Improving the Power of Moderated Multiple Regression to Estimate Interaction Effects[1]

HERMAN AGUINIS
Graduate School of Business Administration
University of Colorado at Denver
http://www.cudenver.edu/~haguinis

CHARLES A. PIERCE
Department of Psychology
Montana State University
http://www.montana.edu/wwwpy/cppage.html

We provide a brief review of moderated multiple regression (MMR) and factors affecting its statistical power. We conclude that, given the low statistical power of MMR, numerous hypothesized interaction effects in organizational science research may have been incorrectly discarded. Drawing erroneous conclusions regarding interaction effects is an important methodological issue that can severely affect the advancement of organizational science.

A variable Z is a moderator of the relationship between variables X and Y when this relationship depends on the value of Z. For instance, motivation is a moderator of the relationship between general cognitive abilities and job performance when this relationship is stronger for employees with higher levels of motivation and weaker for employees with lower motivation. In other words, X and Z interact in affecting Y. In this brief review, we (a) describe moderated multiple regression (MMR) as a technique to estimate interaction effects, (b) note that MMR is one of the most common statistical techniques used to estimate interaction effects, (c) provide a selective review of factors that affect the power of MMR, and (d) offer recommendations regarding various courses of action researchers can take to address MMR's typically inadequate power.

Moderated Multiple Regression (MMR)

Using multiple regression to estimate the effect of a moderator variable Z on the X-Y relationship involves a regression equation that includes Y as a criterion, and X and Z as predictors. In addition, the MMR equation includes a third predictor consisting of the $X \cdot Z$ product. This product term carries information regarding the X by Z interaction (i.e., moderating effect of Z). The MMR equation is the following:

$$\hat{Y} = a + b_1 X + b_2 Z + b_3 X \cdot Z, \qquad (1)$$

where is the predicted value for Y, a is the least squares estimate of the intercept of the surface of best fit, b_1 is the least squares estimate of the population regression coefficient for X, b_2 is the least squares estimate of the population regression coefficient for Z, and b_3 is the least squares estimate of the population regression coefficient for the product term which carries information about the interaction between X and Z (Cohen & Cohen, 1983). Rejecting the null hypothesis that $_3$ (i.e., b_3's population value) $= 0$ indicates the presence of a moderating or interaction effect. Stated differently, rejecting this null hypothesis indicates that the regression of Y on X is unequal across values of Z.

In addition to the regression coefficient associated with the product term, another procedure used to assess the presence of the X by Z interaction is to compute the difference between the multiple correlation coefficient associated with Equation 1 minus the multiple correlation coefficient associated with the same model excluding the product term. The resulting R^2 indicates whether the moderating effect of Z adds explained variance in Y to the model including the main effects of X and Z only.

An MMR analysis can be easily conducted using SPSS, SAS, Minitab, and other commercially-available software packages. A researcher first creates a new variable consisting of the product between X and Z, and then executes the regression procedure. Note that all three predictors should be forced simultaneously in the equation or, alternatively, X and Z should be forced in first followed by $X \cdot Z$ (Aiken & West, 1991; Jaccard, Turrisi, & Wan, 1990; Stone & Hollenbeck, 1989).

Pervasive Use of MMR

MMR seems to be a method of choice for estimating moderating effects in organizational science. Aguinis, Petersen, and Pierce (1999) reviewed articles published in *Academy of Management Journal*, *Journal of Applied Psychology*, and *Personnel Psychology* between 1987 and 1998. This review focused only on articles that reported MMR analyses pertaining to categorical moderator variables (e.g., gender, ethnicity). This twelve-year review revealed that 87 articles published in these outlets reported results of MMR analyses. In addition, each of these 87 articles reported multiple MMR tests. Thus, the Aguinis et al. (1999) review showed that MMR is pervasively used in articles published in three of the most influential journals in organizational science.

Problems Affecting Substantive Research Conclusions Based on MMR

Despite the fact that MMR has been available for virtually half a century and, as described above, is frequently used in organizational science, Monte Carlo simulation results suggest that substantive research conclusions based on MMR may suffer from a serious problem. More precisely, MMR analyses are typically conducted at low levels of statistical power (see Aguinis, 1995, for a review). In practical terms, low power affects substantive research conclusions in that a researcher may incorrectly conclude that the data in hand do not support a hypothesized moderating effect. However, this sample-based conclusion may be incorrect. In fact, the hypothesized moderating effect may be present in the population. Next, we review factors that adversely affect the power of MMR and, consequently, may cause researchers to fail to detect existing interaction effects.

Predictor Variable Variance Reduction

The power of MMR is reduced markedly when the variance of X is smaller in the sample than in the population (Aguinis & Stone-Romero, 1997). This reduction in variance is typical in research conducted in field settings. For example, personnel selection procedures are a major cause of variance reduction in validation research. Decisions regarding which individuals to select for an opening are frequently based on their standing on a predictor variable X (e.g., test of job aptitude); only those who obtain a score that exceeds a specific cutoff point are selected, leading to an X variance in the sample that is smaller than the X variance in the population. Moreover, although Aguinis and Stone-Romero investigated direct range restriction, indirect range restriction is also pervasive in such fields as human resources management (Aguinis & Whitehead, 1997).

Aguinis and Stone-Romero's (1997) Monte Carlo study revealed that a mild ratio of sample to population variance of .80 can have substantial effects on power. Thus, although it may not be feasible in many instances, researchers are advised to not reduce predictor variance in their samples. Alternatively, researchers are advised to obtain information regarding the

extent to which sample-based variances are reduced. A ratio of sample to population variance can be used in computer programs described later in the paper to estimate the resulting statistical power.

Artificial Polychotomization of a Continuous Variable

A special case of predictor variable variance reduction occurs when a continuous variable (e.g., age) is artificially polychotomized (e.g., into a subgroup under 40 and another of 40 and over). Stone-Romero and Anderson (1994) demonstrated that artificial dichotomization of a continuous predictor also leads to substantial power loss in tests of interaction effects. Thus, researchers should not use procedures such as a median split to artificially create subgroups based on a variable measured using a continuous scale.

Scale Coarseness

Russell and Bobko (1992) investigated an issue referred to as criterion variable "scale coarseness." This phenomenon refers to the operationalization of a criterion variable that does not include sufficient scale points. This insufficient number of scale points results in possible information loss and, therefore, prevents an interaction effect from being detected. For instance, if the predictor X and hypothesized moderator Z are measured on 5-point Likert-type scales, the product term $X \cdot Z$ has a possible range of $5 \cdot 5 = 25$ distinct responses. However, if Y is measured on a "coarse" 5-point scale (which is typically the case) rather than on a 25-point scale (which is typically never the case), information regarding the relationship between Y and $X \cdot Z$ is lost, the population interaction effect is underestimated, and power drops inevitably. A way to address the problem of scale coarseness is to use a program by Aguinis, Bommer, and Pierce (1996), which prompts users to provide responses by clicking on a graphic line segment displayed on a computer monitor. Thus, this program overcomes the scale coarseness problem by allowing researchers to gather data using a continuous, as opposed to a coarse, criterion scale.

Issues Pertaining to Categorical Moderator Variables

The estimation of moderating effects of categorical variables faces yet another set of challenges. Most notably, heterogeneous sample sizes and error variances across the moderator-based subgroups (e.g., men and women) can have profound effects on power.

Sample size heterogeneity. When MMR analyses are conducted on categorical moderator variables (e.g, gender, ethnicity), it may be the case that there are unequal sample sizes across the levels of Z (e.g., more Whites than Latinos and African Americans). As a consequence of this situation, which is typical in many management subdisciplines, the power to detect ethnicity or gender as a moderator variable is reduced.

An empirical examination of this issue using Monte Carlo simulations (Stone-Romero, Alliger, & Aguinis, 1994) demonstrated the effect of unequal sample sizes across moderator-based subgroups on the power of MMR. In situations with two subgroups, results showed that there was a considerable decrease in power when the size of Subgroup 1 was .10 relative to total sample size regardless of total sample size (30, 60, 180, 300) and size of the moderating effect in the population (small, medium, large). The effect of unequal subgroup proportions on statistical power was significant above and beyond the effect of total sample size. A proportion of .30, closer to the optimum value of .50, also reduced the statistical power of MMR, but to a lesser extent. Based on these results, researchers should strive to obtain similar sample sizes across subgroups. However, note that gathering similar sample sizes across subgroups when they actually differ in the population increases statistical power but also limits the generalizability of results (Aguinis, 1995).

Error variance heterogeneity. Similar to ANOVA, MMR assumes that the variance in Y that remains after predicting Y from X is equal across k moderator-based subgroups (see Aguinis & Pierce, 1998a, for a review). Violating the homogeneity of error variance assumption has been identified as a factor that can affect the power of MMR to detect categorical moderator variables. In each subgroup, the error variance is estimated by the mean square residual from the regression of Y on X:

$$\sigma_{e_i}^2 = \sigma_{Y(i)} \, (1 - \rho_{XY(i)}^2), \qquad \qquad (2)$$

where $\sigma_{Y(i)}$ and $\sigma_{XY(i)}$ are the Y standard deviation and the X-Y correlation in each subgroup, respectively. In the presence of a moderating effect in the population, the X-Y correlations for the two moderator-based subgroups differ and, thus, the error terms necessarily differ.

Heterogenous error variances can affect both Type I error and statistical power. However, Alexander and DeShon (1994) ascertained that when the subgroup with the larger sample size is associated with the larger error variance (i.e., the smaller X-Y correlation), statistical power is lowered markedly. As noted in Aguinis and Pierce's (1998a) review, this specific scenario in which the subgroup with the larger n is paired with the smaller correlation coefficient is the most typical situation in validation research in a variety of organizational settings. To address the homogeneity of error variance assumption issue, we suggest that MMR users implement a program by Aguinis et al. (1999) to check whether the assumption is violated and compute alternative statistics to MMR's F-test if the assumption is violated.

Conclusions

Given this brief review, it is evident that numerous factors affect negatively the power of MMR to detect moderating effects. In fact, simulations (e.g., Aguinis & Stone-Romero, 1997) show that MMR results published in some of the top journals in organizational science were conducted at power levels of .30 and even lower. Thus, on the one hand, we are not bearers of good news: Many conclusions of previously conducted, and published, research using MMR may be incorrect. On the other hand, we bring a message of optimism: Many unsupported hypotheses regarding interaction effects may in fact receive support if an MMR re-analysis were conducted with adequate statistical power.

We suggest that researchers implement programs to calculate the power of MMR in every instance when a hypothesized moderating effect is not found. Such programs are in the public domain and descriptions, as well as instructions on how to obtain them, can be found in Aguinis, Stone-Romero, and Pierce (1994), Aguinis and Pierce (1998b), and Aguinis, Boik, and Pierce (1998). In addition, we recommend that researchers use a program described by Aguinis et al. (1999) to compute alternatives to MMR's F-test when the homogeneity of error assumption is violated (Aguinis et al., 1999, also provided instructions on how to obtain this program).

In closing, moderating effects are at the core of the scientific enterprise (Aguinis & Pierce, 1998c), and knowledge about moderator variables can be used as an index of the maturity of organizational science as a scientific discipline (Hall & Rosenthal, 1991). We hope the present article will raise awareness regarding the low power of MMR and encourage researchers to consider low power as a possible explanation for a lack of support for hypothesized moderating effects.

Footnotes

[1]We thank Steven Farmer (Wichita State University) and Kurt Kraiger (University of Colorado at Denver) for comments on previous drafts. Correspondence regarding this article should be addressed to Herman Aguinis, Graduate School of Business Administration, University of Colorado at Denver, Campus Box 165, P.O. Box 173364, Denver, CO 80217-3364. E-mail: haguinis@castle.cudenver.edu

References

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management, 21*, 1141-1158.

Aguinis, H., Boik, R. J., & Pierce, C. A. (1998, April). Estimating the statistical power of differential prediction analysis. In J. M. Cortina (Chair), *Greasing the wicket: Addressing some sticky issues in modern data analysis.* Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Aguinis, H., Bommer, W. H., & Pierce, C. A. (1996). Improving the estimation of moderating effects by using computer-administered questionnaires. *Educational and Psychological Measurement, 56,* 1043-1047.

Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). *Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables.* Unpublished manuscript.

Aguinis, H., & Pierce, C. A. (1998a). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1,* 296-314.

Aguinis, H., & Pierce, C. A. (1998b). Statistical power computations for detecting dichotomous moderator variables with moderated multiple regression. *Educational and Psychological Measurement, 58,* 668-676.

Aguinis, H., & Pierce, C. A. (1998c). Testing moderator variable hypotheses meta-analytically. *Journal of Management, 24*, 577-592.

Aguinis, H., Pierce, C. A., & Stone-Romero, E. F. (1994). Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement, 54,* 690-692.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192-206.

Aguinis, H., & Whitehead, R. (1997). Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology, 82,* 528-538.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage

Alexander, R. A., & DeShon, R. P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin, 115*, 308-314.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs, 58,* 437-448.

Jaccard, J. J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-072. Newbury Park, CA: Sage.

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

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.

Stone-Romero, E. F., Alliger, G. M., & Aguinis, H. (1994). Type II error problems in the use of moderated multiple regression for the detection of moderating effects of dichotomous variables. *Journal of Management, 20,* 167-178.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology, 79,* 354-359.

[back to top]