Strategic Management Journal
Strat. Mgmt. J., **28**: 331–343 (2007)
Published online 5 February 2007 in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/smj.582
Received 30 September 2005; Final revision received 12 July 2006

THE USE OF LOGIT AND PROBIT MODELS IN STRATEGIC MANAGEMENT RESEARCH: CRITICAL ISSUES

GLENN HOETKER*
College of Business, University of Illinois at Urbana-Champaign, Champaign, Illinois, U.S.A.

The logit and probit models have become critical parts of the management researcher's analytical arsenal, growing rapidly from almost no use in the 1980s to appearing in 15% of all articles published in Strategic Management Journal *in 2005. However, a review of three top strategy journals revealed numerous areas in their use and interpretation where current practice fell short of ideal. Failure to understand how these models differ from ordinary least squares can lead researchers to misunderstand their statistical results and draw incorrect conclusions regarding the theory they are testing. Based on a review of the methodological literature and recent empirical papers in three leading strategy journals, this paper identifies four critical issues in their use: interpreting coefficients, modeling interactions between variables, comparing coefficients between groups (e.g., foreign and domestic firms), and measures of model fit. For each issue, the paper provides a background, a review of current practice, and recommendations for best practice. A concluding section presents overall implications for the conduct of research with logit and probit models, which should assist both authors and readers of strategic management research.* Copyright © 2007 John Wiley & Sons, Ltd.

INTRODUCTION

The logit and probit models have become critical parts of the management researcher's analytical arsenal, growing rapidly from almost no use in the 1980s to appearing in over 10 percent of *Strategic Management Journal* articles in the 1990s and 2000s (Shook *et al*., 2003). Thus, it is vital that our exemplar journals maintain a high standard of methodological rigor regarding their use. However, a review of recent articles using logit and probit models in 10 strategy journals revealed that current practice can be improved in multiple areas. For example, under half of the papers discussed the magnitude of the effect indicated by the estimated coefficients and, of those that did, almost one-third provided incorrect interpretations. Over half of the papers modeling an interaction between independent variables provided inappropriate or incomplete interpretations of the resulting coefficients. Two-thirds of the papers that compared coefficients across groups (e.g., foreign and domestic firms) did so in ways that can lead to meaningless results.

At issue is the fact that the logit and probit models differ from ordinary least squares (OLS) in ways that many researchers do not fully appreciate. At a minimum, this means that researchers are not conveying their findings in the most meaningful way. More critically, failure to understand these differences can lead to researchers misunderstanding their statistical results and coming to incorrect conclusions regarding the theory they are testing.

Keywords: logit; probit; statistical methods
* Correspondence to: Glenn Hoetker, College of Business, University of Illinois at Urbana-Champaign, 350 Wohlers Hall, 1206 Sixth St., Champaign, IL 61820, U.S.A.
E-mail: ghoetker@uiuc.edu

This paper identifies four critical issues in the use of logit/probit models: interpreting coefficients, modeling interactions between variables, comparing coefficients between groups (e.g., foreign and domestic firms), and measures of model fit. The four issues were selected because they were identified in the methodological literature as potentially highly consequential, appeared reasonably frequently in the published strategy research, and were incorrectly or incompletely addressed by a sizable proportion of researchers.

The next section provides a brief overview of the logit and probit models. I then describe the literature review that led to the identification of the critical issues. Next, I discuss each issue in turn, providing a background, a review of current practice, and recommendations for best practice. A concluding section summarizes and presents overall implications for the conduct of research with logit and probit models.

A BRIEF OVERVIEW OF THE LOGIT AND PROBIT MODELS

The logit and probit models are appropriate whenever modeling which of two alternatives occurs. Consider the archetypical use of the logit model, a company deciding whether to make or buy a component. A given company has some propensity to make a component, y_i^*, linearly related to a vector of observable variables, $\mathbf{x_i}$, e.g., asset specificity and uncertainty, and other factors we cannot observe, the error term, ε_i:

$$y_i^* = \boldsymbol{\alpha}\mathbf{x_i} + \varepsilon_i \tag{1}$$

When y_i^* is greater than zero, the firm decides to make the component. Of course, we cannot observe a company's propensity to make the component, only the actual choice, which we will call y_i and give a value of one when the company makes the component and zero when it buys. The probability that $y_i = 1$ is given by Equation 2, where $\boldsymbol{\beta}$ is the vector of coefficients to be estimated. Positive coefficients mean that the probability of making the component increases with that variable.

$$P(y_i = 1|\mathbf{x}_i) = \begin{cases} \dfrac{\exp(\mathbf{x'}_i\beta)}{1+\exp(\mathbf{x'}_i\beta)} & \text{for logit} \\ \Phi(\mathbf{x'}_i\beta) & \text{for probit} \end{cases} \tag{2}$$

where Φ is the cumulative density function for the standard normal.

These formulae have two favorable characteristics. First, each is limited to between 0 and 1, as appropriate for a probability. Second, the distribution of each (Figure 1) is intuitively attractive.

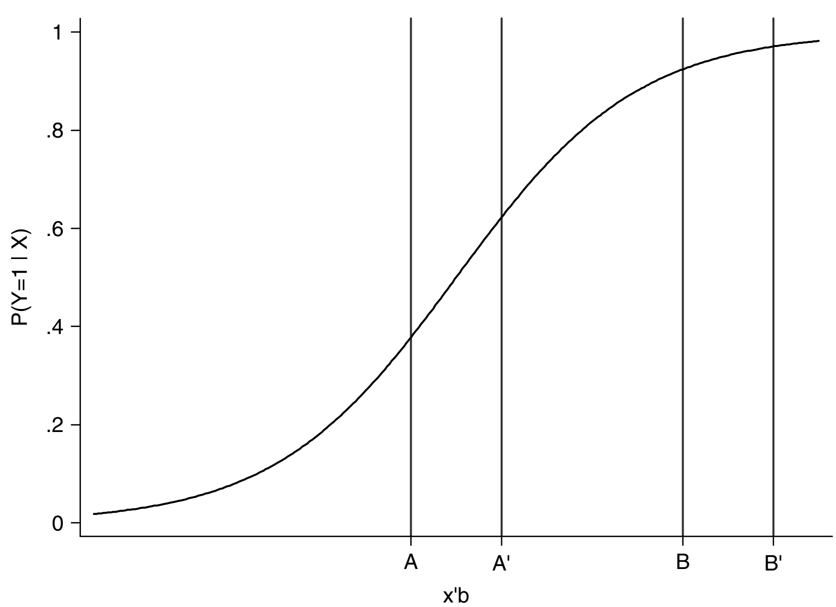

Figure 1. The logit distribution

DOI: 10.1002/smj

The impact in changes in the coefficients on the probability of an event occurring depends on the initial probability of the event. If $\mathbf{x}'\mathbf{b}$ moved from point A to point A′, the probability of the event increases from 0.4 to 0.6. However, a move of equal magnitude from point B to point B′ increases the probability of the event by a much smaller amount (roughly from 0.92 to 0.97). This makes sense: an equal change in asset specificity is much more likely to change the decision of a company with roughly equal propensities to make or buy a component than the decision of a company 90 percent likely to make it.

IDENTIFYING CRITICAL ISSUES: A REVIEW OF THE LITERATURE

I began by reviewing the relevant methodological literature. In addition to the general econometric literature, I also examined methodology papers in the fields of strategic management (e.g., Bowen and Wiersema, 2004), sociology (e.g., Allison, 1999; Long, 1997) and economics (e.g., Train, 2004), with limited references to other fields such as political science (e.g., Huang and Shields, 2000). Having identified issues that were potentially highly consequential, I then reviewed empirical work in strategic management to isolate the issues of greatest practical significance, that is, those that appeared reasonably frequently and were incorrectly or incompletely addressed by a sizable proportion of researchers. I searched articles in the empirical journals rated as outstanding or significant quality in the 'forum for business policy scholars' as identified by MacMillan (1994). Two researchers, the author and an advanced graduate student, coded the articles separately, which yielded a 95.6 percent agreement rate. Discrepancies were subsequently resolved by joint discussion. This review identified four critical issues: interpreting coefficients, modeling interactions between variables, comparing coefficients between groups (e.g., foreign and domestic firms), and measures of model fit.

In all, I searched 5384 articles. Of these, 157 used the basic logit or probit models. There is a strong upward trend in the use of basic logit and probit models, particularly among the top journals. For example, in the first half of 2005, they appeared in 15 percent of all *SMJ* articles and 12.5 percent of all *AMJ* articles. These figures omit models with more than two outcomes, e.g., multinomial logit. While the issues I identified are applicable to these models, the appropriate responses may differ from the basic logit/probit models.

Table 1 provides an overview of areas of concern. My goal is to address issues that are endemic in our field, so I deliberately refrain from identifying individual papers when discussing shortcomings. I only identify papers that offer examples of best practices and papers that meet or exceed existing norms, while providing opportunities to illustrate best practices not prevalent in the current literature. While there is general agreement on most points, there is not universal unanimity and I have noted where significant disagreement exists as to best practice.

Because my primary goal is to identify areas in which current strategy research can improve its use of the logit and probit models, I focus my analysis on the 'recent papers': those published in the period 2000 to June 2005. However, it is also useful to observe trends in the use of these models, so I provide corresponding figures for the period 1995–99.

Table 1. Overview of issues of concern in current practice (2000–2005)

Interpreting coefficients(relevant in all papers)	
Offered no interpretation of the magnitude of a variable's effect	64.9%
Offered an incomplete or incorrect interpretation of the magnitude of a variable's effect	16%
Modeling interactions between variables(relevant in 24 papers)	
Offered an incomplete or potentially incorrect interpretation	63.6%
Comparing coefficients across groups(relevant in 9 papers)	
Offered an incomplete or potentially incorrect interpretation	92%
Measures of model fit(provided in 27 papers)[a]	
Did not fully identify the measure provided	83.3%

[a] Opinions differ as to the value of providing a measure of fit.

DOI: 10.1002/smj

INTERPRETATION OF COEFFICIENTS

Background

While the nonlinear nature of the logit/probit model illustrated in Figure 1 is intuitively attractive, it also complicates interpretation of their results.[1] Researchers often merely report on the significance and sign of logit coefficients (Bowen and Wiersema, 2004). It is more useful, however, to discuss the variable's marginal effect, i.e., how much a change in a variable changes the probability of the focal outcome. However, because the effect of a change in one variable depends on the initial probability of the event occurring (equivalently, on the values of the other variables), interpreting logit/probit coefficients is more subtle than interpreting OLS coefficients. A researcher must identify meaningful values for all of the variables to calculate the impact of changes in a focal variable.

One of the most common approaches is to set the other variables at their mean (Long, 1997). However, Train argues that it is more informative to calculate the response for each observation and then average those responses, because it is unlikely that any single observation actually has the mean value of all variables (Train, 1986: 43, provides a clear graphical illustration of this point). Importantly, the response for the 'average observation' in a sample is *not* the same as the average of responses calculated for each observation, and the two can vary by up to a factor of three (Talvitie, 1976).

Alternatively, the researcher can set the other variables at some theoretically interesting values, e.g., the impact of asset specificity on the probability of vertical integration for a U.S. firm with 5000 employees. Bowen and Wiersema (2004) provide a very accessible description of the various approaches to interpreting logit/probit coefficients. Long's (1997) discussion is also very accessible and emphasizes presenting results graphically.

For the logit model, but not the probit, researchers often report the effect of a variable on the *odds ratio*, where odds of 1 : 1 mean the event is equally likely to occur or not (50% probability) and odds of 2 : 1 mean the event is twice as likely to occur than not (66.7% probability). Doing so has two advantages. First, it is easy: the effect of a one unit change in variable x is to change the odds by a factor of $\exp(\beta_x)$. Values greater than 1 increase the odds of the event occurring and values less than 1 decrease the odds. Second, this calculation applies for *all* observations and does *not* depend on the values of other variables, seemingly avoiding the complexities of the interpretation techniques discussed above (Bowen and Wiersema, 2004).

Unfortunately, a constant change in odds does not imply a constant change in probabilities. As Long (1997: 82) illustrates, if the original odds were 1 : 10, doubling the odds increases the probability from 0.091 to 0.167, a change of 0.076. On the other hand, if the original odds were 1 : 1, doubling the odds increases the probability from 0.5 to 0.667, an increase of over twice as much. Thus, the odds ratio tells us little about the effect of a covariate on the actual probability of an event occurring. Furthermore, the magnitude of the effect on probability is not symmetric around 1 (Long, 1997: 82). The positive impact of multiplying the odds of an event by 5 [$\exp(\beta) = 5$] is the same as the negative impact of dividing the odds by 5 [$\exp(\beta) = 0.2$]. It is far from intuitive that a coefficient of 5 corresponds in magnitude to 0.2, making misinterpretation by the reader likely.

Current practice

Only about one-third of the recent articles examined interpreted the magnitude of variables' effects (35.1%, 12.7% pre-2000). Disturbingly, almost half of the articles that discussed the magnitude of variables' effects did so erroneously or incompletely (45.5%, 62.5% pre-2000). Thus, only 19.1 percent of the recent articles examined provided the reader a complete and correct interpretation of the magnitude of the variables' effect (4.7% pre-2000).

Of the articles discussing the magnitude of variables' effects, 75.8 percent did so by presenting the impact of the variable on the *probability* of an event occurring (75% pre-2000). Of these, 83.4 percent followed common practice and did so at either the mean variable of the other variables or at the mean probability of an event occurring (37.5% pre-2000). Thus, each presented the response of an 'average' observation, rather than the average response for all observations as Train argues for doing.

[1] For simplicity, I will refer to logit models hereafter. Except as noted, probit models are identical.

Of those presenting interpretations, 12.1 percent presented incomplete or irrelevant interpretations (37.5% pre-2000). Mistakes included indicating that the other variables were 'held constant', but not saying at what values. Other papers held the other variables at zero, a value both far from the mean and often empirically irrelevant (e.g., setting the amount of capital invested in a subsidiary to zero). Other papers simply presented the marginal effect of a change in variable on the probability of the event occurring, with no mention of the other variables, and one declared incorrectly that the value of the other variables did not cause variation in the probability of the event occurring.

Of those presenting interpretations, 24 percent of recent papers (25% pre-2000) presented the results using odds ratios.[2] Just under half of these confused changes in odds ratios with changes in the probability or referred to changes in the 'chances' and 'likelihood' of an event occurring. Even two papers with correct interpretations had questionable aspects. Quite confusingly, one described the change of odds when the other predictors were held constant—while technically accurate, the primary advantage of the odds ratio is, of course, that the values of the other variables do not matter. The other referred to the 'multiplier,' but did not identify what was being multiplied (odds or probability).

There were some bright spots. Of recent papers that interpreted coefficients, 6.1 percent (none pre-2000) went beyond mean values to offer interpretations across a range, e.g., mean $\pm$ one standard deviation of key variables. Even more encouragingly, 18.2 percent of recent papers that interpreted the coefficient provided a graphical interpretation of the coefficients' effect (none pre-2000) (Hoang and Rothaermel, 2005; Haas and Hansen, 2005; Hansen and Lovas, 2004; Nicolaou and Birley, 2003; Park, Chen, and Gallagher, 2002). These presentations were intuitive to understand and more informative than the non-graphical interpretations.

Best practice

Although the easiest and most common approach is not to provide any interpretation of effect size, this robs the reader of important information. One possibility is to calculate the effect for several sets of theoretically interesting and empirically relevant values of the variables, rather than trying to calculate an aggregate value for the entire sample. This is both informative and avoids the bias Train discusses.

If an aggregate value is desired, the researcher can follow common practice and calculate the effect with the other variables at their mean value, but should be sure to stress that this is *not* the 'average effect.' Calculating the actual average effect per Train's advice is more informative and is straightforward to do in statistical packages with even rudimentary programmability. Whichever approach is taken, the researcher should describe it clearly and explicitly.

While the change in the odds ratio is frequently presented and easy to calculate, it is not intuitively meaningful to most readers and frequently misinterpreted, even by authors. Odds ratios seem to have little to offer either readers or researchers.

Graphic presentations can provide a richer understanding of variables' effects. Several recent papers (Folta and O'Brien, 2004) provide a variety of useful examples, which can be prepared in most statistical packages or spreadsheets.

INTERACTION TERMS

Background

Interpretation becomes more complicated when there are interactions between variables—and can often be unintuitive. Unlike OLS, the marginal effect of an interaction between two variables in a logit model is not simply the coefficient for their interaction. Indeed, the magnitude and even the sign of the marginal effect can differ across observations (Huang and Shields, 2000).

Both intuitively and mathematically, the nonlinear nature of logit/probit models demonstrated in Equation 2 and Figure 1 causes this counterintuitive result. Consider the interaction of variables x_j and x_k. Even without an explicit interaction term, the value of x_j affects the impact of x_k. Returning to Figure 1, suppose that the distances A–A′ and B–B′ each represent the increase in $\mathbf{x}'\boldsymbol{\beta}$ due to a given increase in x_j. The corresponding increase in the probability of making the component depends, of course, on whether the observation started at A or B. That, in turn, depends on the value of all the variables, including x_k. Holding all else equal, if

[2] Two papers used both methods.

an increase in x_k moved the starting point from A to B, the marginal effect of an increase in x_j would be less.

Mathematically, the interaction effect variables, x_j and x_k, is the cross-partial derivative of π_i with respect to each (Huang and Shields, 2000):

$$\frac{\partial^2 \pi_i}{\partial x_{ij} \partial x_{ik}} = \begin{cases} \phi(\mathbf{x}_i'\boldsymbol{\beta}) \bullet \dfrac{\partial^2 \mathbf{x}_i'\boldsymbol{\beta}}{\partial x_{ij} \partial x_{ik}} - \phi(\mathbf{x}_i'\boldsymbol{\beta}) \bullet (\mathbf{x}_i'\boldsymbol{\beta}) \\ \dfrac{\partial \mathbf{x}_i'\boldsymbol{\beta}}{\partial x_{ij}} \dfrac{\partial \mathbf{x}_i'\boldsymbol{\beta}}{\partial x_{ik}} \quad \text{for probit} \\ \pi_i \bullet (1-\pi_i) \bullet \dfrac{\partial^2 \mathbf{x}_i'\boldsymbol{\beta}}{\partial x_{ij} \partial x_{ik}} + \pi_i \bullet (1-\pi_i) \\ (1-2\pi_i) \bullet \dfrac{\partial \mathbf{x}_i'\boldsymbol{\beta}}{\partial x_{ij}} \dfrac{\partial \mathbf{x}_i'\boldsymbol{\beta}}{\partial x_{ik}} \quad \text{for logit} \end{cases} \tag{3}$$

Not only the magnitude but also the sign of the interaction effect can change according to the sign of $(\mathbf{x}_i'\boldsymbol{\beta})$ or $(1-2\pi_i)$.

Thus, the effect of the interaction is therefore a function of not only the coefficient for the interaction, but also the coefficients for each interacted variable and the values of all the variables. This has several important implications.

First, the sign of the interaction coefficient may not indicate the direction of the interaction effect. Instead, the entire interaction effect must be calculated at a given value. Doing so may indicate that the interaction effect is positive for some observations, null for others, and negative for yet others. Huang and Shields (2000) provide such an example from political science and Norton, Wang, and Ai (2004) provide an example from public health policy.

Second, the significance of the interaction effect cannot be determined just by the significance of the interaction coefficient. There can be a significant interaction effect for some observations even if the interaction coefficient is not significant. Conversely, even if the interaction coefficient is significant, there may not be a significant effect from some observations.

Even more than when interpreting main effects, graphical presentations can be powerful. For example, Figure 2 illustrates representative results from Leiblein and Miller's (2003) study of vertical integration decisions in the semiconductor industry. Their Table 2, Model 5, reports a coefficient of -1.73 for asset specificity (a 0/1 variable), -19.98 for demand uncertainty, and 37.612 for their interaction, where $y = 1$ corresponds to a greater probability of vertical integration. Their interpretation of the positive coefficient for the interaction term is that 'combination of asset specificity and demand uncertainty is strongly associated with vertical

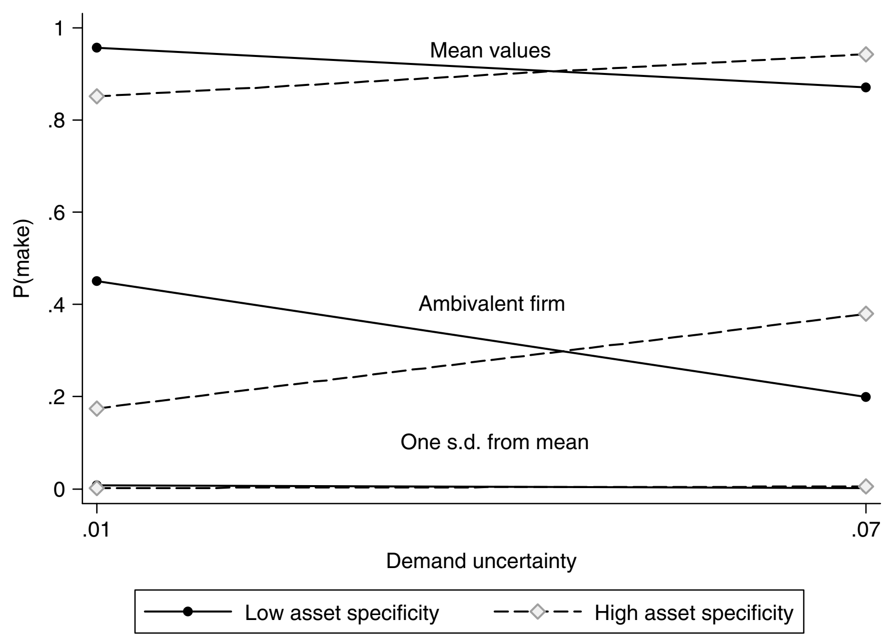

Figure 2. Interpreting interaction terms

DOI: 10.1002/smj

integration of production.' Their interpretation is correct for all observations: calculations show that the interaction effect never becomes negative.[3]

Graphing the interaction effect provides a more nuanced understanding of its practical effect. Figure 2 illustrates the effect of demand uncertainty (where uncertainty ranges to one standard deviation below/above the mean) and asset specificity for three representative firms: one with mean values of each variable, one with variables raised or lowered one standard deviation from their mean so as to minimize the probability of a make decision and a firm that is roughly ambivalent between buying and making. The graphic presentation demonstrates several facts not obvious from examining the coefficients alone. First, for firms with a low base probability of making the component, the interaction effect is very small despite the large value of the associated coefficient. Second, the interaction effect is substantial for firms that are otherwise ambivalent, but much smaller for firms with mean values of the other coefficients. Lastly, in the absence of asset specificity, uncertainty significantly *suppresses* the probability of making a component.

Current practice

Interaction terms were present in 35.1 percent of the recent papers examined (22.2% pre-2000). Most (63.6%, 85.7% pre-2000) offered no interpretation beyond the sign of the interaction term's coefficient. Those that provided purely numeric interpretations did so at a given set of values for the variables, which is an improvement, but still does not illustrate the complex behavior of the interaction effect over the range of variable values.

Encouragingly, 21.2 percent of the papers containing interactions (7.1% pre-2000) interpreted the impact of the interaction graphically. Folta and O'Brien provide a particularly nice graphical interpretation of interaction effects in a study of entry decisions. In particular, they show and report a more comprehensive sense of each interactions effect that would otherwise be possible. For example, they show that *Uncertainty* has a 'monotonically positive effect on entry into industries that require investments that are highly reversible ... Conversely, in industries where investments are more irreversible, *Uncertainty* has a negative effect on entry over 94 percent of its range' (Folta and O'Brien, 2004: 133). Haas and Hansen (2005) provide another example of an extensive and effective graphical interpretation of interaction effects.

Best practice

The first step is to be aware of the issue and not to interpret a positive (negative) sign on an interaction term as meaning that there is always an enhancing (diminishing) relationship between the two variables. Even more than when interpreting a single coefficient, a graphical presentation provides the reader with the most complete understanding of interaction's effect. If the model includes squared terms or interactions among more than two variables, a graphical presentation is almost required. Norton *et al*. (2004) provide Stata code for calculating and graphing the magnitude and significance of the interaction effect over the sample of observations. Huang and Shields (2000) demonstrate an alternative graphic presentation of interaction effects (demonstrated above), in addition to providing an exceptionally lucid explanation of the issue.

If the researcher does not want to use a graphic presentation, he or she should calculate the interaction effect of various meaningful levels of the covariates (see Huang and Shields, 2000).

COMPARING COEFFICIENTS ACROSS GROUPS

Background

Researchers often study strategies and strategic outcomes across groups. For example, Makino and Neupert (2000) asked whether the same factors lead U.S. and Japanese firms to enter overseas markets via wholly owned subsidiaries or joint ventures. Unlike OLS regression, however, comparing covariates' effects across groups is only valid if a little-noted and often violated assumption is true. For cross-group differences in logit coefficients to be meaningful, each group must have the same amount of unobserved variation, that is, the variation in outcomes beyond that explained by the independent variables (the error term ε in

[3] As the examples cited above show, there is no guarantee that this will be the case for a given model.

DOI: 10.1002/smj

Equation 1). If this is not the case, 'Differences in the estimated coefficients tell us nothing about the differences in the underlying impact of x on the two groups' (Allison, 1999: 190). Worse yet, comparisons of coefficients may inappropriately appear informative. They can reveal differences where none exist, conceal differences that do exist, and even indicate differences in the reverse direction of the actual situation.

Intuitively, the problem arises because there is no natural numeric scale for y_i*, the underlying variable—should 'very likely' map to 1 or 10? Lacking a natural scale, certain assumptions are necessary for an estimable model. These assumptions lead to the following relationship between the α terms in Equation 1 and the β terms in Equation 2:

$$\beta = \frac{\alpha}{\sigma} \tag{4}$$

where σ is the standard deviation of the error term or *unobserved variation,* ε. If we could identify σ, we could calculate α, our real theoretical interest, for a given estimate of β. Unfortunately, σ is unobservable (Train, 2004: 44). Monte Carlo simulations show that even small differences in unobserved variation can lead to finding significant differences in β coefficients across groups with identical α coefficients (Hoetker, 2004). Fortunately, Allison (1999) developed an easy-to-apply test of whether unobserved variation differs across groups. Further, if unobserved heterogeneity does differ, Allison also offers a simple model for testing whether *at least one* of the coefficients differs across groups. This approach does an excellent job of indicating when the underlying α coefficients do *not* vary across groups, even if the estimated β coefficients do (Hoetker, 2004).

Researchers commonly investigate cross-group differences in a variable's effect by interacting it with a dummy variable for group membership and estimating the resulting equation for all observations. To test the effect of x (ignoring other variables for clarity), this approach involves setting the dummy variable, G_i, to 1 for observations in Group 1, estimating

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i + \beta_2 G_i + \beta_3 (G_i{}^* x_i) + \varepsilon_i \tag{5}$$

and observing the significance and sign of β_3. Notice, however, that the presence of a single error term, ε_i, forces the unobserved variation for both groups to be the same (Darnell, 1994: 111; Pindyck and Rubinfeld, 1991: 107). Even in OLS, using an interaction term in a single equation when variance differs between groups means the slope coefficients will be found *not* to differ, even if they actually do (Gujarati, 1988: 527).

In the logit case, the outcome can be even worse. Equation 5 will often fail to detect true differences in β_2 and, very disturbingly, may incorrectly find a significant result of the sign opposite the actual relationship (Hoetker, 2004). Intuitively, the problems noted by Gujarati are aggravated because artificially restricting the unobserved variation to be the same also affects the scale of all of the coefficients.

If the model is estimated separately for each group, the researcher can—at a minimum—compare the statistical significance of the coefficients across groups. This is possible because the coefficients and standard errors are consistent *within each group*. One could report, for example, that x has a significant and positive impact for Group 1, but is not significant for Group 2. Obviously, such a statement is more informative if the samples are of roughly the same size, the model appears well specified and the p-values do not straddle a particular significance level, e.g., 0.09 for one group and 0.13 for the other.

Alternatively, suppose we could frame our interest not as whether the absolute effect of a variable x_k differed across groups, but rather as whether the impact of x_k relative to x_j differs across groups. We can do so by comparing the ratio β_k/β_j across groups (Train, 1998: 237). Because β is the underlying coefficient, α, scaled by the standard deviation of the unobserved variation, σ, we find that

$$\frac{\beta_k}{\beta_j} = \frac{\alpha_k \big/ \sigma}{\alpha_j \big/ \sigma} = \frac{\alpha_k}{\sigma}\left(\frac{\sigma}{\alpha_j}\right) = \frac{\alpha_k}{\alpha_j} \tag{6}$$

By taking a ratio, we have removed the impact of unobserved variation and are left with a ratio of the variable's underlying effects, which can now be compared across groups. The statistical significance of the difference in the ratios across groups can be computed with a Wald chi-squared test (Greene, 2000).[4] Unfortunately, even large

[4] In Stata, a combination of the *suest* and *testnl* commands simplifies this process, but the necessary computations are possible

DOI: 10.1002/smj

differences in ratios may not be statistically significant, especially if one or more terms are estimated with poor precision.

Hoetker (2006) uses this method to compare the benefits of prior experience with a supplier relative to the value of the patents a supplier held for low- and high-uncertainty transactions. The ratio $\beta_{years\ prior\ experience}/\beta_{patents}$ was 50.7 for low-uncertainty and 272.2 for high-uncertainty transactions. Thus, under low uncertainty, a buyer values a year of prior experience with a supplier as much as 51 patents. Under high uncertainty, a buyer will be willing to give up almost five times as much technical capability for an additional year of prior experience with a supplier as under low uncertainty.

Current practice

Comparison of coefficients across groups occurred in 10.6 percent of recent papers (14.3% pre-2000). Of those that compared magnitudes across groups, 50 percent (11.1% pre-2000) used interaction terms. None compared the unobserved variation across groups, potentially calling their conclusions into question. One paper compared the impact on the odds ratio across groups, which is econometrically valid but deceptive if the mean probability for an event differs across groups as discussed above in the section, 'Interpretation of coefficients'.

Best practice

Evidence from simulations strongly suggests that researchers should not use an interaction term to compare groups unless there are compelling theoretical reasons to believe that the unobserved variation is the same across groups. It is better to estimate separate equations for each group, first testing for differences in unobserved variation. This will require a change in current practice, but Allison's test is simple to run in any statistical package with programming capabilities.[5]

If a difference in unobserved variation is found, the researcher has several options. Allison's test for at least one coefficient differing between groups is powerful, makes few assumptions, and leads appropriately to more conservative results. The researcher can compare the ratio of coefficients across groups regardless of differences in unobserved variation, if theoretically relevant. Lastly, the researcher can compare the statistical significance of the coefficients across groups.

In terms of data requirement, researchers should gather as complete a set of covariates as possible, because controlling for more variation leaves less unobserved variation to vary across groups. The ratio of coefficients technique requires a sizable sample.

Ultimately, however, researchers may simply not be able to conduct some of the comparisons they are accustomed to doing in the linear setting. While this is frustrating, no results are surely superior to spurious results.

MEASURES OF MODEL FIT

Background

In OLS regressions, it is common to provide a measure of how well the model fits the data, such as R^2. Unfortunately, no direct equivalent to R^2 exists for logit models. A wide range of pseudo-R^2 measures have been proposed. However, these measures have different formulae and will take different values for the same model. For example, Veall and Zimmermann (1996) describe a model in which McFadden's pseudo-R^2 was 0.25, while the McKelvey–Zavoina pseudo-R^2 was 0.5.

The plethora of pseudo-R^2 measures, a *subset* of which are presented in Table 2, leads to the first opportunity to go astray in their use. Authors will often simply report that a model has a 'pseudo-R^2 of 0.45' without identifying which pseudo-R^2 they are reporting. Without that information, the reader can neither interpret the meaning of the measure nor compare it to similar models in other papers.

The second opportunity for misinterpretation is that none of the measures corresponds to the 'percent of variance explained', as R^2 does in OLS.

in almost any statistical package (for detailed examples, see Weesie, 1999). See also the author's *complogit* package for Stata (see footnote 5).

[5] Allison's (1999) paper provides code to perform his tests for multiple statistical programs. In his original code for Stata, replace '\$ml_y1' with '\$ML_y1', noting the capitalization. See also the author's *complogit* package for Stata, which performs all of the tests described in this section. To install from within Stata, type 'ssc install complogit'. 'Help complogit' will then present a comprehensive help file. Stata version 8 or greater is required.

Table 2. Popular pseudo-R^2 measures

Measure	Formula	
Efron's pseudo-R^2	$1-\frac{\sum_{i=1}^{N}(y_i-\hat{\pi}_i)^2}{\sum_{i=1}^{N}(y_i-\overline{y}_i)^2}$,	where $\hat{\pi}$ is the predicted probability
McFadden's pseudo-R^2	$1-\frac{\ln \hat{\mathrm{L}}_{\mathrm{U}}}{\ln \hat{\mathrm{L}}_{\mathrm{R}}}$	where $\hat{L}_U$ and $\hat{L}_R$ are the likelihood of the model with and without regressors respectively
McFadden's adjusted pseudo-R^2	$1-\frac{\ln \hat{\mathrm{L}}_{\mathrm{U}}-K}{\ln \hat{\mathrm{L}}_{\mathrm{R}}}$	where K is the number of regressors
McKelvey and Zavoina's pseudo-R^2	$\frac{\hat{\mathrm{var}}(\hat{y}^*)}{\hat{\mathrm{var}}(\hat{y}^*)+\mathrm{var}(\varepsilon)}$	where $\hat{\mathrm{var}}(\hat{y}^*)=\boldsymbol{\beta}'\hat{\mathrm{var}}(x)\boldsymbol{\beta}$ and $\mathrm{var}(\varepsilon)=\pi^2/3$ for logit and $\mathrm{var}(\varepsilon)=1$ for probit
Maximum likelihood pseudo-R^2	$1-\left(\frac{\ln \hat{\mathrm{L}}_{\mathrm{R}}}{\ln \hat{\mathrm{L}}_{\mathrm{U}}}\right)^{2/N}$	where N is the number of observations
Cragg and Uhler's pseudo-R^2	$\frac{1-[\mathrm{L}_{\mathrm{R}}/\mathrm{L}_{\mathrm{U}}]^{2/N}}{1-L_{\mathrm{R}}{}^{2/N}}$	
Nagelkerke's pseudo-R^2	Identical to Cragg and Uhler's pseudo-R^2	

Intuitively, because we cannot observe the underlying latent variable y^*, it is not possible to calculate what percentage of its variance a model explains. Most pseudo-R^2 measures have no intuitive interpretation for values other than 0 or 1. For example, a McFadden's pseudo-R^2 of 0.6 indicates a 60 percent increase in the log-likelihood function—a figure without obvious meaning.

A popular measure of fit that seems to provide an intuitive sense of the model's quality is the proportion of 'correct predictions.' This measure is problematic because it can give the false impression that the model is predicting well, when it actually is not (Veall and Zimmermann, 1996). By setting the prediction for each observation to the most frequent outcome, one can always achieve at least 50 percent accuracy. A simple adjustment (see Long, 1997: 108) yields an adjusted count R^2, which gives the improvement in correct predictions beyond the number that would be correctly predicted simply by predicting the most common event for all observations. However, the unadjusted proportion of correct predictions is often reported, potentially giving an overly optimistic sense of the model's fit, but also potentially underestimating the improvement in fit across models.

As an example of the benefits of the adjusted count R^2, consider Table 2 of Leiblein and Miller's (2003) study of the make-or-buy decision in the semiconductor industry. They report that a model with just control variables accurately predicted 81 percent of all observations and their full theoretical model 90.2 percent. Their theoretical model seems extremely powerful, although not that large an improvement over their control variables. When we control for the fact that 76 percent of all outcomes were 'make,' a different picture develops. Their full model is actually a 57 percent improvement in prediction over simply predicting 'make' for every observation: impressive, but not quite as impressive as 90 percent. However, the controls-only model was merely a 20 percent improvement over always predicting 'make,' so their theoretical model was actually a much more substantial improvement over the controls than it originally appeared.

Other authors (e.g., Train, 1986) argue that the entire idea of a 'correct prediction' reflects a misunderstanding of what the predicted possibility actually represents. The researcher is not predicting the actual choice the actor will take in a given instance. Rather, he or she is predicting the proportion of the time that a given choice would be taken in repeated trials (or by many people with the same characteristics). If the predicted probability of a 'make' decision occurring is 75 percent, then in over 100 trials we would expect it to occur 75 times, with 25 occurrences of the 'buy' decision (Long, 1997: 41–42). If we actually observed that pattern, the model would have performed perfectly, but the 25 occurrences of 'buy' would be treated as mistakes in calculations of 'percentage of correct prediction.' Additionally, consider two observations for which the predicted probability of 'make' was 50.1 percent and 99.9 percent, respectively. If we observed a 'buy' decision, each observation would count as one 'miss' even though the second represented a much larger surprise (Aldrich and Nelson, 1984).

Current practice

Roughly a quarter of recent papers provided a measure of model fit (25.5%, 12.7% pre-2000). With the exception of one paper that presented the Bayesian Information Criterion, the papers used various pseudo-R^2 measures. Unfortunately, 83.3 percent (75% pre-2000) of those papers did not identify the specific pseudo-R^2. Among recent papers, 10.6 percent (23.8% pre-2000) used percentage of correct predictions as a measure of model fit, but only 20 percent of these (26.7% pre-2000) scaled the percentage of correct predictions as discussed above.

Best practice

Some researchers argue for the routine inclusion of a measure of model fit (Bowen and Wiersema, 2004: 98), while others are more skeptical as to their value (e.g., Long, 1997). If providing a measure of fit, it should be fully identified, e.g., McFadden's pseudo-R^2. Citing a source that explains the derivation and interpretation of the measure is helpful, as even well-informed readers may not be familiar with a given measure.

The proportion of correct predictions is somewhat suspect as a measure of model fit, but the appropriately adjusted version is a vast improvement over the unadjusted proportion. Veall and Zimmermann (1996) document multiple alternative measures of fit based on predictive accuracy.

Lastly, the researcher must not imply any meaning for a measure beyond what it actually represents. Saying that a *pseudo*-R^2 'corresponds to the R^2 in OLS' is overly simplistic and invites misinterpretation of the measure.

SUMMARY

The logit and probit models are increasingly important for the strategy researchers. Almost any statistical package can estimate the models, making their use increasingly routine. Comparing use of the models since 2000 to that before 2000 reveals significantly increased sophistication in their use and presentation. In particular, researchers increasingly interpret the models graphically, which is very valuable given the nonlinearity of the models. However, growth in their application has outstripped growth in expertise in their use, which could lead to significant misunderstanding of empirical results and the associated theoretical relationships.

Fortunately, the difficulties posed by these challenges can be largely ameliorated with the suggested changes to current practice and techniques, all of which are well within the capabilities of most statistical packages. Table 3 summarizes recommendations for doing so. Several common themes appear in the best practices suggested by the literature. First, offer explicit, detailed explanations—what assumptions are behind a calculation of effect size? Which pseudo-R^2 is presented? Second, offer interpretations that correspond to theoretically interesting and empirically relevant cases—which may not be the mean values of the variables. Third, recognize the advantages of presenting information graphically. Lastly, do not rely overly on intuition from OLS models.

Progress in understanding and applying logit and probit models will continue to be scattered across multiple disciplines. While this creates challenges for the individual researcher, it also creates the opportunity for the strategy field, by virtue of its interdisciplinary nature, to advance the sophistication with which these models are applied.

ACKNOWLEDGEMENTS

I gratefully acknowledge the helpful comments of Paul Allison, Rajshree Agarwal, Jongwook Kim, Tim Liao, and Steve Michael. Special thanks go to Editor-in-Chief Dan Schendel and two anonymous

Table 3. Summary of best practice recommendations

Interpretation of coefficients
Use graphical presentations when possible.
Interpret changes in probabilities at meaningful values of the variables.
Be leery of odds ratios.
Consider calculating the average effect over all observations, rather than the effect on the 'average' observation.
Interaction terms
Go beyond interpreting sign and significance of the coefficient for the interaction term.
Use graphical presentations when possible. Their advantages are even greater than in the absence of interactions.
Comparing coefficients across groups
Do not interact variables of interest with dummies indicating group membership. Estimate each group separately.
Test for equality of unobserved variance between groups.
Use robust means of comparison if unobserved variance is unequal, such as Allison's (1999) test.
Compare ratios of coefficients if theoretically relevant.
Measures of model fit
Present a measure of model fit or not. Both approaches have proponents.
Identify fully any measure you are presenting. Prefer 'Efron's (1978) *pseudo*-$R^2 = 0.5$' to '*pseudo*-$R^2 = 0.5$'.
Remember that *pseudo*-R^2 measures do *not* correspond to the OLS R^2.
Be cautious about measuring model fit by the percentage of correct predictions. If you do so, scale to account for the proportion of the most frequent event in the sample.

referees whose counsel led to a much more comprehensive article. Martin Ganco provided valuable research assistance.

REFERENCES

Aldrich JH, Nelson FD. 1984. Linear probability, logit, and probit models. Sage University Papers Series, no. 07–045. Sage: Beverly Hills, CA.

Allison PD. 1999. Comparing logit and probit coefficients across groups. *SMR/Sociological Methods and Research* **28**(2): 186–208.

Bowen HP, Wiersema MF. 2004. Modeling limited dependent variables: methods and guidelines for researchers in strategic management. In *Research Methodology in Strategy and Management*, Ketchen DJ, Bergh DD (eds). Elsevier: New York; 87–134.

Darnell AC. 1994. *A Dictionary of Econometrics*. Edward Elgar: Brookfield, VT.

Efron B. 1978. Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association* **73**: 113–121.

Folta TB, O'Brien JP. 2004. Entry in the presence of dueling options. *Strategic Management Journal* **25**(2): 121–138.

Greene WH. 2000. *Econometric Analysis* (4th edn). Prentice-Hall: Upper Saddle River, NJ.

Gujarati D. 1988. *Basic Econometrics* (2nd edn). McGraw-Hill: New York.

Haas MR, Hansen MT. 2005. When using knowledge can hurt performance: the value of organizational capabilities in a management consulting company. *Strategic Management Journal* **26**(1): 1–24.

Hansen MT, Lovas B. 2004. How do multinational companies leverage technological competencies? Moving from single to interdependent explanations. *Strategic Management Journal*, Special Issue **25**(8–9): 801–822.

Hoang H, Rothaermel FT. 2005. The effect of general and partner-specific alliance experience on joint R&D project performance. *Academy of Management Journal* **48**(2): 332–345.

Hoetker G. 2004. Confounded coefficients: extending recent advances in the accurate comparison of logit and probit coefficients across groups. University of Illinois College of Business Working Paper 03–0100. http://ssrn.com/abstract=609104 [14 August 2006].

Hoetker G. 2006. Do modular products lead to modular organizations? *Strategic Management Journal* **27**(6): 501–518.

Huang C, Shields TG. 2000. Interpretation of interaction effects in logit and probit analyses: reconsidering the relationship between [U.S.] registration laws, education, and voter turnout. *American Politics Quarterly* **28**(1): 72–79.

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.

Long JS. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage: Thousand Oaks, CA.

MacMillan I. 1994. The emerging forum for business policy scholars. *Journal of Business Venturing* **9**(2): 85–89.

Makino S, Neupert KE. 2000. National culture, transaction costs, and the choice between joint venture and wholly owned subsidiary. *Journal of International Business Studies* **31**(4): 705–713.

Nicolaou N, Birley S. 2003. Social networks in organizational emergence: the university spinout phenomenon. *Management Science* **49**(12): 1702–1725.

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

Park SH, Chen R, Gallagher S. 2002. Firm resources as moderators of the relationship between market growth and strategic alliances in semiconductor start-ups. *Academy of Management Journal* **45**(3): 527–545.

Pindyck RS, Rubinfeld DL. 1991. *Econometric Models and Economic Forecasts* (3rd edn). McGraw-Hill: New York.

Shook CL, Ketchen DJ, Cycyota CS, Crockett D. 2003. Data analytic trends and training in strategic management. *Strategic Management Journal* **24**(12): 1231–1237.

Talvitie A. 76. Disaggregate travel demand models with disaggregate data, not with aggregate data, and for what. Working paper 7615, Institute of Transportation Studies, Urban Travel Demand Forecasting Project, University of California, Berkeley.

Train K. 1986. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. MIT Press: Cambridge, MA.

Train KE. 1998. Recreation demand models with taste differences over people. *Land Economics* **74**(2): 230–240.

Train KE. 2004. *Discrete Choice Methods with Simulation*. Cambridge University Press: Cambridge, U.K.

Veall MR, Zimmermann KF. 1996. Pseudo-R^2 measures for some common limited dependent variable models. *Journal of Economic Surveys* **10**(3): 241–259.

Weesie J. 1999. Seemingly unrelated estimation: an application of the cluster-adjusted sandwich estimator. *Stata Technical Bulletin*, No. 52: 19–23.