Strategic Management Journal
Strat. Mgmt. J., **30**: 1335–1348 (2009)
Published online EarlyView in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/smj.783
Received 20 November 2007; Final revision received 16 April 2009

RESEARCH NOTES AND COMMENTARIES

USING SIMULATION TO INTERPRET RESULTS FROM LOGIT, PROBIT, AND OTHER NONLINEAR MODELS

BENNET A. ZELNER*
The Fuqua School of Business, Duke University, Durham, North Carolina, U.S.A.

In a recent issue of this journal, Glenn Hoetker proposes that researchers improve the interpretation and presentation of logit and probit results by reporting the marginal effects of key independent variables at theoretically interesting or empirically relevant values of the other independent variables in the model, and also by presenting results graphically (Hoetker, 2007: 335, 337). In this research note, I suggest an alternative approach for achieving this objective: reporting differences in predicted probabilities associated with discrete changes in key independent variable values. This intuitive approach to interpretation is especially useful when the theoretically interesting or empirically relevant changes in independent variables values are not very small, and also for models that contain interaction terms (or higher-order terms such as quadratics). Although the graphical presentations recommended by Hoetker implicitly embody this approach, they typically fail to include appropriate measures of statistical significance, and may therefore lead to erroneous conclusions. In order to calculate such measures, I recommend and demonstrate an intuitive simulation-based approach to statistical interpretation, developed by King et al. (2000), that has gained widespread adherence in the field of political science. Throughout the article, I provide a running example based on research that has previously appeared in the Strategic Management Journal. Copyright © 2009 John Wiley & Sons, Ltd.

INTRODUCTION

In a recent issue of this journal, Hoetker (2007) identifies four critical issues in the use of logit and probit models in existing management research, and makes recommendations for addressing them.

Keywords: logit; probit; econometrics; nonlinear models; simulation; interaction terms
*Correspondence to: Bennet A. Zelner, The Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708-0120, U.S.A. E-mail: bzelner@duke.edu

In order to address the first two issues, the interpretation of coefficients and the interpretation of interaction effects, Hoetker advises researchers to (1) assess the marginal effects of key independent variables at theoretically interesting or empirically relevant values of the other independent variables in a model, and (2) present results graphically (Hoetker, 2007: 335, 337). The underlying rationale for both recommendations is to provide readers with substantively meaningful interpretations of the estimated effects of interest.

In this research note, I suggest an alternative approach for achieving this objective: reporting differences in predicted probabilities associated with discrete changes in key independent variable values, and using a simulation-based technique developed by political scientists (King, Tomz, and Wittenberg, 2000) to assess statistical significance. This intuitive approach to interpretation is especially useful when the changes in independent variable values of theoretical interest or empirical relevance are not very small, and also for models that contain interaction terms (or higher-order terms such as quadratics), which pose added interpretive challenges. Additionally, I suggest a style of graphical presentation that directly illustrates the measures of central interest for hypothesis testing. I supply sample Stata code for producing such presentations in Appendix 2, and have made available a more general Stata command that can be downloaded from the Internet.

INTERPRETING COEFFICIENTS

In his article, Hoetker reminds readers of the well-known difficulties associated with interpreting coefficients in logit, probit, and other nonlinear models. In linear models, the coefficients have a straightforward interpretation: they represent the estimated change in the value of the dependent variable associated with a unit increase in the corresponding independent variable. In contrast, the nonlinearity of logit and probit models means that the relationship between a change in the value of an independent variable and the estimated change in the probability of a positive outcome cannot be directly discerned from the variable's coefficient. Instead, this magnitude varies depending on the magnitude of the change in the independent variable of interest, the variable's starting value, and the values of the other independent variables in the model (Long and Freese, 2006: 171).

Hoetker's main suggestion for addressing these issues is for researchers to report the marginal effects of key independent variables for several sets of theoretically interesting or empirically relevant values of the other independent variables in the model (Hoetker, 2007: 335). Although this approach is more informative than alternative approaches such as reporting odds ratios or coefficients alone, the fact that a marginal effect is a derivative in some cases limits its utility for answering the questions of substantive interest that motivate statistical analysis in the first place.[1] Such questions typically take the form, 'What is the change in the probability of observing the outcome $Y = 1$ when the independent variable(s) X changes by ΔX?' If the substantively meaningful or empirically relevant value of ΔX is small, then the marginals may provide a reasonable basis for approximating the effect of substantive interest.[2] However, if the amount by which an independent variable is postulated to change is larger—for example, when X is a binary or ordinal variable that necessarily changes in unit increments, or when X is a continuous variable whose empirical distribution is characterized by large differences between observed values—then focusing on marginal effects can give misleading estimates of probability changes (Caudill and Jackson, 1989; Kennedy, 2003: 266–267; Petersen, 1985). In this case, it may be more informative to report the difference in predicted probabilities associated with ΔX (Caudill and Jackson, 1989; Kennedy, 2003: 266–267; Petersen, 1985):

$$\Delta\hat{\pi} = \hat{\pi}(x_j' + \Delta x_j, \bar{x}_{i\neq j}) - \hat{\pi}(x_j', \bar{x}_{i\neq j})$$

where x_j' represents the initial value of the independent variable of interest, Δx_j represents the change in this variable's value, and the other independent variables $x_{i\neq j}$ are held at specific theoretically meaningful or empirically relevant values. Moreover, because interpreting a difference in predicted probabilities requires no specialized knowledge of calculus or advanced statistics (King *et al.*, 2000: 347), an additional benefit of presenting logit and probit results in this manner is increased accessibility to managers, policymakers, and others who might lack training in these areas.

[1] The marginal effect of the independent variable x_j represents the ratio of $\hat{\pi}$, the predicted probability of a positive outcome, to a change in the value of x_j as this change approaches zero, when the other independent variables $x_{i\neq j}$ are held at specific values:

$$\frac{\partial\hat{\pi}}{\partial x_j} = \lim_{\Delta x_j \to 0} \frac{\hat{\pi}(x_j + \Delta x_j, \bar{x}_{i\neq j}) - \hat{\pi}(x_j, \bar{x}_{i\neq j})}{\Delta x_j}$$

[2] The logit or probit function must also be relatively flat in the region of interest. However, the derivative expression in footnote 1 may still take a value that exceeds one (Petersen, 1985: 130) and may also lead to violations of the constraint that the probabilities of exclusive events sum to one (Caudill and Jackson, 1989).

DOI: 10.1002/smj

INTERPRETING INTERACTION EFFECTS

The second topic that Hoetker considers is the interpretation of interaction effects. The issues associated with interacted variables are more serious than those arising in the case of uninteracted variables, and the benefits of reporting differences in predicted probabilities commensurately greater. Consider the following model:

$$\Pr(Y = 1) = \pi = \mathrm{g}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2)$$

where $\mathrm{g}(\cdot)$ is the logit or probit function (or other nonlinear function). If this model were linear, the interaction term coefficient β_3 would have an intuitive interpretation as the amount by which the dependent variable changes in association with a unit change in x_1, conditional on a unit change in x_2 (or symmetrically, the amount by which the dependent variable changes in association with a unit change in x_2, conditional on a unit change in x_1). However, in a logit or probit model, the coefficient on an interaction term lacks such an interpretation as a result of the model's nonlinearity. Indeed, the sign of the interaction term coefficient need not correspond to the direction of the (hypothesized) conditional effect motivating the interaction term's inclusion in the first place, and the standard error of this coefficient conveys no direct information about the statistical significance of the effect (Ai and Norton, 2003; Huang and Shields, 2000).

Hoetker's main suggestions for addressing this issue are twofold. The first is to calculate the cross-partial derivative of $\hat{\pi}$, the predicted probability of a positive outcome, with respect to each of the interaction term's constitutive variables. However, this approach is subject to the same limitation discussed above: because a cross-partial derivative represents a marginal effect, it may give misleading estimates of probability changes when the independent variable value changes of substantive or empirical interest are not very small (e.g., in the case of binary or ordinal variables, or variables whose empirical distributions are characterized by large differences between observed values). Moreover, because strategy researchers typically employ multiplicative interaction terms to assess the effect of a variable x_1, conditional on the level of a second variable x_2, it is often of interest to directly assess this conditional effect, consisting of x_1's 'main' effect through the term $\beta_1 x_1$ and its 'interactive' effect through the term $\beta_3 x_1 x_2$ (Friedrich, 1982: 806). A straightforward way to do so is, again, to calculate a difference in predicted probabilities:

$$\Delta\hat{\pi} = \hat{\pi}(\beta_0 + \beta_1 x_1' + \beta_1 \Delta x_1 + \beta_2 \bar{x}_2 + \beta_3 x_1' \bar{x}_2 + \beta_3 \Delta x_1 \bar{x}_2) - \hat{\pi}(\beta_0 + \beta_1 x_1' + \beta_2 \bar{x}_2 + \beta_3 x_1' \bar{x}_2)$$

where x_1' represents the initial value of the independent variable x_1, Δx_1 represents the change in the value of x_1, and x_2 is held at the value $\bar{x}_2$. This expression can be calculated for different meaningful values of x_2, and the resulting probabilities then compared.

Hoetker's suggestion to portray results graphically (Hoetker, 2007: 336) implicitly embodies the notion of examining differences in predicted probabilities and can be used to illustrate the advantages of this approach. Consider Figure 1, which is reproduced from Hoetker's paper (2007: 336) and summarizes results from Leblein and Miller's (2003) analysis of vertical integration decisions in the semiconductor industry. The binary dependent variable in Leiblein and Miller's (2003) model takes a value of one for a 'make' decision and zero for 'buy,' and the independent variables in the model include a continuous measure of demand uncertainty, a binary measure of asset specificity that takes a value of one when asset specificity is present and zero otherwise, and an interaction term equal to the product of the demand uncertainty and asset specificity variables. The figure shows how the predicted probability of vertical integration changes in association with different combinations of demand uncertainty and asset specificity for three different firm 'types,' defined by setting the remaining independent variables in the model to specific levels.[3]

Leiblein and Miller's (2003) model can be used to answer the research question, 'What is the effect of an increase in asset specificity on the probability of vertical integration, conditional on the

[3] Hoetker defines the 'ambivalent' firm by setting the values of the independent variables other than asset specificity and demand uncertainty to zero, the 'mean values' firm by setting these values to their sample means, and the 'one s.d. from mean firm' by setting these values to one standard deviation above or below the sample mean so as to minimize the probability of a 'make' decision.

DOI: 10.1002/smj

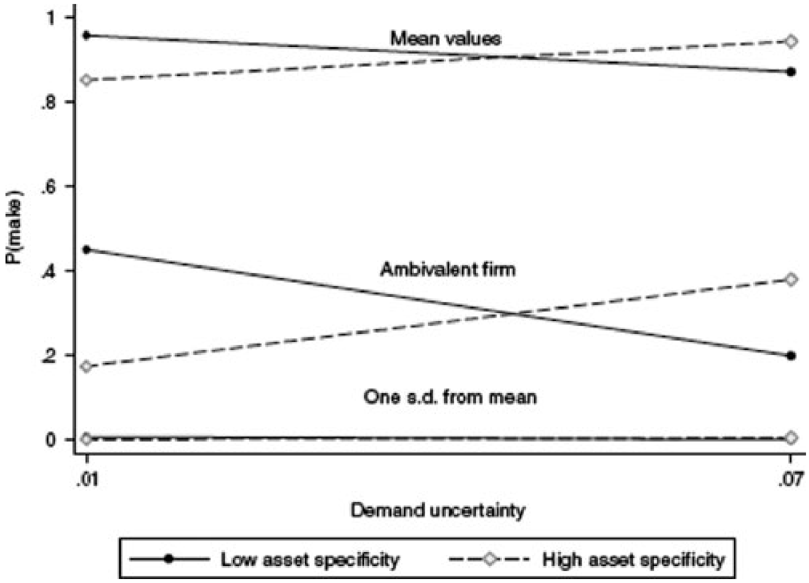

Figure 1. Interaction effects from Leiblein and Miller's (2003) study[6]

level of demand uncertainty?'[4] However, focusing on the cross-partial derivatives associated with the interaction term alone does not provide an answer to this question. Rather, it is the vertical difference between the high and low asset specificity schedules (again, for a given firm type) that is of substantive interest. This conditional effect is equal to the difference in predicted probabilities when the asset specificity variable increases from zero to one, at different observed values of the demand uncertainty variable.[5]

ASSESSING SIGNIFICANCE

Although graphical presentations such as that in Figure 1 are intuitively appealing because they allow readers to visually compare the predicted probabilities associated with different combinations of independent variable values, they do not by themselves provide a sufficient basis on which to draw sound statistical conclusions. The reason is, like the estimated coefficients on which they are based, predicted probabilities are also estimates. Thus, simply observing that a predicted probability changes when the value of an independent variable of interest changes does not provide evidence of a statistical relationship. Instead, it is necessary to test whether the difference in predicted probabilities is statistically different from zero by constructing a confidence interval (typically, a 95% confidence interval) around this estimated quantity and determining whether the interval includes zero.

It is important to recognize that testing a difference in predicted probabilities from a logit or probit model against the null hypothesis does not constitute an additional demand, in terms of rigor, beyond that routinely made for drawing statistical conclusions. Researchers using linear models are expected to test whether the difference in predicted values associated with a unit change in each of a model's independent variables differs statistically from zero. However, in the linear case, this test reduces to a test of each variable's estimated

[4] According to transaction cost theory, higher asset specificity and uncertainty should both be associated with a higher probability of vertical integration (Williamson, 1979, 1991). The specific hypothesis that Leiblein and Miller (2003: 845) consider involves the effect of uncertainty, conditional on the level of asset specificity.

[5] These statements would still hold if the model depicted in Figure 1 were linear. Even though researchers often focus on the magnitude and statistical significance of an interaction term's coefficient, the estimated effect of one of the constitutive variables in an interaction term—as well as its standard error—varies at different levels of the other constitutive variable (Brambor, Clark, and Golder, 2006; Friedrich, 1982; Jaccard, Turrisi, and Wan, 1990).

[6] Figure taken from Glenn Hoetker, The use of logit and probit models in strategic management research: Critical issues; *Strategic Management Journal*, 28(4): 331–343. Reprinted with permission from John Wiley & Sons.

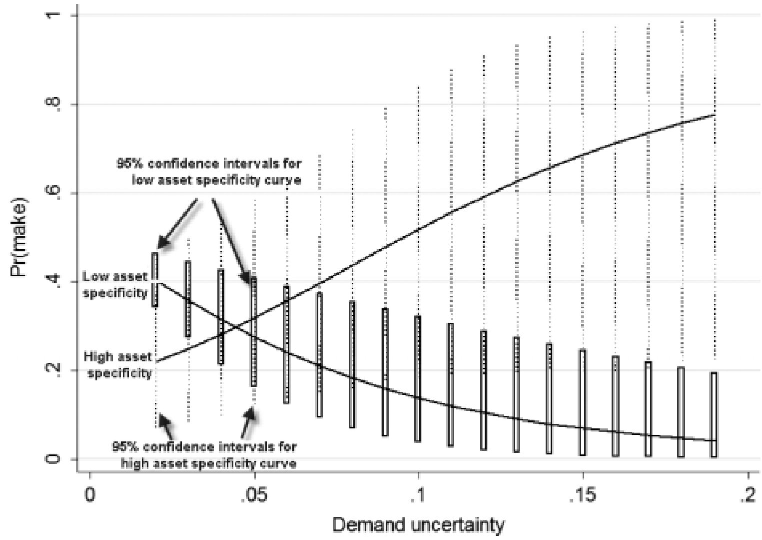

Figure 2. Predicted probabilities and confidence intervals for the 'ambivalent' firm

coefficient against the null hypothesis that the coefficient is equal to zero.

In contrast, in the case of logit and probit models, a hypothesis test based on an independent variable's estimated coefficient alone cannot be interpreted in the same way. The underlying reason is again the nonlinearity of these models. Specifically, just as the magnitude of the estimated effect of an independent variable in one of these models depends on the values of all the independent variables in the model, so too does the confidence interval surrounding this estimated effect (Greene, 2003: 675).[7] Moreover, when an interaction term is present, it is not just the confidence interval surrounding the interaction term's coefficient that is of interest; rather, the confidence interval surrounding the estimated effect of each of the interaction term's constitutive variables, conditional on the level of the other constitutive variable, should also be evaluated (Brambor *et al*., 2006; Friedrich, 1982; Jaccard *et al*., 1990).

Figure 1 can again be used to illustrate. For each firm type, the difference in predicted probabilities associated with an increase in asset specificity is negative at low levels of demand uncertainty and positive at high levels. However, the figure does not indicate whether this difference is ever statistically different from zero (although at the point where the low and high asset specificity schedules for each firm type cross, the difference cannot be statistically different from zero).

Identifying the range of values of demand uncertainty over which the effect of asset specificity is statistically significant has important implications in the substantive context of Leiblein and Miller's (2003) study. If it turns out that the negative effect of asset specificity on the probability of vertical integration at lower levels of demand uncertainty is statistically different from zero, then Leiblein and Miller's (2003) results contradict a central claim of transaction cost theory, which holds that an increase in asset specificity should have a positive effect on the probability of vertical integration (Williamson, 1979, 1991). If, on the other hand, the negative effect of asset specificity at lower levels of demand uncertainty is not statistically different from zero, but the positive effect of asset specificity at higher levels of demand uncertainty is statistically different from zero, then the model's results are consistent with transaction cost theory.

Figure 2 is similar to Figure 1, but it includes confidence intervals in addition to predicted probabilities, reflects the entire range of sample values for the demand uncertainty variable on the x-axis, and depicts only the 'ambivalent' firm. The confidence intervals were calculated using King *et al.'s* (2000) simulation-based approach, discussed in detail below. The ambivalent firm, as defined by

[7] This statement applies to marginal effects as well as differences in predicted probabilities.

DOI: 10.1002/smj

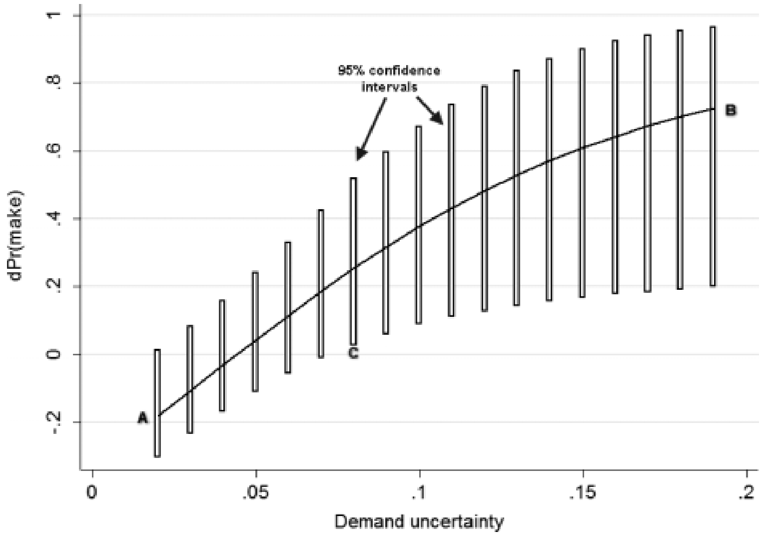

Figure 3. Effect of an increase in asset specificity for the 'ambivalent' firm

Hoetker,[8] is modeled by setting the values of the independent variables other than asset specificity and demand uncertainty to zero, and serves as a useful illustration because the predicted probabilities for this hypothetical firm vary substantially depending on the values of the demand uncertainty and asset specificity variables.

The 95 percent confidence intervals surrounding the predicted probabilities for low asset specificity and high asset specificity in Figure 2 overlap, and the extent of overlap varies with the level of demand uncertainty. Although there is no way to discern from the figure alone whether this overlap is ever great enough so that the null hypothesis (i.e., that the difference in predicted probabilities associated with a change from low to high asset specificity is zero) cannot be rejected, the figure certainly suggests the need to test the effect of asset specificity against the null hypothesis at different levels of demand uncertainty.

Accordingly, Figure 3 depicts the difference in predicted probabilities associated with an increase in asset specificity (i.e., the vertical distance between the low and high asset specificity schedules in Figure 2) at different levels of demand uncertainty, along with the 95 percent confidence interval for this difference. The figure indicates that, even though the estimated effect of asset specificity on the probability of vertical integration is negative at lower values of demand uncertainty—as in Figures 1 and 2—this effect is not statistically significant in this region because the 95 percent confidence interval includes zero. In contrast, at higher levels of demand uncertainty (to the right of point C), not only is the effect of asset specificity positive, but it is also significantly different from zero. Moreover, the null hypothesis that the effect of asset specificity when demand uncertainty is high (point B) does not exceed the effect of asset specificity when demand uncertainty is low (point A) can be rejected at a confidence level of over 99 percent.[9] Thus, Leiblein and Miller's (2003) results—at least for the 'ambivalent' firm—are indeed consistent with transaction cost theory. Failure to assess the confidence interval associated with the difference in predicted probabilities at different levels of demand uncertainty—as in Figure 1—results in an incorrect inference of substantive importance.

Differences in the statistical significance of the effect of asset specificity for a given firm type at different levels of demand uncertainty are due to the fact that the asset specificity and demand uncertainty variables are interacted with one another. However, Leiblein and Miller's (2003) results can also be used to demonstrate that the statistical significance of the effect of asset specificity depends on the values of other independent variables in the model, that is, those with which asset specificity

[8] Personal e-mail exchange with author, 2 July 2007.

[9] This statement holds for both one- and two-tailed tests.

DOI: 10.1002/smj

is not interacted. In Figure 1, the effect of asset specificity—the vertical distance between the high and low asset specificity schedules—at a given level of demand uncertainty is always largest for the 'ambivalent' firm and smallest for the 'one s.d. from mean' firm. Because each of the two firm types is defined by setting the other independent variables in the model to specific values, differences in the effect of asset specificity when demand uncertainty is held constant—as well as the confidence interval surrounding this estimated effect—are entirely attributable to the values of these other variables.

Thus, consider the effect of asset specificity on the probability of a 'make' decision for different levels of the demand uncertainty variable for each of these two firm types. As discussed above, the effect of an increase in asset specificity for the ambivalent firm differs from zero at the 95 percent level or better for values of demand uncertainty of 0.08 and above (point C in Figure 3). However, for the 'one s.d. from mean' firm (not shown in Figure 3), the change in the probability of vertical integration associated with an increase in asset specificity never differs significantly from zero at better than the 69 percent level.

THE SIMULATION-BASED APPROACH TO INTERPRETATION

Although calculating the difference in predicted probabilities associated with a change in the value of one or more independent variables is a straightforward task, calculating the confidence intervals necessary to draw valid statistical conclusions about this quantity requires additional effort. In recent years, applied researchers in the field of political science have begun to address interpretive challenges of this sort for logit, probit, and other nonlinear models by employing the simulation-based approach developed by Gary King and his colleagues (2000).

The starting point for the simulation-based approach to statistical interpretation is the same central limit theorem result underlying conventional hypothesis testing: if enough samples are drawn from the sampling population and used for estimation, the resulting coefficient estimates will be distributed joint-normally (King *et al.*, 2000: 350). The difference is that, instead of constructing confidence intervals based on standard errors derived using multivariate calculus and a normal distribution table, the researcher simulates the distribution of the coefficient estimates directly by repeatedly drawing new values of these estimates from the normal distribution.

Specifically, the simulation-based approach consists of taking M draws from the multivariate normal distribution with mean $\hat{\beta}$, the estimated coefficient vector from a logit, probit, or other nonlinear model; and variance matrix $V(\hat{\beta})$, the estimated variance-covariance matrix for the coefficient estimates in the model. The M draws yield M simulated coefficient vectors. The mean simulated coefficient vector converges to the original estimated coefficient vector, and the distribution of the M simulated coefficient vectors—which is based entirely on information from the original logit, probit, or other nonlinear model, together with the standard assumptions—reflects the precision of the coefficient estimates (King *et al.*, 2000: 349–350). With the M simulated coefficient vectors in hand, the researcher may then calculate simulated predicted probabilities, differences in simulated predicted probabilities, or any other function of interest, as well as the associated confidence intervals.

The simulation-based approach has several advantages over the conventional mathematical technique. First, it produces more accurate results, both because the latter requires use of the 'delta method' (Oehlert, 1992), a calculus-based technique for approximating nonlinear functions of random variables (King *et al.*, 2000: 352–353),[10] and also because the simulation-based approach implicitly corrects for a bias in the formula typically used to calculate predicted probabilities (King and Zeng, 2001).[11] Another advantage of the simulation-based approach is that, in eschewing complex calculus-based approximations in favor of straightforward numerical calculations, it enhances the intuition of researchers and readers with limited knowledge of multivariate calculus. Finally, the simulation-based approach is easy to implement as

[10] Applying the delta method requires the researcher to take derivatives of linearized functions and, in order to calculate standard errors, approximate a complex integral and its variance (King *et al.*, 2000: 352). The difficulty of doing so increases with the number of terms in the model (Liao, 2000), and the accuracy of the results is limited by the number of Taylor series terms used (King *et al.*, 2000: 352).

[11] See Appendix 3.

DOI: 10.1002/smj

the result of a freely available suite of Stata commands known as CLARIFY (Tomz, Wittenberg, and King, 2001), as well as a free standalone program called 'Zelig' (Imai, King, and Lau, 2006).

To be sure, users of Stata whose logit or probit models contain a single interaction term, and in which the interacted variables do not also occur in higher-order terms, may still calculate marginal effects without great difficulty by employing the code written by Norton, Wang, and Ai (2004) and recommended by Hoetker (2007: 337).[12] However, for logit and probit models containing multiple interaction terms or higher-order terms, as well as for ordered logit, ordered probit, multinomial logit, and several other nonlinear models (King *et al.*, 2000: 360), the Norton *et al.* (2004) code will not work, and the simulation-based approach is the most practical alternative.

EXAMPLE

As an example, consider Table 1, the top panel of which reports the estimated coefficients from Leiblein and Miller's (2003) model, 10 sets of simulated coefficients that were generated using CLARIFY, and the mean values of these simulated coefficients. Despite the small number of simulations shown (standard practice is to draw at least 1,000 sets of simulated coefficients), the mean values show strong evidence of converging to the 'original' estimated coefficients, and the distribution of the simulated coefficients can be used to construct confidence intervals for illustrative purposes. For example, the 80 percent two-tailed confidence interval for each coefficient in the table is bounded by the second-lowest and second-highest simulated values for the coefficient.[13] Thus, the null hypothesis that the coefficient on the intercept is zero cannot be rejected at the 80 percent level, because the 80 percent confidence interval (−0.314, 2.201) includes zero. In contrast, the 80 percent confidence interval (0.042, 0.085) for the *FIRM TENURE* coefficient does not include zero, meaning that the *FIRM TENURE* variable is significantly different from zero at the 80 percent level or better. Even though not strictly comparable due to the small number of simulations, these results are nonetheless consistent with Leiblein and Miller's (2003): the intercept is not statistically significant at the 95 percent level in their model, but the *FIRM TENURE* variable is significant at better than the 99 percent level.

To continue with the example, consider the bottom panel of Table 1. The rows for $\hat{\pi}_{LO}$ and $\hat{\pi}_{HI}$ contain predicted probability calculations for the original estimated coefficients and the 10 sets of simulated coefficients when *DEMAND UNCERTAINTY* $= 0.05$ and *ASSET SPECIFICITY* takes values of zero and one, respectively (the remaining variables in the model are still set to zero). The respective 80 percent two-tailed confidence intervals for $\hat{\pi}_{LO}$ and $\hat{\pi}_{HI}$ are (0.178, 0.375) and (0.168, 0.409).[14]

For purposes of hypothesis testing, the values in the next two rows of the table are of central interest. First consider $\Delta\hat{\pi}_{.05}$, the difference in predicted probabilities associated with an increase in *ASSET SPECIFICITY* from zero to one when *DEMAND UNCERTAINTY* continues to take a value of 0.05 (i.e., $\Delta\hat{\pi}_{.05} = \hat{\pi}_{HI} - \hat{\pi}_{LO}$.) The 80 percent two-tailed confidence interval for $\Delta\hat{\pi}_{.05}$, again identified as the second-lowest and second-highest simulated values in the relevant row, is (−0.059, 0.077). Because this confidence interval includes zero, the null hypothesis that an increase in asset specificity is associated with an increased probability of 'make' cannot be rejected at even the 80 percent level when *DEMAND UNCERTAINTY* takes a value of 0.05.

Now consider the row containing $\Delta\hat{\pi}_{.19}$, the difference in predicted probabilities associated with an increase in *ASSET SPECIFICITY* from zero to one when *DEMAND UNCERTAINTY* takes the sample maximum value of 0.19. Even though the details are not shown for the sake of brevity, the figures in this row were calculated in an analogous manner to those in the row for $\Delta\hat{\pi}_{.05}$. The 80 percent two-tailed confidence interval (0.544, 0.932) does not contain zero, indicating that this effect is significant at at least the 80 percent level.

These results are not surprising in light of Figure 3, which is based on 1,000 simulations. According to the figure, the 95 percent two-tailed confidence interval around the effect of an increase in *ASSET SPECIFICITY* from zero to one when *DEMAND UNCERTAINTY* $= 0.05$

[12] The Stata command written by Norton *et al.* is −inteff−.

[13] An 80 percent confidence interval is used for illustrative purposes because it is not possible to construct a 95 percent confidence interval with only 10 simulated values.

[14] The bounds of these confidence intervals will always lie between zero and one in logit and probit models.

DOI: 10.1002/smj

Table 1. Simulation results for Leiblein and Miller's (2003) model

Independent variable	Coefficient estimate	Simulated coefficients												
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Mean	80% Cl	
													Lower	Upper
INTERCEPT	0.433	0.713	1.315	2.565	−0.314	2.201	0.258	0.160	1.702	−1.398	1.915	0.912	−0.314	2.201
FIRM SIZE	0.260	0.111	0.011	0.234	0.201	0.242	0.298	0.386	0.130	0.386	0.454	0.245	0.111	0.386
FIRM TENURE	0.065	0.082	0.099	0.070	0.069	0.050	0.076	0.085	0.083	0.042	0.028	0.068	0.042	0.085
US FIRM	−0.643	−0.259	−0.727	−1.347	−0.228	−0.481	−0.412	−0.199	−0.386	−0.008	−0.571	−0.462	−0.727	−0.199
JAPANESE FIRM	0.037	0.778	0.140	−0.273	1.316	−0.022	0.102	−0.080	0.211	1.990	−0.936	0.323	−0.273	1.316
OTHER ASIAN FIRM	−0.194	0.355	−0.141	−0.078	0.791	−0.578	−1.550	−0.613	0.133	1.012	−1.191	−0.186	−1.191	0.791
EX ANTE SMALL NUMBERS	−0.452	−0.443	−0.254	−0.599	−0.341	−0.477	−0.280	−0.501	−0.592	−0.530	−0.514	−0.453	−0.592	−0.280
SMALL NUMBERS SQUARED	0.013	0.012	0.007	0.016	0.010	0.015	0.008	0.016	0.018	0.017	0.014	0.013	0.008	0.017
ASSET SPECIFICITY	−1.713	−1.860	−1.734	−0.324	−3.395	−2.055	−3.128	−0.825	−2.137	−1.213	−2.499	−1.917	−3.128	−0.825
DEMAND UNCERTAINTY	−19.890	−30.593	−25.020	−8.491	−24.510	−19.607	−37.845	−10.179	−13.888	−12.119	−25.902	−20.815	−30.593	−10.179
ASSET SPECIFICITY X UNCERTAINTY	37.612	44.886	46.637	12.744	60.373	32.251	68.148	19.320	45.110	21.416	52.008	40.289	19.320	60.372
FABRICATION EXPERIENCE HAT	0.315	0.296	0.252	0.354	0.359	0.349	0.311	0.268	0.339	0.227	0.330	0.308	0.252	0.354
SOURCING EXPERIENCE HAT	−0.388	−0.317	−0.544	−0.451	−0.430	−0.388	−0.620	−0.496	−0.330	−0.094	−0.539	−0.421	−0.544	−0.317
DIVERSIFICATION STRATEGY	1.889	2.411	0.936	1.115	2.329	1.110	1.870	1.062	1.331	1.913	1.406	1.548	1.062	2.329
DIVERSIFICATION SQUARED	−0.306	−0.383	−0.130	−0.224	−0.380	−0.206	−0.305	−0.215	−0.226	−0.313	−0.238	−0.262	−0.380	−0.206
$\hat{\pi}_{LO}$	0.270	0.178	0.223	0.395	0.227	0.273	0.131	0.375	0.333	0.353	0.215	0.270	0.178	0.375
$\hat{\pi}_{HI}$	0.304	0.241	0.342	0.472	0.168	0.194	0.166	0.409	0.360	0.321	0.233	0.291	0.168	0.409
$\Delta\hat{\pi}_{.05}$	0.034	0.063	0.120	0.077	−0.059	−0.079	0.035	0.033	0.027	−0.032	0.018	0.020	−0.059	0.077
$\Delta\hat{\pi}_{.19}$	0.817	0.699	0.906	0.453	0.959	0.563	0.932	0.587	0.911	0.544	0.914	0.747	0.544	0.932
$\Delta\Delta\hat{\pi}_{.19,.05}$	0.783	0.635	0.786	0.376	1.018	0.641	0.897	0.553	0.884	0.576	0.897	0.726	0.553	0.897

is approximately zero. However, when *DEMAND UNCERTAINTY* $= 0.19$, the lower bound of the 95 percent two-tailed confidence interval is approximately 0.20, which is well above zero.

The specific steps used to construct Figures 2 and 3 were as follows:

1. Estimate Leblein and Miller's (2003) logit model.
2. Draw 1,000 sets of simulated coefficients from the multivariate normal distribution with mean equal to the estimated coefficient vector and variance equal to the estimated variance-covariance matrix.
3. Calculate the predicted probabilities (Figure 2) or difference between these predicted probabilities (Figure 3) for each set of simulated coefficients at the different combinations of independent variable values depicted in the figures.
4. Plot the means of the simulated predicted probabilities (Figure 2) or simulated differences in predicted probabilities (Figure 3).
5. Identify the lower and upper bounds of the desired confidence intervals for the plotted values. Because the figures show 95 percent two-tailed confidence intervals, the lower and upper bounds are the 26th and 975th simulated values, respectively, when the values are sorted in ascending order.

Appendix 1 gives an overview of the CLARIFY commands used to produce both figures, and Appendix 2 contains the annotated Stata code itself.

Finally, consider the bottom row of Table 1, which contains the double difference $\Delta\Delta\hat{\pi}_{.19,.05} = \Delta\hat{\pi}_{.19} - \Delta\hat{\pi}_{.05}$. This figure is analogous to a cross-partial derivative—a quantity associated with the coefficient on an interaction term—but is based on specific discrete changes in the values of the underlying variables, as opposed to infinitesimal changes. This particular double difference can be used to test against the null hypothesis that the effect of an increase in *ASSET SPECIFICITY* when *DEMAND UNCERTAINTY* takes the sample maximum value of 0.19 is not greater than the effect of an increase in *ASSET SPECIFICITY* when *DEMAND UNCERTAINTY* takes a value of 0.05. The 80 percent two-tailed interval of (0.553, 0.897) is constructed as before by identifying the second-lowest and second-highest values among the 10 simulated double differences. However, the 80 percent one-tailed interval (0.576, 1.018)—constructed by indentifying the third-lowest and maximum simulated double differences, which together bound the upper 80 percent of the distribution—may be more appropriate in this case. Consistent with the discussion of Figure 3 above, the null hypothesis is rejected under either criterion. Appendix 2 again contains the relevant Stata code.

CONCLUSION

Meaningful interpretation of results from logit, probit, and other nonlinear models requires researchers not only to manage these models' technical complexities, but also to link statistical results to the substantive questions motivating a research project in the first place. In many applications, a straightforward, intuitive way to make this link is to report differences in predicted probabilities associated with theoretically meaningful or empirically relevant changes in key independent variable values, when the other independent variables are also set to theoretically meaningful or empirically relevant values. Graphical presentations implicitly embody this notion and are an efficient means of conveying logit and probit results. However, in order to draw valid statistical conclusions, researchers need also to report measures of statistical significance to accompany reported differences in predicted probabilities. The simulation-based approach to interpretation developed by King *et al*. (2000) has gained widespread popularity among political scientists because of its intuitive appeal and ease of use. Strategy researchers who make the small investment of time required to familiarize themselves with this approach stand to realize this technique's benefits as well.

ACKNOWLEDGEMENTS

Thanks to Connie Helfat, Juan Alcacer, Adam Fremeth, Witold Henisz, Glenn Hoetker, Ahmed Khwaja, Gary King, Kristin MacDonald, Edward Norton, Michael Leiblein, Myles Shaver, Charles Williams, and two anonymous referees for helpful comments, conversations or e-mail correspondences. Thanks also to Michael Leiblein and Douglas Miller for sharing their data. Any errors are my own.

REFERENCES

Ai CR, Norton EC. 2003. Interaction terms in logit and probit models. *Economics Letters* **80**(1): 123–129.

Brambor T, Clark WR, Golder M. 2006. Understanding interaction models: improving empirical analyses. *Political Analysis* **14**(1): 63–82.

Caudill SB, Jackson JD. 1989. Measuring marginal effects in limited dependent variable models. *Statistician* **38**(3): 203–206.

Friedrich RJ. 1982. In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science* **26**(4): 797–833.

Greene WH. 2003. *Econometric Analysis* (5th edn). Prentice-Hall: Upper Saddle River, NJ.

Hoetker G. 2007. The use of logit and probit models in strategic management research: critical issues. *Strategic Management Journal* **28**(4): 331–343.

Huang C, Shields TG. 2000. Interpretation of interaction effects in logit and probit analyses: reconsidering the relationship between registration laws, education, and voter turnout. *American Politics Quarterly* **28**(1): 80–95.

Imai K, King G, Lau O. 2006. Zelig: Everyone's statistical software. http://gking.harvard.edu/zelig.

Jaccard J, Turrisi R, Wan CK. 1990. *Interaction Effects in Multiple Regression*. Sage: Newbury Park, CA.

Kennedy P. 2003. *A Guide to Econometrics* (5th edn). MIT Press: Cambridge, MA.

King G, Tomz M, Wittenberg J. 2000. Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science* **44**(2): 347–361.

King G, Zeng L. 2001. Logistic regression in rare events data. *Political Analysis* **9**(2): 137–163.

Leiblein MJ, Miller DJ. 2003. An empirical examination of transaction- and firm-level influences on the vertical boundaries of the firm. *Strategic Management Journal* **24**(9): 839–859.

Liao TF. 2000. Estimated precision for predictions from generalized linear models in sociological research. *Quality & Quantity* **34**(2): 137–152.

Long JS, Freese J. 2006. *Regression Models for Categorical Dependent Variables using Stata*. StataCorp: College Station, TX.

Norton EC, Wang H, Ai CR. 2004. Computing interaction effects in logit and probit models. *Stata Journal* **4**(2): 154–167.

Oehlert GW. 1992. A note on the delta method. *American Statistician* **46**(1): 27–29.

Petersen T. 1985. A comment on presenting results from logit and probit models. *American Sociological Review* **50**(1): 130–131.

Tomz M, Wittenberg J, King G. 2001. CLARIFY: Software for interpreting and presenting statistical results, Version 2.0. http://gking.harvard.edu.

Williamson OE. 1979. Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics* **22**(2): 233–261.

Williamson OE. 1991. Comparative economic organization: the analysis of discrete structural alternatives. *Administrative Science Quarterly* **36**(2): 269–296.

APPENDIX 1. OVERVIEW OF CLARIFY COMMANDS

CLARIFY consists of the following three Stata commands, explained in detail in the software's documentation (Tomz *et al*., 2001):

`-estimsp-` is used to estimate a logit or probit model (or other supported model) and simulate the coefficients the desired number of times.

`-setx-` is used to set variable values for purposes of calculating simulated predicted probabilities or functions of the simulated predicted probabilities (or, more generally, simulated or expected dependent variable values and functions of these values).

`-simqi-` returns the simulated predicted probabilities or differences in simulated predicted probabilities (or, more generally, simulated or expected dependent variable values), along with their standard deviation and (by default) 95 percent confidence intervals.

For example, the command to estimate a logit model with the continuous independent variable *X1* and the binary independent variable *X2* and simulate the coefficients 1,000 times is:

```
. estsimp logit Y X1 X2, sims(1000)
```

In order to set *X1* to its mean and *X2* to zero and simulate the resulting predicted probability that $Y = 1$, along with its standard deviation and 95 percent confidence interval, the commands would be:

```
. setx X1 mean X2 0
. simqi, prval(1)
```

In order to estimate the change in the simulated predicted probability, along with its standard deviation and 95 percent confidence interval, when *X1* remains at its mean and *X2* changes from zero to one, the command would be:

```
. simqi, prval(1) fd(prval(1))
    changex(X2 0 1)
```

APPENDIX 2. STATA CODE

Stata code for producing Figures 2 and 3

In order to generate Figures 2 and 3, additional programming is necessary to set the relevant combinations of variable values and store the results to be plotted. The annotated code used to produce the figures appears below. A generalized version of the code that works with any data set and supports loqit, probit, Poisson, and neqative binomial models containing an interaction term can be installed in Stata by typing 'SSC install intgph' in the Stata command editor.

```
1.  set seed 9999
2.  noisily estsimp logit DVMAKE SIZE TENURE USFIRM JAPANFIRM ASIANFIRM SMALLNO SMALLNOSQ /*
3.      */ ASSETSPEC DEMANDUNC ASSETxUNC FABEXPERIENCE SOURCEXPERIENCE DIVERSIFI DIVERSSQ
4.
5.  foreach var of newlist X Y0 Y1 Y0lb Y1lb Y0ub Y1ub dY dYlb dYub {
6.      gen `var' = .
7.  }
8.
9.  forvalues obs = 1(1)18 {
10.      replace X = .01*(`obs'+1) in `obs'
11.      setx 0
12.      setx DEMANDUNC .01*(`obs'+1)
13.      foreach as_lev in 0 1 {
14.                       setx ASSETSPEC `as_lev' ASSETxUNC `as_lev'*.01*(`obs'+1)
15.                       simqi, genpr(Y`as_lev'_tmp) prval(1)
16.                       sum Y`as_lev'_tmp, meanonly
17.                       replace Y`as_lev' = r(mean) in `obs'
18.                       _pctile Y`as_lev'_tmp, p(2.5,97.5)
19.                       replace Y`as_lev'lb = r(r1) in `obs'
20.                       replace Y`as_lev'ub = r(r2) in `obs'
21.      }
22.      gen dY_tmp = Y1_tmp - Y0_tmp
23.      sum dY_tmp, meanonly
24.      replace dY = r(mean) in `obs'
25.      _pctile dY_tmp, p(2.5,97.5)
26.      replace dYlb = r(r1) in `obs'
27.      replace dYub = r(r2) in `obs'
28.      drop *_tmp
29. }
30.
31. twoway rbar Y0ub Y0lb X, mw msize(1) lcolor(gs0) fcolor(gs16)|| line Y0 X, color(gs0) || /*
32.      */ rspike Y1ub Y1lb X, color(gs0) lp(dot) || line Y1 X, color(gs0) || , yscale (r(0 1)) /*
33.      */ ylabel(0(.2)1) legend(off) xtitle("Demand uncertainty") ytitle("Pr(make)") graphregion(fcolor(gs16))
34.
35. twoway rbar dYub dYlb X, mw msize(1) lcolor(gs0) fcolor(gs16) || line dY X, color(gs0) || /*
36.      */, yscale (r(0 1)) ylabel(-.2(.2)1) legend(off) xtitle("Demand uncertainty") ytitle("dPr(make)") /*
37.      */ graphregion(fcolor(gs16))
```

Lines 1–3 seed Stata's random number generator (so that the simulated coefficients will be the same each time the code is executed), estimate Leiblein and Miller's (2003) logit model, and simulate the coefficients 1,000 times (the default value) using CLARIFY's -estsimp- command. The simulated coefficient values are stored in a series of new variables *b1*, *b2*, ..., *bn*, each corresponding to one of the coefficients in the estimated model.

Lines 5–7 create a series of new variables used to plot Figures 2 and 3. The variable *X* represents the level of demand uncertainty, which appears on the horizontal axis; the variables *Y0* and *Y1* represent the predicted probabilities for low and high asset specificity, depicted in Figure 2; the variable *dY* represents the difference between these predicted probabilities, depicted in Figure 3; and the variables whose names end in 'lb' and 'ub' respectively represent the lower and upper bounds of the 95 percent two-tailed confidence intervals for *Y0*, *Y1* and *dY*.

Lines 9–29 contain a loop that runs 18 times, once for each value of *X*, which ranges from 0.02 to 0.19 in increments of 0.01. The values used to plot Figures 2 and 3 are stored in observations 1–18 of the variables defined in lines 5–7. The macro 'obs' contains the current observation number.

Lines 10–11 assign the appropriate value to the *X* variable in the current observation, and also initially set the values of all the independent variables in the model to zero (to reflect the ambivalent firm) for the purpose of calculating simulated predicted probabilities.

DOI: 10.1002/smj

Line 12 resets the value of the *DEMANDUNC* variable to the current value of *X* for the purpose of calculating simulated predicted probabilities.

Lines 13–21 contain a loop that runs twice, once for each of the two possible values of asset specificity (zero and one). The macro ′as_lev′ contains the current value of asset specificity. (The code here could easily be adapted to assess the effect of a discrete change in the value of a continuous variable, for example, from the variable's sample mean to one standard deviation above the sample mean.)

Line 14 sets the value of the *ASSETSPEC* variable and interaction term *ASSETxUNC* in the model to their current values, based on the macros ′obs′ and ′as_lev,′ for the purpose of calculating simulated predicted probabilities.

Line 15 uses the current variable values defined by the—setx–commands on lines 11, 12, and 14 to calculate 1,000 simulated predicted probability values, one for each set of simulated coefficients. The simulated predicted probabilities are stored in the variable *Y0_tmp* or *Y1_tmp* depending on the current value of asset specificity defined by the macro ′as_lev.′

Lines 16–20 store the mean simulated predicted probability and the lower and upper bounds of the 95 percent two-tailed confidence interval for this quantity in the appropriate observation of *Y0, Y0lb, and Y0ub* (if the current value of asset specificity is zero) or *Y1, Y1lb,* and *Y1ub* (if the current value of asset specificity is one).

Lines 22–27 calculate the simulated difference in predicted probabilities at the current level of *X* along with the lower and upper bounds of the 95 percent two-tailed confidence interval for this quantity, and store these values in the current observation of *dY, dYlb,* and *dYub.*

Lines 31–33 and 35–37 produce Figures 2 and 3, respectively.

Stata code for calculating the double difference in the bottom row of Table 1

Additional programming is also required to calculate the simulated double difference used to test the null hypothesis that the effect of an increase in asset specificity when $DEMANDUNC = 0.19$ is not greater than the effect of an increase in asset specificity when $DEMANDUNC = 0.02$. The annotated code (to be run following the simulation in lines 1–3 above) is as follows.

```
1.  local valnum = 0
2.  foreach unc_lev in .02 .19 {
3.              setx demandunc `unc_lev'
4.              foreach as_lev in 0 1 {
5.                          setx assetspec `as_lev' assetxunc `unc_lev'*`as_lev'
6.                          simqi, genpr(y`as_lev'_tmp) prval(1)
7.              }
8.              gen ydif`valnum' = y1_tmp - y0_tmp
9.              local valnum = `valnum' + 1
10.             drop *tmp
11. }
12.
13. gen ddY = ydif1 - ydif0
14. sum ddY, meanonly
15. noisily display "Double difference = `r(mean)'"
16. _pctile ddY, p(2.5,5,97.5,99)
17. noisily display "95% two-tailed CI: `r(r1)', `r(r3)'"
18. noisily display "95% one-tailed CI: `r(r2)', `r(r4)'"
```

The loop in lines 2–11 cycles through the two values of *DEMANDUNC*, 0.02 and 0.19, to be used for purposes of the simulated double difference. The macro ‘unc_lev’ contains the current value of demand uncertainty.

Line 3 sets the values of *DEMANDUNC* to its current value, based on the macro ‘unc_lev,’ for the purpose of calculating simulated predicted probabilities.

Lines 4–7 contain a loop that runs twice, once for each of the two possible values of *ASSETSPEC* (zero and one). The macro ′as_lev′ contains the current value of *ASSETSPEC*. (The code here could easily be adapted to assess the effect of a discrete change in the value of a continuous variable, e.g., from the variable's sample mean to one standard deviation above the sample mean.)

Line 6 uses the current variable values defined by the -setx- commands on lines 3 and 5 to calculate 1,000 simulated predicted probability values, one for each set of simulated coefficients.

The simulated predicted probabilities are stored in the variable *Y0_tmp* or *Y1_tmp* depending on the current value of asset specificity defined by the macro 'as_lev.'

Line 13 calculates the simulated double difference in predicted probabilities, based on the simulated differences in predicted probabilities stored in *Y0_tmp* and *Y1_tmp*.

Lines 14–18 report the mean simulated double difference in predicted probabilities and the lower and upper bounds of the 95 percent two-tailed and one-tailed confidence intervals for this quantity.

APPENDIX 3. BIAS IN CONVENTIONAL PREDICTED PROBABILITY FORMULAS

For both the logit and probit models, the conventional estimator of the predicted probability $\hat{\pi}$ is a function of the estimated coefficients, $\hat{\beta}$ (King and Zeng, 2001: 137):

$$\hat{\pi} = \begin{cases} \dfrac{1}{1+\exp(-\mathrm{X}'\hat{\beta})} \text{ for logit} \\ \Phi(\mathrm{X}'\hat{\beta}) \text{ for probit} \end{cases} \tag{A1}$$

It is well known that, for large enough samples, $\hat{\beta}$ is an unbiased estimator of the 'true' population coefficients β.[15] However, this does not imply that $\hat{\pi}$ is an unbiased estimator of π, the 'true' population probability. In fact, for a given set of independent variable values, the formulas in (A1) produce biased estimates of π.

The intuition for this result is straightforward: a function of an unbiased estimator, such as $\hat{\beta}$, is generally not unbiased. For example, if the sample mean $\hat{\mu}$ is an unbiased estimate of the population mean,μ, $1/\hat{\mu}$ will not be an unbiased estimate of $1/\mu$ (King and Zeng, 2001: 150).

More formally, the bias in $\hat{\pi}$ arises from the fact that, for a given set of independent variable values, $\hat{\pi}$ represents an estimate that is itself conditional on an estimate (King and Zeng, 2001: 148):

$$\hat{\pi} = \Pr(\mathrm{Y}=1|\hat{\beta}) \tag{A2}$$

Because the bias in $\hat{\pi}$ results from uncertainty about the true value of β, an unbiased estimate of π can be obtained by averaging over $\hat{\beta}$'s distribution (King and Zeng, 2001: 148–149):

$$\check{\pi} = E_{\hat{\beta}}[\Pr(\mathrm{Y}=1)|\hat{\beta}] = \mathrm{E}[\Pr(\mathrm{Y}=1)|\beta] = \mathrm{E}[\pi] \tag{A3}$$

The simulated predicted probability yielded by King and Zeng's (2001) approach corrects for the bias in $\hat{\pi}$ because the last step in the procedure—averaging over the M simulated values of $\hat{\pi}$—is an application of equation (A3).[16]

[15] In many applications, the finite-sample bias in $\hat{\beta}$ is inconsequential when the sample size exceeds approximately 200 observations. However, in datasets with many more zeroes than ones, the bias in $\hat{\beta}$ may result in a substantively meaningful downward bias in predicted probabilities even with sample sizes in the thousands (King and Zeng, 2001: 138).

[16] A separate correction must be applied to correct the bias discussed in Footnote 15 (King and Zeng, 2001: 148).