Strategic Misspecification in Regression Models

Curtis S. Signorino   University of Rochester
Kuzey Yilmaz   Koc University

Common regression models are often structurally inconsistent with strategic interaction. We demonstrate that this "strategic misspecification" is really an issue of structural (or functional form) misspecification. The misspecification can be equivalently written as a form of omitted variable bias, where the omitted variables are nonlinear terms arising from the players' expected utility calculations and often from data aggregation. We characterize the extent of the specification error in terms of model parameters and the data and show that typical regressions models can at times give exactly the opposite inferences versus the true strategic data-generating process. Researchers are recommended to pay closer attention to their theoretical models, the implications of those models concerning their statistical models, and vice versa.

Much of political science research assumes individuals and groups of individuals (e.g., in the form of states) behave strategically. Recent research (Signorino 1999, 2000; Smith 1999) suggests that, when analyzing strategic behavior on the part of individuals or states, failure to reflect that strategic interaction in one's statistical model can result in invalid inferences. Signorino (1999) demonstrates this with a Monte Carlo example in which the inferences from logit regressions are far from (at times completely opposite to) the strategic data-generating process. Signorino (1999), however, is not a complete analysis of the misspecification, but more a warning and demonstration that the misspecification exists. As of yet, the form of the misspecification has not been characterized in a way that most practitioners readily understand or in a manner that allows us to state when the effects of strategic misspecification should be mild versus severe. The primary goal of this article is to do exactly that.

As we demonstrate in this article, strategic misspecification is really an issue of structural (or functional form) misspecification. Because this type of misspecification is not well known among political scientists, we begin in the next section by illustrating the problem of functional form misspecification as it applies to the classical linear regression model. It is easy to show in this case that functional form misspecification is actually a type of omitted variable bias, where the omitted variables are nonlinear terms in a Taylor series approximation of the true functional form.

We then move to the strategic setting and construct what we believe is the simplest model possible: a two-player deterrence game. We assume the data in this case represents whether a particular outcome (war) has occurred or not, and we analyze the misspecification of using logit or probit with the ubiquitous linear XB specification of the latent variable equation. As in the OLS case, it is easy to rewrite the strategic misspecification as a form of omitted variable bias, where the omitted variables are nonlinear terms in a Taylor series expansion of the true functional form and where the nonlinearity is due to the players' expected utility calculations. Because of the misspecification, parameter estimates are not only biased, but inconsistent. Therefore, throwing more data at the problem does not make it go away.

In the third section, we analyze the misspecification when the dependent variable is the action taken by the first player, the attacker. By construction, this choice is

monotonically related to each independent variable. This would seem to be an ideal situation for using logit or probit. In this case, we not only demonstrate that misspecification exists, but also mathematically characterize the specification error in terms of the model parameters and data, showing when it will be better or worse. Interestingly, there are a wide range of very reasonable conditions in which logit or probit will incorrectly indicate that variables in the model have no effect or even the opposite effect of their true values. A Monte Carlo example is provided to illustrate the analytical results.

Although we frame our analysis in terms of strategic interaction and discrete-choice models, our argument is not limited to either, but applies to all parametric estimation in the social sciences. In general, when the functional form of the statistical model is not consistent with the data-generating process, misspecification results. Our analysis therefore emphasizes that more attention should be paid to the functional relationship of the dependent and independent variables in one's theory—and that the lack of a strongly-specified theory does not absolve one of this critical step. The theory, the hypotheses to be tested, and the statistical model all must reflect each other. If they are not identical functional specifications, they should at least be consistent with each other in terms of monotonicity conditions, which we discuss at length in the fourth section.

The Basic Problem Illustrated with the Classical Linear Regression Model

The general problem of functional form misspecification has been well known for quite awhile within the econometrics community, at least in terms of the classical regression model and systems of linear equations. However, it has received relatively little attention within political science. Because most political science scholars are now familiar with the classical linear regression model, we illustrate the basic functional form problem in this section using that model and move to the strategic setting thereafter.[1]

To begin, let us assume that we are analyzing the relationship of a single regressor X to a dependent variable Y, and that this relationship satisfies all the assumptions of the classical linear model concerning Y, X, and error

[1]The illustration in this section is largely based on Kmenta (1986, 449–50).

term ϵ, with the exception that Y is a nonlinear function of X and a parameter β:[2]

$$Y = f(X, \beta) + \epsilon. \tag{1}$$

In analyzing this data, the typical political science scholar might test for heteroskedasticity, autocorrelation, or any number of other deviations from the classical linear model. However, invariably she would assume that

$$Y = B_{0L} + B_{1L} X + \epsilon, \tag{2}$$

i.e., that regression model is a linear function of parameters $\boldsymbol{B}_L$, where we use the L subscript to denote the linear model. More often than not, she would also assume that the regression model is a *first-order* function of some set of substantive explanatory variables $\boldsymbol{X}$.[3] If $f(X, \beta)$ is nonlinear, but the analyst instead employs $\boldsymbol{X}\boldsymbol{B}_L$ (linear in parameters $\boldsymbol{B}_L$ and first-order in $\boldsymbol{X}$), then she has clearly misspecified the functional form of $f(X, \beta)$. The question is: Does that matter?

To analyze the misspecification, let us first take the Taylor series expansion of $f(X, \beta)$ about $\overline{X}$. The resulting Taylor series version of the model will approximate the original nonlinear model, but can be written as linear in the parameters, which then fits the classical linear framework. To simplify notation, we will denote the pth order Taylor series expansion of $f(X, \beta)$ about $\overline{X}$ as $f_T(X, \beta)$. The Taylor series expansion is

$$f_T(X, \beta) = f(\overline{X}, \beta) + (X - \overline{X}) f'(\overline{X}, \beta) + \frac{(X - \overline{X})^2}{2!}$$

$$\times f''(\overline{X}, \beta) + \cdots + \frac{(X - \overline{X})^p}{p!} f^{(p)}(\overline{X}, \beta) + \nu_{p+1},$$

where $f^{(p)}(\overline{X}, \beta)$ is the p-th derivative of $f(X, \beta)$ with respect to X, evaluated at $\overline{X}$. The remainder, ν_{p+1}, can be thought of as the difference between $f(X, \beta)$ and $f_T(X, \beta)$, or equivalently as the Taylor series terms from $p + 1$ to infinity not incorporated in the p-th order expansion. Assuming that $\lim_{p \to \infty} \nu_p = 0$, higher-order

[2]Throughout this article, single variables will be denoted by italicized, and usually uppercase, letters. Bolded letters will represent multiple variables or parameters. Additionally, all utilities, probabilities, and variables have an implicit observation index, which for notational convenience, we will tend to omit.

[3]First-order terms are those including only a single variable X_j raised to the first power. For example, a typical first-order linear specification of $\boldsymbol{X}\boldsymbol{B}_L$ would be $B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_k X_k$. Higher-order terms include quadratic, cubic, and interaction terms. The classical linear model can accommodate higher-order terms as regressors, so long as the model remains linear in the parameters $\boldsymbol{B}_L$. Our general argument does not depend on assuming that $\boldsymbol{X}\boldsymbol{B}_L$ contains only first-order X terms. However, it is by far the most commonly used, and it is mathematically convenient.

Taylor series expansions will better approximate the original function.

Assume that p is large enough that ν_{p+1} is negligible. If we replace the true functional form $f(X, \beta)$ in the regression model with the p-th order Taylor series approximation of it $f_T(X, \beta)$, and rearrange terms, we can write the regression model as a linear function of its parameters:

$$Y = B_0 + B_1 X + B_2 X^2 + B_3 X^3 + \cdots + B_p X^p + \epsilon, \tag{3}$$

where

$$B_k = \sum_{m=k}^{\infty} \binom{m}{k} \frac{(-\overline{X})^{m-k}}{m!} f^{(m)}(\overline{X}, \beta). \tag{4}$$

In this case, the Taylor series model (Equation 3) *is* the true model (Equation 1)—it is just written in a different way.

So what is the problem if the researcher employs the typical first-order linear XB_L regression in Equation 2? The true model, Equation 3, shows that the higher-order terms (X^2, X^3, etc.) are relevant to the relationship between Y and X—they capture the nonlinearity in $f(X, \beta)$. However, the specification in Equation 2 omits the higher-order terms. Therefore in this case, *functional form misspecification is equivalent to omitting relevant variables, where the omitted variables represent the nonlinearities in the relationship between the dependent and independent variables.* Based on what we know about omitted variable bias, we can conclude that the estimate of B_{1L} will be biased to the extent that the higher-order polynomials of X are correlated with X. In the unlikely event that the omitted and included variables are uncorrelated, model fit will still suffer to the extent that the omitted variables have a larger effect on Y. Clearly, the more nonlinear the functional form is, the more the higher-order terms will matter, and thus, the greater the specification error is likely to be.

We have demonstrated the functional form problem when Y is a function of only a single variable. The results are similar (actually worse) when Y is a nonlinear function of multiple independent variables. In that case, the higher-order terms in the Taylor series expansion will consist of higher-order polynomials of the independent variables, as well as higher-order polynomials of their interactions.

Strategic Misspecification as Omitted Variable Bias

The problem of strategic misspecification in parametric models is precisely a problem of functional form misspecification, and its proof is analogous to that just demonstrated with the classical linear model. In this article, we address strategic misspecification in the context of binary choice models, since binary data is so widely used throughout political science. However, the concepts are exactly the same for any binomial, multinomial, or continuous-variable regression model. In this section, we first describe the general latent variable framework for binary data and then introduce a simple strategic model. Using that model, we demonstrate that strategic misspecification is equivalent to omitting relevant variables.

The Latent Variable Specification

Let us assume there exists a latent (i.e., unobservable) variable y^* defined by

$$y^* = f(X, \beta) + \epsilon, \tag{5}$$

where $f(X, \beta)$ is a function of regressors X and parameters β, and ϵ is a random disturbance from some density $h_\epsilon(\epsilon)$ with mean zero. Examples of such latent variables might be a state's utility for war versus peace, a senator's utility for a particular bill in Congress, or the extent to which a shopper prefers apples instead of oranges. Although y^* is unobservable directly, we often can observe whether y^* is above some threshold. For example, we can observe whether a state chooses to go to war, whether a senator votes for a bill, or whether a shopper purchases an apple or an orange. We denote the observable choice as y, which will be our data, and for simplicity we set the threshold to zero:

$$y = \begin{cases} 1 & \text{if } y^* > 0 \\ 0 & \text{if } y^* \leq 0. \end{cases} \tag{6}$$

A particular probability model results from the specification of $f(X, \beta)$ and ϵ.

Unfortunately, the researcher is not omniscient and must therefore make an assumption concerning all of these elements. Of most importance to us is the assumption concerning $f(X, \beta)$. As with the classical linear model, the most common practice is to assume that the functional form is both linear in the parameters and a first-order function of the substantive variables, which we will denote as XB_L. Hereafter when we refer to the "typical," "common," or "traditional" binary specification, it is precisely to this first-order linear latent variable specification to which we will refer, and we will use an L subscript to denote parameters and equations associated

with it.[4] ϵ is generally assumed to be distributed logistic (for logit) or Normal (for probit).

Some mistakenly believe that for logit and probit a first-order XB_L relationship is somehow a "general" or theory-free specification that "allows the data to speak for itself." On the contrary, the first-order XB_L specification is a very specific structural relationship—a straightjacket, if you will—that constrains our analysis and affects our inferences. We are not implying that the first-order XB_L specification is alone in this. Indeed, in the context of parametric models, any assumption concerning $f(X, \beta)$ is a structural assumption. Once specified, the assumed form of $f(X, \beta)$ *imposes* a relationship on the regressors, parameters, and the latent dependent variable y^*. It is not an empirical relationship that we find through data analysis, but a theoretical or modeling relationship in the context of which we conduct our analysis.[5] Nevertheless, for too long the first-order linear XB_L relationship has been taken for granted.

Historically there may not have been valid theoretical reasons for assuming otherwise. However, recent work (Signorino 1999, 2000) suggests that when the data-generating process is the result of strategic behavior, $f(X, \beta)$ will most likely be nonlinear. Moreover, the structure of each "game" implies a particular "strategic" functional form. Given that many of our theories in political science assume strategic behavior and given the prevalence of logit and probit with the XB_L specification, we would like to know how bad the misspecification is.

A Very Simple Strategic Model

To analyze statistical misspecification, we usually start with the simplest possible model and examine how the misspecification affects the results in that situation. More extensive or general results can be derived later if time, space, and mathematical tractability allow.

Consider the simple strategic situation depicted in Figure 1(a), which is typical of deterrence situations, whether in international politics, congress, or market entry. Here, $\{\overline{A}, A, \overline{R}, R\}$ are actions in the game and $\{SQ, Cap, War\}$ are outcomes. Player 1, the attacker, must choose between attacking A and not attacking $\overline{A}$. If she

does not attack $\overline{A}$, then the game ends with the status quo (SQ) as the outcome. If she attacks, then player 2, the defender, must choose between resisting R and not resisting $\overline{R}$, leading to outcomes War and capitulation (Cap), respectively. For each outcome, the observable component of the player's utility is denoted by $U_i(k)$, where i indexes the player and k refers to the outcome. The game in Figure 1 is only partially strategic—only player 1 must condition her behavior on player 2's expected behavior. This is reflected in the fact that there is no need to specify $U_2(SQ)$.

Throughout this article, we will assume the source of uncertainty—i.e., what makes the strategic model probabilistic—is based on agent error (see McKelvey and Palfrey 1998; Signorino 1999, 2000). This assumption is implemented simply for mathematical convenience. The reader who does not find the behavioral assumptions underlying the agent-error model particularly palatable can refer to Signorino (2000) for two other types of statistical (but strictly Nash) strategic models: one based on incomplete information concerning outcome payoffs; and another based on variation in the regressors, which is unobserved only by the analyst. The issues addressed here are not limited to the agent error variant. Indeed, because the primary question addresses the extent to which strategic misspecification affects inferences in binary choice models, the general conclusions of our analysis in this article apply to the payoff perturbation and unobserved regressor variation models as well.

In the current context, player i's true utility for action j is denoted by $U_i^*(j)$ and is divided into two components: an expected utility component $U_i(j)$ that is observable by everyone (including the analyst) and a random component α_j that is observed only by player i. Referring still to Figure 1(a) above, if player 1 attacks, then player 2's utilities for not resisting $\overline{R}$ and resisting R are

$$U_2^*(\overline{R}) = U_2(\overline{R}) + \alpha_{\bar{r}}$$
$$= U_2(Cap) + \alpha_{\bar{r}}, \qquad (7)$$
$$U_2^*(R) = U_2(R) + \alpha_r$$
$$= U_2(War) + \alpha_r. \qquad (8)$$

As we noted before, player 2's calculations are nonstrategic because she does not condition on the attacker's behavior. In contrast, player 1's utilities for actions $\overline{A}$ and A are

$$U_1^*(\overline{A}) = U_1(\overline{A}) + \alpha_{\bar{a}}$$
$$= U_1(SQ) + \alpha_{\bar{a}}, \qquad (9)$$
$$U_1^*(A) = U_1(A) + \alpha_a$$
$$= p_{\bar{r}} U_1(\overline{R}) + p_r U_1(R) + \alpha_a$$
$$= p_{\bar{r}} U_1(Cap) + p_r U_1(War) + \alpha_a. \qquad (10)$$

---

[4]Again, quadratic or cubic terms sometimes appear in published logit or probit regressions. However, the first-order XB specification is far more common in political science research, and, because of its simplicity, it makes the analysis of the specification error much easier. As in the OLS case of the previous section, our general argument does not require this exact specification. Indeed, it does not require any particular specification of $f(X, \beta)$—only that it be misspecified.

[5]See Dubin and Rivers (1989) for a similar statement concerning linear regression.

FIGURE 1   A Very Simple Strategic Model.

The simplest strategic model consists of one actor conditioning her decision on that of another player. Here, player 1 must decide whether to attack (A) or not attack ($\bar{A}$) player 2. If player 1 does not attack, then the status quo (SQ) results. If player 1 attacks, then player 2 must choose between resisting (R) or not resisting ($\bar{R}$), leading to (War) and capitulation (Cap), respectively. Figure 1(a) depicts the game in its more general form. Figure 1(b) shows an example where the individual choice probabilities are all monotonically related to the variables that comprise the utilities.

(a) Strategic deterrence model

(b) Utilities specified with regressors
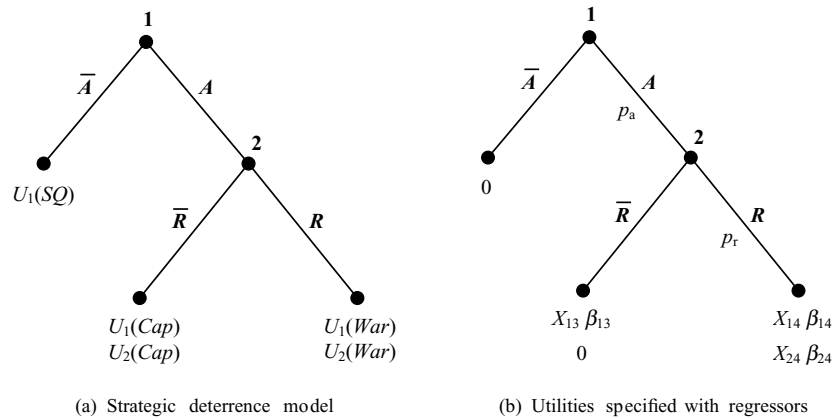
Player 1's calculations *are* strategic. Player 1's observable expected utility for attacking depends on the probability that player 2 resist or not. Let p_j be probability that action j is chosen and p_k be the probability that outcome k is realized. Assuming the α_j are independent and identically distributed according to some density $f(\alpha)$ with finite expectation, then the general form of the equilibrium probabilities is[6]

$$p_a = 1 - p_{\bar{a}} = p_r\big[U_1^*(A) > U_1^*(\bar{A})\big],$$

$$p_r = 1 - p_{\bar{r}} = p_r\big[U_2^*(R) > U_2^*(\bar{R})\big],$$

$$p_{sq} = p_{\bar{a}},$$

$$p_{cap} = p_a\, p_{\bar{r}},$$

$$p_{war} = p_a\, p_r.$$

With an appropriately specified $f(\alpha)$, the above probabilities can be derived and used in data analysis (e.g., MLE). For example, if the α_j are i.i.d. $N(0, \sigma^2)$, then the action probabilities will be Normally distributed. If they

are i.i.d. Type 1 Extreme Value, then the resulting action probabilities will be logistic.

For regression analysis, we must also specify the utilities of Figure 1(a) in terms of regressors. Consider Figure 1(b). Here, the utilities—and, hence, all equilibrium probabilities—are a function of only three regressors: X_{13}, X_{14}, and X_{24}. We have constructed the utilities (1) to make them as simple as possible and (2) to try to ensure wherever possible the monotonicity of the probabilities as a function of the regressors. Although we will discuss this in more detail in a later section, it is important to note here that *all* of the action probabilities ($p_{\bar{a}}$, p_a, $p_{\bar{r}}$, p_r) are monotonic in each of the regressors.

Finally, researchers are often constrained to work with particular forms of dependent variables at their disposal. The best case here would be to have the actual outcome data available for analysis (i.e., whether SQ, Cap, or War occurred in a given observation). However, data might only be available for, say, whether player 1 chose $\bar{A}$ vs. A. Using the appropriate equilibrium probabilities in maximum-likelihood estimation would allow for the analysis of either of these forms of data. Because of the prevalence of both forms of data in political science

[6]For examples of how to derive such probabilities, see Signorino (1999, 2000). Derivations of results in this article are also available upon request from the authors.

research, we analyze the specification error for the model outcomes in this section and for the attacker's actions in the third section.

War Data: Aggregating the Strategic Model's Outcomes

Assume now that the true data-generating process is the deterrence model shown in Figure 1(b). Numerous analyses in international relations have employed logit or probit to analyze data where the dependent variable denotes simply "War" vs. "Not War." Indeed, this type of data, where at least one of the categories aggregates multiple outcomes in some "original" model or process, is actually quite common in political science more broadly. In the context of the deterrence model, the "Not War" category would include both the status quo and capitulation outcomes (*SQ* and *Cap*, respectively).
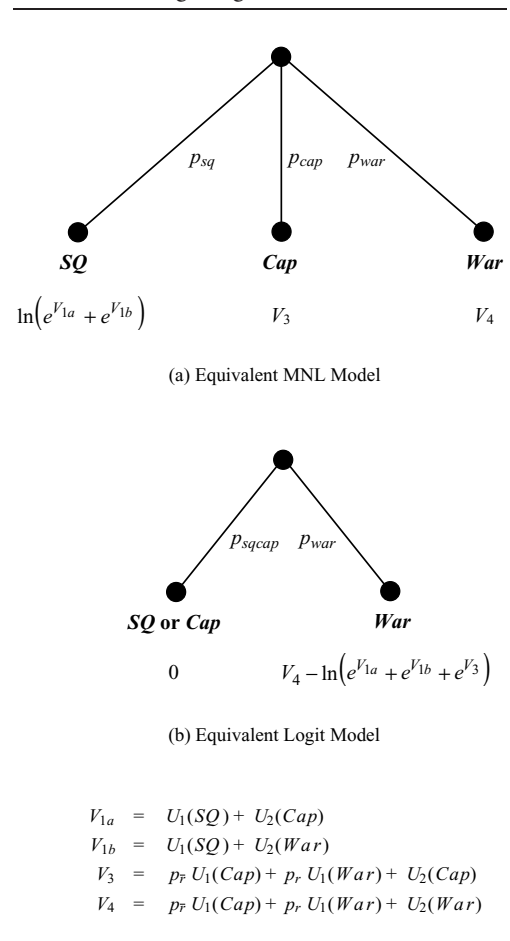
Notice that we have three outcomes in the strategic model (*SQ*, *Cap*, *War*), but only two outcomes (*War*, *Not War*) in the available data. Before we can assess the misspecification of a typical logit or probit model in this context, we need to transform the strategic model in Figure 1 into an equivalent binary model—one where the probabilities over the aggregated outcomes are consistent with the strategic model.

We illustrate this transformation using Figure 2. Figure 2 shows a progression of models in which the probabilities are consistent with the underlying strategic model in Figure 1, here assuming α is distributed Type 1 Extreme Value. Figure 2(a) displays the multinomial logit model with the same number of outcomes and the same outcome probabilities as the strategic model. In some ways it looks strange — and it should! We have taken a two-player strategic model and aggregated it into a single decisionmaker model, where the dyad appears to be the "decisionmaker." Nevertheless, the single decisionmaker model in Figure 2(a), with the accompanying specification of the utilities, is equivalent to the strategic model in terms of the outcome probabilities.

Now consider a dependent variable y_S that is coded $y_S = 1$ if *War* occurred and $y_S = 0$ if either *SQ* or *Cap* occurred. If we aggregate the *SQ* and *Cap* outcomes in Figure 2(a), the resulting binary model can be expressed as in Figure 2(b). Again, even though Figure 2(b) is a binary model, the probabilities are entirely consistent with (i.e., derived from) the strategic data-generating process. We will use the binary model in Figure 2(b) as our referent model for the remainder of this section.

FIGURE 2 **Multinomial and Binomial Equivalents of the Strategic Model.**
Figure (2a) is a multinomial logit model that produces outcome probabilities that are equivalent to the strategic model's. Figure (2b) is the equivalent binomial model and will be used as the referent model for comparison with the typical logit regression. The V_k terms are based on the strategic model's utilities. Notice that the resulting binary model looks nothing like a typical logit regression.

$$\ln\left(e^{V_{1a}} + e^{V_{1b}}\right) \qquad V_3 \qquad V_4$$

(a) Equivalent MNL Model

$$0 \qquad V_4 - \ln\left(e^{V_{1a}} + e^{V_{1b}} + e^{V_3}\right)$$

(b) Equivalent Logit Model

$$
\begin{aligned}
V_{1a} &= U_1(SQ) + U_2(Cap) \\
V_{1b} &= U_1(SQ) + U_2(War) \\
V_3 &= p_{\bar{r}}\, U_1(Cap) + p_r\, U_1(War) + U_2(Cap) \\
V_4 &= p_{\bar{r}}\, U_1(Cap) + p_r\, U_1(War) + U_2(War)
\end{aligned}
$$

Before proceeding, we should point out that many recent studies use multinomial logit or probit to model the strategic decisions of dyads in crises. These studies assume the dyad makes a "decision" among a number of options, including war. However, no consideration is given

in these studies to how one would aggregate an underlying multiplayer strategic model into a dyad-as-player model. The above exercise of aggregating the strategic model into equivalent multinomial and binomial logit models illustrates two important points. First, the specific functional form of the aggregated models depends on the specification of the original strategic model (e.g., the number of players, their sequence of choices, and their utilities) and on which actions or outcomes are aggregated. Second, the resulting specifications of the aggregated multinomial and binomial logit models look nothing like the specifications in typical multinomial and binomial logit analyses. In fact, unless the researcher can prove otherwise, one should generally expect such aggregation of strategic behavior to be *nonlinear* in both the regressors and parameters. The fact that past studies using multinomial logit and probit all employ traditional XB_L specifications for the "utilities" should be a red flag concerning the effects of strategic misspecification on the conclusions reached in those studies.

Specification Error

We can now express the model in Figure 2(b) with its equivalent latent variable representation, and then use that in our analysis of specification error. Figure 3 displays three latent variable specifications for y^*. The first, y_S^*, displays the latent variable equation implied by the referent strategic model, having substituted the utility specification of Figure 1(b). At the expense of repeating the obvious, y_S^* represents the true strategic data generating process. If one believed the strategic model generated the data and only had this aggregated binary data on hand, then y_S^* is the binary model one should employ.

The regression equation for y_S^* is clearly not a typical logit regression, which is denoted by y_L^* in Figure 3. The action probabilities $p_{\bar{r}}$ and p_r are nonlinear functions of X_{24} and β_{24}. Similarly, the R_S term is a nonlinear function of all of the regressors and coefficients.

To demonstrate the misspecification more clearly, we replace $p_{\bar{r}}$, p_r, and R_S with Taylor series approximations, here second-order and taken about zero. The resulting regression equation is denoted by y_T^* in Figure 3.[7] The cross-product and higher-order terms, along with the second-order Taylor series remainder ν_3, are collected into the R_T term to help in comparing the models. Note that, as spec-

ified, the Taylor model y_T^* is equivalent to the strategic model y_S^*—it is simply written in a different way.

Comparing the Taylor and logit regressions, y_T^* and y_L^*, respectively, we see that both contain a constant and first-order terms (i.e., X_{13}, X_{14}, and X_{24}). The most important difference, however, is that the Taylor regression contains the R_T term, which includes the second-order and higher effects, whereas the typical logit regression does not. If the strategic model generated the data, then the second-order and higher terms in R_T are relevant variables—and the logit regression omits these variables. Hence, *the typical logit regression will induce omitted variable bias in the estimated parameters of the included variables.*

The logit equation y_L^* is essentially a first-order Taylor expansion, minus a Taylor expansion remainder ν_2. It is an attempt to capture the linear effects of the variables. However, if the data is generated by a nonlinear (e.g., strategic) process, then the Taylor remainder ν_2 becomes a relevant variable, which is omitted from the equation. The resulting omitted variable bias means that not even the linear effects are correctly estimated. In general, the greater the nonlinearity implied by the strategic model, the larger the Taylor series remainder ν_2 and the greater the omitted variable bias in the typical logit regression.

One might be tempted to look for salvation in a lack of correlation between the included and excluded regressors. However, the excluded higher-order terms are all functions of the included first-order terms, and therefore some correlation will likely be present. In fact, as Yatchew and Griliches (1985) demonstrate for probit, although correlation of the included and excluded variables will exacerbate the omitted variable bias, it is not necessary for bias to result. It is, therefore, unlikely that any of the effects will be correctly estimated by logit even for this simplest of models. Given this, any inferences based on the estimated parameters or on the predicted probability of *War* will almost certainly be invalid.

Characterizing the Specification Error in the (Almost) Ideal Case

We have just demonstrated that strategic misspecification in discrete choice models is equivalent to omitted variable bias. We did so based on a binary aggregation of the outcomes. The strategic equation in Figure 3 shows that the relationship between the dependent variable and the regressors in that case is nonlinear. In fact, it is actually nonmonotonic as well. Some might therefore

---

[7]We could have also taken the Taylor expansion of the entire equation y_S^*. However, it was mathematically more convenient to take the Taylor expansion of only $p_{\bar{r}}$ and p_r. The general result is the same either way.

FIGURE 3   **Strategic, Taylor, and Logit Models of War.**

The figure summarizes the three versions of y^*, representing the likelihood that the attacker and defender will go to war with each other. The first equation is the strategic model, derived from the data-generating process. The second equation is the Taylor series approximation of the strategic model. The third equation is the model commonly estimated to test hypotheses concerning X_{13}, X_{14}, and X_{24}. R_S is a nonlinear term that results from aggregating the original strategic model into a binary model. R_T is the higher-order remainder for the Taylor series approximation. Notice that the commonly estimated Logit model omits these relevant terms.

Data-Generating Process

Strategic $\quad y_S^* \quad = \quad\quad\quad\quad\quad p_{\bar{r}}\,\beta_{13}\,X_{13} \;+\; p_r\,\beta_{14}\,X_{14} \;+\; \beta_{24}\,X_{24} \;+\; R_S \;+\; \epsilon$

Taylor $\quad\quad y_T^* \quad = \quad -\ln(3) \quad + \quad \frac{1}{3}\beta_{13}\,X_{13} \;+\; \frac{1}{3}\beta_{14}\,X_{14} \;+\; \frac{2}{3}\beta_{24}\,X_{24} \;+\; R_T \;+\; \epsilon$

Commonly Estimated Model

Logit $\quad\quad\; y_L^* \quad = \quad B_{0L} \quad + \quad B_{13L}\,X_{13} \;+\; B_{14L}\,X_{14} \;+\; B_{24L}\,X_{24} \quad\quad\quad\quad + \;\; \epsilon$

$$R_S \;=\; -\ln\left[1 + e^{X_{24}\,\beta_{24}} + e^{p_{\bar{r}}\,X_{13}\,\beta_{13} + p_r\,X_{14}\,\beta_{14}}\right]$$

$$R_T \;=\; -\frac{1}{36}\,X_{13}^2\,\beta_{13}^2 - \frac{1}{36}\,X_{14}^2\,\beta_{14}^2 - \frac{1}{9}\,X_{24}^2\,\beta_{24}^2$$
$$\quad\; -\frac{1}{18}\,X_{13}\,X_{14}\,\beta_{13}\,\beta_{14} - \frac{1}{9}\,X_{13}\,X_{24}\,\beta_{13}\,\beta_{24} + \frac{2}{9}\,X_{14}\,X_{24}\,\beta_{14}\,\beta_{24} + \nu_3$$

argue that we have stacked the deck against typical logit or probit models and that more sophisticated researchers would only use logit or probit to analyze data where they believed the dependent and independent variables were monotonically related.[8] Although we would argue (and have argued) that countless scholars have implicitly conducted analyses where the misspecification is similar to that in the previous section, in this section we examine whether the typical logit or probit model is misspecified in the "ideal" case—when the strategic binary dependent variable is monotonically related to all regressors.

Specification Error

Consider again the simple strategic model in Figure 1(b). Suppose now that we do not have data on the defender's

---

[8]Roughly speaking, Y is a monotonically increasing (decreasing) function of X if, holding all other variables constant, Y always increases (decreases) as X increases. Y and X have a nonmonotonic relationship if, holding all other variables constant, as X increases, Y sometimes increases and sometimes decreases.

actions, but only on whether the potential attacker decided to attack or not (A vs $\bar{A}$). We might want to examine the relationship of deterrence success to substantive explanatory variables that we believe affect the incentives of both states to engage in conflict, again represented by X_{13}, X_{14}, and X_{24}. As already noted, we constructed the model so that the attacker's choice probabilities $p_{\bar{a}}$ and p_a are monotonically related to each of the regressors.

To analyze whether the typical logit or probit model is misspecified in this situation, we again recast the model in an equivalent latent variable form. Suppose our observable dependent variable is coded as $y = 1$ if player 1 attacks (A) and $y = 0$ if player 1 does not attack($\bar{A}$). Recall that player 1 will attack if $U_1^*(A) > U_1^*(\bar{A})$. Let $y_S^* = U_1^*(A) - U_1^*(\bar{A})$. Then we observe y as

$$y = \begin{cases} 1 & \text{if } y_S^* > 0 \\ 0 & \text{if } y_S^* \le 0 \end{cases}, \tag{11}$$

where the S subscript denotes the strategic latent variable equation. Substituting Equations 9 and 10 into y_S^* and

then the utilities from Figure 1(b), the strategic equation becomes that shown at the top of Figure 4, where $\epsilon = \alpha_a - \alpha_{\bar{a}}$.

Now consider the typical binary choice regression, y_L^*, displayed in the third equation in Figure 4. Obviously, the functional form differs between the two regression models. However, it is not clear how y_L^* and y_S^* relate to each other—e.g., how the estimators of β_{ij} and B_{ijL} relate to each other or how the regressions differ in their predicted probabilities. The traditional y_L^* is linear in the explanatory variables and coefficients. In contrast, $X_{24}\beta_{24}$ enters y_S^* through $p_{\bar{r}}$ and p_r as part of the expected utility calculation. Although it appears that we have a functional form misspecification, it is not obvious how bad it is. Rather than simply showing that the structural misspecification is equivalent to omitted variable bias, we will instead characterize it in a slightly different way, so we can assess just how bad the specification error will be under different conditions.

Since the main problem in comparing the two models is the nonlinear $p_{\bar{r}}$ and p_r terms in y_S^*, we use a first-order Taylor series expansion of $p_{\bar{r}}$ and p_r about the mean of X_{24}. Let m_{ij} be the mean of variable X_{ij}. Then we can write $X_{ij} = m_{ij} + u_{ij}$, where u_{ij} is the deviation of X_{ij} from its mean, with $E[u_{ij}] = 0$. Greatly abusing notation, we denote $p_{\bar{r}}$ and p_r evaluated at m_{24} by $\bar{p}_{\bar{r}}$ and $\bar{p}_r$, respectively.[9] With the Taylor expansions of p_r and $p_{\bar{r}}$, we can rewrite the strategic equation as its Taylor series equivalent, y_T^*. This is displayed as the second equation in Figure 4, with its coefficients B_{ij} and error term η rewritten at the bottom of Figure 4 in terms of the data and parameters.

To restate the obvious yet again, y_T^* represents the strategic data-generating process. It also happens to be in a form that is comparable to the first-order linear model. Therefore, we can now make a number of statements concerning the effects of estimating the first-order linear model y_L^*, when the data has been generated by our simple strategic model y_T^*.

When will strategic misspecification not be a problem? Perhaps the first question we should ask is: when will strategic misspecification *not* be a problem in our example? Consider the Taylor coefficients B_{ij} and the error term η at the bottom of Figure 4. Notice that when $\beta_{24} = 0$, the Taylor model reduces to $y_T^* = \frac{1}{2}\beta_{13}X_{13} + \frac{1}{2}\beta_{14}X_{14} + \epsilon$. This is simply a linear latent variable model of the same form as the linear logit model, so there is no misspecification. In terms of the underlying choice model, $\beta_{24} = 0$ implies that X_{24} has no effect on the attacker's cal-

culations. Since X_{24} enters the attacker's decision through its assessment of the probability that the defender will resist, this implies that the attacker does not condition her choice on the defender's expected behavior. To put it simply, there is no strategic misspecification if there is no strategic interaction.

If, on the other hand, the attacker takes into account how the defender will respond if attacked, but we estimate a typical binary choice model, then the statistical model will be structurally inconsistent with the data-generating process and some form of misspecification will exist. To assess this misspecification, we will assume $\beta_{24} \neq 0$ for the remainder of the article.

Distributional misspecification and inconsistent estimates. Consider now the "new" error term, η, in the Taylor regression. Perhaps one of the more glaring problems is that η and ϵ cannot have the exact same distribution. As Figure 4 shows, η is composed of two components: the original error term ϵ, and some function of the u_{ij}, determined by the underlying strategic model.

The first potential problem arises if the explanatory variables are correlated. Recall that correlated explanatory variables implies the u_{ij} are correlated (by definition of the u_{ij}). It is straightforward to show that if the u_{ij} are correlated with each other, then the X_{ij} will be correlated with η. Estimated parameters will therefore be biased and inconsistent.

Let us now assume the u_{ij} are independent of ϵ and of each other, then $E(\eta) = 0$ and the explanatory variables will be uncorrelated with η. Denote the variances of ϵ, u_{ij}, and η as σ_ϵ^2, σ_{ij}^2, and σ_η^2, respectively. Then

$$\sigma_\eta^2 = \bar{p}_r^2 \bar{p}_{\bar{r}}^2 \left(\beta_{13}^2 \sigma_{13}^2 + \beta_{14}^2 \sigma_{14}^2 \right) \beta_{24}^2 \sigma_{24}^2 + \sigma_\epsilon^2. \quad (12)$$

The second potential problem is that η is an additive and multiplicative function of ϵ and the u_{ij}. Because of this, there is no reason to expect that η and ϵ will even share the same density but with different variances. Estimated parameters will therefore also likely be inconsistent due to this.

Finally, because η is composed of both ϵ and the u_{ij} terms, η's variance will always be larger than ϵ's. As Equation 12 displays, the difference between η's and ϵ's variances will depend on the variance in the regressors and on the magnitude of the β_{ij} parameters. The larger the variances of the regressors and the greater the magnitude of the parameters, the greater will be the difference in the variance of η versus ϵ. In examining omitted variable bias in probit, Yatchew and Griliches (1985) note that this form of misspecification can affect hypothesis tests.

Having said this, the distributional misspecification is a matter of degree. If the u_{ij} are symmetrically distributed

---

[9]In other words, $\bar{p}_r = p_r|_{X_{24}=m_{24}}$ and $\bar{p}_{\bar{r}} = p_{\bar{r}}|_{X_{24}=m_{24}}$.

FIGURE 4 **Strategic, Taylor, and Logit Models of Deterrence Success vs. Failure.**

The figure summarizes the three versions of y^*, representing the attacker's propensity for attacking. The first equation is the strategic model, derived from the data-generating process. The second equation is the Taylor series approximation of the strategic model. The third equation is the model commonly estimated to test hypotheses concerning X_{13}, X_{14}, and X_{24}. The remaining equations give the values of the Taylor coefficients (B_0, B_{13}, B_{14}, B_{24}) and disturbance η in terms of the original parameters and data. Here, m_{ij} is the mean of explanatory variable X_{ij}, $\bar{p}_j$ is the defender's probability p_j evaluated at m_{24}, and u_{ij} is the deviation of X_{ij} from its mean m_{ij}. The Logit misspecification can be assessed based on assumptions concerning the parameters and data.

Data-Generating Process

Strategic	y_S^*	$=$			$p_{\bar{r}}\,\beta_{13}\,X_{13}$	$+$	$p_r\,\beta_{14}\,X_{14}$		$+$	ϵ
Taylor	y_T^*	$=$	B_0	$+$	$B_{13}\,X_{13}$	$+$	$B_{14}\,X_{14}$	$+$	$B_{24}\,X_{24}$	$+$ η

Commonly Estimated Model

Logit	y_L^*	$=$	B_{0L}	$+$	$B_{13L}\,X_{13}$	$+$	$B_{14L}\,X_{14}$	$+$	$B_{24L}\,X_{24}$	$+$ ϵ

$$
\begin{aligned}
B_0 &= -\bar{p}_{\bar{r}}\bar{p}_r\,(\beta_{14}m_{14} - \beta_{13}m_{13})\beta_{24}m_{24} \\
B_{13} &= \bar{p}_{\bar{r}}\,\beta_{13} \\
B_{14} &= \bar{p}_r\,\beta_{14} \\
B_{24} &= \bar{p}_{\bar{r}}\bar{p}_r\,(\beta_{14}m_{14} - \beta_{13}m_{13})\,\beta_{24} \\
\eta &= \bar{p}_{\bar{r}}\bar{p}_r\,(\beta_{14}u_{14} - \beta_{13}u_{13})\,\beta_{24}u_{24} + \epsilon
\end{aligned}
$$

and do not have a large variance relative to ϵ, then η will tend to be distributed similarly to ϵ and distributional misspecification will have less effect on inferences.

Consistent estimates, but wrong inferences. Although the preceding would seem to offer enough indictment of strategic misspecification, of more interest to us is how the structural misspecification affects our inferences when the parameters are estimated consistently—or at least close to it. For the sake of examining other aspects of strategic misspecification, we will now give the typical first-order linear specification the "benefit of the doubt" and assume that the distributional misspecification is negligible. In this case, the Taylor regression y_T^* takes the same functional and distributional form as the logit regression y_L^*. As usual, the regressor and variance parameters can only be estimated to scale. When ϵ is logistic, consistent estimation (e.g., MLE) of the logit parameters will

converge to

$$
B_{0L} = \frac{B_0}{\sqrt{\frac{3}{\pi^2}\sigma_\eta^2}} \qquad B_{13L} = \frac{B_{13}}{\sqrt{\frac{3}{\pi^2}\sigma_\eta^2}}
$$

$$
B_{14L} = \frac{B_{14}}{\sqrt{\frac{3}{\pi^2}\sigma_\eta^2}} \qquad B_{24L} = \frac{B_{24}}{\sqrt{\frac{3}{\pi^2}\sigma_\eta^2}}. \tag{13}
$$

If we now substitute the equations for the B_{ij} at the bottom of Figure 4 into the above equations, we can determine the effects of using the logit model with the deterrence data.

Interpreting the sign and magnitude of regression coefficients are two common practices in political science research. As B_{13} and B_{14} show (Figure 4), the effects of X_{13} and X_{14} will tend to be estimated correctly by the logit model, at least in terms of the direction of their effect on the attacker's decision. In contrast, inferences concerning X_{24} will depend on idiosyncracies of the data.

In particular, the estimate of B_{24L} is a factor of $\beta_{14}m_{14} - \beta_{13}m_{13}$. So, for example, if we center our data around zero, which is not uncommon, traditional logit or probit would lead us to believe that X_{24} has no effect on whether the attacker chooses to attack. By construction, we know that inference is false. Similarly, if $\beta_{14}m_{14} < \beta_{13}m_{13}$, then logit and probit would lead us to believe that X_{24} has the *opposite* effect it actually does.

Finally, because the parameter estimates in any of these choice models (strategic or nonstrategic) are difficult to interpret by themselves, analysts typically interpret the estimated probabilities for a better understanding of the relationship between the dependent variable and the regressors—not only for the direction of the relationship but the relative magnitude. We would, of course, like to know the extent to which strategic misspecification affects the estimated probabilities. Although the misspecification can be expressed in a fairly general form, its mathematical complexity does not allow for a simple (and useful) analytical expression. We therefore provide the reader with a sense of the misspecification in the following example.

Monte Carlo Example

To demonstrate that the simplifying assumptions of the Taylor series approximation are not unreasonable and to show how the predicted probabilities can differ between the strategic and logit models, we present the results of a simple Monte Carlo analysis. Each replication of the analysis involved generating N = 2000 observations based on the behavioral assumptions of the strategic deterrence model in Figure 1. The explanatory variables were randomly drawn from a uniform distribution on the interval $[-2, 2]$, and the coefficients were set to $\beta_{13} = \beta_{14} = \beta_{24} = 1$. The α disturbances in Equations 7–10 were drawn from a Type 1 Extreme Value distribution with variance $\pi^2/6$, resulting in a logistic ϵ with variance $\pi^2/3$. The strategic and logit regressions were run and their estimates saved. These steps were replicated 2000 times to form densities of the parameter estimates.

Parameter estimates. The densities (not shown here) of the strategic estimates were all approximately normal and centered around one, indicating that the strategic model was able to recover the correct estimates on average.[10] What are our expectations concerning the logit estimates using this data? It turns out that η fairly well approximates a logistic distribution in this case, with approximately the

same variance as ϵ. Therefore, distributional misspecification is not a big concern. Next, note that because the explanatory variables are uniformly distributed between -2 and 2, $m_{13} = m_{14} = m_{24} = 0$. Given the true parameter values and the characteristics of the data, the logit estimators should converge to $B_{0L} = 0$, $B_{13L} = .48$, $B_{14L} = .48$, and $B_{24L} = 0$. In fact, our Monte Carlo analysis produced mean estimates of $\hat{\beta}_{0L} = .000$, $\hat{\beta}_{13L} = .49$, $\hat{\beta}_{14L} = .49$, and $\hat{\beta}_{24L} = .000$. Again, the direction is correct concerning the variables in the attacker's utilities. However, researchers would incorrectly infer that X_{24} has no effect on the attacker's decision to attack. Moreover, although the signs of $\hat{\beta}_{13L}$ and $\hat{\beta}_{14L}$ are correct, we will next show that the picture is more nuanced than the logit model allows.

Estimated probability of attacking. To understand what the parameter estimates imply for the estimated probabilities, Figure 5 displays the attacker's estimated probability of attacking based on (a) the logit model and (b) the strategic model, both as a function of X_{24} and X_{13}.[11] In both cases, the third variable, X_{14}, is held constant at its mean (zero).

Turning to Figure 5(a), we see that the researcher employing the logit model would infer from the estimated probabilities that X_{24} has no effect on the attacker's choice. No matter what X_{13}'s value is, changing X_{24} has no effect on the probability of attacking—the probability remains constant and the first-difference is zero. The researcher would also conclude that increasing X_{13} always increases the probability of attacking, and by the same amount regardless of the value of X_{24}.[12]

Now consider the estimated probabilities in Figure 5(b). The strategic model paints quite a different picture. Moreover, how X_{13} and X_{24} affect the probability of attacking is actually quite intuitive. Recall the strategic model as specified with the regressors in Figure 1(b), and recall that the graph was produced holding X_{14} at its mean, zero. In the context of the strategic model, the observable utilities for the attacker are $U_1(SQ) = 0$, $U_1(Cap) = X_{13}$, and $U_1(War) = 0$. For the defender, they are $U_2(Cap) = 0$ and $U_2(War) = X_{14}$.

To interpret the effects of X_{13} and X_{14}, it may be helpful to consider the three regions (A, B, C) in Figure 5(b)

---

[10]Plots of the densities for all parameter estimates in this section are available upon request from the authors.

[11]The graphs are based on exactly the same Monte Carlo as previously detailed, but with the X_{ij} drawn from $U[-10, 10]$ instead of $U[-2, 2]$. All inferences remain the same as before, including those pertaining to the estimated probabilities. The larger interval for the X_{ij} simply emphasizes the point made here. Plots for both sets of Monte Carlos are available upon request from the authors.

[12]Although not shown here, logit produces similar inferences when the plot is a function of X_{14} and X_{24}, holding X_{13} at zero.

Figures (a) and (b) display the estimated probability that the attacker will attack, based on the logit and strategic models, respectively. The probabilities are plotted as a function of X_{13} and X_{24}, holding X_{14} at its mean (zero). As (5a) displays, logit would lead us to believe that X_{24} has no effect on the attacker's decision, which contrasts with the (true) strategic model (5b). Figure (5b) also shows that the direction of the effect of X_{24} depends on X_{13}. The magnitude of the logit model's error is also fairly high for certain values of (X_{13}, X_{24}): at times off by more than .4.

(a) Logit   (b) Strategic
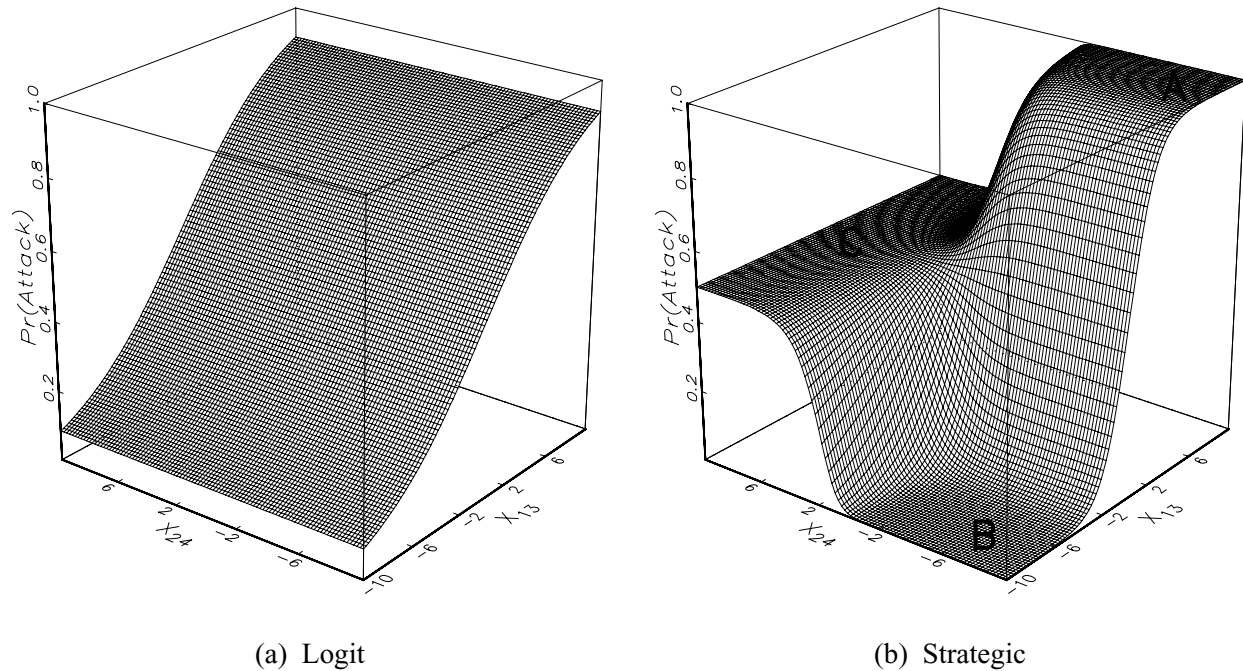
that are "almost flat." Region A reflects a situation where the defender's (observable) utility for war (X_{24}) is much lower than his utility for capitulation. In this case, he will likely capitulate. The attacker knows this, and her utility for capitulation (X_{13}) is much higher than her utility for the status quo. Therefore, in region A, the attacker will almost certainly attack. Region B again reflects a situation where the defender will back down if attacked. However, here the attacker's utility for the defender's capitulation is actually much lower than her utility for the status quo. Therefore, there is almost no chance the attacker will attack. Finally, consider region C. In this case, the defender's utility for war is very high, and he will likely defend himself if attacked. Knowing this, the attacker then has a choice between the status quo and war, both of which have an observable utility of zero. Because the observable utilities are so close (relative to the size of the variance), the analyst is almost completely uncertain as to whether the attacker will attack or not.

The effects of X_{13} and X_{24} can now be thought of as how they change the equilibrium probabilities as the utilities move from one extreme to the other. What is the effect of X_{24}? It depends. When $X_{13} > 0$, then increasing X_{24} decreases the probability of attack. On the other hand, when $X_{13} < 0$, increasing X_{24} increases the probability of attack. What is the effect of X_{13}? Again, it depends. In general, increasing X_{13} always increases the probability of attack. However, when X_{24} is small, X_{13} has a large effect on the probability, whereas when X_{24} is large, X_{13} has very little affect on the probability. Finally, it should be remembered that, because this is an equilibrium model, the effect of each variable depends on the values of the other variables. We have demonstrated how X_{13} and X_{24} affect the probability of attack when X_{14} is held constant at zero. The relationship may be very different when X_{14} is held constant at a different value.

To summarize, in contrast to the researcher using a logit model, a researcher employing the strategic model would conclude that (1) X_{24} affects the probability of attacking, (2) the value of X_{13} affects the direction of X_{24}'s effect, (3) X_{24} affects the magnitude of X_{13}'s effect. It should also be noted that the difference between

the logit and strategic estimated probabilities is off at times by more than .4. Moreover, the researcher employing the strategic model would be able to interpret her results in the context of a causal, equilibrium-based model.

The Monte Carlo examples presented here were intended to provide a concrete demonstration of the analytical results previously derived. They also suggest that the simplifying assumptions made in characterizing the specification error may not be especially egregious. In fact, as we proceeded through our analysis of the specification error, we repeatedly gave logit the benefit of the doubt. When those assumptions are violated, the specification error should only be worse. Finally, the graphs in Figure 5 raise another important issue, one concerning functional form and hypothesis specification and testing. Although the topic deserves a separate monograph of its own, it is relevant to the analysis here, so we provide at least a cursory discussion of it in the next section.

Theories, Functional Form, and Hypothesis Testing

Thus far, we have framed the problem under consideration as one of structural or functional form misspecification. To most practitioners, this may have previously seemed to be a rather arcane technicality, since so little attention is generally paid to it in political science methods training. However, functional form specification is also intimately related to hypothesis specification and testing. We have argued that greater attention should be paid to substantive theory and its implications for functional form. Yet, the converse is true as well: greater attention should be paid to the implications of our hypotheses concerning the functional form of our statistical model and its relationship to the theories we think we are testing.

The Functional Form Implied by Hypotheses

In an ideal world, a researcher would have a well-specified theory. Hypotheses would naturally follow from the theory. And the theory would suggest the appropriate functional form for the statistical model used to test the hypotheses. All three—theory, hypotheses, and statistical model—would be consistent with each other. When they are not consistent with each other, it raises a red flag that a serious problem exists in the research design. After all, hypotheses are supposed to reflect aspects of a theory to be tested, and the statistical model must reflect the hypothe-

ses in order to actually test them. Thinking about it this way—where a well-specified theory provides a functional form that drives the hypotheses and statistical model—the role of functional form in hypothesis specification and in the statistical model seems obvious.

What may not be so obvious are the implications of the above when a strong theory does not exist to specify the functional form and hypotheses. Most political science research still falls into this category. Often scholars invest a great amount of time thinking through the logic of what they are studying—in a very real sense, building models in their minds. However, it remains that, because the model is not written formally (e.g., using mathematical equations or formal logic notation), the paths to the conclusions reached and the relationships between the various parts of the model often remain unclear.

Almost invariably in this case, hypotheses are specified as *unconditionally monotonic* relationships— relationships where as one explanatory variable increases, the dependent variable always increases (or always decreases), regardless of the values of the other explanatory variables. Indeed, often a laundry list of unconditionally monotonic hypotheses is presented. A simple example related to international relations might be:

> H_1: *The stronger a nation is, the more likely it is to enter into war.*
>
> H_2: *The more democratic a nation is, the less likely it is to enter into war.*

Taken together, and in the context of the typical logit or probit model, hypotheses H_1 and H_2 imply that the functional relationship between the regressors and the dependent variable is unconditionally monotonic. The implied model is that the stronger a nation is, the more likely it is to go to war, regardless of how democratic it is; and, similarly, that the more democratic a nation is, the less likely it is to go to war, regardless of how strong it is.

If scholars truly believe the theory being tested is consistent with the jointly unconditionally monotonic relationships, then the common first-order XB_L functional form is exactly what they should use. However, many if not most theories analyzed in political science suggest more complicated relationships. The problem is that many scholars automatically employ a laundry list of monotonic hypotheses and the first-order XB_L functional form, without realizing the implied limitations concerning the theory being tested and the effects of misspecification on their inferences.

Although it is beyond the scope of this article to assess why this is the case, two rationales are frequently offered.

The first is the belief that the first-order XB_L specification is somehow a "general" specification. Apparently because of its simplicity (and perhaps widespread use?), many seem to view the first-order XB_L functional form as a more general functional form, where researchers can at least correctly estimate the first-order effects of variables. This is incorrect. As we have demonstrated, in the context of parametric estimation (whether MLE or Bayesian), it is certainly *not* a more general specification—it is a specific structural specification, which is quite restrictive. Indeed, when the data-generating process is not first-order linear, then it is a misspecification, and our inferences are affected.

To further illustrate why unconditional monotonicity is restrictive, consider the following two plausible situations that violate it. In the first, suppose (1) that war always decreases as democracy increases, regardless of the strength of the nation, but (2) that greater strength leads to less war for democratic regimes and to more war for authoritarian regimes. In this case, war and democracy are still unconditionally monotonically related, but war and strength are now *conditionally monotonically* related—i.e., war and strength are monotonically related for any given level of democracy, but the direction of that relationship changes depending on the level of democracy. Now consider a second example, where the likelihood of war is moderate for authoritarian regimes, very high for regimes that have some moderate level of democratic institutions, but very low for those nations with the highest level of democracy. In this case, war is a *nonmonotonic* function of democracy. All of these relationships seem plausible, but are inconsistent with the typical first-order linear model.

The second rationale for the laundry-list approach is that the researcher lacks a strongly specified theory to provide a functional form. In many ways, this rationale seems reasonable. If one lacks a well-specified theory prior to conducting the empirical analysis, what rationale is there for specifying a complicated relationship between the dependent variable and regressors? Historically, researchers without a firm theoretical justification for including interaction and higher-order polynomial terms have risked accusations of data mining. The lack of a well-specified theory is then held up as an excuse for the list of monotonic hypotheses.

Ironically, within this category there exists a large group of positivist scholars who recognize the importance of competition, incentives, and institutions, and who use statistical analysis to uncover general explanations (or "processes") of political behavior. However, the lack of a well-specified theory is not a "get out of jail free" card. Because the hypotheses and statistical model are formally specified, but the theory is left loosely speci-

fied, one can only conclude one of two things: (1) either the statistical model and hypotheses are inconsistent with whatever (more complicated) theory the researcher has in mind, or (2) they are consistent with a theory that only allows for unconditionally monotonic relationships. In other words, the researcher in this position must either accept that the statistical model does not reflect the theory or that her theory is far more simplistic and restrictive than she might want to admit.

Finally, it is often conjectured that typical logit or probit techniques should be fine for testing models where one can show (e.g., via an analysis of comparative statics) that the relationship being analyzed is monotonic. However, as we noted previously, the received wisdom requires the important qualification that the relationship is not just monotonic, but unconditionally monotonic. When the data-generating process implies unconditionally monotonic relationships between the regressors and dependent variable of interest, then the first-order XB_L specification is appropriate. Nevertheless, it is incumbent upon the researcher to determine (i.e., prove) that the theory implies such a relationship before using that specification.

Is Unconditional Monotonicity Likely the Rule or the Exception?

If most relationships were unconditionally monotonic, then, given current practices, researchers would have little concern about functional form. Unfortunately, in many areas unconditional monotonicity may be the exception, rather than the rule.

Consider again our simple model in Figure 1(b). It is easy to show that, using the least restrictive definition of monotonicity, all action probabilities ($p_{\bar{a}}$, p_a, $p_{\bar{r}}$, p_r) are monotonic in all of the regressors. As we noted before, this would seem to be an ideal situation for the argument that monotonic comparative statics are sufficient justification for the use of traditional logit and probit specifications. Then why did we find misspecification?

$p_{\bar{r}}$ and p_r are functions only of X_{24} and are indeed unconditionally monotonic in X_{24}. $p_{\bar{a}}$ and p_a are also unconditionally monotonic in X_{13} and X_{14}. However, $p_{\bar{a}}$ and p_a are conditionally monotonic in X_{24}. Assuming our data consisted of the attacker's actions, the typical laundry list of unconditionally monotonic hypotheses would be inconsistent with the data-generating process, and the first-order XB_L regression, as a reflection of the hypotheses, would be misspecified. Indeed, that is exactly what we saw in the third section; and Figure 5(b) clearly displays the conditional monotonicity of p_a and X_{24}: for $X_{13} < 0$,

the probability of attack monotonically increases in X_{24}, whereas for $X_{13} > 0$, it monotonically decreases in X_{24}.

Now suppose our data represented whether war occurred or not (i.e., *War* vs. {*SQ or Cap*}). Again, it is easy to show that p_{sq} is unconditionally monotonic in X_{13} and X_{14}, but conditionally monotonic in X_{24}, and that p_{cap} and p_{war} are unconditionally monotonic in X_{13} and X_{14}, but nonmonotonic in X_{24}. Even when each player's individual choice probabilities are monotonic in the regressors, the outcome probabilities may be nonmonotonic in the regressors. Again, this result is inconsistent with the unconditional monotonicity assumption one often finds in lists of hypotheses.

The deterrence model in Figure 1 is one of the very simplest strategic models possible, and, yet, a number of the relationships described by it are nonmonotonic or conditionally monotonic. It does not seem unreasonable to assume that much of the behavior we analyze in political science is at least as complex, if not more, resulting in similar nonmonotonic and conditionally monotonic relationships. This would imply that the traditional practice of specifying hypotheses as a list of unconditionally monotonic relationships may be woefully inconsistent with much of the political behavior we analyze, let alone the loosely-specified theories we think we are testing.

Concluding Remarks

To recap, we have characterized the strategic misspecification that arises from using the typical logit or probit specification—one where the latent dependent variable is a linear function of the parameters and a first-order function of the explanatory variables. In the interest of keeping the analysis as simple as possible and of creating ideal conditions for those interested in applying logit to comparative statics, we have constructed the "simplest strategic model possible" and have ensured that the relationship between all action probabilities and the regressors is monotonic.

Even under these conditions, the typical first-order linear specification is problematic. Ultimately, strategic misspecification is a functional form problem. Strategic models often imply a nonlinear and possibly nonmonotonic relationship between the latent dependent variable and the independent variables, in contrast to the first-order XB_L specification commonly found in multinomial and binomial logit and probit models. We have shown that as a functional form misspecification, strategic misspecification is equivalent to omitting relevant variables, where the omitted variables are nonlinear higher-order terms associated with expected utility calculations, and often with data aggregation as well.

So what is a poor researcher to do? Our use of Taylor series expansions as approximations for the strategic regressions suggests one possible solution to the functional form problem: run polynomial regressions with interaction terms. Data-mining critiques aside, there are at least two other problems with this approach. Most problematic is that increasing the order (and, therefore, the extent to which the polynomial approximates the true functional form) combinatorially increases the number of parameters that must be estimated, becoming impractical to implement with only a relatively small number of independent variables. For example, a typical first-order specification with six variables requires that we estimate only seven parameters—i.e., the coefficients for each independent variable, plus the constant. On the other hand, the full second-order specification for those same six variables requires that we estimate 28 parameters, and the third-order specification requires that we estimate 84 parameters. Secondly, as defined here, a polynomial regression really *is* an exercise in curve fitting—i.e., of correctly approximating the strategic "curve," in order that functional form misspecification not bias estimates. As such, although it may allow us to estimate the functional form correctly, it does not allow us to say anything about causality.

One approach would be to develop better theory and then derive hypotheses and statistical models from that theory, thereby ensuring consistency of one's theory and statistical analysis. Another approach would be to use "theory-less" statistical techniques that do not impose so much structure on the analysis. The benefit of the structural approach is that a particular causal relationship is analyzed. A problem with it is that, if the structure is misspecified, then specification error results. The benefit of the nonstructural approach is that (by definition) it does not rely on structural specification. The problem is that it has no theoretical (or causal) purchase. Still another approach would be to use both methods iteratively: using nonlinear techniques to determine the functional relationship of dependent and independent variables, and then developing and estimating structural models that are informed by the previous stage. In our opinion, the worst situation would be to continue the current practice of using the very simplest of structural statistical models— the first-order XB_L specification—to analyze behavior that we believe *a priori* cannot be consistent with that specification, and which we now know produces invalid inferences.

References

Dubin, Jeffrey A., and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit, and Probit Models." *Sociological Methods and Research* 18(2/3):360–90.

Kmenta, Jan. 1986. *Elements of Econometrics.* 2nd ed. New York: Macmillan.

McKelvey, Richard D., and Thomas R. Palfrey. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics* 1(1):9–41.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–98.

Signorino, Curtis S. 2000. "Structure and Uncertainty in Discrete Choice Models." Presented at the 1999 Summer Political Methodology Meeting.

Smith, Alastair. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43(4):1254–83.

Yatchew, Adonis, and Zvi Griliches. 1985. "Specification Error in Probit Models." *Review of Economics and Statistics* 67(1):134–39.