Developments in Business Simulation and Experiential Exercises, Volume 26, 1999

AN EXAMINATION OF A REANALYSIS OF THE IMPACT OF A MARKET LEADER ON SIMULATION COMPETITORS' STRATEGIES

John R. Dickinson, University of Windsor

ABSTRACT

For empirical research it is obviously important that the conceptual analysis plan be valid and for that plan to be executed faithfully. Where the *raison d'etre* for the study is a reanalysis of previously published theories, hypotheses, and data, that the plan for reanalysis be valid and be executed faithfully become *sine qua non*. Too, it is incumbent on the reanalysts to explain the contribution to knowledge of their study beyond that of the original work. This paper specifically takes issue with a 1997 ABSEL reanalysis on these counts and more generally describes the often inappropriate application of the popular MANOVA-univariate-F-test analysis paradigm.

INTRODUCTION

A study published by Wellington and Faria ("The Impact of an Artificial Market Leader on Simulation Competitors' Strategies," 1997b[1]) is a reanalysis of theories and data of a previously published and copyrighted work ("The Impact of a Market Leader on Simulation Competitors' Strategies," Wellington, Dickinson, and Faria 1990).

Reconsideration, including reanalysis, of previously published work is not only appropriate, it is an imperative for any discipline whose research would advance knowledge. In this instance, however, the reconsideration is inappropriate and invalid. The results reported by W&F generally are numerically impossible and, thus, incorrect. More profoundly fallacious than the execution of their analysis plan, though, are the *raison d'etre* for the reanalysis and the conceptualization of their analysis plan. Augmen-ting this fallaciousness is the failure of the researchers to interpret their results (1) *vis-a-vis* the substantive hypotheses tested and (2) *vis-a-vis* the analyses and results of the previously published study.

The W&F analysis plan and the present examination of that analysis have implications beyond this single study. The W&F analysis plan perpetuates a common analysis paradigm, while this examination contends that that paradigm is, perhaps in the majority of instances, applied when not appropriate.

A SUMMARY OF THE TWO STUDIES' RESEARCH CONTEXTS

The research of WD&F and W&F involved a marketing simulation game. Criterion variables were a mix of 20 marketing strategy decisions, e.g., expenditures on advertising, number of salespeople employed, and so on. Criterion variables were classified as either "push" or "pull" marketing strategy variables according to a well-established principle in marketing management. Experimental treatment variables were (1) the parameters of the simulated marketplace as manipulated by the researchers, i.e., a marketplace more responsive to push decisions versus a marketplace more responsive to pull decisions and (2) the presence or absence of a competing company controlled by the researchers. The strategy decisions of this "artificial leader" or "ringer" company were controlled by the researchers to signal to other

[1] Page references to the Wellington and Faria (1997b) study refer to the complete paper provided by the authors, rather than to the condensed version published in the ABSEL *Proceedings*. In this paper Wellington and Faria (1997b) is abbreviated W&F. The original Wellington, Dickinson, and Faria (1990) paper is abbreviated WD&F.

competitors emphasis on push or pull strategy variables consistent with the manipulated marketplace.

Two examples of hypotheses tested in both the WD&F and W&F studies are:

"H1: Companies in pull environment industries will allocate greater resources to pull variables in artificial leader [ringer] industries than will companies in nonartificial leader [nonringer] industries." (W&F, p. 7; WF&D, p. 36)

"H2: Companies in pull environment industries will allocate fewer resources to push variables in artificial leader [ringer] industries than will companies in nonartificial leader [nonringer] industries." (W&F, p. 8; WF&D, p. 36)

Three relevant properties of these example hypotheses may be noted:

- The hypotheses are based on *a priori* theory, with the pull and push hypotheses being derived from a well-established marketing management principle.
- The hypotheses express a direction of the theorized relationships.
- The hypotheses of W&D had been investigated in the previous WD&F study utilizing the same data.

The W&F analysis comprised a series of eight multivariate analyses of variance (MANOVA) tests plus a total of 76 (nondirectional) univariate F-tests of individual criterion variables as reported in Tables 1 and 2 of their paper (pp. 10,11). Table 1 is reproduced in this paper. The WD&F analysis comprised a series of directional t-tests of individual criterion variables.

NUMERICAL IMPOSSIBILITY

For W&F, as a preliminary step, each criterion variable was "...transformed into T-scores (mean of 50 and standard deviation of 10)..." (W&F, p. 9) Transformation into T-scores is an affine transformation and should have no effect on subsequent MANOVA and univariate F significance tests (Morrison 1967, p. 124). The mean of each transformed criterion variable calculated across the total sample, then, equals 50. When the data are divided by experimental group (specifically four groups in this 2 x 2 design), for any given variable it is then impossible for all four group means to be less than 50. Of the 19 criterion variables tested by W&F (Tables 1 and 2, pp. 10, 11), six do not satisfy this condition.

The total sample in this study comprised 42 companies. For the mean of a given transformed variable to equal 50, the sum of the transformed values across these 42 companies must equal 2100 (_=50=2100/42). The actual sum based on W&F's published mean values may be obtained by multiplying each of the four group means by the relevant experimental group sample size. For example, for the first criterion: 9(50.3)+12(49.3)+9(43.9)+12(45.1)=1980.6.
The sum for the first criterion variable does not equal 2100 and the group means tested by W&F are not possible.

The absolute deviation of 119.4 (=2100-1980.6) is not likely due to rounding error. Rounding errors tend to "average out," with values rounded down offsetting values rounded up. In the unlikely scenario that all 42 of the transformed values were rounded in the same direction, the total absolute deviation could be no greater than 21. If the transformed values for all 42 companies were truncated, the effect on the sum could be no more than a total absolute deviation of 42. Across the 19 sets of four experimental group means each, 16 of the total absolute deviations from 2100 are greater than 21, 14 are greater than 42. A large majority, if not all, of the 76 comparisons of individual means analyzed by W&F are necessarily incorrect.

Within each of W&F's eight MANOVA analyses,

the group mean values for at least five of nine (the push decision variables) or nine of ten (the pull decision variables) criterion variables are not possible and the inferential results for all eight are incorrect.

UNINTERPRETABLE AND LESS POWERFUL MANOVA

Inability to Test Directional Theories

In the context of this research, results of MANOVA may, and normally will, be uninterpretable *vis-a-vis* the conceptual hypotheses. This is the case for all eight of the MANOVA results reported by W&F. The reason is that the conceptual hypotheses are directional, while MANOVA is incapable of testing directional relationships. The MANOVA test for H1 is reported to be significant with p=.011 (Table 1). For six of the criterion variables, the mean value under the artificial leader treatment condition is greater than the mean value under the nonartificial leader condition.

These six results are consistent with the *a priori* theorized direction. For the remaining four criteria, though, the difference between the group means is in the direction opposite that theorized. What interpretation, then, can be made as to whether H1 is supported or not? The significance of .011 simply indicates that the two vectors of means are not equal. However, it is not possible to ascertain whether that significance is due to the differences that are in the hypothesized direction, thus supporting H1, or due to the differences in the direction opposite that hypothesized, thus refuting H1.

The interpretation by W&F is that "...the results shown in Table 1 indicate a significant difference in the overall decision strategies of the companies. Teams in the artificial leader industries did devote more resources to the pull variables than those in the nonartificial leader industries." (p. 12) This interpretation is not warranted. For four of the ten criterion variables, teams in artificial leader industries devoted less, not more, resources to pull variables than those in the nonartificial leader industries. Four of the 10 descriptive results are contrary to H1 and there is no basis for concluding that the statistical significance of the MANOVA somehow reflects directional differences as theorized any more than it reflects the directional differences opposite those theorized.

The use of MANOVA to test directional hypotheses is generally an inappropriate analysis paradigm. The scenarios in which statistical significance is unambiguous are very limited. For the W&F example, even had all ten of the sample differences been in the theorized direction this would be no assurance that H1 was supported. It is possible, for instance, that regardless of being in the theorized direction, that, say, nine of the ten population means are equal with the difference on the tenth criterion accounting for the MANOVA significance. Few would contend that such a result warrants the conclusion that the presence of an artificial leader truly impacts the decisions of simulation participants.

Lesser Statistical Power

"Further, the situations in which MANOVA is more powerful than ANOVA are quite limited; often MANOVA is considerably less powerful than ANOVA." (Tabachnick and Fidell 1996, p. 376) MANOVA is often less powerful than ANOVA. The directional t-tests applied by WD&F are, as explained below, even more powerful than ANOVA. (It is also likely that the nondirectional nature of MANOVA lessens its power even more. Research is presently underway to investigate this possibility.)

In sum, in W&F the MANOVA significance level *per se* is uninterpretable *vis-a-vis* the theorized hypothesis. Insight into the individual criterion variables *may* be drawn from univariate tests. But that insight lies within the univariate tests and is absolutely independent of the MANOVA analyses.

In each of the eight W&F MANOVA analyses at least one mean difference is in the direction opposite that theorized. The W&F MANOVAs do not and can not contribute to knowledge.

INAPPROPRIATE UNIVARIATE F-TESTS

The second type of analysis applied by W&F was "...the univariate F-test results produced by the [SPSS] MANOVA program." (p. 9) Univariate F-tests are clearly inferior to the directional t-tests applied in the original WD&F study and may be readily dismissed as providing no incremental knowledge beyond that provided by the original study.

First, the conceptual hypotheses are directional and the univariate F-test of means is definitionally incapable of testing directional hypotheses. Thus, the theoretical hypotheses and the W&F statistical test are in this way incompatible; the logic of the test is inconsistent with the logic of the hypotheses.

Second, the univariate F-tests are much more susceptible to Type II error than are the directional t-tests; i.e., the nondirectional F-tests are less powerful than the directional t-tests. The test applied by W&F is needlessly weak and in this respect their results are invalid. In the case of two means $F=t^2$ (Guenther 1965, p. 204) and the two tests in this way are identical. However, as mentioned above, the F-test is incapable of testing directional hypotheses. Had the W&F univariate F-tests been executed correctly and had W&F compared their results with the results of WD&F, it would have been apparent that the p-values of the former were exactly twice the p-values of the latter.

The W&F tests are merely the less conceptually appropriate and statistically weaker version of the same tests applied by WD&F. Both logically and statistically, the univariate F-tests applied by W&F are inferior to the directional t-tests applied by WD&F. As with the MANOVA analyses, the W&F tests of individual criterion variables do not and can not advance knowledge beyond the contribution of the original study. To the contrary, the greater susceptibility of W&F's less powerful nondirectional F-tests to Type II errors poses a reduction, not an advancement, of knowledge.

OTHER MISUSES OF MANOVA

One of the statistical advantages commonly ascribed to the use of MANOVA is the control of Type I error. "[MANOVA] solves the type 1 error rate problem by providing a single overall test of group differences across all dependent variables at a specified α level." (Hair, Anderson, and Tatham 1987, p. 150) This control simply stems from there being but a single test of significance compared to the multiple tests of significance with a series of univariate tests. The control only attends the MANOVA significance test and has absolutely no statistical relationship with subsequent F-tests (or t-tests). In the analysis paradigm of W&F, there is absolutely no control of Type I error with respect to the univariate F-tests. (Also, the t-tests of WD&F are exactly equally susceptible to inflation of Type I error as the F-tests of W&F and *vice versa*.)

Additional statistical reasons commonly given for using MANOVA include:

- "The univariate tests ignore important information, i.e., the correlations among the variables. The multivariate test incorporates the correlations (via the covariance matrix) right into the test statistic..." (Stevens 1996, p. 152)
- "Although the groups may not be significantly different on any of the variables individually, *jointly* the set of variables may reliably differentiate the groups." (Stevens 1996, p. 153)

Both of these advantages derive from the information contained in the intercorrelations among the criterion variables. Hypotheses 1 and 2 together theorize differences in a total of 19

criterion variables as a function of the presence or absence of an artificial leader company. In like manner, the same 19 criterion variables comprise the H4 and H5 couplet in W&F (p. 8). H5 and H6 in W&F each incorporate all 19 variables (p. 8).

If the use of MANOVA is to yield the two advantages delineated above, it is obviously necessary for all of the criterion variables to be analyzed together. Yet in all instances of the W&F MANOVAs, the criterion variables are divided into subsets and are in no instance analyzed together. The two delineated advantages of MANOVA are precluded by the W&F analyses.

EPILOGUE

Absence of Comparison With Prior Research

This paper has taken issue with the reanalyses as reported by Wellington and Faria (1997b) in their published paper. It is possible that their reanalysis does contribute to knowledge beyond that made by the original analysis. However, it is incumbent on the researchers to describe that incremental contribution. The original study is not acknowledged by W&F. Rather, its existence appears to be denied: "While no past research has used an artificial industry leader to examine participant responsiveness to simulation environment..." (Wellington and Faria 1997b, p. 2) It follows that no such description of incremental contribution appears in their paper. In response to subsequent inquiries, the researchers have provided no explanation of any incremental contribution to knowledge.

Infeasibility of Scheffe's Multiple Comparison Procedure

In a presentation of their research (Wellington and Faria 1997a), the researchers claimed that univariate F-tests had *not* been conducted, that the paper stating univariate F-tests had been done was a misstatement, and that the results presented in the tables of the paper were the results of directional Scheffe (multiple comparison) tests. However, it is clear from the authors' own interpretation of the tests that the univariate tests were not directional. Too, the very notion of Scheffe's multiple comparison test in the context of this research is questionable. Scheffe's test is a *post hoc* procedure, while this research tests *a priori* hypotheses, conceptually supported, with W&F having available the further support of the previously published WD&F empirical results. In response to subsequent inquiry, the researchers have provided no example of a Scheffe multiple comparison in the context of their research. In fact, there are no multiple comparisons in the context of this research and the basis for the researchers' claim remains unexplained.

CONCLUSION

The essential contribution to knowledge of empirical studies lies in their empirical results. W&F's descriptive empirical results are invalid, virtually completely in their entirety, and their inferential empirical results are invalid, literally completely in their entirety. This is due not only to numeric errors, but to an analysis plan that by nature is uninterpretable or unnecessarily weak. These fatal shortcomings may have become apparent had the researchers expressly undertaken to explain the superiority of their analysis plan over the originally published analysis plan. Invalid empirical results are anathema to the advancement of knowledge, aggravated in the case of W&F by their aim to improve upon an already existing work.

REFERENCES

Hair, Joseph F., Jr., Anderson, Rolph E., & Tatham, Ronald L. (1987). *Multivariate Data Analysis*, Second Edition, New York: Macmillan Publishing Company.

Guenther, William C. (1965). *Concepts of Statistical Inference*. New York: McGraw-Hill Book Company.

Morrison, Donald F. (1967). *Multivariate*

Statistical Methods. New York: McGraw-Hill Book Company.

Stevens, James (1996). *Applied Multivariate Statistics for the Social Sciences*, Third Edition, Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Tabachnick, Barbara G. & Fidell, Linda S. (1996). *Using Multivariate Statistics*, Third Edition New York: HarperCollins College Publishers.

Wellington, William J., Dickinson, John R., & Faria, A. J. (1990). The Impact of a Market Leader on Simulation Competitors' Strategies. In Weinroth, Jay and Hilber, Joe E. (editors), *Simulation in Business and Management, 1991*. Proceedings of the SCS Multiconference on Simulation in Business and Management, San Diego: The Society for Computer Simulation, 33-40.

Wellington, William & Faria, A. J. (1997a). The Impact of an Artificial Market Leader on Simulation Competitors' Strategies. Faculty Research Forum, Faculty of Business Administration, University of Windsor (March 14), presentation.

Wellington, William J. & Faria, A. J. (1997b). The Impact of an Artificial Market Leader on Simulation Competitors' Strategies. In Butler, John K., Jr. and Leonard, Nancy H. (editors), *Developments in Business Simulation and Experiential Learning*. Proceedings of ABSEL, Statesboro, GA: Association for Business Simulation and Experiential Learning, 152-157.

TABLE 1
MEAN STRATEGY DECISION VALUES FOR H1 THROUGH H4

Criterion	Pull Environment		Push Environment	
	Artificial Leader	Nonartificial Leader	Artificial Leader	Nonartificial Leader
Pull Decision Variables		**H1**		**H4**
Broadcast Adv				
S100, T1	50.3	49.3	43.9	45.1
S100, T2	49.8	50.1	43.9	44.3
D200, T1	50.1	49.9	43.7	44.8
D200, T2	49.2	52.8	42.8	43.7
Print Adv				
S100, T1	53.4	46.7**	45.7	45.7
S100, T2	53.1	48.3**	45.2	45.0
D200, T1	53.5	47.3**	45.6	45.5
D200, T2	52.6	48.7	44.7	44.6
R & D				
Standard 100	48.3	48.7	48.8	49.4
Deluxe 200	48.2	49.0	48.6	50.0
MANOVA Results H1			H4	
N		21	N	21
Pillais		.82672	Pillais	.51175
Exact F		4.771	Exact F	1.048
Degress of Freedom		10	Degrees of Freedom	10
Significance		.011**	Significance	.471
Push Decision Variables		**H2**		**H3**
Trade Adv				
S100, T1	45.7	45.3	47.3	45.0**
S100, T2	45.6	45.2	47.0	44.6**
D200, T1	45.6	45.1	47.6	44.7**
D200, T2	45.3	44.8	46.8	44.3**
Co-op Advertising	47.3	51.8	59.3	43.5**
Salesforce Size				
Territory 1	49.4	56.7	48.6	46.8
Territory 2	47.7	57.1	49.5	46.5
Salesforce Salary	44.7	47.4	55.7	46.6**
Sales Commission	45.7	52.3	56.3	47.3**
MANOVA Results H2			H3	
N		21	N	21
Pillais		.30966	Pillais	.82049
Exact F		0.548	Exact F	5.586
Degress of Freedom		9	Degrees of Freedom	9
Significance		.812	Significance	.005**

** = in hypothesized direction, $p < .05$

* = not in hypothesized direction, $p < .05$