



## Research Methods & Analysis

# Simple Question, Not So Simple Answer: Interpreting Interaction Terms in Moderated Multiple Regression

Arthur G. Bedeian  
Kevin W. Mossholder  
*Louisiana State University*

*The appropriate interpretation of interaction terms in studies employing moderated multiple regression (MMR) analysis is considered. It is argued that given a theory-based, a priori hypothesis, an MMR analysis is analogous to a planned statistical comparison and, thus, a significant overall F value is not a prerequisite for interpreting a significant interaction term.*

*Science, like all creative activity, is exploration, gambling, and adventure. It does not lend itself very well to neat blueprints, detailed road maps, and central planning. Perhaps that's why it's fun.*

—Simon, 1964, p. 85

It appeared to be a simple question. A colleague wished to know whether or not we could justify interpreting an interaction term when the overall regression equation of which it was a part failed to reach significance. We had used moderated multiple regression (MMR) analysis to examine the effect of a moderator variable on the relation between a dependent and an independent variable. A restricted model comprised of the independent variable and a hypothesized moderator had been created by entering both terms as a block. Next, a full MMR model had been constructed, adding the focal interaction term (independent variable  $x$  moderator variable) to the restricted model. Standard statistical tests had been used to determine if the incremental variance explained by the interaction term was significant. A simple question, yes; a question with an simple answer, no, as the ensuing quest for justification would soon reveal.

Our colleague's contention was that even though significant, the individual terms in an MMR model could not be meaningfully interpreted unless the

---

Direct all correspondence to: Arthur G. Bedeian, Louisiana State University, Department of Management, Baton Rouge, LA 70803-6312, U.S.A. E-mail: MGBEDE @ LSUVM. The authors are listed alphabetically; both contributed equally to this work.

---

overall  $R^2$  for the model was significant. Our position was that because the theoretical underpinning for our study specified that an interaction effect would occur, only the statistical significance of this effect should be considered in determining whether our hypothesis had been supported. What did the quantitative research literature have to say on this point? In a follow-up conversation with our colleague, we quoted from Baron and Kenny (1986):

The moderator hypothesis is supported if the interaction ... is significant. There may also be significant main effects for the predictor and the moderator ..., but these are not directly relevant conceptually to testing the moderator hypothesis (p. 1174).

There, in black and white, support from the literature. Surely this would satisfy our skeptical colleague.

Not so! Our colleague's reply was prompt and direct. He too had statistical sources that supported his reasoning. Referencing Cohen and Cohen (1983), among other texts, he noted that apparently overall tests of significance in regression analysis apply regardless of whether the independent variables are main effects, interaction terms, or whatever. He did, however, allow that he would be willing to consider other arguments from the statistical community.

The gauntlet had been thrown down and we accepted. Our first recourse was to conduct a literature search. Initially focusing on the moderated regression literature, we were able to locate several articles in which management researchers interpreted significant interactions in the context of nonsignificant equations. The articles tested various interaction effects and represented assorted management fields, including organization theory (McKinley, Cheng, & Schick, 1986), organizational behavior (Sutton & Rafaeli, 1987), strategy (Gupta & Govindarajan, 1986), and production/operations (Snell & Dean, 1992). Only one, however, made even passing mention of the issue at hand (Govindarajan & Fisher, 1990, p. 275n). At the same time, none articulated either a theoretical model or an explanatory framework arguing for interaction effects only (with no main effects proposed).

Turning next to standard regression texts, we found that some contained relevant information, but none addressed our quest in a manner specific enough to be of direct bearing. Because Cohen and Cohen (1983) is one of the most comprehensive applied multiple regression texts with which we are familiar and, perhaps more importantly, our colleague had referenced this text, we combed through it for new insights. Relevant information was found in various chapters and we were thus able to glean sufficient information to support our position.

Our colleague's contention, that an interaction term is uninterpretable unless the overall  $F$  test associated with an MMR model reaches significance, implicitly emphasizes the importance of alpha level protection when multiple comparisons are performed. In this regard, Cohen and Cohen (1983, p. 108) do observe that it is possible to find a nonsignificant  $R^2$  even though  $t$ -tests for individual effects are significant. They indicate that this may occur, for example, when several extremely small effects reduce the average contribution

per predictor variable, making an overall  $F$  non-significant in spite of the apparent significant effects of one or more individual predictor variables. It is critical to recognize, however, that Cohen and Cohen (1983) offer this specific example as it applies to the use of a simultaneous analysis approach to multiple regression in which all independent variables are examined at the same time, with the goal of determining the unique effect of each variable in a set. Had we used a simultaneous regression approach, we would agree with Cohen and Cohen's (and our colleague's) recommendation that any significant effects should be ignored when an overall  $R^2$  is not significant. The usual MMR model, however, involves a hierarchical rather than simultaneous analysis approach. With a hierarchical regression approach, certain predictors are such that an assessment of their contribution to  $R^2$  is meaningful only after related predictors have been partialled. This necessitates entering predictors in a specific, theory-guided order, as occurs in the representation of interactions and curvilinear relations, among others (Cohen & Cohen, p. 227).

When a theoretical, a priori hypothesis predicts an interaction between two variables, MMR is used to test the incremental  $R^2$  associated with a cross-product term from which the effects of both a predictor and moderator variable have been partialled. Main effects are entered first in an MMR model, not for the purpose of testing these effects per se, but rather to remove their effects from the cross-product. This is necessary because a cross-product term carries both main and interaction effect information (Cohen & Cohen, 1983, p. 305). To our knowledge, there is no requirement for the main effects to be significant in an MMR model, for they are seldom relevant to a hypothesized interaction effect.<sup>1</sup> Requiring a significant overall  $F$  before testing the significance of an interaction term allows main effects of any nature to influence an examination of an hypothesized interaction. This would seemingly allow chance elements to influence the intended analysis.

Alpha-level protection for multiple comparisons is most important when the comparisons are post hoc. That is, unplanned in the sense that they are suggested by the outcome of a study and are not specifically anticipated during a study's planning stage. When a single, planned interaction term is examined, the logic of an MMR analysis coincides with that of an "unprotected" or planned comparison, where protection against multiple comparisons is unnecessary because only one theoretically relevant test is conducted. Requiring a significant overall  $R^2$  (i.e., an MMR model which must include two potentially nonsignificant main effects) would seem to be a very restrictive strategy.

The restrictiveness of this strategy can be illustrated in the extreme by considering an MMR model in which the two main effects are nonsignificant, explaining no variance in a dependent variable, and the interaction term is significant, explaining all of the variance in the dependent variable. We examined what could happen in an MMR model if an overall  $R^2$  had to be significant before an interaction could be interpreted, with sample sizes permitting us to identify critical  $F$  values ( $p < .05$ ) from available  $F$  tables and significance tests for overall  $R^2$  and  $R^2$  increment (Cohen & Cohen, 1983, pp. 103-107). Overall  $R^2$ 's were arbitrarily chosen for illustrative purposes and were

identical to  $R^2$  increment given that all variance is attributable to the interaction term.

For an  $n$  of 104 and an overall  $R^2$  value of .07, the  $F$  for a full model is 2.51, whereas the critical  $F$  ( $df = 3, 100$ ) is 2.70, indicating that the overall  $F$  for the full model is not significant. However, an incremental  $F$  test for the interaction term ( $df = 1, 100$ ) is 7.53, which is significant (critical  $F = 3.94$ ). In this scenario, requiring a significant overall  $F$  before examining the interaction would have resulted in ignoring an effect accounting for 7 percent of the variance in the dependent variable. We repeated these calculations, arbitrarily increasing the sample size and decreasing the size of the overall  $R^2$  value, to extend the range of our illustration. With  $ns$  of 129, 154, and 204, interactions accounting for 6, 5, and 3 percent of the variance would be ignored were a significant overall  $R^2$  required. The size of a meaningful interaction is, of course, subject to debate and theoretical dictates. What is not subject to debate is that the significant interactions in the given scenarios, regardless of theory, would be disregarded were a significant overall  $R^2$  required.

To us, the issue of requiring an overall  $R^2$  to be significant turns on whether statistical tests within a regression model are considered planned or unplanned. The strategy of using planned versus unplanned tests is less often discussed in regression than analysis of variance (ANOVA) texts, perhaps because such tests were originally developed in the context of experimentation. The ANOVA omnibus  $F$  test, however, is conceptually related to the MMR overall  $F$  test (McClelland & Judd, 1993). Within an ANOVA framework, it is generally acknowledged that when specific comparisons are planned, there is no logical need to conduct an omnibus  $F$  test (e.g., Keppel & Zedeck, 1989; Kirk, 1968).<sup>2</sup> Indeed, it has been claimed that "the  $F$  test need not and should not be carried out at all" (U.S. Department of Agriculture, 1977, p. 3). This logic recognizes that a blanket  $F$  test averages treatments over comparisons and, if weakened by nonsignificant comparisons, may erroneously yield a nonsignificant  $F$  value. Given a theory-based, a priori hypothesis, we suggest that an MMR analysis is analogous to a planned comparison and, thus, a significant overall  $F$  value is not a prerequisite for interpreting a significant interaction term.

From a broader perspective, an overall  $F$  test gauges how well a single regression line fits its underlying data. A significant interaction term indicates that two or more lines fit the data better than a single regression line. Given an interaction, an overall  $F$  test would be irrelevant in that the underlying data would be best represented by a family of heterogeneous lines.

When researching the  $R^2$  issue in question, we considered a discussion in Cohen and Cohen (1983) that has bearing on the concept of protected tests and MMR. Cohen and Cohen (1983, pp. 172-173) discuss a multiple regression adaptation of Fisher's protected  $t$ -test, using the framework of sets to generalize the protected  $t$ -test to multiple regression. (Fisher's test is the usual overall  $F$  test.) Briefly, their generalization is as follows:

1. Multiple regression analysis proceeds by sets, using whatever analytic structure (e.g., hierarchical, simultaneous, or intermediate) is appropriate.

2. The contribution to dependent variable (DV) variance of each set (or partialled set) is tested for significance at the set  $p$  value by the appropriate  $F$  test.
3. If the  $F$  for a given set is significant, the individual independent variables (IVs) comprising the set are tested for significance at the set  $p$  value by means of a standard  $t$ -test (or equivalently  $F$ , with  $df = 1$ ).
4. If the setwise  $F$  is not significant, no tests on the set's constituent IVs are permitted.

This procedure protects against Type-I error inflation while still providing decent power. It appears that in the case of a MMR model hypothesizing a single interaction, the incremental  $F$  test effectively mimics a protected  $t$  procedure. In an MMR model there are two sets being tested, a set containing two main-effect variables (an independent variable and a moderator variable) and a partialled set containing one cross-product term. Because there is only one variable in the partialled set, the significance level of an  $F$  test for the partialled set is the same as for a  $t$ -test on the variable. Thus, even though we argue that conducting a protected test is unnecessary, in the case of an MMR model involving a single interaction, alpha-level protection is provided.

Though we felt comfortable with what our literature search uncovered, we nevertheless pursued a second means of meeting our colleague's challenge. This involved tapping the collective wisdom of acknowledged experts. We wrote to several individuals whose expertise in areas related to the focal issue was widely known, repeating the question posed by our colleague. They responded:

Although I have no experience with moderated regression per se, my view of contrasts is that any predicted contrast can indeed be interpreted apart from any omnibus or overall test to which it might contribute.

Robert Rosenthal  
Letter dated April 6, 1993

If the purpose of a study is to assess an interaction, then neither the significance of the main effects nor of the overall  $R^2$  is relevant.

Jacob Cohen  
Personal communication, August 17, 1993

Although I would certainly prefer that the R-squared for the equation be significant, I would not be as troubled by it not being significant in moderated regression. The key coefficient is the product and if the main effects are small a non-significant  $R^2$  is not really worrisome.

David A. Kenny  
Letter dated April 19, 1993

In the midst of our dialogue with our colleague and without our instigation, debates concerning the " $R^2$  problem," as it was labelled, spilled over into both the Academy of Management Research Methods Division's electronic network

(RMNET) and the Statistical Consulting electronic network (STAT-L). E-mail voices were heard on both sides—pro and con. Though we are hesitant, from our admittedly biased vantage, to characterize the outcome of either debate, a majority of colleagues seemed to support our position. A sampling of opinions voiced on RMNET have been reproduced in the Research Methods Division's newsletter ("Ask the Experts," 1993).

So, what is the answer to our simple question? We obviously have an opinion, but also have an informed respect for the complexity of the underlying considerations. As with many statistical issues, the answer may depend on the exact context of the question. In an MMR context, we feel that hypothesizing (and examining) an interaction term is justified only to the degree that a sound theoretical case for doing so has been established a priori. For the case of a simultaneous regression model, we agree that requiring a significant overall  $R^2$  before examining individual effects is judicious.

In bringing the aforesaid correspondence to a conclusion, our colleague expressed pleasure in the opportunity to exchange views, but was not completely convinced. A concern of apparent interest to researchers other than the involved parties grew out of a simple but legitimate question. In closing, however, our colleague did offer a final thought: "Why not submit your views to the *Journal of Management's* Research Methods and Analysis section?" It is to be hoped, through this note, we have prompted others to consider what first appeared to be a simple question.

**Acknowledgment:** We thank David V. Day and Edward R. Kemery for discussions leading to this manuscript and Jacob Cohen, David A. Kenny, and Robert Rosenthal for their correspondence.

### Notes

1. The interpretation of significant main effects given a significant interaction effect has been a subject of debate (cf. Aiken & West, 1991); this debate is beyond the scope of this paper.
2. Keppel and Zedeck (1989) were the most expressive regarding this point. They state that the omnibus  $F$  test is "a single statistical test assessing either significance of the differences among all treatment means or the overall association between [treatment conditions]. This test does not tell us, however, which of these differences are significant and which are not. The omnibus  $F$  test evaluates what in effect is an average of all possible pairwise comparisons.

When we plan specific comparisons, we are not generally interested in the outcome of the omnibus test. Indeed, there is no logical need to conduct the test at all! With planned comparisons, our interest is in certain comparisons and not in an average of all pairwise differences. On the other hand, without specific comparisons (or research hypotheses) to guide the analysis—and specific comparisons may be lacking in certain exploratory work—we would probably conduct the omnibus test first and let the outcome of the test determine whether we examine the data in more detail."

"Most experiments in the behavioral sciences are designed to test specific hypotheses, however, and, in our opinion, should be evaluated directly, without reference to the omnibus test. We emphasize this point because one frequently encounters experiments in the research literature that report the result of the omnibus test first, followed by what are in effect planned comparisons. We suspect that in most cases the inclusion of the omnibus text is a habit the experimenter acquired when this two-step procedure was in common use" (pp. 148-149).

### References

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Ask the experts. (1993). *Academy of Management Research Methods Division Newsletter*, (Summer): 13-14.
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173-1182.
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/ correlation analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Govindarajan, V. & Fisher, J. (1990). Strategy, control systems, and resource sharing effects on business-unit performance. *Academy of Management Journal*, 33: 259-285.
- Gupta, A. K. & Govindarajan, V. (1986). Resource sharing among SBUs: Strategic antecedents and administrative implications. *Academy of Management Journal*, 29: 695-714.
- Keppel, G. & Zedeck, S. (1989). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches*. New York: Freeman.
- Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Wadsworth.
- McClelland, G. H. & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114: 376-390.
- McKinley, W., Cheng, J.L.C., & Schick, A. G. (1986). Perceptions of resource criticality in times of resource scarcity: The case of university resources. *Academy of Management Journal*, 29: 623-632.
- Simon, H. A. (1964). Approaching the theory of management. Pp. 77-85 in H. Koontz (Ed.), *Toward a unified theory of management*. New York: McGraw-Hill.
- Snell, S. S. & Dean, J. W., Jr. (1992). Integrated manufacturing and human resource management: A human capital perspective. *Academy of Management Journal*, 34: 776-804.
- Sutton, R., & Rafaeli, A. (1987). Characteristics of work stations as potential occupational stressors. *Academy of Management Journal*, 30: 260-276.
- U.S. Department of Agriculture. (1977). *Comparisons among treatment means in an analysis of variance* (Agricultural Research Service). Washington, DC: U.S. Government Printing Office.