Improving Tests of Theories Positing Interaction*

WILLIAM D. BERRY†
Florida State University

MATT GOLDER‡
Pennsylvania State University

DANIEL MILTON§
Florida State University

ABSTRACT

It is well established that all interactions are symmetric: when the effect of X on Y is conditional on the value of Z, the effect of Z must be conditional on the value of X. Yet the typical practice when testing an interactive theory is to (i) view one variable, Z, as the conditioning variable, (ii) offer a hypothesis about how the marginal effect of the other variable, X, is conditional on the value of Z, and (iii) construct a marginal effect plot for X to test the theory. We show that the failure to make additional predictions about how the effect of Z varies with the value of X, and to evaluate them with a second marginal effect plot, means that scholars often ignore evidence that can be extremely valuable for testing their theory. As a result, they either understate or, more worryingly, overstate the support for their theories.

*Note: We would like to thank Thomas Brambor, William Roberts Clark, Justin Esaray, Robert Franzese, Jeff Gill, Sona Nadenichek Golder, Chris Reenock, David Siegel, members of the Political Institutions Working Group at Florida State University, and three anonymous reviewers for helpful comments on earlier versions of this paper. We also thank Scott Kastner and Mikhail Alexseev for their cooperation and for providing their replication datasets. The data and all computer code necessary to replicate our results will be made publicly available at the authors' homepage upon publication. Stata 10 was used for all statistical analyses.

†Marian D. Irish Professor and Syde P. Deeb Eminent Scholar, Florida State University, Department of Political Science, 531 Bellamy Building, Tallahassee, FL 32306-2230 (wberry@fsu.edu). Tel: 850-644-7321. Fax: 850-644-1367.

‡Associate Professor, Pennsylvania State University, Department of Political Science, 306 Pond Lab, University Park, PA 16802 (mgolder@psu.edu). Tel: 814-867-4323. Fax: 814-863-8979.

§Graduate Student, Florida State University, Department of Political Science, 531 Bellamy Building, Tallahassee, FL 32306-2230 (djm07g@fsu.edu)

Since many political theories assert that the effects of variables vary depending on the social, political, economic, or strategic context, models specifying *interaction* among variables are ubiquitous across all subfields of political science.[1] A consequence is that conditional hypotheses such as "X has a positive effect on Y that gets stronger as Z increases" are extremely common. It is well established that interactive statistical models containing multiplicative terms, such as XZ, are appropriate for evaluating such conditional hypotheses (Wright 1976, Friedrich 1982, Aiken & West 1991, Clark, Gilligan & Golder 2006).[2] A number of authors in recent years have offered valuable advice on how to improve research testing theories positing interaction by properly specifying the expected conditionality in a statistical model, and effectively presenting and interpreting the results (Braumoeller 2004, Brambor, Clark & Golder 2006, Kam & Franzese 2007). However, even researchers following this advice often ignore valuable empirical evidence that can be easily derived from their estimated model, and as a result, fail to assess all of the predictions generated by their theory. The result is that many researchers either understate, or, more worryingly, overstate the empirical support for their conditional theories.

The inadequacy of many empirical tests of conditional theories can be traced to the tendency of scholars positing interaction between two variables to conceive of these variables as having different roles within the theory. One variable, Z, is typically viewed as the "conditioning variable," the role of which is to modify the impact of the other variable, X, on the dependent variable, Y. Certainly, when X and Z interact, it is reasonable to conceive of Z as conditioning the effect of X on Y. However, it makes little sense to view X and Z as having fundamentally different theoretical roles by designating one of the variables as a "conditioning variable" and the other as not. This is because, logically, *all* interactions are symmetric (Brambor, Clark & Golder 2006, Kam & Franzese 2007). In other words, if Z modifies the effect of X on Y, then X *must* modify the effect of Z on Y. Some might view conceiving of one variable as the "conditioning variable" and failing to acknowledge the symmetry of interaction as merely a semantic problem. We demonstrate, however, that this practice can have pernicious consequences, leading researchers to ignore empirical evidence relevant for testing their theory.

In a much-cited 2006 article, Brambor, Clark, and Golder [hereafter BCG] demonstrate that political

[1] In a systematic examination of three leading journals (*American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*) from 1996-2001, Kam and Franzese (2007, 7-8) find that fully 24% of articles employing "statistical methods" tested theories predicting interaction; this amounted to about one-eighth of all the articles published in this time period.

[2] We treat the terms "theory positing interaction," "interactive theory," and "conditional theory" as synonymous.

scientists can greatly increase their ability to impart substantively meaningful information from interactive models by using parameter estimates to construct a *marginal effect plot*, i.e., a graph that shows how the marginal effect of one independent variable varies with the value of another variable. Scholars have responded in large numbers to BCG's call to incorporate marginal effect plots into their analyses. Indeed, within three years of the appearance of BCG's article, at least 44 published papers presented such plots.[3] This has dramatically improved the interpretation of statistical results from interactive models in the literature. Ironically, however, BCG's article may have inadvertently encouraged its readers to make the mistake of viewing one variable as the "conditioning variable." Although BCG (2006, note 9) correctly observe that interactive models are symmetric and that the marginal effect of each independent variable is a "meaningful" quantity of interest, they go on to imply that analysts might reasonably establish one of these quantities as the focus of theoretical interest and produce only one marginal effect plot. Indeed, of the 44 papers we identified that present marginal effect plots to evaluate a theory positing interaction, 39 (89%) present only a single plot, showing how the estimated effect of one variable varies with the other.

We accept as a fundamental principle that scholars estimating a statistical model should use the estimation results to assess as many of the theory's implications as possible. We show that for those testing theories positing interaction between two independent variables, this often means deriving and testing predictions about how the marginal effect of each independent variable varies with the value of the other *not all of which can be evaluated by inspecting a single marginal effect plot*. It is important to note that we are not suggesting that researchers testing a hypothesis about how the marginal effect of X varies with Z should "manufacture" a second hypothesis about how the marginal effect of Z varies with X when their theory generates no predictions beyond those already incorporated in the first hypothesis. We are simply observing that many conditional theories proposed by political scientists generate more predictions than can be tested with a single marginal effect plot, and that in this situation, when the researcher limits consideration to a single plot, she subjects her theory to a weaker test than is possible given the data available.

In the next section, we consider the implications of the inherent symmetry of interactive models for theory testing in more detail. In particular, we demonstrate why it can be dangerous for a researcher with a conditional theory to limit consideration to predictions that can be evaluated with a single marginal effect

[3] As of January 2009, 44 published articles in the ISI Web of Knowledge database cite BCG's (2006) article and present at least one marginal effect plot.

plot. The basic insight is that any observed relationship between Z and the marginal effect of X is always consistent with a wide variety of ways in which the marginal effect of Z varies with X, some of which may be inconsistent with the underlying conditional theory. This means that proposing a hypothesis about how the effect of X varies with Z and assessing it by examining just a marginal effect plot for X often constitutes a weak test of the conditional theory underlying the hypothesis. Supplementing this hypothesis with a second one about how the effect of Z varies with X that can be evaluated by inspecting a marginal effect plot for Z can dramatically narrow the range of relationships that are consistent with one's underlying theory, thereby increasing the power of the empirical test.

In the following section, we provide practical advice on deriving and testing hypotheses from conditional theories. In particular, we discuss issues that arise when evaluating empirical evidence in favor of, or against, conditional theories by examining several prototypical sets of results one might get when estimating an interactive model. Many of these issues have not been adequately addressed in the existing literature, leaving some readers uncertain as to how to evaluate the level of empirical support for conditional theories. Next, we illustrate our central points by replicating two of the numerous published studies that seek to test a conditional theory but that present a marginal effect plot for only one of the two variables predicted to interact. In one replication, constructing a second marginal effect plot reveals additional support for the researcher's theory. In the other, a second plot reveals evidence contrary to the analyst's theory. Throughout the paper, we offer advice on how to maximize the information portrayed in marginal effect plots, and before concluding, we summarize our recommendations.

Implications of the Symmetry of Interaction for Theory Testing

Suppose we have a conditional theory in which X and Z interact in influencing some continuous dependent variable, Y, such that the effects of X and Z can be captured with the following linear-interactive model:[4]

$$Y = \beta_0 + \beta_X X + \beta_Z Z + \beta_{XZ} XZ + \epsilon. \tag{1}$$

This model – involving a single product, or multiplicative, term – is the most common specification of

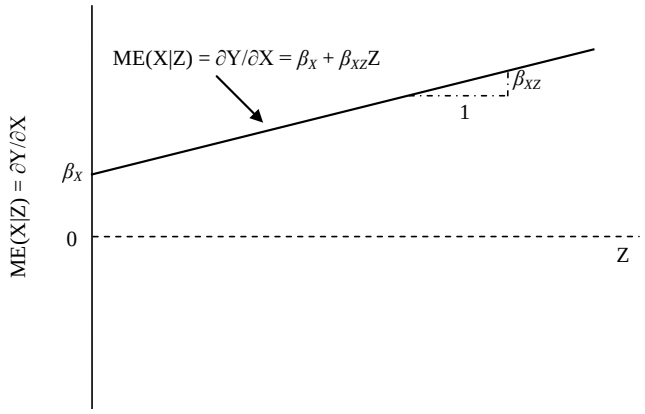
[4]For simplicity, we assume that there are no other covariates in the model. However, all claims in this article hold with any number of additional variables as long as none of them interacts with either X or Z. Although we focus on models with continuous dependent variables, our advice is equally applicable to models with limited dependent variables such as logit and probit.

interaction in political science. In this model, the marginal effect of *X*, $\partial Y/\partial X$, is given by:

$$\frac{\partial Y}{\partial X} = \beta_X + \beta_{XZ} Z. \tag{2}$$

As Eq. (2) clearly indicates, unless the coefficient for the product term, β_{XZ}, is zero, the marginal effect of *X* is conditional on the value of *Z*.[5] To emphasize this conditionality in what follows, we denote the marginal effect of *X* as ME(*X*|*Z*). In turn, we let ME(*X*|*Z* = z) denote the marginal effect of *X* on *Y* when *Z* equals the specific value z. Figure 1 depicts the relationship between ME(*X*|*Z*) and *Z* when β_X and β_{XZ} are both positive. As should be clear, the marginal effect of *X* on *Y* when $Z = 0$ is β_X, the coefficient for *X*. Because β_{XZ} is the constant slope of Eq. (2), it should also be clear that the marginal effect of *X* changes by β_{XZ} for every unit increase in *Z*.

Figure 1: A Plot of the Marginal Effect of *X* on *Y* against *Z* when β_X and β_{XZ} in Eq. (1) are Positive



Note that the interactive model specified in Eq. (1) is *symmetric* in *X* and *Z*. In other words, the fact that the marginal effect of *X* on *Y* is conditional on *Z* logically guarantees that the marginal effect of *Z* on *Y* must be conditional on *X*. Indeed, the marginal effect of *Z* is given by:

$$ME(Z|X) = \frac{\partial Y}{\partial Z} = \beta_Z + \beta_{XZ} X. \tag{3}$$

[5]Note that the expression for $\partial Y/\partial X$ in Eq. (2) implies that the marginal effect of *X* on *Y* is conditional on the value of *Z* but not on the value of *X*. This property stems from the linear functional form for the interaction specified in Eq. (1). In interactive models with a nonlinear functional form, in contrast, the marginal effect of *X* on *Y* necessarily varies with both the value of *X* and the value of *Z* (Brambor, Clark & Golder 2006, 77).

This implies that the marginal effect of Z is β_Z when X is zero and changes by β_{XZ} for every unit increase in X. Thus, it is evident that the coefficient on the product term, β_{XZ}, indicates both the slope of the relationship between ME(X|Z) and Z *and* the slope of the relationship between ME(Z|X) and X. As such, we must recognize that Z conditions the effect of X on Y, and X conditions the effect of Z on Y. It is this inherent symmetry that makes it misleading for scholars to designate X or Z as *the* conditioning variable and the other variable as the one being conditioned.[6] We recognize that in some settings it can be very tempting to conceive of one variable as *the* conditioning variable. Such a conceptualization can be especially attractive when one variable, X, is continuous and the other, Z, is dichotomous. This is because it is often natural to think in terms of the effect of X on Y being different in one context ($Z = 0$) than in another ($Z = 1$). However, the fact remains that the effect of the binary variable Z also varies with X.

Although the inherent symmetry of interactions is well-documented (Brambor, Clark & Golder 2006, Kam & Franzese 2007), the implications of such symmetry for theory testing have been largely overlooked. Recall that the product term coefficient, β_{XZ}, in Eq. (1) indicates both the slope of the relationship between ME(X|Z) and Z and the slope of the relationship between ME(Z|X) and X. This implies that if a researcher with a conditional theory presents a clearly-stated proposition about how the marginal effect of X on Y varies with Z, then she is also implicitly introducing a hypothesis about how the marginal effect of Z on Y varies with X. Thus, on the surface, it would seem unnecessary – even redundant – for the researcher to explicitly state an additional hypothesis about ME(Z|X). However, this intuition is incorrect.

Even a full characterization of how the marginal effect of X on Y varies with Z, as shown in Eq. (2), establishes only the *slope* of the relationship between X and the marginal effect of Z in Eq. (3). Knowing that the slope of the relationship between X and ME(Z|X) is β_{XZ} provides absolutely no information about the value of the intercept, β_Z, in Eq. (3), and, hence, no information about whether the effect of Z is positive or negative at any value of X. This is critically important because different values for this intercept imply quite different ways in which the marginal effect of Z is conditional on X. It may be the case that only some of these ways are consistent with the researcher's underlying conditional theory.

To illustrate, suppose that one has a conditional theory in which X and Z interact to influence Y. In particular, the theory predicts that the marginal effect of X is always positive and that the magnitude of this

[6] We should note that the inherent symmetry of interactions is not the result of the particular linear-interactive specification that we use in Eq. (1); *all* interactive specifications are symmetric (Kam & Franzese 2007, 16).

positive effect increases with *Z*. In other words, both β_X and β_{XZ} in Eq. (1) are expected to be positive. The marginal effect plot in Figure 1 is consistent with these theoretical claims. But what exactly does the fact that β_X and β_{XZ} are positive tell us about the marginal effect of *Z* on *Y*? All we can infer from this information is that the plot of ME(*Z*|*X*) will have the same positive slope as the plot of ME(*X*|*Z*). However, a wide variety of conditional relationships among *X*, *Z*, and *Y* are still possible even after this slope is established.

To see this, suppose that the plot of ME(*X*|*Z*) in Figure 1 has an intercept, β_X, of 0.10 and a slope, β_{XZ}, of 0.004. If we assume arbitrarily that the values of both *X* and *Z* range from 0 to 100 in the population of interest, then this plot implies a conditional relationship in which the marginal effect of *X* is 0.10 when *Z* is at its lowest value and $0.10 + (100)(0.004) = 0.50$ when *Z* is at its highest value. In Figure 2, we depict three quite different conditional relationships among *X*, *Z*, and *Y* that are *all* consistent with this marginal effect plot for *X* where $\beta_X = 0.10$ and $\beta_{XZ} = 0.004$. On the left of Figure 2 are three-dimensional plots of *Y* against *X* and *Z*. These plots permit one to visualize how the two independent variables jointly influence *Y*. To the right of each three-dimensional plot is the associated plot of ME(*Z*|*X*) against *X*. A key feature to note about these marginal effect plots is that the value of the intercept, β_Z, is different in each of them.

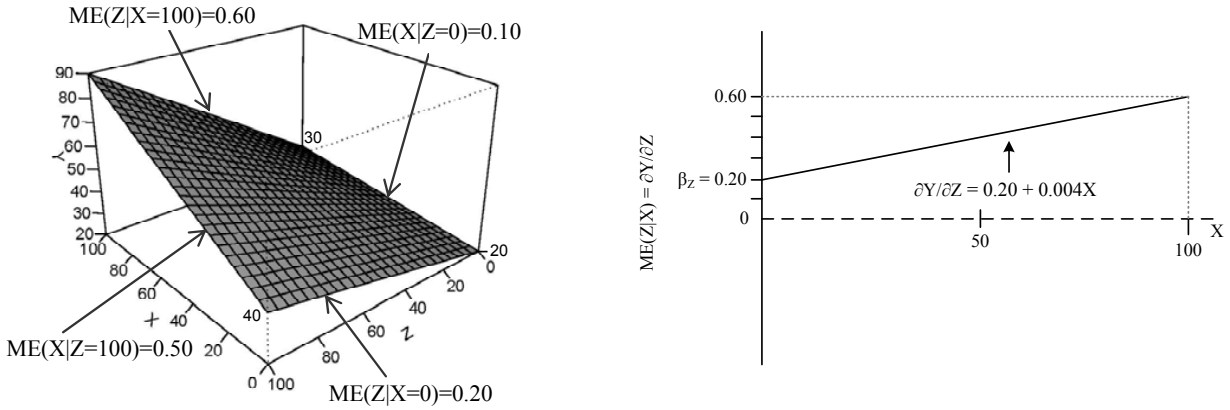
In Figure 2a, β_Z is 0.20, indicating that the marginal effect of *Z* is 0.20 when $X = 0$. The fact that β_{XZ} is also positive means that the marginal effect of *Z* on *Y* is always positive but that this positive effect strengthens as *X* increases, reaching 0.60 when *X* achieves its maximum value of 100. This is reflected in the 3D plot by the slope of *Y* against *Z* being positive both in the left rear vertical plane (i.e., when $X = 100$) and the right front vertical plane (i.e., when $X = 0$), but the slope being more steeply positive in the rear.

In Figure 2b, the intercept, β_Z, is sufficiently negative (-0.60) that the marginal effect of *Z* remains negative at all values of *X* despite the positive value for β_{XZ}. In this scenario, the negative effect of *Z* declines in strength with increases in *X*, reaching -0.20 when *X* obtains its maximum value. This is mirrored in the corresponding 3D plot by the slope of *Y* against *Z* being negative both in the left rear plane (i.e., when $X = 100$) and the right front plane (i.e., when $X = 0$), but the slope being more steeply negative in the front.

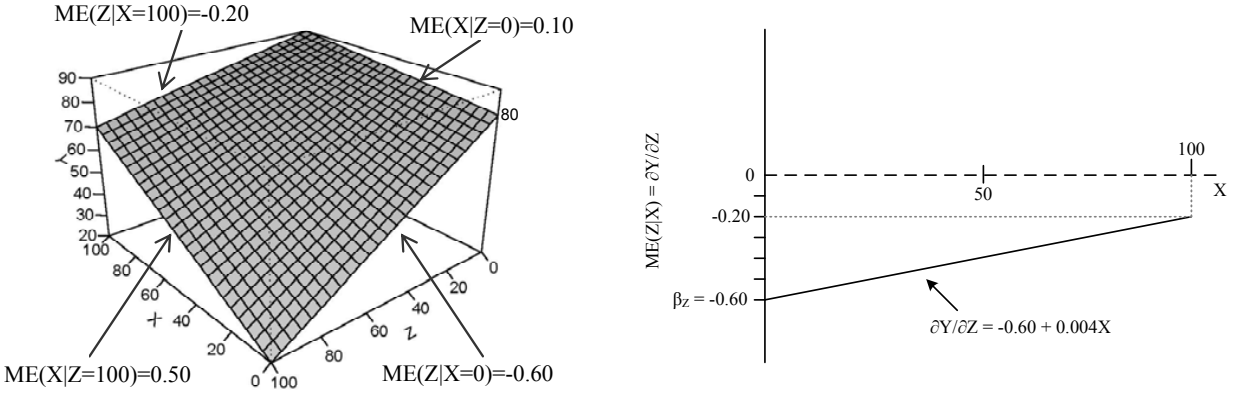
Figure 2c is similar to Figure 2b in that the intercept, β_Z, is negative (-0.20). However, its negative value is sufficiently small in magnitude that the marginal effect of *Z* eventually becomes positive once *X* is large enough. In this scenario, the marginal effect of Z is -0.20 when $X = 0$. *Z*'s negative effect decreases in magnitude as *X* increases until ME(*Z*|*X*) reaches zero when $X = 50$. As *X* increases past 50, the marginal

Figure 2: Three Conditional Relationships Among *X*, *Z*, and *Y* Consistent with the Plot of ME(*X*|*Z*) in Figure 1 (Assuming $\beta_X = 0.10$ and $\beta_{XZ} = 0.004$)

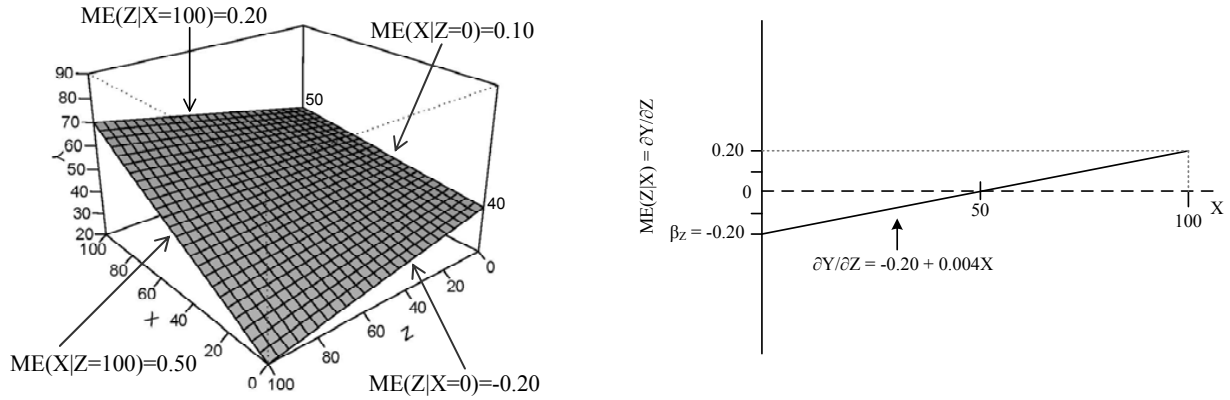
(a)



(b)



(c)



effect of Z becomes positive and grows in strength, reaching 0.20 when $X = 100$. In the associated 3D plot, note that the slope of Y against Z is negative in the right front plane (i.e., when $X = 0$) but positive in the left rear plane (i.e., when $X = 100$).

Figure 2 illustrates quite dramatically how a single marginal effect plot for X can be consistent with very different conditional relationships among X, Z, and Y. It is difficult to imagine someone with a theory predicting that the conditional relationship among X, Z, and Y should be like the one depicted in Figure 2a claiming empirical support for the theory if the *estimated* relationship actually looks like that shown in either Figure 2b or Figure 2c. The plots shown in Figure 2b and Figure 2c depict fundamentally different processes by which Y is jointly determined by X and Z. For example, in Figure 2a, Y is maximized when X and Z are both at their maximum, and Y is minimized when X and Z are both at their minimum. In Figure 2c, Y is also greatest when X and Z are both at their maximum, but Y is smallest when X is minimized while Z is maximized. In Figure 2b, Y is largest when X is maximized and Z is minimized, and Y is smallest when Z is at its maximum and X is at its minimum.

Yet if one limited the empirical evidence examined to an estimated plot of ME(X|Z) showing a positive intercept and a positive slope, as in Figure 1, one might claim support for the conditional theory, ignorant of the inconsistent evidence that would be apparent from an inspection of a plot of ME(Z|X). Thus, even when there is strong empirical support for a hypothesis about how the marginal effect of X on Y varies with Z based on an estimated plot of ME(X|Z), a failure to use one's conditional theory to derive an additional hypothesis about how the marginal effect of Z varies with X (beyond a prediction about the value of β_{XZ}) and inspect a marginal effect plot for Z may mask either (i) additional evidence in support of the theory, or more worryingly, (ii) evidence inconsistent with the theory.

It is important to recognize that once one constructs a theory positing interaction between X and Z in influencing Y specific enough to establish the signs of the intercept and slope of a plot for ME(X|Z), one need not demand a great deal more of the theory to generate additional predictions about ME(Z|X) that would permit a stronger test of the theory. Indeed, if the theory were capable of predicting the sign of the marginal effect of Z at *any* single value of X, at least one – and perhaps two – of the scenarios plotted in Figure 2 would be inconsistent with the theory, thereby creating an additional testable prediction. For example, assume once more that one's theory predicts a plot of ME(X|Z) taking the form of Figure 1, with

both a positive intercept and a positive slope. We have seen that, by itself, this prediction is consistent with all three plots of ME($Z|X$) in Figure 2. But if the theory were to additionally predict that Z has a positive effect on Y when X is at, say, its highest (or, in fact, any) value, then this would imply that Figure 2b – for which ME($Z|X$) is negative throughout – is inconsistent with the theory. If, in contrast, the theory were to predict that Z has a positive effect on Y when X is at its lowest value, then both Figures 2b and 2c would be eliminated as possibilities. In both of these cases, supplementing an estimated plot of ME($X|Z$) with one of ME($Z|X$) would allow for a stronger test of the underlying conditional theory.

Deriving and Testing as Many Predictions as a Conditional Theory Allows

We now offer some practical advice on deriving and testing hypotheses from conditional theories that can be accurately specified with the linear-interactive model of Eq. (1).

Five Key Predictions

Ideally, a theory positing interaction between X and Z in influencing Y would be strong enough to predict the precise magnitude of the effect of each of X and Z at every possible value of the other variable. Of course, theories in political science are very rarely strong enough to generate such specific predictions. However, we believe that conditional theories in the literature are typically strong enough to generate five basic predictions about the marginal effects of X and Z on Y:[7]

1. $\mathbf{P_{X|Z_{min}}}$: The marginal effect of X is [positive, negative, zero] when Z is at its *lowest* value.
2. $\mathbf{P_{X|Z_{max}}}$: The marginal effect of X is [positive, negative, zero] when Z is at its *highest* value.
3. $\mathbf{P_{Z|X_{min}}}$: The marginal effect of Z is [positive, negative, zero] when X is at its *lowest* value.
4. $\mathbf{P_{Z|X_{max}}}$: The marginal effect of Z is [positive, negative, zero] when X is at its *highest* value.[8]
5. $\mathbf{P_{XZ}}$: The marginal effect of each of X and Z is [positively, negatively] related to the other variable.

[7] These predictions are based on the case in which an author's conditional theory conforms to a model of the form shown in Eq. (1). More complex conditional theories would produce testable predictions of a different form and require an alternative model specification.

[8] Two issues regarding these predictions are worth noting. First, when Z is dichotomous, the predictions $P_{Z|X_{min}}$ and $P_{Z|X_{max}}$ should be stated in terms of the response of Y to a *discrete change* in Z rather than in terms of the *marginal effect* of Z. This is because the concept of a marginal effect makes sense only when it is possible to conceive of an infinitesimally small change in Z. The predictions $P_{X|Z_{min}}$ and $P_{X|Z_{max}}$ should be stated similarly when X is dichotomous. Second, when any of these predictions points to a *zero* effect, in which one independent variable has *no* effect at an extreme value of the other, scholars need to think very carefully about whether the functional form of Eq. (1) properly specifies the expected nature of the interaction (see the Appendix).

Note that by calling for researchers to state predictions about what happens when X and Z are at their lowest and highest values, we do not imply that analysts should necessarily focus greatest attention on *estimated* marginal effects at these extreme values. Indeed, as we note below, when there are few observations at these extremes, estimates of marginal effects at these values are less relevant for testing the theory than estimates of marginal effects at values for X and Z around which there are more observations. Rather, we call for predictions at the extremes simply because if one assumes linearity as in Eq. (1), or at least monotonicity, such predictions automatically imply predictions at values between the extremes.[9]

The predictions outlined above need not be presented as five separate hypotheses. Indeed, with careful phrasing, all five predictions can be subsumed in a single hypothesis about how the marginal effect of X varies with Z and a single hypothesis about how the marginal effect of Z varies with X. This is illustrated in the following pair of hypotheses:

- $\mathbf{H_{X|Z}}$: The marginal effect of X on Y is positive at all values of Z; this effect is strongest when Z is at its lowest and declines in magnitude as Z increases.
- $\mathbf{H_{Z|X}}$: The marginal effect of Z on Y is positive when X is at its lowest level. This effect declines in magnitude as X increases; at some value of X, Z has no effect on Y. As X rises further, the effect of Z becomes negative and strengthens in magnitude as X increases.

Note that $\mathrm{H}_{X|Z}$ implies that the marginal effect of X is positive at both the lowest and highest values of Z, thereby offering predictions $\mathrm{P}_{X|Z_{\min}}$ and $\mathrm{P}_{X|Z_{\max}}$. $\mathrm{H}_{Z|X}$ states that the marginal effect of Z is positive at X's lowest value and negative at X's highest value, thereby offering predictions $\mathrm{P}_{Z|X_{\min}}$ and $\mathrm{P}_{Z|X_{\max}}$. There is no need to state a separate hypothesis that each independent variable is negatively related with the marginal effect of the other because such a prediction - that of P_{XZ} - is implicit in both $\mathrm{H}_{X|Z}$ and $\mathrm{H}_{Z|X}$. Thus, in combination, $\mathrm{H}_{X|Z}$ and $\mathrm{H}_{Z|X}$ include all five predictions we recommend and offer as complete a description of the expected interaction between X and Z as one could offer for a linear-interactive model without predicting specific magnitudes for marginal effects at specific values of the independent variables.

In general, scholars who propose a theory should seek to test as many of the theory's implications as possible. When it comes to interactive theories that can be accurately specified by the linear model of Eq. (1), this requires making, and then testing, as many of the five predictions listed above as possible. Later, we illustrate this recommendation by revisiting two recent studies estimating an interactive model –

[9] Of course, when one independent variable - say X - is dichotomous, the highest and lowest values of X are the only two possible values for X, and thus the predictions $\mathrm{P}_{Z|X_{\min}}$ and $\mathrm{P}_{Z|X_{\max}}$ together describe the marginal effect of Z at all possible values of X.

one in comparative politics and one in international relations – and considering whether each utilizes the model's coefficient estimates to test all of the predictions that the author's theory generates. Before we do this, though, we briefly discuss several issues that arise when evaluating empirical evidence in favor of, or against, conditional theories.

Some Prototypical Results When Testing Interactive Models

Suppose we want to evaluate the empirical support for the conditional theory from which hypotheses $H_{X|Z}$ and $H_{Z|X}$ in the previous section are derived following the advice we have offered. We would estimate Eq. (1) and then use the model's coefficients to construct marginal effect plots for both X and Z. Clearly, the evidence in favor of the theory would be greatest in the case where we find strong support for each of the five predictions made by hypotheses $H_{X|Z}$ and $H_{Z|X}$. This would involve finding that the point estimates for ME($X|Z = z_{\min}$), ME($X|Z = z_{\max}$), and ME($Z|X = x_{\min}$) are all positive, statistically significant, and substantively significant; and that the point estimates for ME($Z|X = x_{\max}$) and β_{XZ} are both negative, statistically significant, and substantively significant [where "min" and "max" refer to the minimum and maximum observed values of a variable in the sample]. Below, when we use the term "significant" without any qualification, it is meant to imply that *both* statistical and substantive significance have been established.[10]

But should we require that *all* of these conditions be met before we claim any empirical support for our conditional theory, and reject our theory if any of the conditions is not achieved? Ultimately, we believe that this is an unrealistically strong standard for empirical evidence and that it would be a mistake to treat all situations in which at least one of these conditions fails to be met as equivalent. Although *firm* knowledge that one of the five predictions from earlier is false would be sufficient logical grounds for concluding that the underlying theory is false, it is important to remember that statistical tests cannot tell us with certainty whether any of the predictions is false; all they offer is information about the risks of a false inference when one rejects the null hypothesis that a value equals zero. For this reason, it is inappropriate to establish "hard and fast" rules about what combinations of evidence regarding the five predictions constitute support for the

[10] Unless we explicitly state to the contrary, "statistically significant" in this paper implies significantly different from zero at some specified significance (α) level. When we say that a point estimate is "*substantively* significant," we mean that its value is large enough to be deemed of nontrivial magnitude. We recognize that the minimum magnitude required for substantive significance is subjective and that there is no single correct way of establishing substantive significance. In our replication of a study by Alexseev (2006) later in the paper, we illustrate one potentially useful strategy for demonstrating the substantive significance of interactive relationships. For more on the important difference between statistical and substantive significance, see Achen (1982, 41-51).

underlying conditional theory.

Nevertheless, we can examine several prototypical sets of results one might get when estimating an interactive model taking the form of Eq. (1), and for each, assess the extent to which we would feel comfortable claiming support for the underlying conditional theory given the empirical evidence presented. To ground the discussion, assume we seek to test the theory generating hypotheses $H_{X|Z}$ and $H_{Z|X}$. A strong test would require that we use the model's coefficient estimates to evaluate all five of the predictions contained in $H_{X|Z}$ and $H_{Z|X}$. However, for illustrative purposes, we simplify matters in the discussion that follows by focusing on hypothesis $H_{X|Z}$ and the three predictions that it contains: (i) ME($X|Z = z_{\min}) > 0$, (ii) ME($X|Z = z_{\max}) > 0$, and (iii) $\beta_{XZ} < 0$.

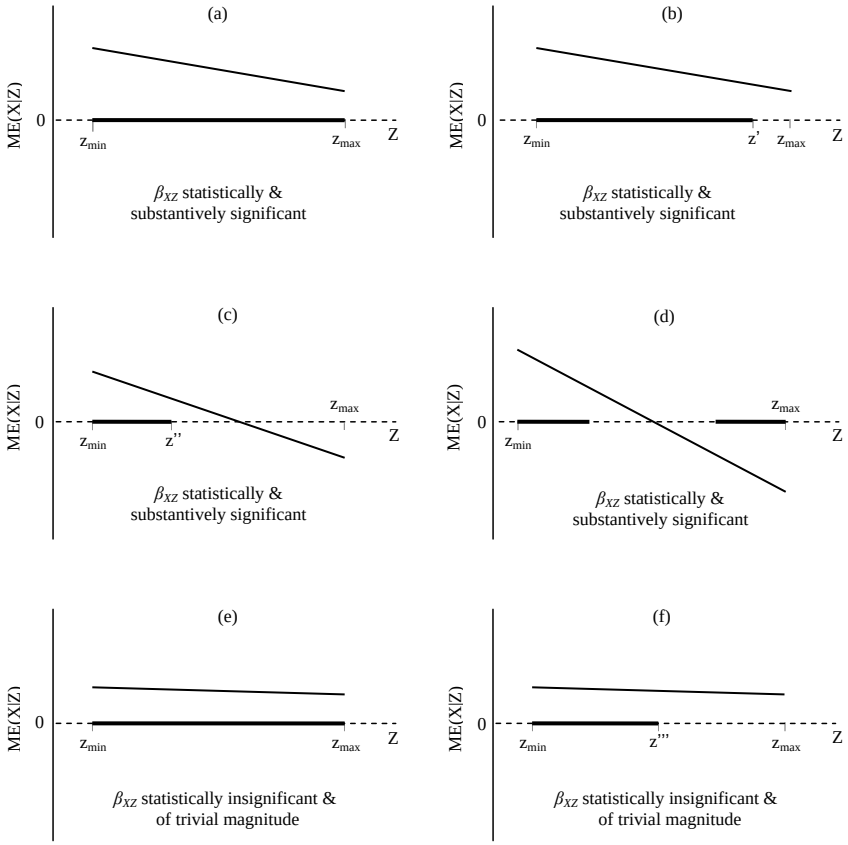
Six different prototypical sets of results are portrayed in Figure 3 in the form of a marginal effect plot for *X*. Although we strongly recommend that authors reporting *substantive* research present confidence intervals in their marginal effect plots, we exclude these intervals in Figure 3 because they complicate the plots with detail that is unnecessary for our specific purposes. Instead, we identify in each plot all the values of *Z* at which the marginal effect of *X* on *Y* is significant (i.e., both statistically and substantively significant) by making the horizontal axis bold.[11] Under each plot, we also indicate whether the coefficient on the product term, β_{XZ}, is statistically significant. This last piece of information is not usually included in published marginal effect plots but is critical for determining whether there is empirical evidence of interaction between *X* and *Z*, i.e., for testing prediction P_{XZ}.[12] This is because, as Eq. (4) reminds us, β_{XZ} indicates the strength of the relationship between both (i) ME(*X*|*Z*) and *Z*, and (ii) ME(*Z*|*X*) and *X*:

$$\frac{\partial Y}{\partial X \partial Z} = \frac{\partial Y}{\partial Z \partial X} = \beta_{XZ}. \tag{4}$$

[11] Making the axis bold to indicate the region in which the marginal effect is significant is a useful convention for this paper because it allows us to discuss prototypical results without having to make an arbitrary choice about the appropriate level (e.g., 99%, 95%) for the confidence intervals, or offer a detailed analysis of substantive significance in the absence of a concrete example. However, we recommend against using this convention in actual research.

[12] Note that whether this is true for models with limited dependent variables like logit and probit depends on the dependent variable of conceptual interest. There are two possible dependent variables of conceptual interest when estimating a binary logit or probit model: (i) an unbounded latent variable, Y^*, assumed to be measured by the observed dichotomous variable, *Y*, and (ii) the probability that *Y* equals one, $\Pr(Y = 1)$. When one's dependent variable of interest is the unbounded Y^*, then the product term coefficient, β_{XZ}, reflects the extent of interaction. However, this is *not* the case when the dependent variable of interest is $\Pr(Y = 1)$. Indeed, when the dependent variable is $\Pr(Y = 1)$, one cannot determine whether there is interaction between *X* and *Z* by inspecting the coefficient on the product term (or any single term). The fact that the marginal effect of each of *X* and *Z* on $\Pr(Y = 1)$ is not linearly related to the other variable means that prediction P_{XZ} must be evaluated by estimating the marginal effects of *X* and *Z* at different values for the independent variables and assessing how they change as the values of the independent variables change (Ai & Norton 2003, Norton, Wang & Ai 2004, Berry, DeMeritt & Esarey 2010).

Figure 3: Plots of ME($X|Z$) Reflecting Several Prototypical Sets of Empirical Results



Notes: The horizontal axis is bold for all values of Z at which the marginal effect of X on Y is statistically and substantively significant. $z_{\min}$ and $z_{\max}$ indicate the lowest and highest observed values of Z.

To facilitate readers seeing as much statistical evidence relevant for testing a theory positing interaction as possible, we therefore recommend that scholars routinely report the estimated product term coefficient and a t-ratio or standard error for this coefficient in their marginal effect plots.

Consider first the plot shown in Figure 3a. The marginal effect of X is positive and significant across the observed range of Z, and β_{XZ} is negative and significant. This plot provides unambiguously strong evidence for hypothesis $\mathrm{H}_{X|Z}$ because each of its three predictions receives strong empirical support. Next, consider the plot shown in Figure 3b. The only difference here is that the marginal effect of X is no longer significant when Z is at its highest value. However, because $\mathrm{H}_{X|Z}$ predicts that the marginal effect of X on Y declines in magnitude as Z increases, which leaves open the possibility of a weak effect by the time Z

gets large, we are not particularly troubled to find that ME($X|Z = z_{\max}$) fails to be significant. Thus, in this situation, we would conclude that there is strong support for $H_{X|Z}$ even though the value for ME($X|Z = z_{\max}$) is not significant.[13]

The plot shown in Figure 3c provides a more ambiguous case. As before, the significant coefficient on the product term represents clear evidence that the marginal effect of X is conditional on Z as predicted. The difference is that the range of values for Z for which the marginal effect of X is positive and significant is now smaller than in Figure 3b and the point estimate for ME($X|Z = z_{\max}$) is actually negative. In this scenario, we are not terribly concerned that the point estimate for ME($X|Z = z_{\max}$) takes the "wrong" sign because it is, we assume here, statistically or substantively insignificant. We would argue that how supportive these results are of hypothesis $H_{X|Z}$ depends on the percentage of observations having values of Z at which the marginal effect of X is positive and significant, i.e., for which $Z < z$" in the figure. The higher this percentage, the more inclined we would be to accept the empirical evidence as supportive. Of course, the minimum percentage high enough to justify a claim of support is subjective. As a result, we recommend that scholars report the percentage of observations that fall within the region of significance. Indeed, it would be very helpful if researchers would provide a frequency distribution for the variable plotted on the horizontal axis so that readers can assess for themselves the relative density of observations across the range of X. We illustrate how such a frequency distribution might be incorporated into a marginal effect plot when we report the results of two replications in the next section.

When it comes to evaluating conditional theories, one practice that we strongly advise against is getting into a "counting game" in which one's conclusion is based strictly on the number of predictions for which there is statistical support. For example, consider the plot shown in Figure 3d. This plot provides statistical confirmation for two of the three predictions contained in $H_{X|Z}$, namely that ME($X|Z = z_{\min}$) is positive and that β_{XZ} is negative. The fact that β_{XZ} is significant provides strong empirical evidence of interaction between X and Z. Importantly, though, the plot suggests that this interaction takes an appreciably different form than that predicted by hypothesis $H_{X|Z}$. Although X has the expected positive effect when Z is low, X has a significant negative effect when Z gets large. We believe that scholars should not sweep this

[13] Ideally, the theory underlying $H_{X|Z}$ would be strong enough to generate a prediction about whether the marginal effect of X on Y should (i) remain strong even when Z reaches its maximum, or (ii) decline to near zero when Z is maximized. In the former case, the theory would predict that X has a significant effect on Y when $Z = z_{\max}$. But in the latter case, it would predict an insignificant effect when $Z = z_{\max}$. Admittedly, theories in political science are rarely capable of yielding such a fine distinction.

kind of inconsistency with the hypothesis "under the rug" by claiming a healthy "batting average" of 0.667, with two of the three predictions confirmed. Evidence that when Z is high, increases in X yield substantial *decreases* in Y rather than the predicted *increases* strikes us as sufficient to raise serious concerns about the conditional theory underlying the hypothesis.

Figure 3e illustrates a more extreme case in which claiming support for $H_{X|Z}$ based on two of the three predictions receiving statistical support would be unwarranted. In this case, the marginal effect of X is positive and significant across the entire observed range of Z, thereby indicating support for the predictions that ME($X|Z = z_{\min}$) and ME($X|Z = z_{\max}$) are positive. However, although β_{XZ} is negative as predicted, it lacks statistical significance and the nearly flat marginal effect line indicates that the magnitude of β_{XZ} is substantively trivial. In essence, there is no evidence of appreciable interaction between X and Z. Indeed, this sort of plot - with a marginal effect line sloped slightly upward or downward - is exactly what we would expect to find if we were to estimate Eq. (1) when each of X and Z has a strong positive effect on Y but their effects are additive rather than interactive. Thus, the evidence in Figure 3e seriously challenges the theory predicting that X and Z interact in influencing Y.

We now consider a final set of prototypical results shown in Figure 3f. Once again, the line plotted is intended to be nearly flat. Assume that the effect of X on Y is *substantively* significant at all values of X, but *statistically* significant only when $Z < z$"'. The fact that the marginal effect of X changes from statistically significant when $Z < z$"' to statistically insignificant when $Z \geq z$"' might seem to suggest that there is interaction between X and Z. Indeed, BCG (2006, 74) imply precisely this when they claim that a situation in which the marginal effect of X on Y is statistically significant for some values of Z but not for others might be interpreted as a sign of meaningful interaction even when the coefficient on the product term is statistically insignificant. However, this is incorrect. The nearly flat line in Figure 3f represents a case in which the marginal effect of X has a t-ratio barely above the threshold for statistical significance when Z is low and a t-ratio barely below the threshold when Z is high. If one capitalizes on the fact that ME($X|Z$) changes from statistically significant to not as Z surpasses z"' to claim evidence of interaction, one is placing too much reliance on an arbitrarily chosen level of statistical significance. If this level were set slightly higher, ME($X|Z$) would be statistically significant over the entire range for Z. If the level were set slightly lower, the marginal effect would not be statistically significant at any value of Z. The more relevant

information is that the coefficient on the product term, β_{XZ}, is not statistically significant and is of small magnitude. As we showed in Eq.(4), this indicates that the marginal effect of *X* varies only trivially with *Z*, and on this basis we should reject the theory positing interaction underlying $\mathrm{H}_{X|Z}$.

Two Replications

We now illustrate our central points by replicating two studies chosen from the many that test a conditional theory but that present a marginal effect plot for just one of the two variables hypothesized to interact. In one replication, examining the second marginal effect plot lends additional support for the researcher's theory. In the other, the second plot provides evidence that contradicts the author's theory.

Revealing Additional Evidence in Favor of the Theory Being Tested

In an article in the *Journal of Conflict Resolution*, Kastner (2007) examines how conflicting interests and the strength of domestic actors with internationalist economic interests affect the level of trade between countries. Previous studies indicate that bilateral trade tends to be lower when countries have conflicting political interests. As Kastner notes, though, there is considerable variation across country dyads in the extent to which conflicting interests lead to reduced bilateral trade. His explanation for this variation centers on the strength of domestic actors who benefit from trade. Specifically, Kastner argues that although leaders generally want to reduce trade with countries that do not share their interests, some leaders are constrained in their ability to do this by the presence of strong domestic actors with internationalist economic interests. As Kastner (p. 670) puts it, "the negative effects of conflict on commerce should be less severe when internationalist economic interests have strong political clout domestically." Unable to measure the strength of internationalist interests in a dyad directly, Kastner uses the extent of trade barriers in the countries (*Trade Barriers*) as a proxy variable that is inversely related to the strength of these interests. If we denote the extent of conflict between two countries by *Conflict* and their level of bilateral trade by *Trade*, Kastner's hypothesis can be stated as follows:

- $\mathbf{H}_{\textit{Conflict}|\textit{Barriers}}$: The marginal effect of *Conflict* on *Trade* is negative at all values of *Trade Barriers*; this negative effect is weakest when *Trade Barriers* is at its lowest level and strengthens in magnitude as *Trade Barriers* increases.

Kastner tests his conditional theory using annual data from 76 countries from 1960 to 1992 and an OLS model with an interactive specification taking the form of Eq. (1):

$$\begin{aligned} \textit{Trade} = \beta_0 + \beta_C \textit{Conflict} + \beta_B \textit{Trade Barriers} + \beta_{CB} \textit{Conflict} \times \textit{Trade Barriers} \\ + \beta \textit{Controls} + \epsilon, \end{aligned} \tag{5}$$

where *Controls* is a vector of control variables. The coefficient on the product term, *Conflict*×*Trade Barriers*, is negative and statistically significant at the 0.01 level, with a t-statistic of -5.26. Using the parameter estimates from his model (Table 1, Model 1, p. 676), Kastner produces a plot showing how the marginal effect of *Conflict* on *Trade* varies with the level of *Trade Barriers*. We reproduce this marginal effect plot in a slightly modified form in Figure 4a.[14] Based on the plot, as well as the statistically significant negative coefficient on the product term, Kastner claims empirical support for his theory.

We advise researchers who propose a theory positing interaction between two variables, *X* and *Z*, to use the theory to generate as many of the five key predictions listed earlier as the theory allows regarding the marginal effects of *X* and *Z* on *Y*. Kastner's hypothesis, $\mathrm{H}_{Conflict|Barriers}$, offers three of these predictions:

- $\mathbf{P}_{\boldsymbol{C|B_{\min}}}$: The marginal effect of *Conflict* on *Trade* is negative when *Trade Barriers* is at its lowest value.
- $\mathbf{P}_{\boldsymbol{C|B_{\max}}}$: The marginal effect of *Conflict* on *Trade* is negative when *Trade Barriers* is at its highest value.
- $\mathbf{P}_{\boldsymbol{CB}}$: The marginal effect of each of *Conflict* and *Trade Barriers* is negatively related to the other variable.

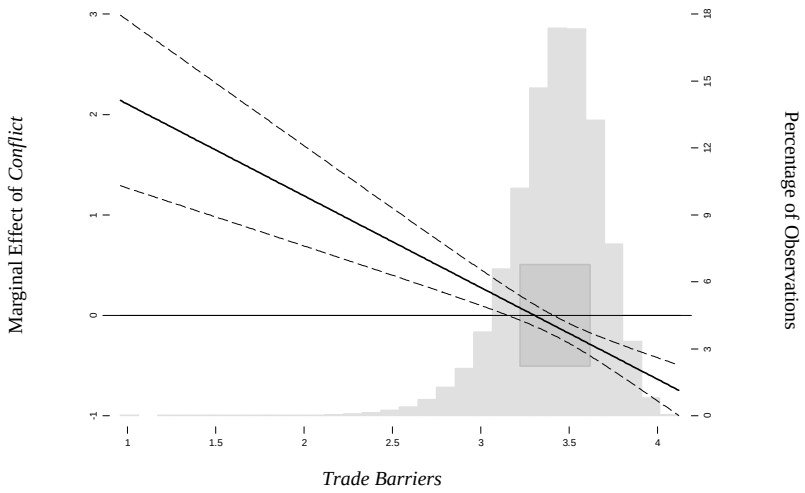
However, Kastner's hypothesis is silent about the expected value (positive, negative, or zero) of the marginal effect of *Trade Barriers* at the highest and lowest values of *Conflict*.

Before we consider the marginal effect of *Trade Barriers* on *Trade*, we reevaluate the empirical support for predictions $\mathrm{P}_{C|B_{\min}}$, $\mathrm{P}_{C|B_{\max}}$, and P_{CB}. Assuming that the statistically significant negative coefficient

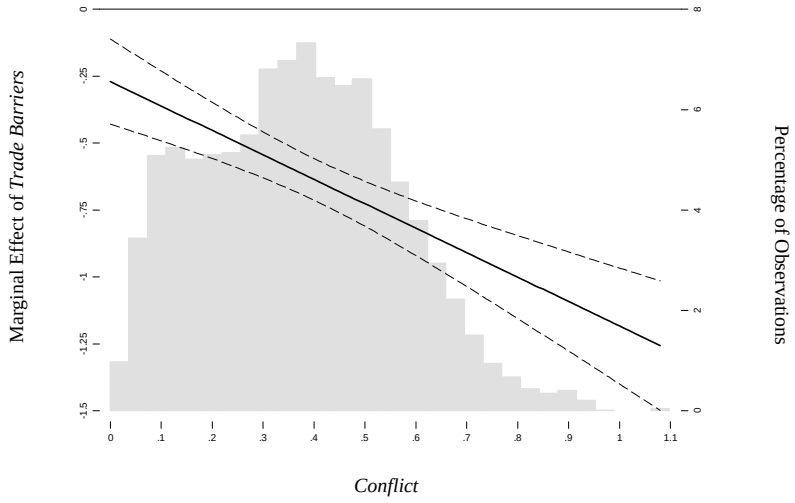
[14] We were able to replicate Kastner's results perfectly. Our marginal effect plot differs from his (Figure 1, p. 677) in four respects. Rather than plotting the marginal effect of *Conflict* on *Trade* on the vertical axis, as we do, Kastner plots the change in *Trade* as *Conflict* increases from its 15^{th} percentile in the sample to its 85^{th} percentile. Given the linear form of Eq. (5), this difference in scaling the vertical axis is superficial because one scaling is a linear transformation of the other. Second, Kastner plots percentiles for *Trade Barriers* in the sample along the horizontal axis. We saw no good reason to distort the scale for *Trade Barriers* by using percentiles rather than the actual values. This difference in scaling for the horizontal axis explains why our plot is linear, but Kastner's is not. Third, we plot the marginal effect of *Conflict* on *Trade* over the entire range of values for *Trade Barriers* in the estimation sample, whereas Kastner plots it only over the values for *Trade Barriers* that fall between the 20^{th} and 80^{th} percentiles. Fourth, we have added shading to our marginal effect plot; we explain the purpose of this shading below.

Figure 4: Marginal Effect Plots Designed to Evaluate the Conditional Theory Presented by Kastner (2007)

(a) Marginal Effect of *Conflict* on *Trade*



(b) Marginal Effect of *Trade Barriers* on *Trade*



- - - - - - - - 95% Confidence Intervals

Coefficient on product term is -0.91 (Standard Error = 0.173; *t*-ratio = -5.26)

Notes: The marginal effect plots are constructed using parameter estimates for Model 1 in Kastner's (2007) Table 1. The vertical axes on the left indicate the magnitude of the marginal effect. The vertical axes on the right are for the histogram and indicate the distribution of observations in the sample on the variable depicted on the horizontal axis. The dark shaded rectangle in panel (a) shows the restricted marginal effect plot presented by Kastner in his Figure 1, which includes values for *Trade Barriers* only between the 20^{th} and 80^{th} percentiles.

for the product term in Eq. (5) is also substantively significant, there is unambiguous support for prediction P_{CB}.[15] In other words, there is clear evidence that the marginal effect of *Conflict* on *Trade* is negatively related to the value of *Trade Barriers*, as Kastner hypothesizes, and (due to the symmetry of interactions) that the marginal effect of *Trade Barriers* is negatively related to the value of *Conflict*.

But is this conditionality consistent with predictions $P_{C|B_{\min}}$ and $P_{C|B_{\max}}$? On the one hand, Figure 4a shows that the marginal effect of *Conflict* is negative and statistically significant when *Trade Barriers* takes on its largest observed value, thereby supporting prediction $P_{C|B_{\max}}$. On the other hand, prediction $P_{C|B_{\min}}$ fails to receive empirical support. Contrary to expectation, the marginal effect of *Conflict* is positive and statistically significant when *Trade Barriers* is at its smallest observed value, and indeed, at all values less than 3.16. Overall, the marginal effect plot for *Conflict* closely resembles the prototypical plot shown in Figure 3d. This raises concerns about the conditional theory underlying the hypothesis being tested because the estimated marginal effect is statistically significant in the "wrong" direction at one end of the horizontal axis. Kastner offers no explanation for why an increase in conflict should lead to increased bilateral trade when domestic actors with internationalist economic interests are strong, i.e., when trade barriers are low. In our view, the statistically significant positive effect of *Conflict* when *Trade Barriers* is less than 3.16 should not be dismissed as a trivial inconsistency; rather, it is an important piece of evidence to consider alongside the support for predictions $P_{C|B_{\max}}$ and P_{CB} when evaluating Kastner's theory.

Our replication of Kastner's analysis illustrates the importance of constructing a marginal effect plot that shows how the effect of X on Y varies over the entire observed range of Z. Kastner plots the marginal effect of *Conflict* only for values of *Trade Barriers* between the 20^{th} and 80^{th} percentiles; this interval is indicated by the shaded rectangle in Figure 4a. Note that in this restricted range for *Trade Barriers*, the estimated marginal effect of *Conflict* on *Trade*, although positive for low values of *Trade Barriers*, is never positive and statistically significant. Thus, although the full marginal effect plot reveals values for *Trade Barriers* at which there is clear evidence of an unexpected positive effect of *Conflict* on *Trade*, the restricted plot masks the existence of these values and makes it appear as if the estimated positive effect of *Conflict* never achieves statistical significance even at the lowest values for *Trade Barriers*. Indeed,

[15] Kastner does not explicitly evaluate the substantive significance of the estimated effects he reports. To facilitate the illustrative value of our replication, we simply assume that "statistical significance" implies "significance" (i.e., both statistical and substantive significance). Accordingly, readers should avoid drawing any substantive conclusions about the forces determining the level of international trade from our analysis.

Kastner's restricted plot more closely parallels the prototypical plot shown in Figure 3c, which we argued earlier potentially offers support for the hypothesis being tested depending on the percentage of sample observations falling into the region of significance.

Consider the results in Figure 4a more closely. The marginal effect of *Conflict* is negative and statistically significant when *Trade Barriers* exceeds 3.41. Superimposed over the marginal effect plot is a histogram portraying the frequency distribution for *Trade Barriers*; the scale for the distribution is given by the vertical axis on the right-hand side of the graph. The histogram shows that 55.4% of the country dyads in Kastner's sample fall into this range of statistical significance. At the other extreme, the effect of *Conflict* is positive and statistically significant when *Trade Barriers* is less than 3.16. 14.5% of the sample observations lie in this range.[16] Although these latter observations, which are inconsistent with Kastner's theory, do not constitute a large percentage of the sample, they are far from being a trivial set of outlier observations.

Of course, the primary point of this article is that it may be a mistake to draw any conclusion about Kastner's theory based solely on the coefficient estimate for the product term and the marginal effect plot shown in Figure 4a. We should also determine whether Kastner's conditional theory generates predictions about the marginal effect of *Trade Barriers* on *Trade* across the range of values for *Conflict* and, if so, determine whether these predictions receive empirical support. Although Kastner provides no explicit hypothesis about the effect of internationalist economic interests on bilateral trade, his underlying theory is not silent on the matter. As Kastner (p. 670) notes, "leaders who depend on support from actors who benefit from trade pay, at the margins, higher domestic political costs for placing restrictions on foreign commerce than do other leaders." This line of reasoning leads to the prediction that stronger internationalist economic interests among domestic groups will prompt increased bilateral trade irrespective of the level of conflict between countries. Given that the proxy variable, *Trade Barriers*, is inversely related to the strength of internationalist economic interests, Kastner's theory implies the following new hypothesis:

- $\mathbf{H}_{\textit{Barriers}|\textit{Conflict}}$: The marginal effect of *Trade Barriers* on *Trade* is negative at all values of *Conflict*. This negative effect is weakest when *Conflict* is at its lowest level and increases in magnitude as

[16] The fact that there are few observations at low levels of *Trade Barriers* means that the evidence in the figure that *Conflict* has a positive effect on *Trade* when *Trade Barriers* is low rests heavily on the model's linearity assumption. Unless one believes that there is a strong a priori theoretical justification for this assumption, one should be skeptical about drawing strong inferences concerning the marginal effect of *Conflict* at low levels of *Trade Barriers*. Note that readers would be completely unaware of this issue in the absence of a histogram illustrating the dearth of sample observations with low values of *Trade Barriers*. This highlights the importance of including in a marginal effect plot information about the distribution of the variable depicted on the horizontal axis.

Conflict increases.[17]

This hypothesis yields two predictions that together with $P_{C|B_{\min}}$, $P_{C|B_{\max}}$, and P_{CB} constitute the full set of five predictions that we delineated earlier:

- $\mathbf{P_{B|C_{min}}}$: The marginal effect of *Trade Barriers* on *Trade* is negative when *Conflict* is at its lowest value.
- $\mathbf{P_{B|C_{max}}}$: The marginal effect of *Trade Barriers* on *Trade* is negative when *Conflict* is at its highest value.

In Figure 4b, we plot the estimated marginal effect of *Trade Barriers* across the observed range of *Conflict* values. This graph provides strong support for the two new predictions, and thus, Kastner's conditional theory. As expected, *Trade Barriers* has a statistically significant negative marginal effect on *Trade* across the entire observed range for *Conflict*. By failing to (i) make explicit some of the predictions ($P_{B|C_{\min}}$ and $P_{B|C_{\max}}$) that are clearly implied by his theory and (ii) construct a marginal effect plot that can be used to evaluate these predictions, Kastner fails to recognize empirical evidence in support of his theory. Readers seeking to assess Kastner's theory should consider both plots shown in Figure 4, as well as the estimated product term coefficient. They should weigh the considerable evidence consistent with the underlying theory against the contradictory finding that the marginal effect of *Conflict* is significantly positive over a range of values for *Trade Barriers* accounting for 14.5% of Kastner's sample observations. Regardless of the importance one attaches to the evidence that is in conflict with Kastner's theory, it is certainly the case that the information derived by constructing a second marginal effect plot adds to the evidence in support of his theory.

Revealing Additional Evidence Contrary to the Theory Being Tested

In an article in *Political Behavior*, Alexseev (2006) examines how changes in the ethnic composition of Russia's regions affect the vote share won by the extreme Russian nationalist Zhirinovsky Bloc in the 2003 elections to the Russian State Duma. Alexseev investigates the ability of three competing theories - the "power threat" model, the "power differential" model, and the "defended nationhood" model - to explain the level of electoral support received by the Zhirinovsky Bloc. Ultimately, Alexseev concludes that the defended nationhood model provides the best explanation. According to this model, support for anti-immigrant

[17] This sentence is implicit in hypothesis $H_{Conflict|Barriers}$ due to the inherent symmetry of interactions.

parties (*Xenophobic Voting*) depends on the percentage of the population in a region belonging to the dominant ethnic group and the change in the percentage of the population accounted for by ethnic minorities (pp. 218-220). More specifically, an increase in the size of the dominant ethnic group should enhance the support for anti-immigrant parties and this positive effect should be greater in regions that have experienced a large influx of ethnic minorities. Moreover, the change in the percentage of the population comprised of ethnic minorities should have a positive effect on support for anti-immigrant parties regardless of the size of the dominant ethnic group. In Russia, Slavs constitute the dominant ethnic group. Thus, in the Russian context, Alexseev's defended nationhood hypothesis can be stated as follows:

- $\mathbf{H}_{\textit{Alexseev}}$: The marginal effect of the size of the dominant ethnic group (*Slavic Share*) on support for the Zhirinovsky Bloc (*Xenophobic Voting*) is always positive; this positive effect grows in strength as the increase in the share of the population comprised by ethnic minorities (Δ*non-Slavic Share*) gets larger (or the decrease in Δ*non-Slavic Share* gets smaller). The marginal effect of Δ*non-Slavic Share* on *Xenophobic Voting* is positive at any value for *Slavic Share*.

Alexseev tests this hypothesis using data from 72 Russian regions and an OLS model with an interactive specification in the form of Eq. (1):

$$\begin{aligned} \textit{Xenophobic Voting} = \beta_0 + \beta_S \textit{Slavic Share} + \beta_N \Delta\textit{non-Slavic Share} \\ + \beta_{SN} \textit{Slavic Share} \times \Delta\textit{non-Slavic Share} + \beta \textit{Controls} + \epsilon, \end{aligned} \tag{6}$$

where *Controls* is a vector of control variables. Using the parameter estimates from this model (Table 2, Test 1, p. 225), Alexseev produces a plot showing how the marginal effect of *Slavic Share* on *Xenophobic Voting* varies with Δ*non-Slavic Share*. We reproduce this plot in Figure 5a.[18] Based on this plot, Alexseev claims empirical support for the defended nationhood model.

Note that Alexseev's hypothesis contains the full set of five predictions we urge scholars with conditional theories to offer readers:

- $\mathbf{P}_{S|N_{\text{min}}}$: The marginal effect of *Slavic Share* on *Xenophobic Voting* is positive when Δ*non-Slavic Share* is at its lowest value.

[18] We were unable to replicate Alexseev's OLS results perfectly. However, our results are extremely close to his. Indeed, the ratio of the coefficient with the larger magnitude across the two estimations to the coefficient with the smaller magnitude was less than 1.01 for all but one regressor; for the one exception, the ratio was 1.016. Not surprisingly, the lines on our respective marginal effect plots are visually indistinguishable. The only other difference between our marginal effect plot and Alexseev's is that we show the marginal effect of *Slavic Share* across the full range of values for Δ*non-Slavic Share* in the sample (including negative values that indicate that the non-Slavic population share is decreasing), whereas Alexseev truncates the horizontal axis at zero.

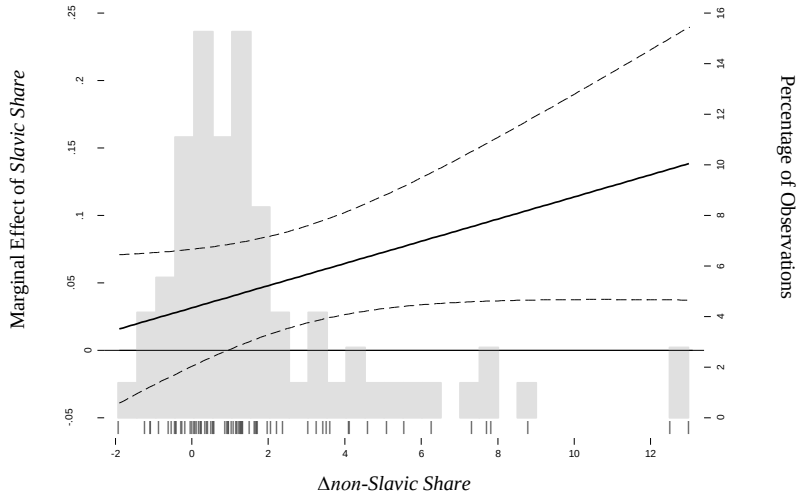
- $\mathbf{P_{S|N_{max}}}$: The marginal effect of *Slavic Share* on *Xenophobic Voting* is positive when Δ*non-Slavic Share* is at its highest value.
- $\mathbf{P_{N|S_{min}}}$: The marginal effect of Δ*non-Slavic Share* on *Xenophobic Voting* is positive when *Slavic Share* is at its lowest value.
- $\mathbf{P_{N|S_{max}}}$: The marginal effect of Δ*non-Slavic Share* on *Xenophobic Voting* is positive when *Slavic Share* is at its highest value.
- $\mathbf{P_{SN}}$: The marginal effect of each of *Slavic Share* and Δ*non-Slavic Share* is positively related to the other variable.

Although Alexseev's hypothesis yields all five of the predictions that we recommend, he evaluates only three of them: $P_{S|N_{min}}$, $P_{S|N_{max}}$, and P_{SN}. We begin by reevaluating the support for these three predictions. In line with prediction P_{SN}, the coefficient on the product term is positive, indicating that the marginal effect of each of *Slavic Share* and Δ*non-Slavic Share* is positively related to the other variable. Although the coefficient on the product term is not statistically significant at the 0.05 level in the two-tail test that Alexseev reports, it is significant at the 0.10 level in a two-tail test or, equivalently, at the 0.05 level in a one-tail test. Given the relatively small sample size ($n = 72$) and the fact that the coefficient is very close to being statistically significant at standard levels, we would not be prepared to reject Alexseev's theory on this ground alone.

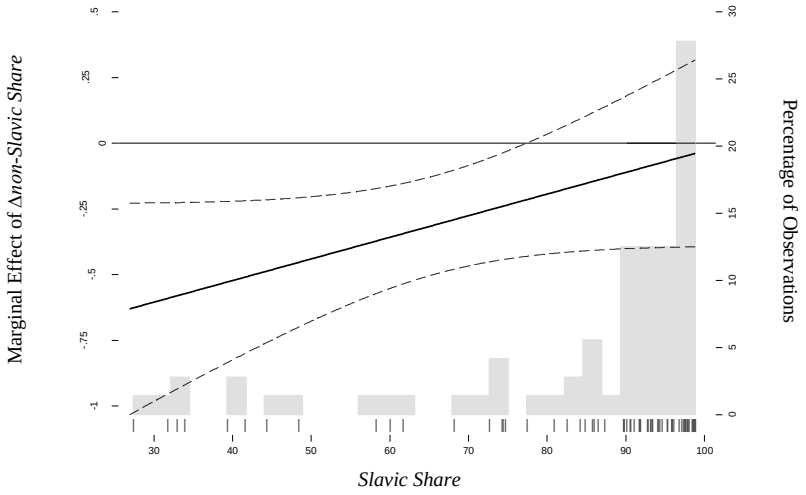
For further relevant information, it is useful to assess the magnitude of the interaction reflected by the point estimate for the product term in more substantive terms. Our goal is to determine whether the estimated marginal effect of *Slavic Share* on *Xenophobic Voting* changes by a nontrivial amount as Δ*non-Slavic Share* changes. We first note that there is substantial variation in *Xenophobic Voting* within Alexseev's sample. For example, the electoral support for the Zhirinovsky Bloc ranges from 2.8% to 19.5% across the Russian regions. The product term coefficient can be used to predict the response of *Xenophobic Voting* to an increase in *Slavic Share* from its lowest value (27.4) in the sample to its highest value (98.9) at both the lowest (-1.93) and highest (12.99) values of Δ*non-Slavic Share*. When Δ*non-Slavic Share* is at its lowest value, a shift across the range for *Slavic Share* produces an expected increase of 1.11 in the Zhirinovsky vote percentage. This expected increase amounts to just 6.7% of the range of *Xenophobic Voting* in the sample and, therefore, indicates a substantively trivial estimated effect. In stark contrast, a shift across the range for *Slavic Share* when Δ*non-Slavic Share* is at its highest value prompts an expected increase of 9.89 in the Zhirinovsky vote percentage, a value equal in magnitude to 59.2% of the range of *Xenophobic Voting* in the sample. This indicates that *Slavic Share* has a strong effect in the expected direction when Δ*non-Slavic*

Figure 5: Marginal Effect Plots Designed to Evaluate the "Defended Nationhood" Model Presented by Alexseev (2006)

(a) Marginal Effect of *Slavic Share* on *Xenophobic Voting*



(b) Marginal Effect of Δ*non-Slavic Share* on *Xenophobic Voting*



- - - - - - - - 95% Confidence Intervals

Coefficient on product term is 0.0082 (Standard Error = 0.0047; *t*-ratio = 1.76)

Notes: The marginal effect plots are constructed using the parameter estimates associated with Test 1, Table 2 in Alexseev (2006, 225). The vertical axes on the left indicate the magnitude of the marginal effect. The vertical axes on the right are for the histogram and indicate the distribution of observations in the sample on the variable depicted on the horizontal axis. Underneath each marginal effect plot is a rug plot, i.e., a set of tick marks indicating the precise location of individual observations for the variable on the horizontal axis.

Share is at its highest. This large variation in the substantive magnitude of the effect of *Slavic Share* across different values of Δ*non-Slavic Share*, along with the near statistical significance of Alexseev's product term coefficient in a small sample, leads us to conclude that there is empirical support for prediction P_{SN}.

Predictions $\mathrm{P}_{S|N_{\min}}$ and $\mathrm{P}_{S|N_{\max}}$ together imply that the marginal effect of *Slavic Share* on *Xenophobic Voting* is positive for all values of Δ*non-Slavic Share*. The plot in Figure 5a shows that the point estimate of the marginal effect of *Slavic Share* is, indeed, positive at all values of Δ*non-Slavic Share*. However, the marginal effect is statistically significant only when the change in the non-Slavic share of the population exceeds 0.93. This marginal effect plot is similar to the prototypical plot in Figure 3b. Given that the positive marginal effect of *Slavic Share* is predicted to decline in magnitude as Δ*non-Slavic Share* decreases, a weak effect of *Slavic Share* at low values of Δ*non-Slavic Share* is not at odds with hypothesis $\mathrm{H}_{Alexseev}$. Thus, we do not view the lack of statistical significance of *Slavic Share*'s effect over part of the range of the plot in Figure 5a as an indication that Alexseev's defended nationhood model lacks empirical support.

Once again, however, our principal point is that researchers should test as many implications of their conditional theories as possible. The marginal effect plot in Figure 5a and the product term coefficient provide the information necessary to evaluate predictions $\mathrm{P}_{S|N_{\min}}$, $\mathrm{P}_{S|N_{\max}}$, and P_{SN}, but not predictions $\mathrm{P}_{N|S_{\min}}$ and $\mathrm{P}_{N|S_{\max}}$. We can evaluate the latter two predictions, though, by producing a marginal effect plot for Δ*non-Slavic Share*. This graph is shown in Figure 5b. According to predictions $\mathrm{P}_{N|S_{\min}}$ and $\mathrm{P}_{N|S_{\max}}$, the marginal effect of Δ*non-Slavic Share* should always be positive. Contrary to expectations, however, the point estimate of the marginal effect of Δ*non-Slavic Share* is uniformly negative. Moreover, it is statistically significant when the Slavic share of the population is less than 77.4%, and in our view, it is substantively significant throughout this range as well.[19] The superimposed frequency distribution for *Slavic Share*, this time shown in the form of a histogram *and* a rug plot, illustrates that 22% (or 16) of Russia's 72 regions fall into this region of significance. The bottom line is that although the defended nationhood model predicts that larger increases in the concentration of ethnic minorities in a population will lead to more extensive xenophobic voting, Alexseev's results actually indicate that larger increases will reduce support for anti-immigrant parties, and significantly so over a nontrivial range of values for *Slavic Share*.

[19] Even at the right edge of this range when *Slavic Share* is 77.4%, an increase in Δ*non-Slavic Share* from its lowest to its highest observed value reduces the expected Zhirinovsky vote percentage by 3.2, a value equal to 19.2% of the range of *Xenophobic Voting*. When *Slavic Share* is at its minimum (27.4%), the same change in Δ*non-Slavic Share* decreases the Zhirinovsky vote percentage by 9.34 – equivalent to 55.9% of the range of *Xenophobic Voting*.

What is the relevance of the new evidence presented in Figure 5b? In our view, the addition of this new information means that although there is evidence that *Slavic Share* and Δ*non-Slavic Share* interact to influence *Xenophobic Voting*, the form of this conditionality is sufficiently different from that predicted to cast substantial doubt on Alexseev's defended nationhood model. To square the results in Figure 5b with the defended nationhood model, one would have to reframe the theory to be consistent with the fact that larger increases in the concentration of ethnic minorities result in less xenophobic voting. Some may consider the inconsistent evidence in Figure 5b to be less important than we do. Ultimately, each reader can come to her own conclusion about this. Nevertheless, it seems indisputable that the level of support afforded Alexseev's theory by the full set of empirical results - including both marginal effect plots - is lower than the apparent level of support based solely on the partial set of results in the published paper. Each researcher testing a theory should present readers with as much as possible of the relevant empirical evidence derivable from the model's coefficient estimates so that readers can make a maximally-informed evaluation about the validity of the theory. In Alexseev's case, this means presenting readers with both of the plots shown in Figure 5.

Maximizing the Information Portrayed in a Marginal Effect Plot

In the replications presented in the previous section, we illustrate several practices regarding the construction of marginal effect plots that we hope will become standard in the political science literature. Most importantly, researchers should make the horizontal axis of a marginal effect plot extend from the minimum observed value in the sample for the variable being plotted to the maximum observed value. Plotting marginal effects over a wider range than this risks misleading readers by portraying out-of-sample inferences, whereas plotting marginal effects over a narrower range ignores information that can be relevant for evaluating hypotheses.

But not all values for the variable depicted on the horizontal axis are equally important. For example, if both the minimum and maximum values are outliers in the sample, estimated marginal effects at the extremes are less relevant for assessing the hypothesis under consideration than marginal effects near the center of the distribution, where the observations are more concentrated.[20] Thus, we encourage analysts to

[20]Even at locations on the horizontal axis closer to the center of the distribution, there may be ranges of values at which there are few observations. It is important to remember that the validity of any inferences about the marginal effect of a variable at such values rests on the linearity assumption of the model being correct. It is noteworthy that the confidence interval for the marginal

superimpose over each marginal effect plot a frequency distribution for the variable on the horizontal axis to give readers information about the relative density of data at different locations. Although it depends to some extent on the context, we believe that a combination of a histogram and a rug plot has many virtues.[21] While a histogram provides readers with a general overview of the frequency distribution and a quick sense of the percentage of observations that fall into various regions, a rug plot can be useful because it provides details about the values of individual observations.

Finally, we encourage authors to report the estimated product term coefficient along with its *t*-ratio or standard error somewhere in each marginal effect plot because this is critical information for evaluating hypotheses about interaction that is not evident from the plot itself.

Conclusion

Since the publication of Brambor, Clark, and Golder's (2006) paper, it has become common for political scientists to present a marginal effect plot when interpreting statistical results for a model positing interaction between two variables. This has led to a rapid and dramatic improvement in the quality of research testing conditional theories. Scholars implementing BCG's advice have nearly uniformly (i) conceived of one of the variables, say Z, as *the* conditioning variable, (ii) developed a hypothesis predicting how the marginal effect of the other variable, X, varies with the value of Z, (iii) estimated a model specifying interaction between X and Z, and (iv) constructed a plot of the relationship between Z and the estimated marginal effect of X designed to test the hypothesis. Only rarely have researchers supplemented this hypothesis with a proposition about how the marginal effect of Z varies with the value of X, and a corresponding plot of the estimated marginal effect of Z against X.

When one's theory is insufficient to yield predictions about the relationship between X and the marginal effect of Z beyond that implicit in one's hypothesis about how the marginal effect of X varies with Z – i.e., whether the marginal effect of Z is positively or negatively related to X – then the restriction of attention to just one marginal effect plot is appropriate. But typically, the conditional theories advanced by

effect shown in Figure 5b is actually narrowest at a location on the horizontal axis at which the data are quite scarce; the width of the confidence interval at this point is being driven primarily by the model's linearity assumption, not the sample observations.

[21] Much of the value of a rug plot can be lost when the sample size is large since individual tick marks blend together and become indistinguishable. This explains why we include a rug plot in Figure 5 for the Alexseev replication ($n = 72$), but not in Figure 4 for the Kastner replication ($n > 60,000$).

political scientists generate additional expectations about the relationship between X and the marginal effect of Z. In such situations, a failure to introduce these predictions and then construct a second marginal effect plot suitable for evaluating them means that researchers are ignoring valuable information relevant to testing their theory. They are, in effect, subjecting their conditional theories to substantially weaker tests than their estimation model permits. The consequence is that the literature exaggerates the empirical support for some theories and understates the support for others. Fortunately, the fix for the problem is straightforward. Researchers positing interaction between two variables should seek to generate hypotheses about how the marginal effect of each variable varies with the value of the other and construct a pair of marginal effect plots to evaluate these hypotheses.

References

Achen, Christopher. 1982. *Interpreting and Using Regression*. London: Sage.

Ai, Chunrong & Edward Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economic Letters* 80:123–129.

Aiken, Leona & Stephen West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. London: Sage Publications.

Alexseev, Mikhail A. 2006. "Ballot-Box Vigilantism: Ethnic Population Shifts and Xenophobic Voting in Post-Soviet Russia." *Political Behavior* 28:211–240.

Berry, William D., Jacqueline H. R. DeMeritt & 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.

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

Braumoeller, Bear. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58:807–820.

Clark, William & Matt Golder. 2006. "Rehabilitating Duverger's Theory: Testing the Mechanical and Strategic Modifying Effects of Electoral Laws." *Comparative Political Studies* 39:679–708.

Clark, William R., Michael Gilligan & Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14:63–82.

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

Kam, Cindy D. & Robert J. Franzese, Jr. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor, MI: University of Michigan Press.

Kastner, Scott L. 2007. "When Do Conflicting Political Relations Affect International Trade?" *Journal of Conflict Resolution* 51:664–688.

Mainwaring, Scott. 1993. "The Presidentialism, Multipartism, and Democracy: The Difficult Combination." *Comparative Political Studies* 26:198–228.

Norton, Edward, Hua Wang & Chunrong Ai. 2004. "Computing Interaction Effects and Standard Errors in Logit and Probit Models." *Stata Journal* 4:103–116.

Wright, Gerald. 1976. "Linear Models for Evaluating Conditional Relationships." *American Journal of Political Science* 2:349–373.

Appendix: When One Variable is Expected to have *No* Effect at an Extreme Value of the Other

We have advised scholars who propose an interactive theory specified in the form of Eq. (1) to use the theory to generate as many of the five key predictions listed in our paper as the theory allows. Four of these predictions relate to the marginal effect of one independent variable at the lowest or highest value of the other. In this appendix, we caution that when one's theory posits that one independent variable has *no* effect at all (i.e., a marginal effect of zero) when the other independent variable is at one of its extremes, one should think very carefully about whether the functional form of Eq. (1) is appropriate.

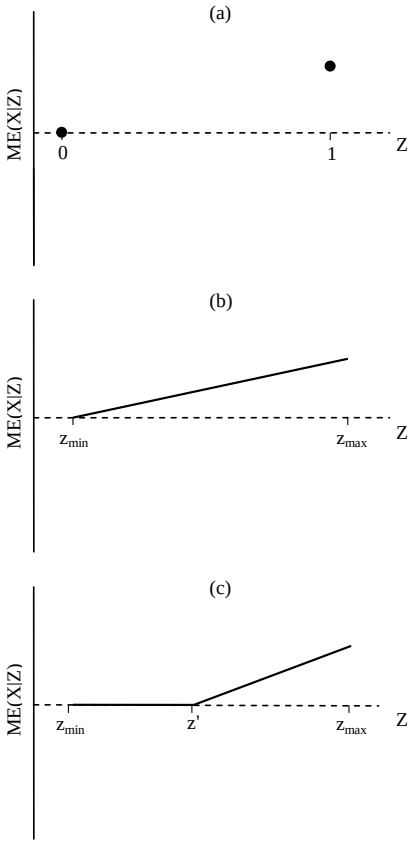
To illustrate why, we will assume that predictions $P_{X|Z_{\min}}$ and $P_{X|Z_{\max}}$ take the following form:

- $\mathbf{P_1^*}$: The marginal effect of X is zero when Z is at its lowest value.
- $\mathbf{P_2^*}$: The marginal effect of X is positive when Z is at its highest value.

For the theory generating these predictions to be accurately specified by Eq. (1) – in which each independent variable is assumed to be linearly related to the marginal effect of the other – *Z's lowest value* must be the *only* value of Z at which X has no effect on Y. There are two ways this could happen. First, Z could be dichotomous (0 or 1) and X could have no effect on Y when $Z = 0$ but a positive effect when $Z = 1$. This situation is depicted in Figure 6a. The second possibility is that Z is continuous and ME(X|Z) increases linearly with Z when $Z > z_{\min}$. This situation is depicted in Figure 6b.

But consider the relationship between Z and the marginal effect of X shown in Figure 6c. Here the value of Z must surpass some threshold, z', for X to have any effect on Y, but once this threshold is achieved, ME(X|Z) grows linearly with Z. Conditional theories that posit some kind of threshold effect similar to that shown in Figure 6c are relatively common in political science (Clark, Gilligan & Golder 2006). For example, Duverger's theory predicts that social heterogeneity increases party system size, but only once the electoral system is sufficiently permissive (Clark & Golder 2006). Similarly, Mainwaring (1993) argues that presidentialism is bad for democratic survival, but only if legislative fragmentation is sufficiently high. The important thing to note is that although the threshold relationship shown in Figure 6c is fully consistent with predictions P_1^* and P_2^*, it is *not* accurately captured by the linear-interactive specification of Eq. (1). This is because the relationship between Z and ME(X|Z) is only piece-wise linear; it is not linear over the entire

Figure 6: Marginal Effect Plots Indicating that X has No Effect When Z is at its Minimum Value



Notes: $z_{\min}$ and $z_{\max}$ indicate the lowest and highest observed values of Z.

range for Z.[1] In this type of situation, an alternative strategy for model specification and testing is needed.

If one's theory generates an *a priori* prediction about the value of the threshold, z', then the predicted value of z' can be used to split the sample into two subsamples. One could then estimate the interactive model specified in Eq. (1) separately in the subsample of observations for which $Z \leq z$' and in the subsample of observations for which $Z \geq z$'.[2] One would predict that in the context in which Z is low, ME($X|Z = z_{\min}$),

[1] Furthermore, note that if X and Z interact as in Figure 6c, the marginal effect of Z on Y is conditional not only on the value of X – as in the interactive model in Eq. (1) – but on the value of Z as well. It is evident from Figure 6c that when $Z \leq z'$, the marginal effect of X is the same regardless of the value of Z; put differently, Z and X are additive in their effects on Y. The symmetry of interaction implies that in this range for Z, the marginal effect of Z is unrelated to the value of X. Figure 6c indicates that when $Z \geq z'$, the marginal effect of X is positively related to Z. The symmetry of interaction implies that in this range for Z, the marginal effect of Z is positively related to X.

[2] An alternative strategy would be to conduct a full-sample estimation of a model specifying three-way interaction among X, Z,

ME($X|Z = z$'), and β_{XZ} are all zero. And one would predict that in the context in which Z is high, ME($X|Z = z$') is zero but both ME($X|Z = z_{max}$) and β_{XZ} are positive.

In the more likely situation in which one's theory is not strong enough to identify the specific value of the threshold, z', the options are less satisfactory. If one is confident, a priori, that the threshold is much closer to z_{min} than to z_{max}, one might reasonably view Eq. (1) as a sufficiently close approximation of the true model to warrant a reliance on this equation for empirical analysis. If one has no expectation about the value of the threshold, one might conduct split-sample estimations of Eq. (1) multiple times, varying the assumed value of the threshold, and then determine the "correct" threshold by comparing the fits of the various models. Still another option would be to approximate the expected functional form with a quadratic specification that assumes that the marginal effect of X changes less abruptly than in Figure 6c, thereby eliminating the need to identify a threshold value for Z altogether. One example of a quadratic specification that provides a reasonably good fit to the functional form shown in Figure 6c is:

$$Y = \beta_0 + \beta_X X + \beta_Z Z + \beta_{XZ} XZ + \beta_{XZ^2} XZ^2 + \epsilon. \tag{7}$$

The marginal effect of X in this interactive model is a nonlinear function of Z:

$$ME(X|Z) = \frac{\partial Y}{\partial X} = \beta_X + \beta_{XZ} Z + \beta_{XZ^2} Z^2. \tag{8}$$

Note that in Eq. (7), the marginal effect of Z is now determined by both X and Z:

$$ME(Z|X) = \frac{\partial Y}{\partial Z} = \beta_Z + (\beta_{XZ} + 2Z\beta_{XZ^2})X. \tag{9}$$

Put differently, the marginal effect of Z is a linear function of X with a different slope at each value of Z.

and a dichotomous variable, D, that equals 1 when $Z > z$' and 0 otherwise. In particular, Y would be regressed on X, Z, D, XZ, XD, ZD, and XZD. This estimation would yield point estimates of marginal effects identical to those obtained using the split-sample approach but the standard errors may be different (Kam & Franzese 2007, 103-111).