Interaction, Nonlinearity, and Multicollinearity: Implications for Multiple Regression

Jose M. Cortina
Michigan State University

Moderated Hierarchical Multiple Regression (MHMR) is typically used to test for the presence of interactions. When an interaction term is composed of correlated variables, linearity and additivity become confounded. The result of this confounding is that an interaction term in MHMR may be statistically significant only because of its overlap with unmeasured nonlinear terms. I recommend that squared terms be used as covariates in such situations and show that the resulting loss of power with respect to the test of significance for the interaction term is limited to that associated with the loss of degrees of freedom and is therefore negligible if it exists at all.

As the social sciences have developed, hypothesized relationships have become more complex. One of the more obvious indicators of this increase in complexity is the increase in the number of moderating relationships that have been hypothesized over the past several decades (Althauser, 1971; Cohen, 1978; Cronbach, 1987). One of the most common methods for testing moderating relationships or interactions has been and is moderated hierarchical multiple regression (MHMR).

Although this method of analysis has a long history of criticism (e.g., Sockloff, 1976; Arnold, 1982; Morris, Sherman, & Mansfield, 1986), its proponents have generally repelled its critics (e.g., Cohen, 1978; Dunlap & Kemery, 1987; Stone & Hollenbeck, 1989). A recent article by Lubinski & Humphreys (1990)[1] has cast a new shadow on the MHMR technique. These authors pointed out that Type I error rates with respect to a product term in MHMR can be greatly inflated because of the overlap between the product term and the usually unmeasured nonlinear terms (i.e., the predictors taken to powers other than 1, −1, or 0). Although some authors have suggested that nonlinear relationships may be more prevalent than the body of psychological research would suggest, especially with respect to the relationship between measures and underlying psychological constructs (Busemeyer & Jones, 1983; Birnbaum, 1973; 1974), the possibility of nonlinear relationships continues to go relatively unexplored. For this reason, interpretation

---

of significant interaction terms in multiple regression may be difficult in certain cases. The purpose of the present article is to analyze carefully the implications and underpinnings of Lubinski & Humphreys (1990), especially in light of past research on the topic, and make recommendations for the future.

The Nature of Interaction

The conceptual meaning of interaction or moderation is well known. One tests for an interaction if one suspects that the effect of one variable, X, on another variable, Y, depends on the level of some third variable, Z. Common examples involve the relationships among ability, motivation, and performance (i.e., ability × motivation = performance) and ability, race, and performance (i.e., ability× race = performance). This possibility is typically tested by examining the variance explained by a product term, $X*Z$, over and above that explained by the constituent parts of the product term (i.e., X and Z). The classic equation is:

$$\hat{Y} = B_0 + B_1(X) + B_2(Z) + \qquad \text{[STEP 1]}$$
$$B_3(X*Z) \qquad \text{[STEP 2]} \qquad (1)$$

where $\hat{Y}$ = the predicted criterion variable, X and Z = linear predictor variables, $X*Z$ denotes the interaction between the linear predictors, and the various B's are structural parameters or regression weights. In other words, we examine the partial correlation between the product term and the criterion with the effects of the parts of the product term partialled out. Because of the presence of the multiplicative interaction term, Equation 1 represents a *multiplicative model*. If STEP 2 were dropped, the equation would represent an *additive* model.

The Nature Of Curvilinearity

A curvilinear effect is very similar in principle to the moderator effect (Cohen, 1978). Whereas the interaction term examines the extent to which the effect of X on $\hat{Y}$ depends on the level of Z, the curvilinear or polynomial term examines the extent to which the effect of X on Y depends on the level of X. The former says that the slope of the X, Y regression line is a function of the third variable Z. The latter says that the slope of the X, Y regression line is a function of X. The curvilinear hypothesis is also tested by examining the variance explained by a product term over and above that explained by individual predictors. One representation of such a test is:

$$\hat{Y} = B_0 + B_1(X) + \qquad \text{[STEP 1]}$$
$$B_2(X*X^a) \qquad \text{[STEP 2]} \qquad (2)$$

where 'a' is nonzero and, thus, $X*X^a$ represents X taken to a power other than 1. Because of the presence of STEP 2 in Equation 2, the equation is multiplicative. Because of the substitution of X^a for Z in STEP 2, Equation 2 is also *nonlinear*. One can easily see the similarity between a test of interaction and a test of curvi-

linearity using MHMR. Nevertheless, they are distinguishable *as long as X and Z are unrelated to each other.*

The Role Of Multicollinearity

When two variables, X and Z, are independent, that is, when $r_{xz} = 0$, the differences among X^2, Z^2, and XZ are clear. (This also applies to X^3, Z^4, etc.) X^2 and Z^2 are quadratic effects, that is, among other things, they are nonlinear. A term which takes a variable to any power other than 1, -1, or 0 is a nonlinear term.[2] This would include cubic terms, root terms, etc. All such terms are used to examine effects above and beyond linear effects. Because such terms involve multiplication, they are nonadditive terms and examine nonadditive effects above and beyond additive effects. Because they involve the multiplication of a variable with a nonzero exponential function of itself, such terms are also nonlinear. $X*Z$ is an interactive effect and is, therefore, also used to examine nonadditive effects. The difference between X^2 or Z^2 and $X*Z$ is that $X*Z$ examines nonadditive, *linear* effects[3] In fact, we can construct a quasi-factorial table[4] such as the following:

	LINEAR	NONLINEAR
ADDITIVE	X,Z	NULL
NONADDITIVE	$X*Z$	X^2,Z^2

The various components of this table are easily distinguished as long as X and Z are independent. Lubinski & Humphreys (1990) point out that, as r_{xz} departs from 0, $r_{(x*z)(x^2)}$ and $r_{(x*z)(z^2)}$ approach 1. In other words, to the extent that X and Z are related, linearity and additivity become confounded in the interaction term such that $X*Z$ is, to some extent, a measure of nonlinear effects. For example, consider the extreme case where the correlation between two standardized variables, X and Z, is 1. In this case, X^2, Z^2, and $X*Z$ are identical. For this reason, a statistically significant interaction term is significant because of a nonlinear multiplicative effect and not because of a linear multiplicative effect. This problem exists in a given test to the extent that X and Z are related.

Why Is This Important?

Our move toward more complex explanations of psychological phenomena could have taken and can still take many different directions. One of these directions has been the investigation of moderator effects. In the 1991 and 1992 volumes of *Journal of Applied Psychology*, no fewer than 123 significant interactions were reported (using MHMR) in a total of 14 different papers. On the other hand, only one of the papers reporting significant interactions also investigated nonlinear trends (Rice, Gentile, & McFarlin, 1991). Our bias against nonlinear hypotheses is evident in virtually every area of applied psychology. For example, Vroom's Expectancy Theory (V*I*E), the Motivating Potential Score of Hackman

& Oldham (1975; [Variety + Identity + Significance)*Autonomy*Feedback]), and many other theories posit complex, nonadditive models without consideration of possible nonlinear effects. While there is nothing wrong with nonadditive linear hypotheses per se, tests of such hypotheses that neglect possible nonlinear trends can be misleading (Cohen, 1978; Birnbaum, 1973; Busemeyer & Jones, 1983; Lubinski & Humphreys, 1990). While many of the problems associated with multicollinearity have been documented (e.g., Althauser, 1971; Sockloff, 1976; Haitovsky, 1969; Zedeck, 1971), this particular problem of confounded interaction terms has been largely ignored. The result of this neglect is unknown. If, however, the analyses of Lubinski & Humphreys (1990) are any indication of the actual importance of nonlinear effects, it is entirely possible that some of our significant results with respect to moderator effects are artifactual. A review of the relevant[5] papers published in the *Journal of Applied Psychology* for the years 1991 and 1992 showed a mean absolute value of the correlation between componenets of significant interaction terms of .21 with values ranging from 0 to .68. For those papers with the larger correlations between predictors, the degree of multicollinearity may be great enough to call into question the interpretation of the interaction terms. Specifically, the significance of their interaction term regression weights may be due only to the overlap between the interaction terms and untested, but significant, nonlinear trends and not to an actual interaction between the variables.

What Is The Solution?

The solution that I present may seem conservative. If one wishes to control for possible nonlinear effects and thus rule out alternative explanations for findings, then one should treat them as covariates. In the case of MHMR, this would mean entering nonlinear terms such as X^2 and Z^2 into regression equations *before* entering interaction terms, thus yielding an equation such as the following

$$\hat{Y} = B_0 + B_1(X) + B_2(Z) + \quad \text{[STEP 1]}$$
$$B_3(X*X^a) + B_4(Z*Z^b) + \quad \text{[STEP 2]}$$
$$B_5(X*Z) \quad\quad\quad\quad\quad\quad \text{[STEP 3]} \quad\quad (3)$$

where 'a' and 'b' are not equal to zero. The reason I say that this solution is conservative is that it involves the addition of terms to the equation that must be partialled out before the assessment of the interaction term. The lack of power associated with the MHMR test of interaction terms is a frequently (and justifiably) cited limitation (Bobko, 1986; Cronbach, 1987; Morris et al., 1986). This addition of terms prior to the assessment of the interaction term is no help to the power problem. However, it may not be as problematic as one might think.

A certain amount of power with respect to a given predictor is always lost when terms are added to a regression equation because of the corresponding loss in degrees of freedom. With the inclusion of nonlinear terms, however, this is the only source of power loss. The linear variance in polynomial terms is removed by the partialling of X and Z in Step 1, leaving only nonlinear variance. Since the interaction term (in the behavioral sciences) almost always deals only with linear

effects, the only power lost by the inclusion of the nonlinear terms in Step 2 of Equation 3 (other than that associated with degrees of freedom) comes from removing the nonlinear variance in $X*Z$, which is artifactual. So, the inclusion of nonlinear terms in a MHMR equation *will not* result in a substantial loss of power with respect to the interaction term.

To reiterate, the only power loss that results from the inclusion of squared terms as suggested above is that associated with the change in degrees of freedom. As an example, consider the following. If I wished to conduct the standard test for interaction using MHMR, I would use Equation 1. The power associated with Step 2 in Equation 1 can be represented as the power associated with the term $R^2_{y.x,z,xz}$ $- R^2_{y.x,z}$ (Cohen, 1977, p.409). If I were to use Equation 3 instead, I would need the power associated with the term $R^2_{y.x,z,x^2,z^2,xz} - R^2_{y.x,z,x^2,z^2}$. Using Cohen's power tables, the power, for a given level of significance, associated with these types of situations (i.e. Case 1 situations) is the value that corresponds to the values of "u" and "L". "u" is the number of variables for which the increment in R^2 is relevant. In both instances, the only variable that is relevant in this sense is XZ, so "u" is equal to 1 in both cases. Therefore, the only difference in power between these two instances must come from differences in the "L" value.

"L" is given by:

$$L = \frac{R^2_{Y.A,B} - R^2_{Y.A}}{1 - R^2_{Y.A,B}} \times (N - u - w - 1) \tag{1}$$

where "A" is the set of variables that is entered before the interaction term is entered, "B" is the interaction term, N is the sample size, u is as described above, and w is the number of variables entered before u (i.e. the number of variables in A). Clearly, the only possible sources of difference in "L" between the two methods are the values of $R^2_{Y.A,B}$, $R^2_{Y.A}$, and "w". Let us first examine the two R^2 terms.

The difference between the two equations in these R^2 terms lies in the difference between the "A" set of variables. In Equation 1, A is composed only of X and Z. In Equation 3, A is composed of X, Z, X^2, and Z^2. As was mentioned before, however, there is no difference in the *linear* variance described by the R^2's in the two equations. The only difference between them is in terms of explained, nonlinear variance, which is usually not considered to be of interest. So, any difference between Equation 1 and Equation 3 in terms of $R^2_{Y.A,B}$ and $R^2_{Y.A}$ is irrelevant for purposes of power analysis.

This leaves us with the difference in the value of "w". In Equation 1, $w = 2$ (X and Z). In Equation 3, $w = 4$ (X, Z, X^2, and Z^2). Hence, the difference in "L", and therefore the difference in power, lies in the difference in the value of "w", which is 2. As might be expected, this leads to a very small difference in power. For example, suppose the difference between $R^2_{Y.A,B}$ and $R^2_{Y.A}$ is .02 because the former is .15 and the latter is .13 (a small effect). Suppose further that we have a sample size of 200 and a significance level of .05 . The value of "L" for Step 2 of

Equation 1 is 3.92. The value of "L" for Step 3 of Equation 3 is 3.88. Since "u" is equal to 1 in both instances, the power estimate for the former is .51 while the power estimate for the latter is .50. This is clearly a very small difference, and one that gets smaller as N increases. If the sacrifice associated with treating nonlinear terms as covariates is so small, then it would seem that any possibility of nonlinearity, no matter how small, would warrant the use of this procedure.

One question that may arise at this point is, Why should we enter the nonlinear term (or any other term) before the interaction term? After all, it is the interaction that we are interested in and that we have reason to believe exists. This is a reasonable question and deserves an answer. The issue of which predictor to enter first only arises when the predictors overlap with each other *and when that overlap is important to the individual predictors themselves.* For example, there is considerable overlap between an interaction term, $X*Z$, and its constituent parts, X and Z. It has been suggested that, when there is sufficient theoretical evidence to do so, an interaction term should be tested without partialling the effects associated with its components (Bobko, 1986). The reason this is suggested is essentially that there is always a considerable correlation between each of two variables, X and Z, and their interaction term, $X*Z$ (i.e., there is overlap), and that the partialling out of X and Z prior to the test for the interaction necessarily removes a large portion of linear variance accounted for by the interaction even if the variance is actually accounted for by the interaction term and not X and Z. In this case, the overlap between the interaction term and its parts is *important* to both. What is suggested in the present research article is that the overlap between polynomial terms and interaction terms is *not* important to the assessment of the interaction term. The reason is that the only overlap involves the nonlinear variance that they share, which is not of interest and, therefore, not important to the interaction term.

The question that remains with respect to the treatment of nonlinear terms as covariates is: Which nonlinear term or terms should be used? There are an infinite number of possibilities (i.e., X^2, X^3, $X^{3.5}$, etc.). My response is that one can feel safe in using only squared terms. There are two reasons for avoiding exponents greater than two. First, psychological phenomena rarely display anything more complex than a quadratic trend (Cohen & Cohen, 1983, p. 225). Second, measurement error becomes so great as imperfectly measured components are added to multiplicative terms that variables taken to powers greater than two may be comprised of measurement error and little else (Busemeyer & Jones, 1983; Lubinski & Humphreys, 1990).

If squared terms (in this group I include any variable taken to a power greater than 1 and less than or equal to 2) are used as covariates, they will identify both concave upward relationships (positive regression weight) and concave downward relationships, negative regression weight(Cohen & Cohen, 1983). For example, if one were examining the effect of test scores and job experience on performance, and one had reason to believe, perhaps because of scaling fidelity issues, that there would be a ceiling with respect to test scores (or performance), then a squared function of test scores would be a plausible covariate and might be expected to have a significant, negative regression weight (i.e., concave downward). If, on the

other hand, a synergistic effect such as that hypothesized by Lubinski and Humphreys (1990) is plausible, then a squared function of the predictor might be expected to have a positive weight. Either way, only the power associated with the change in degrees of freedom is sacrificed, and a more conclusive test of interaction is achieved.

Two final points regarding the solution to this problem of spurious interactions are in order. First, the problem, as I mention above, arises when the correlation between two predictors, X and Z, is nonzero. An obvious solution[6] is to make X and Z independent. In other words, the problem does not exist if the levels of both X and Z are manipulated. Often this is not possible or inappropriate. When the levels of two variables are experimentally manipulated, however, they are statistically independent of each other, and there need not be any concern about spurious detection of moderators.

The second point is that the same logic that I have used to argue for the inclusion of nonlinear terms as covariates when testing for interactions can be used to argue that interaction terms should be entered as covariates when testing for nonlinear effects. Also, since the interaction term, being a linear covariate, would remove none of the nonlinear variance, there would be very little if any loss in power with respect to the nonlinear term. It could be argued, however, that treating nonlinear terms as covariates when testing for interactions is more justifiable than the reverse because a nonlinear relationship represents a more parsimonious explanation than a moderating relationship. The meaning of parsimony is, however, debatable.

The Next Step

The most pressing need with respect to these issues is a determination of the extent to which this confounding of additivity and linearity *has* misled us into concluding that relationships which are actually nonlinear are interactive. Such an assessment could involve two parts. First, it could involve a reanalysis of published data using relevant polynomial covariates in an attempt to change or support the conclusions that were drawn. Second, such an assessment could involve a comparison, perhaps in a Monte Carlo format, of the tests of interaction terms in ANOVA and regression. Because ANOVA, in effect, partials both the linear and nonlinear variance associated with predictors before assessing interaction terms (Winer, 1971; Hays, 1981), regression (without the recommendations of the present paper) should be more "powerful", for spurious reasons, in detecting moderator effects. The Monte Carlo format could also be used to determine whether or not there is a level of multicollinearity below which we needn't worry about nonlinear confounds.

Another pressing need that is brought to light by the present paper is the need of our field to more closely examine the possibilty of nonlinear relationships in general. Although nonlinearity has received some recent attention (e.g., Russell & Bobko, 1992; Lubinski & Humphreys, 1990), very little work has been done in this area. Perhaps the most troubling (and intriguing) suggestion has been that the very relationship between observed and latent variables is nonlinear (Russell & Bobko,

1992; Busemeyer & Jones, 1983). If such a fundamental relationship is indeed nonlinear, other possible nonlinear relationships must, at the very least, be considered. In any case, we may be forced to reconsider more of our conclusions than we would like.

Notes

1. Lubinski and Humphrey's article is, to some extent based on Busemeyer and Jones (1983).
2. For all practical purposes, X is just a particular point on the linearity continuum where the power to which the variable is taken is equal to 1.
3. Although interaction terms certainly can examine nonlinear effects, such as the interaction term X^2*Z, this is seldom if ever done in the behavioral sciences.
4. The top right-hand box of the table is empty because, by definition, a nonlinear term is nonadditive.
5. This was a review of those papers which published significant interaction effects using MHMR and involving at least one nondichotomous variable with at least ordinal properties.
6. Although I say that this is an obvious solution, it did not occur to me, but rather was suggested by a reviewer.

References

Althauser, R. P. (1971). Multicollinearity and non-additive regression models. In H. M. Blalock, Jr. (Ed.), *Causal models in the social sciences.* Chicago: Aldine-Atherton.

Arnold, H. J. (1982). Moderator variables: A clarification of conceptual, analytic, and psychometric issues. *Organizational Behavior and Human Performance, 29:* 143-174.

Birnbaum, M. H. (1973). The devil rides again: Correltations as an index of fit. *Psychological Bulletin, 79:* 239-242.

———. (1974). Reply to the devil's advocates: Don't confound model testing with measurement. *Psychological Bulletin, 81:* 854-859.

Bobko, P. (1986). A solution to some dilemmas when testing hypotheses about ordinal interactions. *Journal of Applied Psychology, 71:* 323-326.

Busemeyer, J. R. & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin, 93:* 549-562.

Cohen, J. (1977). *Statistical power analysis for behavioral sciences.* Hillsdale, NJ: Erlbaum.

———. (1978). Partialled products are interactions: Partialled powers are curve components. *Psychological Bulletin, 85:* 858-866.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102:* 414-417.

Dunlap, W. P. & Kemery, E. R. (1987). Failure to detect moderating effects: Is multicollinearity the problem? *Psychological Bulletin, 102:* 418-420.

Haitovsky, Y. (1969). Multicollinearity in regression analysis: Comment. *The Review of Economics and Statistics, 51:* 486-489.

Hays, W. L. (1981). *Statistics,* 3rd ed. New York: Holt, Rinehart, & Winston.

Lubinski, D. & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin, 107:* 385-393.

Morris, J. H., Sherman, J. D., & Mansfield, E. R. (1986). Failures to detect moderating effects with ordinary least squares-moderated multiple regression: Some reasons and remedy. *Psychological Bulletin, 99:* 282-288.

Rice, R. W., Gentile, D. A., & McFarlin, D. B. (1991). Facet importance and job satisfaction. *Journal of Applied Psychology, 76:* 31-39.

Russell, C. J. & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. *Journal of Applied Psychology, 77:* 336-342.

Sockloff, A. L. (1976). The analysis of nonlinearity via regression with polynomial and product variables: An examination. *Review of Educational Research, 46:* 267-291.

Stone, E. F. & Hollenbeck, J. R. (1989). Clarifying some controversial issues surrounding statistical procedures for detecting moderator variables: Empirical evidence and related matters. *Journal of Applied Psychology, 74:* 3-10.

Winer, B. J. (1971). *Statistical principles in experimental design,* 2nd ed. New York: McGraw-Hill.

Zedeck, S. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin, 76:* 295-310.