INTRODUCTION AND BACKGROUND

My goal is to correctly discuss an interaction term before I die.

—Confidential Dissertator (ca. 2000)

OVERVIEW: WHY SHOULD YOU READ THIS BOOK?

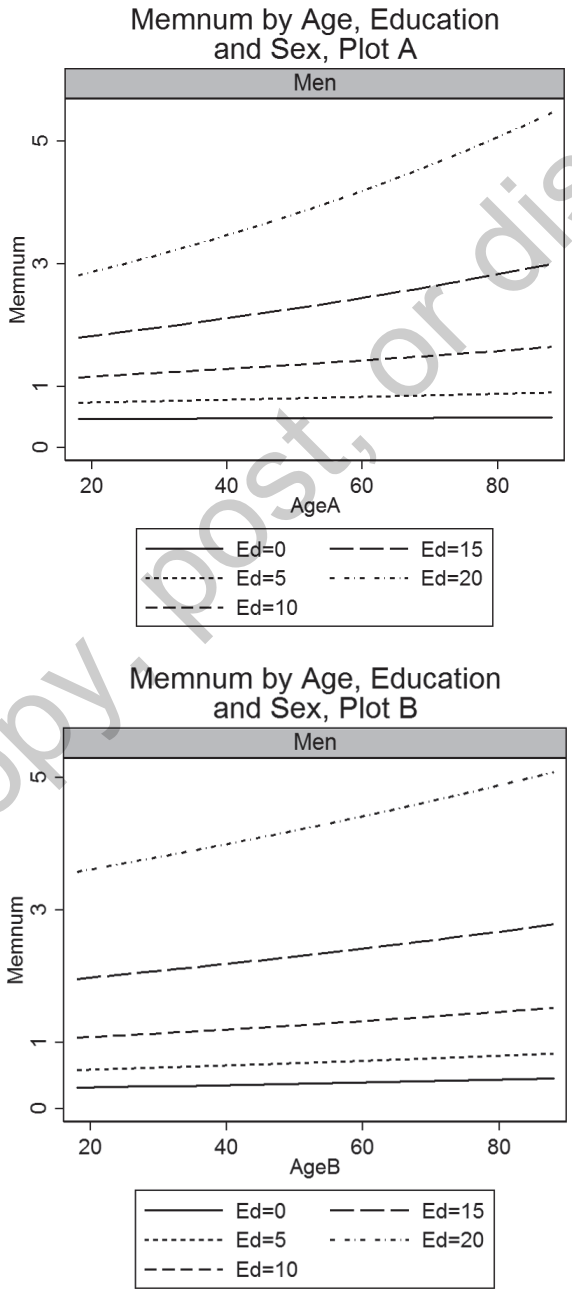
The inevitable question that the author of a statistics book like this has to address is whether another book on interaction effects is necessary—why a reader should read it—given that analyses incorporating both simple and more complicated interaction effects are commonplace. My first answer is embodied in the chapter quote above, which is as true today as it was nearly 20 years ago. Many graduate students as well as post-PhD researchers continue to have difficulty properly testing, interpreting, or specifying interaction effects in linear regression, let alone for nonlinear regression techniques.

This is evident in the apparent short life cycle of publications on how to interpret interaction effects properly. Every 5 to 10 years, there is a renewed call for researchers to follow best practices to avoid common problems. Each explicitly argues that these points are important to continue to reiterate because they still do not consistently inform actual practice (e.g., Aiken & West, 1991; Brambor, Clark, & Golder, 2006; Braumoeller, 2004; Dawson, 2014; Hayes, Glynn, & Huge, 2012; Jaccard, 1998; Jaccard & Turisi, 2003; Kam & Franzese, 2007; Southwood, 1978). When quantitatively oriented colleagues and graduate students learn that I am writing a book on the interpretation of interaction effects, their typical reaction is along the lines of "Great. My students could really use it and so could I. When will it be out so we can read it?"

Second, I identify and discuss solutions to a little recognized problem with tabular and graphical presentations of predicted outcomes derived from analyses using many types of generalized linear models (GLMs). Look at the two plots in Figure 1.1 from a negative binomial regression (a count model) predicting the number of voluntary associations in which the respondent is a member. The only difference is that one model contains only linear terms for age, education, and sex (main effects) and the

other also includes product terms between every pair of these predictors and a product of *Age* × *Education* × *Sex*—that is, a three-way interaction. Just from examining the plots, can you convince yourself which plot shows the main effect model's predictions and which portrays the interactive effect model's predictions? Is it Plot B

FIGURE 1.1 ⬢ PREDICTED VALUE PLOTS FROM MAIN EFFECTS AND INTERACTIVE EFFECTS MODELS



showing interaction effects in which education mutes the effect of age for men? Or is it Plot A representing education as enhancing the effect of age for men?

Moreover, if you were given only the main effects plot (whichever it is), can you honestly say you would identify it as representing a main effects model and not an interactive effects model? My point is that it can be difficult to visually distinguish a graph from a model with an interaction effect from one without an interaction effect in nonlinear models. And this hampers an analyst's ability to interpret visual or tabular displays of interaction effects effectively as well as a reader's ability to understand what is presented. The problem is that a plot or a table of predicted values incorporates and portrays two sources of nonlinearity simultaneously: the nonlinearity of the interaction effect and the inherent nonlinearity of many GLMs in how they model the relationship between the outcome and the predictors. This issue and recommended solutions are the topic of Chapter 5.

Moreover, four limitations of the current didactic literature (Aiken & West, 1991; Brambor et al., 2006; Braumoeller, 2004; Dawson, 2014; Hayes, 2013; Hayes et al., 2012; Jaccard, 2001; Jaccard & Turisi, 2003; Jaccard & Wan, 1996; Kam & Franzese, 2007; Southwood, 1978) motivated me to write this book and shaped its content to overcome and avoid these problems. Specifically, existing treatments are subject to at least one and usually more of the following shortcomings:

- They address a limited range of analytic models—often a single technique, most commonly ordinary least squares (OLS) regression.
- They provide details and examples only for the simplest interaction effect—a focal variable with a single moderator (often a two-category nominal predictor)—leaving readers to extend the approach to more complicated interaction effects themselves.
- They do not cover a wide range of tools for interpreting interaction effects.
- They provide limited, if any, assistance, with rare exceptions, for automating the calculations needed for many of the tools for interpreting interaction effects (i.e., software code/programs or spreadsheet-friendly formulas). And if they do so, it is specific to a particular technique of analysis.

These limitations result in piecemeal knowledge in which practitioners learn how to use some interpretive and calculating tools for one technique but different ones for another. Moreover, there is considerable (wasted) effort as researchers reinvent the wheel by creating their own specialized programs or spreadsheets for doing calculations and creating tables and graphics.

Consequently, my goal in writing this book is to provide a unified approach for interpreting interaction effects, which is more comprehensive in its coverage of interpretive tools as well as applicable across a wide range of techniques of analysis. I have also created a set of Stata routines (ado files) named ICALC—for Interaction CALCulator—to apply the interpretive tools I discuss in this book. ICALC can be downloaded free of charge at **www.icalcrlk.com** as can the data sets and Stata syntax (do-files) for all the book's examples. Readers may find it helpful to follow along in the do-files as they read the application examples. I chose Stata for this platform primarily because I consider this book an intellectual companion to Long and Freese's (2014)

Stata is a registered trademark of StataCorp LLC.

book on interpretive techniques for common nonlinear models. And ICALC uses their SPOST13 suite of tools written for Stata for many of its calculations. Moreover, Stata is, by most indicators, one of the top four package platforms for data analysis, and unlike IBM® SPSS® Statistics and SAS®, its popularity is growing, not declining (Muenchen, 2017). You can download and install the ICALC Toolkit when you are running Stata with online access. In the Stata Command window, type *net search icalc*. The Stata Results window should show four packages. Click the link for *icalc_ado* to install the ICALC program and help files in your PERSONAL directory. Use the same process if you need to install SPOST13, which I would highly recommend. To provide readers the flexibility to apply these tools on other platforms, I strive to present sufficient mathematical details for and application examples of the underlying formulas to enable readers to write their own spreadsheet formulas or software code for the platform they use. Furthermore, ICALC automatically stores graphics during a session as editable-in-Stata graphs (memory graphs) and provides options to save numeric results, tables, and the underlying data used to construct graphics in an Excel spreadsheet. These features give users the flexibility to customize graphics in Stata or to use the saved data to create their own graphics using other platforms.

In the next several sections, I review basic background material about interaction effects, GLMs, and relevant statistical and diagnostic tests. I would recommend that readers well-versed in these topics at least skim through this material to ensure that nothing essential is missed. In particular, the section on confounded nonlinearities in GLMs is likely an unfamiliar issue. I conclude the chapter with a roadmap of the content and organization of the rest of the book.

THE LOGIC OF INTERACTION EFFECTS IN LINEAR REGRESSION MODELS

Let me start by answering four basic questions about interaction effects: (1) What is an interaction effect? (2) Why should you consider including an interaction effect in your analysis? (3) How do you specify an interaction effect in the prediction function of a linear regression model? (4) When is an interaction effect statistically significant?

What Is an Interaction Effect?

Conceptually, an interaction effect is a way of specifying that the relationship between the outcome and a first predictor, call it F, is contingent on the values of another predictor, call it M_1. Or, to put it a different way, the effect of F on the outcome varies with the values of M_1. For example, an economist might argue that the earnings return to work experience is greater for a worker with more education than for a worker with less education. That is, the earnings–experience relationship is different depending on the level of a worker's education; a worker with more years of education would receive a higher payoff to his or her work experience. Or a legal scholar might argue that the effect of education on approval of a legal ban against racial intermarriage differs by race. Specifically, education is more consequential for Whites' approval of a legal ban against racial intermarriage—it reduces their approval more—than it is for Blacks' approval.

The first predictor, F, is often labeled the focal variable in the interaction, and the second predictor, M_1, is often referred to as the moderator or the moderating variable. This corresponds to the fact that interaction effects are frequently developed with a primary theoretical or conceptual focus on one of the predictor's effects and how those effects

are moderated by the second predictor. For this reason, interaction effects are sometimes called moderated effects (e.g., Hayes, 2013; Jaccard, 1998; Jaccard & Turisi, 2003).

This is a very useful heuristic device that I also adopt throughout the book, but the roles of moderating and focal variables should not be reified. They are arbitrary from a statistical point of view because the contingency of the *Outcome*–*F*–M_1 relationship works both ways. That is, specifying that the effect of *F* is contingent on M_1 also specifies that the effect of M_1 is contingent on *F*. Thus, in the first example, it is equally valid to argue that the earnings–education relationship is contingent on a worker's years of work experience. And the flipside in the second example is that the effect of race on approval varies with education.

Interaction effects are not limited to a single pair of predictors. You could specify that *F*'s effect on the outcome changes with two other predictors separately—a two-moderator model—or that it is dependent on the specific combination of values of the other two predictors—a three-way interaction. In the race–education–legal ban example, the legal scholar might extend the original hypothesis to argue that the effect of race varies not only by education but also by region of residence. A two-moderator model would specify that the race-by-education effect on approval is the same in each region, that the race-by-region effect is the same at each level of education, and that there is no interaction between education and region. In contrast, a three-way interaction model would specify that the race-by-education effect on approval differs across regions, that the race-by-region effect varies with education, and that the education-by-region effect varies by race.

Why Should You Consider Including an Interaction Effect in Your Analysis?

With some exceptions in practice, the primary rationale for estimating and testing interaction effects is on theoretical or substantive grounds. That is, you have developed new hypotheses or expectations that certain predictors should have contingent effects. But it is also conventional to include interaction effects to reflect current knowledge in the literature. Additionally, in many of the social sciences (certainly in my discipline of sociology), it is commonplace to propose and analyze outcome differences between groups defined by their social characteristics or statuses such as race/ethnicity, sex or gender, sexual orientation/identity, class, and so on. In the course of developing the rationale for group differences, it is not unusual (and often purposeful) to make an argument that groups diverge in outcome levels because different factors are important for some groups, or the same factors have varying effects. Both these arguments create an expectation that interaction effects exist and set the stage for testing for interactions between the groups and at least some of the predictors.

Diagnostic testing of model fit or for model misspecification sometimes provides the grounds for estimating and testing for interaction effects. For example, finding a significant diagnostic test for the presence of heteroscedasticity might instead indicate the presence of an interaction effect or some other misspecification of the model's functional form (Fox, 2008, p. 274; Greene, 2008, pp. 166–167; Kaufman, 2013, pp. 22–23). Similarly, a residuals analysis could find issues with the functional form of two predictors that might suggest an unspecified interaction effect.

Even the failure to find a significant effect for a predictor might lead you to test for interaction. When a predictor has opposite-signed effects for two groups—or changes sign across the range of its moderator—this can easily average out to a nonsignificant

test of its effect. In such circumstances, I would recommend that you think seriously about the substantive sensibility of specifying an interaction effect before testing for its presence. Other specifications or corrective actions might be more conceptually appealing. And adding interaction effects as a result of data mining runs the risk of overfitting and hence misspecifying the model.

How Do You Specify an Interaction Effect in the Prediction Function of a Linear Regression Model?

To include an interaction effect, you model the outcome Y as a linear function of the focal and moderating variables (F and M_j), their product terms, and a set of other predictors. The coefficients for the focal and moderating variables are commonly referred to as the "main effects" of the predictors, while the coefficients for the product terms are called "interaction effects." Your model should, with rare exceptions, satisfy the principle of marginality (Fox, 2008; Nelder, 1977), also known as the criteria for a hierarchically well-formulated model (Jaccard, 2001; Kleinbaum, 1992):

> [This] specifies that a model including a *higher order term* (such as an interaction) should normally also include the "lower-order relatives" of that term (the main effects that "compose" the interaction). (Fox, 2008, p. 135)

Table 1.1 lists the predictors your model should contain for several forms of interaction effects to adhere to the principle of marginality. Specifically, you add a product term formed by multiplying together all the predictors in the highest order interaction as well as product terms for every lower order relative. For a one-moderator model, you add a predictor defined as the product of F and M_1. Notice that I said "add" a predictor because you keep the individual predictors in the model when you include the product term following the principle of marginality. Similarly, for a two-moderator model, you have three individual predictors (F, M_1, and M_2) plus two product terms ($F \times M_1$ and $F \times M_2$). For a three-way interaction model, you have three individual predictors (main effects), three pairs of product terms among the three predictors (lower order relatives), and a product term for $F \times M_1 \times M_2$ (highest order term).

TABLE 1.1 ⬢ PREDICTORS INCLUDED FOR FORMS OF INTERACTION MODELS

Interaction Form	Predictors in Model for Interaction Specification
One moderator	F, M_1, $F \times M_1$
Two moderators	F, M_1, M_2, $F \times M_1$, $F \times M_2$
Three-way interaction	F, M_1, M_2, $F \times M_1$, $F \times M_2$, $M_1 \times M_2$, $F \times M_1 \times M_2$
Two moderators, M_2 categorical	$F, M_1, D_{M_2=1}, D_{M_2=2}, D_{M_2=3}$ $F \times M_1$ $F \times D_{M_2=1}, F \times D_{M_2=2}, F \times D_{M_2=3}$

How does adding product terms work to create contingent effects of the interacting variables? For simplicity and specificity, let's work with a single-moderator specification:

$$Y = a + b_1F + b_2M_1 + b_3F \times M_1 + \cdots \tag{1.1}$$

Mathematically, the effect of a predictor on Y in a linear regression model is defined as the partial derivative of Y with respect to a predictor,[1] which is the slope of the regression plane. The partial derivative is equal to a predictor's estimated coefficient if the predictor is not part of an interaction specification (or a multiple-variable functional form, e.g., a parabolic effect). For an interaction specification, the partial derivative gives the effect of F as the main effect coefficient for F (b_1) plus the coefficient for the product term $F \times M_1$ (b_3) multiplied by the value of M_1:

$$\text{Effect of } F = \frac{\partial Y}{\partial F} = b_1 + b_3M_1 \tag{1.2}$$

This tells us very concretely how the effect of F on Y changes with the value of the moderator. Similarly, the effect of M_1 found by taking the partial derivative of Y with respect to M_1 is

$$\text{Effect of } M_1 = \frac{\partial Y}{\partial M_1} = b_2 + b_3F \tag{1.3}$$

Using these formulas, we can determine the nature and shape of the moderated effect of F (or M_1) on the outcome and how those change across different values of the moderator. By nature and shape of the effect, I mean the direction of the effect of F (positive or negative), whether it changes sign for different values of M_1, whether it changes significance across the values of M_1, and whether the moderated effect is ordinal or disordinal (see the Aside). How to probe the nature and shape of the interaction is covered in depth in Chapters 2 to 5.

ASIDE: ORDINAL AND DISORDINAL INTERACTIONS

Consider drawing a set of prediction lines for the value of Y plotted against F, each prediction line for a different value of M_1. These prediction lines will all cross at the same value of F (Jaccard & Turisi, 2003, p. 78). The interaction is ordinal if that value of F does not fall within the sample range of values of F. It is disordinal if the crossover value lies within the sample range of F. Conceptually, a disordinal interaction means that the outcome for a given value of the focal variable F and a specific value of the moderator M_1 is sometimes greater but is sometimes less than the outcome for the same value of F and a different value of M_1. And an ordinal interaction means that the outcome for a given value of the moderator M_1 is always greater (or always less) than the outcome for a different value of the moderator, for any value of F in the sample range.

What if one or all of your predictors are categorical? For each categorical predictor listed in Table 1.1, you would replace the single-variable expression with the set of

binary indicators for your predictor. When you create product terms between two predictors, you multiply every term in the first predictor's expression by every term in the second predictor's expression. To illustrate, the last row in Table 1.1 shows the terms included in a two-moderator interaction specification if F and M_1 are interval variables but M_2 is a three-category construct represented by dummy variables for Categories 1 and 2 ($D_{M_2=1}$ and $D_{M_2=2}$; Category 3 is the reference category). To create the list of included predictors, you replace every occurrence of M_2 with its set of dummy indicators. For example, in the expression $F \times M_2$, you replace M_2 with $D_{M_2=1}$ and $D_{M_2=2}$ and multiply by F, which gives two product terms to include in the model: $F \times D_{M_2=1}$ and $F \times D_{M_2=2}$. The corresponding prediction function becomes

$$Y = a + b_1F + b_2M_1 + b_3F \times M_1 + b_4D_{M_2=1} + b_5D_{M_2=2} + b_6F \times D_{M_2=1} + b_7F \times D_{M_2=2} + \cdots \quad (1.4)$$

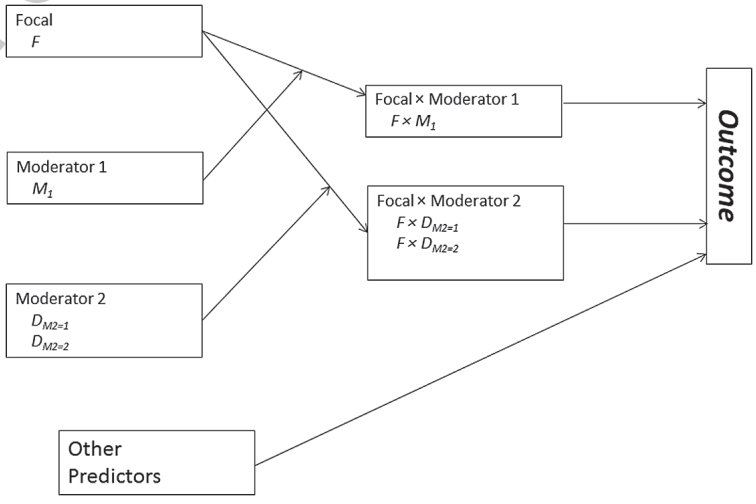
Taking the partial derivative of this prediction function, the moderated effect of F is

$$\text{Effect of } F = \frac{dy}{dF} = b_1 + b_3M_1 + b_6D_{M_2=1} + b_7D_{M_2=2} + \cdots \quad (1.5)$$

Because this is a two-moderator interaction, the effect of F varies corresponding to both the values of M_1 and the categories of M_2. As I elaborate in Chapter 2, the expressions for the moderated effect of F like those in Equations 1.2, 1.3, and 1.5 are an important basis for understanding and interpreting interaction effects.

A path–style diagram can be a useful device to show succinctly the nature and form of your interaction specification, especially for more complicated interaction specifications or when some of your interacting predictors are categorical. Writing out the prediction function in terms of all the component predictors that constitute it is essential for running your analysis, and such an equation communicates the form of the interaction well to mathematically or formulaically inclined readers. But for many readers, a diagram like Figure 1.2 for the two-moderator example is much more comprehensible.

FIGURE 1.2 ⬢ PATH–STYLE DIAGRAM OF INTERACTION EFFECT



The path diagram has boxes for the focal and moderator variables in the leftmost column and for the two-way interaction terms in the second column, and the outcome is a vertical box on the right. There is a box for each construct in the interaction, and the box indicates whether the construct is measured by a single predictor or a set of indicators. The interaction terms are expressed as *Focal* × *Moderator* variable names to make the conceptual and mathematical relationships clearer. Lines connect main effect terms to the relevant two-way interaction terms and connect the two-way terms to the outcome. Intersecting lines—where one line stops with its arrowhead on the other line—indicate that the effects of the corresponding variables interact. This provides a visual and conceptual map of the interaction specification.

The diagram shows that F and M_1 interact because their lines intersect and lead to their corresponding two-way interaction term. Similarly, F and M_2 interact. But M_1 and M_2 do not interact because their lines do not intersect. The diagram shows a nonintersecting line from the "Other Predictors" box to the "Outcome" box as a reminder that a model usually has more than just the interacting predictors. In practice, the estimated coefficients could be shown in the diagram.

When Is an Interaction Effect Statistically Significant?

Part of the answer to why include an interaction specification is that it is statistically significant, but what does that mean in terms of what you specifically test? The essential statistical test is whether or not the coefficient(s) for the highest order term in the interaction are significant. If the coefficient(s) are not significantly different from zero, you would conclude that the lower order effects on Y do not vary at that highest level. The significance or lack of significance of any lower order term constituting the interaction is irrelevant to this decision (Aiken & West, 1991, p. 50).

For example, in a one-moderator model in which F and M_1 are interval level, the highest order term is $F \times M_1$, and if its coefficient is not significant then the effect of F does not depend on M_1 and you should use a model without the interaction term. If the $F \times M_1$ coefficient is significant, you would conclude that the effect of F varies with M_1 and use the interactive model results. The significance of the main effect of F is not relevant because this tests whether F's effect is significantly different from zero when $M_1 = 0$. To see why the main effect coefficient is the effect of F when $M_1 = 0$, look back at Equation 1.2, which defines the effect of F. Substitute 0 as the value for M_1, and this leaves b_1, the main effect coefficient for F:

$$\text{Effect of } F = b_1 + b_3 \times 0 = b_1 \tag{1.6}$$

This means that b_1 is the effect of F when M_1 takes on the value of zero and tells you nothing about the significance of F's effect at any other value of M_1. Its significance may or may not be a substantively interesting finding by itself, but it is not informative about the statistical grounds for including the interaction in your model. The same holds true for testing the main effect of M_1: This tests the effect of M_1 when $F = 0$.

What if you were testing a three-way interaction? You would test the coefficient(s) for the product of the three predictors ($F \times M_1 \times M_2$), and if it is significant keep all the interaction terms in your prediction equation. The significance of the two-way product terms, as well as the main effect terms, is not relevant to this decision because they are tests of effects when the moderating variables equal zero (Aiken & West, 1991, p. 50).

How you test the significance of the highest order term depends on whether or not you have directional hypotheses about the coefficients of the highest order term and

on whether it consists of a single coefficient or multiple coefficients. Table 1.2 summarizes the commonly used test statistics for different combinations of the testing situations. When you have directional hypotheses, there is no choice; you use a z test or a t test, as appropriate for your estimation technique. If you are testing a single coefficient, then the decision rule is straightforward. Keep the highest order interaction term in your model if its coefficient is significant and exclude the term if it is not. If you are testing multiple coefficients, it is typically because the focal or one or more of the moderating variables is nominal with three or more categories. This creates multiple product terms in the highest order interaction, like the two-way interaction of F and M_2 in Equation 1.4. In this case, you need to decide before you conduct the statistical testing what results constitute sufficient grounds for keeping the interaction.

A stringent rule (rarely applied to my knowledge) would be that all of the directional tests are significant. A common approach is to keep the set of coefficients in the model if any one of them is significant, with a Bonferroni or other correction of the significance level for conducting multiple tests.[2] One complication is that a different choice of reference category (or an alternative parameterization) will on occasion lead to a different conclusion. For this reason, an alternative approach is to supplement the tests of individual coefficients with a global nondirectional test such as the likelihood ratio (LR) test or the Wald test. While a global test is nondirectional, it is unaffected by the choice of reference category and provides a guard against excluding a significant interaction by your choice of a reference category. The decision rule in this scenario is to keep the set of coefficients in the model if either the multiple coefficient tests or the global test yields a significant result.

For nondirectional hypotheses, more choices are possible when testing a single coefficient: a z or t test of the coefficient, a Wald test, or the LR test. However, note that the Wald test and the single-coefficient test will always give identical results.[3] Thus, it does not matter which one you use. For testing multiple coefficients, a global test (Wald or LR test) is generally preferred over conducting multiple t tests on the individual coefficients for the reason given earlier: The global test is robust against changing the reference category or other equivalent reparameterizations.

But using just a global test will occasionally find the set of coefficients not significantly different from zero when one or more of the individual coefficients is significant and of substantive interest. Thus, some analysts use the same either/or decision rule described earlier: Include the interaction if the highest order terms are significant

TABLE 1.2 ⬢ TESTS OF THE HIGHEST ORDER INTERACTION TERM(S)

	Directional Hypotheses		Nondirectional Hypotheses	
Test	**Single**	**Multiple**	**Single**	**Multiple**
z test/t test	Preferred	Preferred	Yes	Supplement
Wald test		Supplement	Yes	Yes
LR test		Supplement	Preferred MLE	Preferred MLE

Note: LR = likelihood ratio; MLE = maximum-likelihood estimation.

using the global test or if at least one coefficient is significant from the multiple tests of single coefficients using a z or t test.

The LR test is preferred over the Wald test for nonlinear models and/or techniques of analysis using maximum-likelihood estimation, especially with a large sample size. The minor disadvantage is that you must run the model twice—once with all of the interaction terms and once excluding only the term(s) for the highest order interaction—and then test the change in the log likelihood between the two models. Some analysts prefer the Wald test because it is asymptotically equivalent to the LR test, and you do not have to make strong distributional assumptions as you do when using the LR test. Additionally, the Wald test does not require estimating two models, which is pragmatically an advantage only in those relatively rare instances in which model estimation takes a significant amount of time or if you are applying the test repeatedly.

What should you do if you test the highest order terms of, say, a three-way interaction ($F \times M_1 \times M_2$), and it is not significant? You would conclude that the effect of F does not vary across different combinations of the values of M_1 and M_2. But F's effect might vary with M_1 regardless of M_2's values and with M_2 regardless of M_1. You would test those possibilities by running a two-moderator model and testing each of those two-way terms ($F \times M_1$, $F \times M_2$, $M_1 \times M_2$) and keep a two-way term only if it is individually significant.

Common Errors in Specifying and Interpreting Interaction Effects

For concreteness in describing and discussing these errors, consider a single-moderator interaction specification in which number of children is predicted by the interaction of family income and a dummy indicator of birth cohort (1 = *Pre–Baby Boom*, 0 = *Baby Boom and younger*):

$$\begin{aligned} Children &= a + b_1 Income + b_2 Cohort + b_3 Income \times Cohort + \cdots \\ Children &= 4.4783 - 0.1075 Income - 1.0251 Cohort + \\ &\quad 0.0327 Income \times Cohort + \cdots \end{aligned} \tag{1.7}$$

Excluding Lower Order Terms

The most frequent mistake analysts make in specifying an interaction effect is to not include in the prediction function all of the lower order terms, sometimes referred to as constitutive terms (Aiken & West, 1991; Brambor et al., 2006; Braumoeller, 2004; Jaccard, 2001; Kam & Franzese, 2007). Doing so violates the principle of marginality and fundamentally changes the meaning, estimated values, and/or statistical test results of the coefficients for the other terms in the interaction specification. Perhaps the most consequential issue is the potential for model misspecification (Kam & Franzese, 2007, pp. 100–101). Unless you have both a clear theoretical argument for this exclusion and a high certainty that empirically $b_2 = 0$, you run the risk of an omitted-variable bias affecting the estimates of any predictor that is correlated with the omitted variable.

Suppose the main effect of cohort was excluded from the prediction equation. Because *Income* × *Cohort* is very likely to be positively correlated with cohort, the effect of cohort will be partly attributed (added) to the *Income* × *Cohort* effect. Depending on the sign of the (omitted) effect of cohort and of *Income* × *Cohort*,

this exclusion would either increase or decrease the estimated coefficient for *Income* × *Cohort*. Thus, the significance test of the interactive model of income and cohort versus a linear model of effects will be biased as well. Moreover, this exclusion constrains the prediction line for children plotted against income for each birth cohort to have the same intercept but a varying slope, which Fox (2008) aptly describes, for a different empirical example, as "a specification that is peculiar and of no substantive interest" (p. 138).

A related question that invariably comes up when I first introduce students (or colleagues) to interaction effects is whether you should exclude from the model nonsignificant lower order terms if the highest order term is significant. The proposed rationale for doing so is that keeping the lower order terms in the model unnecessarily inflates the standard errors of the other predictors (decreases the efficiency of estimates). However, the consequences of excluding the lower order terms when they should be included are much more consequential—biased coefficient estimates, as I just discussed.

The more fundamental problem with deciding to exclude nonsignificant lower order terms is that you could well make the opposite decision if you had a mathematically equivalent but different specification of your interaction model. In the example, suppose the main effect of income (b_1) is significant, so you don't think about excluding it from the prediction equation. Someone else does the same analysis except that instead of older cohort (1 = *older cohorts* and 0 = *most recent*) he or she uses recent (1 = *most recent cohorts* and 0 = *older cohorts*) and their prediction equation is

$$Children = a^* + b_1^* Income + b_2^* Recent + b_3^* Income \times Recent + \cdots \tag{1.8}$$

But they find the main effect of income (b_1^*) is not significant, so they decide to exclude the main effect term for income from the prediction equation. The problem with these opposite decisions is that the two prediction equations are mathematically equivalent to each other—you can derive the coefficients of one from the coefficients of the other. To see this, realize that *Recent* = 1 − *Cohort* and rearrange the terms:

$$\begin{aligned} Children &= a^* + b_1^* Income + b_2^* (1 - Cohort) + b_3^* Income \times (1 - Cohort) + \cdots \\ Children &= (a^* + b_2^*) + (b_1^* + b_3^*) Income - b_2^* Cohort - b_3^* Income \times Cohort + \cdots \end{aligned} \tag{1.9}$$

We can now write the coefficients for the original parameterization in terms of the alternative parameterization:

$$a = a^* + b_2^* \qquad b_1 = b_1^* + b_3^* \qquad b_2 = -b_2^* \qquad b_3 = -b_3^*$$

The point is that the two analyses test different things when they test the main effect of income in their prediction equation. The first analysis tests the effect of income when *Cohort* = 0 and the second tests the effect of income when *Cohort* = 1 (i.e., *Recent* = 0). The underlying conceptual problem with excluding the main effect term is that it serves to anchor the moderated effect of income by setting a value for the effect when its moderator (whether cohort or recent) takes on the value of zero. When you change between equivalent parameterizations, the anchor value adjusts so that you get the same moderated effect of income. So you need to keep it in the model as long as the higher order term is in the model. (For a more detailed

discussion, see Aiken & West, 1991; Allison, 1977; Brambor et al., 2006; Kam & Franzese, 2007.)

Interpreting Coefficients as Unconditional Marginal Effects

Because the lower order and higher order terms in the interaction specification are functionally related, you cannot interpret the coefficient for any interaction term without considering how it is related to the other interacting predictors (Aiken & West, 1991; Allison, 1977; Brambor et al., 2006; Braumoeller, 2004). Consider the numeric coefficient of −1.0251 for the main effect of cohort in Equation 1.7. The incorrect way to interpret this would be to say that the predicted number of children for the older cohorts is about one less child than for the younger cohorts. This depicts an unconditional relationship between cohort and children and ignores that the effect of cohort is functionally related to the interaction of *Cohort* × *Income*.

A correct interpretation would be that when *Income* = 0, the predicted number of children for the older cohorts is about one child less than for the younger cohorts. Notice how this statement makes clear that income values define a contingent relationship of cohort to children. But it does not tell the full story of that contingency. A better description would be that the predicted number of children for the older cohorts is about one child less than for the younger cohorts when *Income* = 0 and is predicted to increase by 0.0327 children for each $10K increase in income. Equivalently, and perhaps more informative, would be to replace the end of the sentence with "… *Income* = 0 and is predicted to increase by about one third of a child for a $100K increase in income."

Interpreting Main Effect Coefficients When Not Meaningful and the Myth of Centering

It is always technically correct to interpret a main effect coefficient as the effect of the predictor when its moderator is equal to zero. But that technical interpretation is not always meaningful—in particular, when zero is not a possible value for the moderator or if zero is possible but is not within the sample range for your analysis. In those cases, the interpretation is not meaningful and may be confusing. For example, suppose you are analyzing how household size moderates the effect of rent on savings. The minimum household size is one, so it would not be sensible to interpret the effect of rent for a household with zero people.

This is one reason why some didactic works on interaction effects recommend centering an interval (continuous) predictor by subtracting its sample mean before running the analysis (Aiken & West, 1991; Dawson, 2014; Hayes, 2013; Jaccard & Turrisi, 2003). If a predictor's moderator is centered, then the main effect of the predictor is its effect when the centered moderator equals zero—which is to say, when the moderator equals its mean. In this case, the main effect term will have a meaningful interpretation. In the household size by rent example, the main effect coefficient for rent would now be meaningful because it would refer to the effect of rent when household size is equal to its sample mean. But the numeric value, interpretation, and statistical significance of the moderated effect (Equation 1.5) is the same whether or not you center your predictors (Kam & Franzese, 2007, p. 97).

The myth of centering refers to the second reason often given for centering; namely, that it supposedly reduces problems of collinearity between the components of the interaction specification. The reality is that centering makes no real difference in the estimation of the parameters of the model. Some coefficients and their meanings

change because you changed the measurement of your predictors but not, as noted above, the overall moderated effect of the predictor. The minor and rare exception is that if the collinearity is so extreme that the parameters for the uncentered model cannot be estimated, it is possible but not likely that centering could make enough of a difference to estimate the parameters.

In any case, you can mathematically derive the uncentered coefficients and their standard errors from the centered coefficients, their standard errors, and the covariance between the coefficients, and vice versa (see Kam & Franzese, 2007, pp. 96–98). A point that is often overlooked is that collinearity among the components of an interaction specification is normal and expected. The predictors are functionally related in how they create the overall moderated effect and should be in many instances highly correlated. We should not expect to get precise estimates of the coefficients for the individual components of an interaction because there is by definition a mathematical relationship between the magnitude of the coefficients—each depends on the values of the other predictors in the interaction.

I want to clarify that I am not arguing against examining your data for potential problems of collinearity among your predictors by means of standard diagnostic tools, such as the presence of very high bivariate correlations, variance inflation indices, or the Belsley, Kuh, and Welsch (1980) collinearity diagnostics. Just do not be concerned about potential multicollinearity among the component predictors of an interaction specification.

In sum, there is neither any harm in centering your predictors nor any major advantage. The one gain is that centering can give a main effect coefficient a meaningful interpretation when a moderator value of zero is not possible. In general, the ability to meaningfully interpret the main effect can be useful in some analyses. This is invariably true when the moderator is a dummy variable indicator or a set of dummy variable indicators (which you would not center) because the value of zero corresponds to the reference category. For example, the main effect of income on children is the income effect for recent cohorts (*Cohort* = 0).

Not Interpreting the Moderated Effect of Each Predictor Constituting an Interaction

Brambor et al. (2006) document that many articles in top political science journals that report interaction effect analyses fail to provide an overall picture of the nature and significance of the moderated effect (Equation 1.5). In practice, many analyses only report and discuss the effect and statistical significance of a predictor when its moderator is equal to zero rather than reporting and describing how the effect and its significance changes across the range of the moderator (see the brief example discussed earlier in which I described the effect of cohort as it varies with income). This leads to incomplete and possibly misleading conclusions about the moderated relationship and the hypothesis it represents.

A related issue is a tendency to reify the heuristic device of designating one of the interacting predictors as the focal variable and the others as the moderating variables (also known as conditioning variables). This leads to a failure to reverse the roles of focal and moderating variables by only interpreting the moderated effect of the focal variable. This ignores that a higher order interaction term is symmetric and modifies the effects of all its constitutive predictors. As a result, authors provide and discuss partial and potentially incorrect descriptions of the nature of the complete

interactive relationship (Berry, Golder, & Milton, 2012; Kam & Franzese, 2007). I think this tendency is reinforced by the fact that researchers sometimes propose and test hypotheses about contingent relationships based on asymmetric logic. That is, they initially develop hypotheses explicitly arguing how one factor creates contingent effects of the other and do not explicitly develop hypotheses or rationales for the interaction the other way around.

THE LOGIC OF INTERACTION EFFECTS IN GLMs

What Are GLMs?

They are a class of models that generalize linear regression by relaxing its assumptions to create a common statistical foundation for a wide range of specialized statistical models. To define a GLM, I follow the updated criteria described by Hardin and Hilbe (2012, chap. 2), which includes a wider set of models than the traditional formulation (Nelder & Wedderburn, 1972). Fox (2008, pp. 379–387) concisely frames the GLM approach as three questions whose answers define the type of GLM estimated:

1. What probability distribution characterizes the distribution of the outcome (Y) conditional on the predictors (X)? With limited recent exceptions (Hardin & Hilbe, 2012, p. 12), the choices are from the exponential distribution family and require that the variance be solely a function of the mean. Common choices include the normal, binomial, multinomial, Poisson, negative binomial, gamma, and inverse-Gaussian distributions.
2. What link function $g(\,)$ transforms the expected value of the outcome, $\mu = E(y)$, such that it is a linear function of the predictors? The link function $g(\,)$ must be monotonic, one-to-one, and differentiable. The most frequent options are the identity, log, inverse, inverse-square, square-root, logit, probit, log-log, and complementary log-log functions (see Fox, 2008, p. 380, table 15.1, for the mathematical definition of these link functions).
3. What variables X constitute the linear prediction function for η, the transformed expected value of the outcome—that is, for $\eta = g(\mu) = X\underline{\beta}$? As for a linear model, you specify a linear and additive function of the predictors with coefficients $\underline{\beta}$. Thus, the prediction function may include dummy variables, logged variables, polynomial functions, product terms for interactions, and the like.

For didactic and notational reasons, it is useful to think about a GLM as defined by two component equations. The first is the *linearizing (measurement) equation* that identifies the function $g(\,)$ that transforms the expected value of the observed outcome, $\mu = E(y)$, into the expected value of the modeled outcome (η):

$$\eta = g(\mu) \tag{1.10}$$

Equivalently, we can write μ as a function of η using the inverse[4] of the function g:

$$\mu = g^{-1}(\eta) = g^{-1}\left(\text{linear prediction function}\right) \tag{1.11}$$

For example, in a negative binomial model, $g(\,)$ is the natural log, ln( ), and its inverse function is $g^{-1}(\,) = e^{(\,)}$. That is, $\eta = \ln(\mu)$ and $\mu = e^{\eta}$.

The second component is the *modeling (structural) equation*, which specifies the linear prediction function for the modeled outcome (η). For interaction effects, the modeling equation specifies a linear function of the focal and moderating variables (F, M_1, M_2, . . .), their interaction ($F \times M_1$, $F \times M_2$, . . .), and a set of other predictors heuristically represented by Z:

$$\eta = a + b_1F + b_2M_1 + b_3F \times M_1 + b_4M_2 + b_5F \times M_2 + \cdots + b_zZ \tag{1.12}$$

Note that η can be the expected value of either an observed or an unobserved outcome depending on the link function for the specific GLM estimated. For example, for OLS, η is the observed outcome, Y; while for binomial logistic regression, η is the unobserved outcome, log odds(Y). This expression is sometimes called the "index function."

(Interaction) Effects in the Modeling Component

I labeled the second equation as the modeling or structural component to emphasize that you specify and test your conceptual model of effects in this expression. That is, when you test the coefficients, you do not directly test the effect of a predictor on the observed outcome, except when the link function $g(\,)$ is the identity link. As Equation 1.12 indicates, you directly test a predictor's effect on the modeled outcome η. When you perform a global test of multiple coefficients, you are thus testing whether or not those coefficients are needed in the model predicting the modeled outcome η. Thus, the prior material on how to test the statistical significance of an interaction applies to testing coefficients in the modeling component, not to testing the effects of coefficients on the observed outcome.

Equation 1.12 also means that you can interpret the coefficients for the predictors directly as effects on the modeled outcome. For example, take the partial derivative of η with respect to F in Equation 1.12 to find the effect of F on η:

$$\text{Effect of } F \text{ on } \eta = \frac{\partial \eta}{\partial F} = b_1 + b_3M_1 + b_5M_2 \tag{1.13}$$

This is the counterpart to the moderated effect of F in a linear model shown in Equation 1.2. The difference is that Equation 1.13 is the effect of F on the modeled outcome and Equation 1.2 is the effect of F on the observed outcome. For a GLM with a nonidentity link function, these are not equivalent. As will be discussed throughout the book, this can sometimes be advantageous for interpreting interaction effects.

(Interaction) Effects on the Observed Outcome

This does not mean you cannot interpret the effect of predictors on the observed outcome in a GLM. Rather it means that the estimated coefficients by themselves do not describe that effect. To find the effect of a predictor on the observed outcome, you need to formulate the expression for the expected value of the observed outcome (μ)

as a function of the predictors. The effect of a predictor on the observed outcome is the partial derivative of that expression with respect to the predictor. Start by writing μ as a function of η from Equation 1.11 and then substitute for the linear prediction function from Equation 1.12:

$$\begin{aligned}\mu &= g^{-1}(\eta) = g^{-1}\left(\text{linear prediction function}\right)\\ &= g^{-1}(a + b_1F + b_2M_1 + b_3F \times M_1 + b_4M_2 + b_5F \times M_2 + \cdots + b_zZ)\end{aligned}$$

Using $XB = \mathrm{a} + b_1F + \cdots$ to make the equation more readable gives

$$\mu = g^{-1}(XB) \tag{1.14}$$

Taking the partial derivative with respect to F gives F's effect on the observed outcome as

$$\frac{\partial \mu}{\partial F} = \left(b_1 + b_3M_1 + b_5M_2\right) \times \frac{\partial}{\partial \eta} g^{-1}\left(XB\right)$$

This expression indicates how the effect of F on the observed outcome is related to the effect of F on the modeled outcome. The effect of F on the observed outcome is equal to F's effect on the modeled outcome—the first expression in parentheses—multiplied by a factor defined by the inverse link function and the coefficients and values of every predictor. This makes explicit that the effect of F on the observed outcome expresses two sources of nonlinearity—that specified by the interaction effect and that created by the inverse link function applied to the predictors' coefficients and values. When examining tables or plots of how the effect of F varies with its moderators or how predicted outcome values change with F, it is difficult to parse out how the patterns represent the nonlinearity of the interaction from how they represent nonlinearity induced by the link function. Chapter 5 explores in depth the interpretive complications of the presence of these conflated sources of nonlinearity and proposes solutions.

ASIDE: NONLINEAR EFFECT OF NONINTERACTING PREDICTOR

When F is not part of an interaction, its effect on the observed outcome is

$$\frac{\partial \mu}{\partial F} = b_1 \times \frac{\partial}{\partial \eta} g^{-1}\left(XB\right)$$

The nonconstant effect of F makes the obvious point that there is a nonlinear relationship between the observed outcome and F. Not quite as obvious is that the effect of F will vary with the values of the other predictors because the multiplicative factor calculated by $g^{-1}\left(XB\right)$ depends on those values. This is why the predicted value plot from a main-effects-only model seems to show an interaction effect (in case you were wondering, Plot B shows the main effects predictions). The coefficient for age is multiplied by a different factor at different levels of education yielding varying effects of F.

Common Errors in Using Interaction Effects in GLMs

The common errors concerning interaction effects in linear models all apply in some fashion when using them in GLMs. The errors in specifying interaction effects apply to the specification of the predictors in the modeling component and not directly to the observed outcome unless the link function is the identity link. But the common errors of interpretation apply to interpreting interaction effects on both the modeled outcome and the observed outcome. Beyond these errors, there are other pitfalls when interpreting interaction effects for GLMs with nonlinear link functions.

Improperly Treating Product Terms for an Interaction

The fact that a noninteracting predictor's effect gives the appearance of an interaction with the other predictors in nonlinear models has led some authors to propose that product terms for interaction are not always needed to find and interpret interaction effects (e.g., Berry, DeMerit, & Esarey, 2010). This conflates the nonlinear nature of the statistical model (link function)—what Nagler (1991, p. 1393) describes as an "artifact of the methodology"—with a substantively driven test and interpretation of the form of the relationship. Kam and Franzese (2007) describe this as the difference between implicit interaction resulting from the form of the link function and explicit interaction designed to model hypotheses or expectations. There is a clear and continuing agreement that product terms are necessary to model and test hypotheses about interaction effects in nonlinear models (Berry et al., 2012; Brambor et al., 2006; Braumoeller, 2004; Kam & Franzese, 2007).

A different error concerning the product terms is partly a software-driven error in interpreting the coefficient (and/or its marginal effect) for the product term in isolation from the other components of the interaction. Specifically, the problem occurs when statistical software treats the product term as a unique predictor without taking into account its functional relationship to the component predictors constituting the product term (Ai & Norton, 2003; Greene, 2010). It "mechanically computes a separate 'partial effect' for each variable that appears in the model" (Greene, 2010, p. 292), so the partial effect for the product term $F \times M_1$ is calculated as the partial derivative with respect to the quantity $(F \times M_1)$:

$$\frac{\partial \mu}{\partial (F \times M_1)} = b_3 \times \frac{\partial}{\partial \eta} g^{-1}(XB) \tag{1.15}$$

But what you want is really two different partial effects that involve the product term, the partial effect of F when you describe the moderated effect of F, and the partial effect of M_1 when you interpret the moderated effect of M_1:

$$\begin{aligned} \frac{\partial \mu}{\partial F} &= (b_1 + b_3 M_1) \times \frac{\partial}{\partial \eta} g^{-1}(XB) \\ \frac{\partial \mu}{\partial M_1} &= (b_2 + b_3 F) \times \frac{\partial}{\partial \eta} g^{-1}(XB) \end{aligned} \tag{1.16}$$

This has led some authors to interpret the computed partial effect in Equation 1.15 when it is in fact not a meaningful statistic. Note that the advent of the *margins*

command in Stata is specifically intended to avoid this problem. But it does so only if the interactions are specified using Stata's factor-variable notation when estimating the GLM.

Limited Range of Moderator Values Used to Probe Moderated Effect of Focal Variable

Too often, the interpretation of the effect of *F* discusses its value calculated across a restricted range of the values of its moderators (see the review of published studies in Brambor et al., 2006). Doing so may conceal substantively important—and sometimes unexpected—variation in *F*'s effect. Technically, this point applies to both linear and nonlinear models, but this limitation is much more consequential for the effect of *F* on the observed outcome in nonlinear models because the link function's nonlinearity affects the calculation of *F*'s effect. With limited comparison points, it is more difficult to separate out how much of the difference in *F*'s effect represents the explicit interaction by the moderator from how much is the nonlinearity of the link function. Thus, a recurring recommendation in the didactic literature is to examine how the effect of *F* changes with its moderator(s) across a substantively meaningful range of the moderators' values (Aiken & West, 1991; Brambor et al., 2006; Hayes, 2013; Kam & Franzese, 2007).

Comparing Estimated Coefficients Across Nested Models (for Some GLMs)

Typically, the reason to compare a predictor's coefficient across nested models for the same sample is to examine how other factors mediate (explain) the influence of that predictor by adding predictors in stages. In some GLMs, the model is identified by setting a fixed value for the error variance of a latent outcome. This results in coefficients that are identified only up to a multiplicative scaling factor (Long, 2009, pp. 47, 122–123; Maddala, 1983, p. 23). But the total variance of the (latent) outcome is not fixed as it would be for an observed outcome. Rather, it is the sum of the explained variance and the fixed value of the error variance. The problem is that the explained variance must increase as predictors are added to the model, even if minutely, which results in a larger total variance and hence a bigger scaling factor for the coefficients. Consequently, it is not unusual for the estimated coefficients to increase across stages of the nested model. Because the estimated coefficient can change solely due to the scaling factor changing, you cannot compare the coefficients across the nested models to examine how they are mediated by the other factors nor attribute such changes to suppressor effects.

The simplest and commonly recommended solution is to examine changes in *y*-standardized coefficients, defined as the coefficient divided by the estimated standard deviation of the latent outcome (Long, 1997; Mare & Winship, 1984; Mood, 2010; Williams, 2013). Because the coefficients at each stage are standardized by the latent outcome's changing standard deviation, it counters the differences in the scaling of the estimated coefficients on the condition that the sample analyzed must be the same at each stage. Other solutions include comparisons using predicted probabilities or functions of the predicted probabilities (Long, 2009, 2016) or techniques to decompose effects (Buis, 2010; Kohler, Karlson, & Holm, 2011). Note that this concern is not specific to interaction coefficients but applies to any coefficient in nested models.

ASIDE: MODERATED EFFECTS VERSUS MEDIATED EFFECTS

The discussion of comparing coefficients across models introduced the idea of the mediation of an effect that is sometimes confused with the moderation of an effect. A mediation analysis (also known as path analysis) seeks to understand the causal process through which a predictor *X* affects a penultimate outcome *Y*, drawing a distinction between its direct causal effect and its indirect causal effect. Conceptually, *X*'s indirect (mediated) effect is how *X* affects *Y* through its causal effect on other causes of *Y*. That is, the indirect effect is how *X* affects intermediate outcomes (*Z*), which in turn have their own direct effects on *Y*. This is often described as *Z* mediates (explains) some or all of the total effect of *X* on *Y*.

In contrast, a moderation or interaction analysis does not necessarily adopt a causal analysis framework. Rather, it focuses on how the effect of *X* on *Y* is contingent on (varies with) other predictors (*Z*) in the analysis. This is commonly labeled as *Z* moderates the effect of *X* on *Y*. It is possible to have a model that specifies that *X*'s effect on *Y* is both mediated and moderated by other predictors in the model. (For an excellent introduction to mediation and moderation analysis in OLS regression, see Hayes, 2013.)

DIAGNOSTIC TESTING AND CONSEQUENCES OF MODEL MISSPECIFICATION

Before deciding to include and interpret interaction effects, you should always conduct diagnostic tests of model fit and the validity of the modeling technique's assumptions for your data. Model misspecification can create the appearance of interaction, and vice versa (Aiken & West, 1991; Fox, 2008, p. 274; Greene, 2008, pp. 166–167; Kaufman, 2013, pp. 22–23). In this section, I first describe diagnostic tests that apply broadly to GLMs: testing the link function, assessing model fit/departures, residual analyses for model misspecification, and analysis of influential cases. I then discuss the consequences of misspecifying interaction effects. In each application chapter, I discuss additional diagnostic tests applicable to that specific GLM.

Diagnostic Testing

Link Function Test

The usual way to assess the appropriateness of the link function is by an added variable analysis (Hardin & Hilbe, 2012, p. 55). Specifically, you construct the predicted value of the index function and its square and then estimate your model with those two variables. The significance test for the squared term's coefficient provides an assessment of the adequacy of the link function. If the coefficient is not significant, you would conclude that the link function is appropriate. But if it is significant, you should consider alternative link functions. You can also directly compare two different link functions (for details, see Hardin & Hilbe, 2012, pp. 50–51), but to the best of my knowledge this is not often done.

Assessing Overall Model Fit/Departures

A plot of the standardized deviance residuals against the predicted value of the index function provides a nonspecific diagnostic for systematic departures from the model.

By nonspecific, I mean that it does not identify the potential source(s) of the problem, only that there is evidence that the model fit is not adequate. A well-fit model should show no pattern or trend of the residuals against the predicted index function. If the plot shows a pattern, this indicates a problem with the fit of the model but is not informative about why. Thus, if you find a pattern, you should follow up using the next two diagnostic tests to try to find and correct the sources of the lack of fit.

Residual–Predictor Plots or Partial Residual–Predictor Plots

A residual–predictor plot can help identify if a predictor is creating poor model fit because its functional form is not properly specified. If the plot shows a pattern of the residuals changing with a predictor, it indicates you may need to consider alternative functional forms for the predictor. The pattern you see is net of the predictor's effect in the current model, so it typically represents departures from a linear relationship and can help you decide what type of change to make. It is usually easier to determine an alternative functional form if you create a partial residual–predictor plot because the pattern you see is not net of the predictor's effect in the model; that is, the pattern you see is the pattern you want to reproduce with an alternative functional form. For a given predictor, you create the partial residual by subtracting the predictors' effect as specified in the index function. For instance, for a current model predictor with a

$$\begin{aligned} &\text{Linear effect,} && \textit{Partial residual} = \textit{Residual} - b_{x_j} \times x_j \\ &\text{Quadratic effect,} && \textit{Partial residual} = \textit{Residual} - b_{x_j} \times x_j - b_{x_j^2} \times x_j^2 \end{aligned} \tag{1.17}$$

In principle, such plots could provide information relevant to identifying interaction effects. If two noninteracting predictors each exhibit departures from good model fit, this might suggest a missing interaction effect, but other model misspecifications could also produce that result. If you create separate plots for selected values of one of the predictors in which you plot the residual or partial residual against the other predictor, you may be able to discern an interactive pattern—that is, different slopes. You could also try using a three-dimensional (3D) scatterplot of the residual against the two interacting variables. But in practice either of these plots can be difficult to examine and to discern interaction effects (Fox, 2008, pp. 284–286). They are useful primarily for a two-variable interaction when one of the variables is categorical or has very few interval values. Otherwise, you have to group an interval variable—potentially collapsing across important changes—and need a large sample size to have sufficient cases in the subsamples' separate plots. If your model has two moderators and/or a three-way interaction, you would have to create separate plots for each combination of values of the two moderators.

Residual–Omitted Variable Plots

In any analysis, it is always possible that you omitted a relevant predictor that, if correlated with predictors in the model, would manifest as poor model fit in the prior types of residual plots. A convenient way to determine if a predictor that is not in the model potentially should be included is to plot the residual against any omitted predictors; a systematic patterning of the residuals by the predictor would indicate its inclusion in your model. Alternatively, you could formally test this by adding potential omitted predictors to your model and conduct standard statistical tests of whether or not

the added predictor(s) have significant effects. This diagnostic tool presumes that you have measures of the omitted predictors. In my experience, the main reason for an omitted–predictor bias is that you do not have measures to use in the analysis.

Analysis of Influential Cases

Influential cases are those observations in the estimation sample that have both high leverage values and large residual values. Leverage values measure the distance of an observation's values for the predictors from the typical values in the sample (Fox, 2008, pp. 245, 412; Long, 1997, p. 100; Pregibon, 1981, p. 706). Because cases with high leverage values are unusual relative to the average case in terms of the predictors, they have the potential to affect coefficient estimates especially if a high leverage values case has a large residual. This is particularly concerning for interaction analysis because such cases can "pull" coefficient estimates toward their effects on the outcome. In some instances, this can produce a significant interaction effect that is absent if the sample excludes the influential cases.

Although separate analyses of leverage values and of residuals can be somewhat informative about influential cases, the better choice is to use a direct measure of influence. Perhaps, the most common summary indicator of influence for linear models is Cook's distance measure (see Belsley et al., 1980, for an alternative summary indicator as well as coefficient-specific influence measures). For the *i*th observation, you calculate the sum of the squared change in the predicted value of every observation between the model estimated with the full sample and a model reestimated without case *i*, and normalize it by the product of the sample mean of *y* and the mean squared error of the regression (Fox, 2008, p. 250). Note that Cook's distance measure is computationally intensive since it requires reestimating the model for every observation in your estimation sample.

For GLMs, Pregibon's (1981) approximation of Cook's distance measure is the most commonly recommended diagnostic (e.g., Fox, 2008, pp. 245, 412; Hardin & Hilbe, 2012, p. 49; Long, 1997, p. 101). The approximation does not require reestimating the model for every observation; instead it estimates how much each coefficient would change without the *i*th case in the analysis:

$$C_i = \left(\Delta_i \underline{\hat{\beta}}\right)' \widehat{Var}\left(\hat{\beta}\right)\left(\Delta_i \underline{\hat{\beta}}\right) \tag{1.18}$$

where

$$\Delta_i \underline{\beta} = \widehat{Var}\left(\underline{\hat{\beta}}\right)\underline{x_i}\frac{y_i - \hat{y}_i}{1-h_{ii}},\ \underline{x_i} = \text{column vector of predictor values for case } i,$$

$$\widehat{Var}\left(\hat{\beta}\right) = \text{variance} - \text{covariance matrix of } \hat{\beta}$$

and

$$h_{ii} = \hat{y}_i\left(1-\hat{y}_i\right)\underline{x_i}'\widehat{Var}\left(\hat{\beta}\right)\underline{x_i}$$

The usual criterion for defining a case as problematic is $C_i > \frac{4}{n-k-1}$. You should reestimate your model to decide if the excluded case(s) in fact change your results in a meaningful way and finalize your sample before interpreting the interaction effects.

Consequences of Model Misspecifications

Keep in mind that we never know with certainty the correct specification of the model. If the decision about whether or not to include an interaction specification in your model is not clear-cut, you need to make a decision balancing between two types of model misspecifications and their consequences. First, what happens to your model estimates and results if you include an interaction specification when in reality it is not needed, and second, what are the consequences when you exclude an interaction specification when in actuality it is needed. Last, an area of ongoing debate is the effect of unspecified heterogeneity on the estimation and interpretation of group differences—the relationship between the outcome and a categorical predictor—and how to deal with it. This issue is relevant to the interaction analysis if one or more of your interacting predictors is categorical, which is fairly common.

Including an Interaction Specification When Not Needed

This is a specific instance of the inclusion of irrelevant predictors of any kind in the model. In linear models, it is well-established that the consequence of including superfluous predictors is to increase the standard errors of the coefficients, but the coefficient estimates remain unbiased (Greene, 2008, p. 136). Thus, there is some loss of efficiency but not a problem of bias. The increase in coefficient standard errors also applies to GLMs. But if a GLM has a nonlinear link function, then the coefficient estimates are not necessarily unbiased. Many of the commonly used GLMs can identify the estimated coefficients only up to a multiplicative scale factor, usually the standard deviation of a latent interval outcome (e.g., logistic regression, probit analysis, ordinal regression models, multinomial logistic regression [MNLR]). As a number of authors have pointed out (Allison, 1999; Mood, 2010; Williams, 2009), the scaling factor increases as predictors are added to the model. Consequently, the estimates of all the coefficients are biased by the inclusion of a superfluous predictor—such as a product term for an interaction when it is not needed—because they are scaled by an incorrect factor. This situation is analogous to the problem of comparing coefficients across nested models discussed above, so the same solutions apply. For example, if you use *y*-standardized coefficients to interpret your results, then you remove the bias introduced by the scaling factor. The inefficiency of the estimates (increased standard errors) remains a concern regardless of the type of GLM model you estimate, but it is not always consequential in terms of significance testing and interpretation, particularly if you have a large sample size.

Excluding an Interaction Specification That Is Needed

This is a special case of the general problem of omitted variable bias that has a similar consequence for linear and nonlinear models. In linear models, the coefficients for the other predictors are biased and inconsistent unless the omitted variable is uncorrelated with the included predictors. (Note that a lack of correlation with the omitted variable would be unusual when the omitted variable is a product term of included predictors.) However, in GLMs with nonlinear link functions, the coefficients for the other predictors are always biased and inconsistent even if the omitted variable is uncorrelated with the included predictors (Greene, 2008, p. 787). And for those GLMs whose coefficients are multiplicatively scaled to an identifying factor, the scaling factor is also biased. Moreover, these biases affect the calculations used in a variety of techniques to calculate and probe the relationship of interacting or noninteracting predictors to the outcome, such as marginal and discrete effects, predicted outcomes, and odds ratios.

As with linear models, the potential consequences of excluding relevant interaction terms is worse than the consequences of including superfluous interaction terms for GLMs. That is, your main concern if you include an extraneous interaction term is that your standard errors will be too large, and you will have to be careful to properly take into account the scaling factor bias when interpreting coefficients—which is no easy task. In contrast, if you exclude a relevant interaction term, your coefficients and subsidiary calculations will definitely be biased in an unknown direction. The rational decision in this case would be to err on the side of inclusion of interaction terms when in doubt. The consequence of biased estimates of other coefficients is more certain if you incorrectly exclude the interaction terms than if you incorrectly include them.

Unspecified Heterogeneity and Group Comparisons

A key assumption for GLMs is that the variance function must be a function solely of the mean of the outcome (Hardin & Hilbe, 2012, p. 11). Implicit in this assumption is that there is no additional heterogeneity (heteroscedasticity) in the variance among the sample cases. For linear models, the presence of unspecified heteroscedasticity results in inefficient but unbiased (or consistent) coefficient estimates. The consequences are more serious in GLMs with nonlinear link functions if the unspecified heteroscedasticity is related to any of the model predictors, leading to biased and inconsistent coefficient estimates.

One situation in which heteroscedasticity related to the predictors is plausible, if not likely, is when comparing the outcome across social categories or social groups (Kaufman, 2013, p. 8); that is, when you use a categorical predictor, whether part of an interaction specification or not. This creates a problem akin to the comparison of coefficients across nested models (Allison, 1999; Mood, 2010; Williams, 2009). Making group comparisons is like comparing a coefficient between nested models; the comparison is confounded with the unspecified group heteroscedasticity. There is an ongoing debate over how to solve this problem and no resolution yet (Allison, 1999; Buis, 2010; Kohler et al., 2011; Kuha & Mills, 2018; Long, 2009, 2016; Mood, 2010; Williams, 2009, 2013). Some simulations have shown that using an incorrect heteroscedasticity specification is worse than not adjusting for heteroscedasticity at all (Keele & Park, 2005), while Kuha and Mills (2018) recently argued that the concerns are often irrelevant. Thus, as Williams (2013) put it, "At this point, it is probably fair to say that the descriptions of the problems with group comparisons may be better, or at least more clear-cut, than the various proposed solutions" (p. 11).

This issue reemphasizes the value of the diagnostic testing discussed above to check for problems in the fit of the model to the data and to make modeling changes accordingly. For example, Williams's (2010) reanalysis of Allison's (1999) example demonstrated that adding the square of a predictor corrected for the apparent heteroscedasticity and made moot any concern over confounding of group differences and heteroscedasticity in that analysis. This further points to the importance of models well-grounded in theory and formulated with careful consideration of the appropriate functional form of predictors.

ROADMAP FOR THE REST OF THE BOOK

Overview of Interpretive Tools and Techniques

My approach to understanding interaction effects involves producing and interpreting three types of information to understand the relationship between the outcome

and the interacting predictors. I detail the principles of this approach in Part I and apply them to different GLMs in Part II. I briefly describe the three sets of information here, labeling them with the names of the ICALC tool (command) that you use to calculate and create that information. For a paper or a presentation, you would usually include only a fraction of this information. The complete set is intended to help you fully understand the nature of the interaction effects and to provide you with different options for what you present and discuss. I use the heuristic device of declaring one of the interacting predictors as the focal variable and the other(s) as the moderating variable(s), but a proper interpretation should treat each interacting predictor in turn as the focal variable.

Defining the Moderated Effect of *F* With the GFI (Gather, Factor, and Inspect) Tool

The building block for understanding the nature of the interaction effect is to find the algebraic expression for the effect of the focal variable on the "modeled" outcome. You use this to describe and probe the basic structure of the focal variable's relationship to the outcome. With the GFI tool, you can determine if and when the focal variable's effect changes sign (positive to negative or negative to positive) as it varies with the values of the moderators. And you can create a visual representation of the algebraic expression with a path diagram–like graphic.

Calculating the Varying Effect of *F* and Its Significance: SIGREG (Significance Regions) and EFFDISP (Effect Displays) Tools

In probing the nature of the interaction effect, we are invariably interested in more than whether the moderated effect of *F* changes sign. Typically, we want to know the magnitude of the effect at different values of the moderator(s) and where that effect is significantly different from zero. The SIGREG tool finds, where possible, an analytic solution to define the range of moderator values for which *F*'s effect on the modeled outcome is significant. It also produces an empirically derived significance region table for which it can calculate effect values and significance for alternative types of effects when applicable—factor changes or any of the SPOST13 marginal effects calculated by its *mchange* command (see Long & Freese, 2014, pp. 166–171).

These tables are optionally saved to an Excel file with cell formatting to identify and highlight sign and significance changes. The EFFDISP tool creates visual counterparts to the significance region tables produced by SIGREG, plotting information about the varying magnitude and optionally the significance of the focal variable's moderated effect. Multiple plots are produced if there is more than a single moderator or if the focal variable is categorical with more than two categories.

The Predicted Outcome's Value Varying With the Interacting Predictors: The OUTDISP (Predicted Outcome Displays) Tool

Tables or graphs that display the predicted values of the outcome are probably the most familiar, and in some ways the most understandable, way to present and interpret interaction effects. Earlier in this chapter, I introduced the problem that such displays confound the nonlinearity of the interaction with the nonlinearity of the GLM link function, and I discuss this in detail in Chapter 5. I propose alternative visual displays to deal with this confounding that you can create using the OUTDISP tool.

Is the ICALC Toolkit Necessary?

Stata users might wonder whether the ICALC toolkit is really necessary given the capabilities of the *margins* and *marginsplot* commands in Stata for producing

information about interaction effects. Part of the reason I created ICALC is that the output created by two of the tools cannot be produced from the *margins* or *marginsplot* commands. GFI provides basic information about the pattern of the focal variable's effect, while SIGREG identifies significance regions for the focal variable's effect. Some of the basic output from ICALC for creating the content of tables of the moderated effect (SIGREG), graphics of the conditional effects (EFFDISP), or predicted outcome values (OUTDISP) can in some form be produced by the *margins* or *marginsplot* commands. But the ICALC toolkit provides types of tables, graphs, and options not available through the Stata commands.

Moreover, it reports results that are more compact and simpler to read and uses an interface that is hopefully more user-friendly in much the same way as are the post-estimation results produced by SPOST13.[5] A major advantage of the ICALC toolkit for readers of this book is that the examples and applications throughout the book use ICALC to produce the results, with annotated explanations of the use of the commands in the application chapters.

For analysts who prefer platforms other than Stata for producing tables and graphics, the ICALC toolkit includes options to save the necessary results to an Excel file that can be easily imported into other applications. It will save formatted tables that users can reformat directly in Excel or copy and paste the Excel table into another application. For graphics, ICALC saves the data values used to create the Stata figures in a rectangular data format—a column for the *x*-axis values/categories, one or more columns for the corresponding *y*-axis values for each graphed data series, and columns for the variables defining separate graphics for subsamples (if any).

Organization and Content of Chapters

Part I (four chapters) describes and explains the principles of interpretation. Part II (seven chapters) provides detailed applications of the principles and the ICALC toolkit to a set of commonly used GLMs and ends with a discussion of extensions.

Part I: Principles

Chapter 2: Basics of Interpreting the Focal Variable's Effect in the Modeling Component. This chapter focuses on the derivation and interpretation of the algebraic expression for the moderated effect of the focal variable on the modeled outcome; that is, *F*'s effect contingent on one or more moderators. The emphasis is on a holistic interpretation of the moderated effect rather than interpreting the component coefficients. It introduces a simple mathematical analysis to determine if and under what conditions (values of the moderators) the sign of *F*'s effect changes from positive to negative, or vice versa. And it demonstrates the use of plots of the moderated effect of *F* against the predictors to help understand and interpret the changing sign and magnitude of *F*'s effect.

Chapter 3: The Varying Significance of the Focal Variable's Effect. In this chapter, I discuss how to define the significance region for the moderated effect of *F*. That is, three ways to determine the ranges of the moderators' values for which *F*'s effect is statistically significant are as follows:

- Johnson-Neyman analytically defined boundary values, which mark the boundaries between significance and nonsignificance

- Empirically derived significance region tables reporting *F*'s effect and its significance for user-specified moderator values
- Plots of *F*'s effect against the moderators with confidence intervals demarcating significance regions

Depending on the specific GLM, the last two approaches can be applied to various effects calculated from the moderated effect of *F*, such as factor changes or discrete changes.

Chapter 4: Linear (Identity Link) Models: Using the Predicted Outcome for Interpretation. This chapter and the next cover the nitty-gritty of how to create and interpret tables and plots of the predicted outcome to understand and explain the pattern of the relationship of the interacting predictors with the outcome. I discuss the simpler case of GLMs with an identity link function (e.g., OLS regression) in this chapter to lay a foundation for the more complicated application to nonlinear link functions in Chapter 5. Topics include options for choosing display and reference values for the interacting predictors, reference value options for noninteracting predictors, and the use and interpretation of predicted value tables, charts, and plots.

Chapter 5: Nonidentity Link Functions: Challenges of Interpreting Interactions in Nonlinear Models. This chapter continues the discussion of the use of predicted outcome values. I detail why the approach to interpreting predicted values for identity link functions presented in Chapter 4 is potentially problematic for nonlinear link functions. I revisit the topic of how to choose reference values for other predictors, and then describe and demonstrate the use of three options for handling these problems when creating and interpreting a predicted values table or graphic.

Part II: Applications

Chapter 6: ICALC *Toolkit: Syntax, Options, and Examples.* In this chapter, I describe the ICALC toolkit, explain the syntax and options for using the tools, and provide brief examples of how to apply the five ICALC commands. The INTSPEC command must be run before any of the other commands to save and set up information about the nature of the interaction specification and display/reference values of the predictors. The other four commands do the calculations and create tables and graphics corresponding to the approaches to interpretation described in the Part I principles chapters: gather, factor, and interpret (GFI), significance regions (SIGREG), effect displays (EFFDISP), and outcome displays (OUTDISP).

Chapters 7 to 11: Applications to Frequently Used GLMs. Each technique-of-analysis chapter has the same basic structure. It begins with an overview of the types of data situations for which the technique is typically used and brief descriptions of published examples from multiple disciplines (sociology, political science, economics, criminology, and public health). I chose the examples to illustrate the variety of actual applications, concentrating on publications in disciplinary flagship journals. For the most part, I chose the first examples that I found using search engines, excluding those with egregious errors of estimation or interpretation. This overview is followed by a discussion of the technique's properties as a GLM, the relevant ways to interpret coefficients, technique-specific diagnostics tests, and a brief description of the data used for the application examples. The bulk of each chapter demonstrates the application of the ICALC commands to different specifications of interaction effects and how to interpret the results and output for that specification. Each chapter has at least one

single-moderator empirical example and one multiple-moderator or three-way interaction empirical example. The interpretations always treat each interacting predictor in turn as the focal variable. Many of the chapters conclude with a special topics section. The organization of the application chapters is as follows:

- Chapter 7: Linear Regression Model Applications
- Chapter 8: Logistic Regression and Probit Applications
- Chapter 9: Multinomial Logistic Regression Applications
- Chapter 10: Ordinal Regression Models
- Chapter 11: Count Models

I have organized the application chapters in the order in which I think they should be read. Certainly, readers should first review Chapter 6 on the use of the ICALC toolkit, and then read Chapter 7 on OLS regression even if they do not intend to use it because there is foundational material woven into the chapter. In a similar vein, readers should be familiar with the applications in Chapter 8 before they read about multinomial logistic or ordinal probability models in Chapters 9 and 10.

Table 1.3 presents a simple rubric for knowing when to consider using each of the GLM models in Chapters 7 to 11. This rubric uses the level of measurement of the outcome—nominal/categorical, ordinal, and interval/ratio—and the number of categories (values) in the outcome measure to identify when to choose each technique. There is additional discussion in each chapter of the circumstances and diagnostic tests that might preclude using these recommendations of a technique.

Chapter 12: Extensions and Final Thoughts. In this chapter, I provide very brief extensions to the use of ICALC for the interpretation of interaction effects in three additional analytic techniques—Tobit analysis, the Heckman selection model, and the Cox proportional hazards model for survival analysis—and the interpretation when one of the interacting predictors is a quadratic function. I end with a brief reminder about important dos and don'ts and cautions in specifying and interpreting interaction effects.

TABLE 1.3 ⬢ RUBRIC FOR CHOOSING GENERALIZED LINEAR MODEL TECHNIQUE

Technique	Type of Outcome Measure
Linear regression	Interval or ratio variable
Binomial logistic regression and probit analysis	Categorical with two categories
Multinomial logistic regression	Categorical with three or more unordered categories
Ordinal regression models	Categorical with three or more ordered categories
Count models	Nonnegative integers representing a count of events

CHAPTER 1 NOTES

1. See Section "Mathematical (Geometric) Foundation for GFI" in Chapter 2 for a more detailed discussion.
2. Suppose you are using a significance level of .05. The reason to correct the significance level is that the probability that one or more coefficients will be significant is greater than .05 and increases with the number of multiple tests.
3. For some estimation methods, the Wald test is an $F_{1,df}$ test statistic and the single-coefficient test is a t_{df} statistic. Because $t^2_{df} = F_{1,df}$, you get the same test result from either. Similarly, for other estimation methods, the Wald test is a $\chi^2_{(1)}$ test statistic and the single-coefficient test is a z statistic. Because $z^2 = \chi^2_{(1)}$, you again get the same test result. Given this equivalence, some sources, including Stata, also identify the z and t tests of coefficients as Wald tests.
4. The Oswego City School District Regents Exam prep site has a great explanation of the inverse of a function (http://www.regentsprep.org/regents/math/algtrig/atp8/inverselesson.htm):

   "A function and its inverse function can be described as the 'DO' and the 'UNDO' functions. A function takes a starting value, performs some operation on this value, and creates an output answer. The inverse function takes the output answer, performs some operation on it, and arrives back at the original function's starting value."

5. Long and Freese (2014) provide a similar argument for a parallel issue raised concerning SPOST13 and give an example in which SPOST13 produces 50 lines of well-formatted output compared with more than 1,500 lines of output from the corresponding *margins* command (http://www.indiana.edu/~jslsoc/web_spost13/sp13_whymstar.htm).

PRINCIPLES

PART I