

Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects

Trenton D. Mize

Purdue University

Abstract: Many effects of interest to sociologists are nonlinear. Additionally, many effects of interest are interaction effects—that is, the effect of one independent variable is contingent on the level of another independent variable. The proper way to estimate, interpret, and present these two types of effects individually are well known. However, many analyses that combine these two—that is, tests of interaction when the effects of interest are nonlinear—are not properly interpreted or tested. The consequences of approaching nonlinear interaction effects the way one would approach a linear interaction effect are severe and can often result in incorrect conclusions. I cover both nonlinear effects in the context of linear regression, and—most thoroughly—nonlinear effects in models for categorical outcomes (focusing on binary logit/probit). My goal in this article is to synthesize an evolving methodological literature and to provide straightforward advice and techniques to estimate, interpret, and present nonlinear interaction effects.

Keywords: interaction effects; nonlinearities; categorical models; logit/probit

MANY relationships of interest to sociologists are nonlinear. Even within the context of the linear regression model, nonlinearities of the effects are common, for example, due to polynomial specifications of independent variables, such as both *age* and *age*² being included in the model or due to transformations of the dependent variable (e.g., the log of wages). More commonly, most models for categorical dependent variables (e.g., binary, nominal, and ordinal logistic and probit regression models) produce nonlinearities in the predicted probability metric.

Many relationships of interest to sociologists are also interactive in nature. Interaction effects exist when the effect of one independent variable is contingent on the level of another independent variable. For example, does the effect of obtaining more education depend on whether someone is white or black? Does the effect of aging depend on whether someone is a man or a woman?

The premise of this article is that although correct procedures for modeling nonlinear effects are generally well established and commonly practiced and that correct procedures for modeling and interpreting *linear* interaction effects are also well established and commonly practiced, analyses that combine nonlinearities and interaction effects are often estimated, interpreted, and presented incorrectly in substantive work. Although the methodological literature on nonlinear interaction effects has advanced rapidly in the last 15 years, there has been far too little change in the way most social scientists approach these types of analyses. In their assessment of the economic literature examining nonlinear interaction effects, Ai and Norton (2003) found that “of the 72 articles published between 1980 and 1999 ... none of the studies interpreted the coefficient on the interaction term correctly.” My reading of

Citation: Mize, Trenton D. 2019. “Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects.” *Sociological Science* 6: 81-117.


Received: December 18, 2018

Accepted: December 27, 2018

Published: February 6, 2019

Editor(s): Jesper Sørensen, Olav Sorenson

DOI: 10.15195/v6.a4

Copyright: © 2019 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

the substantive sociological literature does not reveal a more optimistic take of our discipline's recent work.

To confirm this intuition, I examined the 53 articles published in the *American Sociological Review* (ASR) between 2004 and 2016 examining nonlinear interaction effects.¹ Fifty of the 53 referred only to the coefficient on the product term to determine the significance of the interaction effect—an improper test of interaction in terms of the predicted probabilities.² Perhaps because of this fact, the current ASR editors recently published “A Few Guidelines for Quantitative Submissions,” with the editors noting the difficulty of testing for interaction in models for categorical dependent variables, concluding, “The case is closed: Don't use the coefficient on the interaction term to draw conclusions about significance of statistical interaction in categorical models such as logit, probit, Poisson, and so on” (Mustillo, Lizardo, McVeigh 2018:1282). However, despite the definitiveness of this statement about the wrong way to test for interaction, the *correct* way has not been given a thorough treatment aimed at improving the practices of applied researchers.

My primary goal in this article is to provide a set of straightforward best practices for substantive researchers who are interested in estimating, interpreting, and presenting interaction effects when the effects of interest are nonlinear. Methodological advances have provided important frameworks that should guide future work testing nonlinear interaction effects; however, it is clear that these advances have not had much of an impact on most applied sociological research. My goal is to provide a relatively nontechnical overview of the problem and of the current state of best practices to deal with these issues.

Many of the best practices I advocate for are available in the methodological literature across sociology, political science, and economics. My goal is to compile and synthesize these methodological advances in a way that is helpful for applied researchers and that provides straightforward and helpful recommendations for how these methodological insights can be implemented in applied research. Those looking for more statistically oriented overviews of the issues described here can find important methodological progress in the work of Ai and Norton (2003); Berry, DeMerrit, and Esarey (2010); Rainey (2016); Long and Mustillo (2018); and others cited throughout this article.

Scope and Definition of Terms

Definition of Terms

In this article, I define an interaction effect as an effect of one independent variable being contingent on the level of another independent variable. I distinguish between the *coefficient of the product term* and the *interaction effect* itself. As I will show, the coefficient on the product (interaction) term is often a misleading representation of the interaction effect in terms of the predictions. Thus, throughout, any references to *effects* are in terms of the model's predictions.³

Linear effects are those that are constant across the range of a variable (e.g., a change from 10 to 12 years of education has an effect equivalent to a change from 18 to 20 years of education). *Nonlinear effects* are those that are not constant

across the range of an independent variable. Effects can be nonlinear because of the specification of independent variables (e.g., age and age^2) or because of the nature of the dependent variable (e.g., a categorical outcome or a transformation of a continuous outcome [e.g., logged wages]) or both. Note that effects from linear regression models can be nonlinear.

Different approaches (especially visualizations) are more or less effective depending on the measurement level of the independent variables that are the constituent pieces of the interaction effect. My terminology follows how the variables are entered into the regression model, which may or may not directly align with their true measurement level. *Nominal* variables have multiple discrete categories that are not ordered; *binary* variables are a special case when there are only two categories. *Continuous* variables are those that are treated as interval-ratio.

Testing for Interaction in the Natural Metric of the Dependent Variable

The methods I advocate for in this article make one key assumption: The goal of the analysis is to determine whether an interaction effect exists in terms of the natural measured metric of the dependent variable (for example, that the interest is in the effect on wages even when the log of wages is used as the dependent variable). Most commonly, a distinction between the measured and modeled metric occurs in models such as binary, nominal, and ordinal logit/probit models (Kaufman 2018). In this article, my assumption is that the metric of interest from these models is the predicted probabilities—not the odds, log odds, or latent variable metric (for details, see Long 1997; Kuha and Mills 2018). A brief example helps motivate my focus on predicted probabilities as the natural metric of the dependent variable in, for example, binary logit. Consider the various ways to summarize the hypothetical data in Table 1 on gender and voting.

The dependent variable (*voted in last election*) is measured as a binary (voted = 1; did not vote = 0). When examining the effect of gender, the observed metric in the data are proportions: 40 percent of men voted; 60 percent of women voted. The probability that any given man in the sample voted is 0.40; the probability that any given woman in the sample voted is 0.60. Summarizing the effect of gender in the natural metric of the dependent variable would be calculating the differences in these two predicted probabilities ($0.60 - 0.40 = 0.20$). It is possible to convert the probabilities to odds ($odds = \frac{\pi}{1-\pi}$), as row 3 of Table 1 does. Taking the ratio of women's odds of voting to men's odds of voting ($OR = \frac{Odds_{women}}{Odds_{men}}$) produces an odds ratio (OR): Women's odds of voting are larger than men's by a factor of 2.25. Further, we could take the natural log of men's and women's odds and then subtract the two, as the bottom row of Table 1 demonstrates. Indeed, the OR reported above is what you would obtain from exponentiating the coefficient from a binary logit model applied to these data regressing voting on gender (with the difference in log odds being the raw regression coefficient).

The rest of this article demonstrates how to test for interaction effects in the natural metric of the dependent variable: in logit/probit models, the predicted probabilities. I am far from the first to suggest that interpreting the models in this

Table 1: Various ways to summarize the effect of gender (woman = 1) on voting.

	Men	Women	Formula for Effect	Effect of Gender
Proportion	0.40	0.60	$p_{women} - p_{men}$	0.20
Probability	0.40	0.60	$\pi_{women} - \pi_{men}$	0.20
Odds	0.40 / 0.60 = 0.67	0.60 / 0.40 = 1.50	$odds_{women} / odds_{men}$	2.25
Log Odds	-0.41	0.41	$\ln(odds)_{women} - \ln(odds)_{men}$	0.81

metric carries key advantages (e.g., Long 1997; King, Tomz, and Wittenberg 2000; Mood 2010).

Outline of Article

First, I overview the marginal effects framework for summarizing effects in terms of a model's predictions. Next, I illustrate the difficulties of testing nonlinear interaction effects even in the context of the linear regression model. I then spend some time demonstrating why testing for interaction in binary logit/probit requires the techniques advocated for in this article—and why the coefficient on the product term is *not* a test of interaction in terms of the predicted probabilities. Next, I work through a series of applied examples for presenting and testing nonlinear interaction effects that I hope serve as a guide for researchers. I end with a discussion of software considerations and with details for extending the framework advocated for here to nominal, ordinal, and count models.

Marginal Effects As Summaries of Effects in the Natural Metric of the Dependent Variable

Marginal effects summarize an independent variable's effect in terms of a model's predictions (see Long and Freese 2014 for an overview). Marginal effects have several advantages over relying on regression coefficients to summarize an independent variable's effect: (1) They allow for one summary measure of an independent variable's effect even when multiple linked coefficients are in the model (e.g., *income* and *income*²). (2) They avoid the problematic identification (scaling) issues of the coefficients in logit/probit-based models (see Long 1997; Long and Freese 2014; Breen, Karlson, and Holm 2018). (3) As they are based on a model's predictions, they can be expressed and interpreted in a different metric than the regression coefficients.

To explain the various types of marginal effects, I begin with a generic form of a regression model:

$$\eta(\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}). \quad (1)$$

In this notation, \mathbf{x} is a vector of independent variables and $\boldsymbol{\beta}$ is a vector of regression coefficients. Throughout, I use η to denote a prediction that is some function (G) of $\mathbf{x}\boldsymbol{\beta}$. In the linear regression model, $\eta = \hat{y} = \mathbf{x}\boldsymbol{\beta}$. A key benefit of using predictions and marginal effects to summarize a model is the ability to transform $\mathbf{x}\boldsymbol{\beta}$ into a more useful metric when applicable. For example, if the dependent variable is $\ln(y)$ in a linear regression model, the predictions can easily be transformed back into the original y metric by exponentiating the predictions ($\eta = e^{\mathbf{x}\boldsymbol{\beta}}$). In binary logit, $\mathbf{x}\boldsymbol{\beta}$ is in terms of log odds; we can easily transform the predictions into the metric of predicted probabilities with the formula $\eta = \Pr(\mathbf{y}=\mathbf{1}) = \frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1+\exp(\mathbf{x}\boldsymbol{\beta})}$.

The notation $\eta(\mathbf{x})$ demonstrates that the predictions are contingent on certain values of the independent variables in \mathbf{x} . I use the more detailed notation of $\eta(x_k = x_k^*, \mathbf{x} = \mathbf{x}^*)$, where independent variable x_k is the focal independent variable and the variables in \mathbf{x} are other independent variables, such as the control variables.

At its simplest, a marginal effect can be calculated as the difference between two predictions, with only x_k changing and the variables in \mathbf{x} being held constant. However, there are multiple choices as to what values of both x_k^* and \mathbf{x}^* to use as well as a choice of whether to use representative values to generate only two predictions for comparison or to average over multiple sets of predictions. I outline the various options and discuss some pros and cons of each below.

Marginal Effects at the Mean (MEM) or at Other Representative Values (MER)

One option when using marginal effects is to choose values for both x_k^* and \mathbf{x}^* that are representative of values of interest:

$$MER_{x_k} = \eta(x_k = end, \mathbf{x} = \mathbf{x}^*) - \eta(x_k = start, \mathbf{x} = \mathbf{x}^*). \quad (2)$$

The marginal effect indicates the change in the value of η as the focal variable x_k changes from some starting to some ending value. For example, the mean of x_k (\bar{x}_k) and $\bar{x}_k + 1$ are common choices. Any starting and any ending value can be used; commonly, for continuous independent variables, the mean is used as the starting value and increases of one unit or a standard deviation are used to determine the ending value. I focus on *discrete changes*, which are changes in x_k of a fixed value (e.g., +1, +SD, +15, etc.). Instantaneous changes are also possible, which represent the first derivative at a given value of x_k ; however, they are not the focus here. Importantly, if a model includes multiple linked coefficients, such as those due to polynomials or product terms, each associated variable must change at the same time when calculating the marginal effect of x_k (e.g., $income^2$ cannot be held constant while $income$ changes).

For a binary focal variable x_k , the ending value is always 1 and the starting value always 0:

$$MER_{binary} = \eta(x_k = 1, \mathbf{x} = \mathbf{x}^*) - \eta(x_k = 0, \mathbf{x} = \mathbf{x}^*). \quad (3)$$

For example, the effect of gender is the difference in the prediction for women (= 1) and the prediction for men (= 0).

The values in \mathbf{x}^* are commonly held at their sample means, leading to the terminology of a “marginal effect at the mean” (MEM) and an interpretation as the effect for the “average person” in the sample (terminology is adapted from Williams 2012; Long and Freese 2014). However, other values of interest can be used, such as the means for specific subsamples, for example, using the subsample means for those with college degrees when predictions about individuals with college degrees are of interest and using the subsample means for those without college degrees for those predictions.⁴

An awkward aspect of the MEM with \mathbf{x}^* held at the sample means is that this “average person” may not be represented in the sample; for example, no observation may have the mean level of income, mean level of education, and mean level of age. In addition, nominal variables are held at their sample proportion, with, for example, the awkward interpretation of making a prediction at woman = 0.55.⁵ Theoretically, it is more common for the effect *on average* across the sample to be of primary interest rather than the effect for the *average person* (Hanmer and Kalkan 2013). Average marginal effects, which are discussed in the next section, represent an effect on average across the sample.

Average Marginal Effects (AME)

Average marginal effects are estimated by calculating a marginal effect (ME) for every observation in the sample and then averaging these effects. First, consider the marginal effect of x_k for a specific observation i :

$$ME_{x_k} = \eta(x_k = end_i, \mathbf{x} = \mathbf{x}_i) - \eta(x_k = start_i, \mathbf{x} = \mathbf{x}_i). \quad (4)$$

The subscript i 's are used to indicate that all three of the (1) starting, (2) ending, and (3) control variable values can be unique to each observation. An average marginal effect (AME) is the average (mean) of the marginal effects calculated for each observation in the sample:

$$AME_{x_k} = \frac{1}{N} \sum_{i=1}^N \eta(x_k = end_i, \mathbf{x} = \mathbf{x}_i) - \eta(x_k = start_i, \mathbf{x} = \mathbf{x}_i). \quad (5)$$

Equation 5 represents the average effect across the entire sample (all N observations). However, in the context of interaction effects, it is often of interest to calculate an average effect across a subsample, for example, an AME for men and an AME for women for which only the applicable observations are used in the calculation of each.⁶ For AMEs, the most common starting value is the observed value for that observation ($start_i$). As with MERs, for continuous independent variables, discrete changes of increases of 1, a standard deviation, or any other value can be used to determine end_i . For binary variables, predictions are made at 0 and 1 for each observation.

An advantage of the AME is that the observed values of the independent variables are used for each observation-specific prediction; that is, each prediction is based on actual observed values in the data. The AME can be interpreted as the

effect of changing x_k by a given value on average across the sample. In addition to this elegant interpretation, AMEs also have some statistical advantages over MEMs (Hanmer and Kalkan 2013; see also Cameron and Trivedi 2005), although they rarely differ dramatically in practice. For these reasons, my interpretations for the examples used in this article are based on AMEs.

Despite my preference for AMEs as a default choice, the correct choice as to whether to use MERs, MEMs, or AMEs should be guided by which best tests the substantive research question. In addition, whether to use the entire sample or a subsample for the calculations of each marginal effect is best guided by the substantive research question. Long and Freese (2014:280–4) provide additional guidance for these decisions (in particular, see their discussion of global vs. local means). Long and Mustillo (2018) provide additional insights about these decisions in the context of testing for group differences.

Testing the Equality of Marginal Effects: Second Differences

Testing for interaction effects involves testing the equality of multiple effects. For example, when examining possibly interactive effects of the independent variables *age* and *gender*, we may wish to calculate the difference in the effect of age for men and the effect of age for women. For linear effects in linear regression, the coefficient of the product term *ageXgender* provides this test. For nonlinear effects and/or effects in a different metric than the coefficients, tests of the equality of marginal effects can be used. For example, consider a binary logit model regressing *voting* behavior on *age*, *gender*, and an *ageXgender* product term.

Let $\Delta_{age_{men}}$ represent an AME of age for men and $\Delta_{age_{women}}$ be the AME of age for women. Commonly, these are referred to as first differences. A test of second difference is a test as to whether two first differences are equal (Berry et al. 2010; Long and Freese 2014). A Wald test can be used to determine whether the two effects are equal:

$$z = \frac{\hat{\Delta}_{age_{women}} - \hat{\Delta}_{age_{men}}}{\sqrt{\hat{\sigma}_{age_{women}}^2 + \hat{\sigma}_{age_{men}}^2 - 2\hat{\sigma}_{age_{women}, age_{men}}}}. \quad (6)$$

The numerator of Equation 6 represents the difference in the effect size across men and women; the denominator represents the standard error of the difference. Here, $\hat{\sigma}_{k_b}^2$ is the variance estimate of each marginal effect ($\hat{\sigma}_{k_b}$ is the standard error) and $\hat{\sigma}_{k_b=1, k_b=0}$ is the estimate of the covariance between the two effects. The value of z can then be compared to the critical value to determine whether the difference is statistically significant (i.e., the null hypothesis can be rejected).

Although the variances and covariances of regression coefficients are obtained from the variance/covariance matrix of the regression estimates, for marginal effects, they must be calculated post-model estimation. Commonly, the delta method is used (Ai and Norton 2003; Agresti 2013; Dowd, Greene, and Norton 2014; Pitblado 2014), although other methods, such as bootstrapping (Efron and Tibshirani 1994; Dowd et al. 2014) and simulation (King et al. 2000), can also be used.

Nonlinear Interaction Effects in the Linear Regression Model

In this section, I outline how marginal effects and tests of second differences can be used to test for interaction effects in linear regression when the relationship of interest is nonlinear. Consider the relationship between wages (dependent variable), age, and gender. Data for this example come from the General Social Survey, including only employed individuals; typical control variables for respondent demographics are included but not shown.⁷ Age likely has a nonlinear relationship with wages. For example, it is possible that as a person ages his or her wages tend to increase, but this positive effect of age diminishes at older ages. Indeed, in models not shown, I found that including both *age* and *age*² improves the fit of the model (suggesting a nonlinear relationship). We also know from prior work that there is a gender gap in hourly wages. Of interest here is whether the effect of age differs for men and women. To examine this question, I fit a linear regression model regressing hourly wages on *age*, *age*², *woman*, *womanXage*, and *womanXage*². Note that all constituent terms are needed (Brambor, Clark, and Golder 2006).⁸ Note also that although there has been discussion in the literature about the pros and cons of mean-centering continuous variables, such as age, before including them in interaction models, doing so does not affect the predictions or tests of the marginal effects (Kromrey and Foster-Johnson 1998; Dalal and Zickar 2012); none of the models presented in this article mean center any variables.

When estimating a model such as this, I tend to ignore the coefficient estimates because they do not provide any straightforward summary of the effects of interest. Many questions may be of interest regarding the interaction effect that the coefficients alone cannot answer: They do not answer what the effect of age is on average for men, or for women, or whether those effects differ. They do not tell us *where* (if anywhere) across the range of age there are significant gender differences. It is possible that gender differences only exist at certain values of age; alternatively, it is possible that gender differences exist across all levels of age but that the *magnitude* of the difference between men and women is different at various ages. The coefficients do not answer these question for us. Tests of marginal effects, however, can.

The first thing I recommend when examining nonlinear interaction effects is to make a plot of the model's predictions for the focal independent variables. The plot should then guide further interpretation and testing. Figure 1 presents predicted hourly wages across the range of age separately for men and women. Note that as with calculations of marginal effects, the predictions themselves can be made using group-specific observations or means or with the entire sample observations or means; for this example, I have used group-specific observations: I have made the predictions for women using only the women in the sample (and the observed values of their control variables) and the predictions for men using only the men in the sample.

The relationships shown in Figure 1 suggest that getting older is associated with higher wages, but eventually, the additional wage boost of getting even older begins to diminish. The key here beyond the specific pattern for this example is: The effect of age is not constant across the range of the variable; the effect of age

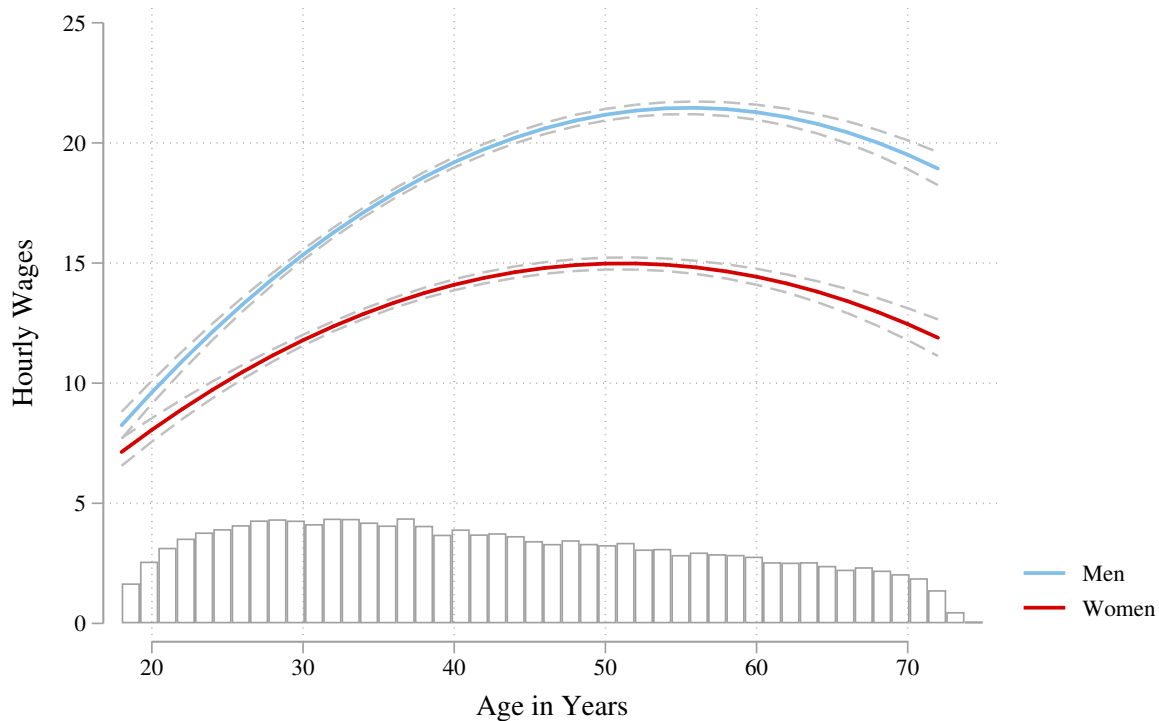


Figure 1: Predicted hourly wages based on age and gender: curvilinear effect of age and interaction effect between age and gender.

depends on which part of the age range we examine. Further, as we can see from Figure 1, the effect of age also depends on whether someone is a man or a woman.

Table 2 shows the results for the effects of age calculated at three different values of age for men and women separately. The top of the table shows that the average marginal effect of an increase in age from 25 to 30 years for men is about a \$2.628 increase in wages ($p < 0.01$), whereas the AME for women is an increase of \$1.688 ($p < 0.01$). We can test the equality of these two AMEs with a test of the second difference: $2.628 - 1.688 = 0.940$, which is significant at the $p < 0.01$ level. Substantively, this indicates that getting older has a significantly larger effect for men than for women if the difference of interest is how changes from 25 to 30 years old are associated with wages. That is, this statement is limited only to the specific level of age at which the second difference was calculated.

In contrast, consider how changes from 65 to 70 years old are associated with wages. For men, this increase in age has a significant negative effect on wages ($\Delta_{men} = -1.118$; $p < 0.01$), and it has a similar negative effect for women ($\Delta_{women} = -1.165$; $p < 0.01$). Here, the second difference indicates that the effects of changes from 65 to 70 years old do not differ for men and women ($-1.165 - -1.118 = 0.046$; $p = \text{not}$

Table 2: Results for how wages are associated with age and gender: tests of average marginal effects (AMEs) and second differences ($N = 30,931$).

	AME_{men}	AME_{women}	Second Difference
<i>Effect of Age</i>			
25 → 30	2.628* (0.081)	1.688* (0.085)	0.940* (0.111)
65 → 70	-1.118* (0.109)	-1.165* (0.120)	0.046 (0.160)
start _{<i>i</i>} + 5	1.149* (0.035)	0.576* (0.035)	0.574* (0.047)

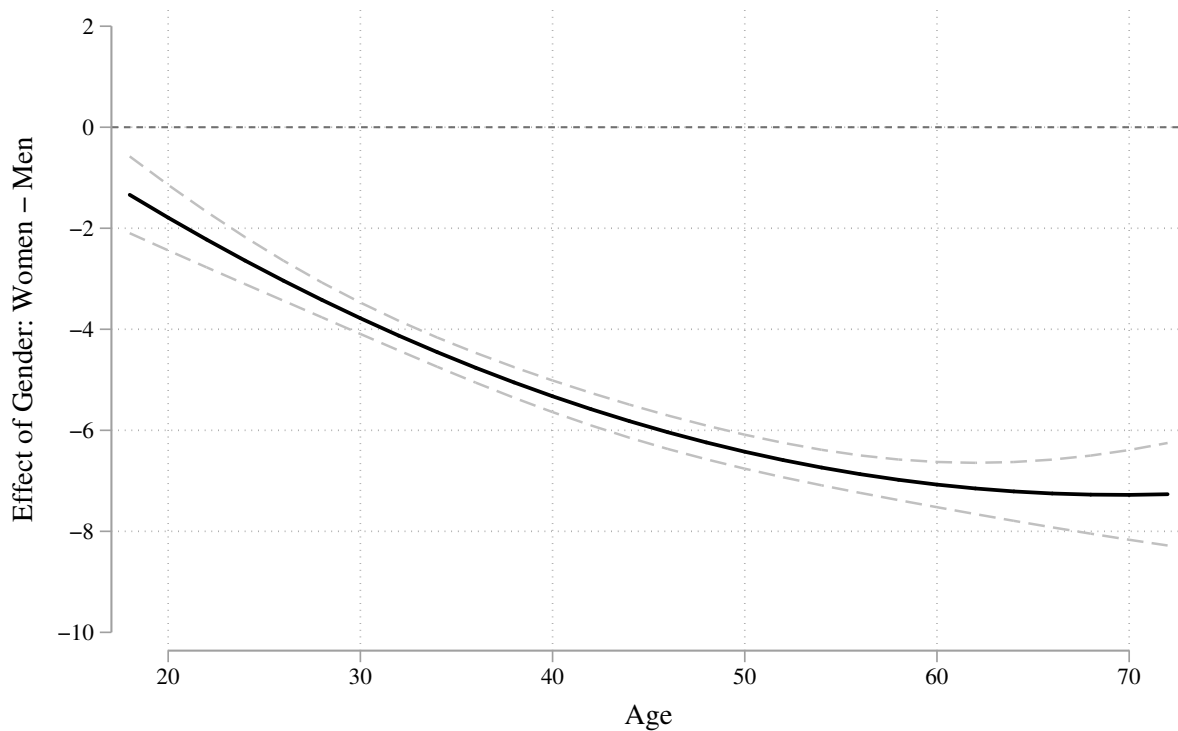
Notes: Standard errors are in parentheses. * $p < 0.05$, two-tailed tests.

significant [ns]). That is, there are no gender differences in the effect of increases in age at these higher levels of age.

When specific values are of substantive or theoretical interest, it is important to test for interaction at those values (e.g., if the research question is about gender differences for men and women in their 20s, the interaction effect should be calculated for those ages). Absent specific values of interest—or in the common case in which a statement about whether an interaction effect exists or not on average across the sample—comparisons of average marginal effects or marginal effects at the means across men and women are recommended. The final row of Table 2 reports such a test for the average effect of a five-year increase in age. On average, a five-year increase in age is associated with a 1.149 increase in wages for men and a 0.576 increase in wages for women; the second difference $1.149 - 0.576 = 0.574$ ($p < .001$) indicates that the average effect of aging five years is larger for men than for women. In other words, an interaction effect between age and gender does exist, on average, in this sample.

Examining Both Sides of the Interaction

Importantly, there are two sides to the interaction effect, and best practice is to examine both sides (Berry, Golder, and Milton 2012). The above paragraph describes the effect of age. However, there are also important insights about the effect of gender and how it varies across the range of age. Figure 1 suggests the gender gap in earnings is relatively small at young ages but increases in size over the range of age. It appears that the 95 percent confidence intervals do not overlap at any age. Here it is worth noting that when confidence intervals *do not* overlap across independent groups, this indicates a significant group difference. However, assessing significance via confidence intervals alone is a risky enterprise. When confidence intervals *do* overlap, this may or may not indicate that the group difference is nonsignificant across independent groups. Moreover, in the presence of clustering or other sources of nonindependence, the “confidence interval overlapping” rule will often fail (Belia et al. 2005). The only way to know for sure whether group differences (e.g., men vs.



Effects below zero indicate women earn less than men at that age

Figure 2: Average marginal effect of gender (woman = 1) across the range of age on predicted wages.

women) are statistically significant is to explicitly test for this difference. Marginal effects provide one useful way to do so.

Figure 2 plots the average marginal effects of gender (predicted $wages_{women}$ – predicted $wages_{men}$) across the range of age. As this is a direct test of the effect of gender, the confidence intervals on the test provide appropriate information to judge the significance of group differences. As is clear from the plot, gender differences are significant across the range of age, with women always being predicted to earn less than men; the gender gap starts off relatively small and increases in size at older ages. For example, the effect of gender at age 55 years ($AME_{woman} = -6.805$; $p < 0.01$) is significantly larger than the effect of gender at age 25 years ($AME_{woman} = -2.841$; $p < 0.01$); the second difference = $-6.805 - -2.841 = -3.964$ ($p < 0.01$).

More Contexts for Nonlinearity in Linear Regression

In this section, I illustrated that even in linear regression models, when the effect of interest is nonlinear, it is useful to use marginal effects and tests of second differences at specific values of interest in order to determine whether an interaction effect exists. There are many other situations in which effects might be nonlinear even in the context of the linear regression model. For example, a model in which

a continuous dependent variable has been transformed (e.g., wages have been logged) will produce effects in terms of the original metric (e.g., wages) that are nonlinear. In addition, other flexible methods beyond simple quadratic terms (e.g., age and age^2 in the above example) can be utilized on the right-hand side of the model to produce nonlinearities in terms of the effect on the predictions. The key point is that just because linear regression is being used to estimate the model does not ensure that the effects of interest are linear.

Nonlinearities in Models for Categorical Dependent Variables

Next, I turn to the necessity of taking the approach outlined above when binary, nominal, ordinal, or count models are used. In contrast to linear regression models, the relationships in models for categorical dependent variables are almost always nonlinear in the natural metric of the dependent variable (e.g., the predicted probabilities in binary logit [Long 1997; Agresti 2013]). For the following section, I focus only on binary logistic regression for simplicity, but the difficulties I note and solutions I recommend apply equally to other models, such as binary probit, ordinal logit, multinomial logit, count models (such as Poisson and negative binomial regression), and others (I discuss these models briefly at the end of this article).

Nonlinearities in Binary Logit

To illustrate the nonlinearities inherent in binary logit and the limitation of using the coefficients to summarize effects, consider the following example using data I simulated. Because the data are simulated, I have specified that the true model is:

$$\ln\left\{\frac{\Pr(y_i = 1)}{1 - \Pr(y_i = 1)}\right\} = -1 + 0.2c + 2b. \quad (7)$$

I have c as a continuous independent variable and b as a binary independent variable. The predictions for $\mathbf{x}\beta$ are in logits, or log-odds units. To transform the predictions from the log-odds metric to the predicted probability metric, we use the formula $\eta = \Pr(\mathbf{y} = 1) = \frac{\exp(\mathbf{x}\beta)}{1 + \exp(\mathbf{x}\beta)}$. Figure 3 shows the effect across the range of c in the predicted probability metric. This demonstrates the classic s-shaped curve that results from the cumulative distribution function of the logistic distribution (Long 1997). Recall that there is only one coefficient estimate for c : $\beta_c = 0.20$. However, the effect on the predicted probabilities is clearly not constant. At particularly low and particularly high values of c , there is almost no effect at all; that is, the curve is almost flat, indicating a slope in the predicted probability metric of almost zero. However, in the middle of the range of c , there is a large effect on the predicted probabilities of a change in c . That is, despite the fact that there is only one coefficient estimate, there is a great deal of variation in the effect on the predicted probabilities.

To complicate things even further, consider that the relationship observed in a particular data set for a particular model will vary greatly even when the regression coefficient estimate is the same. To illustrate this, I have plotted four predicted

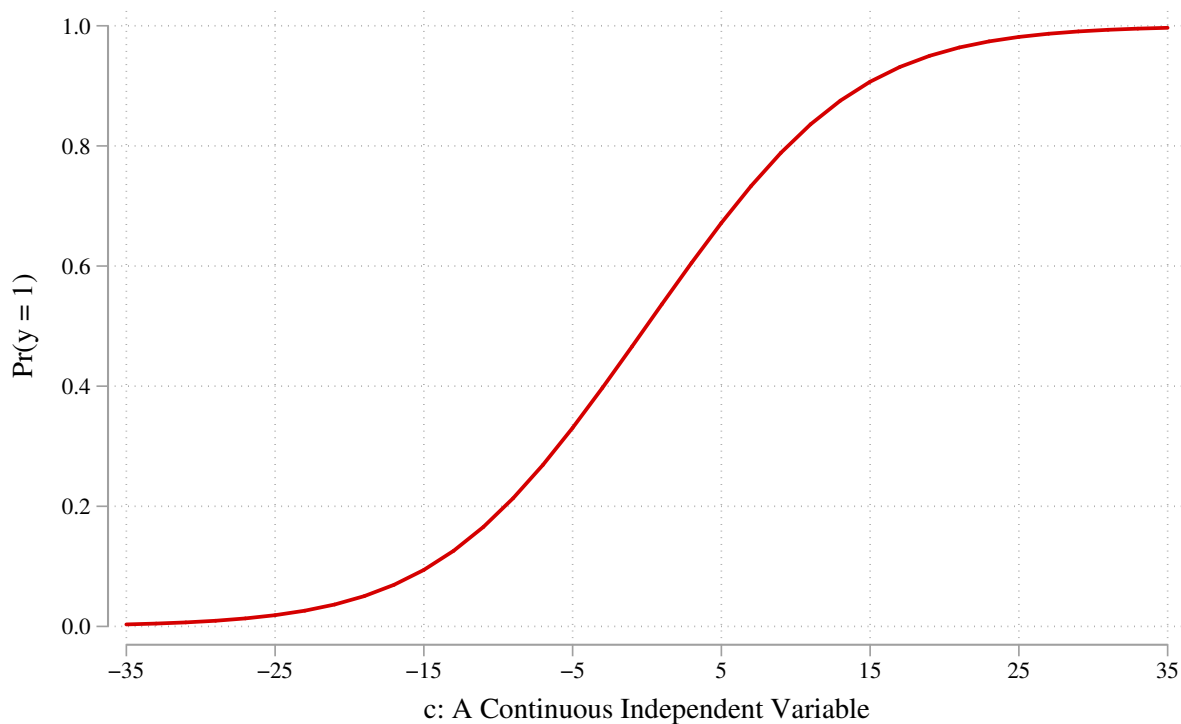


Figure 3: Simulated data: predicted probability of $y = 1$ across the range of c (a continuous variable).

probability curves in Figure 4. Each of these represent a coefficient estimate of $\beta_c = 0.20$. However, depending on where on the full predicted probability curve the observed data fall, the effect in terms of the predicted probabilities can differ drastically. In the top-left panel, there is virtually no effect of c on the predicted probability except for values greater than -20 , and even here the effect is small. In contrast, the top-right panel shows large effects for all values of c but with a highly nonlinear form: The change in the probability is noticeably larger in the center of the figure than at the extremes of c . The bottom-left panel shows an effect that is large and relatively linear in nature. The bottom-right panel shows only small effects on the probability, with the effect diminishing at larger values of c . The key point is that knowing the coefficient estimate ($\beta_c = 0.20$ for all of the probability curves shown in Figure 4) does not indicate an obvious or consistent effect on the predicted probabilities.

Note that interpreting these effects as odds ratios does not solve the problem: We can exponentiate the coefficient for c : $e^{0.20} = 1.22$. That is, a one-unit increase in c is associated with a 1.22-times increase in the odds of $y = 1$, and this effect is the same across all four panels of Figure 4. However, it is clear that the substantive impact on the predicted probabilities varies greatly. If the metric of interest is the

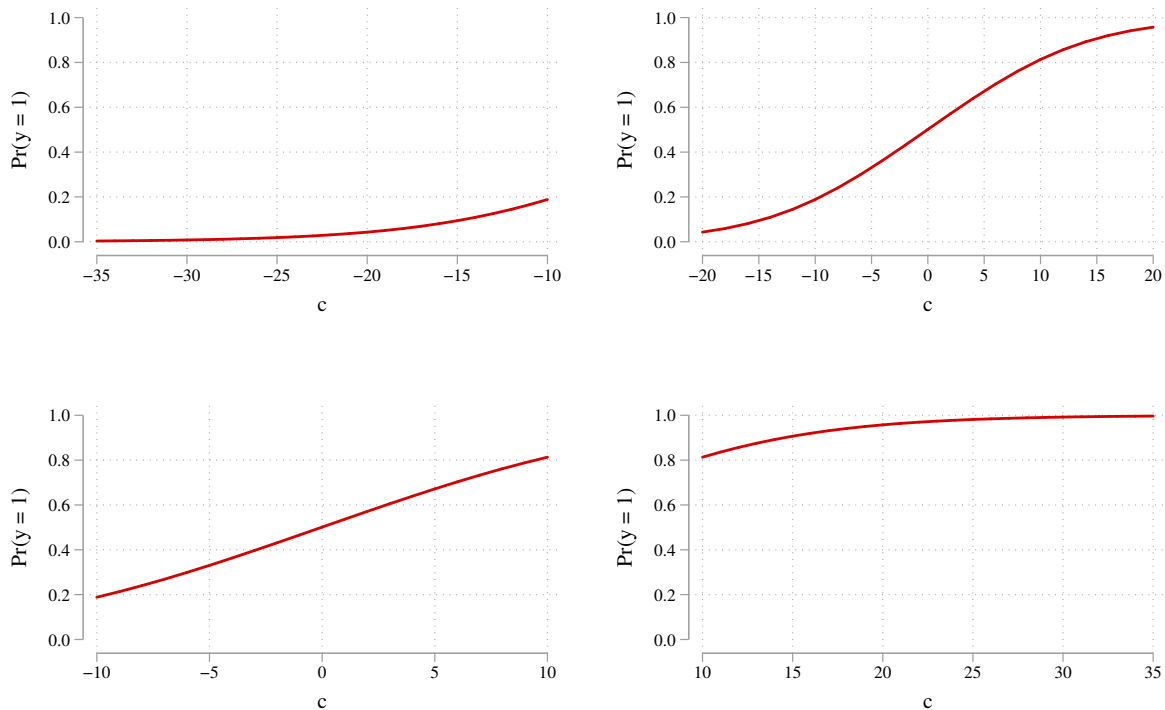


Figure 4: Simulated data: different effects on the predicted probability of $y = 1$ despite the same coefficient estimate of $\beta = 0.20$.

predicted probabilities, the regression coefficients do not provide a straightforward summary of the effects in this metric.

Interaction Effects in Models for Categorical Outcomes

For some of the reasons discussed above—and for others outlined below—it is not possible to determine the nature of an interaction effect on the predicted probabilities in logit/probit models based on the coefficients alone. As Ai and Norton note: “The interaction effect...cannot be evaluated simply by looking at the sign, magnitude, or statistical significance of the coefficient on the interaction term when the model is nonlinear” (2003:129). An additional consideration is that logit/probit models are already interactive in one sense before a product term is even introduced into the model.

To illustrate, consider the effect of an intercept shift for a continuous independent variable c in binary logit produced by a binary independent variable b (also in the model). Figure 5 uses the same simulated data from above with the relationships as specified in Equation 7. From Figure 5, it is clear that the overall shape of the two curves for the effect of c when $b = 0$ and when $b = 1$ are the same. Just as in linear regression, without a product term in the model, the overall slopes of the two

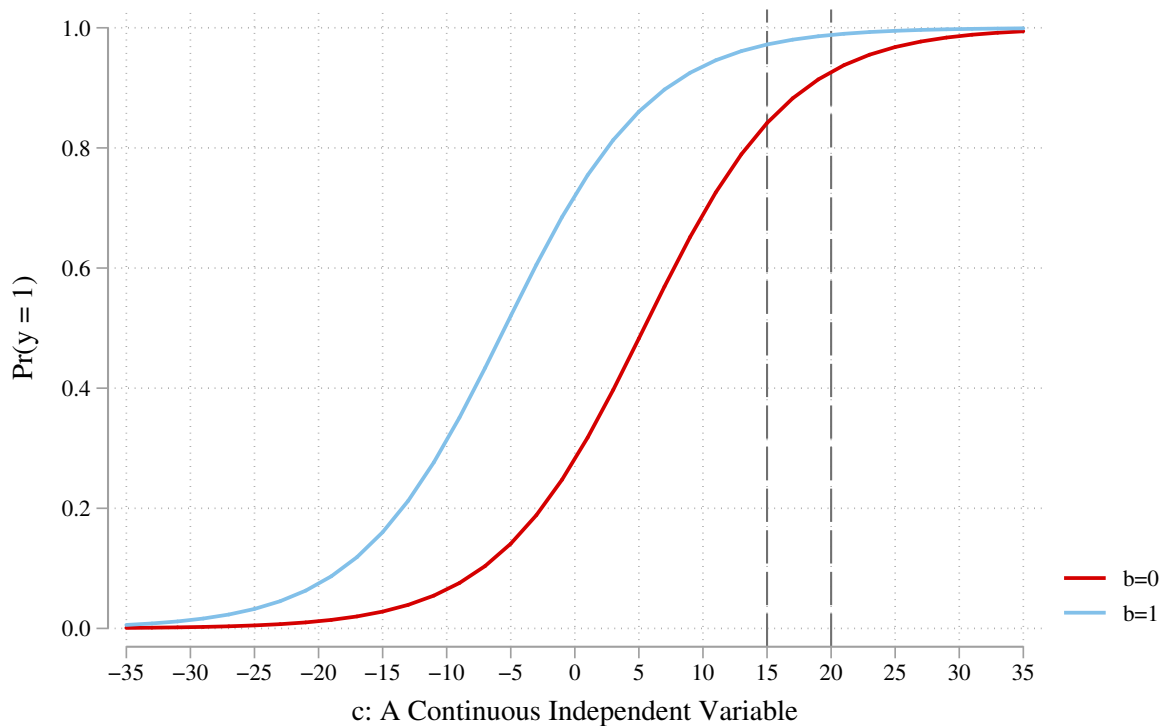


Figure 5: Simulated data: effect of an intercept shift in binary logistic regression.

curves are invariant across the levels of b . However, because of the intercept shift, the effects in terms of the predicted probabilities at certain places on the curves do differ. For example, consider the marginal effect of c increasing from 15 to 20. When $b = 0$, this has a noticeable effect on the predicted probabilities, as the slope of the red curve in Figure 5 is still quite steep. In contrast, when $b = 1$, there is only a very small effect on the predicted probabilities because the curve has almost reached its maximum of one, and thus, the slope is close to zero. This aspect of logit and probit is often referred to as “compression”—that is, because the predicted probability curves are bounded between zero and one, the separate lines for levels of b in Figure 5 must “compress” at some point because they are bounded at zero and one (Berry et al. 2010; Rainey 2016). Importantly, this causes the effects on the predicted probabilities to be interactive: The effect of c is different in some places on the curve depending on whether $b = 0$ or $b = 1$ despite the fact that only a single coefficient produced the slope for these two curves.

To put this more concretely and to show some real data implications, consider the effect that education has on the probability of being employed for parents. In particular, we might want to know whether there are different patterns for mothers and fathers. Data for this example come from Wave IV of the National Longitudinal Study of Adolescent to Adult Health, when respondents were about 28.5 years old,

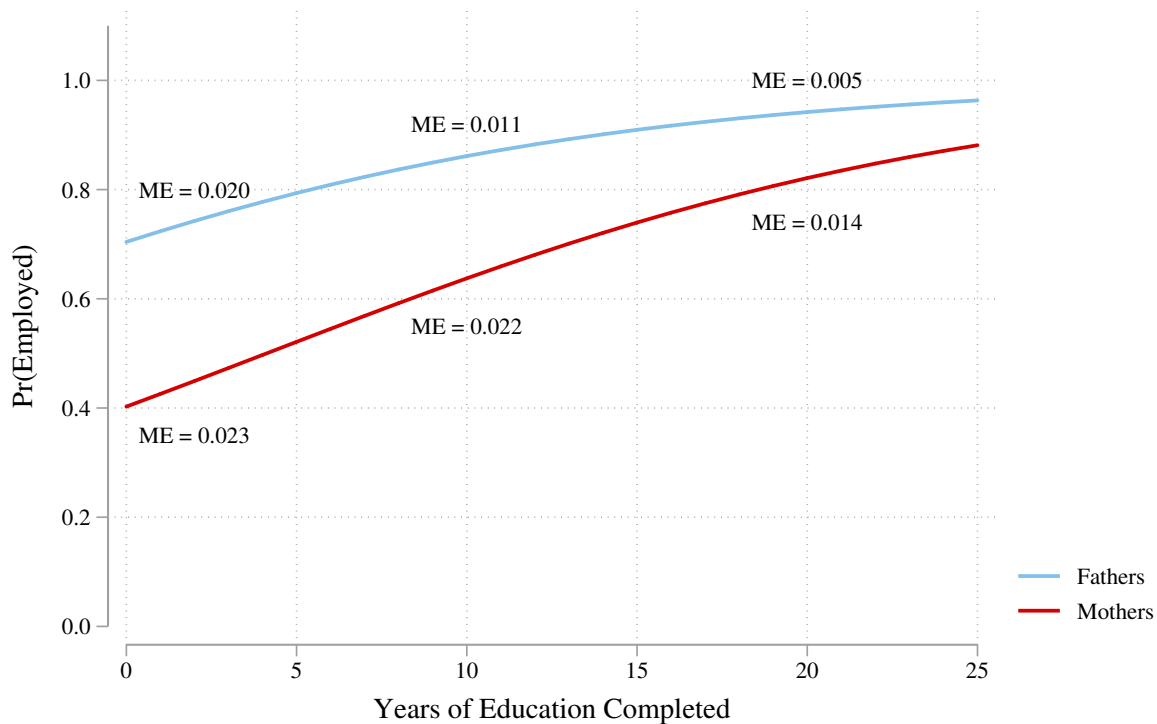


Figure 6: Predicted probability of being employed for mothers and fathers based on years of education completed: intercept shift example (no interaction term in the model).

on average. Here, I fit a binary logistic regression model regressing *employed* (yes or no) on years of *education* completed and on the *gender* of the parent (mothers = 1; childless men and women are excluded from this model; controls for respondent demographics included but not shown).

With linear effects, the effect of education would be equivalent for mothers and fathers, as there is no product term in the model. However, as Figure 6 illustrates, the effect that education has on the predicted probability of employment varies based on both the level of education considered and whether the prediction is for a mother or father. Included on the plot are +1 discrete changes calculated at 0, 10, and 20 years of education for mothers and fathers separately. There are clearly interactive effects in the traditional sense of the term: The effect of education depends on the level of parental gender. At low levels of education, the effects are similar in size for mothers and fathers; at higher levels, the effects appear larger for mothers.

Figure 6 raises the important question of whether a product term is necessary in logit and probit considering that interactive effects exist even without one. Moreover, is it appropriate to interpret the results shown in Figure 6 as evidence of

an interactive relationship between education and parental gender? I turn to this question next.

Is the Product (Interaction) Term Necessary in Models for Categorical Dependent Variables?

There has been some debate in the methodological literature about the necessity of including a product term in logit/probit models because these nonlinear models (as shown above) are already “interactive” in nature to an extent regardless of whether an interaction term is in the model or not (e.g., Nagler 1991; Ai and Norton 2003; Brambor et al. 2006; Berry et al. 2010; Berry, DeMerrit, and Esarey 2016; Rainey 2016). On one side, if “compression” effects are of theoretical interest, the inherently interactive nature of logit and probit models might already handle this type of relationship. For example, it might be reasonable to expect that at extremely high levels of education, both mothers’ and fathers’ probabilities of being employed are almost one, and thus, any further increase in education is unlikely to have much of an effect for either group.

It is worth laying out two important questions about the implications of including or excluding a product term when testing for interaction and considering the pros and cons. First, what are the consequences of including an unneeded product term in a logit or probit model? That is, if compression models the data-generating process well, what harm is there in adding an unnecessary product term to the model? Second, what are the consequences of omitting a needed product term from a logit or probit model? That is, if differential slopes are needed in order to model the data-generating process, what is the harm of leaving out the product term?

To help answer these questions, I first turn to additional simulated data in order to control the data-generating process. The following examples again focus on the case of one continuous and one binary independent variable. Here, I have modified the example first laid out in Equation 7. Equation 8 shows the true model specification I used to generate the simulated data:

$$\ln\left\{\frac{\Pr(y_i = 1)}{1 - \Pr(y_i = 1)}\right\} = -1 + 0.2c + 2b + 0cXb. \quad (8)$$

As Equation 8 shows, there is no interaction effect in terms of the coefficients; that is, the true value of the coefficient on the product term is zero. What effect does this have on the predicted probabilities? Figure 7 plots the predicted probability curves from this simulated data set for two separate models. The left panel of Figure 7 presents a model with a product term excluded from the model; the right panel presents a model with a cXb product term included. As is clear from Figure 7, the implications for inference are unaffected when including the product term. The only loss of including the product term is an additional degree of freedom. Here, I am arguing that that there is little lost by adding the product term to the model even if the researcher theorizes that compression should accurately model the relationships of interest: If there is no interaction effect in the data-generating process, including a product term in the model will not harm inferences, as the

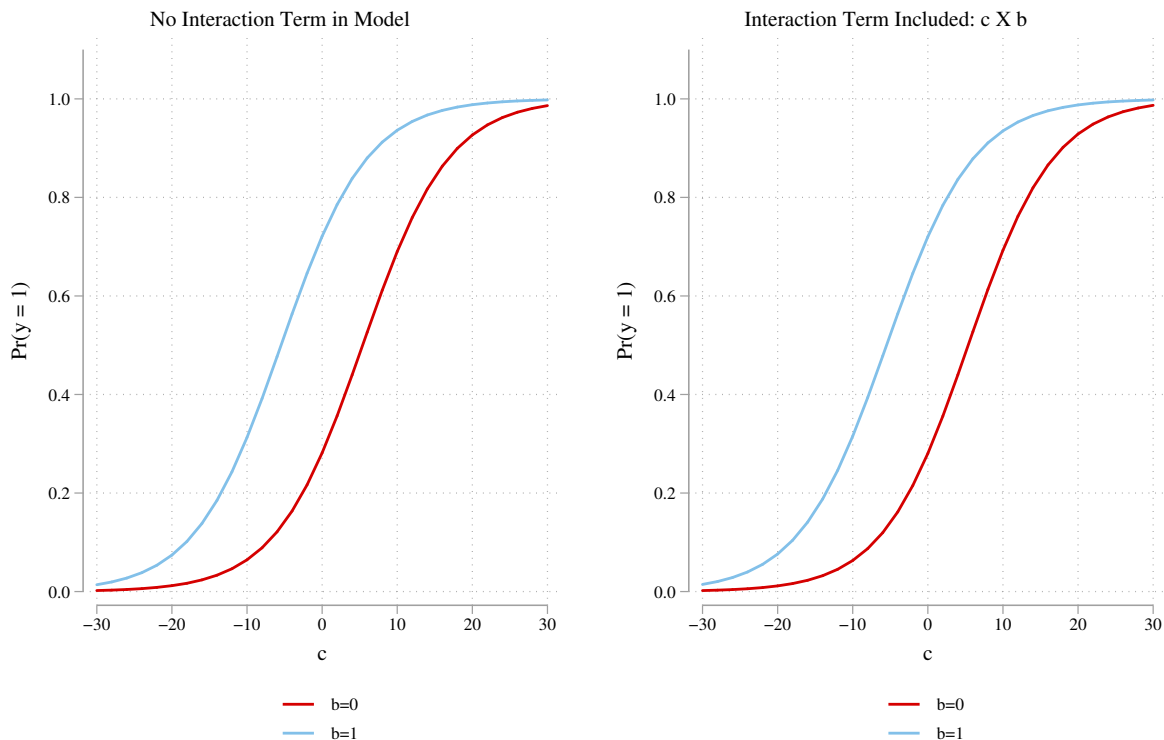


Figure 7: Effect of unneeded product term: the true coefficient on the product term is effectively zero.

estimate of the coefficient on the product term will be effectively zero and will not affect the predictions (see also Brambor et al. 2006).

Importantly, just because the coefficient of the product term is essentially zero, this does not mean that there is no interaction effect in the predicted probability metric. If, for example, effects of c from 20 to 30 years old are of interest, the effects appear larger when $b = 0$ than when $b = 1$ (note that tests of marginal effects and second differences should be used to test this directly if it is of interest). The coefficient of the product term does not provide a test for whether the effect differs in the predicted probability metric.

In contrast to the inconsequential effect of including an unneeded product term in the model, as I will show, the implications of omitting a needed product term from the model are severe and can lead to incorrect conclusions. Here, I again simulate data for which the true data-generating process is:

$$\ln\left\{\frac{\Pr(y_i = 1)}{1 - \Pr(y_i = 1)}\right\} = -1 + 0.2c + 2b + 0.2cXb. \quad (9)$$

Equation 9 specifies that an interaction effect exists in terms of the coefficients and the log-odds metric. To illustrate the effect of excluding or including an interaction term in the model on the predicted probabilities, Figure 8 plots the predicted

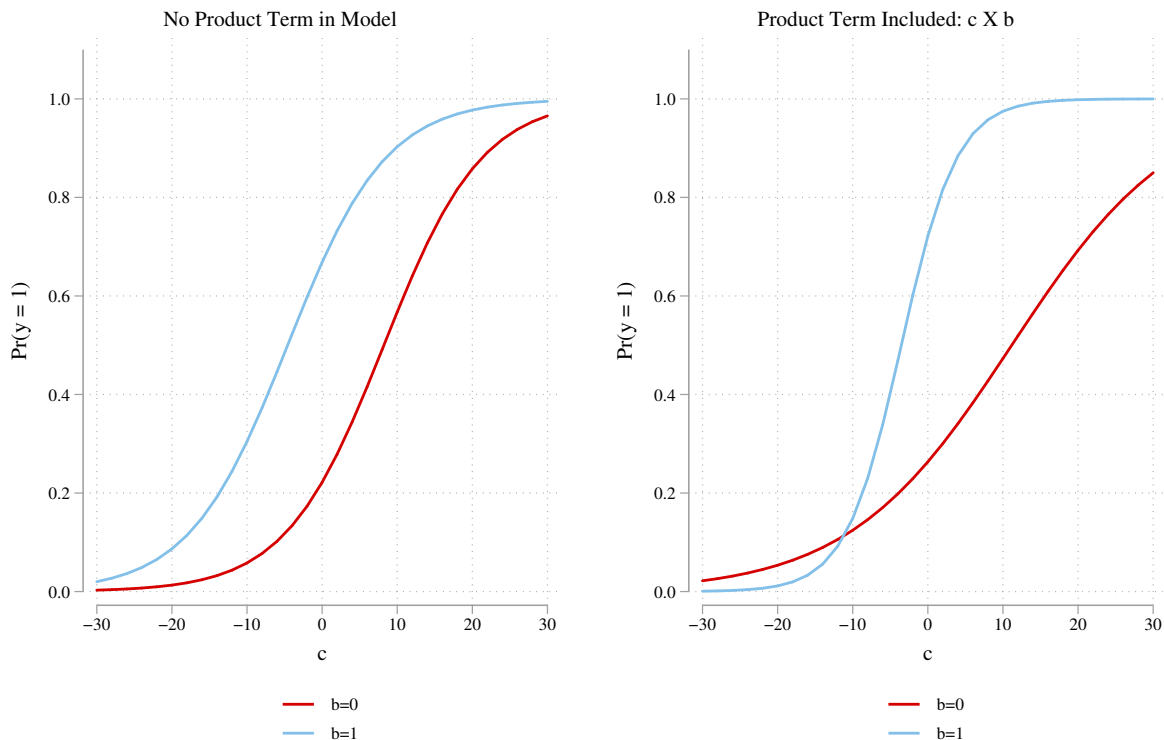


Figure 8: Effect of excluding a needed product term: the true coefficient on the product term is not zero.

probability curves from a model excluding a product term (left panel) and a model including a $c \times b$ product term (right panel). Here, the effect of omitting the needed product term has serious implications. The overall slopes of the curves should be allowed to vary; omitting the needed product term constrains the slopes of the two curves to be equal when they should not be. Therefore, a product term must be included in the model to allow the effect to vary if indeed it should.

Another side of this important question is the consequence of testing for interaction effects in logit and probit models in the absence of a product term in the model. Rainey (2016) presents a systematic and extended investigation into this issue. Although the implications vary based on underlying effect and sample sizes, in extreme circumstances, researchers can find an interaction effect in the predicted probabilities 100 percent of the time even when an interaction effect does not exist in the data-generating process if a product term is omitted from the model. That is, testing for interaction without including a product term can lead to an almost certain type I error (for an example of a product term being necessary in the model to avoid a type I error, see the example of “interactions of one nominal and one continuous independent variable” later in this article). Encouragingly, Rainey (2016) shows that adding a product term to the model—thus relaxing the compression effect if it indeed needs to be relaxed—completely removes this bias,

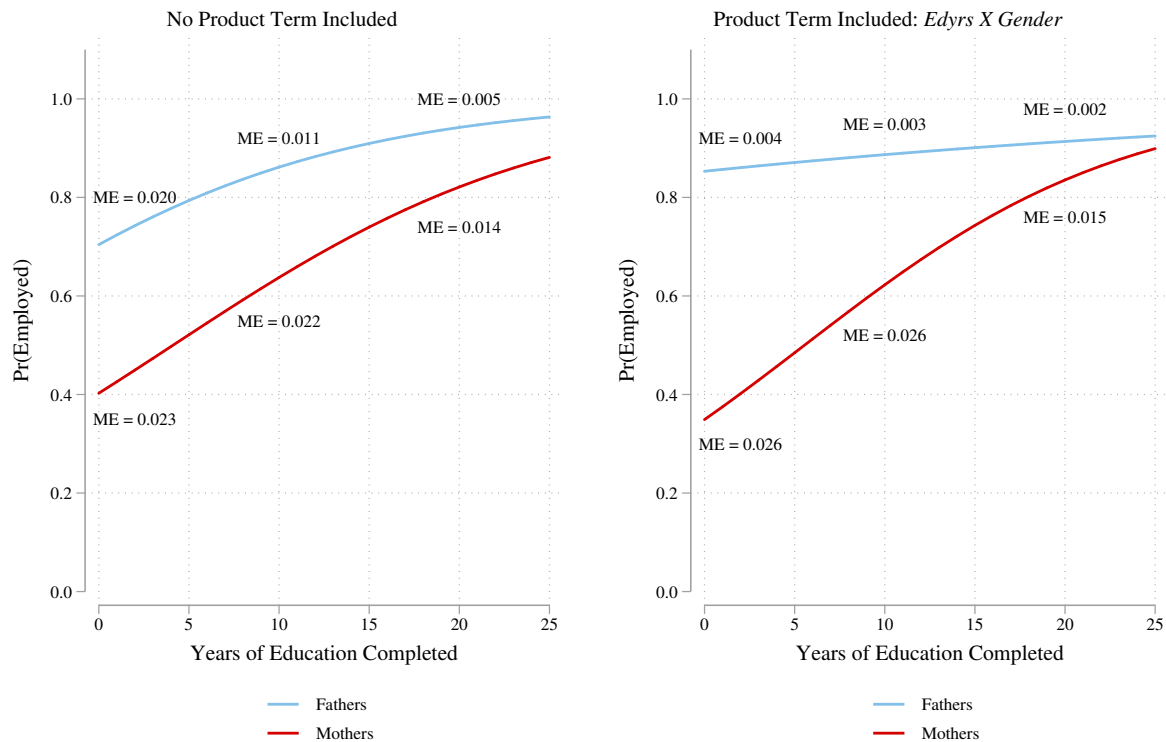


Figure 9: Effect of excluding a needed product term: interaction effect exists in the data-generating process.

and researchers will find evidence of interaction in the predicted probabilities only at levels expected by random chance.

Returning to the model predicting employment status for mothers and fathers illustrates these issues using real data. The left panel of Figure 9 shows the same model considered before, which includes education and parental gender but omits a product term; the model presented in the right panel includes an *education* \times *gender* product term. It is clear that forcing the slopes to be invariant across parental gender leads to quite different implications in terms of the predicted probabilities. Once the overall slopes are allowed to vary via the product term (right panel), the effect of education appears larger for mothers than for fathers; the marginal effects for fathers are fairly small and similar over the range of education. However, for mothers, the effect at 10 years of education is large and almost twice that of the effect for mothers at 20 years of education. A test of the second difference indicates that the average effect of age is larger for women than for men ($p < 0.01$).

In sum, if an interaction effect is of interest, a product term should be included in the model. In addition, even when compression is theoretically reasonable, this should be tested in a model with a product term included—the only loss is a degree of freedom. If compression fits the data, including a product term will not harm the ability to find these relationships. However, excluding the product term is likely

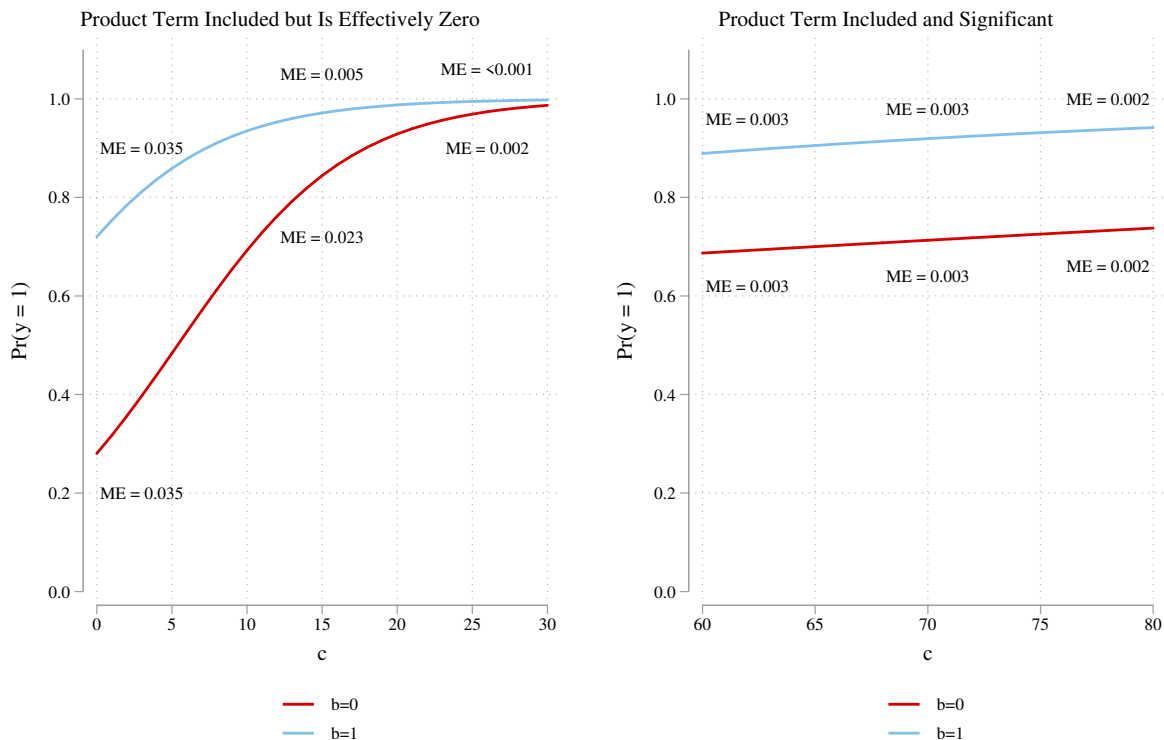


Figure 10: Interaction effect on the predicted probabilities cannot be determined by examining the coefficient of the product term.

to produce a false positive, indicating interaction when there is no true interaction effect in the data.

The Coefficient on the Product Term Is Not a Test of Interaction in the Predicted Probabilities

Above, I went into some depth to defend and promote the inclusion of product terms when testing for interaction in logit/probit models. It is worth expanding on why the coefficient on the product term cannot be interpreted as a test of interaction in the predicted probability metric even though the coefficient itself is needed in the model (for additional details beyond what is presented here, see Ai and Norton 2003). Consider the two models plotted in Figure 10 from simulated data; both data sets contain 2,000 simulated observations, and both models include product terms.

For the model in the left panel, the coefficient of the product term is effectively zero and is nonsignificant ($p = 0.84$). However, the effects in terms of the marginal effects on the predicted probabilities are clearly interactive, with the effect of c varying greatly depending on the level of b . In contrast, the right panel shows the predictions from a model in which the coefficient of the product term is relatively

large and statistically significant ($p < 0.05$). However, the effects in terms of the predicted probabilities for c are all but identical regardless of the level of b . That is, in direct contrast to what we would expect to observe based on the coefficients of the product terms, the effects in terms of the predicted probabilities directly contradict our naive inferences.

The reason for this seeming contradiction again brings up the importance of what logit and probit coefficients can and cannot tell us. Recall the examples shown in Figure 4 in which the effects observed for a single coefficient estimate can be quite different depending on where the observed data lie on the predicted probability curve. An interaction effect allows the entire predicted probability curve to vary across levels of another variable. Combine this fact with the effect of intercept shifts on the predicted probabilities in logit/probit models illustrated earlier. The key point is that how these will affect the predicted probabilities observed in your data is difficult to say without examining the predicted probabilities themselves.

Figure 11 presents predictions from the same models that are shown in Figure 10. However, I have made the predictions at different values of the continuous variable c . Now, the inferences of interaction are quite different. There is a clear interaction effect for the model shown in the right panel of Figure 11 but little to no interaction effect for the model shown in the left panel. The point of showing these different places in the data space is to emphasize that the implications in terms of the predicted probabilities cannot be determined by examining a single coefficient estimate (for the main effects or for the coefficient of the product term). The predicted probabilities themselves must be examined to determine whether an interaction effect exists. The coefficient of the product term—although needed in the model—can safely be ignored when testing for and interpreting interaction effects in the predicted probability metric.

Presenting and Interpreting Interaction Effects

In this section, I give examples and advice about presenting interaction effects with a focus on different techniques that are well suited to a particular type of interaction effect, which depends on the measurement level of the independent variables that are included in the interaction (i.e., nominal or continuous). The examples in this section all use binary logistic regression. I focus on binary logistic regression because models for categorical outcomes require these types of techniques to test interaction effects and because the techniques described here extend straightforwardly to other categorical outcome models, such as those for nominal, ordinal, and count outcomes (see the discussion section for details on extensions to other models).

Interactions of Two Nominal Independent Variables

Interaction effects of two nominal independent variables are the most straightforward types of interaction effects to estimate and to interpret. This is because these produce far fewer possible different values of the focal independent variables at which to test the interaction effect. As an example, consider how being a parent might differentially affect a person's drinking habits depending on whether the

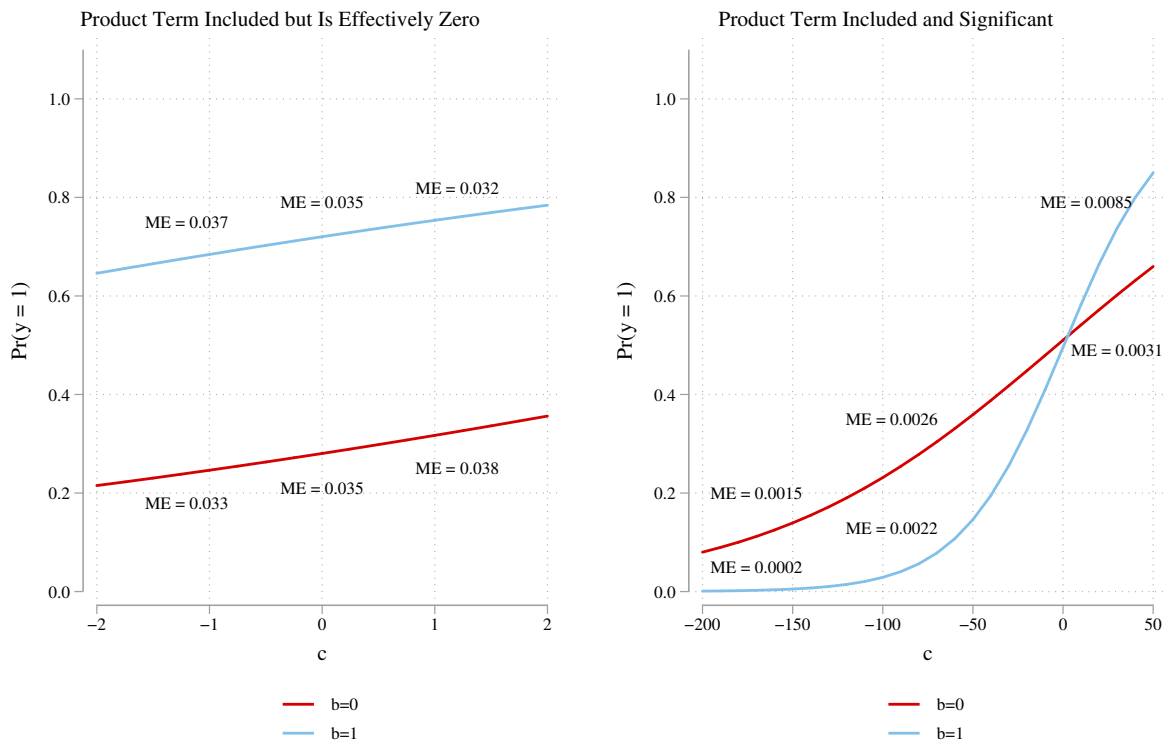


Figure 11: Interaction effect on the predicted probabilities cannot be determined by examining the coefficient of the product term: different parts of data space.

person is a man or a woman. Data for this example come from Wave IV of Add Health. Here, both parental status and gender are coded as binaries. To test for a possible interactive effect, I fit a binary logistic regression model regressing *alcohol use* (any in the last month: yes or no) on *woman*, *parent*, *womanXparent*, and some standard control variables.

These types of effects (especially in the case of nominal variables with only two categories [i.e., binary variables]) can sometimes be conveyed via text and tables alone, although visualizations are always helpful. I recommend researchers make a plot of the model predictions even if the visualization does not end up in the final manuscript; at the least, it is helpful as a guide to interpretation for the analyst. Figure 12 and Table 3 present the same information in different formats. Both show the predicted probability of alcohol use for the four combinations of the gender and parental-status variables; Figure 12 includes a bar chart with standard-error bars to represent uncertainty, and Table 3 presents the same information but also includes the tests of the first (marginal effects) and second differences.

The results show that men without children have a significantly higher probability of drinking alcohol (0.730) than do fathers (0.588; $\Delta = 0.142$; $p < 0.01$). Similarly, women without children have a higher probability of drinking alcohol (0.673) than

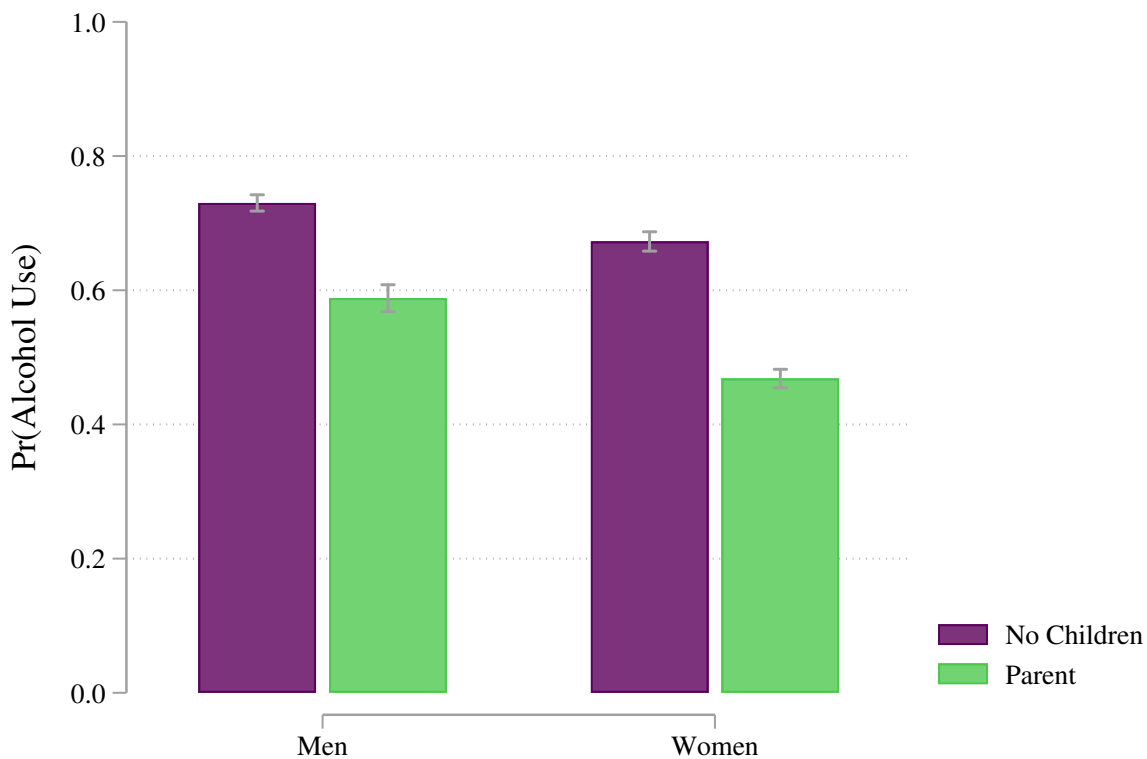


Figure 12: Probability of alcohol use by gender and parental status.

Table 3: Probability of alcohol use by gender and parental status with test of interaction Effect ($N = 4,307$).

	Pr(Alcohol Use)	First Differences	Second Difference
Childless Men	0.730 (0.012)		
Fathers	0.588 (0.020)	$0.730 - 0.588 =$ 0.142*	
Childless Women	0.673 (0.015)		$0.142 - 0.204 =$ -0.063^*
Mothers	0.468 (0.014)	$0.673 - 0.468 =$ 0.204*	

Notes: Standard errors of the predictions in parentheses. * $p < 0.05$, two-tailed tests.

do mothers (0.468; $\Delta = 0.204$; $p < 0.01$). The effect of parenthood—with parents being less likely to consume alcohol—is larger for women than it is for men (second difference $0.142 - 0.204 = -0.063$; $p < 0.05$).

Interaction effects with nominal independent variables become more cumbersome as the number of categories of the independent variables increase but only in the sense that more comparisons are required; the techniques to test the interaction

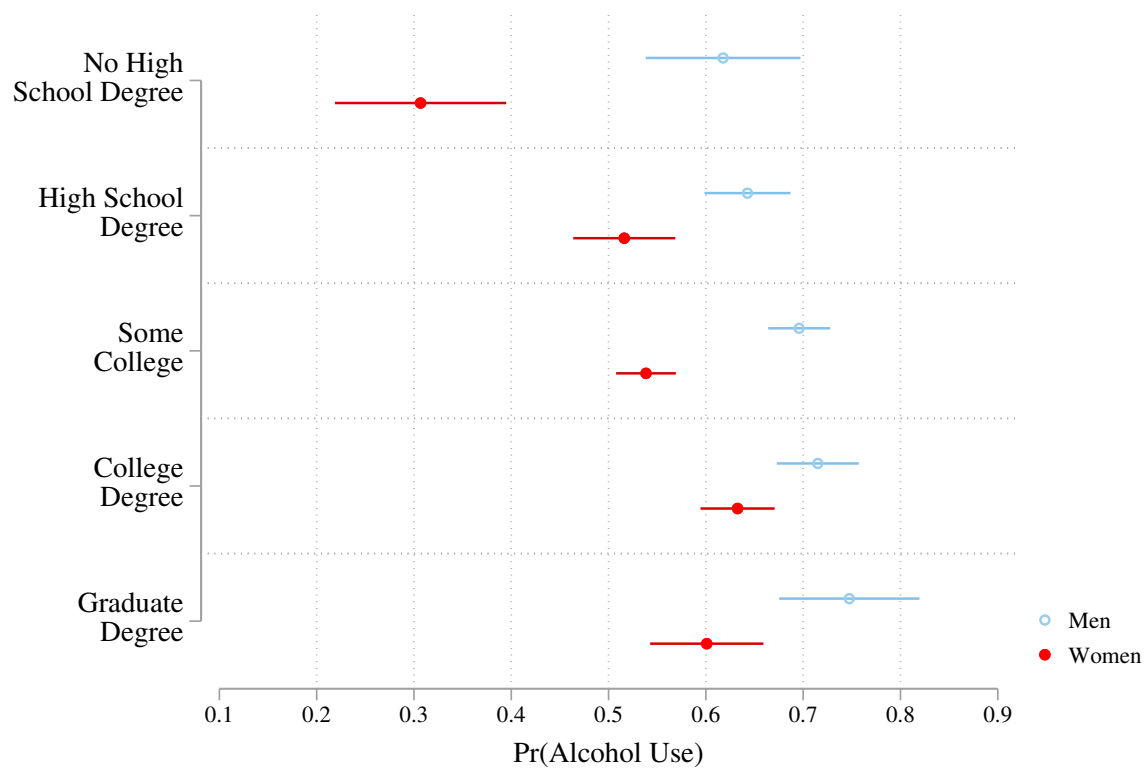


Figure 13: Probability of alcohol use by gender and educational attainment.

effects remain the same. As an example, consider that men are more likely to drink than women. One might be interested in whether this gender gap is bigger or smaller for individuals with different levels of education (here, entered into the regression model as a nominal variable of highest educational degree obtained). To test this, I fit a binary logit model regressing the same *alcoholuse* binary variable on the gender and education categories, and product terms between gender and the education categories (and controls). Figure 13 shows predicted probabilities of alcohol use for each gender and educational attainment category combination. Figure 13 utilizes a different type of figure—a horizontal dot plot with 95 percent confidence intervals—to present the information (Cleveland 1993; Jann 2014); a bar chart, such as Figure 12, could alternatively be used to present this information.

Figure 13 illustrates the gender gap in alcohol use across the various levels of educational attainment by comparing the gap between the predictions for men (in blue) and the predictions for women (in red). With this visual representation, it is easy to notice that the gender gap appears largest for those without a high school degree and smallest for those with a college degree. Table 4 presents the same information along with tests of the gender gap and tests of whether the size of the gender gap differs across levels of education (second differences [i.e., the test of interaction]).

Table 4: Probability of alcohol use by gender and education: marginal effects of gender and differences in effects of gender across levels of education ($N = 4,307$).

	Women	Men	Gender Gap (AME of Gender)	Contrasts
<i>a</i> No High School	0.307	0.618	-0.311*	<i>b, c, d, e</i>
<i>b</i> High School	0.516	0.643	-0.127*	<i>a</i>
<i>c</i> Some College	0.538	0.696	-0.157*	<i>a, d</i>
<i>d</i> College Degree	0.633	0.715	-0.082*	<i>a, c</i>
<i>e</i> Graduate Degree	0.601	0.747	-0.147*	<i>a</i>

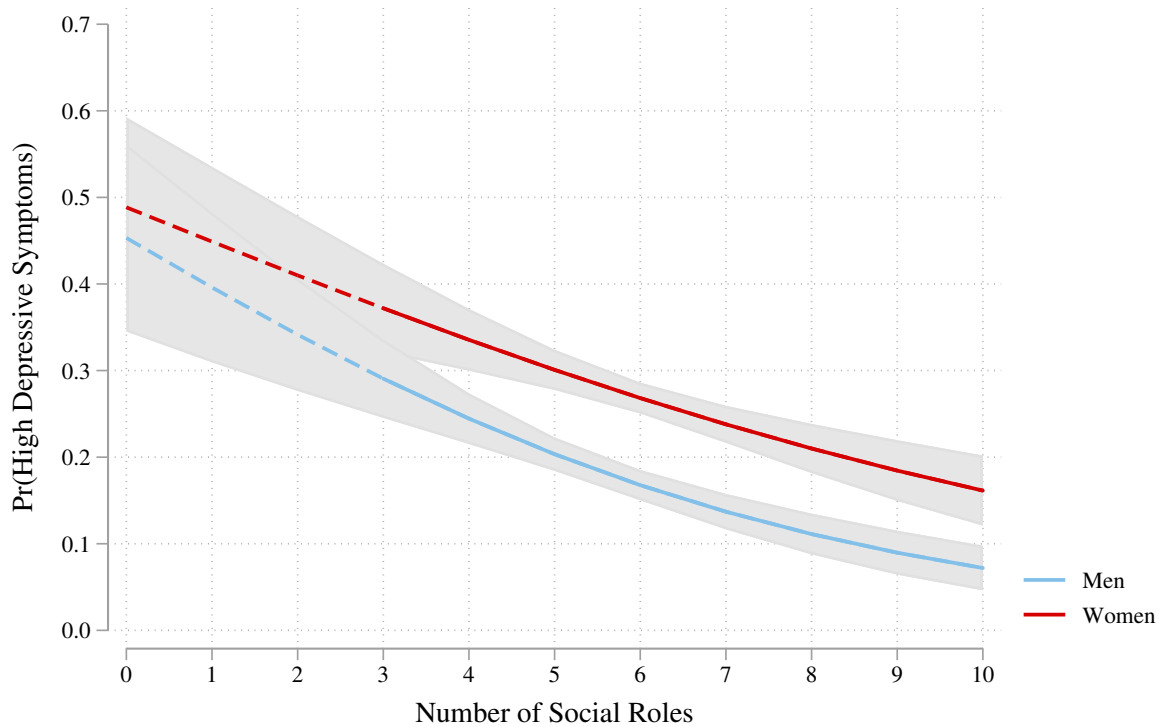
Notes: The "contrasts" column reports which gender gaps are significantly different across levels of education (second differences). * $p < 0.05$, two-tailed tests.

As shown in Table 4, there is a significant gender gap across all educational levels, with men being more likely to report alcohol use than women (all gender gaps $p < 0.01$). Testing whether the effect of gender differs across levels of education requires a test of second difference, presented in the final column labeled "contrasts." For example, the gender gap in the probability of alcohol use is significantly larger for those without a high school degree (-0.311) than it is for those of any other educational level (all second differences $p < 0.05$). Note that these interaction effects are dependent on each contrast being tested. That is, just because the gender gap differs across two given categories of education, this does not necessarily indicate that other contrasts will also differ. Note also that I have only shown one side of the interaction effect for this example (the effect of gender); generally, it is best practice to also examine the other side—in this example, the effect of education. I omit this for space but encourage researchers to examine both sides of the interaction in applied work.

Interactions of One Nominal and One Continuous Independent Variable

For interaction effects when at least one focal independent variable is continuous, a visual presentation of the effects is almost always needed and informative; this is especially true with nonlinear effects. Many of the suggestions below build on those of Long and Mustillo (2018), who focus on examining group differences in nonlinear models. As an example, consider the two well-established findings that women are more likely to report depressive symptoms than are men and that those who are more socially integrated by holding more social roles tend to have better mental health outcomes (for a review, see Thoits 2011). Of interest is whether the positive mental health effects of holding more social roles differ for men and women. To test this possibility, using Add Health Wave IV data, I fit a binary logistic regression model regressing a high level of *depressive* symptoms (binary: high or not) on *gender* (woman = 1), the number of *socialroles* held, and *genderXroles* (and controls).

The results are illustrated in Figure 14. Starting with the social roles side of the interaction, calculating AMEs is helpful: On average, each additional social role is



NOTE: Group difference (men vs women) is significant ($p < 0.05$) when lines are solid

Figure 14: Predicted probability of high depressive symptoms by gender and social roles held: interaction effect between gender and social roles.

associated with a 0.032 decrease in the predicted probability of depressive symptoms for men and a similar 0.030 decrease for women (both AME first differences $p < 0.01$). There is no significant difference in the average effect of social roles for men and women (second difference $p = ns$).⁹

As mentioned above, it is important to test both sides of the interaction—that is, to consider how the interaction effect operates for both variables in the interaction (Berry et al. 2012) (for this example, to test whether there are significant group differences between men and women at different values of social roles held). One way to present this information is to directly plot the marginal effect of gender across the range of social roles (see Figure 2 for an example). An alternative strategy is to incorporate information about the significance of the group difference (men vs. women) directly into Figure 14; indeed, I have already done so. In Figure 14, the lines are dashed when the group differences are not significant and solid when they are significant. That is, the gender gap is significant—with women having a significantly higher probability of high depressive symptoms—when comparing someone who holds between three and 10 social roles (all contrasts $p < 0.05$). There are no gender differences in the probability of depressive symptoms when someone holds between zero and two social roles (all contrasts $p = ns$). This information is

entirely contained within the figure, a solution I find more elegant than a lengthy table of statistics (for a different substantive example using this approach to convey significance of group differences, see Heilman et al. 2011).

Interactions of Two Continuous Independent Variables

The trickiest types of interaction effects to interpret and present are those between two continuous independent variables. However, visual presentations can greatly aid the ability to convey these interaction effects. As an example, consider that 30 years ago, U.S. residents almost universally viewed same-sex relationships as morally wrong; however, public opinion has shifted substantially, with increasing moral approval of same-sex relationships seen over time. Of interest is whether these shifts toward greater social acceptance have been broad based or whether they have been concentrated among individuals with certain political views. To test this, using data from the General Social Survey, I fit a binary logistic regression model predicting views toward same-sex relationships (binary: wrong = 0; not wrong at all = 1) with conservative *political views*, the *year* public opinion was measured (every two years from 1976 to 2016), and *political_views* \times *year*. The measure of *conservative political views* is a standardized scale, with 0 representing neutral, negative values representing more liberal views, and positive values representing more conservative views.

The first method of presenting a continuous-by-continuous interaction is to choose “ideal types” for one of the two continuous independent variables (Weber 1922; Long and Freese 2014). By ideal types, I am referring to particular values of the independent variables of interest that are representative of something either theoretically or substantively interesting. Sometimes, there are clear choices for substantively interesting values of one variable to choose; other times, no particular values may present themselves as obvious ideal types of interest. One solution for any continuous variable is to choose values at systematic percentiles of the variable’s distribution (e.g., the 25th and 75th percentile or the 10th and 90th percentile). As the *conservative political views* scale is standardized, I choose one standard deviation below the mean and term it “liberal” and one standard deviation above the mean and term it “conservative.”

Figure 15 illustrates this ideal-type approach to presenting continuous-by-continuous interaction effects. The effect of the passage of time is represented on the x axis, and the effect of political views is represented with two different lines: one for “liberals” (predictions made at *conservative political views* = -1 ; in blue) and one for “conservatives” (predictions made at *conservative political views* = 1 ; in red). First, both variables have a clear effect: Conservatives are less likely to view same-sex relationships as acceptable ($AME_{\text{conservative_views}} + SD = -0.078$; $p < 0.01$); U.S. residents have become more accepting of same-sex relationships over time ($AME_{\text{year}} + SD = 0.150$; $p < 0.01$). Second, the gap between liberals and conservatives appears to have grown over time. Third, the change in opinion appears slightly larger for liberals than for conservatives. Although I find Figure 15 a clear, intuitive, and helpful way to summarize these effects, it is also important

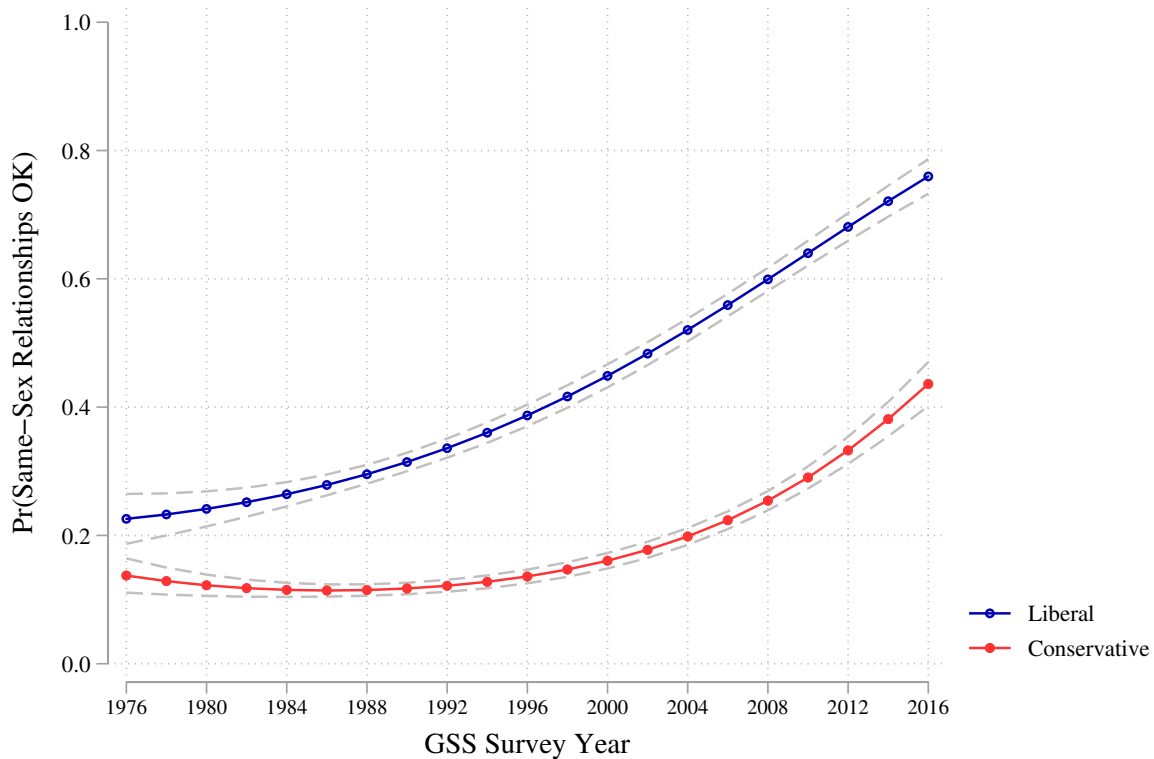
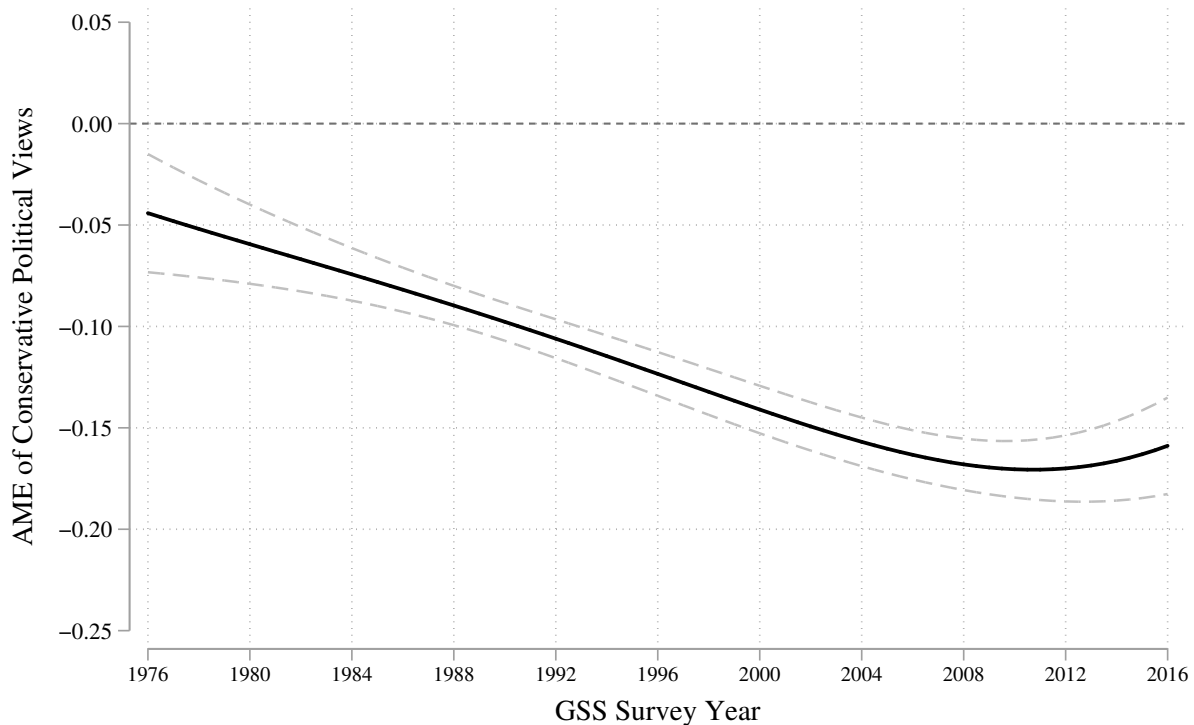


Figure 15: Probability of viewing same-sex relationships as acceptable by political views and survey year: interaction effect between political views and year.

to substantiate these intuitions with additional analyses and tests of statistical significance.

A second method of presenting continuous-by-continuous interaction effects is to plot the marginal effects of one independent variable across the range of the second independent variable—that is, to illustrate how the effect of one variable differs at various levels of the second variable. Figure 16 plots the average marginal effect of conservative political views across the range of survey years (i.e., how the impact of political views on opinions has changed over time).¹⁰ From the plot, we can see that political views have a significant effect in every survey year because the 95 percent confidence intervals on the effects never overlap zero; because the effects are negative, this indicates that the more conservative someone is, the less likely that person is to view same-sex relationships as acceptable.¹¹

In terms of the interaction effect, Figure 16 shows that political views have a larger (more negative) effect on opinions in later years. We can test this directly, for example: The effect of political views in 2016 ($AME_{2016} = -0.159$; $p < 0.01$) is significantly larger than the effect of political views in 1976 ($AME_{1976} = -0.044$; $p < 0.01$) with a second difference of -0.115 ($p < 0.01$).¹²

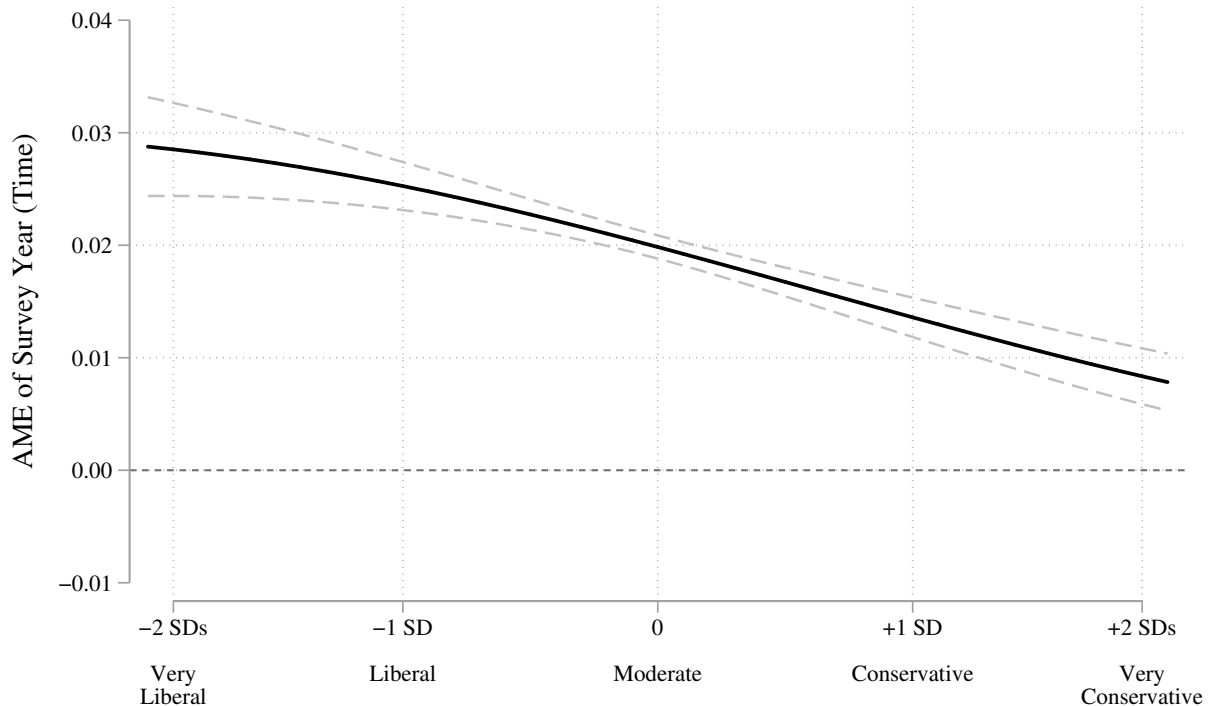


NOTE: Negative effects indicate conservatives are less supportive of same-sex relationships

Figure 16: Average marginal effect of conservative political views on the probability of viewing same-sex relationships as acceptable across GSS survey years.

It is good practice to present both sides of a continuous-by-continuous interaction. Here, it is how the effect of the passage of time has differentially impacted the opinions of those with differing political views. Figure 17 illustrates this side of the interaction. Here, the effect that time has had on opinions is presented across the range of political views. All effects are positive and significant; that is, individuals of all political stripes have become more likely over time to state that they view same-sex relationships as acceptable. However, it is clear that this change is larger for more-liberal individuals than for more-conservative individuals. For example, the average effect of a two-year increase in time for liberals (0.026; $p < 0.01$) is significantly larger than the effect for conservatives (0.015; $p < 0.01$) with a second difference of 0.011 ($p < 0.01$).

A final option for presenting continuous-by-continuous interaction effects is a contour plot (see Huber 2017 for additional examples). Contour plots present the predictions from a continuous-by-continuous interaction with each independent variable on an axis and the value of the prediction as a third dimension of color and/or shade. When color graphics are possible, different colors across the range of predicted probabilities are effective, as they allow the reader to easily differentiate



NOTE: Positive effects = group has become more supportive of same-sex relationships over time

Figure 17: Average marginal effect of time (GSS survey year) on the probability of viewing same-sex relationships as acceptable across political views.

each level of the predictions; for black-and-white graphics, grayscale is generally the most effective choice for displaying a gradient.

Figure 18 plots the predicted probabilities of viewing same-sex relationships as morally acceptable. The x axis is the year of the survey; the y axis is the scale of political views. The colors on the plot represent the predicted probability (scale is defined in the legend on the right side of the plot). An exact predicted probability for any given combination of year and political views can be found by finding the color at the intersection of those two independent variables on the plot. For example, the only individuals with very high (greater than or equal to 0.70; in blue) probabilities of thinking same-sex relationships are acceptable are those who are liberal or very liberal and answered the survey from around 2004 to 2016 (bottom right of the plot). The interaction effect is also clear from the plot. Skim across the very top of the plot (very conservative individuals): The probabilities have changed over time (left to right) but much less (cycling through many fewer colors) compared to the change for very liberal individuals (bottom of the plot). Similarly, political views have a much larger effect in more recent years (right side of the plot) than in prior years (left side). Although contour plots, such as Figure 18, necessitate some space dedicated to explaining the plot, they contain a wealth of information.

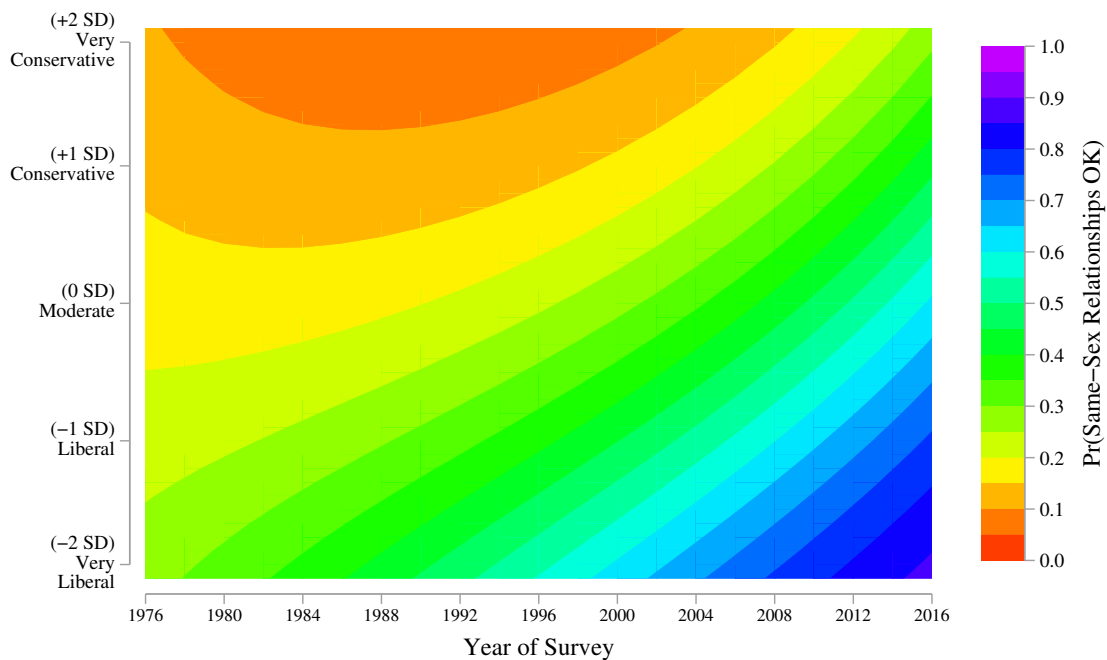


Figure 18: Contour plot of predicted probabilities of viewing same-sex relationships as acceptable at the intersections of political views and survey year.

Summary of Recommendations

Estimating, interpreting, and presenting nonlinear interaction effects requires more effort and consideration than doing the same for linear interaction effects. However, the current methodological literature and statistical software provide the guidance and ability to fully investigate these types of effects. In an effort to summarize the key things to consider:

1. Include a product (interaction) term in the model if you want to investigate whether the effect of one independent variable is contingent on the level of a second independent variable.
2. Ignore the coefficient of the product term: It does not necessarily provide accurate information about the significance, magnitude, or even the direction of the underlying interaction effect on the predictions.
3. Plot the predictions to determine the nature of the underlying interaction effect on the metric of interest.
4. Determine the size and significance of the effects of interest using marginal effects, not regression coefficients.
5. Use tests of second differences (whether two marginal effects are equal) to determine whether an interaction effect is significant for specific values of interest of your independent variables.

6. Absent substantive or theoretically interesting values of the focal independent or control variables to test the interaction effect, use average marginal effects to summarize whether there is an interaction effect present on average in the data.

Software Considerations and Example Code

Replication files to recreate all analyses in this article are available on the author's website at (trentonmize.com/research). Simplified and annotated template files to recreate the key examples in Stata are also available on the same website; I encourage researchers to examine the template files as starting points for conducting analyses similar to those I recommend here. Stata currently has the most robustly implemented procedures for testing nonlinear interaction effects due to the wide-ranging implementation of the `margins` command, which primarily uses the delta method for calculating variances and covariances of predictions and marginal effects (Pitblado 2014). In the example files, I use Long and Freese's (2014) `SPost13` package in many places to simplify the calculations of the marginal effects (`SPost13` is a free, user-written program for Stata). I also use Jann's (2014) `coefplot` command for some of the figures. All of the figures presented throughout the article use my own Stata graphics scheme `cleanplots`, which is freely available at trentonmize.com/software/cleanplots.

Discussion and Conclusion

Extensions to Ordinal, Nominal, and Count Models

Although I have focused on nonlinear interaction effects in linear and binary regression models, the topics I cover here extend straightforwardly to other models for categorical outcomes, such as nominal, ordinal, and count regression models. In each of these models, the effects on the predictions in the natural metric of the dependent variable are not linear, making the tools I advocate here necessary (for details on these models, see Long 1997; Agresti 2013; Long and Freese 2014).

Testing for interaction effects in the predicted probability metric (for example, for ordinal and multinomial logit models) is accomplished as it is described in this article for binary logit/probit models; there are simply more outcome categories for which to test the interaction. For example, for a nominal or ordinal model with four outcome categories, this would necessitate testing for interaction across all four outcome categories.

For count models, analyses of the predicted probability of a specific count (e.g., probability that $y = 0$) can be tested as I outlined for binary logit/probit models. For analyses of predictions of the rate (expected count), these can be tested similarly to nonlinear effects in the linear regression model. The key is to remember that the effects are nonlinear, and thus, the considerations outlined here apply equally.

Conclusion

Getting interaction effects right is important to many sociological inquiries. Many of the topics sociologists study require an analysis of an interaction effect. For example, any theory that predicts that something will operate differently based on a person's gender, race, or class is proposing an interaction effect. There are of course many other types of interaction effects of interest to sociologists beyond these few examples; many social processes are moderated by other factors. Similarly, many relationships of interest to sociologists are nonlinear in nature.

My reading of the substantive sociological literature suggests that although each of these topics individually are well understood, most applied researchers do not treat the combination of these two properly. That is, interaction effects when the relationship of interest is nonlinear are rarely tested in line with current methodological recommendations. I hope it is clear that my intent in this article is not criticism of past work. Instead, my goal is to provide a synthesis of an evolving methodological literature, presented in a way that is helpful for those interested in using these methods in their own substantive work.

Notes

- ¹ For this count, I focus on the specific case of logit- or probit-based models (e.g., binary probit, ordinal logit, etc.). I conducted the search in November 2016 by searching the ASR archives for "interaction" and "logit" or "probit" and examined all articles that report the results of a test of an interaction from a logit- or probit-based model. These articles vary greatly as to the centrality of the interaction effect to the substance of the question at hand.
- ² Only five provided the needed statistical test of the interaction effect in terms of the predicted probabilities either explicitly in text or by implication in a visual representation of the effects. Moreover, even putting aside the presence of the appropriate statistical test, fewer than half reported any information about the predictions or predicted probabilities, which are necessary to assess the substantive nature of the interaction effect (i.e., the direction and magnitude of the effect on the predictions).
- ³ Throughout, the term "effect" does not imply causality.
- ⁴ This can be easily accomplished in Stata by combining the "*over(group)*" option with the "*atmeans*" option when calculating the marginal effects with *margins*.
- ⁵ From a statistical standpoint, this represents that nominal variable's effect being weighted to represent the observed proportions in the sample. Therefore, the difficulty is with the interpretation, not with the statistical properties of the predictions themselves.
- ⁶ Stata's *margins* command easily incorporates this with the "*over(group)*" option.
- ⁷ Tables of regression coefficients for all of the examples presented in this article are included in the online supplement. I do not present the tables in text, as one of the key points of this article is that the coefficient estimates do not necessarily provide useful or accurate information about an interaction when the effects are nonlinear. However, I do consider it good practice to include the regression results as supplementary materials, as this helps readers understand what is included in the model. Appendices and online supplements are particularly helpful for this.

- 8 Stata handles the higher-order and constituent terms elegantly when using factor syntax. For example, `c. age##c. age##i. woman` is automatically interpreted by Stata to indicate each of the five coefficients listed in text.
- 9 Interestingly, the product term is necessary in this example for the effect to *not* be interactive. Without the product term in the model, this second difference is significant (i.e. a type I error).
- 10 The effect plotted is an instantaneous change (first derivative), which is calculated by `margins` in Stata with the `dydx()` option.
- 11 Note that as this is a plot of a marginal effect; a confidence interval that does not overlap zero indicates that the effect is statistically significant.
- 12 Norton, Wang, and Ai (2004) provide an alternative way of testing for the overall interaction effect via cross derivatives.

References

- Agresti, Alan. 2013. *Categorical Data Analysis*. Hoboken, NJ: Wiley.
- Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80:123–9. [https://doi.org/10.1016/S0165-1765\(03\)00032-6](https://doi.org/10.1016/S0165-1765(03)00032-6).
- Belia, Sarah, Fiona Fidler, Jennifer Williams, and Geoff Cumming. 2005. "Researchers Misunderstand Confidence Intervals and Standard Error Bars." *Psychological Methods* 10:389–96. <https://doi.org/10.1037/1082-989X.10.4.389>.
- Berry, William D., Jacqueline H. R. DeMerrit, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54:248–66. <https://doi.org/10.1111/j.1540-5907.2009.00429.x>.
- Berry, William D., Jacqueline H. R. DeMerrit, and Justin Esarey. 2016. "Bias and Overconfidence in Parametric Models of Interactive Processes." *American Journal of Political Science* 60:521–39. <https://doi.org/10.1111/ajps.12123>.
- Berry, William D., Matt Golder, and Daniel Milton. 2012. "Improving Tests of Theories Positing Interaction." *Journal of Politics* 74:653–71. <https://doi.org/10.1017/S0022381612000199>.
- Brambor, Thomas, Williams Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14:63–68. <https://doi.org/10.1093/pan/mpi014>.
- Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2018. "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models." *Annual Review of Sociology* 44:39–54. <https://doi.org/10.1146/annurev-soc-073117-041429>.
- Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CB09780511811241>.
- Cleveland, William. 1993. *Visualizing Data*. Summit, NJ: Hobart Press.
- Dalal, Dev K., and Michael J. Zickar. 2012. "Some Common Myths about Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression." *Organizational Research Methods* 15:339–62. <https://doi.org/10.1177/1094428111430540>.
- Dowd, Bryan E., William H. Greene, and Edward C. Norton. 2014. "Computation of Standard Errors." *Health Services Research* 49:731–50. <https://doi.org/10.1111/1475-6773.12122>.

- Efron, Bradley, and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.
- Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57:263–77. <https://doi.org/10.1111/j.1540-5907.2012.00602.x>.
- Heilman, Julia R., J. Scott Long, Shawna N. Smith, William A. Fisher, Michael S. Sand, and Raymond C. Rosen. 2011. "Sexual Satisfaction and Relationship Happiness in Midlife and Older Couples in Five Countries." *Archives of Sexual Behavior* 40:741–53. <https://doi.org/10.1007/s10508-010-9703-3>.
- Huber, Chuck. 2017. "In the Spotlight: Visualizing Continuous-by-Continuous Interactions with Margins and Twoway Contour." *Stata News* 32. Retrieved December 1, 2018 (<https://www.stata.com/stata-news/news32-1/spotlight/>).
- Jann, Ben. 2014. "Plotting Regression Coefficients and Other Estimates." *Stata Journal* 14:708–37. <https://doi.org/10.1177/1536867X1401400402>.
- Kaufman, Robert L. 2018. *Interaction Effects in Linear and Generalized Linear Models*. Thousand Oaks, CA: Sage.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44:347–61. <https://doi.org/10.2307/2669316>.
- Kromrey, Jeffrey D., and Lynn Foster-Johnson. 1998. "Mean Centering in Moderated Multiple Regression: Much Ado about Nothing." *Educational and Psychological Measurement* 58:42–67. <https://doi.org/10.1177/0013164498058001005>.
- Kuha, Jouni, and Colin Mills. 2018. "On Group Comparisons with Logistic Regression Models." *Sociological Methods and Research*, first published on January 7, 2018, as doi.org/10.1177/0049124117747306.
- Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- Long, J. Scott, and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press.
- Long, J. Scott, and Sarah A. Mustillo. 2018. "Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes." *Sociological Methods and Research*, first published on October 21, 2018, as doi.org/10.1177/0049124118799374.
- Mood, Carina. 2010. "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do about It." *European Sociological Review* 26:67–82. <https://doi.org/10.1093/esr/jcp006>.
- Mustillo, Sarah A., Omar A. Lizardo, and Rory M. McVeigh. 2018. "Editors' Comment: A Few Guidelines for Quantitative Submissions." *American Sociological Review* 83:1281–3. <https://doi.org/10.1177/0003122418806282>.
- Nagler, Jonathan. 1991. "The Effect of Registration Laws and Education on U.S. Voter Turnout." *American Political Science Review* 85:1393–405. <https://doi.org/10.2307/1963952>.
- Norton, Edward C., Hua Wang, and Chunrong Ai. 2004. "Computing Interaction Effects and Standard Errors in Logit and Probit Models." *Stata Journal* 4:154–67. <https://doi.org/10.1177/1536867X0400400206>.
- Pitblado, Jeff. 2014. "How Are Average Marginal Effects and Their Standard Errors Computed by Margins Using the Delta Method?" *Stata FAQ*. <https://www.stata.com/support/faqs/statistics/compute-standard-errors-with-margins/>.

- Rainey, Carlisle. 2016. "Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction." *Political Science Methods and Research* 4:621–39. <https://doi.org/10.1017/psrm.2015.59>.
- Thoits, Peggy A. 2011. "Mechanisms Linking Social Ties and Support to Physical and Mental Health." *Journal of Health and Social Behavior* 52:145–61. <https://doi.org/10.1177/0022146510395592>.
- Weber, Max. 1922. *Economy and Society: An Outline of Interpretive Sociology*. Oakland, CA: University of California Press.
- Williams, Richard. 2012. "Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects." *Stata Journal* 12:308–31. <https://doi.org/10.1177/1536867X1201200209>.

Acknowledgments: I thank J. Scott Long, Bianca Manago, Long Doan, and Josh Doyle for their helpful comments on previous drafts and Dave Armstrong and Shawn Bauldry for the many insightful conversations that influenced the content of the article.

Trenton D. Mize: Department of Sociology and Advanced Methodologies, Purdue University. E-mail: tmize@purdue.edu.