The Annals of Statistics
1975, Vol. 3, No. 4, 797-808

POWER OF ANALYSIS OF VARIANCE TEST PROCEDURES FOR INCOMPLETELY SPECIFIED FIXED MODELS

BY RONALD MEAD, T. A. BANCROFT AND CHIEN-PAI HAN

Iowa State University

Derivations are given of the size and power of a test of a hypothesis for an incompletely specified fixed linear model. Such a testing procedure involves a test of a hypothesis of main interest subsequent to a preliminary test to determine the inclusion or not of a term in the tentatively specified complete fixed linear model. An evaluation of the formulas of the size and power was made over a wide range of values of the parameters involved, and recommendations concerning the choice of a proper significance level are given. One important objective of this study for the fixed model was to parallel as nearly as feasible the study given by Bozivich, H., Bancroft, T. A. and Hartley, H. O. (1956), *Ann. Math. Statist.* **27** 1017-1043 for the random model, and to compare the two. It should be noted that the results and recomendations for the fixed model are quite different from those given in the 1956 paper for the random model.

1. Introduction.

1.1. *Description of a pooling procedure.* The simplest situation of a pooling procedure for the testing hypotheses using analysis of variance techniques may be described as follows: We are given three mean squares, V_1, V_2, V_3, with corresponding degrees of freedom, n_1, n_2, n_3, and expectations σ_1^2, σ_2^2, σ_3^2. We would like to test the hypothesis $H_0: \sigma_3^2 = \sigma_2^2$, which can be done by comparing V_3 with V_2 by the F-test. It is now suspected that $\sigma_1^2 = \sigma_2^2$. If this is indeed the case it would be to our advantage to pool V_1 and V_2 and compare V_3 with $V = (n_1 V_1 + n_2 V_2)/(n_1 + n_2)$. If we are not completely certain that $\sigma_1^2 = \sigma_2^2$ it would seem reasonable to perform a preliminary test. If the hypothesis $H_0': \sigma_1^2 = \sigma_2^2$ is accepted, we would pool V_1 and V_2 and calculate $F = V_3/V$ to test $H_0: \sigma_3^2 = \sigma_2^2$; otherwise we would use $F = V_3/V_2$.

The form of the preliminary test depends on the model being considered. For the random model this pooling situation can be displayed as in Table 1; for the fixed model, as in Table 2. As can be seen from the expected mean squares, for the random model the preliminary test is the test of $H_0': \sigma_2^2 = \sigma_1^2$ against $H_1': \sigma_2^2 > \sigma_1^2$, and uses the statistic $F = V_2/V_1$ with n_2 and n_1 degrees of freedom. In Table 2 $\lambda_i \geqq 0$ is the noncentrality parameter of a noncentral χ^2. For the fixed model the preliminary hypothesis is $H_0': \sigma_1^2 = \sigma_2^2$ versus $H_1': \sigma_1^2 > \sigma_2^2$ and the statistic is $F = V_1/V_2$ with n_1 and n_2 degrees of freedom.

Received August 1973; revised June 1974.

AMS 1970 *subject classifications.* Primary 62J10; Secondary 62F05.

Key words and phrases. Incompletely specified models, preliminary tests of significance, pooling, analysis of variance, fixed models.

TABLE 1
Analysis of variance, random model

Source of Variation	d.f.	Mean Square	E.M.S.
Treatments	n_3	V_3	$\sigma_3^2 \geqq \sigma_2^2$
Error	n_2	V_2	$\sigma_2^2 \geqq \sigma_1^2$
Doubtful Error	n_1	V_1	σ_1^2

TABLE 2
Analysis of variable, fixed model

Source of Variation	d.f.	Mean Square	E.M.S.
Treatments	n_3	V_3	$\sigma_3^2 = \sigma_2^2\left(1 + \frac{2\lambda_3}{n_3}\right)$
Doubtful Error	n_1	V_1	$\sigma_1^2 = \sigma_2^2\left(1 + \frac{2\lambda_1}{n_1}\right)$
Error	n_2	V_2	σ_2^2

Bozivich, Bancroft, and Hartley (1956) discussed the random model case in considerable detail. They derived and evaluated formulas for the size and power of the procedure over a wide range of the parameters and they also made some recommendations concerning the choice of a significance level for the preliminary test. Another frequently occurring type of model is the mixed model. Bozivich, *et al.* showed how to reduce the mixed model to the random model, and the results in their paper could be used in such a situation.

The purpose of this paper is to parallel as nearly as feasible for the fixed model, what Bozivich, *et al.* did for the random model.

1.2. *More precise formulation.* Let us assume a fixed model. For simplicity of illustration, we let

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}\,,$$
$$\textstyle\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0\,, \qquad \text{and}$$
$$\varepsilon_{ijk} \sim NID(0, \sigma_2^2)\,,$$
$$i = 1, 2, \cdots, I, j = 1, 2, \cdots, J,\ k = 1, 2, \cdots, K\,,$$

for which the sum of squares can be partitioned as shown in the structure of the analysis of variance in Table 2. Our example above implies a two-way classification with equal numbers and more than one replication per cell. Considering α as the row effect and β as the column effect, then we may view either α or β as the "treatment" and $(\alpha\beta)$ as the set of terms which make the row $\times$ column interaction mean square V_1 a "doubtful error."

We wish to test the hypothesis H_0: $\sigma_3^2 = \sigma_2^2$ against the alternative H_1: $\sigma_3^2 > \sigma_2^2$. Then, if in Table 2 it is assumed that $\lambda_1 > 0$, the appropriate test procedure is to calculate $F_0 = V_3/V_2$ and reject H_0 if $F_0 \geqq F(\alpha, n_3, n_2)$ where $F(\alpha, n_3, n_2)$ is the upper 100α percent point of the central F-distribution with (n_3, n_2) degrees of freedom. This test proceure is called the never-pool test. If, however, it is assumed that $\lambda_1 = 0$, the expectation of V_1 is σ_2^2. In this case the test criterion would be $F_1 = (n_1 + n_2)V_3/(n_1 V_1 + n_2 V_2)$ and we reject H_0 if $F_1 \geqq F(\alpha, n_3, n_1 + n_2)$. We call this the always-pool test.

Frequently the experimenter is not willing to make any assumptions about λ_1. In such case he may wish to test the hypothesis H_0': $\sigma_1^2 = \sigma_2^2$. This test is referred to as the preliminary test of significance. The final test (of H_0) depends on the outcome of the preliminary test. If H_0' is rejected we use the test statistic $F_0 = V_3/V_2$; otherwise, we use $F_1 = (n_1 + n_2)V_3/(n_1 V_1 + n_2 V_2)$ for testing $\sigma_3^2 = \sigma_2^2$. This test is called the sometimes-pool test.

In our example of the two-way table we realize that having rejected H_0', i.e., having noted that some interactions are present, an investigator may proceed to study the nature of these interactions. This study may cause him to lose interest in the overall differences between rows, and in the test $F_0 = V_3/V_2$, because the presence of interactions implies that the differences between rows vary in some way from column to column, so that a more detailed summary of results is needed. The present investigation applies to cases where overall row comparisons are of interest whether interactions are present or not.

The essential features of the sometimes-pool test procedure related to the fixed model may be summarized by

(i) The error mean square V_2 is distributed as $\chi_2^2\sigma_2^2/n_2$ where χ_2^2 is the central χ^2 for n_2 degrees of freedom, while the other two mean squares V_i ($i = 1$ or 3) are distributed as $\chi_i'^2\sigma_2^2/n_i$ where $\chi_i'^2$ is the noncentral χ^2 based on n_i degrees of freedom and noncentrality parameter

$$\lambda_i = \frac{1}{2}\,\frac{n_i(\sigma_i^2 - \sigma_2^2)}{\sigma_2^2}\,.$$

The three V_i are independent.

(ii) The main purpose of the analysis is to test the null hypothesis $\sigma_3^2 = \sigma_2^2$ against the alternative $\sigma_3^2 > \sigma_2^2$.

(iii) The error mean square V_2 has an expectation σ_2^2 which is smaller than or equal to the expectation σ_1^2 of the doubtful error mean square V_1.

Clearly there are a number of models which satisfy the above conditions in addition to the example we have given above. Another example would be a polynomial regression model with more than one y value for each x value. In this situation the doubtful mean square would be the mean square due to lack of fit. The true error would be the within mean square. This procedure can be extended to a case of test of departure from linear regression in covariance analysis. (See Snedecor and Cochran (1967), page 460.)

1.3. *Related papers and objectives of the present study.* The problem to be discussed here is from the general area of incompletely specified models involving the use of preliminary tests of significance. Work in this area includes studies by Bancroft (1944), (1953), Paull (1950), Bechhofer (1951), Bozivich (1955), Lemus (1955), Bozivich, Bancroft, Hartley, and Huntsberger (1956), and others. Bozivich evaluated the size and power for many combinations of parameters for incompletely specified random models and made recommendations concerning the use of such models. For the fixed model, Bechhofer obtained the power in closed form for only a few special cases so that all numerical comparisons are restricted to small degrees of freedom and to the case $\sigma_1^2 = \sigma_2^2$.

Bancroft (1953) employed Patnaik's approximation to the noncentral χ^2 to develop approximate expressions for the power integrals in the case of Model I. He also made some empirical checks of the agreement between the exact and approximate values. These results are also given in the Wright Air Development Center Report by Bozivich, Bancroft, Hartley, and Huntsberger. Lemus gave a detailed derivation of these formulas and included some tabled values of the power. He also made some comparisons of the results from the approximate formulas with the result of a direct series evaluation. For $[n_1, n_2, n_3] = [4, 6, 2]$, $[6, 6, 6]$, $[8, 6, 8]$ and $[\lambda_1, \lambda_3] = [.6, 1]$, $[.06, .06]$, $[.05, .05]$, $[03, .03]$, the maximal deviations of the approximation from the direct series evaluation for P_1 and P_2 (defined in equation (2)) were .006 and .001 respectively.

The objective of the present study is to use the formulas of Bancroft in evaluating the power and the size of the sometimes-pool test. Numerical computations were carried out on an IBM 360 at the Iowa State University Computation Center for various values of degrees of freedom and parameters. It is shown that the power gain when $n_1 < n_2$ is generally negligible. When $n_1 > n_2$, there may be substantial gain in power as compared with the never-pool test.

2. Exact and approximate formulas for power. Fixed model.

2.1. *Mathematical formulation of the pooling procedure.* As stated in Section 1, we are interested in testing $H_0: \sigma_3^2 = \sigma_2^2$ (see Table 2) against the alternative $H_1: \sigma_3^2 > \sigma_2^2$. The test procedure with sometimes pooling V_1 and V_2 is as follows: Reject H_0 if

$$(1) \qquad \begin{aligned} &\text{either} \quad \{V_1/V_2 \geqq F(\alpha_1, n_1, n_2) \text{ and } V_3/V_2 \geqq F(\alpha_2, n_3, n_2)\} \\ &\text{or} \quad \{V_1/V_2 < F(\alpha_1, n_1, n_2) \text{ and } V_3/V \geqq F(\alpha_3, n_3, n_1 + n_2)\} \end{aligned}$$

where $V = (n_1 V_1 + n_2 V_2)/(n_1 + n_2)$ and $F(\alpha, n_i, n_j)$ is the upper 100α percent point of the F-distribution with n_i numerator degrees of freedom and n_j denominator degrees of freedom.

The power of the test procedure, in general the probability of rejecting H_0, may be written as the sum of the probabilities associated with the two mutually exclusive events given in (1), namely

$$(2) \qquad \begin{aligned} P_1 &= \text{Prob}\{V_1/V_2 \geqq F(\alpha_1, n_1, n_2) \text{ and } V_3/V_2 \geqq F(\alpha_2, n_3, n_2)\} \quad \text{and} \\ P_2 &= \text{Prob}\{V_1/V_2 < F(\alpha_1, n_1, n_2) \text{ and } V_3/V \geqq F(\alpha_3, n_3, n_1 + n_2)\}\,. \end{aligned}$$

This probability is a function of the degrees of freedom, n_1, n_2, and n_3, the noncentrality parameters, λ_1 and λ_3, and the levels of significance α_1, α_2, and α_3. Of the eight parameters only the significance levels can be considered under our control. To simplify computations we will try to keep the size near 0.05; a convenient way of doing this is to set $\alpha_2 = \alpha_3 = 0.05$. We will then consider several values of α_1 with the expectation of determining a value which will increase the power as much as possible without disturbing the size excessively.

2.2. *Recursion formulas for P_1 and P_2.* Patnaik (1949) showed that the sum of squares $n_i V_i$ ($i = 1$ or 3) with a noncentral chi-square distribution is approximately distributed as $\sigma_2^2 C_i \chi^2_{\nu_i}$ where $\nu_i = n_i + 4\lambda_i^2/(n_i + 4\lambda_i)$ and $C_i = 1 + 2\lambda_i/(n_i + 2\lambda_i)$. Following Bancroft (1953), we employ this approximation to derive the recursion formulas for P_1 and P_2.

Defining the following new variables

$$(3) \qquad u_1 = \frac{n_3 V_3}{n_2 V_2 C_3}, \qquad u_2 = \frac{n_1 V_1}{n_2 V_2 C_1}, \qquad w = \frac{n_2 V_2}{2\sigma_2^2}$$

and the new constants

$$(4) \qquad u_1^0 = \frac{n_3}{n_2 C_3} F_{n_3 n_2}(\alpha_2), \qquad u_2^0 = \frac{n_1}{n_2 C_1} F_{n_1 n_2}(\alpha_1),$$

$$u_3^0 = \frac{n_3}{n_1 + n_2} F_{n_3, n_1+n_2}(\alpha_3)$$

and integrating out w, we obtain, from the joint density of $n_i V_i / C_i \sigma_2^2 \sim \chi^2_{\nu_i}$, $i = 1, 2, 3$, the joint density of u_1 and u_2

$$(5) \qquad f(u_1, u_2) = K' u_1^{\frac{1}{2}\nu_3 - 1} u_2^{\frac{1}{2}\nu_1 - 1} (1 + u_1 + u_2)^{-\frac{1}{2}(\nu_1 + n_2 + \nu_3)}$$

where $K' = \Gamma\{\frac{1}{2}(\nu_1 + n_2 + \nu_3)\}[\Gamma(\frac{1}{2}\nu_1)\Gamma(\frac{1}{2}n_2)\Gamma(\frac{1}{2}\nu_3)]^{-1}$.

In terms of the new variables, the rejection region given in (1) may be restated as: Reject H_0 if

$$(6) \qquad \text{either} \quad \{u_2^0 \leqq u_2 < \infty \ \text{ and } \ u_1^0 \leqq u_1 < \infty\}$$

$$\text{or} \qquad \left\{0 \leqq u_2 < u_2^0 \ \text{ and } \ u_3^0 \frac{(1 + C_1 u_2)}{C_3} \leqq u_1 < \infty\right\}.$$

With this definition, P_1 and P_2, the two components of the lower P, may be rewritten as integrals, over the proper limits given in (6), of the joint density of u_1 and u_2. To further simplify evaluation of these expressions Lemus (1955) gave the following set of recursion formulas. For P_1:

$$(7) \qquad P_1(a+1, b) = P_1(a, b) + \frac{1}{(a+1)B(a+1, \frac{1}{2}n_2)} \left(\frac{1 + X_2}{X_1}\right)^{a+1} \times \left(\frac{X_1 + X_2 - 1}{X_1}\right)^{\frac{1}{2}n_2} I_{X_1}(a + 1 + \tfrac{1}{2}n_2, b + 1)$$

where

$$a = \tfrac{1}{2}\nu_3 - 1, \qquad b = \tfrac{1}{2}\nu_1 - 1; \qquad X_1 = \frac{1 + u_1^0}{1 + u_1^0 + u_2^0}, \qquad X_2 = \frac{1 + u_2^0}{1 + u_1^0 + u_2^0}$$

and B and I_X denote the complete Beta function and normalized incomplete Beta function, respectively. The necessary initial value is given by

$$P_1(0, b) = \left(\frac{X_1 + X_2 - 1}{X_1}\right)^{\frac{1}{2}n_2} I_{X_1}(\tfrac{1}{2}n_2, b + 1) . \tag{8}$$

The corresponding formulas for P_2 are

$$\begin{aligned} P_2(a, b) &= (1 - T)P_2(a - 1, b) + TP_2(a, b - 1) \\ &\quad - \frac{T(1 - Q)^b Q^{\frac{1}{2}n_2}}{b\mathrm{B}(b, \frac{1}{2}n_2)} I_{X_3}(b + \tfrac{1}{2}n_2, a + 1) \end{aligned} \tag{9}$$

with

$$Q = \frac{1}{1 + u_2^0}, \qquad T = \frac{C_3}{C_3 + C_1 u_3^0}, \qquad X_3 = \frac{C_3(1 + u_2^0)}{C_3(1 + u_2^0) + u_3^0(1 + C_1 u_2^0)} .$$

The two initial values are

$$P_2(0, b) = \left[\frac{QX_3}{T - X_3(1 - Q)}\right]^{\frac{1}{2}n_2} T^{b+1+\frac{1}{2}n_2} I_{X_1'}(b + 1, \tfrac{1}{2}n_2) , \tag{10}$$

$$\begin{aligned} P_2(a, 0) &= I_{X_2'}(\tfrac{1}{2}n_2, a + 1) - Q^{\frac{1}{2}n_2} I_{X_3}(\tfrac{1}{2}n_2, a + 1) \\ &\quad - \left[\frac{QX_3}{X_3 - T}\right]^{\frac{1}{2}n_2} (1 - T)^{a+1+\frac{1}{2}n_2} \{I_{X_3'}(\tfrac{1}{2}n_2, a + 1) \\ &\quad - I_{X_4'}(\tfrac{1}{2}n_2, a + 1)\} \end{aligned} \tag{11}$$

where

$$X_1' = \frac{X_3(1 - Q)}{T} , \qquad X_2' = \frac{QTX_3}{T - X_3(1 - Q)} ,$$

$$X_3' = \frac{T(X_3 - T)}{(1 - T)(T - X_3(1 - Q))} , \qquad X_4' = \frac{X_3 - T}{1 - T} .$$

These recursion formulas were used to develop a set of master tables for computing P_1 and P_2.

3. Discussion of the size and power of the sometimes-pool test.

3.1. *Size*. The size of the sometimes-pool test can be obtained from P_1 and P_2. The probability of a type I error is computed by setting $\sigma_3^2 = \sigma_2^2$. When it is plotted on a graph, it is called the size curve by Bozivich *et al.* (1956). We shall adopt this name. The behavior of the size curve is illustrated in Figures 1 and 2. The degrees of freedom in Figure 1 are $n_1 = 9$, $n_2 = 16$ and $n_3 = 3$, which is obtained by letting $I = 4$, $J = 4$ and $K = 2$ in the model in Section 1.2. Here the degrees of freedom for the doubtful error (n_1) are smaller than the degrees of freedom of the error (n_2). This is a common case in fixed models. In Figure 2, we let $n_1 = 21$, $n_2 = 8$ and $n_3 = 7$. Here n_1 is larger than n_2. It can occur in a two-way classification with unequal and proportional subclass frequencies. This would occur when the experimenter repeated some, but not all, of the treatments the same number of times in each block in a randomized block

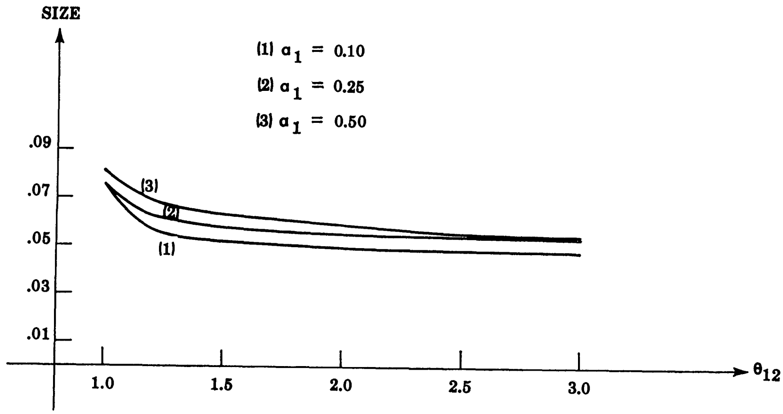

FIG. 1. Size curve $[n_1, n_2, n_3] = [9, 16, 3]$ $\alpha_2 = \alpha_3 = .05$.

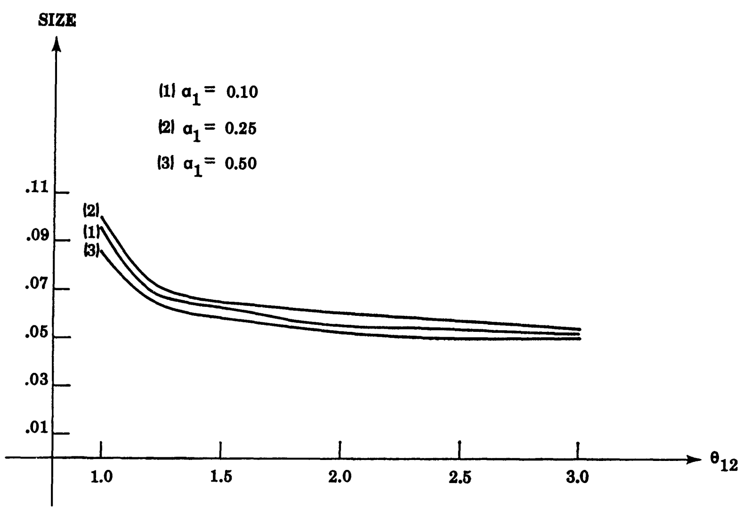

FIG. 2. Size curve $[n_1, n_2, n_3] = [21, 8, 7]$ $\alpha_2 = \alpha_3 = .05$.

design. An example of such a situation is given in Bancroft (1968, problem 1.1 on page 30). In fact we have used that example for setting up the degrees of freedom in Figure 2. In our computation, we let $\alpha_2 = \alpha_3 = .05$ and $\alpha_1 = .10$, .25, .50. By examining these figures (other figures, not shown here, were also examined for values of $[n_1, n_2, n_3] = [2, 6, 2]$; $[6, 12, 2]$; $[6, 12, 3]$; $[9, 8, 3]$), it is seen that the probability of a type I error is the largest (size peak) at $\theta_{12} = \sigma_1^2/\sigma_2^2 = 1$; it decreases rather rapidly to .05 as θ_{12} increases. This behavior is very different from that of the random model. As Bozivich *et al.* (1956) showed, the size curve in the random model has its minimum at $\sigma_1^2 = \sigma_2^2$, then it increases to a maximum before it decreases to .05. In the fixed model, the value of the size at its peak is less than .10 for the cases considered. The size curve is usually above .05 except for the case $[n_1, n_2, n_3] = [9, 16, 3]$ and $\alpha_1 = .10$ where the size curve is slightly less than .05 (it is .048) when θ_{12} is large. Hence we may say that the sometimes-pool test in the fixed model is not conservative in type I error. (By conservative, we mean that the type I error falls below the nominal level .05.)

3.2. *Power.* We now consider the comparison of power between the sometimes-pool test and the never-pool test. In making this comparison we assume that the experimenter will use either the never-pool test procedure or the sometimes-pool test procedure; that is, he will not use the always-pool test procedure, since the never-pool test procedure gives unbiased estimates of error variance whether interaction is present or not and is usually used in practice. In Bozivich *et al.* (1956), the comparison was made when the parameter $\theta_{12} = \sigma_1^2/\sigma_2^2$ is fixed. Similarly, we shall first consider the fixed parameter case, then consider the case when θ_{12} is not fixed.

TABLE 3

Power comparison of the sometimes-pool test and the never-pool test

$[n_1, n_2, n_3] = [21, 8, 7]$ $\alpha_1 = .25$ $\alpha_2 = \alpha_3 = .05$

θ_{12}	Test	θ_{32}					
		1.00	1.20	1.43	1.81	2.15	3.41
1.00	s.p.	.099	.155	.215	.342	.450	.772
	n.p.	.099	.146	.204	.305	.395	.678
1.20	s.p.	.074	.131	.197	.305	.402	.728
	n.p.	.074	.112	.160	.247	.328	.605
1.43	s.p.	.066	.110	.168	.280	.376	.714
	n.p.	.066	.100	.145	.227	.304	.576
1.81	s.p.	.058	.093	.142	.232	.319	.674
	n.p.	.058	.089	.130	.206	.279	.544
2.15	s.p.	.055	.087	.129	.215	.292	.645
	n.p.	.055	.085	.124	.198	.269	.531
3.41	s.p.	.052	.081	.118	.189	.255	.559
	n.p.	.052	.081	.119	.190	.259	.517

When θ_{12} is fixed, the method of power comparison is to evaluate the size of the sometimes-pool test for this fixed θ_{12}, then for this level of size, evaluate the power curve of the never-pool test for given $\theta_{32} = \sigma_3^2/\sigma_2^2$; this power is then directly comparable with that of the sometimes-pool test corresponding to the fixed value of θ_{12}. Table 3 gives an example to illustrate the power comparison for $[n_1, n_2, n_3] = [21, 8, 7]$, $\alpha_1 = .25$ and $\alpha_2 = \alpha_3 = .05$. The power of the sometimes-pool test in Table 3 is always larger than that of the never-pool test except for $\theta_{12} = 3.41$. The difference of the two power curves is the power gain. Figures 3, 4, 5 and 6 demonstrate the power gain as a function of α_1 for $[n_1\ n_2, n_3] = [21, 8, 7]$. It is seen that on the whole the power gain is largest when $\alpha_1 = .25$. By examining the graphs, the power gain increases as θ_{32} increases. The gain is large for small values of θ_{12}, say $1.0 < \theta_{12} < 1.6$, then it decreases to zero as θ_{12} increases.

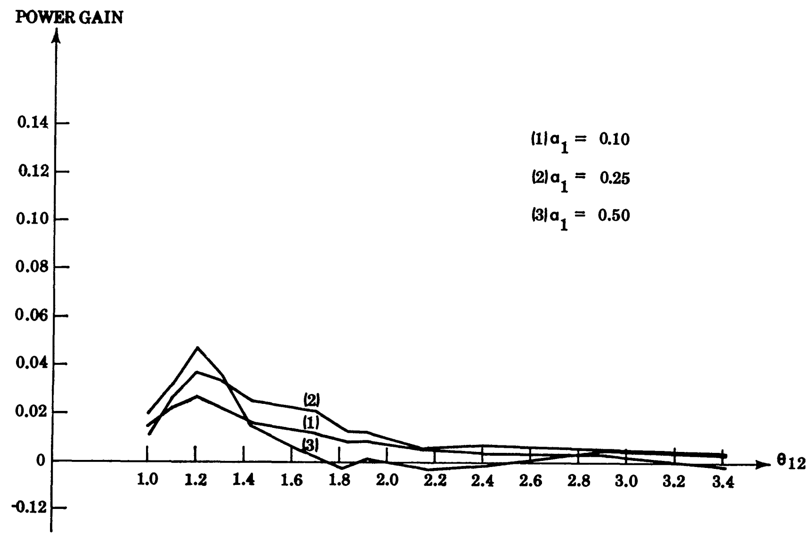

FIG. 3. Power gain $[n_1, n_2, n_3] = [21, 8, 7]$ $\alpha_2 = \alpha_3 = .05$ $\theta_{32} = 1.43$.

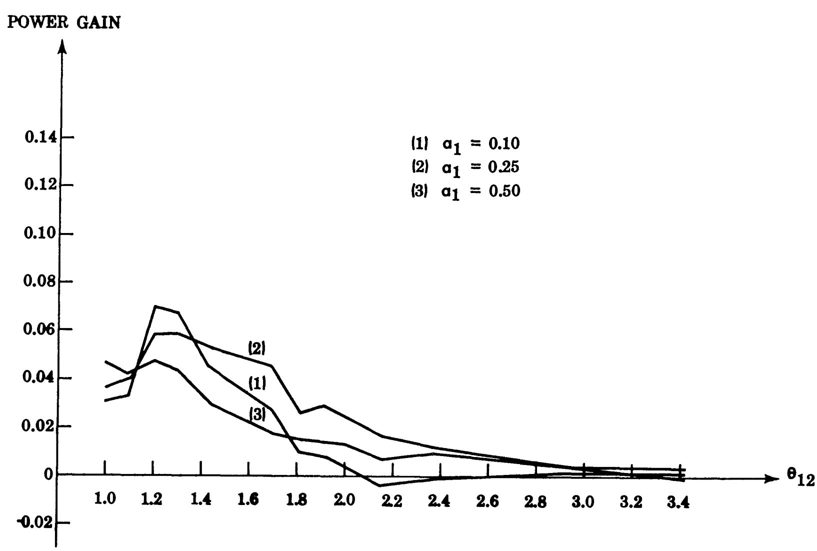

FIG. 4. Power gain $[n_1, n_2, n_3] = [21, 8, 7]$ $\alpha_2 = \alpha_3 = .05$ $\theta_{32} = 1.81$.

The above power comparisons are made for fixed θ_{12}. The procedure used was given in the paper by Bozivich *et al.* (1956). Such a methodological study is important to experienced experimenters in providing information on power gain to be expected for a range of values of θ_{12}. In case the investigator has no knowledge of the value of θ_{12}, he may treat θ_{12} as not fixed. Then it would be necessary to use the size peak in the fixed model at $\theta_{12} = 1$ for the size of the test. In such case the power of the sometimes-pool test is compared with that of the never-pool test at that size level. In Table 3 of our example, the power

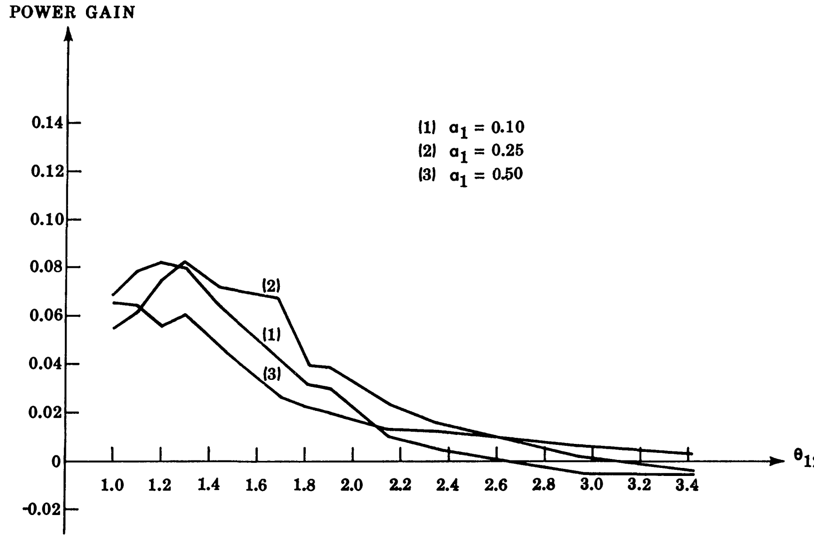

FIG. 5. Power gain $[n_1, n_2, n_3] = [21, 8, 7]$ $\alpha_2 = \alpha_3 = .05$ $\theta_{32} = 2.15$.

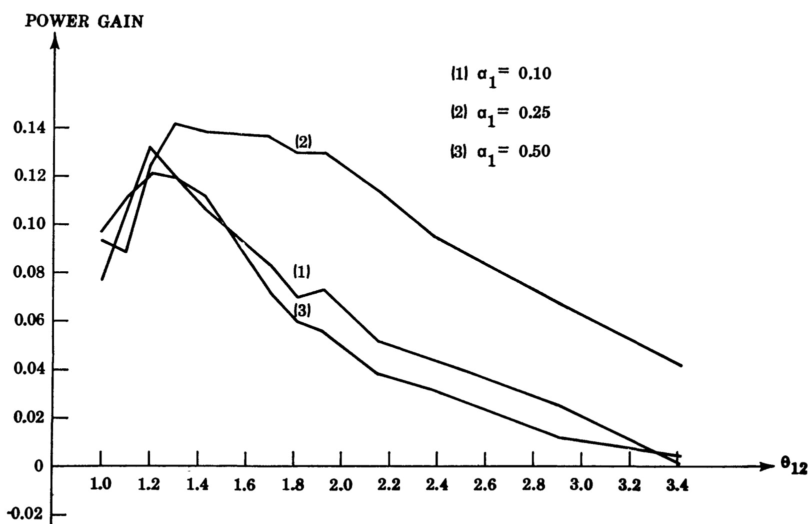

FIG. 6. Power gain $[n_1, n_2, n_3] = [21, 8, 7]$ $\alpha_2 = \alpha_3 = .05$ $\theta_{32} = 3.41$.

of the sometimes-pool test is compared with never-pool at size level .099. Such a comparison shows that the power gain when $\theta_{12} = 1$ remains the same as before. But when $\theta_{12} > 1$, the power gain reduces because the comparison is made at a higher level of the never-pool test when θ_{12} is not fixed. For $\theta_{12} > 1$, the sometimes-pool test has smaller power except when θ_{32} is large.

In the above example we let $n_1 > n_2$, which can occur in designs with unequal subclass frequencies. One other case we examined is $[n_1, n_2, n_3] = [9, 8, 3]$. Here the error and doubtful error degrees of freedom are about the same. We found that the power gain is positive for most parameter values considered and $\alpha_1 = .50$. When the level of the preliminary test is .25 or .10, the power gain is negative for most cases.

We also observed the behavior of power for $[n_1, n_2, n_3] = [2, 6, 2]$; $[6, 12, 2]$; $[6, 12, 3]$; $[9, 16, 3]$. The power gain in these cases is either small or negative. This may be due to the fact that the doubtful error degrees of freedom are small relative to the error degrees of freedom; hence the gain in degrees of freedom for the sometimes-pool test is small. At the same time, the size disturbance has offset the gain completely. Therefore there is little or no gain for the sometimes-pool test.

4. Recommendations for the user. The purpose of examining the behavior of size and power of the sometimes-pool test is to lead us to make recommendations about the level of the preliminary test. Based on the observations we made earlier, we may make the following general recommendation. When the doubtful error degrees of freedom are considerably larger than the error degrees of freedom, say $n_1 > 2n_2$, the level of the preliminary test should be set about .25; when n_1 and n_2 are about equal, one should choose $\alpha_1 = .50$; if n_1 is smaller than n_2 and n_2 is reasonably large, the never-pool test procedure should be used, because our evidence is that the power gain is at best small, while there may be a loss. This recommendation will yield the largest power gain, and on the other hand, the type I error will be slightly greater than the nominal .05.

The above recommendation provides a warning as regards the indiscriminate use of preliminary tests in the case of fixed models. In most fixed models, the doubtful error degrees of freedom are smaller than the error degrees of freedom and the latter are usually reasonably large. The sometimes-pool test usually increases the size of the test and it in turn wipes out the power gain resulting in the small increase in degrees of freedom. This situation is a reverse of the random models considered by Bozivich *et al.* (1956) where the doubtful error degrees of freedom is usually larger.

Acknowledgment. We wish to thank Mr. Wendell Smith of the Statistical Laboratory at Iowa State University for his help in computation. We are grateful to the editors and referees for their valuable comments.

REFERENCES

[1] BANCROFT, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.* **15** 190–204.

[2] BANCROFT, T. A. (1953). Certain approximate formulas for the power and size of a general linear hypothesis incorporating a preliminary test of significance. Unpublished preliminary report. Iowa State University of Science and Technology, Statistical Laboratory.

[3] BANCROFT, T. A. (1968). *Topics in Intermediate Statistical Methods.* Iowa State Univ. Press.

[4] BECHHOFER, R. E. (1951). The effect of preliminary tests of significance on the size and power of certain tests of univariate linear hypotheses. Unpublished Ph. D. thesis. Columbia Univ.

[5] BOZIVICH, H. (1955). Power of test procedures for certain incompletely specified random and mixed models. Unpublished Ph. D. thesis. Iowa State Univ. of Science and Technology.

[6] BOZIVICH, H., BANCROFT, T. A., and HARTLEY, H. O. (1956). Power of analysis of variance test procedures for certain incompletely specified models. *Ann. Math. Statist.* **27** 1017–1043.

[7] BOZIVICH, H., BANCROFT, T. A., HARTLEY, H. O. and HUNTSBERGER, D. V. (1956). Analysis of variance; preliminary tests, pooling, and linear models: preliminary tests of significance and pooling procedures for certain incompletely specified models. Wright Air Development Center Technical Report 55-244 **1** 1–118.

[8] LEMUS, F. (1955). Approximations to distributions in certain analysis of variance tests. Unpublished M.S. thesis. Iowa State Univ.

[9] PATNAIK, P. B. (1949). The non-central χ^2 and F-distributions and their applications. *Biometrika* **36** 202–232.

[10] PAULL, A. E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. *Ann. Math. Statist.* **21** 539–556.

[11] SNEDECOR, G. W. and COCHRAN, W. G. (1967). *Statistical Methods.* Iowa State Univ. Press.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50010