Review of Economic Studies (1980) XLVII, 255–273
0034–6527/80/00120255$02.00

The Bonferroni and the Scheffé Multiple Comparison Procedures

N. E. SAVIN
Trinity College, Cambridge

1. INTRODUCTION

Consider the regression model

$$y = X\beta + u \qquad \ldots(1)$$

where y is a $T \times 1$ vector of observations on the dependent variable, X is a $T \times k$ non-stochastic matrix of rank k, β is an unknown $k \times 1$ parameter vector and u is a $T \times 1$ vector of random disturbances which is distributed $N(0, \sigma^2 I)$ where $\sigma^2 > 0$ is unknown.

Suppose we wish to test the hypothesis

$$H : R\beta - r = \theta = 0 \qquad \ldots(2)$$

where R is a known $q \times k$ matrix of rank $q \leqq k$ and r is a known $q \times 1$ vector. The best linear unbiased estimator of θ is

$$z = Rb - r \qquad \ldots(3)$$

where $b = (X'X)^{-1}X'y$ is the least squares estimator of β. It is easily shown that z is $N(\theta, \sigma^2 V)$ where $V = R(X'X)^{-1}R'$. The usual unbiased estimator of σ^2 is $s^2 = (y - Xb)'(y - Xb)/(T - k)$.

In situations in which we wish only to decide whether H is true or not we can use a direct test of H such as an F test. It is perhaps more common that when H is rejected we want to know which components of θ are different from zero and of the non-zero components which are positive and which negative. In this situation we have a multiple decision problem and a natural solution is to use an induced test. As an example suppose in the case $q = 2$ that we wish to test the hypothesis $H : \theta_1 = \theta_2 = 0$. Since H is true if and only if the separate hypotheses $H_1 : \theta_1 = 0$ and $H_2 : \theta_2 = 0$ are both true, this suggests a sequence of separate tests which will induce a test of H. Testing the two hypotheses H_1 and H_2 where we are interested in whether θ_1 or θ_2 or both are different from zero induces a multiple decision problem in which the four possible decisions are

d^{00} : H_1 and H_2 are both true,

d^{01} : H_1 is true, H_2 is false,

d^{10} : H_1 is false, H_2 is true,

d^{11} : H_1 and H_2 are both false.

Now suppose that a test of H_ν is defined by the acceptance region A_ν^0 and the rejection region A_ν^1 $(\nu = 1, 2)$. These separate tests induce a decision procedure for the four decision problem, this induced procedure being defined by assigning to the decision d^{ij} the region $A_1^i \cap A_2^j$. This induced procedure accepts $H : \theta_1 = \theta_2 = 0$ if and only if H_1 and H_2 are both accepted.

More generally suppose that the hypothesis H is true if and only if the separate hypotheses $H_1, H_2, \ldots$ are true. A procedure for testing H, which Seber (1964) calls the separate induced test, is to accept H if and only if all the separate hypotheses are accepted. The idea of induced tests is discussed in two papers by Lehmann ((1957*a*) and (1957*b*)) and subsequently by Darroch and Silvey (1963) (the source of our example) and by Seber (1964). The separate induced test is closely related to the union–intersection principle of test construction proposed by Roy (1953) and utilized further by Roy and Bose (1953).

Two separate induced procedures which can always be applied to the linear regression model are the Bonferroni B procedure and the Scheffé S procedure. The B procedure is discussed in Miller (1966) and applications to econometrics are found in Jorgenson and Lau (1975), Christensen, Jorgenson and Lau (1975) and Sargan (1976). The S method was first published in Scheffé (1953), elaborated in Scheffé (1959) and reformulated as the S procedure in Scheffé (1977*a*).

Two applications of the S procedure in the econometrics literature are found in Jorgenson ((1971) and (1974)). Reviewing the literature it appears that neither the B nor S procedures are widely used in econometric research. Further, it also appears that how the performance of the B procedure compares with the S procedure is not well understood. Given this state of affairs, the purpose of this paper is to present an exposition of the B and S procedures and to discuss the merits of each. The exposition draws heavily on Miller ((1966) and (1977)) and Scheffé (1959), (1977*a*), (1977*b*).

The organization of the paper is the following. In Section 2 a direct test (F test) and a separate induced test of H are presented and the power of the two tests is compared. In Section 3 the B and S procedures are developed and the length of the simultaneous confidence intervals employed by the two procedures is compared. A more general approach to multiple comparison procedures is considered in Section 4 and this approach is the basis for further comparisons between the B and S procedures. Large sample analogues of the B and S procedures are developed in Section 5 and Section 6 presents two empirical examples. The concluding comments are in Section 7.

2. THE DIRECT TEST AND THE SEPARATE INDUCED TEST

2.1. *The Direct Test*

A direct test of the hypothesis H is obtained by using the familiar F statistic

$$F = z'V^{-1}z/qs^2. \qquad \ldots(4)$$

For an α level F test of H the acceptance region is

$$F \leqq F_\alpha(q, T-k) \qquad \ldots(5)$$

where $F_\alpha(q, T-k)$ is the upper α significance point of an F distribution with q and $T-k$ degrees of freedom. The F test of H is equivalent to one derived from a confidence region

$$(z-\theta)'V^{-1}(z-\theta) \leqq s^2S^2 \qquad \ldots(6)$$

where $S^2 = qF_\alpha(q, T-k)$. The inequality (6) determines an ellipsoid in the θ space with centre at z. The probability that this random ellipsoid covers θ is $1-\alpha$. The F test of H accepts H if and only if the ellipsoid covers the origin.

2.2. *The Separate Induced Test*

Suppose we wish to test the separate hypotheses

$$H_i : \theta_i = 0, \qquad i = 1, 2, \ldots, q, \qquad \ldots(7)$$

and that a test of H_i is defined by the region

$$|t_i| \leqq B, \qquad i = 1, 2, \ldots, q, \qquad \ldots(8)$$

where t_i is the t statistic

$$t_i = z_i/(s^2 V_{ii})^{\frac{1}{2}}, \qquad i = 1, 2, \ldots, q, \qquad \ldots(9)$$

V_{ii} is the ith diagonal element of V and $B = t_{\delta/2}(T-k)$ is the upper $\delta/2$ significance point of a t distribution with $T-k$ degrees of freedom. The separate induced test of H accepts H if and only if all the separate hypotheses $H_1, H_2, \ldots, H_q$ are accepted. The acceptance region of the separate induced test of H is the polyhedron

$$|t_1| \leqq B, |t_2| \leqq B, \ldots, |t_q| \leqq B,$$

i.e. the intersection of the separate acceptance regions (8).

It is easy to calculate the probability that the point $(t_1, t_2, \ldots, t_q)$ falls in the acceptance region when the t_i are independent. In general the t_i are dependent. The separate induced test of H accepts H if and only if the maximum of the $|t_i|$ is less than the critical value, i.e.

$$\max_i |t_i| \leqq B. \qquad \ldots(10)$$

The distribution of the maximum of the $|t_i|$ is not readily available except in a few special cases.

A bound on the probability possessed by the polyhedron is given by the Bonferroni inequality $P(A_1, A_2, \ldots, A_m) \geqq 1 - \sum_{i=1}^{m} P(A_i^c)$ where A_i is an event and A_i^c its complement. Applying the inequality the probability that $(t_1, t_2, \ldots, t_q)$ falls in the polyhedron is $\geqq 1 - \delta q$ when H is true, i.e. the significance level of the separate induced test is $< \delta q$. Hence the probability is $\geqq 1 - \alpha$ that the separate induced test of H accepts H when H is true and $\delta = \alpha/q$. Tables of the percentage points of the Bonferroni t statistic have been prepared by Dunn (1961).

Šidák (1967) has proved a general inequality which gives a slight improvement over the Bonferroni inequality when both are applicable. The Šidák inequality states that if $X = (X_1, \ldots, X_r)'$ is $N(0, \sum)$ where $\sum$ is an arbitrary covariance matrix and if W^2 is an independent random variable distributed as a χ^2 with n degrees of freedom divided by n, $\chi^2(n)/n$, then

$$P\left(\left|\frac{X_1}{W}\right| \leqq c_1, \ldots, \left|\frac{X_r}{W}\right| \leqq c_r\right) \geqq \prod_{i=1}^{r} \left(P\left|\frac{X_r}{W}\right| \leqq c_i\right). \qquad \ldots(11)$$

In words, the probability that a random normal vector (divided by W) with arbitrary covariances falls inside a polyhedron centred at its mean is always at least as large as the corresponding probability for the case where the covariances are zero, i.e. the X_i's are independent. This inequality produces slightly sharper tests or intervals than the Bonferroni inequality because $(1-\delta)^r > 1 - r\delta$. Games (1977) has recently prepared tables of the percentage points of the Šidák t statistic. Charts by Moses (1976) may be used to find the appropriate t critical value with either the Bonferroni or Šidák inequality. In this paper we only employ the Bonferroni inequality.

The separate induced test of H is equivalent to one based on a confidence region. The probability is $\geqq 1 - \alpha$ that simultaneously for all $i = 1, 2, \ldots, q$,

$$|z_i - \theta_i| \leqq Bs(V_{ii})^{\frac{1}{2}} \qquad \ldots(12)$$

where $B = t_{\alpha/2q}(T-k)$. The inequalities define a polyhedron in the θ space with centre at z. The probability that this random polyhedron covers the true parameter point θ is $\geqq 1 - \alpha$. The separate induced test of H accepts H if and only if the polyhedron covers the origin.

The following illustrative example of a direct test and a separate induced test of H is taken from Miller (1966, pp. 12–22). Suppose θ_1 and θ_2 are the parameters of primary interest and $\sigma^2 V = I$ which is known. The direct test of H is now a χ^2 test. The acceptance region of an α level χ^2 test of H is

$$z_1^2 + z_2^2 \leqq S^2 \qquad \ldots(13)$$

where $S^2 = \chi^2_\alpha(2)$ is the upper α significant point of a χ^2 distribution with 2 degrees of freedom. The acceptance region is a circular region in the z_1 and z_2 space with centre at (0, 0) and radius S as shown in Figure 1. For a 0·05 level test $S^2 = 5{\cdot}991$ so that the radius $S = 2{\cdot}448$. As a confidence region, the circular region is centred at (z_1, z_2) in the θ_1 and θ_2 space.

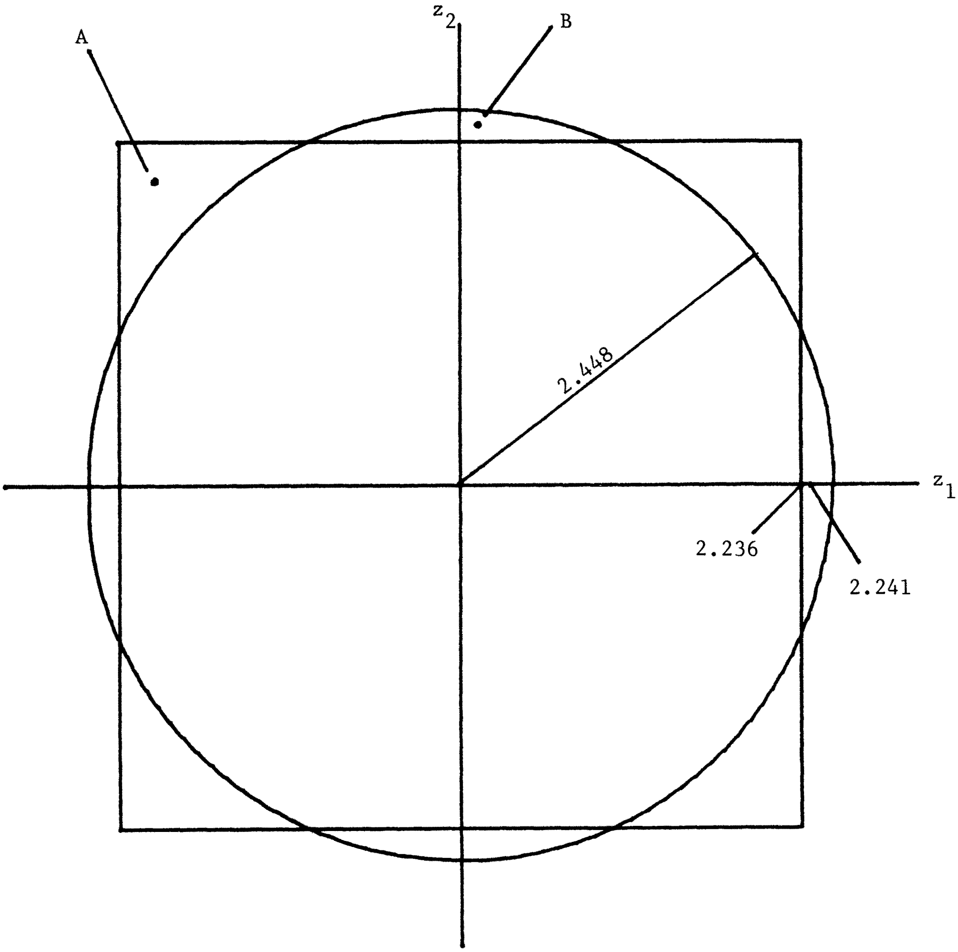

FIGURE 1

The acceptance region of the separate induced test of H is the square region

$$|z_1| \leqq B, \; |z_2| \leqq B \qquad \ldots(14)$$

where B is the upper $\delta/2$ significance point of a $N(0, 1)$ distribution. Since z_1 and z_2 are independent we can determine the exact $100(1-\alpha)$ per cent acceptance region. For an exact 0·95 per cent acceptance region $\delta = 0{\cdot}025$ since $(1-\delta)^2 = (0{\cdot}9747)^2 = 0{\cdot}95$ and hence $B = 2{\cdot}236$. The acceptance region of an exact 0·05 level separate induced test of H is the square region centred at (0, 0) with sides $2B = 4{\cdot}472$ in Figure 1. As a confidence region the square region is centred at (z_1, z_2) in the θ_1 and θ_2 space.

Applying the Bonferroni inequality given $\delta = 0{\cdot}05/2 = 0{\cdot}025$ the probability is $\geqq 0{\cdot}95$ that the separate induced test of H accepts H when H is true. Since $\delta = 0{\cdot}025$ implies that $B = 2{\cdot}241$ the acceptance region for a nominal 0·05 level Bonferroni separate induced test is a square region with sides $2B = 4{\cdot}482$. There is an increase of 0·01 in the sides of the square when the inequality is used. The true overall significance level is $1-(0{\cdot}975)^2 = 0{\cdot}0494$, which is quite close to the bound 0·05.

The square acceptance region of the separate induced test of H using either the exact or the Bonferroni intervals does not lie completely inside of the circular region of the χ^2 test. Hence H can be accepted by the separate induced test and rejected by the direct (χ^2) test and *vice versa*. In Figure 1, H is accepted by the separate induced test and rejected by the χ^2 test when point A is the point (z_1, z_2) and *vice versa* when point B is the point (z_1, z_2).

2.3. *Power of the Tests*

Christensen (1973) has calculated the power of the separate induced test and χ^2 test of H for the case where θ_1 and θ_2 are the parameters of primary interest, σ is known and

$$V = \frac{1}{1-r^2}\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}, |r| < 1. \qquad \ldots(15)$$

The acceptance region of the separate induced test is a square with sides $4{\cdot}482\,(1-r^2)^{-\frac{1}{2}}\sigma$ and the acceptance region of the χ^2 test is an elliptical region $z'V^{-1}z \leqq \sigma^2\, 5{\cdot}991$. The probability of a Type I error is no larger than 0·05 for the Bonferroni separate induced test so that 0·05 can be used as the nominal significance level, and exactly 0·05 for the direct (χ^2) test.

The power calculations show that neither test is more powerful against all alternatives. For example, when $r = 0$ the separate induced test is more powerful against the alternative $\theta_1 = \theta_2 = 1{\cdot}585\sigma$. This is not surprising since neither of the acceptance regions contain the other. Despite this, though, Christensen found that when $|r|$ was small the power of the two test procedures was approximately the same regardless of the alternative. However, when $|r|$ was high the separate induced test had very little power against any alternatives. If only θ_1 or θ_2 is different from zero, then the χ^2 test has good power regardless of the value of r. When both θ_1 and θ_2 are different from zero the power of the χ^2 test is mixed: against some alternatives the power is extremely good–increasing with $|r|$. On the other hand, the power against other alternatives decreased badly with increasing $|r|$.

One of the reasons for the poor power of the separate induced test is that the actual significance level decreases as $|r|$ increases. For $r = 0$ the actual significance level is $1-(0{\cdot}975)^2 = 0{\cdot}0494$. As $|r|$ approaches one, the actual significance level approaches 0·025. Hence what is being calculated is the probability of not making a Type II error for a significance level smaller than the nominal one, which is not strictly the power required to be compared with that of the direct test. For the more general discussion of the power of the separate induced test see Darroch and Silvey (1963) and Seber (1964).

3. THE B AND S MULTIPLE COMPARISON PROCEDURES

3.1. *Preliminaries*

Let L be the set of linear combinations $\psi = a'\theta$ such that $\theta \in R^q$. The minimum variance unbiased estimator of any ψ in L is $\hat{\psi} = a'z$ where the usual estimator of its variance is $\hat{\sigma}^2_{\hat{\psi}} = s^2 a'Va$.

In the multiple decision problem the ψ in L are divided into two sets where one set consists of the ψ of primary interest and the other the ψ of secondary interest. The θ_i are always treated as ψ of primary interest and, in addition, there may be other ψ of primary interest. It is assumed that the ψ of primary interest are selected before analysing the data.

Let G be the set of ψ in L of primary interest and the complement of G relative to L, denoted by $L-G$, be the set of ψ in L of secondary interest. The set G is either a finite or an infinite set. If G is an infinite set then G is either a proper subset of L or equal to L. In the latter case all the ψ in L are of primary interest. The nature of G defines three possible situations:

(i) G finite, $L-G$ infinite,
(ii) G infinite, $L-G$ infinite,
(iii) G infinite, $L-G$ empty.

In the B procedure G is a finite set and in the S procedure G is infinite and equal to L. There appear to be no procedures for situation (ii).

As noted earlier there is the difficulty of computing the significance level of the separate induced test. If G is a finite set an upper bound on the significance level of the test can always be calculated using the Bonferroni or Šidák inequality. The exact significance level can also be easily calculated for a few cases in the one-way lay-out of the analysis of variance. These include Tukey's procedure (see Scheffé (1959, Theorem 1, p. 731)) and the Dunnett (1955) procedure. In the Tukey procedure G is the set of $\frac{1}{2}q(q-1)$ differences between means $(\theta_i-\theta_j)$, $i, j=1, 2, \ldots, q$, $i \neq j$, and in the Dunnett procedure G is the set of $(q-1)$ differences $(\theta_1-\theta_i)$, $i=2, \ldots, q$, between a control mean θ_1 and the other means. In a one-way lay-out where G is any finite set Hochberg and Rodriqúez (1977) propose using a second-order Bonferroni approximation due to Siotani (1964) to calculate an upper bound on the significance level of the test. In the case of the S procedure the exact significance level of the separate induced test is known due to the relation between the S procedure and the F test.

Before turning to the B and S procedures it is worth remarking that when the number of ψ in G is greater than q the separate induced procedure produces decisions which at first sight may appear puzzling. As an example suppose $q=2$ and that the three ψ in G are $\psi_1=\theta_1$, $\psi_2=\theta_2$ and $\psi_3=\theta_1+\theta_2$. Testing the three separate hypothesis $H_i:\psi_i=0$, $i=1, 2, 3$, induces a decision problem in which one of the eight possible decisions is:

$$H_1 \text{ and } H_2 \text{ are both true, } H_3 \text{ is false.}$$

This decision may appear puzzling since if H_1 and H_2 are both true, then H_3 is also true. The puzzle is resolved by observing that this decision is not made if it is known that H_1 and H_2 are both true since in this case there is no decision problem.

3.2. *The* B *Procedure*

Let G be a set of finite number m of ψ in L of primary interest and denote the ψ in G by $\psi_i=a_i'\theta$, $i=1, 2, \ldots, m$. Since the θ_i are of primary interest the hypothesis $H:\theta=0$ is equivalent to

$$H:\psi_1=\psi_2=\ldots=\psi_m=0. \qquad \ldots(16)$$

As a consequence there is a separate induced test of $H:\theta=0$ for each choice of G. A separate induced test of H can accept H for one set G and reject H for another set. Hence it is important that G should be selected before analysing the data.

Suppose we wish to test the separate hypotheses

$$H(a_i):\psi_i=a_i'\theta, \qquad i=1, 2, \ldots, m, \qquad \ldots(17)$$

and that a test of $H(a_i)$ is defined by the acceptance region

$$|t_0(a_i)| \leqq B, \qquad i=1, 2, \ldots, m, \qquad \ldots(18)$$

where $t_0^2(a_i)$ is the squared t ratio

$$t_0^2(a_i) = [a_i'z]^2/s^2a_i'Va_i, \qquad i = 1, 2, \ldots, m, \qquad \ldots(19)$$

and where $B = t_{\delta/2}(T-k)$. The separate induced test of H accepts H if and only if all the separate hypotheses $H(a_1)$, $H(a_2), \ldots, H(a_m)$ are accepted. Applying the Bonferroni inequality given $\delta = \alpha/m$ the probability is $\geqq 1-\alpha$ that the separate induced test of H accepts H when H is true.

Simultaneous confidence intervals can be constructed for all ψ in G. The probability is $\geqq 1-\alpha$ that simultaneously for all ψ in G

$$\hat{\psi} - B\hat{\sigma}_{\hat{\psi}} \leqq \psi \leqq \hat{\psi} + B\hat{\sigma}_{\hat{\psi}} \qquad \ldots(20)$$

where $B = t_{\delta/2}(T-k)$ when applying the Bonferroni inequality given $\delta = \alpha/m$, and these intervals are called B intervals. The intersection of the B intervals for all ψ in G, which may be called the induced confidence region, is a polyhedron in the θ space. The separate induced test of H accepts H if and only if all the B intervals cover zero.

An estimate $\hat{\psi}$ of ψ is said to be significantly different from zero (sdfz) according to the B criterion if the B interval does not cover $\psi = 0$, i.e. if $|\hat{\psi}| > B\hat{\sigma}_{\hat{\psi}}$. Hence H is rejected if and only if the estimate of at least one ψ in G is sdfz according to the B criterion.

As an example of a Bonferroni separate induced test when $m > q$ consider the case where $m = 3$, $q = 2$ and $\sigma^2V = I$ which is known. Suppose that the three ψ in G are $\psi_1 = \theta_1$, $\psi_2 = \theta_2$ and $\psi_3 = \theta_1 + \theta_2$ and that tests of the three separate hypotheses $H_i : \psi_i = 0$, $i = 1, 2, 3$, are defined by the three separate acceptance regions

$$\begin{gathered} |z_1| \leqq 2{\cdot}39, \, |z_2| \leqq 2{\cdot}39, \\ |z_1 + z_2| \leqq (2)^{\frac{1}{2}} 2{\cdot}39 = 3{\cdot}380, \end{gathered} \qquad \ldots(21)$$

respectively, where $2{\cdot}39$ is the upper $0{\cdot}05/2(3) = 0{\cdot}00833$ significance point of a $N(0, 1)$ distribution. Suppose the hypothesis H is accepted if and only if the separate hypotheses H_1, H_2 and H_3 are accepted. Applying the Bonferroni inequality the probability is $\geqq 0{\cdot}95$ that the separate induced test of H accepts H when H is true. The acceptance region of the separate induced test of H, which is the intersection of the three separate acceptance regions, is shown in Figure 2. When A is the point (z_1, z_2) the hypothesis H is rejected and the decision is that H_1 and H_2 are both true and H_3 is false.

For comparison consider the case where $m = q = 2$. The tests of the two separate hypotheses $\psi_1 = \theta_1$ and $\psi_2 = \theta_2$ are now defined by the two acceptance regions

$$|z_1| \leqq 2{\cdot}24, \, |z_2| \leqq 2{\cdot}24, \qquad \ldots(22)$$

respectively, where $2{\cdot}24$ is the upper $0{\cdot}05/2(2) = 0{\cdot}0125$ significance point of a $N(0, 1)$ distribution. Applying the Bonferroni inequality the probability is $\geqq 0{\cdot}95$ that this separate induced test of H accepts H when H is true. The acceptance region of this separate induced test of H is the inner square region shown in Figure 2. With this region we accept H when A is the point (z_1, z_2). When B is the point (z_1, z_2) the hypothesis H is accepted if ψ_3 is of primary interest and rejected if ψ_3 is of secondary interest. This comparison shows that the separate induced test of H can accept H for one set of ψ of primary interest and reject H for another set.

3.3. *The* S *Procedure*

In the S procedure all ψ in L are of primary interest. Suppose for all ψ in L we wish to test the separate hypotheses

$$H(a) : \psi = a'\theta = 0 \qquad \ldots(23)$$

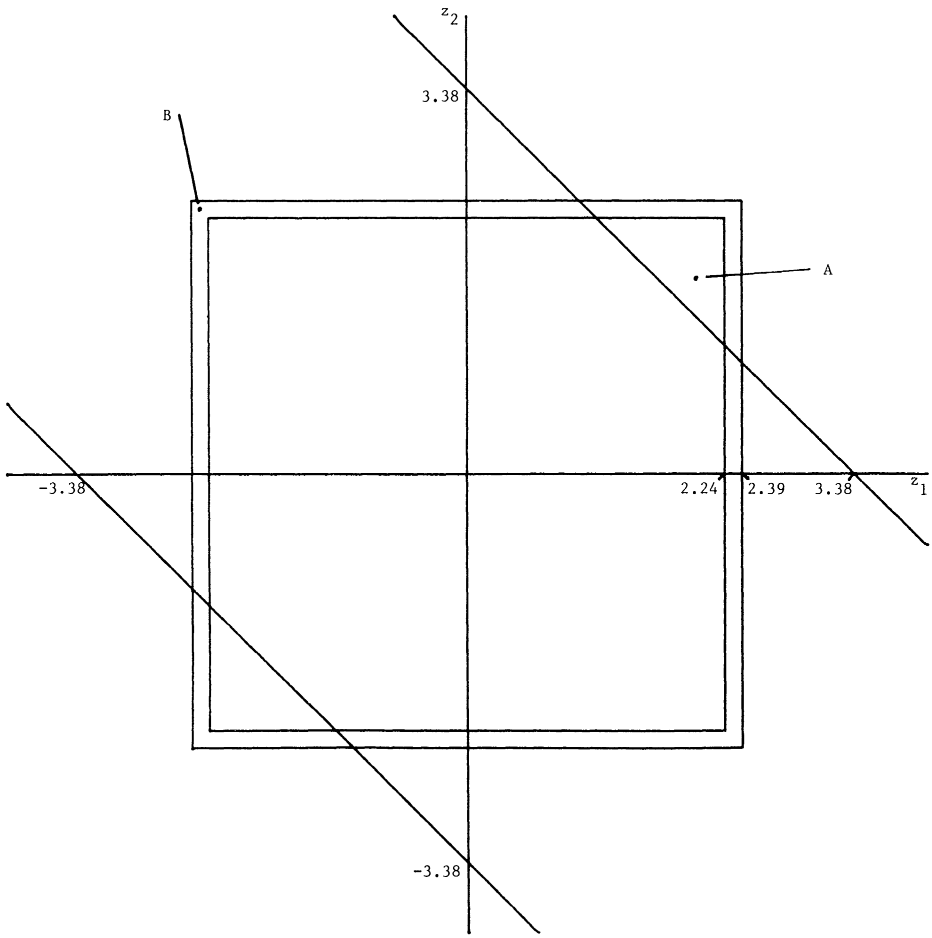

FIGURE 2

and that a test of $H(a)$ is defined by the acceptance region

$$|t_0(a)| \leqq S \qquad \ldots(24)$$

where $t_0(a)$ is the squared t ratio

$$t_0^2(a) = [a'z]^2/s^2a'Va \qquad \ldots(25)$$

and where $S = (qF_\alpha(q, T-k))^{\frac{1}{2}}$. Suppose we accept H if and only if we accept the separate hypotheses $H(a): \psi = 0$ for all ψ in L. A remarkable fact is that the intersection of the separate acceptance regions (24) for all ψ in L is the acceptance region of an α level F test of H. In other words, the acceptance region of the α level Scheffé separate induced test of H is the same as the acceptance region of the α level direct test of H, i.e. the F test. As a consequence we can start the Scheffé separate induced test procedure with an F test of H. If the F test of H rejects H, the next step is to test the separate hypotheses in order to decide which are false, i.e. which are responsible for the rejection of the separate induced test of H.

The S procedure is based on the following theorem which is proved by Scheffé (1959, pp. 69–70). The probability is $1-\alpha$ that simultaneously for all ψ in L

$$\hat{\psi}-S\hat{\sigma}_{\hat{\psi}} \leqq \psi \leqq \hat{\psi}+S\hat{\sigma}_{\hat{\psi}} \qquad \ldots(26)$$

where $S=(qF_\alpha(q, T-k))^{\frac{1}{2}}$. These intervals are called S intervals.

The theorem states that the probability is $1-\alpha$ that simultaneously for all non-null a

$$t^2(a) \leqq S^2 \qquad \ldots(27)$$

where $t^2(a)$ is the squared t ratio

$$\begin{aligned} t^2(a) &= (\hat{\psi}-\psi)^2/\hat{\sigma}^2_{\hat{\psi}} \\ &= [a'(z-\theta)]^2/S^2 a'Va \end{aligned} \qquad \ldots(28)$$

and where (26) is trivially satisfied for $a=0$. The inequality (27) is satisfied if and only if

$$\max_a t^2(a) \leqq S^2. \qquad \ldots(29)$$

Thus, the theorem is proved if we show that the maximum of the squared t ratio $t^2(a)$ is distributed as $qF(q, T-k)$.

We maximize the squared t ratio $t^2(a)$ subject to the normalizing constraint $a'Va=1$ since $t^2(a)$ is not affected by a change of scale of the elements of a. Therefore we form the Lagrangian

$$L(a, \lambda) = [a'(z-\theta)/s]^2 - \lambda(a'Va-1) \qquad \ldots(30)$$

where λ is the Lagrange multiplier. Setting the derivative of $L(a, \lambda)$ with respect to a equal to zero gives

$$[(z-\theta)(z-\theta)' - \lambda s^2 V]a = 0. \qquad \ldots(31)$$

Premultiplying (31) by a' and dividing by $s^2 a'Va$ shows that $\lambda = t^2(a)$. Hence the determinantal equation

$$|(s^2V)^{-1}(z-\theta)(z-\theta)' - \lambda I| = 0 \qquad \ldots(32)$$

is solved for the greatest root λ^*. Since (32) has only one non-zero root–the matrix $(z-\theta)(z-\theta)'$ has rank one–the greatest root is

$$\begin{aligned} \lambda^* &= \text{trace } (s^2V)^{-1}(z-\theta)(z-\theta)' \\ &= (z-\theta)'(s^2V)^{-1}(z-\theta) \end{aligned} \qquad \ldots(33)$$

which is distributed as $qF(q, T-k)$. The solutions to (31) are proportional to $(s^2V)^{-1}(z-\theta)$ and the normalized solution is $a^* = V^{-1}(z-\theta)/(s^2\lambda^*)^{\frac{1}{2}}$.

We now show the relation between the S procedure and the F test of H. Let a_0 be the vector a which maximizes $t_0^2(a)$, i.e. $t^2(a)$ with $\theta=0$. From (33) it follows that

$$t_0^2(a_0) = z'(s^2V)^{-1}z \qquad \ldots(34)$$

which is distributed as $qF(qT-k)$ when H is true. In the S procedure we accept

$$H(a_0): \psi_0 = a'_0\theta = 0 \qquad \ldots(35)$$

if and only if

$$t_0^2(a_0) \leqq S^2. \qquad \ldots(36)$$

Hence the α level Scheffé separate induced test of H accepts H if and only if the α level F test of H accepts H since we accept the separate hypotheses $H(a)$ for all ψ in L if and only if we accept the hypothesis $H(a_0)$.

In terms of S intervals the F test of H looks at ψ_0 and accepts $H(a_0)$ if and only if the S interval for ψ_0 covers zero. Since the S intervals for all ψ in L cover zero if and only if the S interval for ψ_0 covers zero, it follows that the α level F test of H accepts H if and only if for all ψ in L the S intervals cover zero. The intersection of the S intervals for all ψ in L is the confidence region (6), which is an ellipsoidal region in θ space with centre at z.

An estimate of $\hat{\psi}$ of ψ is said to be sdfz if the S interval does not cover $\psi = 0$, i.e. if $|\hat{\psi}| > S\hat{\sigma}_{\hat{\psi}}$. Hence H is rejected if and only if the estimate of at least one ψ in L is sdfz according to the S criterion.

We can begin the S procedure by first calculating the S interval for ψ_0. If this interval covers zero then there is no need to calculate other S intervals since all other S intervals will also cover zero. If ψ_0 is sdfz we will want to examine the S intervals to find $\hat{\psi}$ which are sdfz. Beginning with the S interval for ψ_0 is the same as beginning with the F test of H. If the F test of H rejects H we will want to find which $\hat{\psi}$ are sdfz. Since a_0 can be calculated from (31) we can always find at least one linear combination which is responsible for rejection namely ψ_0. Of course, there is no guarantee that ψ_0 has an economic interpretation of interest. In practice, we generally begin with the F test since computer programmes for regression analysis generally calculate the F statistic, but not the S interval for ψ_0.

When the hypothesis H is that all the slope coefficients are zero the components of the a_0 vector have a simple statistical interpretation. Suppose that the first column of X is a column of ones and let D be the $T \times (k-1)$ matrix of deviations of the regressors (excluding unity) from their means. Since z is simply the least squares estimator of the slope coefficients, $z = (D'D)^{-1}D'y$. Hence $a_0 = (D'D)z/(s^2qF)^{\frac{1}{2}} = D'y/(s^2qF)^{\frac{1}{2}}$ so that the components of a_0 are proportional to the sample covariances between the dependent variable and the regressors. If the columns of D are orthogonal, then the components of a_0 are proportional to the least squares estimates of the slope coefficients, i.e. z. Thus in the orthogonal case $\hat{\psi}_0$ is proportional to the sum of the squares of the slope coefficients.

For an example of the S procedure we again turn to the case where $q = 2$ and $\sigma^2 V = I$ which is known. The acceptance region of a 0·05 level separate induced test of H when all ψ in L are of primary interest is the circular region of the 0·05 level direct (χ^2) test shown in Figure 1. Recall that the square region in Figure 1 is the acceptance region of a 0·05 level Bonferroni separate induced test of H when the only ψ in L of primary interest are $\psi_1 = \theta_1$ and $\psi_2 = \theta_2$. As noted earlier these two acceptance regions can produce conflicting inferences and hence the same is true for the Bonferroni and the Scheffé separate induced tests of H.

3.4. *An Extension of the* S *Procedure*

The S procedure has been extended by Gabriel ((1964) and (1969)) to include testing multivariate hypotheses

$$H_*: R_*\beta - r_* = \theta_* = 0 \qquad \ldots(37)$$

where $[R_*r_*]$ consists of any q_* rows of $[Rr]$ defined in (2). Let F_* be the F statistic for testing H_* and let $t_0^2(a_*)$ be the squared t ratio for testing

$$H_*(a_*): a'_*(R_*\beta - r_*) = a'_*\theta_* = 0 \qquad \ldots(38)$$

where $a_* \in R^q$. With no loss of generality we may let $[R_*r_*]$ consist of the last q_* rows of $[Rr]$. From the development (30) to (33) we find that

$$\max_{a_*} t_0^2(a_*) = q_*F_* = \max_{a \in I} t_0^2(a) \qquad \ldots(39)$$

where I is the set of all non-null a vectors such that the first $q - q_*$ elements are zero. Hence

$$q_*F_* \leqq qF \qquad \ldots(40)$$

since the constrained maximum of $t_0^2(a)$ is less than or equal to the unconstrained maximum. This establishes that when H is true the probability is $1-\alpha$ that the inequality

$$q_* F_* \leqq q F_\alpha(q, T-k) = S^2 \qquad \ldots(41)$$

is simultaneously satisfied for all hypotheses H_* defined in (37) where F_* is the F statistic for testing H_*.

The implication is that using the S procedure, i.e. acceptance region (41), we can test any number of multivariate hypotheses H_0 with the assurance that all will simultaneously be accepted with probability $1-\alpha$ when the hypothesis H is true. The hypotheses H_* may be suggested by the data. When we begin the S procedure with an α level F test of H, this is a special case of H_* when $q_* = q$. For further discussion see Scheffé (1977*a*).

3.5. *The Lengths of the* B *and* S *Intervals for* ψ *of Primary Interest*

The ratio of the length of the B intervals to the length of the S intervals for ψ in G is simply the ratio of B to S. When q is fixed, S is a constant and B increases as m increases so that for sufficiently large m the B intervals are longer for all ψ in G than the S intervals. On the other hand, for sufficiently small m the B intervals for all ψ in G are shorter than the S intervals. Suppose $q=2$ and $\sigma^2 V = I$ which is known. If G consists of $m=4$ linear combinations and if nominally $\alpha = 0{\cdot}05$, then applying the Bonferroni inequality gives $B = 2{\cdot}50$. Since $S = 2{\cdot}448$ the S intervals are shorter than the B intervals for all ψ in G; the ratio of B to S is $1{\cdot}02$. If G consists of $m=2$ linear combinations and if $\alpha = 0{\cdot}05$ exactly, then from Figure 1 we see that $B = 2{\cdot}236$ so that the ratio of B to S is $0{\cdot}913$. If instead of calculating the exact 95 per cent confidence region we use the Bonferroni inequality, then $B = 2{\cdot}241$ which is also less than S. See Figures 4 and 5 in Miller (1966, pp. 15–16).

In the case where $m=q$ and $\alpha = 0{\cdot}05$ calculations by Christensen (1973) show that the B intervals are shorter than the S intervals regardless of the size of q. Similar results are reported by Morrison (1976, p. 136) for 95 per cent Bonferroni and Roy–Bose simultaneous confidence intervals on means. The Roy–Bose simultaneous confidence intervals are the same as S intervals in the case of the classical linear normal regression model. Games (1977) has calculated the maximum number of ψ of primary interest (the number m) such that the B intervals (applying the Šidák inequality) are shorter than the S intervals for all ψ in G.

4. SIMULTANEOUS CONFIDENCE INTERVALS AND THE GENERALIZED B PROCEDURE

4.1. *Confidence Sets and Simultaneous Confidence Intervals*

Suppose after inspecting the data we wish to make inferences about linear combinations of secondary interest. We now consider how the B procedure can be generalized so that inferences can be made about all ψ in L.

Scheffé (1959, pp. 81–83) has suggested a more general approach to methods of multiple comparison. Let G be a set of ψ in L of primary interest and suppose we have a multiple comparison procedure which gives for each ψ in G an interval

$$\hat{\psi} - h_\psi s \leqq \psi \leqq \hat{\psi} + h_\psi s \qquad \ldots(42)$$

where h_ψ is a constant depending on the vector a but not the unknown θ. The inequality (42), which may be written

$$|a'(\theta - z)| \leqq h_\psi s, \qquad \ldots(43)$$

can be interpreted geometrically to mean that the point θ lies in a strip of the q dimensional space between two parallel planes orthogonal to the vector a, the point z being midway between the planes. The intersection of these strips for all ψ in G determines a certain convex set C and (43) holds for all ψ in G if and only if the point θ lies in C. Thus the problem of multiple comparison can be approached by starting with a convex confidence set C instead of a set G of ψ in L. From any convex set C we can derive simultaneous confidence intervals for the infinite set of all ψ in L by starting with the relation that the point θ lies in C if and only if it lies between a parallel pair of supporting planes of C.

Let L^* be the set of ψ in L for which $a'Va = 1$ and G^* be a set of m linear combinations ψ in L^* of primary interest. This normalization is convenient since the B intervals for all ψ in G^* have length $2Bs$ and the S intervals for all ψ in L^* have length $2Ss$. We now define the confidence set C of the B procedure to be the intersection of the B intervals for all ψ in G^* and the set C of the S procedure to be the intersection of the S intervals for all ψ in L^*. In the B procedure C is a polyhedron, and in the S procedure C is the confidence ellipsoid defined by (6). When $q = 2$ the region C is a polygonal region in the B procedure and an elliptical region in the S procedure. In addition, if $m = 2$ and if $\sigma^2 V = I$, then C is a square region in the B procedure and circular region in the S procedure, as depicted in Figure 1.

Consider the case where the confidence region C is a square with sides $2Bs$. Starting with a square we can derive simultaneous confidence intervals for all ψ in L^*, not just for θ_1

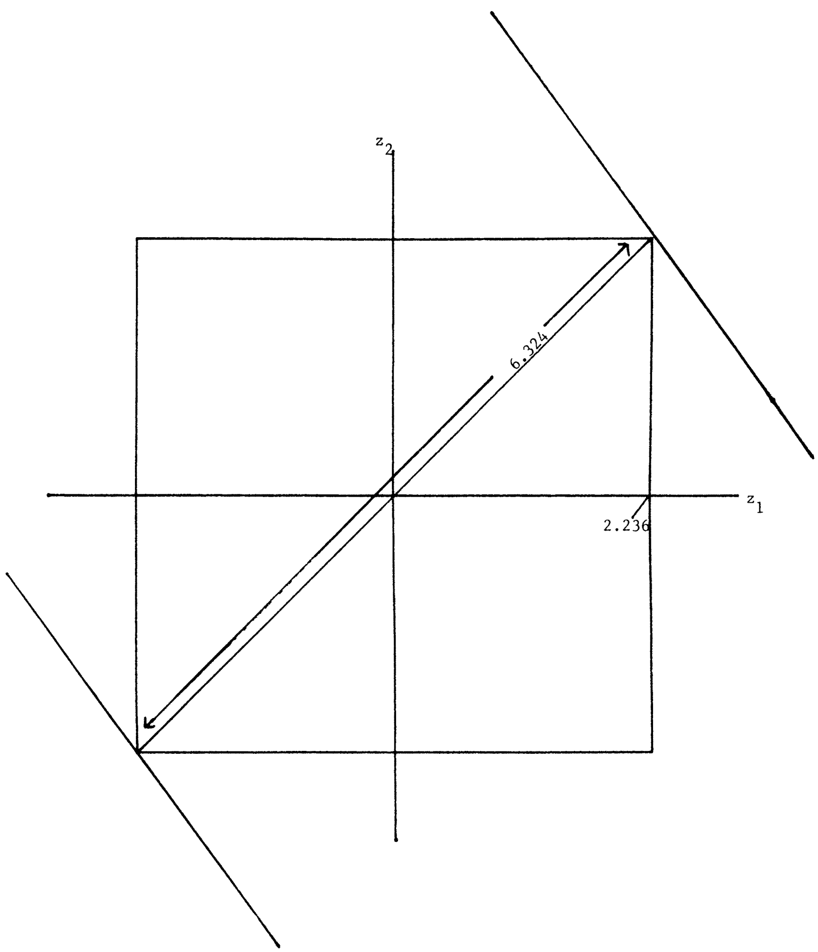

FIGURE 3

and θ_2. The square has four extreme points, which are the four corner points. There are only two pairs of parallel lines of support where each supporting line contains two extreme points. These two pairs of lines define the B intervals for the ψ of primary interest, i.e. θ_1 and θ_2, respectively, and contain all the boundary points of the square. In addition to these two pairs of parallel lines of support there are an infinite number of pairs of parallel lines of support where each line contains only one extreme point. One such pair is shown in Figure 3. This pair defines a simultaneous confidence interval for some ψ of secondary interest. We can derive a simultaneous confidence interval for every ψ of secondary interest by taking into account pairs of supporting lines where each line contains only one extreme point.

4.2. *The Generalized* B *Procedure*

Let G be a set of finite number m of ψ of primary interest and as before denote the ψ in G by $\psi_i = a_i'\theta$, $i = 1, 2, \ldots, m$, $a_i \in R^q$. Any ψ in L is a linear combination $c_1\psi_1 + c_2\psi_2 + \ldots + c_m\psi_m$ of the ψ of primary interest where $c = (c_1, c_2, \ldots, c_m)'$ and $c \in R^m$. The generalized B procedure is based on the following theorem which is proved by Richmond (1979, Theorem 1, p. 5). The probability is $\geqq 1-\alpha$ that simultaneously for all ψ in L

$$\hat{\psi} - B\sum_{i=1}^{m}(s^2 a_i' V a_i)^{\frac{1}{2}}|c_i| \leqq \psi \leqq \hat{\psi} + B\sum_{i=1}^{m}(s^2 a_i' V a_i)^{\frac{1}{2}}|c_i| \qquad \ldots(44)$$

where $B = t_{\delta/2}(T-k)$, when applying the Bonferroni inequality given $\delta = \alpha/m$, and these intervals are called B intervals. When $c = (0, \ldots, 0, 1, 0, \ldots, 0)'$, the 1 occurring in the ith place, the B interval is for ψ_i, a ψ of primary interest.

As an example of the generalized B procedure suppose that $q = 2$ and $\sigma^2 V = I$ which is known. Consider the B interval for $\psi = (1/2)^{\frac{1}{2}}(\psi_1 + \psi_2) = (1/2)^{\frac{1}{2}}(\theta_1 + \theta_2)$. Suppose $G = m = 2$ so that ψ is of secondary interest. If $\alpha = 0{\cdot}05$ exactly the length of the B interval is $2(2)(1/2)^{\frac{1}{2}}(2{\cdot}236) = 6{\cdot}324$ which is the case shown in Figure 3. Applying the Bonferroni inequality given $\delta = 0{\cdot}05/2$ the length of the B interval is $2(2)(1/2)^{\frac{1}{2}}(2{\cdot}24) = 6{\cdot}336$. Now suppose $m = 3$ and ψ is of primary interest. Applying the Bonferroni inequality given $\delta = 0{\cdot}05/3$ we have $B = 2{\cdot}39$ so that the length of the B interval for ψ is $2(2{\cdot}39) = 4{\cdot}78$, which is considerably less than when ψ is of secondary interest. This shows that the length of a B interval for a ψ in L can vary considerably depending on whether ψ is of primary or secondary interest.

4.3. *The Tradeoff Between the* B *and* S *Procedures*

In general the B intervals are shorter for the ψ of primary interest and the S intervals are shorter for least some ψ of secondary interest. Hence there is a tradeoff between the generalized B procedure and the S procedure. It is instructive to compare the length of the simultaneous confidence intervals derived from the square region with sides $2B = 4{\cdot}472$ with the intervals derived from the circular region with diameter $2S = 4{\cdot}895$. The generalized B procedure is the procedure which gives for each ψ in L^* an interval derived from the square region. The B intervals for ψ in L^* include the B intervals for θ_1 and θ_2, respectively, which are the ψ of primary interest. The length of the shortest B interval is equal to the length of the side of the square region and the length of the longest B interval is equal to the length of the diagonal which is $6{\cdot}324$. Since the length of the S intervals for all ψ in L^* is $4{\cdot}895$ the S intervals are shorter than the B intervals for some ψ in L^*; in particular, the S interval is shorter for $\psi = (1/2)^{\frac{1}{2}}(\theta_1 + \theta_2)$, the B interval for this ψ being the one shown in Figure 3.

As noted earlier, when G is finite there are a few cases in the one-way lay-out of the analysis of variance where the exact significance level of the separate induced test of H can be easily calculated. In these cases the probability that simultaneously for all ψ in L the confidence intervals cover the true values can also be easily calculated. These cases include the generalized Tukey procedure (see Scheffé (1959, Theorem 2, p. 74)) where the

ψ of primary interest are the pairwise comparisons $(\theta_i - \theta_j)$, $i, j = 1, 2, \ldots, q$, $i \neq j$, and the "extended Dunnett procedure" developed by Schaffer (1977) where the ψ of primary interest are the differences $(\theta_1 - \theta_i)$, $i = 2, \ldots, q$. Schaffer (1977) found that the Tukey intervals are shorter than the S intervals for the ψ of primary interest in the generalized Tukey procedure and likewise that the Dunnett intervals are shorter than the S intervals for the ψ of primary interest in the extended Dunnett procedure. On the other hand, the S procedure generally gives shorter intervals for the ψ of secondary interest. Richmond (1979) obtained similar results when extending the Schaffer study to include the generalized B procedure where the ψ of primary interest in this procedure are taken to be the same as in the extended Dunnett procedure and the B intervals are calculated by applying the Sidák inequality. For further comparisons between Tukey and S intervals see Scheffé (1959, pp. 75–77) and Hochberg and Rodriqúez (1977).

The S procedure is well adapted to "data snooping", i.e. making inferences about ψ that are suggested by the data as opposed to the relatively small set of ψ of primary interest which are selected before the experiment is performed. Scheffé (1959, p. 80) discusses a method which permits the data snooping valid under the S procedure while avoiding the wide S intervals for the ψ of primary interest. This consists of allocating beforehand the chosen α into components $\alpha_1, \alpha_2, \ldots, \alpha_m$ for the B intervals for the ψ of primary interest and a component α_0 for the S intervals so that $\alpha = \sum_{i=0}^{m} \alpha_i$. The probability that all the B intervals and S intervals simultaneously cover the true values is then $\geqq 1 - \alpha$.

5. LARGE SAMPLE B AND S PROCEDURES

Large sample analogues of the generalized B procedure and the S procedure can be employed in a general regression framework. Consider the model

$$y_t = g(X_t : \beta) + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \qquad \ldots(45)$$

where X_t is a $k \times 1$ vector which may contain lagged dependent variables and where the ε_t are independent $N(0, \sigma^2)$. Suppose we wish to test the hypothesis

$$H : R\beta - r = \theta \qquad \ldots(46)$$

where R and r are defined as in (2) and let L be the set of $\psi = a'\theta$ such that $a \in R^q$. From any convex set C we can derive simultaneous confidence intervals for all ψ in L. Starting with a finite set G of ψ in L of primary interest the set C can be defined as the intersection of large sample Bonferroni t intervals for all ψ in G. The set C can also be the acceptance region of a large sample direct test such as a Wald, likelihood ratio or Lagrange multiplier test. A large sample analogue of the S procedure is based on the Wald test of H.

As an example consider model (1) where the assumption of non-stochastic regressors is replaced by the assumption that the last regressor is a lagged dependent variable and where the assumption of independent disturbances is replaced by the assumption that the disturbances u_t follow a first order autoregressive process

$$u_t = \rho u_{t-1} + \varepsilon_t, \; |\rho| < 1, \qquad t = 0, \pm 1, \pm 2, \ldots, \qquad \ldots(47)$$

where the ε_t are independent $N(0, \sigma^2)$. The unrestricted maximum likelihood (ML) estimators of β, σ^2 and ρ are denoted by b, $\hat{\sigma}^2$ and $\hat{\rho}$, respectively. To obtain the ML estimators we may use the search procedure described in Dhrymes (1971, pp. 69–70).

Suppose the conditions of Dhrymes Theorem 7.1 (1971, p. 199) are satisfied. Then we have asymptotically that $T^{\frac{1}{2}}(b - \beta)$ is $N(0, \sigma^2 \Omega^{-1})$ where Ω is given by equation (7.83) in Dhrymes (1971, p. 200). It follows that asymptotically

$$T^{\frac{1}{2}}[(Rb - r) - \theta] = T^{\frac{1}{2}}(z - \theta) \qquad \ldots(48)$$

is $N(0, \sigma^2 V)$ where $V = R\Omega^{-1}R'$. A consistent estimator of V is $\hat{V} = R\hat{\Omega}^{-1}R'$ where $\hat{\Omega}$ is

Ω with the unknown parameters β and ρ replaced by their ML estimators b and $\hat{\rho}$. Hence, asymptotically $T(z-\theta)'[\hat{\sigma}^2\hat{V}]^{-1}(z-\theta)$ is $\chi^2(q)$.

Let G be a set of a finite number m of ψ in L of primary interest and denote the ψ in G by $\psi_i = a_i'\theta$, $i = 1, 2, \ldots, m$. Suppose we wish to test the separate hypotheses

$$H(a_i): \psi_i = a_i'\theta, \qquad i = 1, 2, \ldots, m, \qquad \ldots(49)$$

and that a test of $H(a_i)$ is defined by the large sample acceptance region

$$|t_0(a_i)| \leqq B, \qquad i = 1, 2, \ldots, m, \qquad \ldots(50)$$

where $t_0^2(a_i)$ is the large sample squared t ratio

$$t_0^2(a_i) = T[a_i'z]^2/\hat{\sigma}^2 a_i'\hat{V}a_i, \qquad i = 1, 2, \ldots, m, \qquad \ldots(51)$$

and where B is the upper $\delta/2$ significance point of a $N(0, 1)$ distribution. The large sample Bonferroni separate induced test of H accepts H if and only if all the separate hypotheses $H(a_1), H(a_2), \ldots, H(a_m)$ are accepted. When applying the Bonferroni inequality given $\delta = a/m$ the probability is asymptotically $\geqq 1-\alpha$ that the separate induced test of H accepts H when H is true. The large sample analogue of the B procedure exploits the fact that $T^{\frac{1}{2}}(\hat{\psi}-\psi) = T^{\frac{1}{2}}a'(z-\theta)$ is asymptotically $N(0, \sigma^2 a'Va)$. Large sample simultaneous confidence intervals for the ψ of primary and secondary interest can be developed using a large sample analogue of the generalized B procedure.

The large sample analogue of the S procedure is developed in association with the Wald test of H. The Wald statistic for testing H is

$$W = Tz'\hat{V}^{-1}z/\hat{\sigma}^2. \qquad \ldots(52)$$

For a large sample Wald test of H at level α the natural choice of acceptance region is

$$W \leqq \chi_\alpha^2(q). \qquad \ldots(53)$$

The large sample analogue of the S procedure is based on the following theorem. The probability is asymptotically $1-\alpha$ that simultaneously for all ψ in L

$$\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \leqq \psi \leqq \hat{\psi} + S\hat{\sigma}_{\hat{\psi}} \qquad \ldots(54)$$

where now $S = (\chi_\alpha^2(q))^{\frac{1}{2}}$ and $\hat{\sigma}_{\hat{\psi}}^2 = \hat{\sigma}^2 a'\hat{V}a$. The proof consists in showing that the probability is asymptotically $1-\alpha$ that

$$\max_a t(a) \leqq S^2 \qquad \ldots(55)$$

where $t^2(a)$ is the large sample squared t ratio

$$t^2(a) = T[a'(z-\theta)]^2/\hat{\sigma}^2 a'\hat{V}a. \qquad \ldots(56)$$

By an argument essentially the same as that of (30) to (33) it can be shown that

$$\max_a t^2(a) = T(z-\theta)'[\hat{\sigma}^2\hat{V}]^{-1}(z-\theta). \qquad \ldots(57)$$

Since $\max_a t^2(a)$ is asymptotically distributed as $\chi^2(q)$ this completes the proof.

The relation between the Wald test of H and the large sample S procedure is that the large sample α level Wald test of H accepts H if and only if for all ψ in L the large sample S intervals cover zero.

In the case of a higher order disturbance process we may wish to test linear restrictions on the ρ's as well as the β's. For example we can use the S procedure to investigate the autoregressive disturbances process as well as the lag structure of the regression equation.

6. EMPIRICAL EXAMPLES

6.1. *Textile Example*

Our first empirical illustration is based on the textile example of Theil (1971, Chapter 3, p. 103). This example considers an equation of the consumption of textiles in the Netherlands 1923–1939:

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u$$

where y = logarithm of per-capita textile consumption, x_1 = logarithm of real per-capita income and x_2 = logarithm of the relative price of textile goods. The estimated equation is reported by Theil (p. 116) as

$$\underset{(0\cdot 31)}{y=1\cdot 37}+\underset{(0\cdot 16)}{1\cdot 14x_1}-\underset{(0\cdot 04)}{0\cdot 83x_2}$$

where the numbers in parentheses are standard errors.

Theil tests the hypothesis that the income elasticity (β_1) is unity, and that the price elasticity (β_2) is minus unity. This hypothesis is

$$H: R\beta-r=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}-\begin{bmatrix}1\\ -1\end{bmatrix}$$

$$=\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}=\theta=0.$$

As a direct test we use the F test. The 0·01 level F test rejects H since the value of the F ratio is 11·2 and the upper 1 per cent significance point of an $F(2, 14)$ distribution is 6·51.

Consider the Bonferroni separare induced test of H where the linear combination of primary interest are θ_1 and θ_2. The t ratio's for θ_1 and θ_2 are

$$t_1=z_1/(s^2V_{11})^{\frac{1}{2}}=0\cdot 1430/0\cdot 16=0\cdot 89$$

and

$$t_2=z_2/(s^2V_{22})^{\frac{1}{2}}=0\cdot 1711/0\cdot 04=4\cdot 28,$$

respectively. The nominal 0·01 level separate induced test of H rejects H since $B=t_\delta(14)=2\cdot 51$ when $\delta=0\cdot 01/4=0\cdot 0025$. Clearly the separate hypothesis $\beta_2=-1$ is responsible for the rejection of the separate induced test of H. The 0·01 level Scheffé separate induced test of H also rejects H since the 0·01 level F test rejects H.

We now calculate simultaneous confidence intervals for θ_1 and θ_2. The B interval for θ_1 is $0\cdot 1430\pm 0\cdot 16(2\cdot 51)$ and for θ_2 is $0\cdot 1711\pm 0\cdot 04(2\cdot 51)$ so that the B intervals are $-0\cdot 26\leqq\theta_1\leqq 0\cdot 54$ and $0\cdot 07\leqq\theta_2\leqq 0\cdot 27$, respectively. The S interval for θ_1 is $0\cdot 1430\pm$ $16(3\cdot 61)$ and for θ_2 is $0\cdot 1771\pm 0\cdot 04(3\cdot 61)$ since $S=(2F_{0\cdot 01}(2, 14))^{\frac{1}{2}}=3\cdot 61$. Hence the S intervals are $-0\cdot 43\leqq\theta_1\leqq 0\cdot 72$ and $0\cdot 03\leqq\theta_2\leqq 0\cdot 32$, respectively. We see that the S intervals are longer than the B intervals, both intervals for θ_1 cover zero and both intervals for θ_2 cover only positive values. We infer that the income elasticity β_1 is unity and that the price elasticity β_2 is greater than minus one. In this example the hypothesis $\beta_2=-1$ is responsible for the rejection of the Scheffé as well as the Bonferroni separate induced test of H. This result also follows from the fact that the absolute value of the t statistic for θ_2, $|t_2|$, is larger than either B or S, i.e., $|t_2|>B$ and $|t_2|>S$.

Next we calculate the normalized a vector

$$a_0=[R(X'X)^{-1}R']^{-1}(Rb-r)/(s^2qF)^{\frac{1}{2}}$$

where $a_0'Va_0 = 1$. From Theil we have that

$$s^2[R(X'X)^{-1}R']^{-1} = \begin{bmatrix} 43\cdot2 & 41\cdot6 \\ 41\cdot6 & 807\cdot0 \end{bmatrix}$$

so that

$$a_0 = \frac{1}{s(4\cdot733)} \begin{bmatrix} 43\cdot2 & 41\cdot6 \\ 41\cdot6 & 807\cdot0 \end{bmatrix} \begin{bmatrix} 0\cdot1430 \\ -0\cdot1711 \end{bmatrix} = \begin{bmatrix} 207\cdot5 \\ 2247\cdot7 \end{bmatrix}$$

where $s^2 = 0\cdot0001833$. This confirms Theil's conclusions (p. 145) that the specification $\beta_2 = -1$ for the price elasticity is mainly responsible for the F test (Scheffé separate induced test) of H rejecting H, i.e. any linear combination with a sufficiently large weight on θ_2 is responsible for rejection.

Suppose in the B procedure that $\psi = \theta_1 - \theta_2$ is of secondary interest. The B interval for ψ is $0\cdot3141 \pm 0\cdot20(2\cdot51)$ or $-0\cdot19 \leqq \psi \leqq 0\cdot82$. The S interval for ψ is $0\cdot3141 \pm 0\cdot023(3\cdot61)$ or $0\cdot23 \leqq \psi \leqq 0\cdot40$ so that the S interval is shorter than the B interval. Also notice that $\hat{\psi} = z_1 - z_2$ is sdfz according to the S criterion, but not the B criterion. Hence the Scheffé separate induced test of H is rejected by the separate hypothesis that the income and price elasticities are the same except for sign: $\beta_1 = -\beta_2$. Theil (p. 134) objects to the length of the S intervals for the ψ of primary interest. In fact in the textile example the S intervals give interesting results for both the ψ of primary and secondary interest.

6.2. *Klein's Model I Example*

Our second example is based on the unrestricted reduced form equation for consumption expenditures from Klein's Model I of the United States economy 1921–1941:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + u$$

where y = consumption, x_1 = government wage bill, x_2 = indirect taxes, x_3 = government expenditures, x_4 = time (measured as year -1931), x_5 = profits lagged one year, x_6 = end of year capital stock lagged one year and x_7 = private product lagged one year. For the purpose of this example all regressors are treated as non-stochastic. The data is taken from Theil (1971, p. 456). The estimated equation is

$$\begin{array}{l} y = 58\cdot3 + 0\cdot193x_1 - 0\cdot366x_2 + 0\cdot205x_3 + 0\cdot701x_4 + 0\cdot748x_5 - 0\cdot147x_6 + 0\cdot230x_7 \\ \quad\;\; (1\cdot90)\ (0\cdot079)(-0\cdot871)\quad (0\cdot541)\quad (0\cdot930)\quad (1\cdot49)\quad (-1\cdot27)\quad (0\cdot842) \end{array}$$

where now the numbers in parentheses are t ratios. Our estimates of the β's agree with those reported in Goldberger (1964, p. 325). (Note that Goldberger uses $x_1 - x_3$ in place of x_1 so that his estimate of β_1 is $0\cdot19327 - 0\cdot20501 = -0\cdot01174$.)

Consider testing the hypothesis that all the slope coefficients are zero:

$$H: \beta_i = \theta_i = 0, \qquad i = 1, 2, \ldots, 7.$$

The slope coefficients are multipliers so we are testing the hypothesis that all the multipliers in the reduced form equation for consumption are zero. The $0\cdot05$ level Scheffé separate induced test of H rejects H since the $0\cdot05$ level F test overwhelmingly rejects H. The F ratio is $28\cdot2$ which is much larger than $2\cdot83$, the upper $0\cdot05$ significance point of the $F(7, 13)$ distribution. Suppose that the linear combinations of primary interest in the B procedure are the slope coefficients: $\psi_i = \theta_i$, $i = 1, 2, \ldots, 7$. Then the critical t value for a nomial $0\cdot05$ level Bonferroni separate induced test of H is $B = t_{\delta/2}(13) = 3\cdot20$ where $\delta = 0\cdot05/14 = 0\cdot00357$. The t ratio with the largest absolute value is the one for lagged profits (β_5). Since this is only $1\cdot49$ the Bonferroni separate induced test of H overwhelmingly accepts H. Thus in this example the Scheffé and Bonferroni separate induced test of H produce conflicting inferences.

We now apply the S procedure to find which linear combination of the multipliers led to rejection of the Scheffé separate induced test of H. In this example none of the individual multipliers are responsible for rejection since none of the t ratios have an absolute value greater than S. The largest t ratio is $1{\cdot}49$ and $S=(7F(7, 13))^{\frac{1}{2}}=4{\cdot}45$. To find linear combinations of the multipliers which are responsible for rejection we calculate the normalized vector a_0. This vector has components

$$a_1=5{\cdot}82, \quad a_2=4{\cdot}81, \quad a_3=7{\cdot}37, \quad a_4=19{\cdot}44$$
$$a_5=12{\cdot}13, \quad a_6=14{\cdot}33, \quad a_7=35{\cdot}84$$

where these are proportional to the sample covariances between the dependent variable and the regressors. It appears that the linear combinations which are responsible for rejection give some positive weight to all the multipliers and especially to the multiplier (β_7) for lagged private product. In this example the linear combinations responsible for rejection of the Scheffé separate induced test of H do not seem to have an interesting economic interpretation.

7. CONCLUDING COMMENTS

If we are only interested in a direct test of H, then we should use the direct test which has the greatest power against the alternatives of interest. This direct test may be the F test or some other test. For example, if the critical region of a Bonferroni separate induced test has the greatest power against the alternatives of interest, then this test can be employed as a direct test. Similarly, the F test can be be used as a direct test or as the first stage in the Scheffé separate induced test procedure. This is because the acceptance region of the F test of H is the same as the acceptance region of the Scheffé separate induced test of H. Hence before performing an F test we should decide whether we have a two decision or a multiple decision problem. If we have a multiple decision problem, then the relative lengths of simultaneous confidence intervals are of interest. When some linear combinations are of primary interest while others are only of secondary interest the B procedure may be preferred to the S procedure. This is because the B intervals are generally shorter than the S intervals for the linear combinations of primary interest. On the other hand, if we are interested in data snooping the S procedure may be preferable. The trade-off between the lengths of the B and S intervals depends on the particular regression problem and the linear combinations of primary interest. Since model building in economics involves multiple comparisons there appears to be no shortage of opportunities to use either the B or the S procedure.

I wish to thank G. B. A. Evans, R. W. Farebrother and J. Richmond as well as D. F. Hendry, D. W. Jorgenson, G. E. Mizon, J. D. Sargan and J. Stewart for helpful comments. I am especially indebted to an anonymous referee for many valuable suggestions.

REFERENCES

CHRISTENSEN, L. R. (1973), "Simultaneous Statistical Inference in the Normal Multiple Linear Regression Model", *Journal of the American Statistical Association*, **68**, 457–461.

CHRISTENSEN, L. R., JORGENSON, D. W. and LAU, L. J. (1975), "Transcendental Logarithmic Utility Functions", *American Economic Review*, **65**, 367–383.

DARROCH, J. N. and SILVEY, S. D. (1963), "On Testing More Than One Hypothesis", *Annals of Mathematical Statistics*, **27**, 555–567.

DHRYMES, P. J. (1971) *Distributed Lags* (San Francisco: Holden Day).

DUNN, O. J. (1961), "Multiple Comparisons Among Means", *Journal of the American Statistical Association*, **56**, 52–64.

DUNNETT, C. W. (1955), "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control", *Journal of the American Statistical Association*, **50**, 1096–1121.

GABRIEL, K. R. (1964), "A Procedure for Testing the Homogeneity of All Sets of Means in the Analysis of Variance", *Biometrics*, **40**, 459–477.

GABRIEL, K. R. (1969), "Simultaneous Test Procedures–Some Theory of Multiple Comparisons", *Annals of Mathematical Statistics*, **40**, 224–250.

GAMES, P. A. (1977), "An Improved Table for Simultaneous Control on g Contrasts", *Journal of the American Statistical Association*, **72**, 531–534.

GOLDBERGER, A. S. (1964) *Econometric Theory* (New York: John Wiley and Sons).

HOCHBERG, Y. and RODRIQUEZ, G. (1977), "Intermediate Simultaneous Inference Procedures", *Journal of the American Statistical Association*, **72**, 220–225.

JORGENSON, D. W. (1971), "Econometric Studies of Investment Behaviour: A Survey", *Journal of Economic Literature*, **9**, 1111–1147.

JORDENSON, D. W. (1974), "Investment and Production: A Review", Intriligator, M. D. and Kendrick, D. A. (eds.) *Frontiers in Quantitative Economics*, II (Amsterdam: North Holland).

JORGENSON, D. W. and LAU, L. J. (1975), "The Structure of Consumer Preferences", *Annals of Social and Economic Measurement*, **4**, 49–101.

LEHMANN, E. L. (1957*a*), "A Theory of Some Multiple Decision-Problems, I", *Annals of Mathematical Statistics*, **28**, 1–25.

LEHMANN, E. L. (1957*b*), "A Theory of Some Multiple Decision Problems, II", *Annals of Mathematical Statistics*, **28**, 547–572.

MILLER, R. G. Jr. (1966) *Simultaneous Statistical Inference* (New York: McGraw-Hill).

MILLER, R. G., Jr. (1977), "Developments in Multiple Comparisons, 1966–1976", *Journal of the American Statistical Association*, **72**, 779–788.

MORRISON, D. F. (1976) *Multivariate Statistical Methods*, 2nd Edition (New York: McGraw-Hill).

MOSES, L. E. (1976), "Charts for Finding Upper Percentage Points of Student's t in the Range ·01 to ·0001" (Technical Report No. 24 (5 Rol GM 21215.02), Stanford University).

RICHMOND, J. (1979), "A General Method for Constructing Simultaneous Confidence Intervals" (Department of Economics Discussion Paper, University of Essex).

ROY, S. N. (1953), "On A Heuristic Method of Test Construction and its Uses in Multivariate Analysis", *Annals of Mathematical Statistics*, **24**, 220–239.

ROY, S. N. and BOSE, R. C. (1953), "Simultaneous Confidence Interval Estimation", *Annals of Mathematical Statistics*, **24**, 415–536.

SARGAN, J. D. (1976), "The Consumer Price Equation in the Post War British Economy: An Exercise in Equation Specification Testing" (Mimeo, London School of Economics, 1976).

SCHAFFER, J. P. (1977), "Multiple Comparisons Emphasizing Selected Contrasts: An Extension and Generalization of Dunnett's Procedure", *Biometrics*, **33**, 293–303.

SCHEFFÉ, H. (1953), "A Method for Judging All Contrasts in the Analysis of Variance", *Biometrika*, **40**, 87–104.

SCHEFFÉ, H. (1959) *The Analysis of Variance* (New York: John Wiley and Sons, Inc.).

SCHEFFÉ, H. (1977*a*), "A Note on a Reformulation of the S-Method of Multiple Comparison", *Journal of the American Statistical Association*, **72**, 143–144.

SCHEFFÉ, H. (1977*b*), "A Note on a Reformulation of the S-Method of Multiple Comparison–Rejoinder", *Journal of the American Statistical Association*, **72**, 146.

SEBER, G. A. F. (1964), "Linear Hypotheses and Induced Tests", *Biometrika*, **51**, 41–47.

ŠIDÁK, Z. (1967), "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions", *Journal of the American Statistical Association*, **62**, 626–633.

SIOTANI, M. (1964), "Interval Estimation for Linear Combination of Means", *Journal of the American Statistical Association*, **59**, 1141–1164.

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