THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ON TESTING SOME LINEAR RELATIONS AMONG VARIANCES

by

James M. Davenport

Technical Report No. 28
Department of Statistics THEMIS Contract

April 12, 1969

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

ON TESTING SOME LINEAR RELATIONS

AMONG VARIANCES

A Thesis Presented to the Faculty of the Graduate School

of

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Master of Science

with a

Major in Statistics

by

James Melton Davenport
(B.A., West Texas State University, 1964)

April 11, 1969

Davenport, James Melton B.A., West Texas State University, 1964

On Testing Some Linear Relations Among Variances

Adviser: Associate Professor John T. Webster

Master of Science degree conferred May 25, 1969

Thesis completed April 11, 1969

In statistical methods, it is often desirable to test that the relation

$$\sum_{m=1}^{k} c_m \theta_m = \sum_{j=k+1}^{p} c_j \theta_j$$

holds among the variances θ_i . An assumption made in this paper is that there exists independent mean square estimates v_i of the variances θ_i such that $n_i v_i / \theta_i$ follows the chi-square distribution for each i = 1, 2, $\cdots$, p . An approximate test of the above relation is Satterthwaite's approximate F-test.

A solution, when testing the relation $\theta_3 = \theta_1 + \theta_2$, is developed for finding the true probability of being in the rejection region, when the Satterthwaite approximate F-test is used, and these probabilities are presented for several values of the parameters involved. A comparison of the results obtained is made with the work done by W. G. Cochran in this area.

In addition, methods are developed for finding the true probabilities of being in the rejection region, when using Satterthwaite's approximate F-test for testing the relations $\theta_1 + \theta_2 = \theta_3 + \theta_4$ and $\theta_1 = \theta_2 + \theta_3 - \theta_4$. However, in these cases none of the probabilities are presented.

ACKNOWLEDGMENTS

I would like to thank Dr. John T. Webster for his assistance in the preparation of this paper and Dr. Donald B. Owen for his assistance in the preparation of the computer program used for the numerical calculations. I am grateful to all my teachers; in particular, Dr. Webster, Dr. Owen, Dr. C. H. Kapadia, Dr. Richard K. Williams, Dr. Richard P. Bland and Dr. H. L. Cook, for his guidance in my undergraduate mathematics studies.

I am deeply indebted to Dr. Paul D. Minton for his efforts in obtaining a National Defense Education Act Title IV Fellowship for my financial support.

I wish to thank Mrs. Linda White for her excellent job in typing this manuscript.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In statistical methods, it is often desirable to test the general null hypothesis that the relation

$$C_1\Theta_1 + \cdots + C_m\Theta_m = C_{m+1}\Theta_{m+1} + \cdots + C_s\Theta_s \tag{1}$$

holds among the variances Θ_i , where the C_i are known real numbers. An assumption which is essential in deriving the results in the sequel for testing relation (1), is that there exists a number of independent estimates v_i of each of the variances Θ_i , respectively, $(i = 1, 2, \cdots , s)$. We assume that each estimate v_i is based on n_i degrees of freedom, and $n_i v_i/\Theta_i$ follows the chi-square distribution with n_i degrees of freedom.

This problem has been encountered several times in the application of statistical methods, especially in the analysis of variance. Several methods are well known for testing some particular cases of relation (1). These are as follows:

1. The F-test. This tests the relation $\Theta_1 = \Theta_2$.
2. Bartlett's [1] test of homogeneity of variances. This tests the series of k-1 relations $\Theta_1 = \Theta_2 = \cdots = \Theta_k$.
3. The two-tailed Behrens-Fisher problem. This tests the relation that

$$\Theta_3 = \Theta_1 + \Theta_2 \tag{2}$$

holds among the variances, as is shown by Cochran [2].

4. Satterthwaite's [10] F'-statistic, which is an approximate F-test. This tests relation (1) when the $C_i = \pm 1$, $(i = 1, 2, \cdots, s)$.

When the $C_i = \pm 1$, $(i = 1, 2, \cdots, k)$, relation (1) can always be reduced to the form

$$\Theta_1 + \Theta_2 + \cdots + \Theta_r = \Theta_{r+1} + \Theta_{r+2} + \cdots + \Theta_k . \tag{3}$$

Satterthwaite [10] has suggested the test criterion

$$F' = \frac{v_1 + v_2 + \cdots + v_r}{v_{r+1} + v_{r+2} + \cdots + v_k} , \tag{4}$$

where the v_i are the independent mean square estimates based on n_i degrees of freedom of the variances Θ_i . When relation (3) is true, the null distribution F' follows the F-distribution, approximately. The degrees of freedom of F' , ν_1 and ν_2 , are found by a rule also suggested by Satterthwaite.

$$\nu_1 = \frac{(v_1 + v_2 + \cdots + v_r)^2}{\frac{v_1^2}{n_1} + \frac{v_2^2}{n_2} + \cdots + \frac{v_r^2}{n_r}} \; : \; \nu_2 = \frac{(v_{r+1} + v_{r+2} + \cdots + v_k)^2}{\frac{v_{r+1}^2}{n_{r+1}} + \frac{v_{r+2}^2}{n_{r+2}} + \cdots + \frac{v_k^2}{n_k}} \tag{5}$$

It should be noted that relation (2) is a special case of relation (3). Hence, an approximate test procedure for relation (2) is one of the F'-statistics.

Many different ratios in (4) can be formed by testing different cases of relation (3). Those that will be considered in the sequel will be those that arise in testing the significance of a specific factor in the analysis of variance of the three factor factorial Eisenhart [4] model II and the mixed model. Letting Θ_i represent the variance components or expected mean squares in the analysis of variance tables for these two models, testing the hypotheses of no main effects can be reduced to testing one of the following four relations:

$$\Theta_1 + \Theta_2 = \Theta_3 + \Theta_4 \tag{6}$$

$$\Theta_1 - \Theta_2 = \Theta_3 - \Theta_4 \tag{7}$$

$$\Theta_1 = \Theta_2 + \Theta_3 - \Theta_4 \tag{8}$$

$$\Theta_1 + \Theta_2 - \Theta_3 = \Theta_4 \quad . \tag{9}$$

The elements on the right and left hand sides of the equal signs indicate the structure of the test statistic.

In the sequel, results will be derived for testing relations (2), (6) and (8). For convenience, F_1' will be used to represent the Satterthwaite's approximate F-statistic used to test relation (2). Likewise, F_2' will be associated with relation (6) and F_3' with relation (8).

In general, the alternative hypothesis will be that the left side of the relation will be greater than the right side of the relation. However, if the alternative hypothesis is two-sided, the numerator of F' will be whichever of the estimates of the variance is larger, so that the alternative will again be one-sided as above. The resulting probability will be doubled.

In each case, the quantity obtained is $P(F' > F_{(1-\alpha)}[\nu_1 , \nu_2])$. $F_{(1-\alpha)}[\nu_1 , \nu_2]$ is the $1-\alpha$ percent level point from the F-distribution with ν_1 and ν_2 degrees of freedom, where ν_1 and ν_2 are obtained using equation (5). In practice, this probability is assumed to be α , which is not true since F' is only an approximate F-statistic.

Cochran [2] has developed a method for finding the exact probability of the type-I error for testing relation (2), and he has tabulated this for several degrees of freedom, several values of the nuisance parameter, and at the apparent .05 significance level.

In the sequel, this method of testing relation (2) is presented, and most of the values that Cochran tabulated were recalculated using a computer for comparison. In addition, the true probability of the type-I error is calculated for several more degrees of freedom and at three different apparent significance levels for testing relation (2). Comparisons of Cochran's conclusions and those obtained in the sequel are also presented.

In Chapters III and IV, methods are presented for testing relations (6) and (8); however, no numerical integration was accomplished. This is planned for a later time, including methods for testing relations (7) and (9), and comparisons of the true probabilities of the type-I error will be made to find the "best" Satterthwaite's approximate F-statistic in the analysis of variance of the three factor factorial models. The power functions of these tests will also be investigated.

CHAPTER II

TESTING THE RELATION $\Theta_3 = \Theta_1 + \Theta_2$ USING F_1'

As stated before, testing relation (2) is a special case of testing (3). Hence, an F'-statistic can be used. F_1' is defined to be

$$F_1' = \frac{v_3}{v_1 + v_2} \quad , \tag{10}$$

where v_1 , v_2 and v_3 are independent mean square estimates of Θ_1 , Θ_2 and Θ_3 , respectively, and each is based on n_1 , n_2 and n_3 degrees of freedom, respectively. The degrees of freedom of F_1' , ν_1 and ν_2 , are

$$\nu_1 = n_3 \tag{11}$$

$$\nu_2 = \frac{(v_1 + v_2)^2}{\frac{v_1^2}{n_1} + \frac{v_2^2}{n_2}} = \frac{(1 + u)^2}{\frac{u^2}{n_1} + \frac{1}{n_2}} \tag{12}$$

where $u = v_1/v_2$.

Let A represent the event that $F_1' > F_{(1-\alpha)}[\nu_1 , \nu_2]$. Note that the event A actually depends on the realized value of u . Hence, if α is the significance level of the test, α is actually equal to $P(A|u)$. That is, any approximation found is conditional. Consequently, applying the rules of conditional probability it is seen that

$$P(A) = \int_0^\infty P(A|u)f(u)du \quad , \tag{13}$$

where f(u) is the density function of the random variable u , and P(A) is the desired quantity. This may be obtained by routine methods.

An alternate method, given by Cochran [2], gives the conditional distribution more quickly. First, let us consider the three following random variables: 1) $n_3 v_3/\Theta_3$, 2) $n_2 v_2/\Theta_2$ and 3) $n_1 v_1/\Theta_1$. It is well known that each is distributed as a central chi-square with n_3 , n_2 and n_1 degrees of freedom, respectively, and that they are all independent. Therefore, the random variable

$$Q = \frac{v_3(n_1 + n_2)}{\Theta_3\left[\frac{n_1 v_1}{\Theta_1} + \frac{n_2 v_2}{\Theta_2}\right]}$$

is distributed as a central F-distribution with n_3 and $n_1 + n_2$ degrees of freedom. Also, Q is independent of u , from the well known fact that the ratio of two independent chi-squares is independent of the F-statistic in which the sum of these two chi-squares appears as either the numerator or denominator. A derivation of this is given in Appendix I.

Hence, when relation (2) is true, the result follows that:

$$F_1' = \frac{v_3}{v_1 + v_2} \cdot \frac{Q}{Q} = \frac{Q(\Theta_1 + \Theta_2)\left[\frac{n_1 v_1}{\Theta_1} + \frac{n_2 v_2}{\Theta_2}\right]}{(n_1 + n_2)(v_1 + v_2)}$$

$$= \frac{Q[1 + U]\left[\frac{n_1 u}{U} + n_2\right]}{(1 + u)(n_1 + n_2)} \qquad (14)$$

where $U = \Theta_1/\Theta_2$

Substituting equation (14) into equation (13) gives

$$P(A) = \int_0^\infty P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + u)(n_1 + n_2)}{(1 + U)\left[\frac{n_1 u}{U} + n_2\right]} \;\middle|\; u \right] f(u)\,du \ .$$

But, since Q and u are independent, the result reduces to

$$P(A) = \int_0^\infty P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + u)(n_1 + n_2)}{(1 + U)\left[\frac{n_1 u}{U} + n_2\right]} \right] f(u)\,du \ . \tag{15}$$

By applying the probability integral transformation $dp = f(u)du$ or $p = \int_0^{u'} f(u)du$, to equation (15), the work can be simplified. Note that the random variable u/U is distributed as an F-distribution with n_1 and n_2 degrees of freedom. Letting $u = x \cdot U$, the probability integral transformation becomes

$$p = \int_0^{x'} g(x)\,dx \ , \tag{16}$$

where g(x) is the density function of an F-statistic, with n_1 and n_2 degrees of freedom. Therefore, for fixed values of n_1 , n_2 , U and p , a value for x and consequently u is determined. Then applying the transformation (16), equation (15) becomes

$$P(A) = \int_0^1 P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + u)(n_1 + n_2)}{(1 + U)\left[\frac{n_1 u}{U} + n_2\right]} \right] dp \ , \tag{17}$$

which is the form that will be used to evaluate the probability.

The fact that equation (17) is symmetric, when U = 1.0 , with respect to its parameters n_1 and n_2 can be verified by using the fact that $F_{(\alpha)}[m_1 , m_2] = 1/F_{(1-\alpha)}[m_2 , m_1]$. If we switch the roles of n_1 and n_2 and let U = 1.0 in the previous development, equation (15) becomes

$$P(A) = \int_0^{\infty} P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + u_1)(n_2 + n_1)}{2[n_2 u_1 + n_1]}\right] f(u_1)\,du_1 , \tag{18}$$

where $u_1 = v_1/v_2$,

ν_1 and ν_2 are now given by

$$\nu_1 = n_3 \quad ; \quad \nu_2 = \frac{(1 + u_1)^2}{\frac{u_1^2}{n_2} + \frac{1}{n_1}} , \tag{19}$$

and Q is the same F-statistic as before. Since U = 1.0 , u_1 is an F-statistic with n_2 and n_1 degrees of freedom. Now applying the transformation (16) to equation (18), it becomes

$$P(A) = \int_0^{1} P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + u_1)(n_2 + n_1)}{2[n_2 u_1 + n_1]}\right] dp \tag{20}$$

Recall that the values for u_1 are obtained by giving a value for p between zero and one, say β , and then finding the β^{th} cumulant point of the F-distribution with n_2 and n_1 degrees of freedom. Then set $u_1 = F_{(\beta)}[n_2 , n_1]$. Using the inverse identity for the F-distribution, then $u_1 = 1/u$, where u is the same quantity used in the previous development. That is, u is the $(1-\beta)^{th}$ cumulant point of an F-distribution with n_1 and n_2 degrees of freedom. Notice, however, that this is the $(1-\beta)^{th}$ cumulant, not the β^{th} cumulant. Therefore, if the right side of equation (20) is multiplied by a minus one, integrated from one to zero, and 1/u is substituted for u_1 , the equality of equation (20) is preserved, and it becomes

$$P(A) = (-1) \int_1^{0} P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + \frac{1}{u})(n_1 + n_2)}{2\left[\frac{n_2}{u} + n_1\right]}\right] dp . \tag{21}$$

Likewise, if u_1 is set equal to $1/u$ in equation (19), it reduces to

$$\nu_2 = \frac{(1 + \frac{1}{u})^2}{\frac{1}{n_2 u^2} + \frac{1}{n_1}} = \frac{(1 + u)^2}{\frac{u^2}{n_1} + \frac{1}{n_2}} ,$$

which is the same as equation (12). Therefore, equation (21) becomes

$$P(A) = \int_0^1 P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + u)(n_1 + n_2)}{2[un_1 + n_2]}\right] dp ,$$

which is equation (17) with $U = 1.0$. Hence, the symmetry of equation (17) with respect to n_1 and n_2 is established. However, this is not true if U is not equal to one.

The limiting cases of equation (17) as $U \to 0$ and $U \to \infty$ can be considered by investigating $U = \Theta_1/\Theta_2$ and relation (2). Since in practical situations infinite variances can be ignored, $U \to 0$ if and only if $\Theta_1 \to 0$. If $\Theta_1 \to 0$, then relation (2) reduces to $\Theta_3 = \Theta_2$, and F_1' becomes a true F-statistic with n_3 and n_2 degrees of freedom. Hence, $P(A) \to \alpha$. Likewise, $U \to \infty$ if and only if $\Theta_2 \to 0$, and relation (2) then reduces to $\Theta_3 = \Theta_1$. F_1' then is a true F-statistic with n_3 and n_1 degrees of freedom. Again $P(A) \to \alpha$. Therefore, as $U \to 0$ or $U \to \infty$, then $P(A) \to \alpha$. Various values for $P(A)$ were calculated for varying values of U and fixed values of n_1 , n_2 and n_3 . The results are presented in Figure 1 and Figure 2.

The values of $P(A)$ are obtained by numerically integrating equation (17) using a digital computer program, which is outlined in Appendix II. Cochran [2] has calculated several values for $P(A)$, for several values of n_1 , n_2 , n_3 and U . Some of these values were computed, using the program, to check the accuracy of the method and the program. The results are

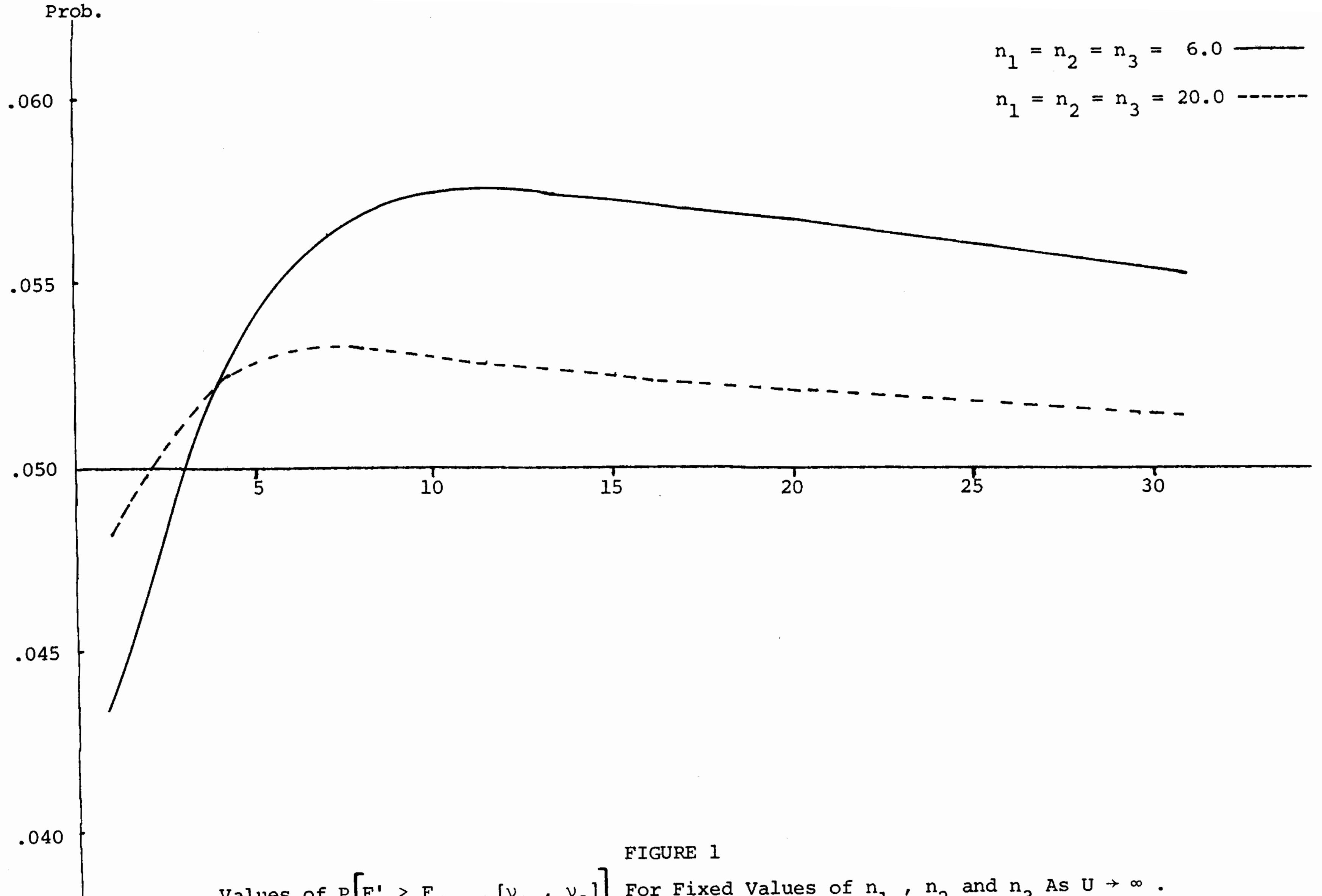

FIGURE 1

Values of $P\left[F_1' > F_{(.95)}[\nu_1, \nu_2]\right]$ For Fixed Values of n_1, n_2 and n_3 As $U \to \infty$.

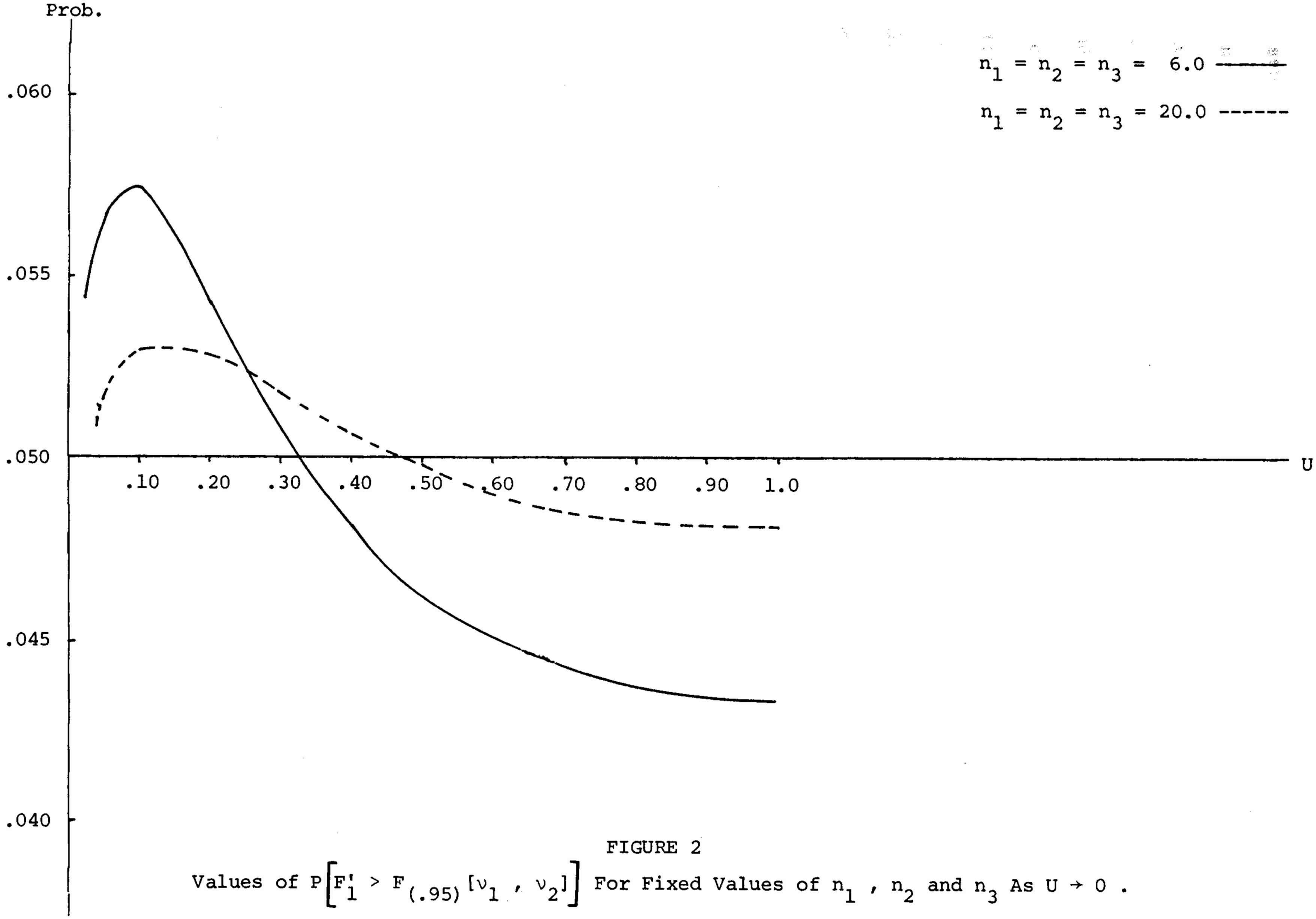

FIGURE 2

Values of $P\left[F_1' > F_{(.95)}[\nu_1, \nu_2]\right]$ For Fixed Values of n_1, n_2 and n_3 As $U \to 0$.

presented in Table 1, where Cochran's values are given in parentheses. When the values of U and the degrees of freedom are relatively large, the convergence of the numerical integration is slower than when these values are relatively small. Since Cochran's values were probably calculated by hand, they may tend to be in error for these slower converging integrations. In each case, where the values arrived at using the computer program are different from Cochran's values, the numeral integration program arrived at Cochran's value and continued to a more accurate answer; since Cochran's value was not within the tolerance limit set for convergence of the numerical integration.

TABLE 1

Comparisons of the Values of $P\left[F_1' > F_{(.95)}[\nu_1 , \nu_2]\right]$ with Cochran's Work (Cochran's Values Given in Parentheses)

$n_1=n_2$	n_3	$U = \theta_1/\theta_2$			
		1	2	4	16
12	6	.0477 (.047)	.0491 (.049)	.0516 (.052)	.0525 (.054)
	12	.0468 (.046)	.0490 (.049)	.0529 (.054)	.0538 (.056)
24	6	.0493 (.050)	.0498 (.050)	.0506 (.052)	.0508 (.053)
	12	.0489 (.049)	.0498 (.050)	.0512 (.052)	.0513 (.054)

In addition, it was noted that the values of P(A) appear to reach their maximum deviations from α , when the value of U is in the neighborhood of 0.1 , 1.0 and 10.0 .

Further investigation to determine some of the properties of equation (17) was accomplished for several values of n_1 , n_2 and n_3 with U fixed at the value of one.

By letting n_1 and n_2 assume fixed values and letting n_3 vary, the following results were obtained:

1. For small values of n_1 and n_2 , P(A) departs from α , as $n_3 \to \infty$.
2. For large values of both n_1 and n_2 , the P(A) departs from α , as $n_3 \to \infty$, but at a much slower rate than the above.
3. For unequal extreme values of n_1 and n_2 , the P(A) still departs from α , as $n_3 \to \infty$, but at a more rapid rate than either of the above cases.

Also by letting n_3 be fixed, and letting n_1 and n_2 vary, the following conclusions were made:

4. Irregardless of the value of n_3 , if n_1 and n_2 are extremely different, then the P(A) departs drastically from α .
5. If n_1 and n_2 become large simultaneously, the P(A) approaches α .
6. If n_1 and n_2 are both small, the P(A) is approximately α .

Special attention should be made to the cases when the P(A) departs from α . They are:

7. If n_1 and n_2 are both small and n_3 is large.
8. If n_1 and n_2 assume extremely different values.

In both the above cases, drastic departures from α , can be realized for P(A) .

Graphic results of all of the above conclusions are given in Figures 3, 4, and 5.

The properties of equation (17) observed by Cochran [2] include the limiting cases of P(A) as $U \to 0$ and $U \to \infty$ and items 1, 2, 5, 6 and 7 above. He also noted that the significance probabilities appear to increase to a maximum which is not far from the value realized for $U = 16$. However, here it was noted that the values for P(A) actually achieve their maximum in a neighborhood for $U = 10$. Other points of maximum departure from α were noted to be in the neighborhoods of $U = 1.0$ and $U = 0.1$.

Additional results not noted by Cochran include the fact that equation (17) is symmetrical with respect to n_1 and n_2 when $U = 1.0$, and items 3, 4 and 8 above.

Equation (17) was tabulated for several values of n_1 , n_2 and n_3 , which are common degrees of freedom that arise in most practical designs; for four values of U ; and for three values of α , the apparent significance level of F_1' . These results are presented in Tables 2, 3 and 4. Values for n_1 and n_2 both small were not included, since $n_1 + n_2$ is the degree of freedom of the denominator of an F-statistic. Such F-statistics with small degrees of freedom in the denominator have low power; therefore, such cases are usually avoided in most practical experiments.

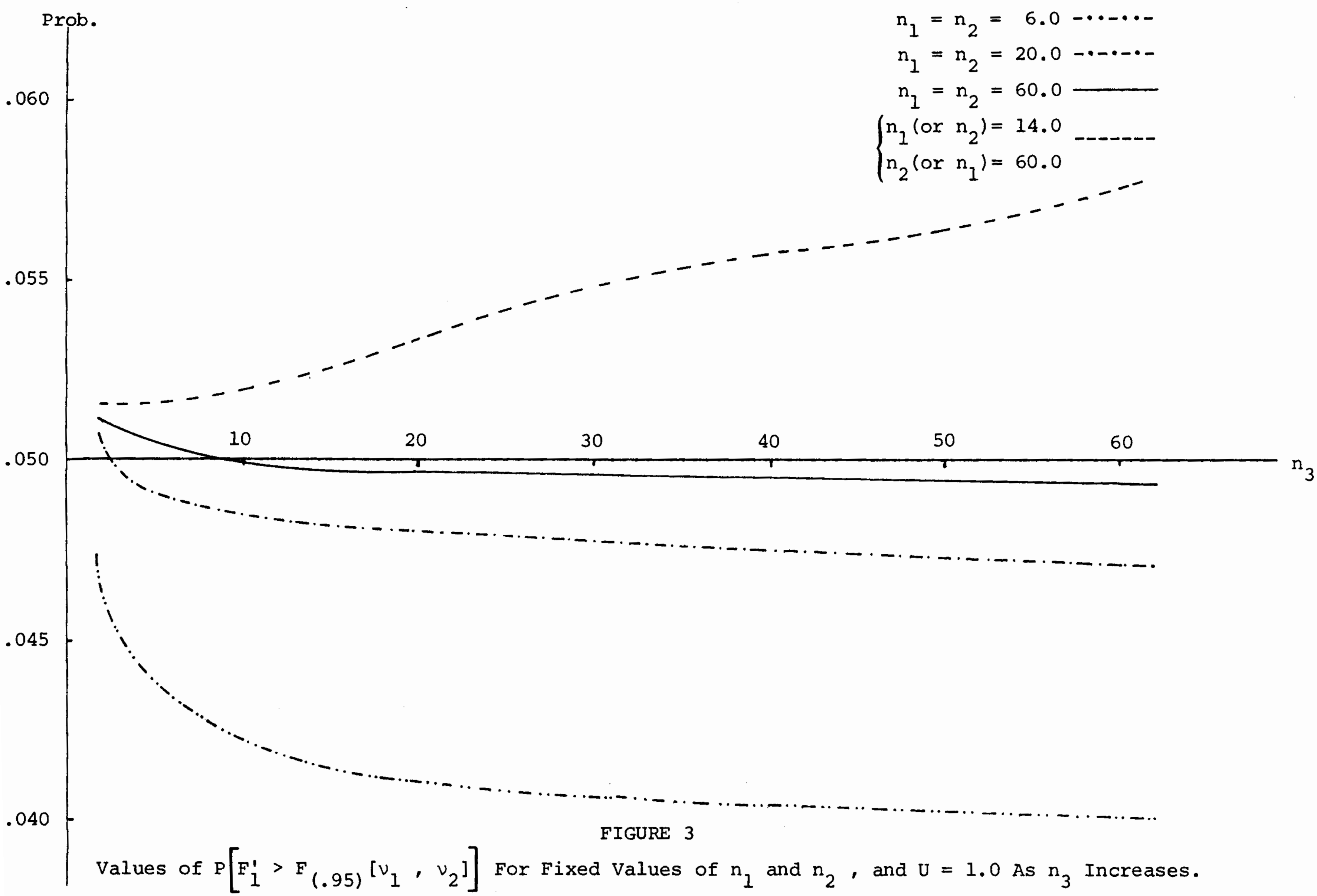

FIGURE 3

Values of $P\left[F_1' > F_{(.95)}[\nu_1, \nu_2]\right]$ For Fixed Values of n_1 and n_2, and $U = 1.0$ As n_3 Increases.

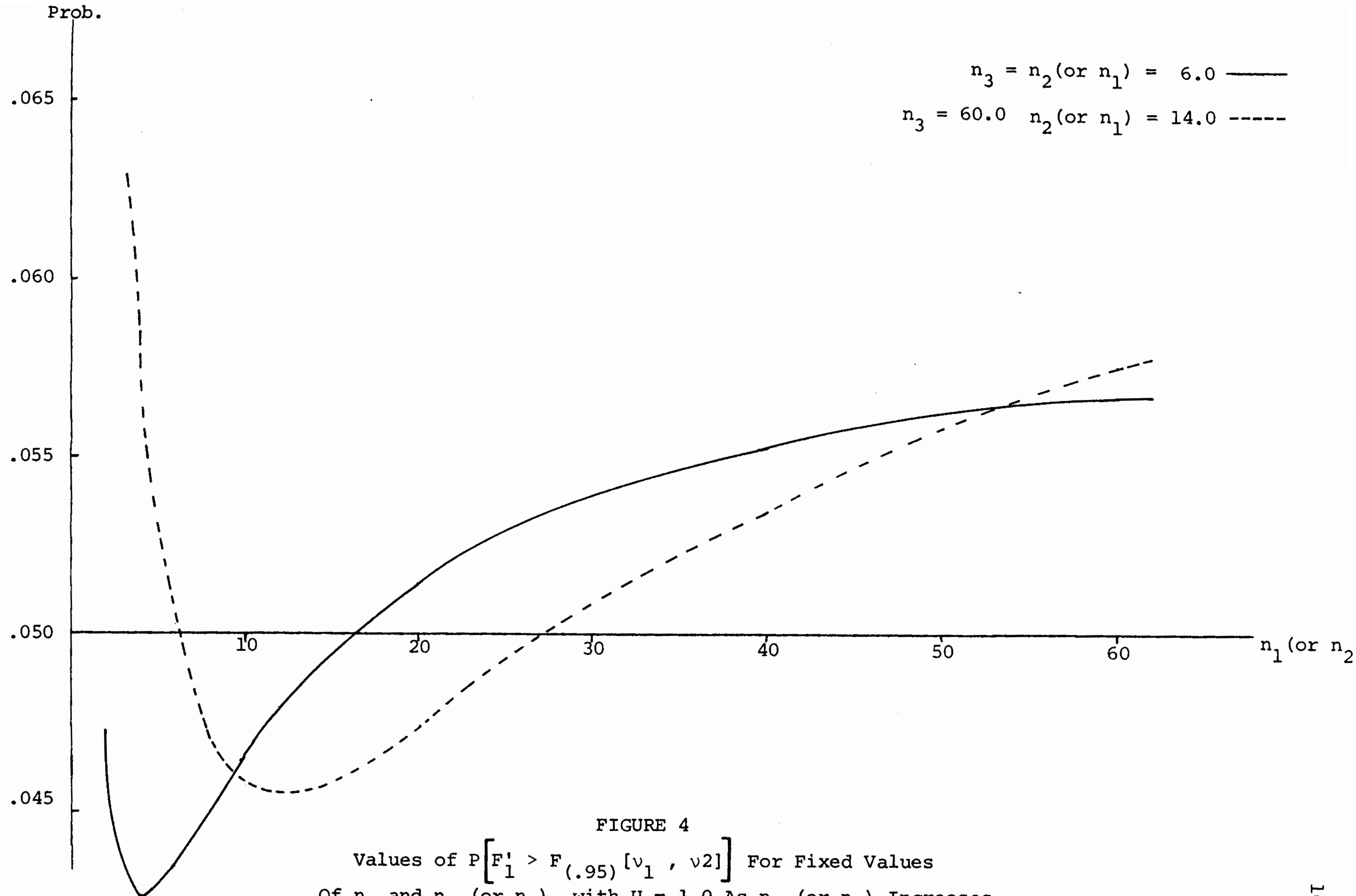

FIGURE 4

Values of $P\left[F_1' > F_{(.95)}[\nu_1 , \nu 2]\right]$ For Fixed Values Of n_3 and n_2 (or n_1), with U = 1.0 As n_1 (or n_2) Increases.

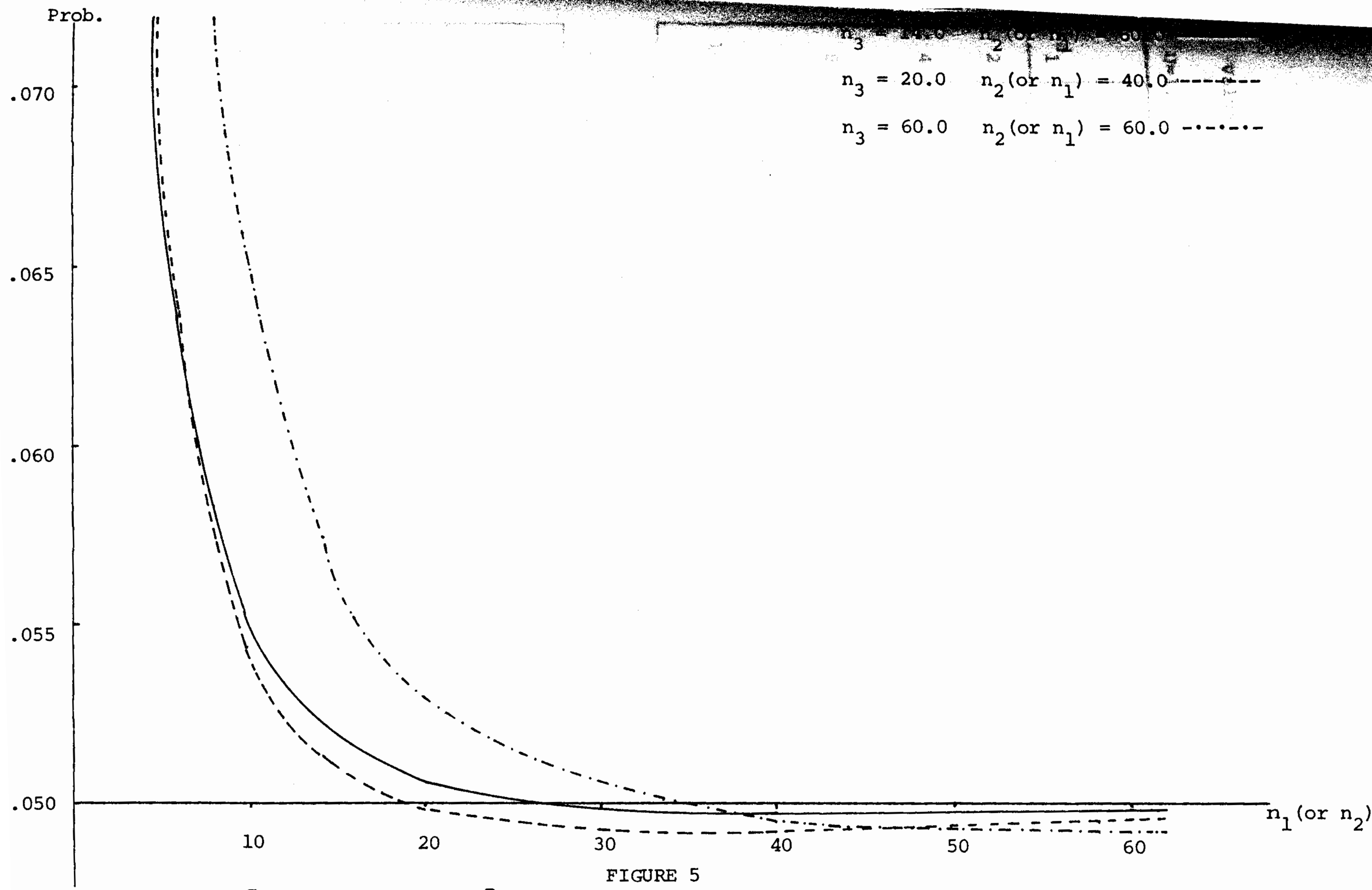

FIGURE 5

Values of $P\left[F_1' > F_{(.95)}[\nu_1, \nu_2]\right]$ For Fixed Values of n_3 and n_2 (or n_1) with U = 1.0 As n_1 (or n_2) Increases.

TABLE 2

Values of $P\left[F_1' > F_{(1-\alpha)}[\nu_1, \nu_2]\right]$ at the Apparent .01 Level

U=1.0

n_1	n_2	n_3 = 2	4	8
2	8	.0125	.0115	.0117
	16	.0166	.0172	.0197
4	8	.0103	.0089	.0086
	16	.0125	.0118	.0125
8	2	.0125	.0115	.0117
	4	.0103	.0089	.0086
	8	.0103	.0087	.0083
	16	.0112	.0098	.0095
16	2	.0166	.0172	.0197
	4	.0125	.0118	.0125
	8	.0112	.0098	.0095
	16	.0111	.0097	.0094

U=2.0

n_1	n_2	n_3 = 2	4	8
2	8	.0186	.0190	.0210
	16	.0253	.0296	.0376
4	8	.0127	.0119	.0124
	16	.0159	.0166	.0195
8	2	.0098	.0084	.0081
	4	.0097	.0081	.0077
	8	.0109	.0095	.0093
	16	.0123	.0114	.0121
16	2	.0121	.0112	.0116
	4	.0109	.0095	.0093
	8	.0109	.0094	.0090
	16	.0114	.0100	.0099

U=4.0

n_1	n_2	n_3 = 2	4	8
2	8	.0279	.0317	.0378
	16	.0364	.0461	.0608
4	8	.0162	.0166	.0188
	16	.0196	.0222	.0277
8	2	.0094	.0078	.0074
	4	.0105	.0090	.0087
	8	.0122	.0112	.0116
	16	.0135	.0133	.0149
16	2	.0106	.0092	.0090
	4	.0107	.0092	.0088
	8	.0113	.0099	.0097
	16	.0118	.0107	.0109

U=16.0

n_1	n_2	n_3 = 2	4	8
2	8	.0397	.0480	.0571
	16	.0451	.0555	.0657
4	8	.0188	.0206	.0239
	16	.0203	.0228	.0269
8	2	.0112	.0098	.0096
	4	.0123	.0115	.0119
	8	.0131	.0126	.0137
	16	.0135	.0133	.0146
16	2	.0111	.0096	.0093
	4	.0116	.0103	.0102
	8	.0118	.0107	.0108
	16	.0120	.0109	.0112

TABLE 3

Values of $P\left[F_1' > F_{(1-\alpha)}[\nu_1, \nu_2]\right]$ at the Apparent .025 Level

U=1.0

n_1	n_2	n_3 = 2	4	8
2	8	.0277	.0274	.0282
	16	.0337	.0363	.0413
4	8	.0244	.0227	.0222
	16	.0276	.0274	.0289
8	2	.0277	.0274	.0282
	4	.0244	.0227	.0222
	8	.0246	.0226	.0217
	16	.0259	.0244	.0242
16	2	.0337	.0363	.0413
	4	.0276	.0274	.0289
	8	.0259	.0244	.0242
	16	.0259	.0243	.0238

U=2.0

n_1	n_2	n_3 = 2	4	8
2	8	.0370	.0399	.0443
	16	.0453	.0537	.0662
4	8	.0280	.0278	.0290
	16	.0322	.0346	.0395
8	2	.0233	.0216	.0211
	4	.0235	.0214	.0204
	8	.0255	.0239	.0235
	16	.0274	.0269	.0281
16	2	.0270	.0265	.0276
	4	.0253	.0237	.0234
	8	.0255	.0238	.0232
	16	.0263	.0249	.0247

U=4.0

n_1	n_2	n_3 = 2	4	8
2	8	.0490	.0570	.0673
	16	.0581	.0725	.0912
4	8	.0327	.0347	.0386
	16	.0367	.0415	.0493
8	2	.0228	.0207	.0198
	4	.0249	.0231	.0225
	8	.0273	.0266	.0274
	16	.0290	.0294	.0320
16	2	.0247	.0231	.0228
	4	.0252	.0234	.0227
	8	.0261	.0246	.0244
	16	.0268	.0258	.0263

U=16.0

n_1	n_2	n_3 = 2	4	8
2	8	.0593	.0700	.0800
	16	.0641	.0764	.0869
4	8	.0352	.0382	.0425
	16	.0367	.0404	.0454
8	2	.0261	.0246	.0244
	4	.0276	.0270	.0277
	8	.0285	.0284	.0299
	16	.0289	.0291	.0310
16	2	.0259	.0242	.0238
	4	.0266	.0253	.0253
	8	.0269	.0258	.0262
	16	.0271	.0261	.0266

TABLE 4

Values of $P\left[F_1' > F_{(1-\alpha)}[\nu_1, \nu_2]\right]$ at the Apparent .05 Level

U=1.0

n_1	n_2	n_3 = 2	4	8
2	8	.0522	.0528	.0546
	16	.0595	.0643	.0719
4	8	.0481	.0462	.0455
	16	.0522	.0526	.0548
8	2	.0522	.0528	.0546
	4	.0481	.0462	.0455
	8	.0487	.0464	.0451
	16	.0504	.0489	.0485
16	2	.0595	.0643	.0719
	4	.0522	.0526	.0548
	8	.0504	.0489	.0485
	16	.0506	.0490	.0483

U=2.0

n_1	n_2	n_3 = 2	4	8
2	8	.0639	.0695	.0769
	16	.0731	.0851	.1016
4	8	.0527	.0531	.0552
	16	.0576	.0613	.0680
8	2	.0463	.0444	.0436
	4	.0471	.0445	.0429
	8	.0499	.0482	.0477
	16	.0522	.0520	.0537
16	2	.0512	.0513	.0531
	4	.0495	.0478	.0473
	8	.0501	.0482	.0473
	16	.0510	.0497	.0495

U=4.0

n_1	n_2	n_3 = 2	4	8
2	8	.0773	.0890	.1028
	16	.0863	.1041	.1253
4	8	.0582	.0616	.0669
	16	.0624	.0688	.0783
8	2	.0460	.0433	.0417
	4	.0491	.0471	.0462
	8	.0521	.0517	.0530
	16	.0539	.0550	.0582
16	2	.0486	.0468	.0462
	4	.0496	.0476	.0465
	8	.0508	.0494	.0490
	16	.0516	.0509	.0515

U=16.0

n_1	n_2	n_3 = 2	4	8
2	8	.0851	.0970	.1072
	16	.0893	.1025	.1130
4	8	.0604	.0641	.0690
	16	.0618	.0662	.0717
8	2	.0508	.0494	.0492
	4	.0526	.0522	.0533
	8	.0535	.0537	.0556
	16	.0539	.0545	.0567
16	2	.0505	.0489	.0483
	4	.0514	.0502	.0503
	8	.0518	.0509	.0513
	16	.0519	.0512	.0519

CHAPTER III

TESTING THE RELATION $\Theta_1 + \Theta_2 = \Theta_3 + \Theta_4$ USING F_2'

The hypothesis that relation (6) holds among the variances can be tested by using an F'-statistic. F_2' , which will be associated with relation (6) in the sequel, is defined to be

$$F_2' = \frac{v_1 + v_2}{v_3 + v_4} \quad , \tag{22}$$

where v_1 , v_2 , v_3 and v_4 are independent mean square estimates of Θ_1 , Θ_2 , Θ_3 and Θ_4 , respectively, and each is based on n_1 , n_2 , n_3 and n_4 degrees of freedom, respectively. The degrees of freedom of F_2' , ν_1 and ν_2 are given by

$$\nu_1 = \frac{(v_1 + v_2)^2}{\frac{v_1^2}{n_1} + \frac{v_2^2}{n_2}} = \frac{(1 + u)^2}{\frac{u^2}{n_1} + \frac{1}{n_2}}$$

$$\nu_2 = \frac{(v_3 + v_4)^2}{\frac{v_3^2}{n_3} + \frac{v_4^2}{n_4}} = \frac{(1 + w)^2}{\frac{w^2}{n_3} + \frac{1}{n_4}} \quad ,$$

where $u = v_1/v_2$ and $w = v_3/v_4$.

Let A_2 denote the event that $F_2' > F_{(1-\alpha)}[\nu_1 , \nu_2]$. As in Chapter II with the F_1'-statistic, the probability that the event A_2 has occurred is a conditional probability. Hence, if α is the significance level of the test, then $P(A_2|u , w) = \alpha$.

Consequently,

$$P(A_2) = \int_0^\infty \int_0^\infty P[A_2 | u , w] g(u , w) du dw \quad , \tag{23}$$

where $g(u , w)$ is the joint density function of the random variables u and w .

Using the method which was developed in Chapter II, this conditional distribution can be found. From the fact that $n_i v_i / \Theta_i$ is distributed with the chi-square distribution with n_i degrees of freedom for, $(i = 1, 2, 3, 4)$, it can be shown that the random variable

$$Q = \frac{(n_3 + n_4)\left[\frac{n_1 v_1}{\Theta_1} + \frac{n_2 v_2}{\Theta_2}\right]}{(n_1 + n_2)\left[\frac{n_3 v_3}{\Theta_3} + \frac{n_4 v_4}{\Theta_4}\right]}$$

is distributed as a central F-distribution with $n_1 + n_2$ and $n_3 + n_4$ degrees of freedom. Also, Q is independent of u and w .

Therefore, the result follows that

$$F_2' = \frac{v_1 + v_2}{v_3 + v_4} \cdot \frac{Q}{Q} = Q \frac{(\Theta_1 + \Theta_2)(v_1 + v_2)(n_1 + n_2)\left[\frac{n_3 v_3}{\Theta_3} + \frac{n_4 v_4}{\Theta_4}\right]}{(\Theta_1 + \Theta_2)(v_3 + v_4)(n_3 + n_4)\left[\frac{n_1 v_1}{\Theta_1} + \frac{n_2 v_2}{\Theta_2}\right]} \quad ,$$

and when relation (6) is true, this becomes

$$F_2' = Q \frac{(1 + \phi_2)(1 + u)(n_1 + n_2)\left[\frac{n_3 w}{\phi_2} + n_4\right]}{(1 + \phi_1)(1 + w)(n_3 + n_4)\left[\frac{n_1 u}{\phi_1} + n_2\right]} \quad , \tag{24}$$

where $\phi_1 = \Theta_1/\Theta_2$ and $\phi_2 = \Theta_3/\Theta_4$.

Substituting equation (24) into equation (23) gives

$$P(A_2) = \int_0^\infty\int_0^\infty P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + \phi_1)(1 + w)(n_3 + n_4)\left[\frac{n_1 u}{\phi_1} + n_2\right]}{(1 + \phi_2)(1 + u)(n_1 + n_2)\left[\frac{n_3 w}{\phi_2} + n_4\right]} \middle| u,w\right] g(u,w)\,dudw \;.$$

However, since Q is independent of u and w , the result reduces to

$$P(A_2) = \int_0^\infty\int_0^\infty P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + \phi_1)(1 + w)(n_3 + n_4)\left[\frac{n_1 u}{\phi_1} + n_2\right]}{(1 + \phi_2)(1 + u)(n_1 + n_2)\left[\frac{n_3 w}{\phi_2} + n_4\right]}\right] g(u,w)\,dudw \;. \quad (25)$$

The probability integral transformation

$$dp_1 dp_2 = g(u,w)\,dudw \quad \text{or} \quad P_1 \cdot P_2 = \int_0^{w'}\int_0^{u'} g(u,w)\,dudw \;,$$

applied to equation (25) will simplify the numerical integration. Note that u and w are independent, and that u/ϕ_1 has an F-distribution with n_1 and n_2 degrees of freedom. Likewise, w/ϕ_2 has an F-distribution with n_3 and n_4 degrees of freedom. Then by letting $x = u/\phi_1$ and $y = w/\phi_2$, the probability integral transformation becomes

$$P_1 \cdot P_2 = \int_0^{x'} g_1(x)\,dx \int_0^{y'} g_2(y)\,dy \;, \quad (26)$$

where $g_1(x)$ and $g_2(y)$ are density functions of F-statistics with n_1 and n_2 ; and n_3 and n_4 degrees of freedom, respectively. Then applying the transformation (26), equation (25) becomes

$$P(A_2) = \int_0^1\int_0^1 P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](1 + \phi_1)(1 + w)(n_3 + n_4)\left[\frac{n_1 u}{\phi_1} + n_2\right]}{(1 + \phi_2)(1 + u)(n_1 + n_2)\left[\frac{n_3 w}{\phi_2} + n_4\right]}\right] dp_1 dp_2 \;,$$

which can be easily numerically integrated.

CHAPTER IV

TESTING THE RELATION $\Theta_1 = \Theta_2 + \Theta_3 - \Theta_4$ USING F_3'

Relation (8) can be tested by using the F_3'-statistic, which is defined to be,

$$F_3' = \frac{v_1}{v_2 + v_3 - v_4} \quad , \tag{27}$$

where the v_i , (i = 1, 2, 3, 4) are independent mean square estimates of Θ_i , respectively. Each v_i is based on n_i degrees of freedom. The degrees of freedom of F_3' , ν_1 and ν_2 are given by

$$\nu_1 = n_1$$

$$\nu_2 = \frac{(v_2 + v_3 - v_4)^2}{\frac{v_2^2}{n_2} + \frac{v_3^2}{n_3} + \frac{v_4^2}{n_4}} = \frac{(u + w - 1)^2}{\frac{u^2}{n_2} + \frac{w^2}{n_3} + \frac{1}{n_4}} \quad ,$$

where $u = v_2/v_4$ and $w = v_3/v_4$

Let A_3 denote the event that $F_3' > F_{(1-\alpha)}[\nu_1 , \nu_2]$. As before, this is a conditional event. Hence, $P(A_3|u , w) = \alpha$ and, consequently,

$$P(A_3) = \int_0^\infty \int_0^\infty P[A_3|u , w]g(u , w)\,dudw \quad , \tag{28}$$

where g(u , w) is the joint density function of the random variables u and w .

Again, from the fact that $n_i v_i/\Theta_i$ is distributed with a chi-square distribution with n_i degrees, and the v_i are independent for, $(i = 1, 2, 3, 4)$, it can be shown that the random variable

$$Q = \frac{(n_2 + n_3 + n_4)v_1}{\Theta_1\left[\frac{n_2 v_2}{\Theta_2} + \frac{n_3 v_3}{\Theta_3} + \frac{n_4 v_4}{\Theta_4}\right]} ,$$

is distributed with a central F-distribution with n_1 and $n_2 + n_3 + n_4$ degrees of freedom. Also Q is independent of u and w.

Therefore, when relation (8) is true, the result follows that,

$$F_3' = Q \frac{(\phi_1 + \phi_2 - 1)\left[\frac{n_2 u}{\phi_1} + \frac{n_3 w}{\phi_2} + n_4\right]}{(n_2 + n_3 + n_4)(u + w - 1)} , \qquad (29)$$

where $\phi_1 = \Theta_2/\Theta_4$ and $\phi_2 = \Theta_3/\Theta_4$.

Substituting equation (29) into equation (28) gives

$$P(A_3) = \int_0^\infty\int_0^\infty P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](n_2 + n_3 + n_4)(u + w - 1)}{(\phi_1 + \phi_2 - 1)\left[\frac{n_2 u}{\phi_1} + \frac{n_3 w}{\phi_2} + n_4\right]} \middle| u,w\right] g(u,w)\,dudw .$$

But, since Q is independent of u and w, the result reduces to

$$P(A_3) = \int_0^\infty\int_0^\infty P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](n_2 + n_3 + n_4)(u + w - 1)}{(\phi_1 + \phi_2 - 1)\left[\frac{n_2 u}{\phi_1} + \frac{n_3 w}{\phi_2} + n_4\right]}\right] g(u,w)\,dudw . \qquad (30)$$

Let $x = u/\phi_1$ and $y = w/\phi_2$, in the probability integral transformation

$$h(p_1 , p_2) = \int_0^{w'}\int_0^{u'} g(u,w)\,dudw .$$

It then becomes

$$h(p_1 , p_2) = \int_0^{y'} \int_0^{x'} g(x,y)\,dxdy \quad ,$$

where x follows the F-distribution with n_2 and n_4 degrees of freedom, and y follows the F-distribution with n_3 and n_4 degrees of freedom. In this case, however, x and y are not independent. Hence, the points x' and y' must be found from a bivariate F-distribution with the corresponding degrees of freedom and non-zero correlation.

Once this has been accomplished, equation (30) can be numerically integrated in the form

$$P(A_3) = \int_0^1\int_0^1 P\left[Q > \frac{F_{(1-\alpha)}[\nu_1 , \nu_2](n_2 + n_3 + n_4)(u + w - 1)}{(\phi_1 + \phi_2 - 1)\left[\frac{n_2 u}{\phi_1} + \frac{n_3 w}{\phi_2} + n_4\right]}\right] d[h(p_1 , p_2)] \ .$$

APPENDIX I

Let X_1 , X_2 and X_3 be independent random variables, each following the chi-square distribution with n_1 , n_2 and n_3 degrees of freedom respectively. Then it is well known that

$$F = \frac{X_1(n_2 + n_3)}{n_1(X_2 + X_3)}$$

follows a central F-distribution with n_1 and $n_2 + n_3$ degrees of freedom. The problem is to show that the random variable F and the random variable $Z = X_2/X_3$ are independent. The joint density function of X_1 , X_2 and X_3 is

$$g(X_1 , X_2 , X_3) = \frac{X_1^{\left(\frac{n_1-2}{2}\right)} X_2^{\left(\frac{n_2-2}{2}\right)} X_3^{\left(\frac{n_3-2}{2}\right)}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{n_3}{2}\right)2^{\left(\frac{n_1+n_2+n_3}{2}\right)}} e^{-\frac{X_1+X_2+X_3}{2}}$$

for $X_1 > 0$, $X > 0$ and $X_3 > 0$. Letting F and Z be defined as above and letting $W = X_3$, it can easily be shown that $X_2 = ZW$ and $X_1 = n_1FW \times (1 + Z)/(n_2 + n_3)$. Then the Jacobian of the transformation is $J = n_1W^2(1 + Z)/(n_2 + n_3)$. Therefore the joint density of F, Z and W is as follows:

$$h(F,Z,W) = \frac{F^{\left(\frac{n_1-2}{2}\right)}(1+Z)^{\frac{n_1}{2}} Z^{\left(\frac{n_2-2}{2}\right)} W^{\left(\frac{n_1+n_2+n_3-2}{2}\right)}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{n_3}{2}\right)2^{\left(\frac{n_1+n_2+n_3}{2}\right)}}$$

$$\times \left(\frac{n_1}{n_2+n_3}\right)^{\frac{n_1}{2}} e^{-\frac{1}{2}\left\{\left[\frac{n_1F}{n_2+n_3}+1\right](1+Z)W\right\}} ,$$

for $F > 0$, $Z > 0$ and $W > 0$. The joint density of F and Z is given by

$$h_{1,2}(F,Z) = \int_0^\infty h(F,Z,W)\,dW \quad .$$

Hence,

$$h_{1,2}(F,Z) = \frac{F^{\left(\frac{n_1-2}{2}\right)}(1+Z)^{\frac{n_1}{2}} Z^{\left(\frac{n_2-2}{2}\right)}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{n_3}{2}\right)2^{\left(\frac{n_1+n_2+n_3}{2}\right)}}\left(\frac{n_1}{n_2+n_3}\right)^{\frac{n_1}{2}}$$

$$\times \int_0^\infty W^{\left(\frac{n_1+n_2+n_3-2}{2}\right)} e^{-\frac{1}{2}\left\{\left[\frac{n_1F}{n_2+n_3}+1\right](1+Z)W\right\}} dW \quad .$$

This last integral is of the gamma distribution class; therefore,

$$h_{1,2}(F,Z) = \frac{\Gamma\left(\frac{n_1+n_2+n_3}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2+n_3}{2}\right)}\left(\frac{n_1}{n_2+n_3}\right)^{\frac{n_1}{2}} \frac{F^{\left(\frac{n_1+2}{2}\right)}}{\left[\frac{n_1F}{n_2+n_3}+1\right]^{\left(\frac{n_1+n_2+n_3}{2}\right)}}$$

$$\times \frac{\Gamma\left(\frac{n_2+n_3}{2}\right) Z^{\left(\frac{n_2-2}{2}\right)}}{\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{n_3}{2}\right)(1+Z)^{\left(\frac{n_2+n_3}{2}\right)}}$$

$$h_{1,2}(F,Z) = \frac{\Gamma\left(\frac{n_2 + n_3}{2}\right) Z^{\left(\frac{n_2-2}{2}\right)}}{\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{n_3}{2}\right)(1 + Z)^{\left(\frac{n_2+n_3}{2}\right)}} g(F) \quad ,$$

where $g(F)$ is the density function of an F-statistic with n_1 and $n_2 + n_3$ degrees of freedom. Hence, F and Z are independent.

If $X_2 = n_2 v_2/\Theta_2$ and $X_3 = n_3 v_3/\Theta_3$, then $Z = v_2 n_2 \Theta_3 / v_3 n_3 \Theta_2 = u(n_2\Theta_3/n_3\Theta_2)$, where $u = v_2/v_3$. Since Z is independent of F , then u is independent of F , also.

APPENDIX II

The numerical integration was actually accomplished by means of a digital computer program, written in Fortran IV, for an IBM 360/44. The main body of the program is the Romberg numerical integration technique. The functional values of the integrand were calculated by a series of subroutines as the variable of integration, p, assumed values between zero and one. They are as follows:

1. Given a value of p, the standard normal p^{th} cumulant point was found by means of an approximation given by Hastings [5] (page 192).
2. Then with this value, the p^{th} cumulant of the F-distribution with n_1 and n_2 degrees of freedom was found using the Cornish-Fisher [3] method.
3. This F-cumulant then specifies a value for u, given a value for U. With the value of u, values of ν_1 and ν_2 were calculated.
4. Now with a specified standard normal α^{th} cumulant point and the values of ν_1 and ν_2, the Cornish-Fisher method was again used to calculate the α^{th} critical point in $F(\nu_1, \nu_2)$.
5. Given all of the above values, a constant is determined for this specified value of p. The remaining cumulative probability of an F with n_3 and $n_1 + n_2$ degrees of freedom was found by using the relation between the F-distribution

and the Beta-distribution, and then using the relation between the incomplete Beta and the binomial. This cumulative value was set equal to the functional value of the integrand at the value, p .

6. Control was then returned to the main body of the program to repeat the process as often as necessary.

LIST OF REFERENCES

1. Bartlett, M. S. "Properties of Sufficiency and Statistical Tests." Proceedings of the Royal Society of London, A, Vol. 160 (1937). pp. 268-282.

2. Cochran, W. G. "Testing a Linear Relation Among Variances." Biometrics, Vol. 7 (1951) pp. 17-32.

3. Cornish, E. A. and Fisher, Sir Ronald A. "The Percentile Points of Distributions Having Known Cumulants." Technometrics, Vol. 2 (1960) pp. 209-225.

4. Eisenhart, C. "The Assumptions Underlying the Analysis of Variance." Biometrics, Vol. 3 (1947) pp. 1-21.

5. Hastings, Cecil Jr. Approximations for Digital Computers. Princeton, New Jersey: Princeton University Press, 1955.

6. Kelly, Truman Lee. The Kelley Statistical Tables. Cambridge, Mass.: Harvard University Press, 1948.

7. Mood, Alexander M. and Graybill, Franklin A. Introduction to the Theory of Statistics. New York: McGraw-Hill Book Co., Inc., 1963.

8. Ostle, Bernard. Statistics in Research, 2nd Edition. Amos, Iowa: The Iowa State University Press, 1963.

9. Owen, Donald B. Unpublished percentile points of the F-distribution. Southern Methodist University, 1968.

10. Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components." Biometrics, Vol. 2 (1946) pp. 110-114.

11. Smith, H. Fairfield. "The Problem of Comparing the Results of Two Experiments with Unequal Errors." Journal of the Council for Scientific and Industrial Research, Vol. 9 (1936) pp. 211-212.