \leqq 0,\ \theta\hat{\gamma}_2 + \hat{\gamma}_1 \leqq 0;$$

$$\begin{pmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{pmatrix} = \frac{\theta\hat{\gamma}_2 + \hat{\gamma}_1}{1 + \theta^2} \begin{pmatrix} 1 \\ \theta \end{pmatrix} \quad \text{if} \quad \theta\hat{\gamma}_2 + \hat{\gamma}_1 \geqq 0,\ \hat{\gamma}_2 - \theta\hat{\gamma}_1 \leqq 0.$$

The statistic ξ_{LR} has four different forms:

$$\xi_{LR} = \hat{\gamma}_1^2 + \hat{\gamma}_2^2, \quad \text{if} \quad \hat{\gamma}_1 \geqq 0,\ \hat{\gamma}_2 - \theta\hat{\gamma}_1 \geqq 0;$$

$$\xi_{LR} = \hat{\gamma}_2^2, \quad \text{if} \quad \hat{\gamma}_1 \leqq 0,\ \hat{\gamma}_2 \geqq 0;$$

$$\xi_{LR} = 0, \quad \text{if} \quad \hat{\gamma}_2 \leqq 0,\ \theta\hat{\gamma}_2 + \hat{\gamma}_1 \leqq 0;$$

$$\xi_{LR} = \frac{(\theta\hat{\gamma}_2 + \hat{\gamma}_1)^2}{1 + \theta^2}, \quad \text{if} \quad \theta\hat{\gamma}_2 + \hat{\gamma}_1 \geqq 0,\ \hat{\gamma}_2 - \theta\hat{\gamma}_1 \leqq 0.$$

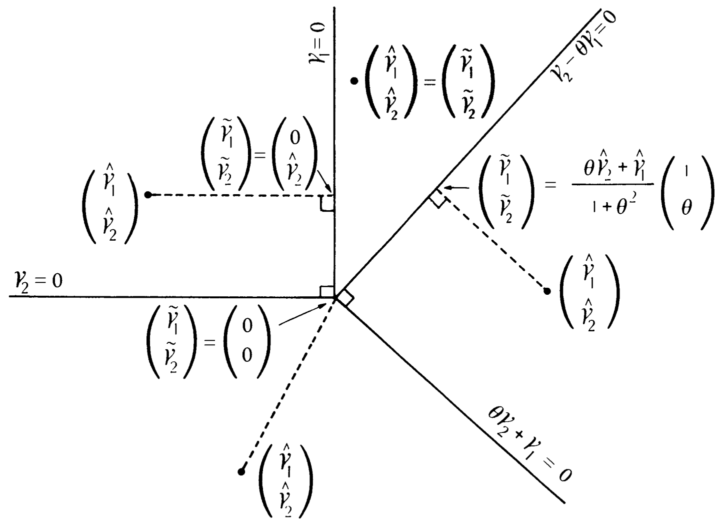

FIGURE 1

The distribution of ξ_{LR} under $H_0(\gamma_1 = \gamma_2 = 0)$ is given by:

$$P[\xi_{LR} = 0] = P[\hat\gamma_2 \leqq 0; \theta\hat\gamma_2 + \hat\gamma_1 \leqq 0] = q \qquad (0 \leqq q \leqq \tfrac{1}{2})$$

and

$$P[0 < \xi_{LR} < x] = P_1 + P_2 + P_3$$

where

$$P_1 = P[0 < \hat\gamma_1^2 + \hat\gamma_2^2 < x; \hat\gamma_1 \geqq 0; \hat\gamma_2 - \theta\hat\gamma_1 \geqq 0],$$

$$P_2 = P[0 < \hat\gamma_2^2 < x; \hat\gamma_1 \leqq 0; \hat\gamma_2 \geqq 0],$$

$$P_3 = P\left[0 < \frac{(\theta\hat\gamma_2 + \hat\gamma_1)^2}{1+\theta^2} < x; \theta\hat\gamma_2 + \hat\gamma_1 \geqq 0; \hat\gamma_2 - \theta\hat\gamma_1 \leqq 0\right].$$

Let us derive the values of P_1, P_2, and P_3:

$$P_1 = P[0 < \hat\gamma_1^2 + \hat\gamma_2^2 < x]P[\hat\gamma_1 \geqq 0; \hat\gamma_2 - \theta\hat\gamma_1 \geqq 0]$$

since $\hat\gamma_1^2 + \hat\gamma_2^2$ and the angle between the vectors

$$\begin{pmatrix}\hat\gamma_1\\ \hat\gamma_2\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}1\\ 0\end{pmatrix}$$

are independent under H_0; therefore

$$P_1 = (\tfrac{1}{2} - q)\phi_2(x)$$

where ϕ_2 is the c.d.f. of the $\chi^2(2)$ distribution.

$$P_2 = P\left[0 < \hat{\gamma}_2^2 < x; \hat{\gamma}_2 \geqq 0\right] P\left[\hat{\gamma}_1 \leqq 0\right]$$

since, under H_0, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are independent, or

$$P_2 = \tfrac{1}{2} P\left[0 < \hat{\gamma}_2^2 < x; \hat{\gamma}_2 \geqq 0\right]$$
$$= \tfrac{1}{4} P\left[0 < \hat{\gamma}_2^2 < x\right]$$

since, under H_0, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ have symmetrical distributions; therefore

$$P_2 = \tfrac{1}{4}\phi_1(x)$$

where ϕ_1 is the c.d.f. of the $\chi^2(1)$ distribution. The computation of P_3 is similar to that of P_2 because of the independence of $\theta\hat{\gamma}_2 + \hat{\gamma}_1$ and $\hat{\gamma}_2 - \theta\hat{\gamma}_1$:

$$P_3 = \tfrac{1}{4}\phi_1(x).$$

Finally the distribution of ξ_{LR}, under H_0, is:

$$w(2,0)\chi^2(0) + w(2,1)\chi^2(1) + w(2,2)\chi^2(2)$$

where $w(2,0) = q$; $w(2,1) = \frac{1}{2}$; $w(2,2) = \frac{1}{2} - q$. q is equal to $(1/2\pi)\cdot \cos^{-1}[-\theta/(1+\theta^2)^{1/2}]$, and its expression in terms of the original data is

$$\frac{1}{2\pi}\cos^{-1}\left\{ R_1(X'\Omega^{-1}X)^{-1}R_2' \Big/ \left[R_1(X'\Omega^{-1}X)^{-1}R_1'\right]^{1/2} \times \left[R_2(X'\Omega^{-1}X)^{-1}R_2'\right]^{1/2}\right\}$$

where $R_j (j = 1,2)$ is the jth row of R.

Note that q is also equal to $(1/2\pi)\cos^{-1}\rho(R_1\hat{\beta}, R_2\hat{\beta})$, where $\rho(R_1\hat{\beta}, R_2\hat{\beta})$ is the correlation coefficient between $R_1\hat{\beta}$ and $R_2\hat{\beta}$. The critical region at level α has the form $\{\xi_{\text{LR}} > c\}$, where c is given by $P[\xi_{\text{LR}} > c] = \alpha$. Therefore c is the solution of: $q + \frac{1}{2}\Phi_1(c) + (\frac{1}{2} - q)\Phi_2(c) = 1 - \alpha$ (with $1 - \alpha \geqq q$).

Let us briefly mention how the value of c may be obtained. We first observe that

$$q + (1-q)\Phi_2(c) < q + \tfrac{1}{2}\Phi_1(c) + (\tfrac{1}{2} - q)\Phi_2(c) < q + (1-q)\Phi_1(c).$$

Therefore c belongs to the interval $[c_1, c_2]$, where c_j is equal to

$$\Phi_j^{-1}\left(1 - \frac{\alpha}{1-q}\right) \qquad (j = 1,2).$$

Next, for given values of q and α, the value of c is obtained by a search procedure in the interval $[c_1, c_2]$.

Table 1 provides the critical values c for some values of q and for levels $\alpha = 0.05$ and $\alpha = 0.10$. Note that the standard theory would lead to the critical

TABLE I

$\alpha \backslash q$	0.0	0.1	0.2	0.3	0.4	0.5
5%	5.2	4.8	4.4	4.0	3.4	2.7
10%	3.8	3.5	3.1	2.8	2.2	1.8

region $\{\xi_{LR} > c'\}$ where c' is given by $P(Z > c') = \alpha$, $Z \sim \chi^2(2)$. Since c' is greater than c, the region of acceptance would be too large; for instance, $c' = 5.99$ for $\alpha =$ 5 per cent and $c' = 4.60$ for $\alpha = 10$ per cent (compare with the values of the table).

5. THE THREE TEST STATISTICS WHEN Ω DEPENDS ON A FINITE NUMBER OF UNKNOWN PARAMETERS

Typically the Ω matrix is unknown and its elements are continuous functions of a finite number of parameters which are not related to β. In this Section we assume that $\Omega = \sigma^2 Q$ where σ^2 is also unknown. Moreover, we assume that the maximum likelihood estimators of σ^2, Q, and β under H_0 and under H exist and are unique; they are respectively denoted by $\hat{\sigma}_0^2$, $\hat{Q}_0$, $\hat{\beta}_0$, and $\tilde{\sigma}^2$, $\tilde{Q}$, $\tilde{\beta}$.

Since the log-likelihood is equal to

$$L = -(T/2)\log 2\pi - (T/2)\log \sigma^2 - (1/2)\log \det Q$$
$$- (1/2\sigma^2)(y - X\beta)'Q^{-1}(y - X\beta),$$

we have

$$\tilde{L} = -(T/2)\log(2\pi) - (T/2)\log \tilde{\sigma}^2 - (1/2)\log \det \tilde{Q} - T/2$$

and

$$\hat{L}_0 = -(T/2)\log 2\pi - (T/2)\log \hat{\sigma}_0^2 - (1/2)\log \det \hat{Q}_0 - T/2.$$

The likelihood ratio test statistic ξ_{LR} which is equal to $2(\tilde{L} - \hat{L}_0)$ may be written as

$$\xi_{LR} = T \log \frac{S(0,0)}{S(1,1)} \tag{3}$$

with

$$S(0,0) = \hat{\sigma}_0^2(\det \hat{Q}_0)^{1/T}$$

and

$$S(1,1) = \tilde{\sigma}^2(\det \tilde{Q})^{1/T}.$$

The Wald test statistic ξ_W is based on the inequality constrained estimators and is equal to

$$\xi_{\mathrm{W}} = (R\tilde{\beta} - r)'\left[R\left(X'\tilde{Q}^{-1}X\right)^{-1}R'\right]^{-1}(R\tilde{\beta} - r)/\tilde{\sigma}^2. \tag{4}$$

ξ_{W} is the optimum value of the objective function of a primal problem (P) in which Ω is replaced by $\tilde{\Omega} = \tilde{\sigma}^2\tilde{Q}$. Consequently, ξ_{W} may be written

$$\begin{aligned}\xi_{\mathrm{W}} &= -(y - X\tilde{\beta})'\tilde{Q}^{-1}(y - X\tilde{\beta})/\tilde{\sigma}^2 + (y - X\tilde{\tilde{\beta}}_0)'\tilde{Q}^{-1}(y - X\tilde{\tilde{\beta}}_0)/\tilde{\sigma}^2 \\ &= -T + (y - X\tilde{\tilde{\beta}}_0)'\tilde{Q}^{-1}(y - X\tilde{\tilde{\beta}}_0)'/\tilde{\sigma}^2\end{aligned}$$

where $\tilde{\tilde{\beta}}_0$ is the equality constrained generalized least squares estimator of β under the constraint $R\beta = r$ and using the covariance matrix $\tilde{\Omega} = \tilde{\sigma}^2\tilde{Q}$.

If we define

$$S(0,1) = (1/T)(y - X\tilde{\tilde{\beta}}_0)'\tilde{Q}^{-1}(y - X\tilde{\tilde{\beta}}_0)(\det \tilde{Q})^{1/T},$$

then ξ_{W} may be written

$$\xi_{\mathrm{W}} = -T + T\frac{S(0,1)}{S(1,1)}. \tag{5}$$

Let us now define the Kuhn-Tucker multiplier statistic. Let $\tilde{\lambda}_0$ be the K.T. multiplier associated with the primal problem (P) in which Ω is replaced by $\hat{\Omega}_0 = \hat{\sigma}_0^2\hat{Q}_0$. ξ_{KT}, the K.T. multiplier statistic, is defined as

$$\xi_{\mathrm{KT}} = (\tilde{\lambda}_0 - \hat{\lambda}_0)'R\left(X'\hat{Q}_0^{-1}X\right)^{-1}R'(\tilde{\lambda}_0 - \hat{\lambda}_0)\hat{\sigma}_0^2/4. \tag{6}$$

Note that ξ_{KT} is the optimum value of the objective function of (P) in which Ω is replaced by $\hat{\Omega}_0$ which is equal to

$$\xi_{\mathrm{KT}} = -(y - X\tilde{\beta}_0)'\hat{Q}_0^{-1}(y - X\tilde{\beta}_0)/\hat{\sigma}_0^2 + T$$

where $\tilde{\beta}_0$ is the inequality constrained generalized least squares estimators when the covariance matrix of the disturbances is set equal to $\hat{\Omega}_0 = \hat{\sigma}_0^2\hat{Q}_0$. If we define

$$S(1,0) = (1/T)(y - X\tilde{\beta}_0)'\hat{Q}_0^{-1}(y - X\tilde{\beta}_0)(\det \hat{Q}_0)^{1/T}$$

then we may write

$$\xi_{\mathrm{KT}} = T\left[1 - \frac{S(1,0)}{S(0,0)}\right]. \tag{7}$$

Let us now consider the asymptotic distribution of these three test statistics under the null hypothesis. Let us assume that $\hat{\sigma}_0^2$ and $\tilde{\sigma}^2$ are consistent estimators of σ^2 and that $\hat{Q}_0$ and $\tilde{Q}$ are consistent estimators of Q. By continuity arguments it may be verified that the three test statistics defined in this section have the same asymptotic distribution as those which are evaluated with the true covariance matrix $\Omega = \sigma^2 Q$. According to the results mentioned in the previous Section the three test statistics are asymptotically distributed under H_0 as

$$\sum_{i=0}^{p} w(p,i)\chi^2(i)$$

where $w(p,i)$ are probability weights.

Inequalities between the three test statistics may also be established which are similar to the inequalities obtained by Berndt and Savin [**6**] (see also Savin [**20**] and Breusch [**7**]). More precisely, we have

$$(8) \qquad \xi_{\mathrm{W}} \geqq \xi_{\mathrm{LR}} \geqq \xi_{\mathrm{KT}}.$$

Let us first consider the relationship between ξ_{LR} and ξ_{W}. We may write

$$\xi_{\mathrm{LR}} = T \log\left[1 + \frac{S(0,0) - S(1,1)}{S(1,1)}\right]$$

and from the inequality $\log(1+z) \leqq z$ if $z \geqq 0$, we have

$$\xi_{\mathrm{LR}} \leqq T \frac{S(0,0) - S(1,1)}{S(1,1)}.$$

Moreover $-S(0,0)$ is obtained through the maximization under H_0 of the log-likelihood and $-S(0,1)$ is obtained by maximizing under H_0 the same function with Q being replaced by $\tilde{Q}$. As a consequence we have

$$-S(0,1) \leqq -S(0,0)$$

and hence

$$\xi_{\mathrm{LR}} \leqq T \frac{S(0,1) - S(1,1)}{S(1,1)}.$$

The right hand side of this inequality is equal to ξ_{W}. We have thus proved that

$$\xi_{\mathrm{LR}} \leqq \xi_{\mathrm{W}}.$$

Similarly, by using the inequalities $\log(1+z) \geqq z/1+z$ if $z \geqq 0$ and $S(1,0) \geqq S(1,1)$ we may prove that

$$\xi_{\mathrm{LR}} \geqq \xi_{\mathrm{KT}}$$

and hence (8) is proved.

We now indicate an alternative way of evaluating the Kuhn-Tucker multiplier statistic. It is similar to the interpretation of the Lagrange multiplier test statistic as a determination coefficient (see Breusch and Pagan [**8**] and Engle [**11**]).

In effect it may be observed that ξ_{KT} is the optimum value of objective function in the primal problem (P) associated with $\hat{\Omega}_0 = \hat{\sigma}_0^2 \hat{Q}_0$:

$$\max_{\beta} -(y - X\beta)'\hat{\Omega}_0^{-1}(y - X\beta) + (y - X\hat{\beta}_0)'\hat{\Omega}_0^{-1}(y - X\hat{\beta}_0)$$

$$\text{subject to} \qquad R\beta \geqq r.$$

Let $\hat{e}_0$ be the estimated residuals when the parameters of the model are estimated by maximum likelihood under the null hypothesis $R\beta = r$. We have

$$\hat{e}_0 = y - X\hat{\beta}_0$$

and the problem above may be written

$$\max_{\gamma} -(\hat{e}_0 - X\gamma)'\hat{\Omega}_0^{-1}(\hat{e}_0 - X\gamma) + \hat{e}_0'\hat{\Omega}_0^{-1}\hat{e}_0$$

$$\text{subject to} \qquad R\gamma \geqq 0$$

with $\gamma = \beta - \hat{\beta}_0$. Let $\tilde{\gamma}$ be the optimal value of γ. We have

$$\xi_{KT} = -(\hat{e}_0 - X\tilde{\gamma})'\hat{\Omega}_0^{-1}(\hat{e}_0 - X\tilde{\gamma}) + \hat{e}_0'\hat{\Omega}_0^{-1}\hat{e}_0. \tag{9}$$

According to this last equality we may adopt the following two-stage procedure in order to evaluate the K.T. statistic.

(i) Estimate the model by the maximum likelihood method under the constraint $R\beta = r$ and compute the estimated residuals $\hat{e}_0$.

(ii) Estimate the model

$$\hat{e}_0 = X\gamma + v$$

$$\text{subject to} \qquad R\gamma \geqq 0,$$

with the covariance matrix of v being set equal to $\hat{Q}_0$. We define the determination coefficient R^2 by

$$R^2 = 1 - \frac{(\hat{e}_0 - X\tilde{\gamma})'\hat{Q}_0^{-1}(\hat{e}_0 - X\tilde{\gamma})}{\hat{e}_0'\hat{Q}_0^{-1}\hat{e}_0}.$$

According to (9), we have

$$\xi_{KT} = -(\hat{e}_0 - X\tilde{\gamma})'\hat{Q}_0^{-1}(\hat{e}_0 - X\tilde{\gamma})/\hat{\sigma}_0^2 + \hat{e}_0'\hat{Q}_0^{-1}\hat{e}_0/\hat{\sigma}_0^2$$

and since $\hat{\sigma}_0^2 = \hat{e}_0'\hat{Q}_0^{-1}\hat{e}_0/T$, we obtain

$$\xi_{KT} = TR^2. \tag{10}$$

We may conclude this paper by noting that the three test statistics here defined possess almost all the properties of the corresponding test statistics when the regression parameters are unconstrained under the maintained hypothesis. However, there is an important difference between the usual case and the inequality constraint case. In the former, the asymptotic distribution of the test statistics is a chi-square distribution with p degrees of freedom (where p is the number of constraints) whereas in the latter the asymptotic distribution of the test statistics is a mixture of chi-square distributions with degrees of freedom ranging from 0 to p.

Université de Paris IX Dauphine, France,

Université de Lausanne, Switzerland, and Laboratoire d'Econométrie, Ecole Polytechnique, France,

and

E.N.S.A.E., Malakoff, France.

Manuscript received November, 1979; revision received January, 1981.

APPENDIX

The general case of p inequality constraints can be treated by the same method as in Section 4.2. We shall prove that, when Ω is known, the distribution of the LR test statistic under H_0 is still a mixture of chi-squared distributions. Incidentally, it follows from the arguments presented in Section 5 that this result holds asymptotically when the elements of Ω are continuous functions of a finite number of parameters.

Consider the general linear model:

$$y = X\beta + u,$$

$$R\beta \geqq r,$$

u is an n-dimensional vector $\mathfrak{N}(0, \Omega)$.

The null hypothesis is $H_0 : R\beta = r$.

As proved in Section 4, this is equivalent to the linear model

$$\hat{\beta} = \beta + v,$$

$$R\beta \geqq r,$$

v is a K-dimensional vector $\mathfrak{N}\left(0, (X'\Omega^{-1}X)^{-1}\right)$.

By applying an invertible affine transformation we obtain the following model:

$$\hat{\gamma} = \gamma + w,$$

$$\gamma_1 \geqq 0,$$

$$\gamma_2 - \theta_{21}\gamma_1 \geqq 0,$$

$$\vdots$$

$$\gamma_p - \theta_{p,p-1}\gamma_{p-1} - \cdots - \theta_{p1}\gamma_1 \geqq 0,$$

w is a K-dimensional vector $\mathfrak{N}(0, I_K)$.

Let $\bar{\gamma}$ be the $p \times 1$ vector such that

$$\bar{\gamma}' = (\gamma_1, \gamma_2, \ldots, \gamma_p).$$

The above inequalities can be written, in obvious notation,

$$\theta_1'\bar{\gamma} \geqq 0,$$

$$\vdots$$

$$\theta_p'\bar{\gamma} \geqq 0.$$

It is readily verified that the null hypothesis is transformed into $H_0: \gamma_1 = \gamma_2 = \ldots = \gamma_p = 0$, i.e., $\bar{\gamma} = 0$.

The test statistic ξ_{LR} is

$$\xi_{\mathrm{LR}} = \|\hat{\gamma} - \hat{\gamma}_0\|^2 - \|\hat{\gamma} - \tilde{\gamma}\|^2$$

where $\hat{\gamma}_0$ and $\tilde{\gamma}$ are the M.L. estimates of γ under H_0 and under the maintained hypothesis respectively. Since $\hat{\gamma}_0$ is the orthogonal projection of $\tilde{\gamma}$ on the subspace which defines H_0, we have

$$\xi_{\mathrm{LR}} = \|\tilde{\gamma} - \hat{\gamma}_0\|^2$$
$$= \tilde{\gamma}_1^2 + \ldots + \tilde{\gamma}_p^2.$$

Let $\hat{\bar{\gamma}}$ be the $p \times 1$ vector consisting of the first p components of $\hat{\gamma}$. The $p \times 1$ vector $\tilde{\bar{\gamma}}$ is the orthogonal projection of $\hat{\bar{\gamma}}$ on the cone $C_{\{1,2,\ldots,p\}}$ defined by

$$C_{\{1,2,\ldots,p\}} = \{\bar{\gamma} \in \mathbb{R}^p \mid \theta_j'\bar{\gamma} \geqq 0, j = 1, \ldots, p\}.$$

We have:

$$\xi_{\mathrm{LR}} = \|\tilde{\bar{\gamma}}\|^2$$

and this test statistic takes different forms (exactly 2^p) according to the location of $\hat{\bar{\gamma}}$ in $\mathbb{R}^p$. More precisely, let us associate with each vector θ_i, $i = 1, \ldots, p$, a vector $\varphi_i \in \mathbb{R}^p$ such that φ_i is orthogonal to any θ_j for any $j \neq i$ and $\theta_i'\varphi_i < 0$ (i.e., θ_i and φ_i are not in the same half-space determined by the θ_j's for any $j \neq i$). Now, with each subset S of the set $\{1, \ldots, p\}$ we associate the cone C_S defined as

$$C_S = \left\{ \bar{\gamma} \in \mathbb{R}^p \mid \bar{\gamma} = \sum_{i=1}^{p} \alpha_i \eta_i;\ \alpha_i \leqq 0;\ \eta_i = \theta_i \text{ when } i \notin S \text{ and } \eta_i = \varphi_i \text{ when } i \in S \right\}.$$

Let E_S be the linear subspace of $\mathbb{R}^p$ spanned by the vectors φ_i, $i \in S$ (by convention, $E_\phi = \{0\}$). If $\hat{\bar{\gamma}}$ belongs to C_S then $\tilde{\bar{\gamma}}$ is the orthogonal projection of $\hat{\bar{\gamma}}$ on E_S which we call $\Pi_S(\hat{\bar{\gamma}})$; therefore,

$$\xi_{\mathrm{LR}} = \|\Pi_S(\hat{\bar{\gamma}})\|^2 \qquad \text{if} \quad \hat{\bar{\gamma}} \in C_S.$$

Note that this defines ξ_{LR} everywhere.

Let A be any Borel set of $\mathbb{R}$. We have,

$$\text{(A1)} \qquad P\{\xi_{\mathrm{LR}} \in A\} = \sum_S P\{\hat{\bar{\gamma}} \in C_S;\ \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A\}.$$

We wish to factorize each of the terms $P\{\hat{\bar{\gamma}} \in C_S;\ \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A\}$. To this end, we first note that

$$C_S = C_S^\varphi + C_S^\theta$$

where

$$C_S^{\varphi} = \left\{ x \in \mathbb{R}^p \mid x = \sum_{i \in S} \alpha_i \varphi_i ;\ \alpha_i \leqq 0 \right\}$$

and

$$C_S^{\theta} = \left\{ x \in \mathbb{R}^p \mid x = \sum_{i \notin S} \alpha_i \theta_i ;\ \alpha_i \leqq 0 \right\}.$$

Next, we observe that $C_S^{\varphi} = E_S \cap C_S$ and $C_S \subset E_S^{\perp}$. Therefore, an alternative definition of C_S is:

$$C_S = \left\{ \bar{\gamma} \in \mathbb{R}^p \mid \Pi_S(\bar{\gamma}) \in C_S^{\varphi} ; \Pi_S^{\perp}(\bar{\gamma}) \in C_S^{\theta} \right\}$$

where $\Pi_S^{\perp}(\bar{\gamma})$ is the orthogonal projection of $\bar{\gamma}$ on $E_S^{\perp}$. It follows that

(A2) $$P\left\{ \hat{\bar{\gamma}} \in C_S ; \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A \right\} = P\left\{ \Pi_S(\hat{\bar{\gamma}}) \in C_S^{\varphi} ; \Pi_S^{\perp}(\hat{\bar{\gamma}}) \in C_S^{\theta} ; \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A \right\}.$$

Under the null hypothesis $\hat{\bar{\gamma}}$ has a $\mathfrak{N}(0, I_p)$ distribution which implies that $\Pi_S(\hat{\bar{\gamma}})$ and $\Pi_S^{\perp}(\hat{\bar{\gamma}})$ are independent. Hence, the right hand side of (A2) can be factorized as:

(A3) $$P\left\{ \Pi_S(\hat{\bar{\gamma}}) \in C_S^{\varphi} ; \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A \right\} \cdot P\left\{ \Pi_S^{\perp}(\hat{\bar{\gamma}}) \in C_S^{\theta} \right\}.$$

Since $\Pi_S(\hat{\bar{\gamma}})$ is spherically distributed, $\|\Pi_S(\hat{\bar{\gamma}})\|$ and $\Pi_S(\hat{\bar{\gamma}})/\|\Pi_S(\hat{\bar{\gamma}})\|$ are independent (see Dempster [**10**, Theorem 12.2.2.]). In addition, the events $\{\Pi_S(\hat{\bar{\gamma}}) \in C_S^{\varphi}\}$ and $\{\Pi_S(\hat{\bar{\gamma}})/\|\Pi_S(\hat{\bar{\gamma}})\| \in C_S^{\varphi}\}$ are identical. Therefore, we may write (A3) as:

(A4) $$P\left\{ \Pi_S(\hat{\bar{\gamma}}) \in C_S^{\varphi} \right\} \cdot P\left\{ \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A \right\} \cdot P\left\{ \Pi_S^{\perp}(\hat{\bar{\gamma}}) \in C_S^{\theta} \right\}.$$

Gathering the first and the third term, this expression becomes:

(A5) $$P\left\{ \hat{\bar{\gamma}} \in C_S \right\} \cdot P\left\{ \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A \right\}$$

which implies that

$$P\left\{ \xi_{\mathrm{LR}} \in A \right\} = \sum_S P\left\{ \hat{\bar{\gamma}} \in C_S \right\} \cdot P\left\{ \|\Pi_S(\hat{\bar{\gamma}})\|^2 \in A \right\}$$

as we wished to prove.

Finally, since the distribution of $\|\Pi_S(\hat{\bar{\gamma}})\|^2$ under H_0 is $\chi^2(\dim E_S)$, the distribution of ξ_{LR} is

(A6) $$\sum_S P(\hat{\bar{\gamma}} \in C_S) \cdot \chi^2(\dim E_S).$$

Since $\hat{\bar{\gamma}}$ is spherically distributed the weight $P(\hat{\bar{\gamma}} \in C_S)$ is equal to the volume of $C_S \cap \Sigma$ where Σ is the sphere centered at the origin whose volume is equal to 1. This remark may serve as a basis of a numerical computation of these weights by simulation techniques.

Let us briefly consider the particular "orthogonal" case in which the maintained hypothesis is $\gamma_1 \geqq 0, \gamma_2 \geqq 0, \ldots, \gamma_p \geqq 0$. We have:

$$\tilde{\gamma}_i = \begin{cases} \hat{\gamma}_i & \text{when } \hat{\gamma} \geqq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad (i = 1, \ldots, p).$$

It follows from the theory of Section 4.1 that the distribution of $\tilde{\gamma}_i^2$ is, under $H_0, \frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$.

Furthermore, the $\tilde{\gamma}_i$'s are independent. Hence, the distribution of $\xi_{LR} = \sum_{i=1}^{p} \tilde{\gamma}_i^2$ is the convolution product

$$\left(\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)\right)^{*p} = \frac{1}{2^p}\sum_{i=1}^{p}\binom{p}{i}\chi^2(i)$$

and the weights $w(p, i)$ are equal to

$$\binom{p}{i}/2^p \qquad (i = 0, \ldots, p).$$

Note that we are in the "orthogonal" case if, and only if, the initial constraints are orthogonal for the scalar product associated with $(X'\Omega^{-1}X)^{-1}$, i.e., if, and only if

$$R_i(X'\Omega^{-1}X)^{-1}R_j' = 0 \qquad \forall i \neq j.$$

REFERENCES

[1] Aitchison, J., and S. D. Silvey: "Maximum Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29 (1958), 813–828.

[2] Barlow, R. E., D. J. Bartholomew, J. N. Bremner, and H. D. Brunk: *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression.* New York: John Wiley & Sons, 1972.

[3] Bartholomew, D. J.: "A Test of Homogeneity of Ordered Alternatives, I," *Biometrika*, 46 (1959), 36–48.

[4] ———: "A Test of Homogeneity of Ordered Alternatives, II," *Biometrika*, 46 (1959), 328–335.

[5] ———: "A Test of Homogeneity of Means Under Restricted Alternatives," *Journal of the Royal Statistical Society*, Series B, 23 (1961), 239–281.

[6] Berndt, E. R., and N. E. Savin: "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45 (1977), 1263–1278.

[7] Breusch, T. S.: "Conflict Among Criteria for Testing Hypotheses: Extensions and Comments," *Econometrica*, 47 (1979), 203–206.

[8] Breusch, T. S., and A. R. Pagan: "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47 (1980), 239–253.

[9] Chacko, V. J.: "Testing Homogeneity Against Ordered Alternatives," *Annals of Mathematical Statistics*, 34 (1963), 945–956.

[10] Dempster, A. P.: *Continuous Multivariate Analysis*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1969.

[11] Engle, R.: "The Lagrange Multiplier Test as a Regression Diagnostic," mimeo, University of California, San Diego, March, 1978.

[12] Ekeland, I.: *La Théorie des Jeux et ses Applications à l'Économie Mathématique*. Paris: Presses Universitaires de France, 1974.

[13] Judge, G. C., and T. Takayama: "Inequality Restrictions in Regression Analysis," *Journal of the American Statistical Association*, 61 (1966), 166–181.

[14] Kudô, A.: "A Multivariate Analogue of the One-Sided Test," *Biometrika*, 50 (1963), 403–418.

[15] Liew, C. K.: "Inequality Constrained Least-Squares Estimation," *Journal of the American Statistical Association*, 71 (1976), 746–751.

[16] Nüesch, P. E.: "Multivariate Tests of Location for Restricted Alternatives," Doctoral Dissertation, Swiss Federal Institute of Technology, Zurich, 1964.

[17] ———: "On the Problem of Testing Location in Multivariate Problems for Restricted Alternatives," *Annals of Mathematical Statistics*, 37 (1966), 113–119.

[18] Perlman, M. D.: "One Sided Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, 40 (1969), 549–567.

[19] Rao, C. R.: "Large Sample Tests of Statistical Hypothesis Concerning Several Parameters with Applications to Problems of Estimation," *Proceedings of the Cambridge Philosophical Society*, 44 (1948), 50–57.

[20] Savin, N. E.: "Conflict Among Testing Procedures in a Linear Regression Model with Autoregressive Disturbances," *Econometrica*, 44 (1976), 1303–1315.

[21] SHORACK, G. R.: "Testing Against Ordered Alternatives in Model I Analysis of Variance; Normal Theory and Nonparametric," *Annals of Mathematical Statistics*, 38 (1967), 1740–1753.
[22] SILVEY, S. D.: "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30 (1959), 389–407.
[23] WALD, A.: "Tests of Statistical Hypothesis Concerning Several Parameters when the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54 (1943), 426–482.