Compression and Conditional Effects

Carlisle Rainey[*]

Assistant Professor
University at Buffalo (SUNY)
Department of Political Science
carlislerainey.com
rcrainey@buffalo.edu

For the most recent draft of this manuscript,
see the project page at crain.co/compress.

Abstract

Previous research in political methodology argues that researchers do not need to include a product term in a logistic regression model to test for interaction if they suspect interaction due to compression alone. I disagree with this claim and offer analytical arguments and simulation evidence that when researchers incorrectly theorize interaction due to compression, models without a product term bias the researcher, sometimes heavily, toward finding interaction. However, simulation studies also show that models with a product term fit a broad range of non-interactive relationships surprisingly well, enabling analysts to remove most of the bias toward finding interaction by simply including a product term.

Introduction

Many theories in political science suggest an "interactive" relationship in which the effect of one explanatory variable X depends on the value of another explanatory variable Z (Clark, Gilligan, and Golder 2006; Berry, Golder, and Milton 2012). While earlier research lays out a clear and uncontroversial method for testing these claims in the context of linear regression (e.g., Kam and Franzese 2007; Brambor, Clark, and Golder 2006; and Friedrich 1982), political methodologists disagree about the best approach for researchers using logistic regression, particularly over

whether a product term must be included in the model (Nagler 1991; Frant 1991; Berry and Berry 1991; and Berry, DeMeritt, and Esarey 2010).[1]

The recent debate hinges around the substantive work of Wolfinger and Rosenstone (1980), who theorize that easing registration restrictions should have a smaller positive effect on the probability of voting among individuals with more education. They do not evaluate the presence of interaction by testing the statistical significance of the product term's coefficient. Indeed, they do not even include a product term in the model. Instead, they simply note that, even without a product term, a logistic regression model can represent the kind of interaction suggested by their theory because the S-shaped response curve creates "compression," requiring that all variables have smaller effects as the probability of an event approaches zero or one.

> As a statistical model, [logistic regression] is a more faithful representation of our substantive theory than [linear regression]. As we will see, the impact of most demographic variables is not constant across all types of individuals. Rather, the effect of a variable depends on the probability that the individual would vote. For example, a high-status occupation or a high-income has less impact on a college graduate who is 90 percent likely to vote, than it has on a high school dropout, who is only 55 percent likely to vote... With [logistic regression] a variable has very little impact on those who are either very unlikely or nearly certain to vote. It has the greatest impact in the middle of the distribution, on those who are between 40 and 60 percent likely to vote and are the most susceptible to the forces pushing them to vote or not to vote (11).

Instead of including a product term and testing the statistical significance of its coefficient, Wolfinger and Rosenstone (1980) argue for interaction by calculating a second-difference, which is the difference in the effect of easing registration restrictions between individuals with a high and low level of education. Formally, a second-difference $\Delta\Delta$ is the difference in the effect on $\Pr(Y)$ of changing X from a low value to a high value as Z changes from a low value to a high

---

[1]Throughout the manuscript, I focus on logistic regression models in particular, but the conclusions generalize to a wide range of models of binary outcomes that rely on S-shaped link functions, including probit models.

value, which can be defined mathematically as follows:

$$\Delta\Delta = [\Pr(Y|X = x_{hi}, Z = z_{hi}) - \Pr(Y|X = x_{lo}, Z = z_{hi})]$$
$$- [\Pr(Y|X = x_{hi}, Z = z_{lo}) - \Pr(Y|X = x_{lo}, Z = z_{lo})],$$

where x_{lo}, x_{hi}, z_{lo}, and z_{hi} take on substantively interesting values of the explanatory variables X and Z.[2] Using this approach, they find, as expected, that easing registration restrictions has a larger effect on the probability of turning out among less educated individuals.

However, Nagler (1991, see 1994 as well)) rejects this finding because the S-shaped curve automatically builds interaction into the model. He claims that "examining predicted probabilities generated by non-linear models such as [logistic regression] may produce spurious results when used to determine interactive effects between two [explanatory] variables (1393)." Instead of just calculating a second-difference to quantify the amount of interaction, Nagler prefers to argue for interaction by including a product term and testing the statistical significance of its coefficient. When he does this, the coefficient for the product of education and registration restrictions is incorrectly signed (suggesting that easing restrictions has a larger effect on the propensity to vote among the *most* educated), leading Nagler to conclude that the data do not support Wolfinger and Rosenstone's theory.

But Berry, DeMeritt, and Esarey (2010) defend Wolfinger and Rosenstone's approach against Nagler's challenge, arguing that "compression should not be viewed as theoretically irrelevant." Instead, they suggest that "it can be a strong theoretical rationale for expecting interaction between variables in their influence on [the probability of an event] (255)." Further, BDE suggest that if researchers expect interaction due to compression alone, then "there is no need to include a product term in the model (261)." In this situation, BDE recommend following

---

[2]For more details, see Berry, DeMeritt, and Esarey (2010). For a discussion of using simulation to calculate confidence intervals for second-differences, see King, Tomz, and Wittenberg (2000) and Hanmer and Kalkan (2013) for an alternative.

Wolfinger and Rosenstone's strategy of (1) excluding a product term, (2) computing a second-difference, and (3) calculating a confidence interval and verifying that it contains only values consistent with the research hypothesis (i.e., is statistically significant).[3]

Despite compression's theoretical relevance, I argue that researchers should *always* include a product term, even when they expect interaction on the basis of compression alone. A product term is not necessary to allow the model to represent interaction–it can do that with no product term (Berry, DeMeritt, and Esarey 2010). A product term must be included because it allows the model to better represent a *non-interactive* relationship. I present an analytical argument and simulations showing that if no product term is included, a logistic regression model cannot represent a non-interactive relationship and thus overstates the evidence for interaction. Fortunately, I find that simply including a product term eliminates most of this bias.

---

[3]Most of the debate about testing for interaction between two variables in influencing the probability of an event has occurred in political science. An earlier, but less detailed, debate between Berry and Berry (1990, 1991) and Frant (1991) almost exactly parallels the debate among Wolfinger and Rosenstone (1980), Nagler (1991), and Berry, DeMeritt, and Esarey (2010). The substantive application focuses on lottery adoption in the U.S. states, but Frant (1991) makes an argument similar to Nagler (1991) and Berry and Berry (1991) make an argument similar to Berry, DeMeritt, and Esarey (2010). Also in political science, Tsai and Gill (2013) and Huang and Shields (2000) consider interaction in the context of logistic regression, but focus on the calculation of marginal effects on the probability of an event without addressing the need for a product term or discussing contexts in which researchers should care about the probability of an event as opposed to the latent outcome. While the most detailed discussions of interaction in the context of logistic regression have occurred in political science, especially relating quantities of interest and the need for a product term to substantive theories, similar discussions have emerged in related disciplines. Work in economics focuses on the relationship between the sign and significance of the product term in logit and probit models and the sign and significance of the second-derivative $\dfrac{\partial^2 \text{Pr}(Y)}{\partial X \partial Z}$ (Ai and Norton 2003; Norton, Wang, and Ai 2004; and Greene 2010). However, this work offers little guidance as to when applied researchers should focus on $\dfrac{\partial^2 \text{Pr}(Y)}{\partial X \partial Z}$ as opposed to $\dfrac{\partial^2 Y^*}{\partial X \partial Z}$ and does not discuss the consequences of including or excluding a product term. Research in epidemiology works to define interaction in the context of logistic regression models, usually focusing on differences in risk-ratios rather than second-differences or second-derivatives. Again, though, this literature offers little guidance about whether researchers should include or exclude a product term in logistic regression models used to test for interaction. Also outside of political science, Bowen (2012) draws a sharp distinction between the interaction induced by the inclusion of a product term and the interaction induced by the link function and offers a statistical test to distinguish the two forms of interaction, but does not address whether a product term is necessary to conclude that interaction is present and does not offer advice about the advantages of focusing on the probability of an event instead of the latent outcome variable.

Current Advice to Applied Researchers

Prior to the publication of Berry, DeMeritt, and Esarey (2010), the literature suggested that when analysts faced a dichotomous outcome and suspected interaction, there was only one path forward: include a product term and conclude that there is interaction between X and Z if and only if the coefficient for the product term was statistically significant (Nagler 1991; Brambor, Clark, and Golder 2006, though see Berry and Berry 1991). However, Berry, DeMeritt, and Esarey (2010, 2014) have added much needed nuance to the advice this advice. In particular, Berry, DeMeritt, and Esarey (2010) lay out an initial framework describing the contexts in which researchers should hypothesize in terms of the latent variable Y^* or $\Pr(Y)$, describe the implications for whether one performs hypothesis tests on the coefficient of the product term or a second-difference, and make suggestions about when to include a product term. Table 1 summarizes the current advice to applied researchers.

I find the advice summarized in Table 1 sound, useful, and large improvement over the coarse practice prior to Berry, DeMeritt, and Esarey (2010). However, I disagree with the recommendation to exclude a product term when the analyst puts forward a strong theory predicting interaction on the basis of compression alone (the bold "No" in Table 1). In this paper, I explain why I find Berry, DeMeritt, and Esarey's (2010) argument for excluding a product term incomplete.

The Argument for Excluding a Product Term

Berry, DeMeritt, and Esarey's (2010) (hereafter, BDE) overarching claim, which I wholeheartedly agree with, is that "a statistically significant product term is neither necessary nor sufficient for concluding that there is substantively meaningful interaction among independent variables in their influence on [the probability of an event]" (261). Given this point though, BDE ask

Situation	Description	Include a Product Term?	Quantity of Interest	Source
Interaction in Influencing the Latent Outcome	Guided by a strong theory, the analyst hypothesizes that X and Z interact in influencing the latent outcome variable Y^*. For example, it sometimes makes sense to conceptualize Y^* as utility and derive a probit model using a random utility framework (Train 2009). See Berry, DeMeritt, and Esarey (2010, esp.pp. 261-262) for more details and an example.	Yes	$\dfrac{\partial^2 Y^*}{\partial X \partial Z}$	Nagler (1991)
Interaction Due to Compression Alone	Guided by a strong theory, the analyst hypothesizes that X and Z interact in influencing $\Pr(Y)$ due to compression alone. That is, as the probability of an event approaches zero or one, the effect of any explanatory variable (including X and Z) have smaller effects. While researchers such as Frant (1991) and Nagler (1991) sometimes dismiss compression as an unimportant form of interaction, Berry, DeMeritt, and Esarey (2010) make a strong case that this type of interaction is often theoretically meaningful.	**No**	$\dfrac{\partial^2 \Pr(Y)}{\partial X \partial Z}$	Berry, DeMeritt, and Esarey (2010)
Specification Ambiguity	Guided only by weak theoretical intuition, the analyst hypothesizes that X and Z interact in influencing $\Pr(Y)$, but have no strong theoretical rational for the functional form. In this situation, the analyst lacks the theoretical guidance necessary to theorize about interaction in terms of the latent variable or on the basis of compression alone.	Yes	$\dfrac{\partial^2 \Pr(Y)}{\partial X \partial Z}$	Berry, DeMeritt, and Esarey (2014)

TABLE 1: This table summarizes three situations in which analysts might theorize interaction in the context of logistic regression. Notice that the quantity of interest varies across these situations, as well as whether the literature recommends including a product term. In this paper, I address the situation in which an analyst suspects interaction due to compression alone. In particular, I agree with Berry, DeMeritt, and Esarey's (2010) characterization of the theoretical value of compression and the implied quantity of interest, but argue that analysts must still include a product term.

whether it is always necessary to include a product term in the model. To decide whether to include a product term, they suggest that analysts draw a sharp distinction between theories that focus on the unbounded latent variable, denoted as Y^* (perhaps conceptualized as "utility"), and theories that focus on the probability of an event, denoted as $\Pr(Y)$.[4]

> *BDE Recommendation*: When guided by a strong theory, the decision to include a product term must be based on the theorized effects of variables on the unbounded latent dependent variable Y^*, assumed by the model.

> *Part (a)*: If the theory assumes that variables interact in influencing the latent unbounded variable Y^*, then the researcher should include a product term. (See the situation "Interaction in Influencing the Latent Outcome" in Table 1.)

We agree on *Part (a)*. As both BDE and Nagler (1991) note, this situation is entirely analogous to linear regression (e.g., $Y^* = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$), making a product term essential.

> *Part (b)*: If the theory assumes that variables should interact in influencing $\Pr(Y)$ strictly on an expectation of compression, then there is no need to include a product term in the model. (See the situation "Interaction Due to Compression Alone" in Table 1.)

BDE explain:

> [I]n any logit or probit model, the marginal effect of X on $\Pr(Y)$ depends on the values of all [explanatory] variables; this marginal effect is greatest when $\Pr(Y)$ is 0.5 and declines when a change in either variable pushes $\Pr(Y)$ toward 0 or 1. We refer to the phenomenon ... as compression, because deviations of $\Pr(Y)$ away from [0.5] compress further possible change in Pr(Y) to ever-smaller ranges. (251)

They continue by suggesting:

> [A] researcher...may base his hypothesis that independent variables interact in influencing $\Pr(Y)$ strictly on an expectation of compression. In this case, there is

---

[4]In BDE's notation, Y^* refers to the so-called "latent variable" implied in models of binary outcomes. Y^* simply equals the value of the linear predictor, often denoted by $\eta = X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$ in the notation of generalized linear models, and is transformed into $\Pr(Y)$ by an inverse link function that maps the real line to $[0, 1]$. Sometimes researchers' theories lead to hypotheses about Y^* rather than $\Pr(Y)$. For example, theorists often link formal models of decision-making to random utility models, a common framework for deriving models such as logistic regression, particularly the almost identical probit model. Using this approach, researchers can evaluate explanations of utility, represented by the value of $Y^* = \eta = X\beta$.

no need to include a product term in the model. Put differently, no product term is required if the analyst's only reason for positing interaction between X and Z in influencing $\Pr(Y)$ is an expectation that when the value of X is extreme, $\Pr(Y)$ is near its limit of zero or one and thus there is little room for $\Pr(Y)$ to change as Z changes. In any event, the analyst should use the parameter estimates for the model (with no product term) to generate estimates of the effects of variables on $\Pr(Y)$ to test his hypothesis; the same tools used for models with a product term–second-differences or marginal effect plots–can be employed. (261)

While I find BDE's suggestion to exclude a product term problematic, I take two other points from their paper as a starting point.

1. Compression can be substantively interesting and theoretically meaningful. Consider the example from Wolfinger and Rosenstone (1980) discussed above. Recall that their model does not have a product term, the only interaction is due to compression. Yet if their model is correct , then restrictive voting rules not only reduce turnout, but also increase the inequality of political participation because restrictive rules have a larger negative effect on the probability of voting among the least educated. This relationship has enormous normative consequence and policy relevance.

2. If analysts are interested in interaction due to compression, then the quantity of interest is not the product term, but the second-difference (or second-derivative). Several authors show that the sign and significance of the second-difference does not depend on the sign and significance of the product term (e.g., Tsai and Gill 2013; Ai and Norton 2003).

Although, I find BDE's argument for these two points compelling, I disagree with the advice to exclude a product term when theorizing interaction on the basis of compression alone.

The Argument for Including a Product Term

Researchers must employ a statistical model that can represent (1) relationships that are *consistent* with the theory and (2) plausible relationships that are *inconsistent* with the theory. It

is obvious that models should be able to represent relationships that are consistent with theory. But if researchers use models that cannot also represent situations *inconsistent* with the theory, then they have assumed their theories are correct. Though the theoretical rationale may be strong, it has not been made vulnerable to the data.

Thus, BDE's analysis is incomplete when they write that "no product term is required if the analyst's only reason for positing interaction between X and Y in influencing $\Pr(Y)$ is an expectation that when the value of X is extreme, $\Pr(Y)$ is near its limit of zero or one, and thus there is little room for $\Pr(Y)$ to change as Z changes." While they correctly note that the theorized relationship can be represented by a model with no product term, they fail to discuss whether this model can also represent relationships that are inconsistent with their theory.

Without a product term, a logistic regression model cannot represent many reasonable situations that are inconsistent with an interactive theory. In particular, it cannot represent a situation in which two explanatory variables have non-zero effects and the probability of an event is greater or less than one-half, but there is no interaction. Specifically, for a logistic regression model including X and Z, but excluding their product, if all of the following conditions hold, then X and Z must interact in influencing $\Pr(Y)$. [5]

1. X has a non-zero effect.

2. Z has a non-zero effect.

3. $\Pr(Y) < 0.5$ or $\Pr(Y) > 0.5$.

This can easily be seen by inspecting the second derivative of the logistical regression model with no product term, given by

$$\frac{\partial^2 \Pr(Y)}{\partial X \partial Z} = \Pr(Y)[1 - \Pr(Y)][1 - 2\Pr(Y)]\beta_x \beta_z. \tag{1}$$

---

[5]For further discussion, see Nagler (1991) and Berry, DeMeritt, and Esarey (2010).

If any of Conditions 1-3 hold, then the second derivative is exactly zero. Otherwise, the second derivative is non-zero, which implies an interactive relationship the model. Consequentially, if the researcher incorrectly theorizes compression, but 1-3 hold, then a logistic regression model without a product term will overstate the evidence for interaction.[6]

With a product term, however, a logistic regression model can represent *some* plausible situations inconsistent with an interactive theory. For a logistic regression model including X, Z, and XZ, if all of the conditions (1-3) identified above hold, then X and *Z may or may not* interact in influencing $\Pr(Y)$. The intuition of this claim can be seen by simply inspecting the second derivative of a logistic regression model including a product term, given by

$$
\begin{aligned}
\frac{\partial^2 \Pr(Y)}{\partial X \partial Z} &= \Pr(Y)[1 - \Pr(Y)]\beta_{xz} \\
&+ \Pr(Y)[1 - \Pr(Y)][1 - 2\Pr(Y)](\beta_x + \beta_{xz}Z)(\beta_z + \beta_{xz}X).
\end{aligned}
\tag{2}
$$

The presence of the product term creates another coefficient that allows the second derivative to be zero, even if Conditions 1-3 hold. Thus, this model is capable of representing situations in which 1-3 hold, but no interaction is present. This means that a product term "uncouples"

---

[6]Responding to Frant's (1991) suggestion to include a product term, Berry and Berry (1991) write:

> Frant claims that this result [that the effect of election proximity on the probability of lottery adoption depends on the state's fiscal health] "sounds like a *conclusion* (and the authors' treat it as such); but it is actually an *assumption* of the model." It is true that interaction among among variables in their effects on the probability of adoption is an assumption of our probit model. But all multivariate models make assumptions about the function form of effects of independent variables; yet substantive conclusions are drawn from empirical analyses of these models. Indeed, an interactive regression with a multiplicative term (e.g., $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$)–similar to the form that Frant proposes we use–also *assumes* interaction among the independent variables (provided $\beta_3 \neq 0$) (574).

My analytical argument supports Frant's (1991) position. While Berry and Berry (1991) are correct that the linear model including a product term assumes a form of interaction (e.g., the marginal effect of X_1 is a *linear* function of X_2), it does not assume that interaction is present (hence their caveat that $\beta_3 \neq 0$). A linear model with a product term can represent situations that are consistent and inconsistent with the interactive theory ($\beta_3 \neq 0$ and $\beta_3 = 0$, respectively). This allows the model to effectively assess the evidence against the null hypothesis and for the research hypothesis. As the analytical argument shows, the logistic regression model with no product term, while it can represent an interactive relationship, cannot represent a non-interactive relationship. Both my anlytical arguments and simulations strongly suggest that Frant (1991) is correct when he advises researchers to avoid making empirical claims about interaction between variables in influencing the probability of an event when no product term is included in the model.

hypotheses about the effects of X and Z from their interaction.

While including a product term does not entirely remove the bias toward confirming interactive hypotheses, it does greatly reduce the bias. Thus, including a product term makes the empirical argument more compelling by making the theory more vulnerable to the data.[7]

Simulations

To illustrate the potential consequences of excluding a product term, I report two simulation studies. The first examines a single relationship in detail. The second looks at a wide range of potential relationships. Because I am particularly interested in what happens if the theory (or researcher's hypothesis) is wrong, I imagine that a researcher incorrectly theorizes interaction due to compression and follows BDE's suggestion to exclude the product term. In almost all cases, excluding the product term biases the researcher toward finding support for her theory.

But why focus specifically on what happens if the researcher is wrong? Why not look at what happens if the researcher correct theorizes interaction on the basis of compression alone. Careful readers recognize the tradeoff between the size of a hypothesis test (i.e., the probability of finding interaction when there is none) and power (i.e., the probability of finding interaction when it is present). While analysts usually face a tradeoff between these two, I focus my attention squarely on the size of tests for interaction.

Hypothesis test form compelling statistical arguments because, by design, they rarely lead a researcher to reject the null hypothesis when they null hypothesis is true. This is because hypothesis tests are constructed by fixing $\alpha = \Pr(\text{reject null}|\text{null is true})$ at an acceptably low value, typically $\alpha = 0.05$. This is important because researchers and readers interpret decision to reject the null hypothesis as strong evidence against the null hypothesis (or for the research hypothesis).

---

[7]This analytical argument provides a formal justification for Huang and Shield's (2000) claim that "[o]f course, adding multiplicative terms to explicitly model this interactive relationship is necessary (86)."

Social scientists typically use hypothesis tests because they want to conclude that their research hypothesis is correct only when the data offer strong evidence in favor of their research hypothesis (or equivalently, against their null hypothesis). If a test is improperly sized (e.g., $\alpha = 0.75$), then the decision to reject the null hypothesis is no longer compelling. Making matters worse, readers and reviewers are still likely to interpret the results of this over-sized test as compelling. I focus squarely on the size of tests for interaction because the logic of hypothesis testing requires that the size be fixed at a suitably small value, usually 0.05 in the social sciences, while allowing the power of a test to fluctuate with other variables, such as the number of observations. My primary focus on size, or false-positive errors, such not be taken as meaning that power is unimportant. To address the secondary issue of power, I include a discussion of the statistical power of models that exclude and include product term in the Appendix.[8] One might also worry about the impact of including a potentially unnecessary product term XZ in the model (i.e., $\beta_{XZ} = 0$) on the estimated effects of other variables in the model. In the Appendix, I show that including an unnecessary product term leads to no bias in the estimated coefficients for W or the first differences as W moves from a low value to a high value.

Study #1: A Fixed Relationship

I first consider the non-interactive relationship (i.e., data-generating process) $\Pr(Y) = 0.3 + 0.2X - 0.3Z$ shown in the left panel of Figure **??**. I imagine that the analyst correctly suspects that X has a positive effect on $\Pr(Y)$ and Z has a negative effect on $\Pr(Y)$. She also theorizes that because the event is relatively rare and Z pushes $\Pr(Y)$ even closer to zero, there is less room

---

[8]These simulations reported in the Appendix show that the ability of a model with a product term to "detect" an interactive relationship (i.e., statistical power) is nearly equal if the amount of interaction is substantial or the sample size is large. However, it is important to point out that when the sample size is small (e.g., less than 1,000) *and* the amount of interaction is small (e.g., second-difference of $0.04 - 0.09 = -0.05$), then including a product term might lead to a substantial drop in statistical power. However, it is important to note that the gain in statistical power comes from *assuming* an interactive relationship, not *detecting* one.

for X to affect $\Pr(Y)$. Therefore, she expects that X should have a smaller effect when $Z = 1$. But this hypothesis is not correct–she *incorrectly* expects that X and Z interact in influencing $\Pr(Y)$ due to compression. Still, she notices that a logistic regression model can represent the kind of interaction suggested by her theory and decides that no product term is necessary (Berry, DeMeritt, and Esarey 2010).

Because X is approximately uniformly distributed from zero to one and Z is dichotomous with half of the observations equal to one and the other half equal to zero, she decides to test for interaction by estimating a logistic regression model with no product term, calculating the second-difference as X changes from 0.25 to 0.75 and Z changes from zero to one. Her hypothesis suggests that this second-difference should be negative, since X should have a smaller effect when $Z = 1$.
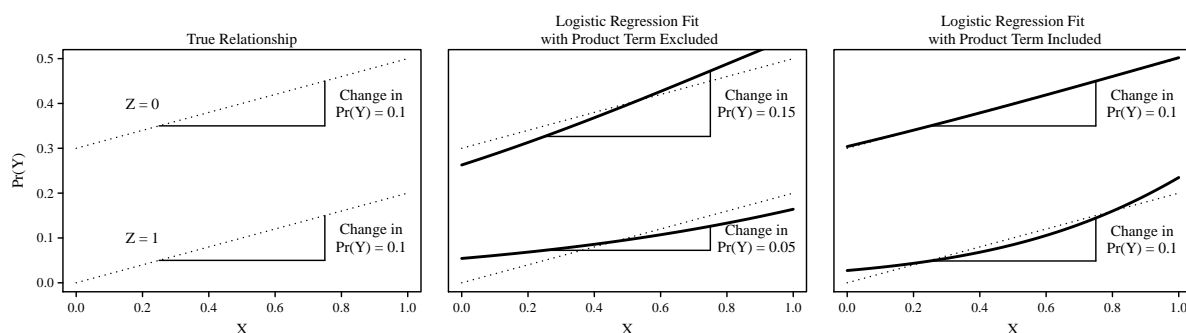

FIGURE 1: This figure shows the true relationship in the first simulation study (dotted lines) as well as how logistic regression models with and without a product term fit the process (solid, bold lines). Notice that X and Z have large effects on $\Pr(Y)$ and $\Pr(Y)$ is always less than 0.5. In this case, excluding a product term forces the model to point toward interaction, despite the fact that the relationship is non-interactive. Including a product term allows the model to fit this non-interactive process surprisingly well.

Notice the middle panel of Figure 1, which shows how a logistic regression model without a product term might fit the true relationship. Although there is no interaction in the actual data-generating process, the logistic regression model with no product term suggests that there

is substantial interaction between X and Z in influencing $\Pr(Y)$. When $Z = 1$, the estimated effect of moving X from 0.25 to 0.75 is 0.05. When $Z = 0$, however, the effect of X triples to 0.15 (the second-difference is $0.05 - 0.15 = -0.1$). Although the hypothesis of interaction is wrong, the logistic regression model with no product term suggests there is strong evidence in favor of it.

Now notice the right panel of Figure 1, which shows the estimates when the model contains a product term. While this model does not exactly capture the true data-generating process (i.e., notice the non-linearity when $Z = 1$), it estimates the amount of interaction very well. The estimated effect of X when $Z = 0$ is 0.1. When $Z = 1$, the estimate is identical. Therefore, the estimated second-difference is zero ($0.1 - 0.1 = 0$) and the model correctly suggests that there is no interaction.

Rather than examining how models fit the data-generating process, it is perhaps more informative to evaluate each approach in repeated trials. I simulate 3,000 data sets containing 1,000 observations each. I use each data set to estimate a second-difference and its confidence interval using a model with and without a product term. Finally, I estimate the probability of finding interaction when none is present (i.e., a Type-I or false-positive error) by calculating the proportion of data sets that pointed toward interaction with each approach (i.e., the confidence interval does not contain zero). By convention, this probability should be close to 0.05. When the model does not include a product term, the probability of finding interaction is 0.98. Thus, finding a statistically significant second-difference is not a all compelling evidence against the null when the model does not include a product term, since the researcher will likely find this even when there is no interaction. However, the probability of finding interaction drops to 0.06 when the researcher includes a product term in the model. This suggests that finding a statistically significant second-difference is compelling when the model includes a product term since the researcher would rarely find this when the null is true.

To help understand why excluding a product term biases the researcher toward finding

interaction, Figure 2 shows the estimates and confidence intervals from the first 50 simulated data sets, sorted by the size of the estimated second-difference. Notice the left panel of Figure 2, which shows that the models without a product term always estimate a substantial amount of interaction in the expected direction and that the confidence intervals rarely overlap zero. *This happens because compression serves as the theoretical motivation for the hypothesis and is <u>assumed</u> by the model.* But notice the right panel of Figure 2, which shows that most of this bias disappears when a product term is included in the model.
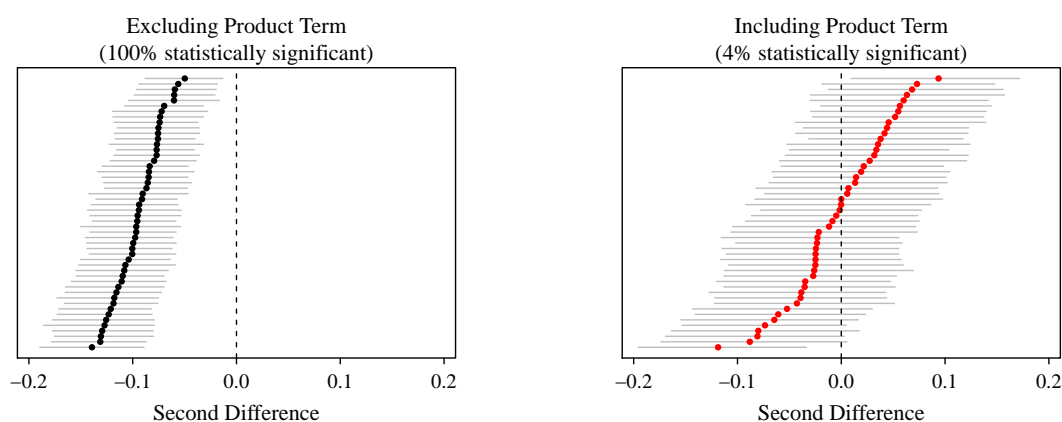
FIGURE 2: This figure shows fifty simulations of estimated second-differences and confidence intervals using models with and without a product term. Although second-difference is actually zero (a non-interactive relationship), the model with no product term consistently finds interaction. Including a product term removes almost all of this bias.

To get a sense of the robustness of this conclusion, I systematically vary the sample size, the distribution of X, and the change in X used to calculate the second-difference while the true relationship remains the same. For each combination of the simulation parameters, I simulate 3,000 data sets. For each data set, I estimate a logistic regression model with and without a product term and compute the confidence interval around the second-difference, as BDE recommend. I then calculate the proportion of simulations in which each model points
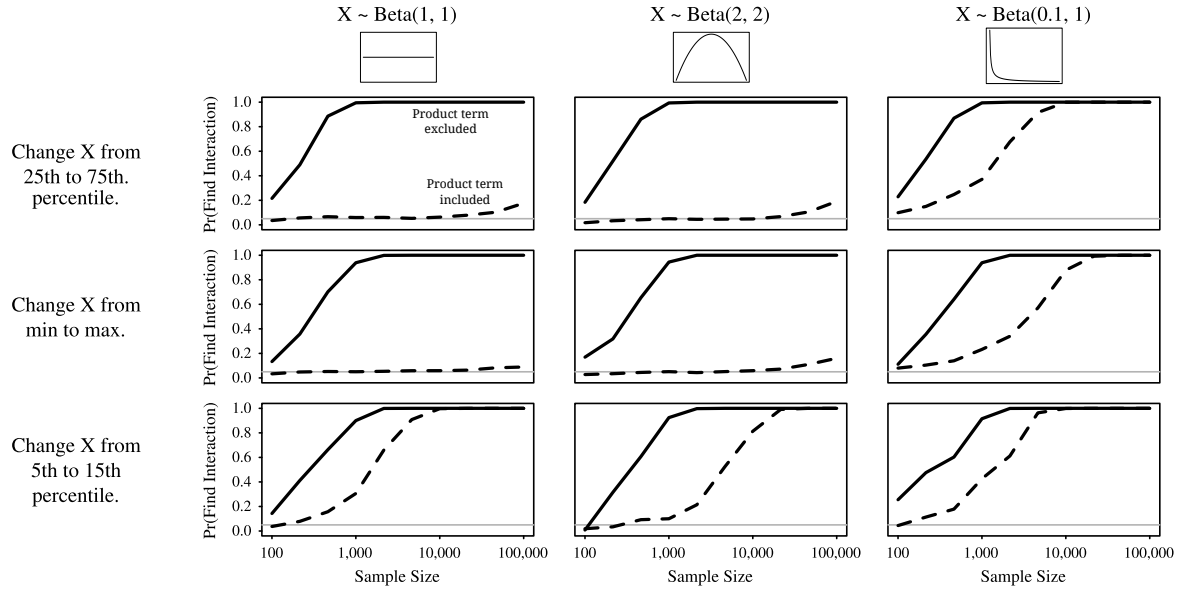
FIGURE 3: This figure shows the probability of concluding interaction when the true relationship is given by $\Pr(Y) = 0.3 + 0.2X - 0.3Z$ (i.e., there is no interaction) using logistic regression models with and without a product term as the sample size, the distribution of X, and the change in X considered in computing the second-difference varies. Notice that the model including a product term performs remarkably well, although the improvement diminishes as the simple size gets large, especially when the distribution of X is skewed and the researcher considers the effect of changing X near its extremes.

incorrectly toward interaction to estimate the probability of finding interaction when none is present. The results are presented in Figure 3.

Across all values of the simulation parameters, the models without a product term bias the researcher toward finding interaction and sometimes strongly. The bias is reduced, usually dramatically, when the researcher includes a product term. Indeed, for symmetric distributions of X and centered second-differences, adding a product term to the model eliminates nearly all of the bias, lowering the probability of finding interaction from nearly 1.00 (the worst situation) to about 0.05 (the ideal situation).[9]

---

[9]As a robustness check, I repeated this study using a slight different data generating process. For this alternative, "rare events" process, the probability of an event never rises above 0.08. The full results are discussed in the Appendix, but the basic findings suggest an even more optimistic conclusion about the ability of the product term to remove the bias toward finding interaction. This is as expected since I chose the fixed process discussed above to serve as a worst case scenario.

16

Study #2: A Diverse Set of Relationships

To demonstrate that the patterns seen in Figure 3 do not depend on any particular relationship, I now examine a diverse collection of 1,500 simulated analyses. Each analysis is generated through a random process (described in the Online Appendix) that varies in the following ways:

1. The relationship between $\Pr(Y)$, X, and Z varies, but follows the form of $\Pr(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X^2 + \beta_4 Z^2$ and is always monotonic across the range of X and Z. Importantly, X and Z never interact in changing $\Pr(Y)$.[10]

2. The distributions of X and Z vary and might be binary, flat, unimodal, or skewed to the right.

3. The values of x_{lo}, x_{hi}, z_{lo}, and z_{hi} used to construct the second-difference vary from zero to one. All combinations are possible, though if a variable is binary, its values are set to zero and one.

4. The sample size varies, ranging from 500 to 100,000, though smaller samples are more likely.

Each "analysis" consists of features the researcher totally controls (e.g., values of x_{lo}, x_{hi}, z_{lo}, and z_{hi}), somewhat controls (e.g., sample size), and does not control (e.g., distributions of explanatory variables and the true data-generating process). For each simulated analysis, I use the same procedure as before to estimate the probability of finding interaction using models with and without a product term. Because none of the true relationships are interactive, this probability should be near 0.05.

The left panel of Figure 4 shows that when the researcher uses a logistic regression model without a product term, the probability of finding interaction when none exists is less than 0.1 for only 40 percent of the simulated analyses. It is greater than 0.9 for over 20 percent of the analyses. This suggests that excluding a product term strongly biases the researcher

---

[10]I use a quadratic functional form here to include a more general set of curves rather than restrict my processes to straight lines. The only condition that I want to place on the processes is that they be monotonic (i.e., the marginal effects never change signs) and there be no interaction. The quadratic form allows a nice mixture of different relationships within these restrictions (though not exhaustive, of course). The restriction disallowing interaction is necessary because I am interested mainly in the probability of finding interaction when their is none present.

toward finding interaction for many plausible relationships. On the other hand, the right panel of Figure 4 shows that, when using a model with a product term, the probability of finding interaction is less than 0.1 for almost all of the simulated analyses (over 98 percent). This suggests that including a product term can eliminate most of the bias toward finding interaction for many plausible relationships and analyses.
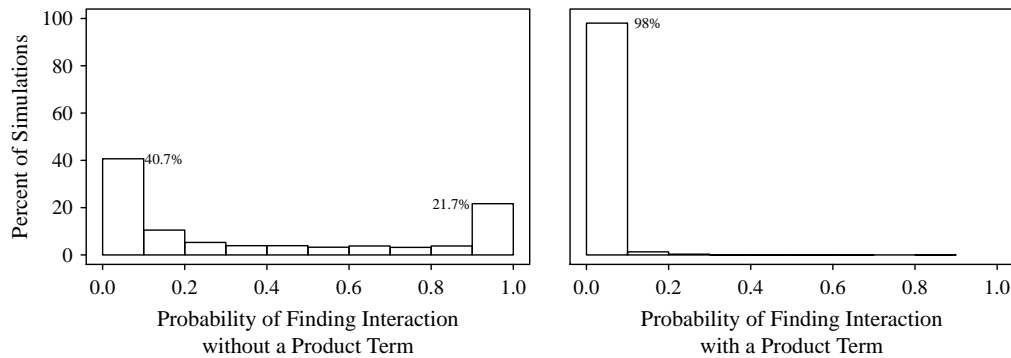

FIGURE 4: The histogram on the left shows that logistic regression models without a product term tend to find interaction when none exists. The histogram on the right shows researchers can eliminate most of this bias by simply including a product term.

Figure 5 directly compares the two approaches. Notice that every point falls below the diagonal line, except when both approaches fall below 0.1 (i.e., both perform well). This means that in every situation in which excluding a product term led to substantial bias, the model with a product term reduced this bias. In no case did the model without a product term perform substantially better than the model with a product term. Regardless of the absolute performance of each, the model with a product term performs at least as well as the model without a product term in all the analysis generated in my simulation, except when both perform well. In cases in which the model with a product term performs somewhat poorly (e.g., the probability of finding interaction is about 0.3, for example), the model without a product term performs terribly (e.g., the probability of finding interaction is nearly 1.00).
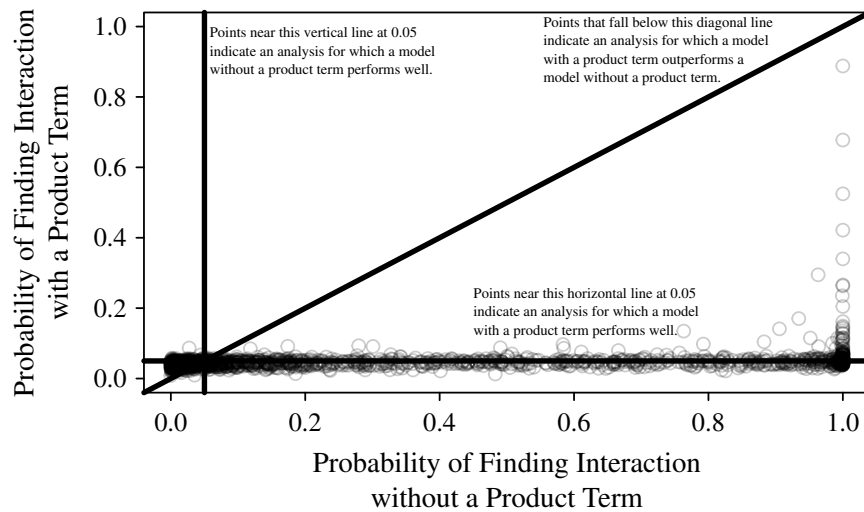
18

FIGURE 5: This figure shows the probability of concluding interaction when none exists from logistic regression models with and without a product term for many different analyses. Notice that many points fall away from the vertical line at 0.05 indicating poor performance of models without a product term. Also notice that most points fall near the horizontal line at 0.05, indicating that models with a product term tend to perform well. Finally, notice that all points fall below the diagonal line, indicating that models without a product term are more likely to mislead researchers in every simulated analysis.

Recommendations for Applied Researchers

I suggest that applied researchers take several steps to ensure their claims of interaction are empirically meaningful.

1. Clearly explicate your interactive theory (Clark, Gilligan, and Golder 2006; Berry, Golder, and Milton 2012) and carefully consider which relationships are consistent and inconsistent with the theory. Only use empirical models that can represent both relationships.

2. As Berry, DeMeritt, and Esarey (2010) suggest, specify whether the relevant outcome variable is the latent variable Y^* or the expected value $\Pr(Y)$. If you use the concept Y^*, then test for interaction by testing the statistical significance of the product term. If you rely on $\Pr(Y)$, then test for interaction by simulating confidence intervals around a carefully chosen second-difference or second-derivative. See Berry, DeMeritt, and Esarey (2010) for the details of calculating a second-derivative or second-difference and King, Tomz, and Wittenberg (2000) for the details of the simulation procedure. Note that

Hanmer and Kalkan (2013) offer an alternative approach in which researchers average the quantity of interest across the observed data rather than choosing a specific comparison of interest.

3. Regardless of whether the theoretically relevant outcome is Y^* or $\Pr(Y)$, and even if you theorize interaction on the basis of compression alone, you must include product terms in order to draw compelling substantive conclusions about interaction from logistic regression models. Otherwise, you are simply pulling assumptions through data and treating them as conclusions.

4. As Nagler (1991) and Frant (1991) note, if you do not include a product term, then avoid making claims about interaction. The interaction is assumed and you should understand and clearly identify it as such.[11]

Democracy, Distance, and the Probability of Conflict

To illustrate how to conduct a compelling test for interaction due to compression, I replicate analyses by Oneal and Russet (2001).[12] Berry, DeMeritt, and Esarey (2010) explain that the probability of war between two democracies is nearly zero, but much higher if one or both countries is non-democratic. Thus, on the basis of compression alone, Berry, DeMeritt, and Esarey (2010) argue that one can meaningfully hypothesize that explanatory variables should have smaller effects on the probability of conflict when both countries are democratic, simply because this probability is so close to zero. Using the example of geographic distance, the authors note that we should expect distance to have a much smaller effect on the probability of conflict when both countries are democracies than when at least one country is not democratic. Importantly, we have no reason to expect that democracy and distance interact in influencing

---

[11]Note that while researchers might suspect interaction between two variables (perhaps these are control variable), it is not necessary to include the product term as long as no empirical claim is made about the interaction of these variables. For example, it is not necessary for future studies of conflict to include the product of democracy and distance in their logistic regression models unless the researchers are specifically interested making compelling empirical arguments that the two interact in influencing the probability of conflict.

[12]While Oneal and Russet (2001) estimate several different model using generalized estimating equations, I estimate only their simplest logistic regression model (p. 314, Table A3.1) using maximum likelihood and the usual standard errors. Using their more complex approach does not substantively alter the results, though it does increase the standard errors slightly.

the latent outcome (e.g., the utility of conflict). Thus, we have no reason to theorize about the sign of the product term. Further, the only reason to include a product term is to allow the model to represent a situation in which we are wrong (i.e., there is no interaction between democracy and distance in influencing the probability of conflict).

This leads to three hypotheses about the effects of distance and democracy on the probability of conflict. First, as has been commonly suggested (e.g., Ray 1998; Tomz and Weeks 2013), democracies are much less likely to fight each other.

> DEMOCRACY HYPOTHESIS: Regardless of the distance between the two countries in a dyad, as the joint level of democracy in the dyad increases, the probability of conflict in that dyad decreases.

Second, other literature in international relations points out that states that are located geographically further from one another are less likely to engage in conflict (e.g., Diehl 1991; Bremer 1992).

> DISTANCE HYPOTHESIS: Regardless of the level of democracy in the dyad, as the distance between the countries in the dyad increases, the probability of conflict in that dyad decreases.

As Berry, DeMeritt, and Esarey (2010) point out, these two hypotheses imply a third hypothesis. Because democratic dyads are much less likely to engage in conflict, we should expect the negative effect of distance to be much smaller in these dyads simply due to compression.

> INTERACTION HYPOTHESIS: The negative effect of distance on the probability of conflict is larger (more negative) in non-democratic dyads than in democratic dyads.

Results

To test these three hypotheses, I use data from Oneal and Russet (2001). Their logistic regression model of conflict (militarized interstate disputes) includes variables for democracy (smaller of

the two Policy IV scores in each dyad), distance (the number of miles between the capitals), and several control variables. I estimated one model using Oneal and Russet's exact specification (with no product terms) and another model that includes the product of democracy and distance.[13] See the Appendix for the details and the coefficient estimates. The estimates from models with and without a product term are presented in the Appendix. Since each of the hypotheses are stated in terms of the probability of conflict, I compute and plot estimates and 90% confidence intervals for predicted probabilities, first-differences, and second-differences using CLARIFY-like simulation (King, Tomz, and Wittenberg 2000)

Figure 6 provides a nice overview of the intuition and the evidence for the hypotheses by plotting the estimated probably of conflict in both democracies and non-democracies from each empirical model.[14] Notice that the results are consistent with all three hypotheses. First, increasing the distance between the states in a dyad reduces the probability of conflict in that dyad in both democratic (smallest Polity IV score is 10) and non-democratic (smallest Polity IV score is -10) dyads. Second, notice that the probability of conflict is lower in more democratic dyads, regardless of the distance between the states in the dyad. Finally, notice that the pacifying effect of distance is much smaller (i.e., less negative) in democratic dyads.

---

[13] Oneal and Russet use generalized estimating equations (GEE) to estimate their models, but I rely on the simpler approach of maximum likelihood. The two methods produce substantively similar results, but the standard errors from the GEE approach are slightly larger.

[14] Notice that the estimated relationships are similar across both modeling strategies. However, I have shown above that the model with no product term forces an interactive relationship into the results if both variables have an effect and the relevant probabilities are above or below one-half.
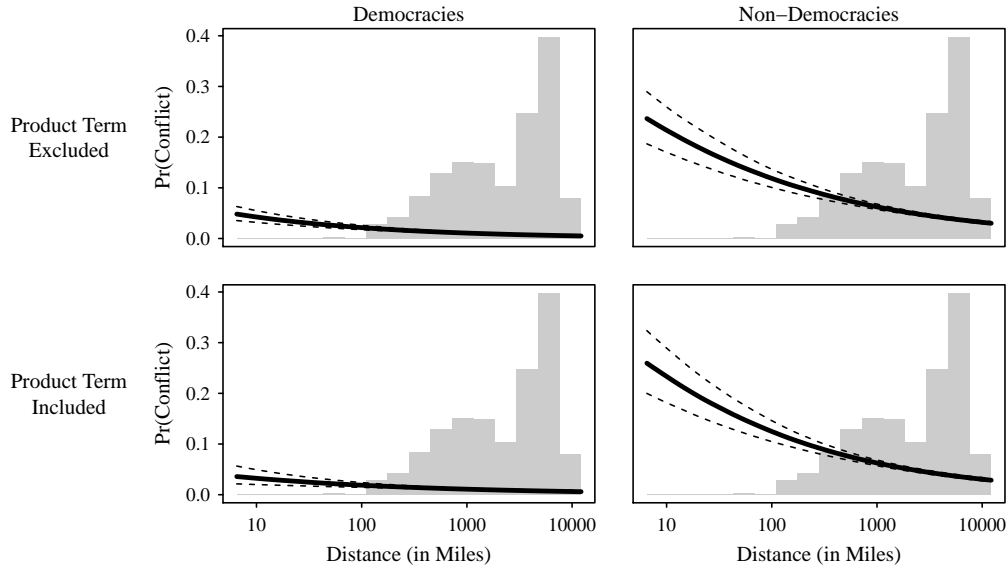
FIGURE 6: This figure compares the estimated probability of conflict in democracies (Polity IV score of 10) and non-democracies (Polity IV score of -10) from models with and without a product term. Notice that, as expected, there is a much smaller change in the probability of conflict in democracies. However, the model with no product term forces this pattern of compression into the estimates. The model including a product term allows the model to represent non-interactive relationships as well. Because of this, the estimates from the model with a product term are more compelling.

While the pattern of the predicted probabilities is consistent with the three hypotheses, I now compute and plot the quantities of direct substantive interest and the confidence intervals. First, Figure 6 plots the effect of moving from a non-democratic dyad (lower Polity IV score is -10) to a democratic dyad (lower Polity IV score is 10) on the probability of conflict as distance varies. The Democracy Hypothesis suggests that this effect should be negative regardless of the distance between the states. Figure 6 shows that the expected negative effect holds for across all distances and the confidence interval never overlaps zero. This pattern holds across the model with and without a product term.
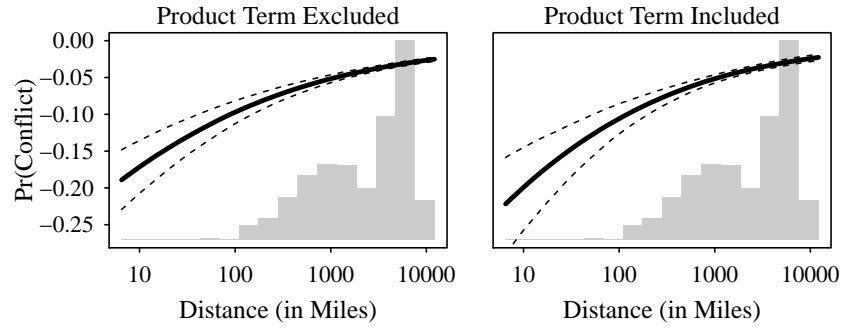
FIGURE 7: This figure shows how the negative effect of distance on the probability of conflict shrinks toward zero as dyads become more democratic. This expected pattern is consistent with compression, but the model without a product term forces this pattern into the results while the model with a product term allows other types of relationships to emerge.

Next, Figure 8 plots the effect of moving from a "nearby dyad" (about 850 miles apart, which is the 25th percentile in the data set) to a "distant dyad" (about 8,600 miles apart, which is the 75th percentile) as the democracy of the dyad varies. The Distance Hypothesis predicts that this effect should be negative regardless of the level of democracy in the dyad. Figure 8 shows that the effect is indeed negative across the range of observed levels of democracy and the confidence interval never overlaps zero. Again, this pattern holds across the model with and without a product term.
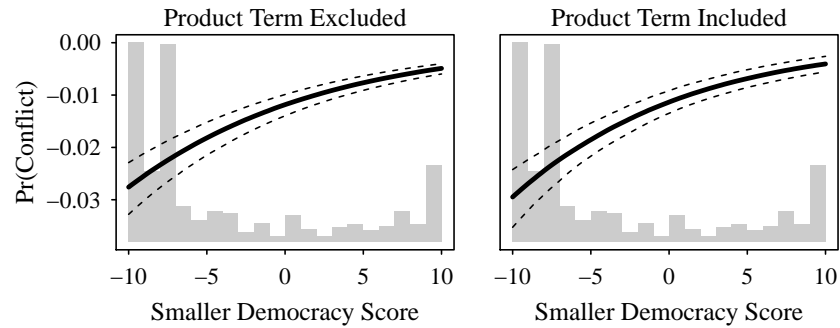
FIGURE 8: This figure shows how the negative effect of distance on the probability of conflict shrinks toward zero as dyads become more democratic. This expected pattern is consistent with compression, but the model without a product term forces this pattern into the results while the model with a product term allows other types of relationships to emerge.

Finally, and most importantly for my purposes, Figure 9 shows the estimated second-differences–the difference in the effects of distance among democratic and non-democratic states. The negative estimates suggest that distance has a smaller effect among democratic states, which supports the Interaction Hypothesis. Notice that, as before, the results are substantively similar across the model with and without a product term. Overall, then it seems that both models provide strong support for all three hypotheses.

Is Parsimony Preferable?

In this situation, applied researchers might be tempted to discard the model with a product term for three reasons.

1. The logistic regression model with no product term captures the relationship (democracy and distance have negative effects and interact due to compression) suggested by the strong theory. Berry, DeMeritt, and Esarey (2010) recommend excluding the product term in this situation.[15]

---

[15]Berry, DeMeritt, and Esarey (2010) write: "[A] researcher may have no reason to believe that the effects of independent variables on the latent variable Y^* are interactive and may base his hypothesis that independent variables interact in influencing $\Pr(Y)$ strictly on an expectation of compression. In this case, there is no need to include a product term in the model."
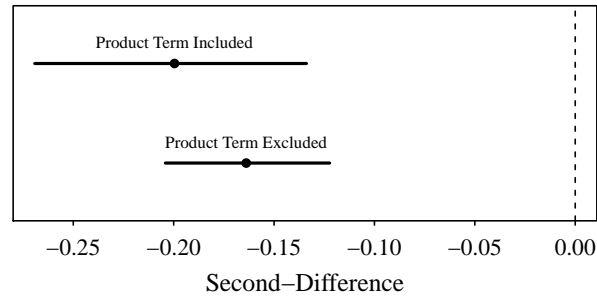
FIGURE 9: This figure shows the estimated second-differences from both the models with and without a product term. While the estimates are similar, the results from the model with a product term are more compelling, because excluding a product term strongly biases the second-differences away from zero (see the left panel of Figure 2), while including removes most of this bias (see the right panel of Figure 2).

2. The AIC suggests the two models fit equally well (AIC = 13,520 for both models) and the BIC points toward the simpler with no product term (BIC = 13,580 for the model without the product term and BIC = 13,589 for the model with the product term).

3. The product term is not statistically significant ($p = 0.23$).

Researchers might be tempted to drop the product term for *any* of these three reasons. In spite of these temptations, however researchers must include the product term XZ when making arguments about X and Z interacting in influencing $\Pr(Y)$.

In the applied example, it is important to include the product term because it "uncouples" the Interaction Hypothesis from the Distance and Democracy Hypotheses. If I do not include a product term in the model, then the large negative effect of democracy and distance guarantee that the two interact in influencing $\Pr(Y)$ (see the left panel of Figure 2 and Equation 1). With not product term in the model, the Distance and Democracy Hypotheses *imply* the Interaction Hypothesis and nothing is learned from examining the second-difference. By including the product term, however, I am able to uncouple the Democracy and Distance Hypotheses from the Interaction Hypotheses (see the right panel of Figure 2 and Equation 2). When the model includes a product term, the Distance and Democracy Hypotheses no longer imply the Interaction

Hypothesis and something is learned from examining the second-difference. While the two estimates of the second-difference presented in Figure 9 are similar, they represent different amounts of evidence for the Interaction Hypothesis. The estimate from the model with no product term is assumed by the structure of the empirical model. The estimate from the model with a product term, on the other hand, represents a meaningful relationship based on the data.

Conclusion

When researchers are interested in how explanatory variables interact in influencing the probability of an event, my study shows that (1) models *without* a product term point toward interaction very often when none exists, (2) models *with* a product term exhibit much less bias toward interaction, and (3) models *with* a product term almost always reduce and usually eliminate the bias found in models *without* a product term. Although logistic regression models require "compression" to keep the probability of an event bounded between zero and one, researchers can limit the impact of this requirement on tests for interaction by simply including a product term, which allows the models to represent a broad range of non-interactive relationships surprisingly well. Therefore, even when theorizing interaction due to compression alone, researchers should include a product term, making their theories more vulnerable to the data.

References

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80(1):123–129.

Berry, Frances S., and William D. Berry. 1991. "Specifying a Model of State Policy Innovation." *American Political Science Review* 85(2):573–579.

Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84(2):395–415.

Berry, William, Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54:248–266.

Berry, William, Jacqueline H. R. DeMeritt, and Justin Esarey. 2014. "Bias and Overconfidence in Parametric Models of Interactive Processes." *American Journal of Political Science* .

Berry, William, Matt Golder, and Daniel Milton. 2012. "Improving Tests of Theories Positing Interaction." *Journal of Politics* .

Bowen, Harry P. 2012. "Testing Moderating Hypotheses in Limited Dependent Variable and Other Nonlinear Models: Secondary Versus Total Interaction." *Journal of Management* 38(3):860–889.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14:63–82.

Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution* 36(2):309–341.

Clark, William Roberts, Michael J. Gilligan, and Matt Golder. 2006. "A Simple Multivariate Test for Assymetric Hypotheses." *Political Analysis* 14(3):311–331.

Diehl, Paul F. 1991. "Geography and War: A Review and Assessment of the Empirical Literature." *International Interactions* 17(1):11–27.

Frant, Howard. 1991. "Specifying a Model of State Policy Innovation." *American Political Science Review* 85(2):571–573.

Friedrich, Robert. 1982. "In Defense of Mulitiplicative Terms in Multiple Regression Equations." *American Journal of Political Science* 26:797–833.

Greene, William. 2010. "Testing Hypotheses About Interaction Terms in Nonlinear Models." *Economics Letters* 107(2):291–296.

Hanmer, Michael, 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(1):263–277.

Huang, Chi, and Todd G. Shields. 2000. "Interpretation of Interaction Effects in Logit and Probit Analyses." *American Politics Research* 28(1):80–95.

Kam, Cindy D., and Robert Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. University of Michigan Press.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44:341–355.

Nagler, Jonathan. 1991. "The Effect of Registration Laws and Education on U.S. Voter Turnout." *American Journal of Political Science* 85:1393–1405.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38(1):230–255.

Norton, Edward C., Hua Wang, and Chunrong Ai. 2004. "Computing Interaction Effects and Standard Errors in Logit and Probit Models." *The Stata Journal* 4(2):154–167.

Oneal, John R., and Bruce Russet. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W. W. Norton.

Ray, James Lee. 1998. "Does Democracy Cause Peace?" *Annual Review of Political Science* 1(1):27–46.

Tomz, Michael R., and Jessica L. P. Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107(4):849–865.

Train, Kenneth E. 2009. *Discrete Choice Models with Simulation*. New York: Cambridge University Press.

Tsai, Tsung-han, and Jeff Gill. 2013. "Interaction in General Linear Models: Theoretical Issues and an Application to Personal Vote Earning Attributes." *Social Sciences* 2(2):91–113.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* Yale University Press.

Online Appendix

A   Statistical Power

Including a product term does reduce statistical power in tests of interaction using second-differences. If the researcher is correct and there is interaction due to compression (i.e., the logistic regression model is correct), then including a product term reduces the probability that the researcher will find a statistically significant interaction effect (i.e., the confidence interval for the second-difference does not contain zero). However, as I explain in the main text, statistical arguments proceed by fixing the size of the test at a suitably small value (usually 0.05) and then maximizing power. Maximizing power by increasing the size of one's test over-represents the evidence for one's hypotheses because readers expect the researcher to only accept their research hypothesis when the evidence overwhelmingly supports it. If power is maximized at the expense of size, then researchers accept their research hypothesis unless their is overwhelming evidence against it. Thus, I focus on the size of tests for interaction in the main text. Nonetheless, attentive researchers are conscious of the power of their tests, so I conduct a brief analysis here.

To estimate the power of tests for interaction with and without a product term, I simulate from the model (i.e., data-generating process) $\Pr(Y) = \text{logit}^{-1}(\beta_x X - \beta_z Z)$, where X is a continuous variable ranging from zero to one and Z dichotomous variable that equals zero or one. I imagine that an analyst correctly suspects that X has a positive effect on $\Pr(Y)$, Z has a negative effect on $\Pr(Y)$, and that X and Z interact in influencing $\Pr(Y)$ due to compression. That is, because the event is relatively common ($\Pr(Y) > 0.5$) and Z pushes $\Pr(Y)$ even closer to zero, there is less room for X to affect $\Pr(Y)$. Therefore, the analyst expects that X should have a smaller effect when $Z = 1$. All three hypothesis are correct because logistic regression is the correct functional form. For varying values of β_x, β_z, and the sample size, I simulate 1,000 data sets and compute the proportion that have a statistically significant second-difference. Because β_x and β_z vary, the magnitude of this second-difference varies from trivially small (-0.02) to quite large (-0.21).

Because X is approximately uniformly distributed from zero to one and Z is dichotomous with half of the observations equal to one and the other half equal to zero, I test for interaction in each simulated data set by estimating a logistic regression model with and without a product term, calculating the second-difference as X changes from 0.25 to 0.75 and Z changes from zero to one.

Notice that for trivial amounts of interaction (i.e., $\Delta\Delta = 0.07 - 0.09 = -0.02$), the model with a product term dramatically reduces the power of the test, especially for sample sizes around 1,000. However, for larger sample sizes and larger interaction effects, adding a product term to the model decreases the power of the test only slightly, while fixing the size of the test at a more typically value near 0.05 (see Figure 5).
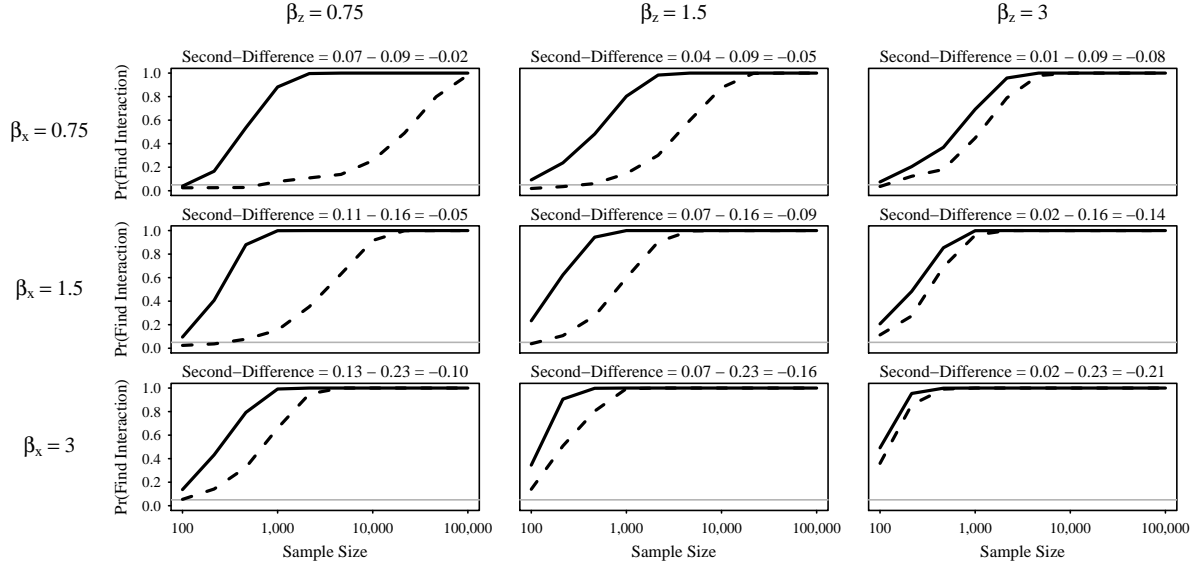
FIGURE 10: A figure showing the statistical power of models including a product term (dashed line) and excluding a product term (solid line) for detecting interaction. Notice that adding a product term decreases the power appreciably only for trivially small interaction effects (i.e., $\Delta\Delta = 0.07 - 0.09 = -0.02$).

B The Impact of Including an Unnecessary Product Term on Estimates of Effects of Other Variables

One potential concern is that if the correct model is given by $\Pr(Y) = \text{logit}^{-1}(\beta_{cons} + \beta_X X + \beta_Z Z + \beta_W W)$, then including an unnecessary interaction term XZ will lead to biased estimates of the effect of other variables, such as W. Simulations show that this is not the case. No bias exists in estimates of the coefficient β_W or the first difference $\Pr(Y|W = w_{hi}) - \Pr(Y|W = w_{lo})$. I only summarize a brief simulation here, but these result are identical for a range of other simulations I performed.

1. Set $N = 1000$. Setting the sample size to other values has no impact on the results.

2. Set $\beta_{cons} = 0$ and $\beta_X = \beta_Z = 1$. Setting these coefficients to other values has no impact on the results.

3. Draw X, Z, and W from a Uniform$(0, 1)$ distribution to create the covariates. Using other distributions does not impact the results. Introducing correlations between the covariates does not impact the results.

4. As β_W ranges from -1 to 1, do the following:

a. Compute $\pi = \Pr(Y)$.

b. To get an estimate of the expected value of the estimated effect of W, do the following 100,000 times.

    i. Simulate $Y \sim \text{Bernoulli}(\pi)$.

    ii. Estimate a logistic regression model with and without a product term.

    iii. Store the coefficient for W.

    iv. Calculate and store the first difference $\Pr(Y|W = 0.75) - \Pr(Y|W = 0.25)$.

c. Calculate the mean of the estimated coefficients.

d. Calculate the mean of the estimated first-differences

As Figure 11 shows, the estimates of β_W are not affected by the inclusions of an unnecessary product term. Particularly, there is no bias in either estimate. The mean estimates from the model with and without a product term entirely overlap with the true values. Figure **??** shows that the same hold for the first difference.
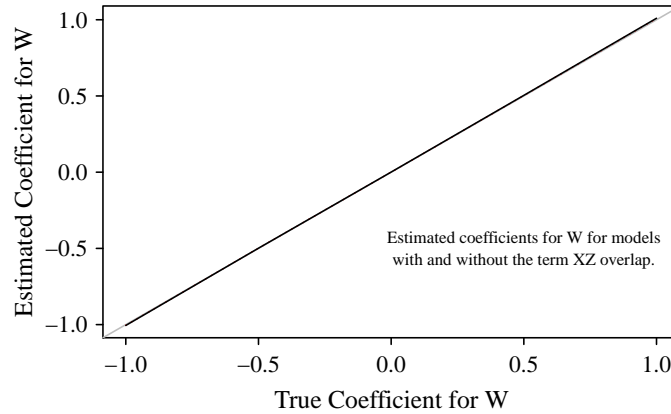

FIGURE 11: A figure showing the expected coefficient $E(\hat{\beta}_W)$ as the true coefficient β_W changes for the models with and without a product term. Note that all three overlap exactly, which means that including a product term does not bias the coefficients of other explanatory variables.
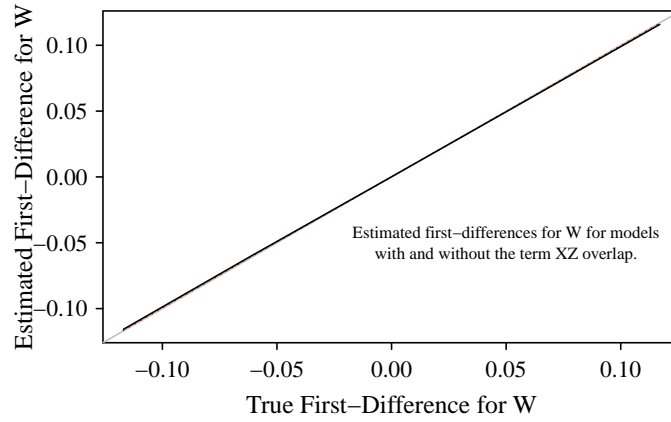
FIGURE 12: A figure showing the expected estimate of the first-difference as the true first-difference changes for the models with and without a product term. Note that all three overlap exactly, which means that including a product term does not bias the first-differences of other explanatory variables.

C Replication of the Fixed Process Simulation for Rare Events

Particularly given the application of my ideas to the probability of conflict, one might wonder whether including a product term removes the bias as effectively when the event of interest is relatively rare. As a simple test case, I use the process $\Pr(Y) = 0.04 + 0.04X - 0.02Z$, shown in Figure 13 as a robustness check to the fixed process presented in Figure 1. Notice, as before, that the model with a product term fits this non-interactive process surprisingly well and the model with no product term forces interaction into the estimates.
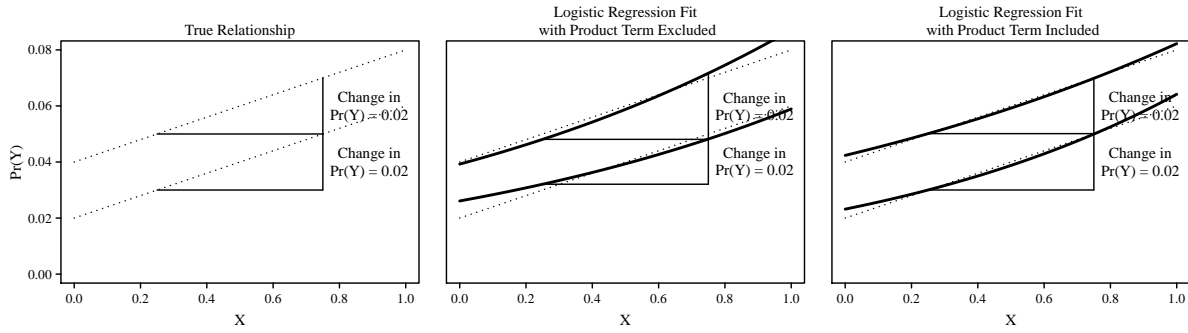
FIGURE 13: This figure shows the true relationship in the "rare events" simulation study (dotted lines) as well as how logistic regression models with and without a product term fit the process (solid, bold lines). Notice that X and Z have small effects on $\Pr(Y)$ and $\Pr(Y)$ is always less than 0.1. In this case, excluding a product term forces the model to point toward interaction, despite the fact that the relationship is non-interactive. Including a product term allows the model to fit this non-interactive process surprisingly well. I use this process as a simple robustness check for the results presented in Figure 3.

Figure 14 presents the results from the fixed "rare events" process. For this rare events process, the product term does a much better job at removing the bias toward interaction than the process examined in the main text. Except for extremely large sample sizes near 100,000, the adding a product term removes nearly all of the bias and removes much of the bias even for the sample sizes near 100,000. The results suggest that the results presented in the main text represent a "worst case" scenario (that is what I had in mind when choosing the fixed process) and that the suggestion to include a product term applies to extremely rare (and, by symmetry, extremely common) as well as to events with probabilities closer to one-half.
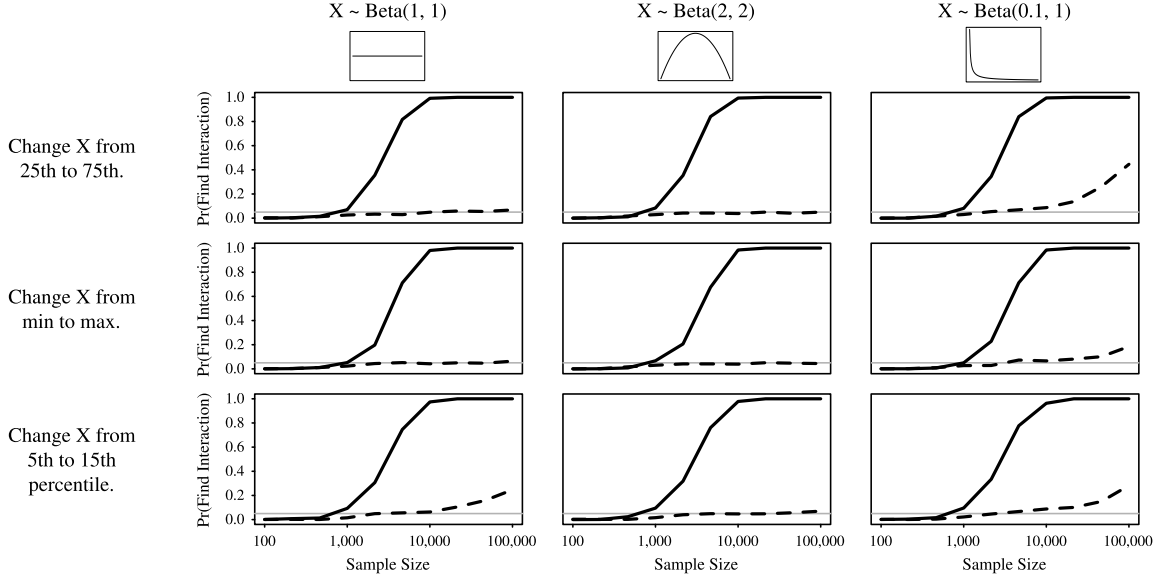
FIGURE 14: This figure shows the probability of concluding interaction when the true relationship is given by $\Pr(Y) = 0.04 + 0.04X - 0.02Z$ (i.e., there is no interaction) using logistic regression models with and without a product term as the sample size, the distribution of X, and the change in X considered in computing the second-difference varies. Notice that the model including a product term performs remarkably well, removing almost all of the bias except for extremely large samples.

D   Generate a Random Analysis

1. Choose a random relationship that is non-linear and non-interactive.

   a. Draw A and B from a uniform distribution between 0.3 and 0.7.

   b. Draw C from a uniform distribution between A and B if $A < B$ or between B and A if $B < A$.

   c. Draw D from a uniform distribution between $A - 0.3$ and $A + 0.3$.

   d. Draw E from a uniform distribution between 0 and D if $D > 0$ or D and 0 if $D < 0$.

   e. The true relationship is then given by $Pr(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X^2 + \beta_4 Z^2$, where
   - $\beta_1 = A$
   - $\beta_2 = -3A - B + 4C$
   - $\beta_3 = 4E - D$
   - $\beta_4 = 2A + 2B - 4C$
   - $\beta_5 = 2D - 4E$

   f. Check the resulting relationship is monotonic as X and Z range from zero to one. If not, repeat until a monotonic relationship is obtained.

35

Figure 15 illustrates how this procedure works graphically and Figure 16 shows several randomly simulated relationships using this procedure.

2. Choose a sample size by drawing from a Cauchy distribution, multiplying by 1,000, and rounding to the nearest whole number. The Cauchy distribution has very heavy tails so that most samples fall below 2,000, but very large samples (up to 100,000) occur as well. Table 2 provides the quantiles from this distribution.

3. Choose the distribution of X and Z. Perform the following to generate X and Z.

   a. Choose whether X is binary or continuous. Set as binary with probability 0.5.

   b. If X is continuous, choose the parameters of the Beta distribution from which it will be drawn from a uniform distribution from 0.7 to 3. Figure 17

   c. If X is binary, choose the proportion of ones from a uniform distribution from 0.1 to 0.9.

   d. Repeat for Z.

4. Choose the changes and X and Z used to test for interaction.

   a. If X is binary, set $\langle x_{lo}, x_{hi} \rangle$ to $\langle 0, 1 \rangle$.

   b. If X is continuous, then draw and sort two values from a uniform distribution from zero to one. Set these sorted draws as $\langle x_{lo}, x_{hi} \rangle$.

   c. Repeat for Z.

Percentile	Sample Size
0%	500
10%	646
20%	815
30%	1,019
40%	1,276
50%	1,614
60%	2,105
70%	2,896
80%	4,427
90%	8,896
100%	100,000

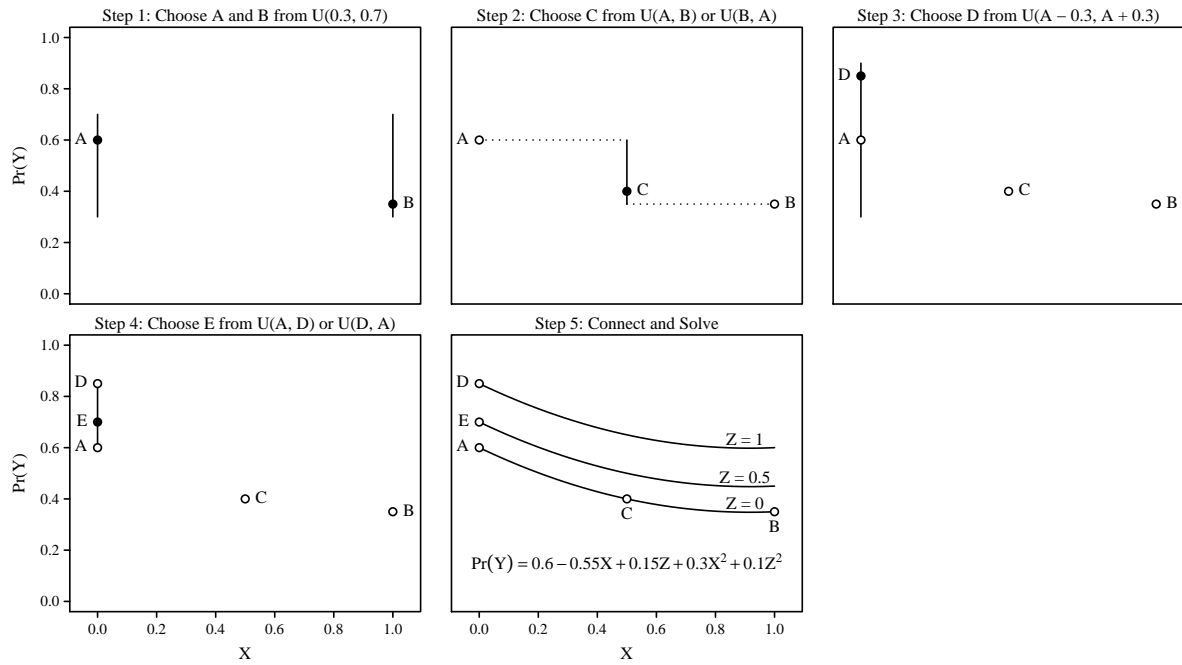TABLE 2: The distribution of sample sizes I used to simulate analyses.

FIGURE 15: This figure graphically illustrates how I generate random relationships that are non-interactive and non-linear using the procedure describe in Step 1.
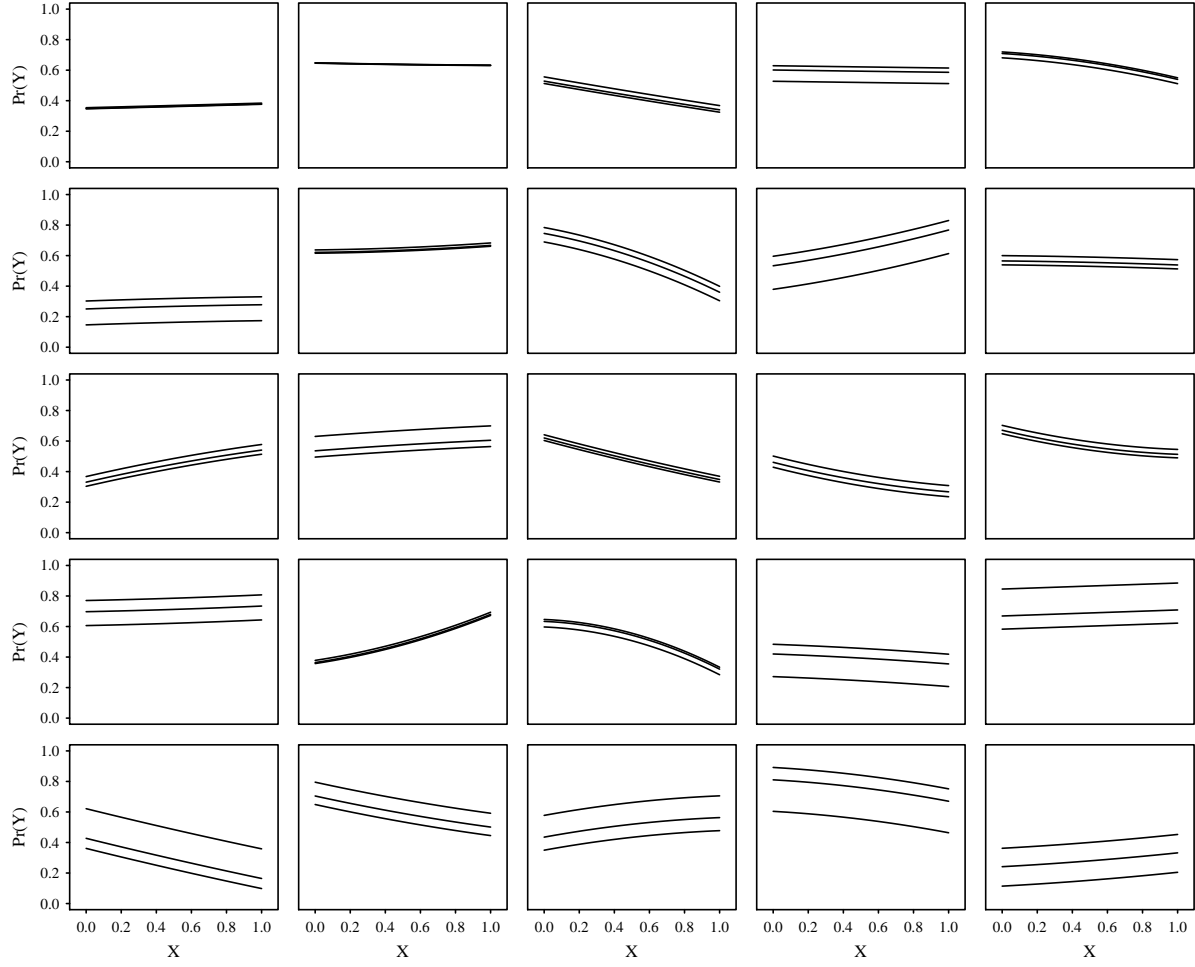
FIGURE 16: This figure shows several relationships randomly generated using the steps described in Figure 15 and Step 1 of the algorithm above. The three lines represent the relationship between X and $\Pr(Y)$ when Z equals zero, one-half, and one, respectively. Due to the assumption of monotonicity described in Step 1(f), the line for which $Z = 0.5$ must fall between the other two.
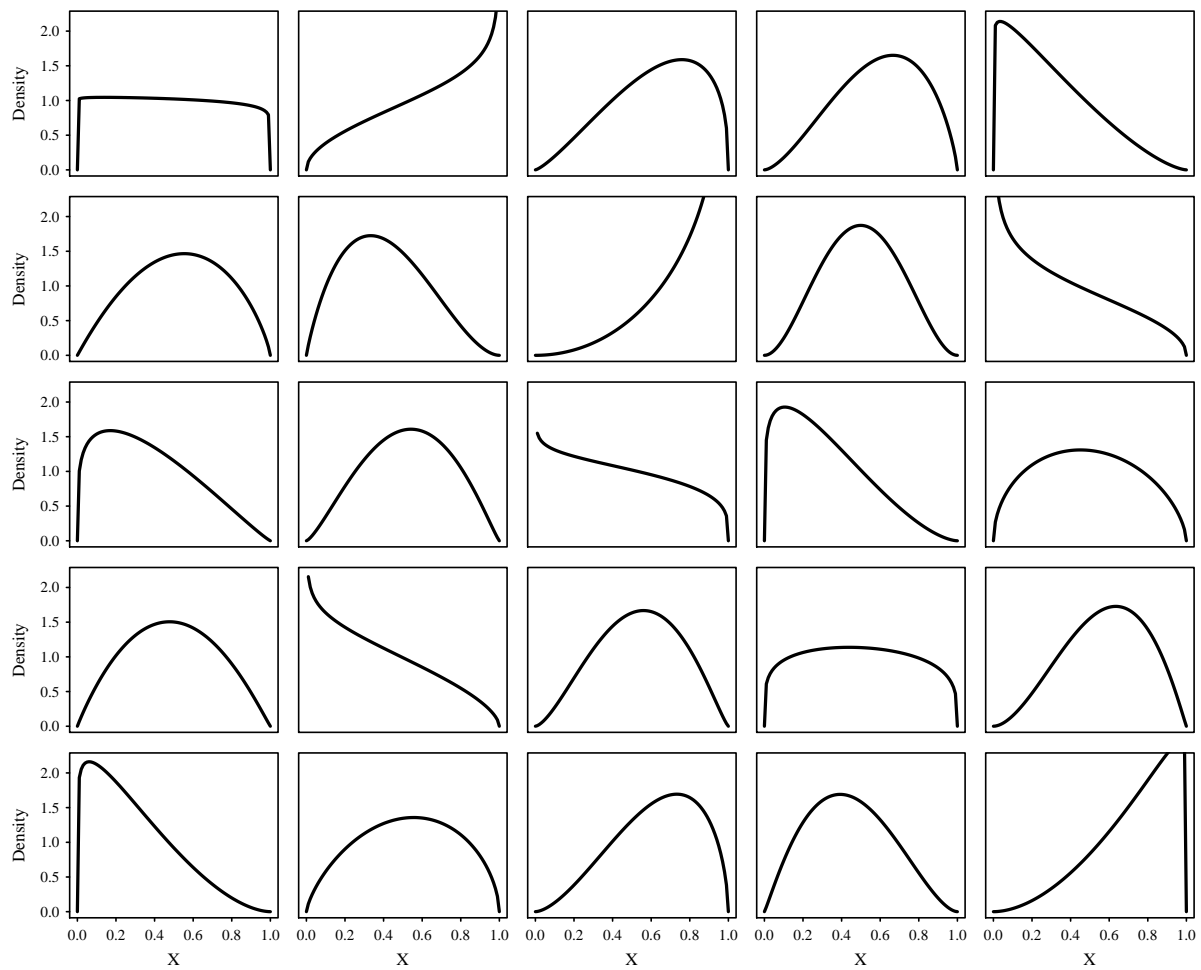
FIGURE 17: This figure shows several possible continuous distributions of X (or Z) randomly generated using Step 3(b).

Model Estimates

As Brambor, Clark, and Golder (2006) and Berry, DeMeritt, and Esarey (2010) explain, examining tables of coefficients is not always helpful for evaluating interactive hypotheses with logistic regression models. In the main text I present plots of predicted probabilities (Figure 6), first-differences (Figures 7 and 8), and second-differences (Figure 9). Table 3 presents the coefficients and standard errors for reference.

Relying on the model with no product term, we can immediately evaluate the evidence for the Democracy and Distance Hypotheses by looking at the coefficients for Lower Democracy and log(Distance). We can see that both of these estimates are negative (as expected) and statistically significant. Further, based on this model, we can already conclude that the negative effect of log(Distance) is smaller in more democratic dyads since (1) both coefficients are negative and (2) conflict is relatively rare (i.e., the probability of conflict falls well below 0.5).

	No Product Term	Product Term
Lower Democracy	−0.091***	−0.121***
	(0.005)	(0.025)
Log(Distance)	−0.306***	−0.285***
	(0.026)	(0.031)
Lower Democracy × Log(Distance)		0.004
		(0.004)
Allies	−0.694***	−0.697***
	(0.063)	(0.063)
Power Ratio	−0.264***	−0.262***
	(0.019)	(0.019)
Noncontiguity	−0.980***	−0.982***
	(0.073)	(0.073)
Only Minor Powers	−0.467***	−0.475***
	(0.070)	(0.071)
Constant	−1.222***	−1.370***
	(0.202)	(0.237)
AIC	13,520	13,520
BIC	13,580	13,589
Num. obs.	39,996	39,996

$^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$

TABLE 3: Logistic regression models of the probability of conflict in each dyad-year. The model in the left column excludes the product term, as suggested by Berry, DeMeritt, and Esarey (2010). The model in the right column includes a product term. Notice that although the product term is not statistically significant and the BIC for this model is higher (suggesting a worse fit), including this product term allows a more compelling test of the Interaction Hypothesis.