AD-A104 004 COLORADO UNIV AT BOULDER CENTER FOR RESEARCH ON JUDG--ETC F/G 12/2
DETECTION OF MULTIPLICATIVE SYNERGISMS IN SIMULATED DATA FOR NO--ETC(U)
AUG 81 J SHANTEAU N00014-77-C-0336
UNCLASSIFIED CRJP-235 NL

1 of 1
AD A 104004

END
DATE FILMED
10-81
DTIC

AD A104004
LEVEL II
12

(12)

Detection of Multiplicative Synergisms in Simulated Data for Nonorthogonal Designs: What Lies Beyond Linearity?

James Shanteau

UNIVERSITY OF COLORADO

INSTITUTE OF BEHAVIORAL SCIENCE

and

DEPARTMENT OF PSYCHOLOGY

KANSAS STATE UNIVERSITY

Center for Research on Judgment and Policy

Report No. 235

August 1981

This research was supported by the Engineering Psychology Programs, Office of Naval Research, Contract N00014-77-c-0336, Work Unit Number NR 197-038 and by BRSG Grant #RR07013-14 awarded by the Biomedical Research Support Program, Division of Research Resources, NIH. Center for Research on Judgment and Policy, Institute of Behavioral Science, University of Colorado. Reproduction in whole or in part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) CRJP-235	2. GOVT ACCESSION NO. AD-A204 004
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)* (6) Detection of Multiplicative Synergisms in Simulated Data for Nonorthogonal Designs: What Lies Beyond Linearity?	5. TYPE OF REPORT & PERIOD COVERED (9) Technical rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) James Shanteau	8. CONTRACT OR GRANT NUMBER(s) (15) N00014-77-C-0336, PHS-RR-07023
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Research on Judgment and Policy Institute of Behavioral Science University of Colorado, Boulder, CO 80309	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (12) 69
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Street Arlington, VA 22217	12. REPORT DATE (11) August, 1982 13. NUMBER OF PAGES 68
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*	15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; Distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

synergism, simulation, regression model, hierarchical test, cue intercorrelation

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A three stage simulation was conducted to evaluate the impact that multiplicative synergisms have within nonorthogonal (regression) designs. In the first stage, various designs with differing cue intercorrelations and number of stimulus cases were generated. In the second stage, several response strategies were simulated, including adding, multiplying, and adding-multiplying strategies. In the final stage, various research techniques, involving both descriptive and inferential analyses, were applied to

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

4-20647

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

analyze the simulated data. The results revealed that (1) while synergisms produce notable effects throughout the range of intercorrelation values, the most striking influence was found for negative intercorrelations, (2) descriptive indices (such as R^2) were relatively unrevealing about the presence of synergisms, (3) in contrast, inferential indices (particularly the hierarchical test) were much more revealing, and (4) there were considerable between-stimulus differences in terms of how easily synergisms could be accurately detected. These findings imply that, because synergisms have been frequently overlooked by investigators, they should be specifically tested for in future research using nonorthogonal designs.

Detection of Multiplicative Synergisms in Simulated Data for Nonorthogonal Designs: What Lies Beyond Linearity?

According to Webster's (1976) dictionary, a synergism is an action "such that the total effect is greater than the sum of the effects taken independently." In psychological terms, synergisms are usually represented by a multiplicative interaction between two or more variables. For instance,

$$R = f\ (A \times B), \qquad (1)$$

where the response R is a function of the multiplicative combination of variables A and B. A number of behavioral models and theories take the form of Equation 1. Examples of synergisms in psychology can be found in learning (performance = drive x habit strength), industrial (performance = ability x motivation), and social (level of aspiration = desirability of goal x expectancy of success). For a more complete discussion of the importance of synergism, along with numerous psychological examples, see the review by Shanteau (1981).

In experimental designs which are orthogonal, such as factorial designs, detection of a mulitplicative synergism is relatively straightforward. In analysis of variance terms, a synergism will lead to an interaction. Moreover, the interaction will take a specific bilinear form, i.e., the data will plot as a diverging fan of lines. While several statistical procedures are available, the most generally useful technique is to evalute the linear x linear component of the interaction. The various procedures are discussed and compared in Shanteau (1978, 1981).

In contrast to the orthogonal case, the detection of synergisms in nonorthogonal designs is much less clear cut. Nonorthogonal designs typically arise when the levels of variables are assigned in some non-controlled fashion, e.g., as in a representative design Brunswik, 1956). Typically, this

means that the variables will be intercorrelated with each other. The problems caused by such intercorrelations are well recognized in the literature on multiple-regression analyses (e.g., Darlington, 1969). However, the difficulties introduced by intercorrelation in the detection of synergisms have yet to be analyzed. Therefore, the purpose of the present study is to evaluate the effects of synergisms on nonorthogonal designs.

Research Strategy

While it would be desirable to use real data sets for analyzing synergisms, this is impractical on two grounds: (a) It is difficult, if not impossible, to know in advance whether a synergism does or does not exist in a given set of data. Without such knowledge, any further analyses would be fruitless for present purposes. (b) There is no feasible way to obtain data sets which correspond to all the conditions which would be desirable to examine. Moreover, any real data sets are likely to have been influenced by a variety of other unknown and probably idiosyncratic factors.

Because of these problems, an alternative research strategy was developed based on the construction of simulated data sets. There are three noteworthy features to such data sets: First, the presence (or absence) of a synergism can be built into simulated data. That is, a synergism can be specifically included or excluded in the simulated data-generating model. Thus, the "truth" about the simulated data is known a priori.

Second, various properties of an experimental design can be easily and systematically manipulated with simulated data. For instance, the degree of cue intercorrelation between two variables can be varied from high positive to high negative with numerous steps in between. Other properties, such as the size of the design can also be easily specified. Therefore, an advantage of simulated data is that potentially important properties of the design can be varied in predetermined ways.

Third, nonorthogonal designs with the same properties, can be produced in many different ways. That is, a variety of stimulus sets can have the same degree of intercorrelation values, etc. This lack of uniqueness in nonorthogonal designs makes it desirable to compare alternative stimulus sets with the same correlations. Thus, simulated data sets allow a direct way to evaluate the consistency of any results involving synergisms.

Of course, the use of simulations also has its disadvantages. Some of these will be taken up in more detail in the Discussion. At this point, however, it is worth emphasizing that every effort was made to produce data sets which "look like" real data. Towards this end, a variety of models were used which resemble those which are known to be used by subjects. In addition, error was added to the data at levels which are similar to the values found in typical response sets. In short, the simulated data had all the outward appearances of realistic data.

The remaining sections of the paper begin with a detailed description of the simulation technique used. This also includes a consideration of the procedures, analyses, etc., employed. Then the results from the analyses of over 14,000 simulated data sets are presented. Finally, the last section contains a discussion of the implications, as well as qualifications, of the present results.

Simulation Approach

The basic goal behind the present research approach was to separate a typical experimental study into three stages and to simulate each stage separately. As shown in Table 1, these stages correspond to the construction of an experimental design (the environment), the formation of the responses according to some strategy (the subject), and the analysis and interpretation of the results (the experimenter). Since independent algorithms were constructed for each stage, the simulation techniques will be considered separately.

Insert Table 1 about here

Environmental Simulation

The experimental design specifies the stimulus environment in which research evidence is collected. Obviously, a subject can only provide responses to the particular stimulus combinations presented. This means that the experimental design can play a crucial role in determining whether a specific relation, such as a synergism, can or cannot be detected in the data.

In the present case, the construction of each stimulus design involved a two-step process. The first step was based on tentatively constructing a stimulus set intended to reflect various prespecified conditions. The second step was based on examining the tentative set to see if, in fact, the desired conditions had been met. For instance, if a prespecified level of cue intercorrelation was desired, then the observed intercorrelation value for a tentative set was compared to the desired value. If the stimulus set met all the conditions, then it was used. If not, and if minor adjustments did not produce a satisfactory stimulus set, then the set was discarded and the process started over. Both of these steps will now be considered in more detail.

Cue generation. A computer algorithm incorporated in the program CUEGEN (see the Appendix for details) was used to construct sets of nonorthogonal stimuli with specifiable properties. Some of these properties were arbitrarily fixed for purposes of this research project, while others were varied. The following four properties were held constant across all stimulus sets: (1) Only two-cue stimulus sets were used (see Table 2 for an example). This allowed complete freedom to vary intercorrelations from +1 to -1; the

use of three or more cues would have reduced the range of correlation values possible. (2) The range of the cue values was restricted from 0 to 100. This restriction was purely for convenience and involves no loss of generality. (3) A uniform sampling distribution was specified for each cue. Although other distributions such as normal were considered, it proved to be easier to construct appropriate stimulus sets using the uniform. (4) The mean and standard deviation were specified to be 50.0 and 20.0, respectively. As in the case for the other fixed properties, these arbitrary values appeared to have little impact on the pattern of results observed.

Insert Table 2 about here

There were three properties which were varied systematically in the construction of the stimulus sets: (1) The intercorrelation values were specified to be +.90, +.75, +.50, +.25, .00, -.25, -.50, -.75, and -.90. These values both covered the range of possible intercorrelations and were reasonably dense and well-spaced. (2) The number of stimuli in each stimulus set was specified to be either 25 or 100. These numbers are representative of what is typically used in "small" and "large" nonorthogonal designs. (3) Nine independent stimulus sets were generated for each combination of correlation value and stimulus set size. There were thus a total of 9 (intercorrelations) x 2 (stimulus set sizes) x 9 (independent sets) = 162 stimulus sets generated.

Testing the stimulus sets. Each constructed set was subjected to a series of tests to check whether it was close to the desired properties, such as the specified intercorrelation value. Only if a stimulus set satisfied all the tests was it kept. Otherwise, the cue values were randomly

adjusted and the test repeated. If after five adjustments, the stimulus set was still not satisfactory, then it was discarded and a new set was generated using the CUEGEN program.

There were four types of tests that each stimulus set had to pass at the .05 level in order to be acceptable: (1) The intercorrelation between the cue values (i.e., between paired entries in Table 2) had to be nonsignificantly different from the desired value. (2) The mean for each cue separately (i.e., each column in Table 2) had to show nonsignificant deviations from the desired value. (3) Similarly, the standard deviation for each cue had to be nonsignificantly different from the specified valve. (4) Finally, the distribution of values for each cue separately had to approach the uniform distribution. (To make this test, each distribution was divided into segments and deviations from the uniform were computed for each segment. A chi-square procedure was then used to test for any discrepancies.)

In all, each constructed stimulus set had to pass seven tests: one for the intercorrelation value and two each for means, standard deviations, and uniform distribution. In practice, almost all stimulus sets satisfied the intercorrelation restriction. Tests on means and standard deviations resulted in a few rejected sets. However, the distribution test was the most demanding and produced by far the highest proportion of rejected sets. However, it was possible in all cases to construct nine independent stimulus sets which satisfied each of the tests.

Subject Simulation

To simulate the subjects' behavior in an experiment, a two-step process was followed. First, a model was specified, e.g., multiplying, for each simulated subject. This model represents the "truth" to be detected by the

subsequent analysis. Second, random error was introduced to [illegible] more realistic responses. Each of these two steps will now be considered in more detail.

Model. Three different models were used in these simulations. The first was a multiplying model in which the two cues, X_1 and X_2, were combined as follows:

$$Y_m = X_1 \times X_2. \quad (2$$

For the example shown in Table 2, the first pair of values for X_1 and [illegible] 67.4 and 44.3, respectively. The product, Y_m, would be 67.4 x 44.3 = [illegible]. This model represents a "pure synergism" in the form of crossproduct.

The second model involved adding the two cues:

$$Y_a = X_1 + X_2. \quad (3)$$

For the first pair of values in Table 2, Y_a would be 67.4 + 44.3 = 111.7. This model provides a baseline or control condition in which a synergism is known not to exist.

The third model was a combination adding-multiplying model:

$$Y_c = X_1 + X_2 + (X_1 \times X_2). \quad (4)$$

For the example above, Y_c would be 67.4 + 44.3 + (67.4 x 44.3) = 3097.5. This combination model allows examination of a synergism in the context of an adding process.

Error. To produce realistic data, error obviously must be introduced. In considering error, two major choices have to be made which correspond to the location and the size of the error term. In regard to location, error can be added either before or after the cue values are combined in Equations 2 to 4. Introducing error before the combination process corresponds to variability produced by stimulus or perceptual uncertainty. Introducing error after the combination corresponds to variability caused by response or output uncertainty.

For two reasons, it was decided to add error after the combination of stimulus values. That is, random error, E, was added to each Y value in Equations 2 to 4 so that the simulated data, Y', can be expressed as:

$$Y' = Y + E. \quad (5)$$

The first reason is that statistical models typically assume a post-combination additive error. Since one of the goals of this study was to evaluate various statistical techniques, it seemed preferable to make the data consistent with assumptions made by the statistical analyses.

A second reason is that there is some limited empirical support for the error-after location. Shanteau (1970) examined within-cell variability in a task known to produce synergistic behavior, i.e., a gambling task. The cell variances were found to be unrelated to the cell means; this is consistent with Equation 5, but inconsistent with a before-combination of error.

The other major choice involves the <u>size</u> of the error term to be added. It should be clear, since the ranges of the Y values for the three models are quite different, that the same size error cannot be used for all three models. Instead, the size of the error was calibrated individually to reflect each set of Y values.[6] Thus, the Y' values were produced by adding to each Y a random normal error value, with mean zero and standard deviation equal to c. After trial-and-error exploration, it was found that a c value equal to one-half of the coefficient-of-variation for the Y values produced the most reasonable looking results, i.e., c = ½ x Standard Deviation ÷ Mean. A variety of other ways of defining c were explored, and in general the pattern of the results did not appear to depend on the definition of c used. (Some additional comments on issues related to error appear in the Discussion).

Using the approach outlined above, ten independent data sets were generated for each stimulus set. In other words, from each set of Y values, different

combinations of random error were added to form ten set of Y' values.[1] In summary, ten simulated subjects were created for each of the Y sets of stimulus values.

Experimenter Simulation

For present purposes, the experimenter's role is that of analyzing and comparing the results for various statistical methods. Since there are a variety of methods which can be used, the experimenter's question becomes: which of the various statistical techniques is most sensitive to the presence (or absence) of a synergism? Before dealing with this question, however, it will first be necessary to review the overall multiple-regression approach used.

Multiple-regression analysis. Each of the simulated subjects (i.e., each set of Y' values) was analyzed by three types of multiple-regression models. These models correspond to the three models used to create the data in Equations 2, 3, and 4. Thus, each data set was analyzed using a multiplicative or pure cross—product regression model:

$$\hat{Y}_m = \beta_1 (X_1 \times X_2), \tag{6}$$

where $\hat{Y}_m$ is a predicted value derived from the product of the cue values X_1 and X_2. Similarly each data set was analyzed using an additive or linear regression model:

$$\hat{Y}_a = \beta_1 X_1 + \beta_2 X_2, \tag{7}$$

where the predicted value Y_a is obtained from a weighted sum of the stimulus values. Finally each data set was analyzed by a multilinear (combined additive-multiplicative) regression model:

$$\hat{Y}_c = \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2), \tag{8}$$

where the predicted $\hat{Y}_c$ value is a weighted sum of the various terms.

In the actual analyses, ordinary least squares procedures were used to obtain estimates of the β weights. These estimates minimize the discrepancy between the original Y' values and the derived $\hat{Y}$ values. The raw regression analyses were performed on the data for each simulated subject. For convenience, however, the results have been averaged across the ten subjects generated for each stimulus set. (As an aside, the zero intercept and its weight, β_0, have been omitted for clarity from Equations 6 to 8. These values were invariably near zero and contributed nothing to the interpretation of the results.)

Statistical indices. Based on the multiple-regression analyses, a number of statistical measures were computed. These measures are considered in some detail in the Appendix. However, because the results for many of the indices were redundant, only the most relevant measures will be considered in the results. These indices, which are listed across the top of Tables 3 and 4, will be described as necessary in the Results section.

Results

Since 14,580 separate multiple-regression analyses were run, it is obviously necessary to be highly selective in presenting results. For brevity, graphical summaries will be presented of just the most relevant statistical results. In addition, only short descriptions will be given of the indices (see the Appendix for a more complete description of the statistical procedures).

Descriptive Indices

The squared multiple-correlation value, R^2, provides a standard measure of the variance-accounted-for by a regression model. Figure 1 shows the average R^2 values for the fit of an additive (linear) regression model (Equation 7); the left panel gives the results for the 25-stimulus condition

and the right panel gives the results for the 100-stimulus condition. The three lines in each panel correspond to the three data-generating models in Equations 2 to 4, with the nine levels of cue intercorrelation listed along the horizontal axis. Each of the plotted values in Figure 1 is the average of 90 multiple-regression analyses: ten simulated subjects in each of nine stimulus sets.

Insert Figure 1 about here

Three general observations follow directly from this and the other figures. First, the results appear to be reasonably smooth and lawful. Moreover, several trends both between curves and across correlation values are easily discernable. Thus, at least at a surface level, the present simulation approach has produced orderly data.

Second, the multiplying data model (Equation 2) and the adding-multiplying model (Equation 4) lead to essentially identical results. With few exceptions, the statistical indices produced by these two models are virtually indistinguishable. This suggests that what a synergism is combined with is not as important as the fact that a synergism is present.

Third, the adding data model (Equation 3) led to results which are consistently lower than the results for the other two data-generating models. In some circumstances, this might be expected since the adding model serves as a baseline for many of the analyses. For indices such as R^2, however, the lower values were unexpected. The reason for this result apparently lies in the procedure used to add error to the simulated data. Using the coefficient of variation to determine the size of the error seems to have introduced relatively more error into the adding data than into the

other two data models. Hence, the R^2 are proportionally lower. Fortunately, this difference does not influence any of the major findings reported here (although it does suggest that some attention be given to the issue of compatible error values in any future simulation research).[2]

One unique finding in Figure 1 is that the results in the two panels are remarkably similar. For instance, the curves for the synergistic models are nearly straight for positive intercorrelations with a gradual increase to around .70. Thus, for positively correlated cues, an additive (linear) regression model seemingly produces a stable fit to synergistic data.

For negative intercorrelations, on the other hand, the results reveal quite a different picture. In both panels, there is a sharp "elbow" in the top curves with the R^2 values dropping down to around .20. That is, the fit of a linear regression model to synergistic data is very much influenced by the degree of negative intercorrelation. (The dip in the curves around .00 intercorrelation in the right panel will be taken up later.)

Figure 2 shows the R^2 values for the fit of a multilinear regression model (Equation 8) to the same data. The two panels, representing the results for 25 and 100 stimulus cases, respectively, are also quite similar. The top curves are relatively flat for positive intercorrelations with an asymptote of around .75. For negative intercorrelations, there is again an elbow in the R^2 values. Compared to Figure 1, the top curves are consistently higher; the fit to the adding data, however, is virtually unchanged.

Insert Figure 2 about here

A more revealing view of the difference between Figures 1 and 2 can be obtained by computing the improvement in R^2 gained from a multilinear regression

model over an additive model. These ΔR^2 values are shown in Figure 3. As before, the two panels, for 25 and 100 stimulus cases, are fairly similar. For positive intercorrelations, the ΔR^2 values stabilize at about .05, i.e., the improvement in variance-accounted-for when going from an additive to a multilinear model is almost constant. For negative intercorrelations, however, the improvement in fit increases dramatically with higher negative values. At the most extreme, the ΔR^2 values approach .25 when $r = -.90$ in the right panel.

Insert Figure 3 about here

To summarize, the R^2 results show that a linear regresssion model appears to do reasonably well in describing synergistic data when the cues are positively correlated. But, when the "correct" multilinear regression model is used, there is an improvement of about .05 in the R^2 values. However, when the cues are negatively correlated, a linear regression model is definitely inferior to a multilinear regression model. Moreover, the larger the negative intercorrelation, the greater the improvement by using the "correct" regression model.

Beyond variance-accounted-for measures, the most frequently used descriptive indices are the regression weights. Figure 4 shows the standardized regression (Beta) weights for the crossproduct terms in Equation 8. Except for a falloff with highly negative cue correlations, the Beta weights are relatively sizable. The values range from around .65 for highly negative intercorrelations in the left panel to nearly .90 for highly positive intercorrelations in the right panel.

For comparative purposes, the Beta weigh s for the cross product of a multilinear model fit to adding data are shown in the middle of the figure.

As expected, these values are uniformly near zero. Thus, the [illegible] weights provide suggestive indications that the presence of a synergism does make a consistent difference.

Insert Figure 4 about here

While descriptive indices can be quite useful in summarizing data, they are of course inherently incapable of saying whether a synergism is present or not. That is, they cannot be used to answer yes-no questions. Therefore, such indices can be quite deceptive if, say, a high R^2 value is used to support a linear regression model. There is never any way of knowing whether the observed R^2 value is good, bad, or mediocre. For a more complete discussion of this issue, see Anderson and Shanteau (1977; also see Shanteau, 1977).

The appropriate way to deal with such issues is to employ inferential test statistics. Accordingly, the next section will deal with various inferential indices designed to detect the presence of synergisms.

Inferential Indices

Besides providing descriptive information, the Beta weights for the crossproduct terms can also be tested for significance. The proportions of significant Beta weights (out of 90 for each point) are plotted in Figure 5. For 100 stimulus cases, the right panel shows that the weights are uniformly significant. In contrast, the left panel for 25 cases reveals that only highly negative intercorrelations produced proportions near one. For positive intercorrelations, the proportions drop to near .75. Thus, the results for regression weights reveal that they are generally sensitive to the presence of a synergism. However, this sensitivity is reduced when 25 stimulus cases are used and when the intercorrelations are positive.

Insert Figure 5 about here

There are some well-known problems with using regression weights to examine the importance of terms such as crossproducts. One problem is that when the cues are intercorrelated, the order in which the terms are examined can influence the magnitude of a weight (Darlington, 1969). Also, the scaling of the stimulus metric can influence the apparent importance of even standardized weights (Anderson & Shanteau, 1977). While such problems were controlled for in the present simulations, these shortcomings would generally make regression weights impractical for testing synergisms in real data.

One procedure which avoids these problems is the hierarchical test proposed by Cohen and Cohen (1975). Briefly, this procedure involves testing the ΔR^2 presented in Figure 3 with an F-ratio. As can be seen in Figure 6, the F values are considerably higher for the 100 stimulus cases in the right panel. Moreover, there is a pronounced decline in the F values as the intercorrelations go from negative to positive.

Insert Figure 6 about here

When these F values are tested for significance, the proportions for the 100 stimulus cases in Figure 7 are uniformly at 1.0. However, the proportions for 25 cases in the left panel range between .75 and 1.0. In addition, positive intercorrelations are less likely to produce significant results than negative intercorrelations.

Insert Figure 7 about here

On the whole, the hierarchical procedure is quite sensitive to the presence of a synergism, especially in the larger stimulus set. In the smaller set, there is a slightly higher chance of detecting synergisms when the cues are negatively correlated. By and large, however, the hierarchical test results were little affected by any of the present environmental manipulations.

While the hierarchical test was quite good at detecting synergisms, it is not capable of saying whether there is more involved than simple bilinearity. That is, showing that a crossproduct is present does not rule out the presence of other more complex terms. One way to check that is to examine the lack-of-fit after the additive and multiplicative terms have been extracted. As described by Draper and Smith (1966), F-ratios for lack-of-fit can be used to test the unaccounted-for-variance. The average F-ratios are shown in Figure 8 for the 25 and 100 stimulus cases, respectively. The curve shapes are for the most part similar across the two panels, with the largest F-ratio found for negative intercorrelations.

Insert Figure 8 about here

The proportions of F-ratio that were significant are shown in Figure 9. While the proportions are considerably higher for the 100 stimulus cases, the results are uniformly greater than the chance levels observed for the adding model. Of course, since deviations from multilinearity are being tested, and since the synergistic models contain no such deviations, the proportions should all be at the chance level. It would thus appear that

this procedure leads to an inflated type I error rate. Based on the present results, the lack-of-fit test is apparently overly sensitive to nonexistent deviations and therefore should not be used.

Insert Figure 9 about here

Post-Hoc Analyses

Aside from the planned analyses reported to this point, several additional analyses were performed based on some unanticipated results. These involved the surprising influence of some atypical stimulus sets and the anomalous findings observed with zero-intercorrelation sets.

Between-set differences. In the results presented so far, the findings have been averaged over the nine stimulus sets constructed for each intercorrelation condition. Since particular efforts were made to ensure the comparability of these sets, there was little reason to expect any sizable differences between them. Nevertheless, there were some notable between-set differences in the values for various indices.

One of the most striking examples is summarized in Table 3; the results were taken from the .00 intercorrelation/100-stimulus-case condition with the data generated by the adding-multiplying model (Equation 4). The row entries present the results for each of the nine constructed stimulus sets. In the first column, the observed cue intercorrelation values can be seen to be extremely close to the prespecified .00 value. Although, not shown, the means, standard deviations, and distributions were also quite close to their prespecified values (see Table 2 for an example). The fit of the linear regression model (Equation 7) led to average R^2 values in the second column which range from .61 to .66--with one notable exception.

The fifth stimulus set has an R^2 value of only .37. A similar discrepancy can be observed in the third column for the R^2 values from the fit of a multilinear model (Equation 8).

Insert Table 3 about here

The average Beta weights for the crossproduct term are given in column four with the number significant (out of 10) in parentheses. The average weight falls between .77 and .88, again with the exception of the fifth set which has a value of .66. Even larger differences can be seen for the average hierarchical F-ratios in column five and the lack-of-fit F-ratios in column six. In both cases, the results for the fifth stimulus set are far out of line from the other eight stimulus sets.

A more typical set of data is presented in Table 4 for -.90 correlation/ 100-stimulus-case condition. For these results, the eighth stimulus set, and to a lesser extent the fourth set, stand apart. As an example, the R^2 values for the multilinear regression model range from .45 to .60 with the exception of .30 for the eighth set and .34 for the fourth. Most of the other conditions, although not shown to save space, produced results similar to Table 4 in that one or two of the stimulus sets stood out from the others.

Insert Table 4 about here

In an effort to localize the source of these discrepancies, the means and standard deviations of the simulated data were computed. As can be seen from the averages reported in column seven of Tables 3 and 4, the means are

quite similar across all the stimulus sets. However, the standard deviations in column eight are another matter. The values are markedly smaller for the fifth set in Table 3 and the eighth set in Table 4. That is, the range, variability of the data generated for these sets is compressed relative to the other stimulus sets. This in turn apparently produced smaller values for various statistical indices. In short, it was more difficult to detect the presence of a synergism in the data generated from these stimulus sets. Obviously, this would make the generalizability of the results obtained from these sets highly suspect.

Zero-correlation results. In several of the figures, the .00 intercorrelation/100-stimulus-case condition appears to have produced anomalous results. For instance, in Figures 1 and 2 the top curves in the right panels show a marked dip for the .00 intercorrelation value. While the reason for this anomaly is not entirely clear, .00 intercorrelated stimulus sets have also been observed to be unusual in other studies (Stewart, 1980).

One contributing factor may be the relative homogeneity of the stimulus sets with .00 intercorrelation. In the first column of Table 3, for instance, the observed intercorrelation values are all extremely close to .00. Such close similarity was not observed for any of the other intercorrelation conditions. In Table 4, for example, the observed intercorrelation values range from -.85 to -.93. It would thus appear in the .00 intercorrelation condition, the stimulus sets were much closer to criterion value than was the case elsewhere. It would therefore appear that the greater homogeneity of the .00 intercorrelation sets apparently accentuated any differences from the other conditions.

Since the stimulus sets in the .00 intercorrelation are so similar, this emphasizes all the more the uniqueness of the fifth set in Table 3. That is,

of all the intercorrelation conditions, this would seem to be the least likely to produce an anamolous stimulus set. In fact, the fifth set, with a value of .01, was the only one not to have an intercorrelation value of .00. It appears that even this very slight deviation in the fifth set may have contributed to the anomalous results. This finding suggests that atypical stimulus sets can occur even in the most unexpected and tightly controlled situations.

Discussion

There are three noteworthy findings in the present study. First, multiplicative synergisms do make a major difference in nonorthogonal designs. When trying to account for synergistic data, it matters a great deal whether a "correct" or an "incorrect" regression model is used. While the effect is more pronounced for negative intercorrelations, the impact of a synergism can be seen throughout the range of cue intercorrelations. Thus, investigators who continue to ignore the possibility of synergisms may be overlooking some potentially very important relationships.

Second, much of what might be considered common practice in the analysis of nonorthogonal data is called into question by the present results. For instance, some of the indices regularly used to analyze judgmental data, e.g., R^2, were found to be insensitive to the presence of synergisms. On the other hand, there were some less common measures which were sensitive to synergisms, e.g., ΔR^2, and which could be routinely incorporated into judgment analyses. In addition, a rather surprising result was that some stimulus cue sets were better than others at revealing the presence of synergisms. Since all sets had to meet some rather stringent qualification requirements, this suggests that there may be some unappreciated difficulties in generalizing results obtained from nonorthogonal designs.

Finally, the success of the present simulation approach [illegible] on several grounds. On the one hand, the approach proved quite useful in investigating some issues which would have been difficult, if not impossible, to study empirically. On the other hand, several new research issues were raised which can be addressed in empirical investigations. For instance, the role of error in judgmental data might become of special concern in future analyses. Although not without limitations, the simultaneous simulation of environmental conditions, subject behavior, and real-world structures can provide a fruitful basis for investigating many other issues. The implications of each of these three findings will be taken up in the remainder of the Discussion.

The Impact of Synergisms

The present results are quite clear in showing that the presence of a synergism does make a difference. Regardless of the environmental conditions investigated, there was always an improvement in the fit of a multilinear regression model over a linear model for synergistic data. While the size of the improvement did vary depending on which indices and conditions were used, there was not a single occasional in any of the present simulations where an improvement failed to appear.

Cue intercorrelations. The pervasiveness of the influence of synergisms was somewhat unexpected. Perhaps most surprising was the persistent effect found for the high positive intercorrelation conditions. When cues are closely bound together, e.g., by a correlation of +.90, the fit might be expected to be insensitive to the form of the regression model. That is, for highly correlated cues, high values will follow from high cues and low values will follow from low cues regardless of whether a linear or a multilinear model is used. Moreover, the degree of insensitivity might be expected to increase as the correlation approaches unity.

Instead, the present results revealed an almost constant difference between the fit of linear and multilinear regression models for positive intercorrelations. It would appear that the degree of intercorrelation, when it is positive, is of little relevance. Thus, up to the limits of the present simulation analyses, the influence of a synergism appears to be quite consistent for positive cue correlations.

Some rather curious results appeared for the zero-intercorrelation condition. Basically, the results are quite similar to those observed for positive correlations. However, several of the figures revealed "dips" and other discontinuities for zero intercorrelations. While the source of this irregularity is not entirely understood, the important result is nevertheless unchanged: synergisms have just as much effect of this intercorrelation condition as in any positive condition.

A rather different picture emerges for negative cue intercorrelations. The greater the size of the negative intercorrelation, the greater the disparity between the fit of linear and multilinear models to synergistic data. That is, the size of the negative correlation influenced how much was lost by ignoring the presence of a synergism.

While some difference between positive and negative intercorrelations had been anticipated, the magnitude of the difference was not. With high negative correlations, the discrepancy between the fit of linear and multilinear models approaches, and in some cases even exceeds, the fit of the linear model. That is, the variance-accounted-for can actually be doubled by shifting from a linear to a multilinear regression model. Because of the disparity between these results and the general acceptance of linear models, the role of such models will next be considered in some depth.

Linear models in decision making. A great deal of knowledge has been accumulated about the use of linear models in summarizing human judgments.

Much of what has been found suggests that neither correct model form nor correct weights are important for getting an adequate description of judgment data (Slovic, Fischhoff, & Lichtenstein, 1977). The present results for synergistic data would imply the need for some modifications in this view. (The issue of weights will be taken up separately below.)

There are a number of papers in the literature which have argued that linear models can do a good job of describing nonadditive data (e.g., Yntema & Torgerson, 1961). In perhaps the best known of these papers, Dawes and Corrigan (1974, p. 98) concluded that "linear models are good approximations to all multivariate models that are conditionally monotonic in each predictor variable."[4] The authors go on to add that the linear approximations improve with increasing error. Hence, there has been a widespread feeling that distinguishing model form is unimportant for most multiple regression analyses of judgmental data. Indeed, many of the approaches which use regression procedures, such as the Len's model approach (Hammond, Stewart, Brehmer, & Steinmann, 1975), routinely ignore anything but a linear regression model.

In contrast, Anderson and Shanteau (1977) cautioned against the routine use of linear models. Among other shortcomings, they offered an example of multiplying data which satisfies conditional monotonicity but which is clearly nonlinear. While a best-fitting linear model correlated .885 with the data, the fit was far from adequate. On a 100-point scale, the linear model was providing estimates which ranged from -25 to +25 for a data value of 0. The basic problem was that the data displayed a diverging pattern of lines and a linear model can only produce parallel lines. In short, a linear model was not adequate to describe multiplying data.

Anderson and Shanteau (1977) went on to point out that linear models can be useful in applications involving data prediction. When the goal is to understand psychological processes, however, linear models can be deceptive. For instance, synergistic processes can be easily misinterpreted as being additive if linear models are used exclusively in data analysis (for an actual example, see Shanteau, 1977).

The present findings extend the arguments against the routine use of linear models in two ways. First, even the predictive use of linear models can be questioned when, at best, the loss in variance-accounted-for is 5%. At worst, a linear model can account for only half of the systematic variance. Moreover, in all cases, the loss in variance-accounted-for was significant when a linear model was used. Of course, whether losses of this magnitude are within the realm of a "good approximation" might still be subject to some debate. Nevertheless, the present findings suggest that the predictive ability of linear models should not be uncritically accepted.

Second, previous evidence against linear models, such as that offered by Anderson and Shanteau, has been primarily based on analyses of variance applied to orthogonal designs. The results offered here demonstrate that synergisms also make an important difference in regression-based analyses of nonorthogonal designs. Thus, neither the type of statistical analysis nor the experimental design are relevant to the argument that linear models can be misleading when applied to synergistic data.

Weights for linear models. Several recent papers have pointed out that previous arguments about the insensitivity for weights for linear models may be inappropriate. Previously, investigators such as Wainer (1976), as well as Dawes and Corrigan (1974), had argued that equal weights (or even random weights in some circumstances) can do about as well as optimal weights

in linear models. However, Newman (1977) demonstrated that, while equal weights can provide good approximations when cues are positively correlated, equal weights are generally inferior when cues are negatively correlated. Thus, previous arguments about equal weights in linear models apparently cannot be generalized to conditions of negatively-correlated cues (also see John & Edwards, 1978).

The results here expand and complement the findings of Newman (1977). While Newman was concerned with showing that weights in a linear model matter for negative intercorrelations, the present results show that model form also matters. Thus, when cues are negatively correlated, neither the weights nor the form of the model should be taken for granted.

Positive-vs-negative intercorrelations. Since negatively intercorrelations produce such different results, it is appropriate to ask about the conditions under which such intercorrelations might be observed. Perhaps, negatively correlated cues are relatively rare in reality and so would not be much cause for concern. After all, perceptual judgments for instance are made in the context of numerous largely redundant, i.e., positively correlated, cues (Brunswik, 1956; also see Hammond, 1981).

However, negative correlations may in fact be the rule not the exception in decision making. Many decision problems are only problems because the cues are inversely related. For example, selecting a new car would be trivial if the attributes were positively correlated, i.e., if the cheapest car was also the best looking. In reality, however, such a car does not exist and instead we are forced to make tradeoffs between conflicting attributes. Thus, negative intercorrelations will be found precisely in those situations where decisions are most likely in reality.

A more elegant discussion of the role of negative intercorrelations can be found in McClelland (in press). He shows that if the set of alternatives

is restricted to those that are nondominated, i.e., on the Pareto frontier, then the cues will necessarily be negatively correlated. That is, only by including dominated alternatives can a set have anything but negatively correlated cues. Therefore, excluding inferior alternatives produces negative intercorrelations.

Moreover, decision situations involving negative intercorrelations are also quite likely to lead to synergistic processing rules. For instance, in risky decision making, payoffs and probabilities are generally negatively correlated, i.e., high payoffs have low probabilities and low payoffs have high probabilities. Further, both the optimal decision rule (Edwards, 1954) and the strategy generally used by subjects to make risky decisions (Slovic et al, 1975) involve the multiplication of probability and payoff. Similarly, other synergistic rules are likely to be found in precisely those conditions where negative intercorrelations are observed (Hammond, 1981).

Another way of looking at positive versus negative correlations is in terms of the assumed resources available. In a world involving positively-correlated attributes, there is no implied limit on resources, i.e., it is theoretically possible to get the most on all attributes at the same time. Thus, positive intercorrelations suggest an unlimited-resources view of the world. In contrast, negatively-correlated attributes imply a limit on available resources, i.e., it's not possible to simultaneously get the most on every attribute. This latter view may be much more reasonable in a world that, in fact, requires choices involving limited resources.

Statistical Issues and Synergisms

For purposes of detecting synergisms, two findings at a statistical level stand out. The first concerns the comparison of various statistical indices. The second relates to experimental design and the influence of atypical stimulus

sets. The implications of each of these findings will be discussed in this section.

Comparison of statistical indices. As can be seen in the figures, the various indices were all sensitive to some extent to the presence of a synergism. Regardless of the index, the values were uniformly higher when the "correct" regression model was applied to synergistic data. This was true across all intercorrelation values and across the two stimulus set sizes.

In practice, however, several of the indices in common use may provide uncertain information about the presence of a synergism. For instance, a R^2 value of .70 for the fit of a linear model may look good--until it is discovered that a multilinear model leads to a R^2 value of .75 for the same data (these values are taken from Figures 1 and 2). The difficulty with indices such as R^2 (and other correlation-based measures) is that it is impossible to know by looking at a single value whether the fit is good or not. Only by comparing the fits for various models can any evaluative statements be made.

Unfortunately, comparative analyses using alternative models are seldom performed. Worse yet, there is no limit to the number or variety of alternative models that might be considered. In short, measures such as R^2 do not provide an adequate basis for detecting synergisms (also see Shanteau, 1977).

Other common descriptive measures, such as the size of the regression weight for the crossproduct term, are also inadequate. The problem is that when cues are intercorrelated, the order in which the analysis is conducted can influence the size of the weights (Darlington, 1968). This means that the size of the weights depends on a variable that is under direct control of the investigator. Thus, regression weights provide an uncertain indicator of the presence of a synergism.

The only descriptive measure that can be recommended on the basis of the simulations here is ΔR^2. This measure, obtained from the difference in fit of linear and multilinear models, was generally sensitive to the presence of a synergism. In addition, the hierarchical test discussed below provides a test of significance that is related to this measure. Therefore, ΔR^2 would be the preferred descriptive index.

The present results, therefore, show that some widely used measures in regression analyses are not suitable for detecting synergisms. This may well explain why synergisms have been so infrequently reported in previous analyses of nonorthogonal designs. In contrast, numerous instances of synergisms have been found in studies involving orthogonal designs (e.g., Shanteau, 1981).

Of the inferential indices considered here, the measure of choice appears to be the hierarchical test proposed by Cohen and Cohen (1975). The test correctly detected the presence of synergisms in all the 100 stimulus-case conditions. A slight decrement in detectability rates was observed for the 25 stimulus-case conditions. But even at its worst, the hierarchical procedure detected 75% of the synergisms (see Figure 7). Equally important, the false-alarm rates for the hierarchical tests were consistently below 10%. In short, the hierarchical test was quite sensitive to the presence of synergisms.

The hierarchical procedure has been advocated (Arnold & Evans, 1979) precisely because of its potential for evaluating multiplicative components. A problem in previous regression analyses has been how to analyze multilinear models when the independent variables are not measured on a ratio scale. In those cases, an additive component is introduced into the regression model and direct tests of the crossproduct may not be revealing. However, as

shown here, the presence of an additive term has little influence on the hierarchical test results (see Figures 6 and 7). Thus, the present recommendation would be to include routinely ΔR^2 and the hierarchical procedure in multiple-regression analyses of judgmental data (also see Stahl & Harrell, 1981).

In contrast, the lack-of-fit test is notable because of its poor performance. The major problem appears to be an overly large sensitivity to nonexistent deviations from the multilinear model. Of course, it's possible that the inflated detection rate may be due to some facet of the present simulation analyses. That is, would real data produce a better-behaved test statistic?

To examine this question further, it is worth considering the study by Shanteau and Nagy (1979). They applied a comparable lack-of-fit procedure to test the adequacy of a multilinear model for dating decisions. The model described the data quite well and was highly accurate in predicting actual dating choices. Despite the apparent good fit it, however, the lack-of-fit tests revealed significant discrepancies for over two-thirds of the subjects. Additional analyses revealed no discernable locus to the discrepancies, and the deviations seemed to be quite small. It thus appears that the test is overly sensitive to small deviations in real data. Taken together with the simulation results, it would appear that until more is known about the properties of the test, it cannot be recommended for regular use.

Idiosyncrasies in nonorthogonal designs. Perhaps the most unexpected finding to come out of the simulations was the occasional occurence of an atypical stimulus set. What made this so surprising was that all stimulus sets had to pass a number of stringent qualification tests before being accepted. The goal was to produce stimulus sets that were as homogeneous as possible.

Yet, one or two atypical sets were found in nearly every condition. Among other differences, such sets led to nonconforming results concerned the sensitivity of various indices to the presence of a synergism. What this means is that the ability to detect a synergism depends on the particular stimulus set selected.

Of course, it is possible that even more stringent qualification tests would have led to more homogeneous sets. Indeed, based on hindsight, many of the present atypical sets could have been eliminated by checking the variability of the expected (pre-error) response values. However, demanding stricter selection criteria would not be feasible on two grounds. First, at a practical level, nonorthogonal designs are frequently used in research settings which have little or no flexibility. For instance, a marketing researcher has little if any control over the product alternative set. In such settings, the researcher may have no choice but to use the available stimulus cases.

Second, at a theoretical level, unless the stimulus sets are identical, they can never be homogeneous for all purposes. While it might be possible to construct qualification tests to insure that the stimulus sets are equivalent in regard to detecting synergisms, the sets might still be dissimilar for other purposes. That is, there is no way to select stimulus sets that are homogeneous for all applications. Moreover, since many analyses are impossible to anticipate, there is no way to preselect stimulus sets. In short, it is not feasible to develop a general-purpose qualification procedure.

The inability to develop general preselection criteria means that a researcher can never be certain whether a particular stimulus set is atypical or not. As the present study demonstrates, even using as many as 100 stimulus

cases is no protection (see Table 3). This means that the results from even fairly large designs may not generalize. Unfortunately, the possibility of atypical stimulus sets raises questions about the generalizability of many previous results obtained using nonorthogonal designs.

There are at least three options for alleviating the problem of lack of generality. The first option is to replicate all findings using different nonorthogonal designs. This would greatly minimize, but not eliminate, the possibility that the observed results will fail to replicate because of an atypical design. While such experimental replications are of course always desirable, they may be impractical in many settings.

Second, there are many investigations in which orthogonal (factorial) designs might be used instead of nonorthogonal designs. Factorial designs avoid almost all of the problems discussed above for nonorthogonal stimulus sets. Specifically, factorial designs lead to optimal parameter estimates and model tests. While factorials do have shortcomings for judgment research, the disadvantages have frequently been overemphasized relative to the advantages.[5]

The final option is to pretest the nonorthogonal design using the types of simulation analyses performed here. That is, the anticipated behavioral models, along with the method(s) of analysis, can be simulated in advance. In this way, the suitability of the stimulus design for answering the research question(s) can be established a priori.

The Role of Simulation Analyses

There are both important advantages and important disadvantages to the use of simulation analyses to study issues such as synergisms. On the positive side, simulations allow the study of psychological problems that would be intractable using an empirical approach. On the negative side, any simulation

is only as good as its assumptions and some of the assumptions made here can certainly be questioned. Before considering these pluses and minuses in detail, there are some important distinctions that need emphasis.

Simulations as a psychological research tool are hardly unique. There are numerous applications in the literature of useful simulation analyses. For instance, computer models of subject behavior have been frequently employed to analyze cognitive processes in problem solving behavior (e.g., Newell & Simon, 1963). Similarly, Monte Carlo simulations have long been used to analyze the properties of various statistical procedures (e.g., Lindquist, 1953, pp. 78-90).

What separates the present approach from these earlier simulation analyses is the effort here to simulate all stages of an experiment. As outlined in Table 1, the three-part approach involves separate simulations of environment, behavior, and analyses. While prior approaches have concentrated on simulations of behavior or analyses, the present approach is to view the research process as a whole. Therefore, these three elements are all included in what might be termed an experiment-simulation. The advantages and disadvantages of this approach will now be considered.

Advantages. There are at least four advantages to experiment simulations. The first is that it is possible to address research questions that could not be practically investigated in any other way. The present study, for instance, involves 1,620 separate conditions. To run even one subject in each of these conditions would clearly be prohibitive. Thus, studies that would be impossible to conduct empirically can still be approximated through experiment simulations.

A second advantage is that, with the present approach, the "truth" is always known. That is, the true state of the environment, the true behavioral

model, and the true analytic answer are always known. Knowledge of these "truths" allows a number of analyses to be performed that could not be conducted otherwise. For instance, various designs, models, and analytic techniques can be directly compared because the correct answers are known. Therefore, the researcher is in the highly enviable position of knowing the truth at every stage.

The third advantage is that experiment simulations can be used to generate more precise empirical investigations. That is, simulations can point out research issues and areas where empirical research is lacking. For instance, some marked differences between positive and negative intercorrelations conditions were demonstrated in the present simulations. However, there is as yet relatively little empirical evidence to demonstrate how subjects respond under the range of conditions studied here. Its not even clear whether subejcts would use synergistic rules under all intercorrelation conditions. Empirical research is clearly necessary to answer such questions generated by the experiment simulations.

The final advantage is that experiment simulations can be employed to investigate entirely new research issues. Problems which not have yet been considered may be highlighted by performing a simulation. In the present study, for example, the location of the error term in the subject's model became a major issue. Yet, because no prior research was available, this issue could not be addressed empirically. However, synergisms provide a unique opportunity to separate stimulus error from resonse error and to investigate the relative magnitude of each. Thus, experiment simulations have helped to focus attention on a new and potentially quite interesting research issue.

Disadvantages. There are two noteworthy problems in the experimental simulation approach used here. The first is that there is no criterion for whether the simulations are successful or not. While the results produced look reasonable, that is no assurance that there are not important difficulties. As a check, what is needed is to compare some of the present findings against specific empirical results. Until such checks have been performed, there is always the possibility that the present simulations may be unrelated to empirical reality.

A second disadvantage is that this or any simulation is only as good as its assumptions. If the assumptions are faulty, then so necessarily will be the results from the simulation. In the present three-stage approach, there would seem to be relatively little cause for concern in regard to the simulations of the environment and the experimenter. The assumptions made for these stages were largely noncontroversial and in line with standard research procedures.

The status of the subject simulation is not as clear, however, since several rather arbitrary assumptions had to be made. Most notably, the way in which error was incorporated may be a special source of concern. As noted above, there is little empirical evidence in the literature concerning how error enters into a subject's behavior. Without such evidence, there was no choice except to make some "seat of the pants" assumptions. Specifically, it was assumed that error enters in after the stimuli have been combined and that the coefficient of variation provides a useful rule-of-thumb as to the size of the error component.[6]

While these assumptions led to reasonable-looking data, they are certainly subject to further analysis. Additional simulation and empirical research can be used to investigate error in more detail. Simulations can be

employed, for instance, to explore alternative assumptions about error. And empirical research, as noted above, can be directed at such issues as the size and location of error. Thus, the simulation approach taken here, while not without its problems, has raised some interesting questions and suggested some new directions for future research.

Reference Notes

1. Boyle, P. A note on correlated score generation procedures for multiple-cue judgment tasks. Unpublished manuscript, Institute of Behavioral Science, University of Colorado, 1970.

2. Hammond, K. R. Personal Communication, June 27, 1980.

3. McClelland, G. H. Equal versus differential weighting for multiattribute decisions: There are no free lunches. Unpublished manuscript, Institute of Behavioral Science, University of Colorado, 1978.

4. Shanteau, J. Comparison of methods for analyzing bilinear (multiplicative) data. Paper presented at the meeting of the Psychonomic Society, San Antonio, November 1978.

5. Shanteau, J. How to analyze synergistic (multiplicative) relations in psychology. Manuscript submitted for publication, 1981.

6. Stewart, T. R. Personal communication, July 18, 1980.

References

Arnold, H. J., & Evans, M. G. Testing multiplicative models does not require ratio scales. Organizational Behavior and Human Performance, 1979, 24, 41-59.

Anderson, N. H. How functional measurement can yield validated interval scales of mental quantities. Journal of Applied [illegible], 1979, [illegible], 677-692.

Anderson, N. H., & Shanteau, J. Weak inference with [illegible]. Psychological Bulletin, 1977, 84, 1155-1170.

Brunswik, E. Perception and the representative design of psychological experiments. (2nd ed.). Berkeley: University of California Press, 1956.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Wiley, 1975.

Darlington, R. Multiple regression in psychological research and practice. Psychological Bulletin, 1969, 69, 161-182.

Dawes, R. M., & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Draper, N. R., & Smith, H. Applied regression analysis. New York: Wiley, 1966.

Edwards, W. The theory of decision making. Psychological Bulletin, 1954, 51, 380-417.

Hammond, K. R. The integration of research in judgment and decision making. (Rep. No. 226). Boulder: University of Colorado, Center for Research on Judgment and Policy, July 1980.

Hammond, K. R., Stewart, T. R., Behmer, B., & Steinmann, D. O. Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press, 1975.

John, E. S., & Edwards, W. Importance weight assessment for additive, riskless preference functions: A review. (Research Rep. 78-5). Los Angeles: University of Southern California, Social Science Research Institute, December 1978.

Kaiser, H., & Dickman, K. Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. Psychometrika, 1962, 27, 179-182.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Newell, A., & Simon, H. A. Computers in psychology. In R. D. Luce, R. R. Bush, & E. Galanter (eds.), Handbook of mathematical psychology, Vol. 1. New York: Wiley, 1963.

Naylor, T., Balintfy, J., Burdick, D., & Chu, K. Computer simulation techniques. New York: Wiley, 1965.

Newman, R. J. Differential weighting in multiattribute utility measurement: When it should not and when it does make a difference. Organizational Behavior and Human Performance, 1977, 20, 312-325.

Phelps, R. H., & Shanteau, J. Livestock judges: How much information can an expert use? Organizational Behavior and Human Performance, 1978, 21, 209-219.

Shanteau, J. C. Component processes in risky decision judgments. Unpublished doctoral dissertation, University of California, San Diego, 1970.

Shanteau, J. An information integration analysis of risky decision making. In M. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press, 1975.

Shanteau, J. POLYLIN: A FORTRAN IV program for the analysis of multiplicative (multilinear) trend components of interactions. Behavior Research Methods & Instrumentation, 1977, 9, 381-382.

Shanteau, J., & Nagy, G. F. Probability of acceptance in dating choice. Journal of Personality and Social Psychology, 1979, 37, 522-533.

Slovic, P. Analyzing the expert judge: A description of a stockbroker's decision processes. Journal of Applied Psychology, 1969, 53, 255-263.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Stahl, M. J., & Harrell, A. M. Modeling effort decisions with behavioral decision theory: Toward an individual differences model of expectancy theory. Organizational Behavior and Human Performance, 1981, 27, 303-325.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Webster's new collegiate dictionary. Springfield, Mass.: Merriam, 1976.

Yntema, D. B., & Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, HFE-2, 20-26.

Footnotes

Work on this project was supported in part by Office of Naval Research contract N00014-77-C-0336 to the Center for Research on Judgment and Policy at the University of Colorado. The author wishes to acknowledge the invaluable contributions of Michael O'Reilly in the development of the computer programs. The author also wishes to thank Kenneth Hammond, Gary McClelland, and Thomas Stewart for their advice during the conceptual development of this project. Finally, the helpful comments of Gary Gaeth on portions of the manuscript are gratefully acknowledged. This work was carried out during the author's appointment as a Visiting Research Associate at the Institute of Behavioral Science.

[1]The data for each simulated subject were created by randomly perturbing the Y' values to generate two response replications, Y_1' and Y_2'. These two replications represent responses obtained from two independent presentations of the stimulus set to a subject. The reason for having two replications is that one of analyses required a separate estimate of error; such estimates can only be obtained by having response replications. For analyses which do not require independent error estimates, the Y_1' and Y_2' values were averaged to produce $\bar{Y}'$ values.

[2]Ken Hammond (personal communication, 1980) has indicated that similar differences have been observed in analyses run for other purposes. Moreover, the downward trend for the adding-model results across negative intercorrelations is also commonly observed. Thus, the present results are apparently in line with results found in other settings.

[3]To illustrate the effect that a reduced range can have on detecting a synergism, consider the following two stimulus sets: In set one, the paired

cue values are (1,5), (3,5), (5,5), (7,5), and (9,5). In set two, the cue values are (3,5), (4,5), (5,5), (6,5), and (7,5). Looking at just the first cue, the means are obviously equivalent in the two sets, but the standard deviation for set two is less. (Note that setting second cue to 5 is for simplicity and convenience; however, it does not effect the generality of the argument in any way.)

If a multiplying model (Equation 2) is applied to set one, the (errorless) data will be 5, 15, 25, 35, and 45. If an adding model (Equation 3) is applied to the same set, the data will be 6, 8, 10, 12, and 14. The sum-of-squared deviations between the models is $1^2 + 7^2 + 15^2 + 23^2 + 31^2 = 1765$, with the major difference between adding and multiplying occuring at the upper end. For set two, the data for the multiplying model is 15, 20, 25, 30, and 35. The data for the adding model is 8, 9, 10, 11, and 12. The sum-of-squared deviations is then $7^2 + 11^2 + 15^2 + 19^2 + 23^2 = 1285$, or 480 less than in set one.

Two points are worth emphasizing. First, the range of the data is considerably larger in set one than set two. This, of course, follows directly the construction of the stimulus sets. Second, the difference between adding and multiply is less in set two than in set one. Assuming comparable error values, that implies that detecting the synergism in set two will be more difficult.

[4]There is an interesting asymmetrical relation between intercorrelational values and conditional monotonicity. If the cues are positively correlated, then if one cue is conditionally monotone so must be the other. If the cues are negatively correlated, then if one cue is conditionally monotone in one direction the other will be conditionally monotone in the opposite direction. Thus, knowledge of the intercorrelation value allows inferences to be drawn about conditional monotonicity.

However, knowing that each cue is conditionally monotone does not imply any restrictions on the intercorrelation values. In fact, two cues can both be conditionally monotone in the same direction and still be highly negatively correlated. Thus, statements about conditional monotonicity do not allow inferences to be drawn about cue correlations.

[5]Factorial designs have frequently been criticized on the grounds that (1) they are unrepresentative (Brunswik, 1956), and (2) they frequently require too many stimulus cases. However, both of these criticisms can be met by the use of fractional factorial designs. Such partial designs allow for both reduction in the number of stimuli and control of unrepresentative stimuli. Some illustrative applications of fractional designs can be seen in Phelps and Shanteau (1977) and Slovic (1969).

It is noteworthy that most judgment researchers are aware of the difficulties of nonorthogonal designs. For instance, the problem of estimating weights with intercorrelated cues is well known. However, instead of turning to orthogonal designs, many researchers have used nonorthogonal designs with zero intercorrelations. As the present results make clear, the .00 correlation condition shares many of the same shortcomings as the other conditions. Also, it is not clear how using uncorrelated stimulus cue sets can be any more representative than using factorial designs.

[6]As an aside, some comment should be made about the possibility of normalizing all the data to cover the same range. Normalization has the advantage of allowing the same size error term to be used with the three data-generating models. Then, a single treatment of error could be applied to all three data sets.

While this approach seems attractive in theory, there are two practical problems. First, the pre-error data did not cover any consistent range of

values. That is, due to randomness in the construction of the stimulus sets and the restrictions necessary to satisfy the constraints of intercorrelation value, etc., the various data sets differed widely in their range. Thus, even if the same data-generating model is used, that is no assurance that the resulting data will have similar ranges. Therefore, normalization using the range (or any other sample statistic) would not lead to equivalent data sets.

Second, even if the data could somehow be normalized on range, another problem remains. The distribution of the data varies systematically between the three models. For the adding model, the data is symmetrically distributed around the midpoint of the range. However, for the multiplying and adding-multiplying models, the data is skewed towards the lower end of the range. This asymmetry is a direct consequence of the multiplying operation in that a high response can only result form the multiplication of two high cues; otherwise, relatively low responses will result. Thus, the use of a constant error term would have dissimilar effects on range-normalized data for the three models. Proportionally, the error contribution would be less for the adding model than for the other two models.

Because of such difficulties, it was decided not to normalize the data values. Instead, the error component was individually calibrated to match each data set. To reflect both range variation and distributional differences, error was made proportional to the coefficient of variation. Since this coefficient depends on both the mean and the standard deviation, it avoids most of the problems outlined above.

Table 1

Chart of the Simulation Strategy Employed

Conceptual Stages	Procedural Steps	Choice/Options
Environmental Simulation	Stimulus Construction	Intercorrelation Values Number of Stimulus Cases Number of Replications
	Test of Properties	Intercorrelation Between Cues Means for Each Cue Standard Deviations for Each Cue Distribution of Each Cue
Subject Simulation	Data Generation	Multiplying Model Adding Model Adding-Multiplying Model
	Introduction of Error	Location of Error Term Size of Error Term
Experimenter Simulation	Multiple Regression Analysis	Multiplicative Regression Model Linear Regression Model Multilinear Regression Model
	Descriptive Indices	R^2 for Fit of Linear Model R^2 for Fit of Multilinear Model Improvement in R^2 Regression Weight for Crossproduct
	Inferential Indices	Test of Regression Weight Hierarchical Test Lack-of-Fit Test

Table 2

Illustrative Output From Stimulus Construction Program[a]

Parameter Specified	Requested Value	Observed Value
Number of Cues	2	2
Number of Stimulus Cases	100	100
Minimum Value, Cue 1	1	6.4
Maximum Value, Cue 1	100	100.0
Minimum Value, Cue 2	1	2.0
Maximum Value, Cue 2	100	97.4
Mean, Cue 1	50	49.9
Standard Deviation, Cue 1	20	19.6
Mean, Cue 2	50	50.1
Standard Deviation, Cue 2	20	19.7
Chi Square Distribution Test, Cue 1	-	12.4
Significance Level (df = 19), Cue 1	≥.05	.87
Chi Square, Distribution Test, Cue 2	-	18.4
Significance Level (df = 19), Cue 2	>.05	.50
Intercorrelation Between Cues 1 and 2	0	.00

Stimulus Case Number[b]	Cue 1	Cue 2
10	67.4	44.3
20	45.5	77.2
30	57.6	24.0
40	23.4	58.7
50	76.2	44.7
60	31.2	12.4
70	51.1	87.0
80	36.5	22.9
90	54.8	19.3
100	95.8	2.0

[a]Output adapted from CUEGEN program (see Kaiser & Dickman, 1962; Naylor, et al, 1965).

[b]Every tenth case selected for illustrative purposes.

Table 3

Average Results for the .00 Intercorrelation, 100-Case Condition,

Data Generated by Adding-Multiplying Model: $X_1 + X_2 + (X_1 \times X_2)$[a]

Correlation X_1 and X_2	R^2 Fit for Linear[b]	R^2 Fit for Multilinear[c]	β Weight[d]	F Hierarchical[e]	F Lack-of-Fit[f]	Data Mean	Data St. Dev.
.00	.63	.72	.88 (10)	63.40 (10)	1.85 (10)	2609	2455
.00	.64	.75	.86 (10)	89.06 (10)	1.91 (9)	2596	2585
.00	.64	.71	.84 (10)	49.52 (10)	1.60 (9)	2581	2443
.00	.61	.70	.88 (10)	58.96 (10)	1.80 (9)	2594	2357
.01	.37	.42	.66 (10)	15.34 (10)	1.15 (1)	2676	1751
.00	.64	.71	.81 (10)	50.53 (10)	1.61 (8)	2650	2497
-.00	.63	.71	.77 (10)	52.87 (10)	1.75 (9)	2626	2444
.00	.64	.73	.84 (10)	70.17 (10)	1.73 (8)	2595	2500
.00	.66	.75	.81 (10)	72.23 (10)	1.83 (7)	2592	2602

[a] Results averaged over 10 simulated subjects

[b] Linear regression model: $\beta_1 X_1 + \beta_2 X_2$

[c] Multilinear regression model: $\beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2)$

[d] Multilinear cross-product weight

[e] Test on increment in R^2

[f] Test on residual from multilinear

Table 4

Average Results for the -.90 Intercorrelation, 100-Case Condition

Data Generated by Adding-Multiplying Model: $X_1 + X_2 + (X_1 \times X_2)$[a]

Correlation X_1 and X_2	R^2 Fit for Linear[b]	R^2 Fit for Multilinear[c]	β Weight[d]	F Hierarchical[e]	F Lack-of-Fit[f]	Data Mean	Data St. Dev.
-.89	.21	.45	.64 (10)	87.26 (10)	1.87 (10)	1904	1308
-.89	.22	.48	.66 (10)	102.30 (10)	2.27 (10)	1927	1313
-.90	.26	.53	.72 (10)	111.80 (10)	2.13 (10)	1844	1375
-.85	.21	.34	.59 (10)	38.31 (10)	1.40 (5)	2122	1338
-.93	.33	.60	.76 (10)	128.95 (10)	2.59 (10)	1817	1405
-.89	.32	.52	.70 (10)	81.73 (10)	1.82 (10)	1934	1423
-.89	.25	.47	.67 (10)	82.01 (10)	1.87 (10)	1907	1345
-.87	.13	.30	.50 (10)	49.13 (10)	1.67 (8)	2061	1208
-.90	.23	.49	.69 (10)	104.45 (10)	2.07 (10)	1857	1336

[a] Results averaged over 10 simulated subjects

[b] Linear regression model: $\beta_1 X_1 + \beta_2 X_2$

[c] Multilinear regression model: $\beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2)$

[d] Multilinear cross-product weight

[e] Test of improvement in R^2

[f] Test on residual from multilinear

Figure Captions

Figure 1. Average R^2 values for the fit of a linear regression model (Equation 7) to three data-generating models: (0) a multiplying model (Equation 2), (1) an adding model (Equation 3), and (2) an adding-multiplying model (Equation 4). (Left panel = 25 stimulus cases, right panel = 100 stimulus cases.)

Figure 2. Average R^2 values for the fit of a multilinear regression model (Equation 8) to three data-generation models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

Figure 3. Average improvement in R^2 values for a multilinear regression model over a linear regression model for three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel= 25 cases, right panel = 100 cases.)

Figure 4. Average standardized regression weights (β) for crossproduct term in the fit of a multilinear model to three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

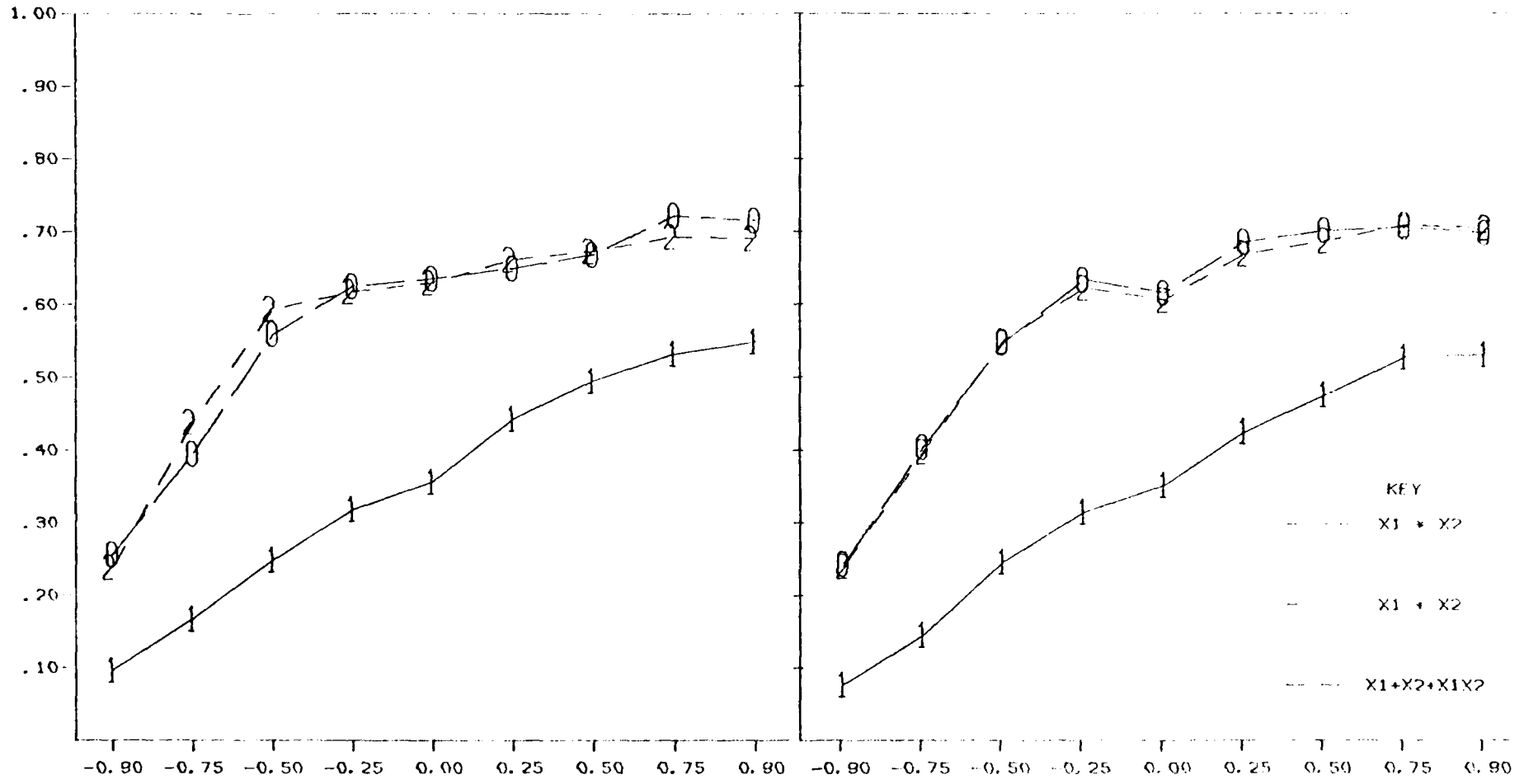
Figure 5. Proportion of significant regression weights for crossproduct term in the fit of a multilinear model to three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

Figure 6. Average F-ratios for hierarchical test applied to three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

Figure 7. Proportion of significant hierarchical tests for three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

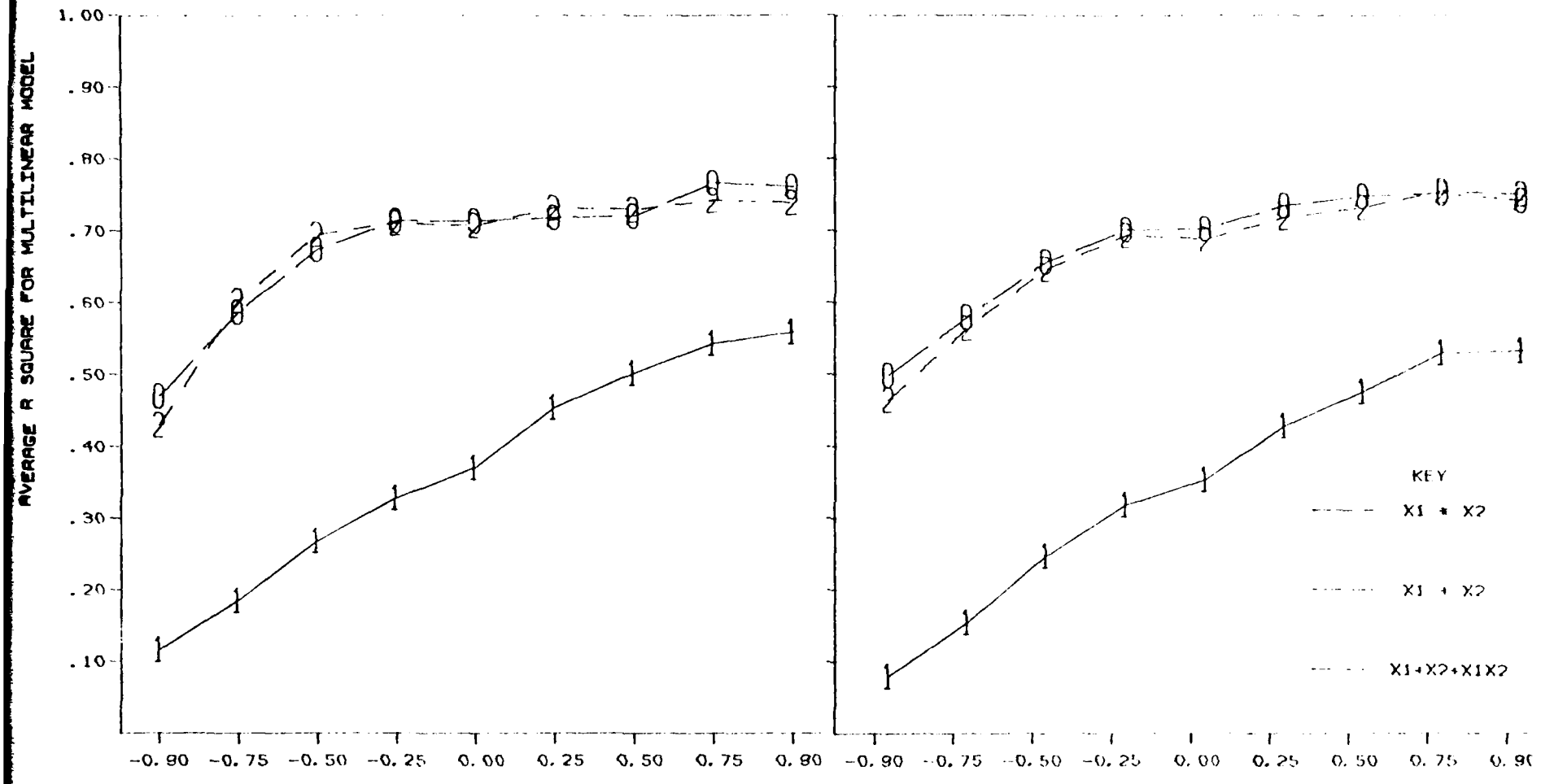
Figure 8. Average F-ratios for lack-of-fit test applied to three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

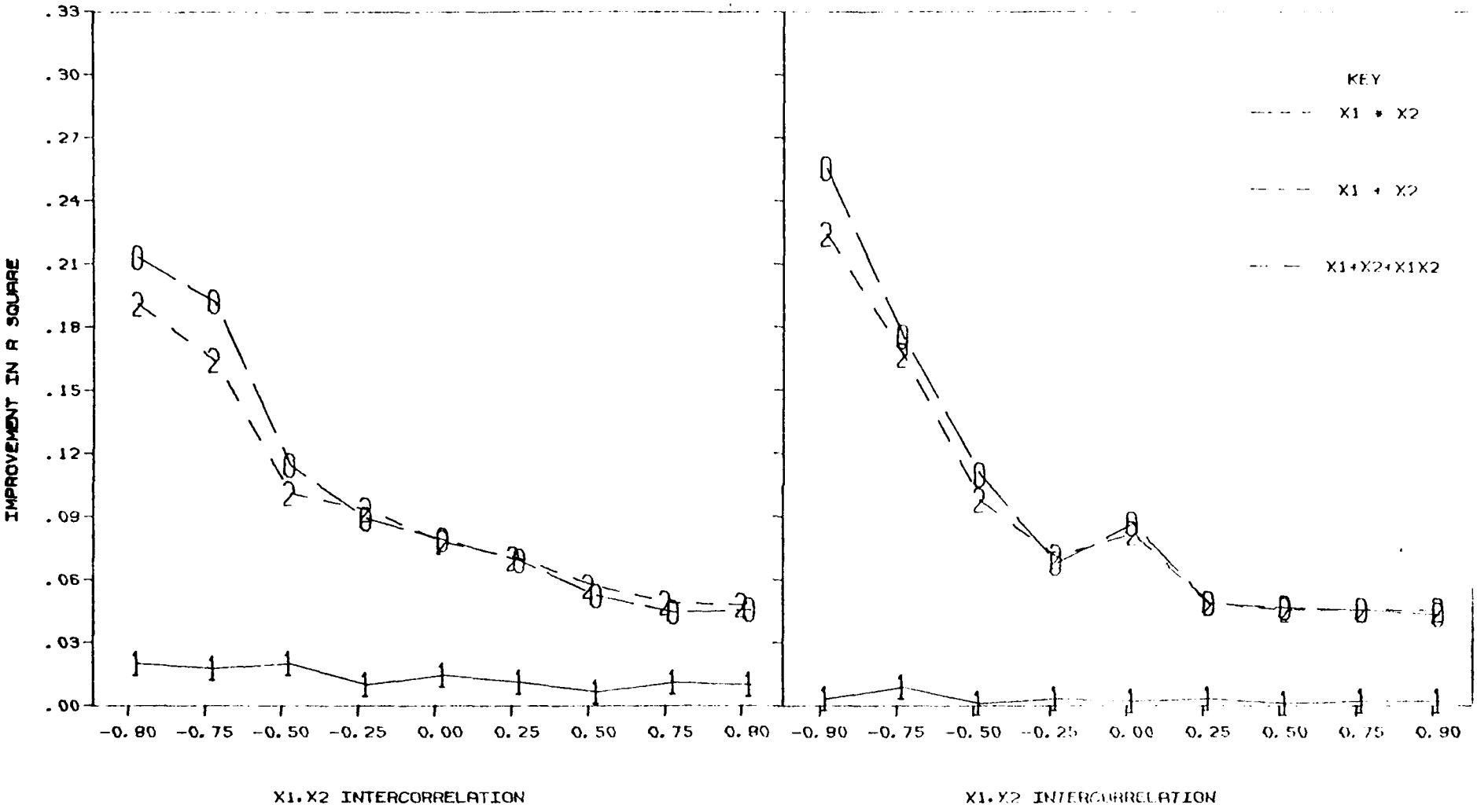
Figure 9. Proportion of significant lack-of-fit tests for three data-generating models: (0) multiplying, (1) adding, and (2) adding-multiplying. (Left panel = 25 cases, right panel = 100 cases.)

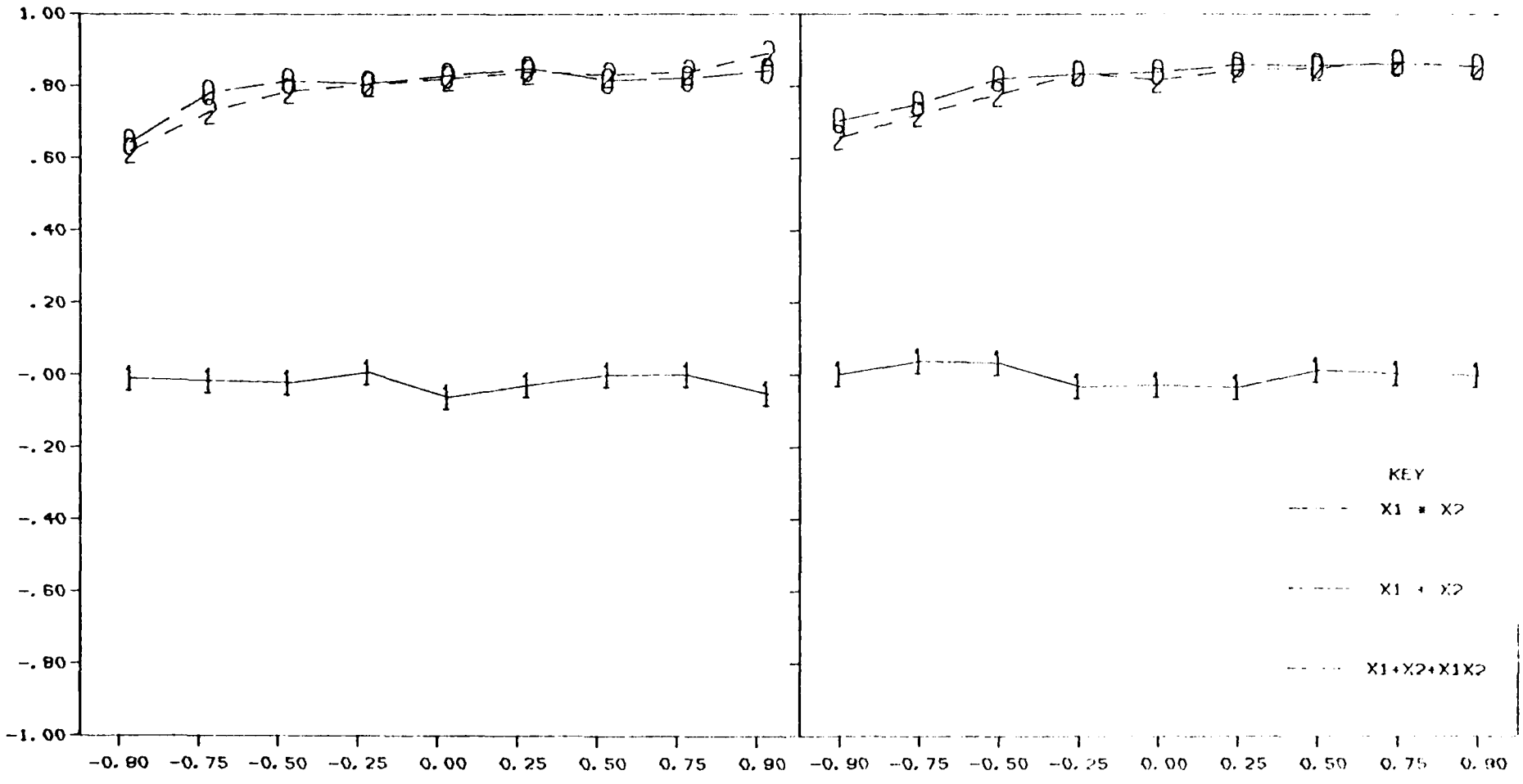


X1.X2 INTERCORRELATION

X1.X2 INTERCORRELATION

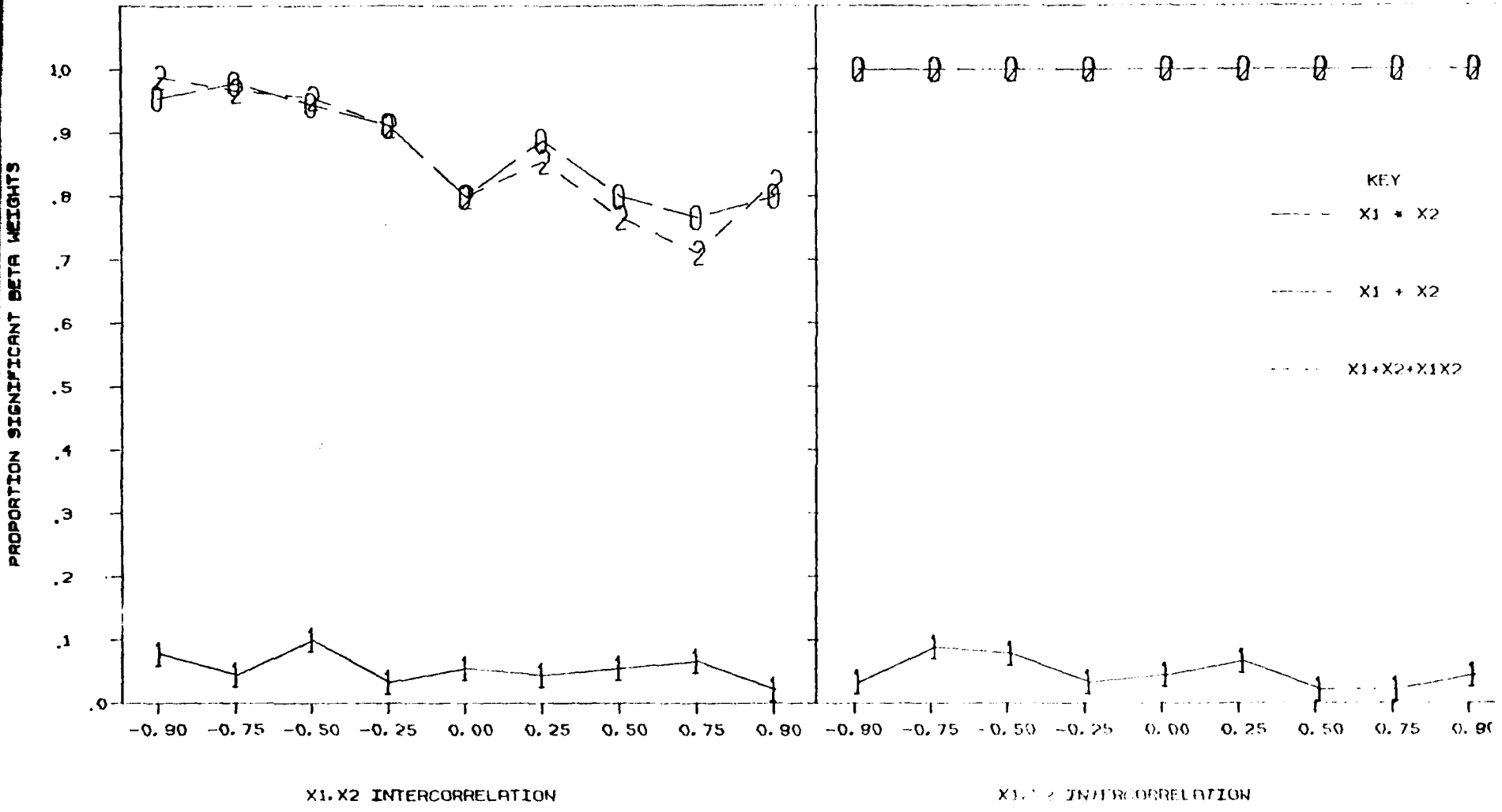

AVERAGE R SQUARE FOR MULTILINEAR MODEL
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.80
X1.X2 INTERCORRELATION
KEY
X1 * X2
X1 + X2
X1+X2+X1X2
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75
X1.X2 INTERCORRELATION

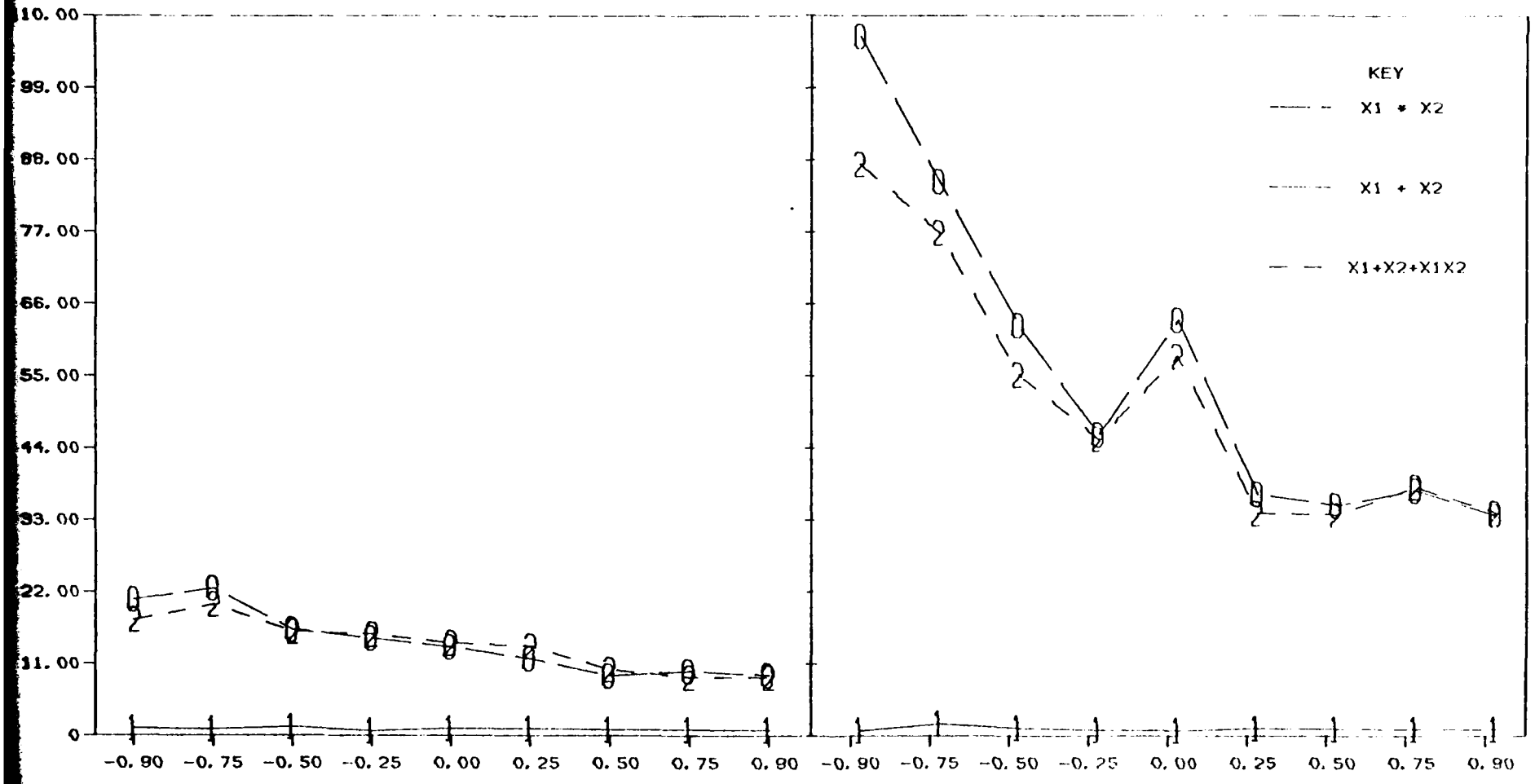

IMPROVEMENT IN R SQUARE
.33
.30
.27
.24
.21
.18
.15
.12
.09
.06
.03
.00
-0.80 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.80
X1.X2 INTERCORRELATION
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
X1.X2 INTERCORRELATION
KEY
X1 * X2
X1 + X2
X1+X2+X1X2



X1.X2 INTERCORRELATION

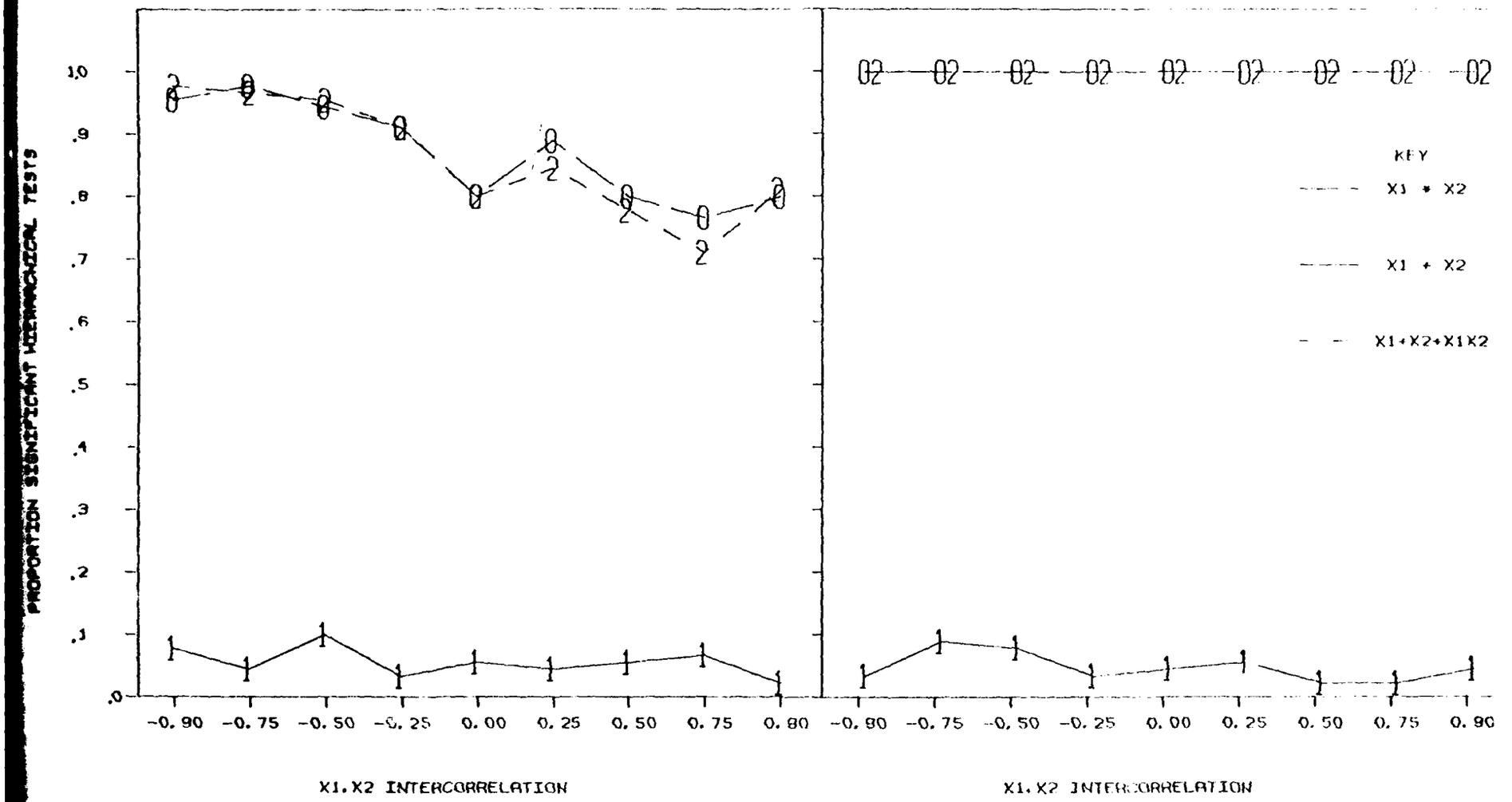
X1.X2 INTERCURRELATION

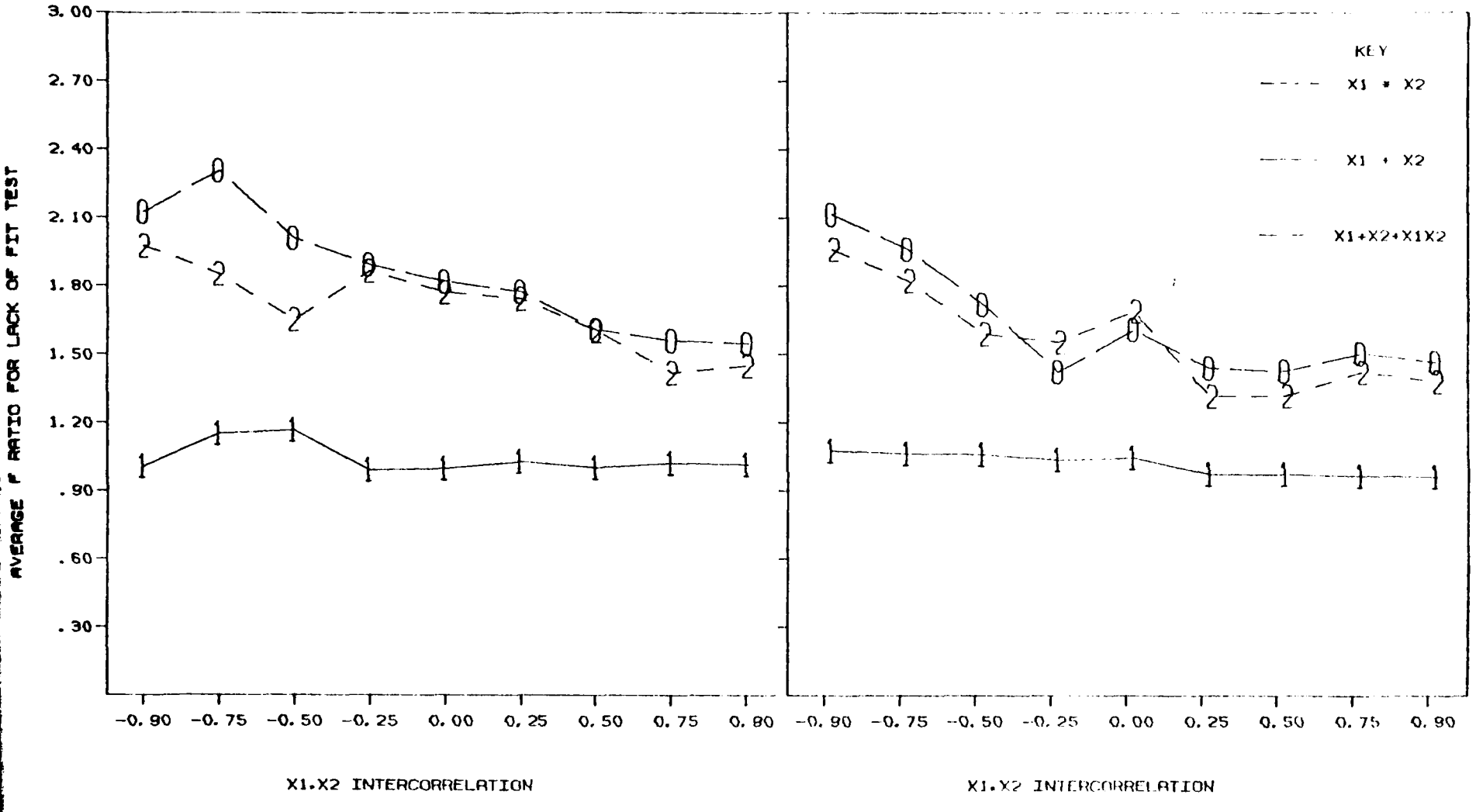

PROPORTION SIGNIFICANT BETA WEIGHTS
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
KEY
X1 * X2
X1 + X2
X1+X2+X1X2
X1.X2 INTERCORRELATION

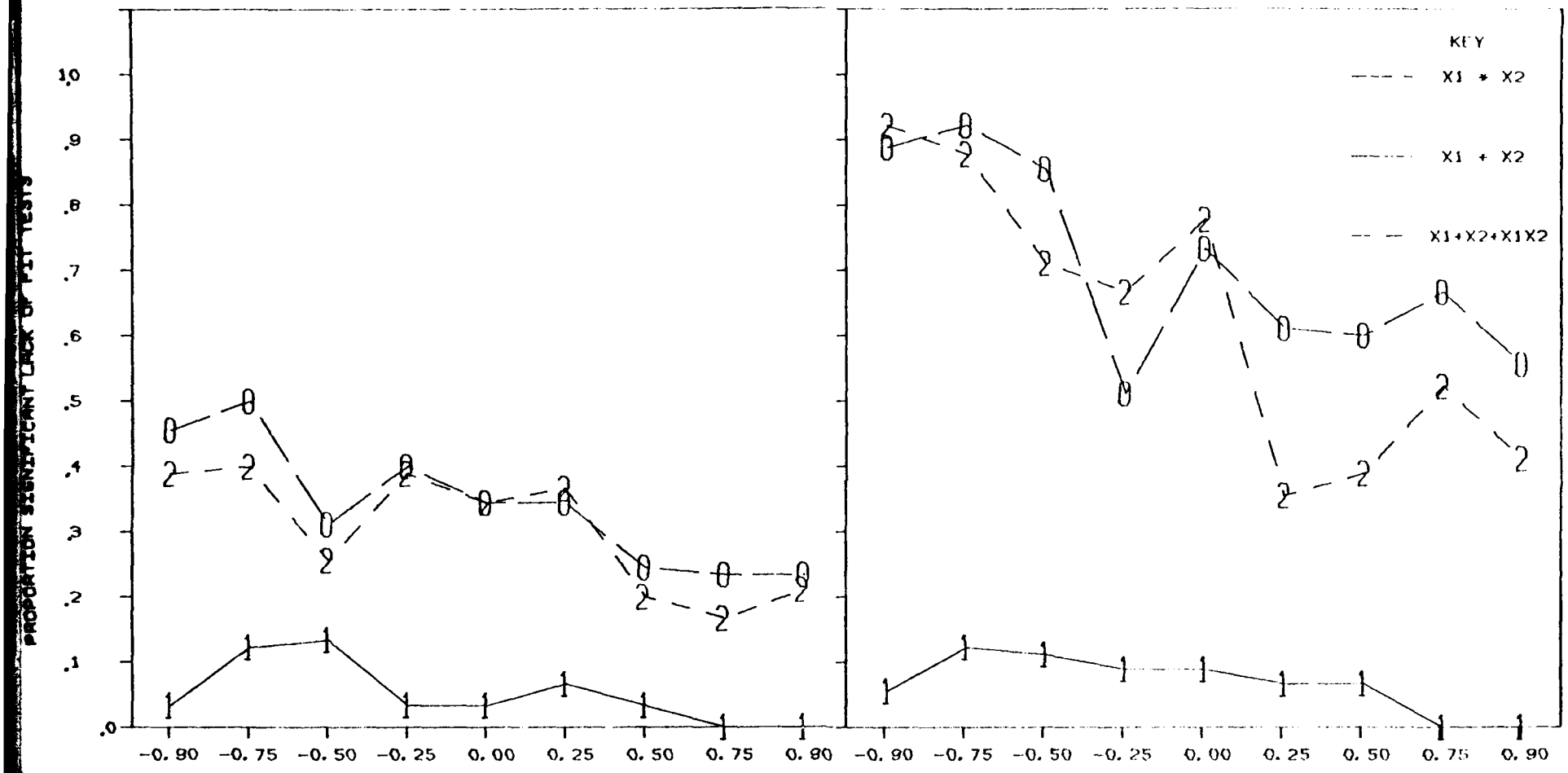


X1.X2 INTERCORRELATION

X1.X2 INTERCORRELATION


PROPORTION SIGNIFICANT HIERARCHICAL TESTS
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
X1.X2 INTERCORRELATION
KEY
X1 * X2
X1 + X2
X1+X2+X1X2
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
X1.X2 INTERCORRELATION


AVERAGE F RATIO FOR LACK OF FIT TEST
3.00
2.70
2.40
2.10
1.80
1.50
1.20
.90
.60
.30
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.80
X1.X2 INTERCORRELATION
KEY
X1 * X2
X1 + X2
X1+X2+X1X2
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
X1.X2 INTERCORRELATION


PROPORTION SIGNIFICANT LACK OF FIT TESTS
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
X1.X2 INTERCORRELATION
KEY
X1 * X2
X1 + X2
X1+X2+X1X2
-0.90 -0.75 -0.50 -0.25 0.00 0.25 0.50 0.75 0.90
X1.X2 INTERCORRELATION

Appendix

Two of the major components of the present simulation analysis deserve greater elaboration. The first concerns specification of the CUEGEN program used to construct stimulus cases. The second involves discussion of the various statistical procedures used to analyze the data. Each of these will be addressed in following supplementary material.

CUEGEN Program

This program is based on procedures described in Kaiser and Dickman (1962) with modifications outlined in Boyle (1970). Relevant programming information can be found in Naylor, Balintfy, Burdick, and Chu (1965). Additional modifications were incorporated especially for this research project by Michael O'Reilly.

Program description. CUEGEN will generate sample stimulus cases which will approximate user specified values for the means, standard deviations, and intercorrelations. The sample cases will satisfy with maximulm accuracy (in a least-squares sense) the specified values within the limits of computational accuracy, computing time, etc. The user may also specify either a uniform or a normal distribution and this property will also be maximally satisfied within limits.

The algorithm used by CUEGEN starts with the desired (population) correlation matrix. Through the use of component analysis, random sample matrices are generated from the population correlation matrix. If a sample matrix does not meet the intercorrelation requirements, then adjustments are made in the elements of the matrix to bring it more in line with the requested matrix. Once a satisfactory sample correlation matrix is achieved, linear transformations are applied to produce the desired means and

standard deviations. All properties, including the distributions, are statistically checked, and a sample matrix is rejected if any property is not satisfied at the .05 level.

Technical description. The remaining material provides a more technical discussion of the CORGEN program. A fundamental postulate of component analysis states that:

$$\underline{Z} = \underline{F}\,\underline{X},$$

where $\underline{F}$ (or order n x n) is a factoring of $\underline{R}$, the desired correlation matrix, and $\underline{X}$ (of order n x n) is a population matrix derived from the components in $\underline{F}$. The program begins by generating an arbitrary $\hat{\underline{X}}$, sampling randomly from uncorrelated populations with any distribution and with zero mean and unit variance. Then

$$\hat{\underline{Z}} = \underline{F}\ \hat{\underline{X}}$$

can be found, where $\hat{\underline{Z}}$ represents a matrix of observations from a multivariate population with zero means, unit variances, and correlations $\underline{R}$.

If the absolute correlational error is larger than some acceptable value, then an element of $\hat{\underline{Z}}$ is chosen at random and adjusted by a preset step (default value of 1.0) to reduce the error. The direction of the adjustment is chosen by examining the effects of the change on the correlation matrix. The process is repeated (within specifiable limits) until the desired intercorrelation values are obtained. Finally, the rows of $\hat{\underline{Z}}$ may, if necessary, be linearly tranformed to reflect specified means and standard deviations.

It is possible for a user to request a pattern of properties which, upon analysis, leads to a correlation matrix which is non-Gramian. To produce a correlation matrix most like the specified correlation matrix, negative eigen values (if any are found during the Gramian factoring) are set to zero.

The program can generate stimulus cases having either a uniform or a normal distribution. Consequently, a test of distribution is performed by setting up equal probability intervals. A chi-square test is used to compare the actual number of cases in each interval with the expected number. Sample matrices which fail the distribution test are randomly altered and reentered into the program as necessary (default value of 5 reentries). More information on this or any of the other programs is available upon request from the author.

Statistical Procedures

A number of statistical procedures were evaluated in the researcher simulation stage. Some of these indices were discussed in detail in the text (e.g., R^2), while others provided redundant information and so were not discussed (e.g., r). In the following material, the descriptive indices will first be described followed by the inferential indices. (Except for the indices which are not widely known, the computational formulas have been omitted.)

Descriptive indices. (1) Probably the simplest of the descriptive indices is the ordinary product moment correlation coefficient, r, between the set of predicted values, $\hat{Y}$, and the set of observed values, Y'. (2) A closely related measure is the squared multiple correlation value, R^2, which describes the variance-accounted-for by a given multiple-regression model. (3) Based on the R^2 values for the separate multiple-regression models, a ΔR^2 can be obtained from the improvement in variance-accounted-for in going from a linear model (Equation 7) to a multilinear model (Equation 8). This reflects the increase in R^2 from adding a crossproduct term to a linear model. (4) For the two regression models with crossproduct terms (i.e., Equations 6 and 8), the unstandardized regression weight, b, can be obtained for the $X_1 \times X_2$ terms. (5) Similarly, the standardized regression (Beta) weight, β, can be evaluated for the crossproduct terms.

(6) Using the derived regression values, $\hat{Y}$, the difference between derived and observed values, Y', can be summarized by mean-absolute-deviation (MAD) scores. (7) The same comparison can be made in terms of root-mean-squared deviations (RMSD). (8) The data generated by the various response models can be described by means, and (9) standard deviations. (10) Finally, a coefficient-of-variation (standard deviation ÷ mean) can be easily computed from the preceeding measures.

Inferential indices. (11) The significance of the standardized regression weight (computed in step 5) can be determined for the crossproduct terms in Equations 6 and 8. (12) The hierarchical multiple regression approach (Cohen & Cohen, 1975, chapter 8) is based on testing the improvement in R^2 in going from a linear model to a multilinear model (see step 3). The computation formula used was:

$$F = \frac{(R^2_8 - R^2_7)\ (N - k_7 - k_8 - 1)}{(1 - R^2_8)\ k_8}$$

where R^2_8 and k_8 refer to the variance-accounted-for and number of independent variables added (=1), respectively, for the multilinear model in Equation 8. R^2_7 and k_7 (=2) are the comparable values in Equation 7. The F-ratio, based on k_8, $(N - k_7 - k_8 - 1)$ degrees of freedom, can be tested directly for significance (see Arnold & Evans, 1979, p. 44).

(13) The lack-of-fit test (Draper & Smith, 1966, sec. 1.5) is based on splitting the residual sum-of-squares (SS) in a regression analysis into two parts: a lack-of-fit SS and a "pure error" SS. The F-ratio for lack-of-fit can be computed directly following standard analysis-of-variance logic. (14) To obtain an estimate of pure error, a complete repliation of the data was generated. The difference between the replications, Y_1' and Y_2', provides an independent error estimate for use in the lack-of-fit test (step 13). For all other analyses, the difference in the replicates was ignored (see footnote 1 for further details).

Distribution List

CDR Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Dr. Stuart Starr
Office of the Assistant Secretary
of Defense (C3I)
Pentagon
Washington, D.C. 20301

Director
Engineering Psychology Programs
Code 455
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Communication & Computer Technology
Code 240
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR M. Curran
Manpower, Personnel and Training
Code 270 Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mr. Randy Simpson
Operations Research Programs
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Dr. Edward Wegman
Statistics and Probability Program
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division

Dr. Michael Melich
Communications Sciences Division
Code 500
Naval Research Laboratory
Washington, D.C. 20375

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Dr. W. Mehuron
Office of the Chief of Naval
Operations, OP 987
Washington, D.C. 20350

Dr. Jerry C. Lamb
Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02840

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813

Dr. Albert Colella
Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02840

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 0722 - Rm. 508
800 North Quincy Street
Arlington, VA 22217

CDR C. Hutchins
Naval Air Systems Command
Human Factors Programs
NAVAIR 340F
Washington, D.C. 20361

Mr. Milon Essoglou
Naval Facilities Engineering Command
R & D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

CDR Robert Biersner
Naval Medical R & D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Dr. James Regan, Technical Director
Navy Personnel Research and
Development Center
San Diego, CA 92152

Dr. Julie Hopson
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Mr. Merlin Malehorn
Office of the Chief of Naval
Operations (OP-115)
Washington, D.C. 20350

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Mr. Leslie Innes
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsview, Ontario M3M 3B9
Canada

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology
Establishment
Teddington, Middlesex TW11 OLN
ENGLAND

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314

Dr. Judith Daly
Cybernetics Technology Office
Defense Advanced Research Projects
Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Gary McClelland
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, CA 94025

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

Dr. Arthur I. Siegel
Applied Psychological Services, Inc.
404 East Lancaster Street
Wayne, PA 19087

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. Robert T. Hennessy
NAS - National Research Council
JH #819
2101 Constitution Avenue, N.W.
Washington, DC 20418

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, CA 90007

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Richard W. Pew
Information Sciences Division
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

Dr. Hillel Einhorn
University of Chicago
Graduate School of Business
1101 E. 58th Street
Chicago, IL 60637

Mr. Tim Gilbert
The MITRE Corporation
1820 Dolly Madison Blvd.
McLean, VA 22102

Dr. John Payne
Duke University
Graduate School of Business
Administration
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
University of Virginia
School of Engineering and Applied
Science
Charlottesville, VA 22901

Dr. Leonard Adelman
Decisions and Designs, Inc.
8400 Westpark Drive, Suite 600
P.O. Box 907
McLean, VA 22101

Dr. Lola Lopes
Department of Psychology
University of Wisconsin
Madison, WI 53706

Mr. Joseph G. Wohl
Alphtech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Rex Brown
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Mr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19190