Vlerick Leuven Gent Management School

the Autonomous Management School
of Ghent University and Katholieke Universiteit Leuven

Vlerick Leuven Gent Working Paper Series 2003/30

MODELLING LIMITED DEPENDENT VARIABLES: METHODS AND GUIDELINES FOR RESEARCHERS IN STRATEGIC MANAGEMENT

HARRY P. BOWEN

Harry.Bowen@vlerick.be

MARGARETHE F. WIERSEMA

D/2003/6482/31

MODELLING LIMITED DEPENDENT VARIABLES: METHODS AND GUIDELINES FOR RESEARCHERS IN STRATEGIC MANAGEMENT*

HARRY P. BOWEN

Vlerick Leuven Gent Management School

MARGARETHE F. WIERSEMA

Graduate School of Management,

University of California - Irvine

Contact:

Harry P. Bowen

Vlerick Leuven Gent Management School

Vlamingenstraat 83, 3000 Leuven, Belgium

Tel : +32 16 32.35.86

Fax : +32 16 32.85.81

email: harry.bowen@vlerick.be

* Forthcoming in *Research Methodology in Strategy and Management*, D. Bergh and D. J. Ketchen, Jr, Series Co-Editors, Elsevier Press, 2004.

ABSTRACT

Research on strategic choices available to the firm are often modeled as a limited number of possible decision outcomes and leads to a discrete limited dependent variable. A limited dependent variable can also arise when values of a continuous dependent variable are partially or wholly unobserved. This chapter discusses the methodological issues associated with such phenomena and the appropriate statistical methods developed to allow for consistent and efficient estimation of models that involve a limited dependent variable. The chapter also provides a road map for selecting the appropriate statistical technique and it offers guidelines for consistent interpretation and reporting of the statistical results.

INTRODUCTION

Research in strategic management has become increasingly sophisticated and more specialized in terms of the range and depth of issues addressed and the theoretical frameworks applied. However, methodological rigor has often not kept pace with theoretical advances. Several areas of weakness with respect to statistical methods employed in past strategy research, as well as methodological issues such as the validity of measures, have recently been the subject of a number of articles (Bergh and Fairbank, 1995; Bergh and Holbein, 1997; Bowen and Wiersema, 1999; Lubatkin, Merchant, and Srinivasan, 1993; Robins and Wiersema, 2003). The recent concerns raised about statistical and methodological issues are well-founded since the use of appropriate statistical techniques is critical for generating valid statistical conclusions (Scandura and Williams, 2000). This chapter adds to this stream of methodological introspection by examining a set of statistical issues likely to arise in the analysis of strategic choice at the firm level. In particular, in such settings the researcher is often faced with a limited dependent variable (LDV) that takes a limited number of (usually discrete) values. In such cases discrete LDV methods such as Logit and Probit are used since the use of ordinary Least Squares (OLS), the most common statistical technique used in management research,[1] will produce biased and inconsistent estimates of model parameters.

The use in strategy management research of methods such as Logit and Probit has increased significantly in recent years.[2] Despite the growing popularity of such methods, there appears to be widespread problems in the application and interpretation of these methods within the literature. One frequent problem is the use of an inappropriate research design to examine the phenomenon of interest. For example, strategy researchers interested in explaining strategic choices often model such choices as a simple binary dependent variable. Given the wide array of strategic alternatives considered by the firm's management, a binary construct may not adequately capture the full set of choices available. In addition, a review of studies that utilize LDV methods indicates that researchers often present incomplete or inconsistent analytical results. In many cases the researcher limits their interpretation of results to the significance and direction of an explanatory variable without any attempt to assess the magnitude of the effect that an explanatory variable has on the dependent variable. As discussed here, the sign and magnitude of a coefficient estimated in a LDV model is almost never an accurate guide to the

[1] OLS is used by 42% of all research studies in management (Scandura and Williams, 2000).
[2] In a review of the articles appearing in the *Strategic Management Journal* we found LDV techniques used in twelve articles in 2002 versus four articles in the 1999.

direction and magnitude of the underlying relationship between the dependent variable and an independent variable. The problems evident in the past use of LDV techniques provides the basis for highlighting here what researchers need to know when modeling a discrete LDV.

While a LDV can arise because the strategic choices themselves are represented by a limited number of discrete options, more subtle instances of a LDV arise when values of a dependent variable are censored or truncated. A censored dependent variable occurs when values of a variable above or below some threshold value are all assigned the same value. An equivalent form of censoring is when the phenomenon of interest exhibits a significant number of observations for which the dependent variable takes only a single value. An example of the latter could arise in a study of the level of firm diversification since the diversification measure computed for single business firm takes a single common value.

A truncated dependent variable arises when values of the dependent variable are excluded from the sample, either by choice of the researcher to use a (non-randomly) selected subset of the population of firms or because some firms in the population are not observed unless another variable is observed. The latter case is known as the "sample selection problem," and if not properly handled leads to a "sample selection bias." An example of this might be a study of the performance of firms in a joint venture in relation to their level of equity participation. Since firms first make the decision to undertake a joint venture, only firms undertaking a joint venture will be observed in the sample. If one does not account for how a firm "selects" itself to enter into a joint venture, and hence to be observed in the data sample, the estimated coefficients in the performance equation may be biased.

The cases of a LDV that arise from censoring, truncation, or particular forms of nonrandom sample selection have received little attention in the empirical strategic management literature. However, these cases are potentially a widespread problem with respect to the issues commonly studied by researchers in strategic management. The issue of bias that arises from the sample selection problem is, in particular, a problem that we feel has been severely neglected in strategy research, as evidenced by the almost non-existent use in the literature of the techniques that deal with this problem.[3]

This chapter highlights statistical methods that allow for consistent and efficient estimation of models involving a LDV that arises from an underlying model of choice, or from censoring, truncation, or non-random sampling. We first discuss some research design issues associated with a discrete LDV and offer a roadmap for selecting the appropriate statistical technique in such cases. We then follow with a detailed discussion of the most common

[3] For example, the Sample Selection model discussed later has rarely appeared in published research.

techniques used to model a discrete LDV that arises in a choice based framework, and a continuous LDV that arises from censoring, truncation or nonrandom sampling. Where appropriate, our discussion concludes with an overview, in table format, of key elements regarding the use and interpretation of alternative methods. These elements include the statistical assumptions underlying a technique, what to report when presenting results, and how the results can be interpreted. Our hope in raising awareness of the statistical, methodological, and interpretation issues for the most common LDV models is that strategic management researchers who adopt such models will utilize appropriate research designs, standardize their presentation and interpretation of results, and ultimately conduct analyses that offer sound and statistically correct conclusions.

RESEARCH DESIGN ISSUES

A crucial aspect of any empirical research is to develop a research design to understand the phenomenon of interest and to guide the selection of an appropriate statistical method. A first step toward the choice of statistical method is deciding what measure of the dependent variable can best represent the concept of interest. To arrive at the appropriate measure, the researcher will need to determine the range of variation of the phenomenon of interest, the nature of its distribution, and how fine or gross to make the distinction between particular attributes of the phenomenon. It is these considerations, in conjunction with the purpose of the research, that drive the final choice of measure for the dependent variable. It is essential that the dependent variable be well-measured, well-distributed, and have enough variance so that there is indeed something to explain.

For many strategy phenomena there can exist numerous ways the construct of interest can be operationalized and thus measured. If one is interested in whether or not a firm engages in a specific activity (e.g. to invest overseas or not) then a simple binary outcome may be appropriate. However, a firm (or rather its managers) rarely faces a binary decision choice. More likely, there is an array of options for deciding to engage in a particular activity (e.g. the decision to invest overseas can occur through joint venture, strategic alliance, acquisition, or Greenfield). Our review of LDV studies conducted in the strategic management literature revealed a predominate use of a binary dependent variable. Yet based on the phenomenon of interest this rarely seemed appropriate. In many cases researchers collapsed richer data into a simple binary decision or they insufficiently identified and measured the variation in the phenomenon of interest. For example, one study (Toulan, 2002) examined the scope of outsourcing by operationalizing the outsourcing decision as a simple binary choice (increase vs. decrease in outsourcing activities). Yet it was clear from the study that most firms increased their extent of outsourcing and that the extent and type of activities being outsourced varied widely. This was not captured by the simple binary dependent variable. If managers do not view their strategic choices as binary, then why should researchers?

In other studies, researchers gathered survey data on multiple items along a Likert scale but then collapsed the data into two extremes (high and low) to arrive at a binary dependent variable. In such cases the use of a binary variable is throwing away valuable information about the phenomenon of interest. If the phenomenon of interest occurs along a range of variation then the phenomenon should be operationalized to minimize loss of pertinent information and increase the predictive power of the model. The extent of variation lost by collapsing the data depends on the number of categories selected for the new (i.e.: collapsed) variable; the fewer the number of categories the more variation lost. The researcher's ability to understand and explain the phenomenon of interest can thus be compromised if the dependent variable is operationalized using too gross a categorization when recoding data. To capture the complete range of decision outcomes, an ordinal or interval scaled dependent measure may allow the researcher to provide much greater explanation.

Once the researcher has operationalized the concept of interest as a discrete dependent variable, other issues will determine the choice of appropriate statistical technique. The flow chart given in FIGURE 1 can help in this regard. This asks a series of questions on the nature of the data and based on the answers to these questions, the chart leads to a statistical technique appropriate to the research situation.

Insert Figure 1 About Here

The discussion of LDV models that follows implicitly assumes that they will be applied to a cross-sectional data sample. While cross-sectional data is most often used to estimate LDV models, there is nothing to prevent one from applying these models to a longitudinal data set. More generally, these models can also be estimated using panel (cross-section, time-series) data. One limitation that arises in a panel data setting is that, for some models, one cannot model heterogeneity across cross-section units (firms). For example, in a standard regression analysis that uses a panel data set on a number of firms over time one might model differences across firms using a set of dummy variables that allow the model's intercept to vary across firms (see Bowen and Wiersema (1999) for discussion of regression models in a panel data setting). This type of modeling is not possible for some of the models discussed here (e.g., Multinomial Probit) due to statistical issues. If one has panel data and wants to model heterogeneity across firms using, for example, dummy variables, then one is encouraged to consult more advanced presentations (e.g., Greene, 2002, Chapter 21) of the LDV models discussed here before proceeding.

CHOICE BASED LIMITED DEPENDENT VARIABLES

This section discusses models for the predominant case of a LDV that arises from an underlying model of discrete choice by the firm. We begin with the most frequently used LDV models in the empirical strategy literature, the binary Logit and Probit models. In these models the dependent variable takes one of two values, either a 0 or a 1. As we will discuss, the use of OLS to examine such a dependent variable is not appropriate. Our discussion of these binary choice models serves to introduce notation, summarize underlying assumptions, and to indicate a desired framework for the presentation and interpretation of results. Following this, we discuss more general models of choice among multiple alternatives, where these choices can be unordered or ordered. An example of an unordered set of choices would be the mode of entry into a new market (e.g., Greenfield, acquisition, or joint venture). An example of an ordered set of choices would be discrete levels of equity participation (e.g., low, medium, high) for a firm entering a joint venture. The basic methods of interpretation and analysis for the binary models will, in most cases, also apply the more general multiple choice models.

Binary Outcomes

Strategic decisions involving only two choices (outcomes) are the most common type of LDV studied in strategy research. Examples include a firm's choice of whether or not to strategically refocus its corporate portfolio (Chatterjee, et al, 2003); enter a new market by acquisition or internal expansion (Chang, 1996; Chang & Singh, 1999); expand overseas via a start-up or acquisition (Vermeuelen & Barkema, 2001); exit an existing market via divestiture or dissolution (Chang & Singh, 1999); or enter into a strategic alliance (Gulati, 1999; Chung, et al, 2000).

The two models commonly used to model binary choice are the binary Logit and binary Probit models. Which model one chooses is largely arbitrary. In practice, the models produce the same qualitative results, and there is a fairly well established relationship between the coefficients estimated from the two models. A distinct advantage of the Logit model is that the results are easily interpretable in terms of the odds in favor of one choice versus the other, and how these odds change with changes in an independent variable. In contrast, calculating changes in odds from a Probit model requires a number of indirect calculations.

Model Specification

To understand the development of the binary Logit and Probit models we first consider the problems that arise if a standard regression approach is used to model a binary dependent variable. Let y be the dependent variable of interest. By assumption, y takes only two values, 0 or 1, where the value y = 1 represents a choice of one of the two outcomes. The researcher is interesting in explaining the observed choice and proceeds to specify a set of explanatory variables. Let **x** be a vector of k explanatory variables plus a constant term **x** = $(1, x_1, \ldots, x_k)$ where "1" represents the constant term, and denote the probability of outcome "*A*" as Pr(*A*). The probability of outcomes y = 1 and y = 0, conditional on **x**, can then be written

$$\Pr(y=1 \mid \mathbf{x}) = F(\mathbf{x}, \boldsymbol{\beta})$$
$$\Pr(y=0 \mid \mathbf{x}) = 1 - F(\mathbf{x}, \boldsymbol{\beta}) \tag{1}$$

In (1), b is a vector of k+1 coefficients $(\beta_0, \beta_1, \ldots, \beta_k)$ and F(**x**, b) is some function of the variables **x** and parameters b. Since y takes only the values 0 or 1, the conditional expectation (conditional mean) of y, denoted E[y | **x**], is simply the probability that y = 1:

$$E[y \mid \mathbf{x}] = [1 \times \Pr(y=1 \mid \mathbf{x}) + 0 \times \Pr(y=0 \mid \mathbf{x})]$$
$$E[y \mid \mathbf{x}] = \Pr(y=1 \mid \mathbf{x}) = F(\mathbf{x}, \boldsymbol{\beta}) \tag{2}$$

The standard regression model postulates that the conditional mean of the dependent variable is a linear function of **x**, that is, $E[y \mid \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$. Adopting this specification gives the *Linear Probability Model* (LPM):

$$y = E[y \mid \mathbf{x}] + e$$
$$y = \mathbf{x}'\boldsymbol{\beta} + e \tag{3}$$

where ε is the error (i.e., ε = y – E[y | **x**]). From (2), setting $E[y \mid \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}$ implies that $F(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}'\boldsymbol{\beta}$. But since the value of F(**x**, b) is the probability that y = 1, one problem with the LPM is immediately clear: nothing guarantees that values of $\mathbf{x}'\boldsymbol{\beta}$ will lie between 0 and 1.

Hence, given estimates **b** of the b, there is nothing to prevent $\mathbf{x}'\mathbf{b}$ from yielding predicted probabilities outside the [0, 1] interval. In addition to this problem, there are two other issues concerning the LPM:

- the variance of the error (ε) depends on **x** and is therefore not constant, that is, the error variance is heteroscedastic.[4]
- since y takes only two values, so also do the errors. Hence the errors cannot have a Normal distribution.[5]

Despite efforts to correct the problems of the LPM, this model is essentially a dead end. The preceding problems with the LPM are resolved if a form for the function F(**x**, b) is chosen such that its values lie in the [0, 1] interval. Since any cumulative distribution function (cdf) will do this, one can simply choose from among any number of cdfs for F(**x**, b).[6] Choosing the Normal cdf gives rise to the Probit model and choosing the Logistic cdf gives rise to the Logit model. For the Probit model the probability that y = 1 is

$$\Pr(y = 1 | \mathbf{x}) = F(\mathbf{x}, \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} f(t)\, dt = \Phi(\mathbf{x}'\boldsymbol{\beta}) \qquad (4)$$

where φ(·) denotes the *standard* Normal density function and Φ(·) denotes the *standard* Normal cdf. For the Logit model the probability that y = 1 is

$$\Pr(y = 1 | \mathbf{x}) = F(\mathbf{x}, \boldsymbol{\beta}) = \frac{\exp[\mathbf{x}'\boldsymbol{\beta}]}{1 + \exp[\mathbf{x}'\boldsymbol{\beta}]} = \Lambda(\mathbf{x}'\boldsymbol{\beta}) \qquad (5)$$

where Λ(·) denotes the *standard* Logistic cdf and exp[·] is the exponential function. The assumed probability distribution then applies directly to the conditional distribution of the error. Both models assume E[ε | **x**] = 0. For the Probit model, the choice of a *standard* Normal cdf involves the nonrestrictive assumption Var[ε | **x**] = 1. For the Logistic model, the choice of a *standard* Logistic cdf involves the nonrestrictive assumption Var[ε | **x**] = $\pi^2/3$.[7] The *standard* Normal and *standard* Logistic distributions are chosen because they are simple and easily manipulated functions of the variables. The assumed value for the variance of the error distribution is an identifying restriction needed to pin down the values of the coefficients in either model (see Long, 1997, pp. 47-49).

---

[4] The error variance is Var[ε | **x**] = F(**x**¢b)(1 - F(**x**¢b)) = **x**¢b(1 - **x**¢b)

[5] This only precludes hypothesis testing, not estimation.

[6] The cumulative distribution function (cdf) of a random variably Z gives the probability of observing values of Z less than or equal to some chosen value (z*), that is cdf(z*) = Pr(Z ≤ z*).

[7] The value $\pi^2/3$ is the variance of the *standard* Logistic distribution.

Estimation

Estimation of binary Logit and Probit models (and almost all the other models discussed here) is made using the method of Maximum Likelihood, which we assume is familiar to the researcher (see Eliason, 1993). In all cases, one first determines the form of the likelihood function for the model.[8] Once determined, the estimates **b** for parameters b are then derived by maximizing the likelihood function with respect the parameters b. This involves setting the first derivatives of the likelihood function to zero and solving for the coefficients. In general, the first derivative equations (called the Likelihood Equations) are nonlinear, so an exact analytical solution for the coefficients cannot be obtained. Instead, the values **b** that maximize the likelihood function are obtained using an iterative numerical method. This simply means one starts with an initial set of estimates $\mathbf{b}_0$, computes the value of the likelihood function using $\mathbf{b}_0$ and then, using some method to update the values $\mathbf{b}_0$, one obtains new values $\mathbf{b}_1$. One then computes the value of the likelihood function using the new values $\mathbf{b}_1$. This iterative process of updating the coefficients **b** and calculating the value of the likelihood function continues until convergence, the latter being a stopping rule for when the computer is told to believe that it has obtained the values of **b** for which the likelihood function is at its maximum.

Statistical programs such as LIMDEP, SAS, SPSS and STATA provide "point and click" routines to estimate Logit and Probit models. Hence, we need not dwell further on the intricacies of the numerical methods used to obtain Maximum Likelihood estimates (Greene (2002, Chapter 17) has extensive discussion). However, three general points are worth noting. First, for computational simplicity, one maximizes the natural logarithm of the model's likelihood function and not the likelihood function itself. As a result, the computer printout will report the maximized value of the log-likelihood function and not the maximized value of the likelihood function. This presents no special issues and is in fact convenient since the maximized value of the log-likelihood function is a number used to test hypotheses about the model and the estimated coefficients. Second, Maximum Likelihood estimates are consistent, normally distributed and efficient. However, these are asymptotic properties that hold as the sample size approaches infinity. In practice, this means using relatively large samples. Given the focus on organizations rather than individuals, strategy researchers often lack such large samples. Long (1997, pp. 53-54) suggests samples sizes of at least 100 observations with 500 or more

[8] For the binary models, each observation is assumed to be an independent Bernoulli trial with success probability $\Pr(y = 1 \mid \mathbf{x}) = F(\mathbf{x}¢b)$ and failure probability $\Pr(y = 0 \mid \mathbf{x}) = [1 - F(\mathbf{x}¢b)]$. Given "n" independent observations, the likelihood function takes the form $L(b \mid \mathbf{Y}, \mathbf{X}) \propto \prod_{i=1}^{n} [F(\mathbf{x}_i' ß)]^{y_i} [1 - F(\mathbf{x}_i' ß)]^{1-y_i}$

observations being desirable. But since the number of parameters in the model is also important, a rule of at least 10 observations per parameter is suggested, keeping in mind the minimum requirement of at least 100 observations. Finally, variables measured on widely different scales can cause computational problems. One should therefore scale the variables so their standard deviations have about the same order of magnitude.

Interpreting Results

As with standard regression, a researcher is first interested in assessing the overall significance and “goodness of fit” of the model. After that, support for or against one’s hypotheses is usually made by examining the significance, the sign, and possibly the magnitude, of one or more estimated coefficients. In OLS these aspects are quite straightforward, with key results such as the F-test, R^2, coefficient estimates, t-statistics, etc. reported in the computer output. As a result, researchers tend to be consistent in their interpretation and reporting of standard regression results. Unfortunately, this is not the case for models that involve a LDV. Our review of the recent use of the Logit model in strategy research indicated that most studies do not provide adequate reporting of results.[9] Researchers tend to focus on the individual significance and direction of the coefficients to support or refute their hypotheses without also providing a test of the overall significance of the model. However there is also almost a total absence of discussion about the marginal impact of an explanatory variable on the dependent variable. The following sections discuss the appropriate methods for interpreting the estimation results of the binary Logit model.[10]

To facilitate discussion we estimated a binary Logit model for the nature of CEO succession to illustrate the presentation and interpretation of results. To model the choice by a firm’s board to hire either an individual from outside the organization or from within the organization as replacement CEO we define the dependent variable, CEO Replacement Type. This variable takes the value 1 if the replacement CEO came from the outside the organization and equals 0 if the individual was promoted from within. The explanatory variables are “Succession Type” and “Pre-Succession Performance.” Succession Type is a dummy variable that equals 1 if the former CEO was dismissed and equals zero otherwise (i.e., routine succession). The variable Pre-Succession Performance is the average change in the total return

[9] Studies often fail to report basic statistics to indicate overall model significance, and most studies do not go beyond reporting the model and individual coefficient significance.

[10] Most of what is said here also applies to the binary Probit model.

to a shareholder of the firm during the two years prior to the year of CEO succession.[11] The results are shown in TABLE 1 and TABLE 2.

Insert Table 1 & 2 About Here

Assessing Model Significance

In the standard regression framework an F-statistic is used to test for overall model significance. The null hypothesis being tested is that all explanatory variable coefficients are jointly equal to zero. If the model passes this test then the researcher proceeds to examine the significance and sign of individual coefficients to support or reject hypotheses about the phenomenon of interest. In the context of Maximum Likelihood estimation, the same null hypothesis of overall model significance is tested using a Likelihood Ratio (LR) test.

In general, a LR test is conducted by comparing the maximized value of the log-likelihood function of an unrestricted (full) model to the maximized value of the log-likelihood function of a model in which some restrictions have been imposed on some or all of the model's coefficients. Let LL_R denote the log-likelihood value of the restricted model and let LL_U denote the log-likelihood value of the (full) unrestricted model with all variables included. The LR test statistic is calculated as $LR = -2 \times [LL_R - LL_U]$. This test statistic has a Chi-square distribution with degrees of freedom equal to the number of coefficient restrictions imposed on the full model.

To conduct a LR test of overall model significance two models are estimated. The first is the full model that includes all variables and the second is a restricted model that contains only a constant term. Using the values of the log-likelihood for each model, one computes the statistic $LR = -2 \times [LL_R - LL_U]$. The p-value for LR is obtained from a Chi-square distribution with degrees of freedom equal to the number of explanatory variables.

TABLE 1 shows that the maximized value of the log-likelihood function for the full model is -77.925 and -98.601 for the null model (results for the null model are normally not reported, but are reported here for illustration). The LR statistic is $LR = -2\times(-98.601 - 77.925) = 41.35$ (this value is also reported in TABLE 1, and it is commonly reported in the usual

[11] All the discrete LDV models presented in this chapter were estimated using the program STATA.

computer output). Since the full model contains two explanatory variables and the null model contains none, the number of restrictions being imposed on the full model is 2. From a Chi-square distribution with 2 degrees of freedom one finds that the probability of observing a LR value greater than 41.35 is 1.048E-09. Hence, the hypothesis that the variable coefficients are jointly equal to zero can be rejected, providing support that the overall model is significant.

The LR ratio test extends to cases where one is interested in testing the significance of subsets of the variables. In a standard regression framework, strategy researchers often present their models by starting from a minimal "base" model (e.g., constant and control variables) to which they then add different groups of variables resulting in several models. This is usually presented as a Stepwise Regression where at each step the contribution to R^2 is evaluated using an F-statistic that tests if the coefficients on the group of variables just added to the model are jointly equal to zero. The analogue to this for a model estimated by Maximum Likelihood is to start with the full model with all variables included and to then successively test, using the LR statistic, if the coefficients on a subgroup of variables are jointly equal to zero. In all cases, the LR statistic is $LR = -2 \times [LL_R - LL_U]$ where LL_R is the log-likelihood value for the restricted model that excludes the subgroup of variables and LL_U is the log-likelihood value for the model with all variables included. The p-value for the LR value obtained is derived from a Chi-square distribution with degrees of freedom equal to the number of variables excluded from the full model. Note that this procedure is always testing a partial model that excludes some subgroup of variables against the full model with all variable included.[12]

In addition to testing for model significance, some measure indicating the overall "goodness of fit" of the model should be reported. Strategy researchers that use Logit or Probit models rarely report a goodness of fit measure. This may be explained, in part, by the fact that Maximum Likelihood estimation does not lead to a natural measure of goodness of fit, unlike R^2 for OLS. This arises because Maximum Likelihood estimation is not based on maximizing explained variation whereas OLS seeks to maximize R^2. However, one obvious measure of "fit" in the context of Maximum Likelihood is the maximized value of the log-likelihood function and this should always be reported. This number is always negative so that smaller (absolute) values indicate a higher likelihood that the estimated parameters fit the data.[13] Use of this log-

[12] One might think to compare each incremental model (Model 2, 3, 4, etc.) to the base model (Model 1). However, this is an inappropriate use of the LR test. The LR test assumes one is imposing restrictions on the coefficients of a full model with all variables included. Hence, for models estimated by Maximum Likelihood researchers should not perform the type of "incremental R^2" analysis often done in the standard regression framework.

[13] This is true if the number of variables in the model remains constant. Like the standard regression model, where adding more variables increases R^2, the likelihood value also raises if more variables are added to the model.

likelihood value is only made when one compares different models, since its value for a single model tells us nothing about how well that model "fits."

Two additional goodness of fit measures often reported are the pseudo R-square and the percentage of correctly predicted choices.[14] The pseudo R-square, or Likelihood Ratio Index (McFadden, 1973), is computed as

$$\text{LRI} = 1 - \frac{\text{LL}_U}{\text{LL}_R}$$

where LL_U is again the log-likelihood value for the full model and LL_R is the log-likelihood value for a null model that includes only a constant term. Computer programs often report this pseudo R-square. For our logit example the pseudo R-square is 0.21 (see TABLE 1). This does **not** mean that the full model explains 21% of the variation in the dependent variable. No such interpretation is possible. Instead, this number is only a benchmark for the value of the log-likelihood function of the full model compared to that for the restricted model. The pseudo R-square will be higher the more "significant" is the full model compared to the null model, but otherwise no further interpretation can be given.[15] Hence, reporting this value serves mainly as a benchmark for comparing other models of the same phenomena that might be estimated and presented in the literature.

Whether the model correctly predicts the observed sample choices is another commonly used measure of "fit." This involves computing the predicted probability ($\hat{y}_i$) that y = 1 for each firm in the sample and then comparing this predicted probability to some threshold probability, usually 50% for the case of a binary dependent variable. If the predicted probability exceeds the threshold probability then the prediction is that $\hat{y}_i$ = 1, otherwise $\hat{y}_i$ = 0. The predicted choice is then compared to the actual choice (y = 0 or 1) and the proportion of correct predictions is then taken as an indicator of how well the model fits in terms of predictive ability. For our logit example the percentage of correctly classified choices is 80.3%, which can be calculated from the table of predicted vs. actual choices shown in TABLE 2. A contentious aspect of this predictive fit measure is the choice of the threshold value beyond which the predicted choice is assumed to be $\hat{y}_i$ = 1. The threshold probability 50% is often used. But in an unbalanced sample where the sample proportion of successes is far from 50% it is recommended that one instead choose the threshold value to be the actual proportion of observations for which y = 1 in the

[14] Several other measures have been proposed (see Long, 1997, pp. 102-113).

[15] Since the pseudo R^2 uses the log-likelihood values of the restricted and unrestricted models, values of this measure can be linked to the Chi-Square test of model significance.

sample.[16] In our data the sample proportion of outsiders (y = 1) is 20.58%. When this number is used as the prediction threshold the percent of correct predictions is 73.4%.

Individual Effects

Once overall model significance is assessed the researcher can examine specific hypotheses regarding individual variables. In studies that use OLS, researchers usually discuss the significance of each explanatory variable and the effect that a unit change in a variable will have on the dependent variable in terms of its direction and magnitude (i.e. the sign and size of a variable's coefficient). Since Maximum Likelihood estimates are asymptotically normally distributed all the familiar hypothesis tests regarding individual coefficients, including the usual test that an individual coefficient is zero, can be performed based on the estimated coefficient standard error. However, unlike OLS, the ratio of a coefficient to its standard error is not a t-statistics but is instead a normal z-value, so that p-values are based on the normal distribution.[17] The interpretation of the directional impact (+ or -) of a change in an explanatory variable in the binary Logit (Probit) Model is identical to that for OLS, except that one should keep in mind that the direction of the effect refers to the change in the probability of the choice for which y = 1.

Strategy researchers who use the binary Logit (or Probit) Model often limit their interpretation of results to the significance and direction of the coefficient and rarely calculate the impact of an explanatory variable. In studies where the individual impact of an explanatory variable is discussed it is often done erroneously, by directly referring to size of the estimated coefficient. This is not correct. In general, the coefficient estimated in the context of a discrete LDV model does not indicate the size of the effect on the dependent variable due to a unit change in an independent variable. This is because the relationship between the dependent and independent variables is nonlinear. Instead, one needs to compute what is called the "marginal effect" for each independent variable. In general, the marginal effect will vary with the value of the variable under consideration and also with the values of all other variables in the model. Hence, unlike the coefficients in standard linear regression, the marginal effect of a change in an independent variable on the decision outcome $\Pr(y = 1 \mid \mathbf{x})$ is not a constant.

[16] Greene (2002) discusses the arbitrariness of such fit measures and the tradeoffs inherent in their application.
[17] Since normality of Maximum Likelihood estimates is an asymptotic property, computer programs sometimes report the z-values as "asymptotic t-statistics."

Marginal Effects

The marginal effect due to a change in an independent variable on the probability that y =1 is calculated either from the expression for the partial derivative of the logit (probit) function or as the discrete change in the predicted probability when the variable of interest undergoes a discrete change. The latter discrete method must be used to compute the marginal effect for a dummy independent variable.

Taking first the derivative approach, the marginal effect on the probability that y = 1 is:

$$\frac{\partial \mathrm{E}[y \mid \mathbf{x}]}{\partial x_k} = \frac{\partial \Pr[y=1 \mid \mathbf{x}]}{\partial x_k} = f(\mathbf{x}'\boldsymbol{\beta})b_k \qquad (6)$$

where $f(\mathbf{x}'b)$ is the density function associated with either the Probit (standard Normal) or Logit model (standard Logistic).[18] There are three important things to notice about the marginal effect given in (6). First, unlike OLS, the marginal effect is not the estimated coefficient β_k. Second, the sign of the marginal effect is the same as the sign of the estimated coefficient β_k (since $f(\mathbf{x}'b)$ is always positive). Thirdly, the size of the marginal effect depends on the estimated coefficients and the data for all other variables. Hence, to calculate values of the marginal effect (6), one must choose values for all the other variables. Stated differently, the marginal effect for a change in a variable x_k is computed holding fixed the values of all other variables.

There are two common approaches to calculating a marginal effect (these approaches apply to all discrete choice models, not just the binary models discussed here). The first is to compute the value of $f(\mathbf{x}'b)$ using as data the mean of each x variable and to then multiply this value times the estimated coefficient β_k as in (6). This effect is called the "marginal effect at the mean."[19] The value of $f(\mathbf{x}'b)$ needs to be calculated only once since the same value of $f(\mathbf{x}'b)$ multiples each coefficient (β_k).

For our sample model, the "marginal effect at the mean" for each variable is shown in TABLE 1. For the variable Pre-Succession Performance the value of $f(\mathbf{x}'b)$ was calculated holding fixed the values of Succession Type and Pre-Succession Performance at their sample means. The value of $f(\mathbf{x}'b)$ in this case was 0.15744549. As shown in TABLE 1, the resulting

[18] The term $f(\mathbf{x}'b)$ appears in the marginal effect since $f(\mathbf{x}'b)$, being the derivative of the cdf, indicates the steepness of the cdf at the value $\mathbf{x}'b$, and the steeper is the cdf the larger will be the increment in the probability for a given change in x_k .

[19] Another approach to calculating a marginal effect is to compute the values $f(\mathbf{x}'b)$ for each observation and to the average these values across observations. This average value of $f(\mathbf{x}'b)$ is then multiplied times the estimated coefficient for the variable of interest to obtain the "average marginal effect" for that variable.

marginal effect for Pre-Succession Performance is -0.00276. That means that a one unit (one percentage point) rise in Pre-Succession Performance above its mean value lowers the probability of an outsider being chosen as the replacement CEO by 0.00276 (.28%) – a relatively small value.

The marginal effect for a dummy variable like Succession Type is not calculated using formula (6). Instead, the marginal effect for a dummy variable must be computed using the discrete change in the probability due to a discrete change in the variable. Again, one needs to fix the values of all other variables, usually at their mean levels (denoted collectively below as $\bar{\mathbf{x}}$). The effect of a discrete change in a variable x_k of size "δ" on the predicted probability is

$$\frac{\Delta \Pr(y=1 \mid \bar{\mathbf{x}})}{\Delta x_k} = \Pr(y=1 \mid \bar{\mathbf{x}}, (\bar{x}_k + d) - (\Pr(y=1 \mid \bar{\mathbf{x}}, \bar{x}_k) \qquad (7)$$

The choice for the size (δ) of the change in a variable is up to the researcher; common values are $\delta = 1$ (a one unit change) and $\delta = \sigma_k$ where σ_k is the sample standard deviation of variable x_k (a one standard deviation change). In all cases the incremental change in a variable is measured starting from the mean of that variable.

Calculation of a discrete change in the probability is necessary to assess the effect of a change in a dummy variable. For the case of a dummy variable that changes from 0 to 1 the formula is:

$$\frac{\Delta \Pr(y=1 \mid \bar{\mathbf{x}})}{\Delta x_k} = \Pr(y=1 \mid \bar{\mathbf{x}}, x_k = 1) - \Pr(y=1 \mid \bar{\mathbf{x}}, x_k = 0) \qquad (8)$$

The marginal effect for Succession Type in our example was calculated using (8) where the predicted probability was computed with the value of Pre-Succession Performance held fixed at its mean value. As shown in TABLE 1, the calculated marginal effect is 0.2923. This means that, holding the firm's Pre-Succession Performance fixed at its mean value, the probability that the Board will select an outsider as the replacement CEO increases by 0.2923 (28.2%) if the former CEO was dismissed, a significant and important finding.

Odds Effects

For the Logit model there is another useful interpretation of the estimated coefficients: the effect that a change in a variable will have on the *odds* in favor of outcome y = 1 versus y =

0.[20] One can show that the change in the *odds* in favor of choice y = 1 versus choice y = 0 when a variable x_k changes by $\Delta x_k = \delta$ units is

$$\frac{\Delta(\text{Odds of Y=1 versus Y = 0})}{\Delta x_k} = \exp(d b_k) \qquad (9)$$

This states that the effect of a one unit change (i.e., $\delta = 1$) in variable x_k on the odds is just the exponential of that variable's coefficient.[21] The values of exp($\delta\beta_k$) are always positive, but can be greater or less than one. A value greater than one indicates that the odds in favor of y = 1 rise as x_k rises, while values less than one indicate that the odds instead move in favor of y = 0 as x_k rises. A key advantage of considering the odds effect of a change in a variable is that this effect, unlike the marginal effect, does not depend on the values of any of the variables in the model, and they are also easy to compute from the estimated coefficients. In addition, formula (9) is also used to calculate the odds effect for a change (from 0 to 1) in a dummy variable.

The last column of TABLE 1 lists the odds effects for Succession Type and Pre-Succession Performance. For Dismissal Succession Types, the value 6.762 means that the odds in favor of an outsider being selected as the replacement CEO are almost 7 times higher if the former CEO was dismissed. For Pre-Succession Performance, the value 0.979 means that a one unit (one percentage point) increase in performance lowers the odds in favor of an outsider being selected as the replacement CEO. Specifically, a ten unit (ten percentage point) increase in this variable would reduce the odds in favor of an outsider being selected as the replacement CEO by a factor of 0.811 (= $\exp(\delta\beta_k) = \exp(10 \times -0.021)$).

As illustrated by the CEO succession example there is rich set of interpretations one can make about the relationship between the independent variables and the phenomenon of interest beyond simply the direction of the effect. For the Logit model one should, at a minimum, compute and discuss the odds effect for each variable. The calculation and interpretation of marginal effects takes more care, but these are also useful numbers, and they are needed if one is to know how changes in variables affect the probability of making the choice for which y = 1.

Summary of Binary Model Methods

TABLE 3 gives an overview of the elements discussed in this section and which one needs to be aware of when using Binary Logit or Probit models. The table also states key

---

[20] To calculate the change in the odds in a Probit model one needs to compute probabilities at different values of a variable and then compute the odds before and after a change in the variable. Since this involves many indirect computations, the analysis of odds is rarely done for the Probit Model.

[21] The effect of a one-standard deviation change in x_k is computed by setting δ equal to the sample standard deviation of variable x_k.

assumptions underlying the models as well as what researchers should minimally report when presenting the results of their analysis. Our recommendation to report the pseudo R-square and the percentage of correct predictions is made to achieve a consistency of reporting across papers, like that done for OLS results. But in making these recommendations we do not ignore that these "fit" measures have problems of interpretation.

Insert Table 3 About Here

Multiple Outcomes

Strategy researchers are often interested in the nature of the strategic choices made by corporate managers and the factors underlying these choices. However, such choices are rarely binary. Examples include the numerous options for entering a new market (Kogut and Singh, 1988), the choice to expand, hold, or exit an industry (Eisenmann, 2002), and choice regarding the level of patent litigation (Somaya, 2003). In addition, researchers who examine firm performance as a dependent variable often use categorical rather than continuous performance data (Pan and Chi, 1999). Therefore, much of the strategic choice phenomenon that strategy research has often operationalized as binary should instead be broadened to consider the full array of options a firm can pursue. Doing so may offer a greater chance to explain variation in decision outcomes and lead to a better understanding of the real world wherein managers contemplate an array of options before making one strategic choice.

Strategic choices that involve multiple discrete alternatives pose a different set of challenges for the researcher. This section discusses models where the dependent variable involves multiple discrete outcomes. The choice outcomes represented by discrete values of the dependent variable can be either ordered or unordered. We first discuss the case of unordered outcomes.

Unordered Outcomes

FIGURE 1 shows there are five basic models for the case of an unordered discrete LDV: Multinomial Logit, Multinomial Probit, Nested Logit, Conditional Logit, and Mixed Logit. Our discussion will focus on Multinomial Logit since this model is the most widely used in the strategic management literature. Of course, one can also specify a Multinomial Probit model, which has the advantage that it imposes less restrictive assumptions on the probabilities than do the Logit based models, an issue we discuss further below in the section entitled "The Independence of Irrelevant Alternatives." [22]

Multinomial Logit

The Multinomial Logit model is the most widely used model when a researcher has a limited dependent variable with multiple unordered alternatives. The model assumes J+1 unordered and mutually exclusive alternatives numbered from 0 to J. For a given observation the value taken by the dependent variable is the number of the alternative chosen. In this model the probability that decision maker "i" chooses alternative j, denoted $\Pr(y_i = j \mid \mathbf{x}_i)$, is

$$\Pr(y_i = j \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{\sum_{j=0}^{J}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)} \qquad j = 0,1,2,\ldots,J \qquad (10)$$

The vector $\mathbf{x}_i$ in (10) contains a set of firm specific variables thought to explain the choice made. The coefficient vector $b_{\mathbf{j}} = [\beta_{0j}, \beta_{1j}, \ldots, \beta_{kj}, \ldots \beta_{KJ}]$ contains the intercept β_{0j} and slope coefficients β_{kj}. Note that the set of coefficients $b_{\mathbf{j}}$ is indexed by "j." This means there is one set of coefficients for each choice alternative and that the effect each variable x_k has on the probability of a choice varies across the choice alternatives. The model given in (10) has J+1 equations but only J of these equations can be estimated due to an identification problem with respect to model coefficients (discussed below). Therefore, estimation of the model will result in J equations, one for each of J choice alternatives, and the estimated coefficients for one particular choice alternative (may) differ from those of any other choice alternative.

If one were to insert the coefficient vector $\tilde{\boldsymbol{\beta}}_j = \boldsymbol{\beta}_j + \mathbf{z}$, where $\mathbf{z}$ is any vector, in (10) the probability would not change. Hence, some restriction on the coefficients is needed. The usual assumption is to restrict $b_{\mathbf{0}} = \mathbf{0}$ (remember $b_{\mathbf{0}}$ is a vector of coefficients for the choice alternative coded as "0"). Restricting all coefficients to equal zero for the choice y = 0 means that this choice is selected as the "base choice" for the model. Imposing the constraint $b_{\mathbf{0}} = \mathbf{0}$ in (10) gives

$$\Pr(y_i = j \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{\sum_{j=0}^{J}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)} = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{1+\sum_{j=1}^{J}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)} \qquad j = 0,1,2,\ldots,J \text{ and } \boldsymbol{\beta}_0 = \mathbf{0} \qquad (11)$$

[22] One issue that has limited the use of the Multinomial Probit model is the difficulty of numerically computing the value of multivariate normal integrals. But the attractiveness of this model in terms of its assumptions should not be ignored when deciding on which model, Probit or Logit, to use. Moreover, recent computational advances now

where the final expression arises since $\exp(\mathbf{x}_i'\mathbf{ß}_j) = \exp(0) = 1$. Effectively, each of the J equations in (11) is a binary logit between alternative j and the base choice, that is, the choice whose coefficients are restricted to equal zero. Which choice alternative is selected to be the base choice is arbitrary and only affects how one interprets the resulting coefficient estimates. Note that while all coefficients in the base choice equation are restricted to equal zero, the probability that the base choice is selected can still be computed, as can the marginal effects.[23]

Interpreting Results

As with the binary Logit model, a researcher using a Multinomial Logit model is first interested in assessing the overall significance and "goodness of fit" of the model. In addition, hypotheses testing will require examining the significance, the sign, and possibly the magnitude of the coefficients. In Multinomial Logit the number of choice alternatives increases the number of binary comparisons to be made. Our review of the use of multinomial models in strategy research indicates that most studies again fail to provide an adequate reporting of results. Unlike a binary model, a multinomial model has the added problem that the sign of a coefficient need not indicate the direction of the relationship between an explanatory variable and the dependent variable. Only by calculating the marginal effects in the Multinomial Logit model can one arrive at a valid conclusion about the direction and magnitude of the relationship between the dependent variable and an explanatory variable.

To illustrate results and their interpretation, we estimate the earlier binary Logit model of CEO succession as a Multinomial Logit model. To do this we constructed a new dependent variable as the interaction of CEO succession type and CEO replacement type. This resulted in four succession outcomes coded as follows: y = 0 if the CEO succession is routine and an insider is hired as the replacement CEO; y = 1 if the CEO succession is routine and an outsider is hired as the replacement CEO; y = 2 if the CEO succession is a dismissal and an insider is hired as the replacement CEO; and y = 3 if the CEO succession is a dismissal and an outsider is hired as the replacement CEO. The explanatory variables are "Pre-Succession Performance" and "Succession Year Performance." The first variable is the same variable used for the binary Logit example; it captures the change in stockholder return in the two years prior to the succession year. The second variable is the total return to a shareholder of the firm in the year of succession.

permit estimation of a Multinomial Probit model with up to 20 choice alternatives (e.g., the most recent version of LIMDEP). Hence, the use of this model may be expected to increase in the future.

[23] If y = 0 is the base choice the probability of this alternative being chosen is $\Pr(y = 0 \mid \mathbf{x}) = 1 \Big/ \left(1 + \sum_{j=1}^{J} \exp(\mathbf{x}_i'\mathbf{ß}_j)\right)$.

To estimate the model the choice y = 0 (i.e., routine succession and insider replacement CEO) was selected as the base choice. The estimated results for each choice, including the base choice y = 0, are shown in TABLE 4. Normally, only results for the unrestricted choice options (here, y = 1, 2 and 3) would be reported. However, it is to be noted that the full model encompasses all choices including the base choice. The Maximum Likelihood procedure jointly estimates all choice equations and therefore results in a single log-likelihood value for the joint model, not a separate log-likelihood value for each choice equation. As shown in TABLE 4, this single log-likelihood value for our model is -175.55.

Insert Table 4 About Here

Assessing Model Significance

Assessing the overall significance for a Multinomial Logit model is, as in the binary case, determined using the Likelihood Ratio (LR) test that compares the maximized value of log-likelihood of the full model to the maximized value of log-likelihood of a null model that includes only a constant term. Since the Multinomial Logit model has J equations the null model refers to the case in which each of the J choice equations contains only a constant term. The resulting LR statistic therefore has a Chi-square distribution with degrees of freedom equal to the number of variables times the number of unrestricted choice equations (J). In the present example, the Chi-square distribution used to test overall model significance has 6 degrees of freedom (3 equations times 2 variables in each equation). As shown in TABLE 4, the Chi-square value is 65.16, which is highly significant as indicated by its associated p-value. Therefore, the hypothesis that the coefficients on all variables across all equations are jointly equal to zero can be rejected.

As for goodness of fit, the pseudo R-square for the model is 0.156. Again, this does **not** mean that the full model explains 15.6% of the variation in the dependent variable since this number is only a benchmark for the value of the log-likelihood function of the full model compared to the null model. To determine the percentage of correct predictions the following procedure is used. For each observation (firm) one computes the predicted probability for each of the four choices (y = 0, 1, 2, and 3). The choice option with the highest predicted probability is then selected to be predicted choice for that observation (firm). Doing this for our model produces the table of predicted vs. actual choices shown in. Summing the diagonal elements in and dividing by the total number of observations (= 188) gives the percentage of correctly classified choices as 65.43%

Insert Table 5 About Here

Individual Effects

If the overall model is significant, specific hypotheses regarding individual explanatory variables can be examined. Determining the significance of individual explanatory variables in the Multinomial Logit model differs from that done in the binary model since there are J coefficients for each variable. Therefore, a test of the overall significance of a variable requires testing the hypothesis that the J coefficients for that variable are jointly equal to zero. This hypothesis is tested using a LR statistics test that compares the maximized value of the log-likelihood of the full model to the maximized value of the log-likelihood of the model that excludes the variable of interest. The resulting LR statistic has a Chi-square distribution with J degrees of freedom. While the LR procedure tests the significance of a variable for the model as a whole, the individual z-statistic and associated p-value reported for a given variable in any one particular choice equation is used to test if that variable is significant in determining the probability of that particular choice.

To test overall variable significance in our model we estimate two restricted models. The first (Model 1) excludes the variable Succession Year Performance, the second (Model 2) excludes the variable Pre-Succession Year Performance. The resulting log-likelihood values are -195.50 for Model 1 and -195.15 for Model 2. To test the significance of Succession Year Performance we compute the LR statistic as $LR = -2 \times [-195.50 - (-175.55)] = 160.1$. The associated p-value is 1.74E-34 based on a Chi-square distribution with 3 (= J) degrees of freedom. This variable is therefore highly significant. For Pre-Succession Year Performance the LR statistic is $LR = -2\times[-195.15 - (-175.55)] = 159.3$ and the associated p-value is 2.60E-34 (again, from a Chi-square distribution with 3 degrees of freedom). The variable Pre-Succession Year Performance is also highly significant.

When considered on a individual basis, each variable is significant in each choice equation except for the variable Succession Year Performance for choice y = 1. This most likely reflects that this choice was made by only 10 of the 188 firms in the sample.

In the Multinomial Logit model, the direction (sign) of an estimated coefficient cannot be used to ascertain the direction (+ or -) of the relationship between an explanatory variable and the probability of a specific choice. The directional relationship and the relative impact of an explanatory variable instead depend on the values of all variables and their estimated coefficients across all choice alternatives. To be specific, it can be shown that the marginal effect of a change in variable x_k on the probability that alternative j is chosen is

$$\frac{\partial \Pr(y_i = j \mid \mathbf{x})}{\partial x_k} = \Pr(y_i = j \mid \mathbf{x}) \left[b_{kj} - \sum_{m=1}^{J} b_{km} \Pr(y_i = m \mid \mathbf{x}) \right] \quad (12)$$

If one chooses to actually calculate (12) one usually sets the values of all variables to their sample mean value and calculates the "marginal effect at the mean," similar to the calculation of a marginal effect in the binary case. Fortunately, most computer programs include an option to compute these marginal effects. The critical thing to notice about (12) is that the sign of the marginal effect need not be the same as the sign of the estimated coefficient β_k. This fact is often overlooked by researchers when reporting their findings and often leads to confusion. Many researchers erroneously assume that the sign of an estimated coefficient specifies the direction of the relationship between a variable and the probability of a given choice, and they use this to support or refute their hypotheses.[24]

To illustrate that the sign of the estimated coefficient for a variable need not to be the same as the sign of its marginal effect, we can turn to the results shown in TABLE 4. For choice y = 1, the sign of the coefficient on Succession Year Performance is negative while this variable's marginal effect (at the mean) is positive. This result does not depend on the fact that this variable is not significant in this choice equation.

The effect of a change in a variable in the Multinomial Logit model can also be interpreted in terms of its effect on the odds in favor of a given choice relative to the base choice. The computation of this odds effect is same as in the binary case, that is, for a change of size δ in variable x_k the odds effect is computed as $\exp(\delta\beta_{kj/0})$ where $\beta_{kj/0}$ is the coefficient on variable k in equation j and the "0" subscript indicates that y = 0 is the base choice. Setting $\delta = 1$ gives the effect on the odds in favor of choice j versus the base choice (y = 0) for a unit change in variable x_k. It is important to take note that the odds effect for a variable refers to a change in the odds in favor of a particular choice versus the base choice.[25]

---

[24] In addition, the standard errors of the marginal effects will differ from the standard errors of the model coefficients since the former will depend on the variable values used when calculating the marginal effect.

[25] All computer programs that estimate the Multinomial Logit model impose the restriction $b_m = \mathbf{0}$ for some choice m. Here m = 0, that is, choice y = 0 is the base choice. The use of y = 0 as the base choice may not always be the case, so one must check which choice is taken to be the base choice by one's computer program.

TABLE 4 lists the odds effects for our model. These numbers show how each one unit change in a variable would affect the odds in favor of a given choice versus the choice y = 0. Specifically, for choice y = 2 (dismissal and insider replacement CEO), the effect of a one unit (one percentage point) increase in Succession Year Performance would lower the odds in favor of choice y = 2 versus choice y = 0 by a factor or 0.9707. Stated differently, the odds in favor of choice y = 2 (dismissal and insider hired) versus choice y = 0 (routine and insider hired) would decline by -0.0203 (= 0.9707 – 1). Since choices y = 2 and y = 0 both involve hiring an insider, one interpretation of this result is that higher stock performance in the succession year reduces the odds that the incumbent CEO would have been dismissed.

The Independence of Irrelevant Alternatives

An important assumption of the Multinomial Logit model is that the odds of one choice versus another choice do not depend on the number of choice alternatives available. In other words, adding choices to the existing set of choices (or subtracting choices from the existing set) does not affect the odds between any two alternatives. This feature of the Multinomial Logit model is derived from the formal equation for the odds in the model and is called the Independence of Irrelevant Alternatives (IIA) (McFadden, 1973). The practical advice often given is that when the alternatives are close substitutes the IIA assumption may be violated and the Multinomial Logit model may not give reasonable results.[26] Hausman and McFadden (1973) devised a test to assess if the IIA assumption is violated (for details see Long, 1997, pp. 182-184). As indicated in FIGURE 1, one should test for the validity of this assumption. If the IIA assumption is rejected then one possibility is to use the Multinomial Probit model since this allows the errors across choices (i.e., equations) to be correlated and hence does not impose the IIA assumption. Another alternative is the Nested Logit model discussed below.

Nested Logit

The Nested Logit model partially relaxes the IIA assumption by using a tree structure for the decisions that can be characterized as a set of branches and twigs (Greene, 2002, pp. 725-727). Each branch is a set of first level choices while each twig along a given branch represents a final choice. Take as an example the decision to undertake expansion using a joint venture or Greenfield investment. The branches might represent the choice of whether or not to expand in the domestic market or to expand abroad. For each branch, the twig level decisions are joint

[26] The IIA assumption derives from an assumed independence of the errors across the alternatives and is effectively the assumption that the error variance is homoscedastic.

venture and Greenfield. There are total of four decision outcomes, but these four decisions are partitioned. The nested specification does not impose the IIA assumption for the choice among branches but does maintain the IIA assumption among the twigs on a given branch. In estimating the Nested Logit model one can test the assumption of separating the decisions into branches and twigs or if the model can instead be collapsed into a standard Multinomial Logit model of choice among all twigs. Further details of this model can be found in Greene (2002, pp. 725-727). For an application of the model see Belderbos and Sleuwaegen (2003).

We now briefly discuss two additional Logit based models that can be used to model choice among multiple alternatives. While these models have yet to appear in the strategic management literature, they are of potential use and therefore deserve mention.

Conditional Logit Model

The Conditional Logit model is due largely to McFadden (1973).[27] This model is often (and confusingly) referred to as a Multinomial Logit model. The key distinction is that the variables used to explain the choices in the Conditional Logit model are characteristics of the choices themselves, rather than characteristics of the individual decision makers (firms). For example, in a study of the mode of foreign market entry, one might use variables that measure characteristics of the entry modes. If so, then one is estimating a Conditional Logit and not a Multinomial Logit model. In the Conditional Logit model the characteristics of the choices are the data, but these data may also vary across individual decision makers. For example, one might construct a "cost" variable that measures, for each firm, its cost for each entry mode. The values of this cost variable vary across the choices and also across firms.

To contrast the Conditional and Multinomial Logit models, we can consider each model's specification for the probability that firm i makes choice j (i.e., $y_i = j$). For the Multinomial Logit the specification is

$$\Pr(y_i = j \mid \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{ß}_j)}{\sum_{j=0}^{J}\exp(\mathbf{x}_i'\boldsymbol{ß}_j)} \qquad j = 0,1,2,...,J \qquad (13)$$

For the Conditional Logit model the specification is

$$\Pr(y_i = j \mid \mathbf{x}) = \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{ß})}{\sum_{m=0}^{J}\exp(\mathbf{x}_{im}'\boldsymbol{ß})} \qquad j = 0,1,2,...,J \qquad (14)$$

[27] This model is also called the Discrete Choice model.

where $\mathbf{x}_{ij}$ is a set of variables for firm i that relate to choice j. In the Multinomial Logit model one equation is estimated for each choice and there is one set of coefficients for each choice. In the Conditional Logit model only one set of coefficients is estimated over all choices.

In the Multinomial Logit model choice is modeled in terms of variation in firm characteristics while in the Conditional Logit model choice is modeled in terms of the variation in the characteristics of the choices (which may also vary with the firm). These are just different ways to view the process of how decision choices are made, and it seems reasonable to think these two models could be combined such that both characteristics of the decision outcomes and characteristics of the decision maker (firm) variables could determine the choice. This combined model is known as the Mixed Logit model.

Mixed Logit Model

The Mixed Logit model augments the Conditional Logit model to include variables on decision maker characteristics. In the combined data set the characteristics of the firm (i.e., the decision maker) do not vary across the alternatives (e.g., firm size, performance). To incorporate these variables a set of J dummy variables is used, one for each of the J choices, and these dummy variables are then interacted with the firm level characteristics to create a set of choice specific variables. For example, one might include firm size as a variable together with the cost to the firm of each mode of entry. In this case, the firm size variable would be incorporated using dummy variables, where the value taken by dummy variable j (corresponding to choice j) for firm i would be the size of firm i. The estimated dummy coefficient for a particular entry mode then indicates how firm size influences a firm's choice of that particular mode of entry. Details of the Mixed Logit model can be found in Powers and Xie (2000). It should be noted that both the Conditional and Mixed Logit models assume Independence of Irrelevant Alternatives (IIA) and therefore this assumption should be tested to assess model adequacy.

Ordered Outcomes

When the discrete values taken by the dependent variable can be rank ordered one can use an Ordered Logit or Ordered Probit model (McKelvey and Zaviona, 1975; McCullagh, 1980). Similar to the unordered multinomial models discussed previously, the Ordered Logit model arises if the choice probability is modeled in terms of a standard Logistic cdf while the Ordered Probit model arises if the choice probability is modeled in terms of the standard Normal cdf. The Ordered Logit and Ordered Probit models give essentially the same results, so the choice of model is up to the researcher.[28]

The ordered model assumes there are J rank ordered outcomes y = 1, 2, …, J. Since the choices are ranked order, the model for the probability of any particular choice can be formulated as the difference between cumulative probabilities. This is a key difference between ordered and unordered models. In particular, it means that only one set of variable coefficients is estimated in the ordered model, in contrast to the J coefficients estimated for each variable (one for each of the J choice outcomes) in an unordered model. While the ordered model estimates only one coefficient for each variable, it also estimates J-1 "intercepts" or "cut-points" that serve to differentiate the choices. Denote these J-1 ordered cut-points as $t_1 < t_2 < \ldots < t_{J-1}$ and let $F(\mathbf{x}, b)$ denote either the Logit or Normal cdf. The probabilities in the ordered model are then given as:

$$\begin{aligned}
&\Pr(y = 0 \mid \mathbf{x}) = F(-\mathbf{x}'\boldsymbol{\beta}) \\
&\Pr(y = 1 \mid \mathbf{x}) = F(t_1 - \mathbf{x}'\boldsymbol{\beta}) - F(-\mathbf{x}'\boldsymbol{\beta}) \\
&\vdots \\
&\Pr(y = j \mid \mathbf{x}) = F(t_j - \mathbf{x}'\boldsymbol{\beta}) - F(t_{j-1} - \mathbf{x}'\boldsymbol{\beta}) \\
&\vdots \\
&\Pr(y = J \mid \mathbf{x}) = 1 - F(t_{J-1} - \mathbf{x}'\boldsymbol{\beta})
\end{aligned} \tag{15}$$

The cut-point values are not observed but are instead estimated along with the variable coefficients using Maximum Likelihood.[29] The values estimated for the cut-points are only needed to compute predicted probabilities for each outcome and are otherwise of little interest with respect to model interpretation. Lastly, the formulation above assumes the structural model does not contain a constant term.[30]

While an ordered model is easily estimated, interpretation of the results requires careful attention. The marginal effect for a continuous variable in the ordered model is:

$$\frac{\partial \Pr(y = j \mid \mathbf{x})}{\partial x_k} = b_k \left[f\left(t_{j-1} - \mathbf{x}'\boldsymbol{\beta}\right) - f\left(t_j - \mathbf{x}'\boldsymbol{\beta}\right) \right] \tag{16}$$

---

[28] Since the Logit formulation lends itself to interpretation in terms of odds this may be one basis for choosing between the models.

[29] If the cut-point values are known the model is the Grouped Data regression model (Stewart, 1983). For example, one might have performance data (e.g. return on assets) on firms in terms of intervals rather than continuous performance data. In this case one might set y = 1 for firms with an ROA between 0 and 5%, y = 2 for firms with ROA between 5% and 10%, etc... Since the cut-points are known values the ordered model is not needed. Programs such LIMDEP and STATA estimate the Grouped Data model using Maximum Likelihood.

[30] As in other multiple response models, a restriction is required to identify model coefficients. One choice restricts the first cut-point value to equal zero, i.e., $t_1 = 0$ and to estimate the model with a constant term (LIMDEP uses this restriction). Another choice restricts the model's constant term to equal zero and to then estimate all J-1 cut-points (STATA uses this restriction). The restriction used does not affect the estimated variable coefficients. Our example model uses this second restriction and hence does not contain a constant term.

where f(·) is the pdf associated with F(**x,** b). From (16) it can be seen that the sign of the marginal effect depends on the values of all coefficients and variables, and it need not be the same as the sign of the estimated coefficient (β_k) for variable x_k. In addition, the sign of the marginal effect could switch depending on the values of the variables.[31] Given this, the interpretation of marginal effects in ordered models is tricky and requires careful analysis. Finally, we note that the marginal effect for a dummy variable must be computed, as before, as the discrete change in a predicted probability.

A review of the use of ordered models in the literature indicates that most researchers often report only the estimated model coefficients and do not compute marginal effects. Since the marginal effect indicates the directional relationship between the choice probabilities and an independent variable, reporting only estimated model coefficients conveys little if any information about the nature of the model. This implies that most researchers who have used an ordered model fail to pay proper attention to how the results in this model are interpreted.

If an Ordered Logit model is used then one can discuss the effect of variable changes in terms of changes in the odds in favor of one choice versus the remaining choices. However, unlike the binary and multinomial Logit models, the odds in the Ordered Logit are interpreted as the odds for cumulative probabilities. This means that a change of size δ in a variable x_k will change the odds in favor of outcomes *less than or equal to* alternative j versus those outcomes *greater than* alternative j by the amount $\exp(-\delta\beta_k)$, holding all other variables constant.[32] An example of this interpretation is given below.

To illustrate interpretation of an Ordered Logit model we estimate the CEO Succession model of the previous section but now treat the succession outcomes y = 0, 1, 2, and 3 as rank ordered. The results are presented in TABLE 6. Overall model significance is again tested using the LR statistic and, as indicated by the Chi-square p-value given in TABLE 6, the overall model is significant. Also reported are the pseudo R-square and percent of correctly classified choices (62.23%).

Insert Table 6 About Here

---

[31] This is because the values $f(t - \mathbf{x}'b)$ in (16) are the height of the pdf. As **x** changes the heights represented by the two values $f(t_{j-1} - \mathbf{x}'b)$ and $f(t_j - \mathbf{x}'b)$ can change relative to each other.·

TABLE 6 reports the estimated coefficients as well as the "marginal effect at the mean" for each variable. These marginal effects indicate the effect each variable has on the probability of each succession type. For choice y = 0 (routine succession and insider CEO replacement) the marginal effect is positive for both variables. This indicates that a rise in either variable will raise the probability of choice y = 0 by the indicated amount. For all other choices the marginal effect is negative. This indicates that a rise in either variable will lower the probability of choosing y = 1, 2 or 3. The magnitude of the decline in the probability of a given choice is indicated by the size of the marginal effect.[33]

If the only information given about the model is the estimated coefficients in TABLE 6 then the only conclusion one can reach is that, since the sign of each coefficient is negative, a rise in either variable would lower the probability of the last choice (y =3) and raise the probability of the first (choice y = 0). Of course, the marginal effects for these two choices also reveal this information. However, without these marginal effects, one could not say if a rise in either variable would also lower the probability of the intermediate choices y = 1 and y = 2.

Finally, we can consider the effect of variable changes on the cumulative odds. These effects are shown in the last column of TABLE 6. The calculated odds effect is 1.019 for Succession Year Performance and is 1.024 for Pre-Succession Performance. What do theses numbers tell us? Each number indicates, for a one unit rise in a variable, the change in the odds in favor of all choices less than or equal to one choice alternative versus all other choices greater than that choice alternative. For example, a unit increase in Current Year Performance will raise the odds in favor of y = 0 versus choices y =1, 2 and 3 combined by the factor 1.019. Similarly, a unit increase in Current Year Performance will raise the odds in favor of the choices y = 0 and y = 1 versus choices y = 2 and y = 3 by the factor 1.019. Finally, a unit increase in Current Year Performance will raise the odds in favor of choices y = 0, 1 and 2 versus choice y = 3 by the factor 1.019. Notice that the change in the odds is the same no matter which choice is the focus of the analysis. This result is called the *proportional odds* assumption, and it is a feature of the Order Logit model. Whether this assumption makes any sense needs to be considered by the researcher. However, one can test if this assumption is valid, much as one can test the Independence of Irrelevant Alternatives assumption. If the assumption of proportional odds is rejected, then the Ordered Logit model is called into question and an alternative model should be

---

[32] Note the negative sign in front of the estimated coefficient, in contrast to odds effect for the unordered Logit model.

[33] Since the values of the marginal effects can vary widely depending on the values chosen for the variables, the magnitude for the change in probability is strictly valid only for a small change in a variable. A more complete analysis would consider the discrete change in the predicted probability and also calculate the marginal effect for a

sought. In this regard, the Multinomial Logit model for unordered choices could instead be used since it does not impose the proportional odds assumption.

Summary of Multinomial Model Methods

An overview of the elements discussed for ordered and unordered multinomial limited dependent variable techniques is provided in TABLE 7. The table provides insights on key assumptions underlying the models as well as what we feel researchers should report when presenting the results of their analysis.

Insert Table 7 About Here

CENSORED AND TRUNCATED LIMITED DEPENDENT VARIABLES

Despite the increasing application of Logit and Probit in empirical strategy research, most strategy research still utilizes continuous rather than discrete measures for the dependent variable. Strategy researchers, for example, routinely seek to examine factors that may explain the extent of a specific strategic activity (e.g., corporate refocusing or diversification). Yet not all the firms in the population of interest may chose to engage in the activity of interest. For example, when examining the extent of diversification among firms, many firms will not pursue diversification. This results in a data sample for which a significant number of observations have a single common value for the dependent variable. Samples wherein the dependent variable has the same specific value for several observations is also likely when examining, for example, performance outcomes that fall below a certain target level (Reuer and Leiblein, 2000) or when examining equity ownership since many firms will own 100% of their foreign operations (Delios and Beamish, 1999).[34] In such situations, the researcher is faced with a censored dependent variable that takes a common value for many of the observations as well a set of continuous values for other observations. In such cases, OLS will fail to account for the different nature of the observations that take the single common (discrete) value and those observations with continuous values and will result in estimates that are biased and inconsistent. Consequently,

---

wide range of values of a given variable. See Long (1997, pp. 127-138) for discussion of the different types of analyses one can undertake with respect to the marginal effects in an Ordered model.

[34] Censoring of the dependent variable can also arise when interval values of the dependent variable are reported. This is common with financial performance data where the researcher reports performance intervals rather than

one's inferences about the relationship between the dependent variable and the independent variables are unlikely to be valid.

In addition to the case of a censored dependent variable, strategy researchers often non-randomly select a subset of the broader population and thus use data samples that do not encompass the entire population of interest. For example, it is common to limit a sample to only the largest of public firms (e.g. *Fortune 1000*) but to then interpret the findings as if they apply to the whole population of public companies. Another frequent research design for which a truncated dependent variable arises is when the researcher deliberately selects a sample based only on certain observed values of the dependent variable, e.g. studying only firms that exhibit CEO turnover (Zajac and Westphal, 1996) or IPO firms that are only covered by financial analysts (Raghuram and Servaes, 1997). By excluding a subset of the population (e.g. firms that do not engage in or exhibit a particular phenomenon), values of the dependent variable are not observed over a range of its population values, resulting in a truncated dependent variable. When a dependent variable is truncated, the use of OLS to estimate the model leads to biased and inconsistent estimates of the parameters. Without accounting for truncation of the dependent variable, one cannot directly infer from the truncated sample how firms not represented in the sample would respond, and the coefficient estimates one obtains will not represent the estimates that one would obtain if one had sampled values of the dependent variable from entire population.

When a data sample comprises only truncated values of the dependent variable the key issue that arises is that the mean of this variable will not equal its population mean. The correction for this problem leads to the Truncated Regression model. If the dependent variable is instead censored, one must also model the discrete distribution of the data represented by the common limit value of the dependent variable and the continuous distribution of dependent variable values that lie above (below) the common limit value. The model that arises in such cases is the Censored Regression or Tobit model.

Truncated Regression Model

To understand the Truncated Regression model we need to first understand how truncation of a variable biases inferences about the population mean of the variable. The distribution of a truncated variable is that part of a variable's population distribution that lies above (or below) some particular value of the variable. This particular value is called the

---

continuous performance data. For example, one might report ROA as: less than zero, 0 to 10%, 11 to 20%, etc. In such cases the Grouped Data model can be used (see footnote 29).

truncation point and is denoted below as "t." Assume the variable y* has a Normal distribution with mean μ and variance σ^2 and denote the values of y* actually observed as y. With truncation, we only observe the values y when the underlying population variable y* takes values greater than t. Given this, it can be shown that the mean of the observed y values is

$$E(y|y^* > t) = \mu + \sigma \lambda(\alpha) \qquad (17)$$

where $\alpha = (\mu - t)/\sigma$. The ratio $\lambda = \phi(\alpha)/\Phi(\alpha)$ is called the inverse Mills ratio. This ratio is the ratio of the Normal pdf to the Normal cdf evaluated at the standardized value α.[35] Since both σ and λ in equation (17) are positive, this equation confirms that the mean of the truncated variable y (i.e., E(y | y* > t)) exceeds the population mean (μ). Equation (17) is a general expression for the mean of a truncated random variable and it will be used below when deriving both the truncated and censored regression models.

The Truncated Regression model arises when one takes into account that the observed sample values of y represent values from a truncated distribution. As is now shown, the reason OLS is inappropriate when the dependent variable is truncated is because the error term in the usual regression model will not have zero mean.

In a standard regression model the mean of the dependent variable in the population as whole is assumed to be a linear function of variables $\mathbf{x}_i$:

$$E(y_i^* \mid \mathbf{x}_i) = \mu_i = \mathbf{x}'_i \beta$$

However, if the dependent variable is truncated, not all values of y* are observed. Instead, only the values for which y* > t are observed. The model for the observed values of y* (i.e., y) is then the usual regression model that includes an error term:

$$y_i = \mathbf{x}'_i \beta + \varepsilon_i \quad \text{for } y_i = y_i^* > t$$

Taking expectations of this model gives

$$E(y_i \mid y_i^* > t, \mathbf{x}) = E(\mathbf{x}'_i \beta + \varepsilon_i \mid y_i^* > t, \mathbf{x})$$

$$E(y_i \mid y_i^* > t, \mathbf{x}) = \mathbf{x}'_i \boldsymbol{\beta} + E(\varepsilon_i \mid y_i^* > t, \mathbf{x}) \qquad (18)$$

In the standard regression model the expected value of the error, $E(\varepsilon_i \mid \mathbf{x})$, is zero. In this standard case the error term in (18) would drop out and we would have the usual result that the expectation of the dependent variable equals its population mean $\mathbf{x}'_i \beta$. However the expectation of the error in (18) is not over the population distribution associated with y* but instead only over the truncated distribution of the observed values y. To determine $E(\varepsilon_i \mid y_i^* > t, \mathbf{x}_i)$ we use the fact that $\varepsilon_i = y_i - \mathbf{x}'_i\beta$ to evaluate $E(y_i - \mathbf{x}'_i\beta \mid y_i^* > t, \mathbf{x}_i)$. Doing this gives

$$E(y_i - \mathbf{x}_i'\beta \mid y^*_i > t, \mathbf{x}_i) = E(y_i \mid y^*_i > t, \mathbf{x}_i) - E(\mathbf{x}_i'\beta \mid y^*_i > t, \mathbf{x}_i)$$

$$E(y_i - \mathbf{x}_i'\beta \mid y^*_i > t, \mathbf{x}_i) = E(y_i \mid y^*_i > t, \mathbf{x}_i) - \mathbf{x}_i\beta$$

The first term on the RHS of this equation is just the mean of a truncated distribution. The expression for this mean is given by (17). Using this result the above becomes

$$E(y_i - \mathbf{x}_i'\beta \mid y^*_i > t, \mathbf{x}_i) = \mathbf{x}_i'\beta + \sigma\lambda(\alpha) - \mathbf{x}_i'\beta$$

$$E(y_i - \mathbf{x}_i'\beta \mid y^*_i > t, \mathbf{x}_i) = \sigma\lambda(\alpha)$$

Inserting this expression into (18) then gives

$$E(y_i \mid y^*_i > t, \mathbf{x}_i) = \mathbf{x}_i'\beta + \sigma\lambda\left(\frac{(\mathbf{x}_i'\beta - t)}{\sigma}\right) \qquad (19)$$

Let $\gamma_i = (\mathbf{x}_i'\beta - t)/\sigma$ so we can write $\lambda_i = \lambda(\gamma_i)$. Since λ_i varies across observations it can be treated it as a variable with coefficient σ. This suggests writing (19) as the following regression model

$$y_i = \mathbf{x}_i'\beta + \sigma\lambda_i + \hat{e} \qquad (20)$$

Equation (20) is the Truncated Regression model. Since this model includes the variable λ_i, it indicates that a standard regression of y_i on $\mathbf{x}_i$ alone would exclude λ_i, and therefore result in biased estimates of the β due to an omitted variables bias. The Truncated Regression model in (20) is estimated using Maximum Likelihood after one specifies the value of the truncation point t.

Model Significance and Interpretation

Examining the goodness of fit and significance of the Truncated Regression model proceeds as for any model estimated using Maximum Likelihood. That is, one reports the maximized value of the log-likelihood and pseudo R-square, and tests overall model significance using the LR statistic that compares the full model to the model with only a constant term.

If the overall model is significant then one can consider the significance and interpretation of individual variables. All the usual tests for coefficient significance apply. Interpretation centers on the marginal effects for the model. For the Truncated Regression model there are two marginal effects to consider: one is the effect of a change in a variable x_k on the value of the dependent variable y* in the population and the second is the effect on the value of the dependent variable y in the truncated sub-population. Since the mean of the population

[35] The inverse Mill's ratio appears consistently in the analysis of truncated and censored distributions. This ratio goes to zero as the truncation point t moves further and further to the left (assuming truncation from below) so that,

variable y* is linearly related to $\mathbf{x}_i'b$, i.e., E(y* | $\mathbf{x}_i$) = $\mathbf{x}_i'b$, the first marginal effect is just the estimated coefficient β_k - this is the marginal effect that applies to the population as whole.[36] The marginal effect of a change in x_k on y in the *sub-population* is not β_k. This marginal effect is instead

$$\frac{\partial \mathrm{E}\,(\mathrm{y}\,|\,\mathrm{y}^* > \mathrm{t}, \mathbf{x})}{\partial \mathrm{x}_\mathrm{k}} = b_\mathrm{k}\left[1 + g_\mathrm{i}\, l\,(g_\mathrm{i}) - l\,(g_\mathrm{i})^2\right] \qquad (21)$$

where $g_\mathrm{i} = (\mathrm{t} - \mathbf{x}_\mathrm{i}'\mathbf{ß})/s$. Greene (2002, p. 760) shows that the term in square brackets lies between 0 and 1. Since the term in square brackets in (21) is positive, the directional effect (i.e. sign) of a change in an independent variable on the dependent variable in the sub-population (y) is the same as that for the dependent variable in the full population (y*). In addition, since the term in square brackets is less than one, the marginal effect is less that the corresponding coefficient (β_k).

Censored Regression Model

When a continuous dependent variable has a cluster of observations that take a specific value the Censored Regression or Tobit model applies (Tobin, 1958). In the standard Censored Regression model the relationship between the population variable y* and the observed values y is as follows

$$\mathrm{y}_\mathrm{i} = \begin{cases} \mathrm{y}_\mathrm{i}^* = \mathbf{x}_\mathrm{i}'\mathbf{ß} + e_\mathrm{i} & \text{if } \mathrm{y}^* > \mathrm{t} \\ \mathrm{t}_\mathrm{y} & \text{if } \mathrm{y}^* \le \mathrm{t} \end{cases}$$

Here, “t” is the censoring point and t_y is the value taken by the dependent variable if the value of y* is at or below the censoring point. This is like the case of a truncated variable except values of y* at or below the truncation point are not discarded, they are instead all assigned the same limit value t_y. As with truncation, the issue is the expression for the mean of the censored variable y. Using arguments similar to the case of a truncated variable, one can show that the mean of the censored variable is:

$$\mathrm{E}(\mathrm{y}_\mathrm{i}\,|\,\mathrm{y}^* \ge \mathrm{t}, \mathbf{x}_\mathrm{i}) = \Phi(g_\mathrm{i})\,\mathbf{x}_\mathrm{i}\mathbf{ß} + s\,f\,(g_\mathrm{i}) + \Phi(-g_\mathrm{i})\,\mathrm{t}_\mathrm{y} \qquad (22)$$

where $g_\mathrm{i} = (\mathbf{x}_\mathrm{i}'\mathbf{ß} - \mathrm{t})/s$. As with the Truncated Regression model, the conditional mean of the censored variable is a nonlinear function of **x (**since it involves both the cdf and the pdf of the Normal distribution). Also like the case of truncation, (22) implies that an OLS regression of

in the limit, the mean of a truncated variable will equal the mean in the full population.

[36] Important to note is that β_k is not an OLS estimate but is instead the Maximum Likelihood estimate derived by taking into account the truncation of the population variable y*.

y on **x** alone would exclude the "variables" $f(g_{\text{i}})$ and $\Phi(-g_{\text{i}})$ and hence result in coefficient estimates that are both biased and inconsistent due to an omitted variables bias.

Maximum Likelihood is used to estimate the Censored Regression model. The likelihood function for the model is a mixture of a discrete distribution (when y takes the censored value t_y) and a continuous distribution (when y takes values above (below) the censored value t_y). Details of the likelihood function and its estimation can be found in Greene (2002, pp. 766-768) and Long (1997, pp. 204-206), among others.

Model Significance and Interpretation

To examining the goodness of fit and significance of the Censored Regression model one reports the maximized value of the log-likelihood and the pseudo R-square, and tests overall model significance using the LR test that compares the full model to the model with only a constant term.

If the overall model is significant then one can consider the significance and interpretation of individual variables. All the usual tests for coefficient significance apply, and interpretation centers on the marginal effects for the model. As with the Truncated Regression model, there are two marginal effects to consider: the one that applies to the population variable y* and the one applies to the observed values y. The marginal effect for y* is again just the estimated coefficient $b_{\text{k}} = \partial \text{E}(\text{y}^*|\,\mathbf{x})/\partial \text{x}_{\text{k}}$. The marginal effect for the observed values of y (both censored and uncensored) is obtained by differentiating (22) with respect to variable x_k. The result is

$$\frac{\partial \text{E}(\text{y}|\,\mathbf{x})}{\partial \text{x}_{\text{k}}} = \Phi(g)\,b_{\text{k}} + (\text{t} - \text{t}_{\text{y}})\Phi(g)\,\frac{b_{\text{k}}}{s} \qquad (23)$$

In the standard Tobit model the truncation point (t) and the limit value (t_y) are assumed to equal zero. Setting $t = t_y = 0$ in (23), the marginal effect in the sub-population of uncensored observations is

$$\frac{\partial \text{E}(\text{y}|\,\mathbf{x})}{\partial \text{x}_{\text{k}}} = \Phi\left(\frac{\mathbf{x}_{\text{i}}'\boldsymbol{\beta}}{s}\right)b_{\text{k}} \qquad (24)$$

Hence, the marginal effect in this case is just a variable's coefficient multiplied by the proportion of uncensored observations in the sample (which is the probability that an observation is uncensored).

A recent study by the authors (Bowen and Wiersema, 2003) that examined the effect of import competition on the level of firm diversification can be used to illustrate the Tobit model. In their study, censoring arose because the data sample included a number of single business firms whose measured value of the dependent variable (diversification) was zero. Since the sample contained a high proportion of zero values (60%) for the dependent variable this dictated the use of a Tobit model rather than OLS.

Insert Table 8 About Here

TABLE 8 shows TOBIT estimates for one version of the model along with two sets of OLS estimates derived using two alternative data samples: the full sample that includes both single business and diversified firms (the sample used for the Tobit) and a sample that excludes the censored observations (i.e., single business firms).[37] For comparison to the OLS estimates, the marginal effects associated with the continuous variables in the Tobit model are also reported. While the sign of the estimated coefficient for the key variable of interest, import penetration, is the same for the Tobit and OLS models, the sign and significance of other variables is often different, and indicates the extent to which the estimates are sensitive to the estimation procedure used.

Numerous extensions have been made to the original Tobit model to allow, for example, both lower and upper censoring of the dependent variable and the limit value to vary by observation. In addition, the basic model assumes homoscedastic error variances but this assumption is easily relaxed to allow for a relatively general form of heteroscedasticity. Most damaging to the Tobit model is violation of the assumption of normality of y*, since violation of this assumption produces inconsistent Maximum Likelihood estimates (see Greene, 2002, pp. 771-772).

Sample Selection Model

In the Truncated Regression model one knows the value of the truncation point beyond which values of the dependent are not observed. However, in some cases one may be able to say more about the nature of the truncation of the dependent variable. In particular, one may be able specify a mechanism that systematically explains how the truncated observations arise. If so, the model that incorporates this mechanism is called the Sample Selection model.

[37] Excluding the censored observations creates a truncated dependent variable.

The basic Sample Selection model contains two equations (Heckman, 1976). The first, as in the truncated model, is the equation for the population variable y* that is of primary interest to the researcher

$$y_i^* = \mathbf{x}_i'\mathbf{ß} + e_i \qquad (25)$$

The second equation is the "selection equation" which determines when values of y* are observed:

$$z_i^* = \mathbf{w}_i'\mathbf{?} + h_i \qquad (26)$$

The variables **w** that determine the z* may include the same variables as in **x**. The rule adopted is that values of y* are observed when z* > 0. The model then assumes that the errors ε and η have a bivariate Normal distribution with mean zero and correlation coefficient ρ. Using results for a truncated bivariate Normal distribution, the following equation for the mean of the observed variable y can be derived:

$$\mathrm{E}(y_i \mid z_i^* > 0) = \mathbf{x}_i'\mathbf{ß} + d\,l\,(a_i) + n_i \qquad (27)$$

where $a_i = -\mathbf{w}_i'\mathbf{?}/s_h$ and $l\,(a_i) =_i f\left(\mathbf{w}_i'\mathbf{?}/s_h\right)/\Phi\left(\mathbf{w}_i'\mathbf{?}/s_h\right)$. As in the truncated regression model, λ is the inverse Mills ratio, but this time evaluated at values of the selection variable z* (compare (27) with (20)). Also like the Truncated Regression model, (27) implies that not accounting for the selection mechanism, and so regressing y on **x** alone, will result in biased and inconsistent estimates of the b due to an omitted variables bias. However, unlike the truncated model, even if the OLS regression were restricted to the sample of truncated observations, the estimates obtained would not be efficient since the error ν_i in (27) can be shown to be heteroscedastic (Greene, 2002, p. 783).

In practice the values of z* are rarely observed. Instead, only the "sign" of z* is observed. This means, for example, that one only observes if a firm has or has not entered a joint venture. In such cases the selection equation (26) is then modeled as a binary Probit[38] where the observed values of z are: z = 1 if z* > 0 and z = 0 when z* < 0. This leads to a reformulation of equation (27) in terms of the observed values z:

$$\mathrm{E}(y_i \mid z_i = 1, \mathbf{x_i}, \mathbf{w_i}) = \mathbf{x}_i'\mathbf{ß} + r\,s_e\,l\,(\mathbf{w}_i'\mathbf{?})$$

This expression can be written more compactly as

$$y_i = \mathbf{x}_i'\mathbf{ß} + d\,l_i + n_i \qquad (28)$$

[38] A Probit is used since the error η in the selection equation (26) is assumed to be normally distributed.

where $\delta = \rho\sigma_\varepsilon$. Estimation of (28) is usually based on a two-step estimator.[39] In the first step a Probit model for z using variables **w** is estimated. The estimated values of the coefficients g together with the data **w** are then used to calculate the values $l_i = f\left(\mathbf{w}_i'\mathbf{?}\right)/\Phi\left(\mathbf{w}_i'\mathbf{?}\right)$. The (consistently) estimated values of the λ_i are then used as data, along with the variables in **x**, to estimate (28) using OLS. Since OLS is used in the second step the interpretation and testing of the estimated coefficients proceeds as usual, the only difference being that the estimates are now unbiased and consistent having been "corrected" for the selection bias. Finally, since the coefficient δ is directly related to the correlation between the errors in the selection model and the structural model, if this coefficient is not significantly different from zero it suggests that the selection mechanism plays no role in generating the values of the observed dependent variable y.

The focus of the Sample Selection model is that observed values of a dependent variable may arise from some form of systematic non-random sampling (i.e., the selection equation) and the deleterious effect of the selection bias that results if the structural model of interest is estimated with OLS. The issue of systematic non-random sampling has important implications for many of the issues studied by researchers in strategic management. To understand why, consider the frequently examined relationship between firm performance and diversification strategy in which researchers are interested whether or not firms that pursue a diversification strategy outperform firms that do not diversify. The structural relationship is usually modeled as a linear relationship between firm performance and the level of diversification (as one of several independent variables) and the model is estimated using OLS. Will the estimated coefficient on the diversification variable accurately indicate the impact of being diversified on firm performance? The answer is "no" if a firm's decision to be diversified is related to its performance.

The issue that leads to this negative answer is called the problem of "self-selection," and it is a direct application of the sample selection problem studied by the Sample Selection model. In terms of the performance/diversification relationship, the problem is that if a firm's choice to become diversified is a response to poor (good) performance then the sample of firms will be biased in favor of poorly (well) performing firms. In terms of the Sample Selection model, if one fails to account for how firms "self-select" themselves to become diversified then selection bias is an issue. This implies that the simple OLS estimate for the coefficient on the diversification variable in the structural model will be biased. Since theoretically it has been argued that one

[39] The model can also be estimated using Full Information Maximum Likelihood.

reason firms choose to diversify is as a defensive response to declining market and profit opportunities in their core businesses, the issue of self-selection bias is directly relevant.

A self-selection bias may in fact account for the widely different findings that have been reported in the literature for the relationship between firm performance and diversification. In particular, if poor performance is a factor that influences the firm's decision to become diversified then the errors in the selection equation and in the structural model (between performance and diversification) are negatively correlated (i.e., $\rho < 0$).[40] This implies that the coefficient δ in (28) would be negative. Hence, a simple OLS estimate for the coefficient on the diversification variable could be positive or negative, depending on the sign and relative magnitudes of the true effect β compared to the negative selection coefficient δ in equation (28). To overcome this bias, one should apply the Sample Selection model. This would mean, as per equation (28), first modeling the (selection) decision to be diversified (where $z = 1$ if a firm is diversified) in terms of performance and perhaps other variables, and to then use the estimated values from this equation to compute the values λ_i that are then used as data in the structural model between performance and diversification.

The problem of self-selection bias can arise whenever a firm can choose to undertake a particular strategic action based on an outcome of the firm (e.g., performance), and the focus of one's study is to determine the effect of that particular strategic action on that outcome of the firm (e.g., performance). Given this, the problem of self-selection, and more generally non-random sample selection, may be endemic to many of the questions examined in strategic management research, since much of this research seeks to understand the consequences of strategic choices on firm performance. Researchers in strategic management have given little, if any, attention to the important issue of self-selection and the bias it introduces if a model is estimated by OLS. We feel strongly that the Sample Selection model should be an integral part of any future empirical work that seeks to model an outcome for the firm, such as performance, in relation to the strategic choices of the firm.

[40] The issue of a self-selection bias is different from the issue of simultaneity bias that often plagues the models examined by researchers in strategic management. A simultaneity bias arises when the dependent variable and one or more of the dependent variables are jointly determined. For the performance/ diversification relationship, simultaneity means that performance determines diversification and diversification determines performance. Failure to account for simultaneity leads to biased OLS estimates. However, any simultaneity bias that might arise is additional to the bias induced by self-selection, since the self-selection problem deals with the issue of a non-randomly selected data sample.

Summary of Censored and Truncated Limited Dependent Variables

The continuing prevalence of continuous dependent variables in empirical strategy research makes the issues of censoring, truncation, and sample selection bias important statistical issues that need to be confronted. For many phenomenon, the researcher will have a cluster of responses that take a common value which raises the issue of censoring. In such cases one should use the Censored Regression model to properly account for the mixture of discrete and continuous data that arises due to the censored nature of the dependent variable.

A researcher whose dependent variable is only observed over a restricted range of the total population of values is faced with a truncated dependent variable. In such cases the appropriate model is the Truncated Regression model.

Perhaps the most serious issue facing researchers in strategic management is the issue of sample selection (bias). As discussed, the issue of a biased sample induced by the problem of self-selection may be endemic to strategy research given that strategic choices (e.g. to expand overseas, to enter a joint venture, to replace the CEO) may themselves depend on the dependent variable that is the focus of one's study. The researcher should therefore carefully consider the relationship they intend to study to assess if a sample selection problem might exist, regardless if the problem is due to self-selection or to the more general form of non-random sample selection. If a selection problem is suspect, one should use the Sample Selection model to account of the way the observations arise, and to then obtain unbiased and consistent estimates for the parameters that are the focus of their study.

CONCLUSION

The use of discrete limited dependent variable models has grown in recent years as researchers increasingly examine strategic phenomenon that can be represented as discrete choices or organizational outcomes. Researchers therefore need to learn the proper use of discrete LDV techniques and the methods for interpreting the results obtained. Based on a review of studies that have used LDV techniques in recent issues of the *Strategic Management Journal*, many researchers do not fully and accurately report their results, and in many instances make erroneous interpretations about the relationship studied. The problem may be due to a lack of familiarity with these techniques and confusion over how to interpret the direction and magnitude of the relationship between the dependent and independent variables. Unlike OLS, the coefficients estimated in a discrete LDV model are almost never an accurate indicator of the nature of the relationship modeled. Our discussion of alternative discrete LDV models was therefore intended to address the observed shortcomings that past strategy research has displayed, and to illustrate and recommend how researchers can interpret and report the results from such models.

While the use of discrete LDV models is growing in the literature, the majority of studies continue to examine a dependent variable that takes continuous values. In this context, we discussed three important cases in which a LDV can arise: censoring, truncation, and non-random sample selection. For each of these cases it was shown that the use of OLS would lead to biased and inconsistent estimates of model parameters. The most important issue considered was the general problem of sample selection. In particular, biased samples due to self-selection may be a problem endemic to the kinds of issues commonly addressed by researchers in strategic management. We again stress that the issue of sample selection, and in particular self-selection, and the bias it introduces needs to be taken much more seriously by researchers.

By providing an investigation of the common errors that have prevailed in the use of these methods this chapter sought to motivate researchers to be accurate and consistent in how they estimate and interpret LDV models. In raising awareness of statistical and interpretation issues for common discrete LDV models, as well the issues of censoring, truncation and sample selection in the context of a continuous LDV, we hope that strategy researchers can conduct analyses that offer sound and statistically correct conclusions.

REFERENCES

Amemiya, T. (1981). Qualitative Response Models: A Survey. Journal of Econometrics, 19(4):481-536.

Barkema, H. and F. Vermeulen. (1998). International Expansion Through Start-Up or Acquisition: A Learning Perspective. *Academy of Management Journal*, 41:7-26.

Belderbos R. and L. Sleuwaegen (2003), International Plant Configuration Strategies: A Structured Decision Making Approach and Product Level Test, *Vlerick-Leuven-Gent Management School Working Paper Number 2003/02*.

Bergh, D. and J. Fairbank. (1995). Measuring and testing change in strategic management research. Strategic Management Journal, 23:359.

Bergh, D. and G. Holbein. (1997). Assessment and redirection of longitudinal analysis: Demonstration with a study of the diversification and divestiture relationship. *Strategic Management Journal*, 18:557-571.

Bowen, H. and M. Wiersema. (1999). Matching Method to Paradigm in Strategy Research: Limitations of Cross-Sectional Analysis and Some Methodological Alternatives. *Strategic Management Journal, Vol.* 20:625-636.

Bowen, H. and Wiersema, M. (2003). The Impact of Foreign-Based Competition on Firm Diversification: A Resource-Based Perspective. *Vlerick-Leuven-Gent Management School Working Paper Number 2003/20*.

Chang, S.J. (1996). An Evolutionary Perspective on Diversification and Corporate Restructuring; Entry, Exit, and Economic Performance During 1981-89. *Strategic Management Journal*, 17:587-611.

Chang, S.J. and H. Singh. (1999). The Impact of Modes of Entry and Resource Fit on Modes of Exit by Multibusiness Firms. *Strategic Management Journal*, 20:1019-1035.

Chatterjee, S. and J. Harrison, D. Bergh. (2003). Failed Takeover Attempts, Corporate Governance and Refocusing. *Strategic Management Journal*,24:87-96.

Chung, S., H. Singh, and K. Lee. (2000). Complementarity, Status Similarity and Social Capital as Drivers of Alliance Formation. *Strategic Management Journal*, 21:1-22..

Davidson, R. and J. MacKnnon (1984). Convenient Specification Tests for Logit and Probit Models. *Journal of Econometrics*, 25:241-262

Delios, A. and P.W. Beamish. (1999). Ownership Strategy of Japanese Firms: Transactional, Institutional, and Experience Influences. *Strategic Management Journal*, 20:915-933.

Eisenmann, T.R. (2002). The Effects of CEO Equity Ownership and Firm Diversification on Risk Taking. Strategic Management Journal, Vol.23:513-534.

Eliason, S. (1993). *Maximum Likelihood Estimation: Logic and Practice*. Sage Publication. Newbury Park, CA.

Greene, William H.(2002). *Econometric Analysis, Fifth Edition (International Edition).* Prentice-Hall.

Gulati, R. (1999). Network Location and Learning: The Influence of Network Resources and Firm Capabilities on Alliance Formation. *Strategic Management Journal*, 20:397-420.

Harzing, A. (2002). Acquisitions Versus Greenfield Investments: International Strategy and Management of Entry Models. *Strategic management Journal*, 23:211-227.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5:475-492.

Kogut, B. and H. Singh. (1988). The Effect of National culture on the Choice of Entry Mode. *Journal of International Business.* Vol. 19:411-433.

Long, J. Scott. (1997). *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oaks, CA: Sage.

Lubatkin, M., Merchant, H., and N. Srinivasan. (1993). Construct validity of some unweighted product-count diversification measures. *Strategic Management Journal*, 14:433-449.

Maddala, G. (1983). *Limited Dependent Variables in Econometrics*. New York: Cambridge University Press

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of Royal Statistical Society*, 42:109-142.

McDonald, J. and R. Moffitt (1980). The Uses of Tobit Analysis. *Review of Economics and Statistics*, 62:318-321.

McKelvey, R. and W. Zaviona. (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology*, 4:103-120.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers of Econometrics*. New York: Academic Press.

McNamara, G. and P. Bromiley. (1997). Decision Making in an Organizational Setting: Cognitive and Organizational Influences on Risk Assessment in Commercial Lending. *Academy of Management Journal*, 40:1063-1088.

Pan, Y. and P.S.K. Chi. (1999). Financial Performance and Survival of Multinational Corporations in China. *Strategic Management Journal*, 20:359-374.

Powers, Daniel A. and Yu Xie. (1999). *Statistical Methods for Categorical Data Analysis*, Academic Press, Inc.

Raghuram, R. and H. Servaes. (1997). Analyst Following of Initial Public Offerings. *The Journal of Finance*, 52(2):507-530.

Reuer, J.J. and M. J. Leiblein. (2000). Downside Risk Implications of Multinationality and International Joint Ventures. *Academy of Management Journal*, 43:203-214.

Robins, J. and M. Wiersema (2003). The Measurement of Corporate Portfolio Strategy: Analysis of the Content Validity of Related Diversification Indexes. *Strategic Management Journal*, 24:39-59.

Scandura, T. and E. Williams. (2000). Research Methodology in Management: Current Practices, Trends, and Implications for Future Research. *Academy of Management Journal*, 43:1248-1264.

Somaya, D. (2003). Strategic Determinants of Decisions Not To Settle Patent Litigation. *Strategic Management Journal*, 24:17-38.

Song, J. (2002). Firm Capabilities and Technology Ladders: Sequential Foreign Direct Investments of Japanese Electronics Firms in East Asia. *Strategic Management Journal*, 23:191-210.

Thiel, H. (1969). A Multinomial Extension of the Linear Logit Model. International Economic Review, 10:251-259.

Toulan, O. (2002). The Impact of Market Liberalization on Vertical Scope: The Case of Argentina. *Strategic Management Journal*, 23:551-560.

Vermeulen, F. and H. Barkema. (2001). Learning Through Acquisitions. *Academy of Management Journal*, 44:457-476.

Zajac, E. and J. Westphal. (1996). Who Shall Succeed? How CEO/Board Preferences and Power Affect the Choice of New CEOs. *Academy of Management Journal*, 39:64-91.

TABLE 1

Binary Logit Results for Predicting CEO Replacement Type

Variables	**Null Model**	**Full Model**	**Marginal Effects**	**Odds Effects**
Succession Type		1.911***	0.2923	6.762
Pre-Succession Stock Performance		-0.021***	-.00276	0.979
Constant	-1.277***	-2.10***		
Log likelihood	-98.601	-77.925		
χ^2 value (2 dof)		41.35		
p-value		0.000		
Pseudo R^2		0.210		
Percent Correctly Classified		80.3 %		
Observations	188	188		

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

TABLE 2

Tabulation of Actual versus Predicted Choices for CEO Replacement Type

Replacement CEO Type		**Predicted Choice**		**Totals**
		Insider	Outsider	Total
Actual Choice	Insider	135	12	147
	Outsider	25	16	41
	Totals	160	28	188

TABLE 3

Summary of Issues for Binary Models

	Binary Logit	Binary Probit
Key Assumptions	▪ Model error has standard Logistic distribution ▪ Error variance homoscedastic	▪ Model errors have standard Nor distribution ▪ Error variance homoscedastic
What to Report	▪ Maximized value of log-likelihood ▪ Pseudo-R-square ▪ Percent of correct predictions ▪ Chi-square value and p-value for likelihood ratio test of full model against model with only a constant term	▪ Maximized value of log-likeliho ▪ Pseudo-R-square ▪ Percent of correct predictions ▪ Chi-square value and p-value fo likelihood ratio test of full model against model with only a constant term
Interpreting Coefficients	▪ Sign of coefficient indicates directional effect on probability that Y =1 ▪ Size of coefficient does not indicate size of effect on probability of Y = 1. Need to compute marginal effect. ▪ Exponential of a coefficient indicates change in odds in favor of Y = 1 due to a one unit change in a variable.	▪ Sign of coefficient indicates directional effect on probability that Y =1 ▪ Size of coefficient does not indic size of effect on probability of Y =1. Need to compute marginal effect.
Marginal Effects	▪ Depend on values of all variables and coefficients ▪ Compute using 1) expression for derivative of Logit function or 2) as discrete change in probability. ▪ Must hold fixed values of all other variables	▪ Depend on values of all variable and coefficients ▪ Compute using 1) expression fo derivative of Probit function or 2) as discrete change in probability. ▪ Must hold fixed values of all oth variables

TABLE 4

Multinomial Logit Model Predicting CEO Succession

Choice Alternative / Variables	Model Coefficients	Marginal Effects (at variable means)	Odds Effect
Y=0 : Routine/Insider CEO [a]			
Succession Year Performance	zero	0.0055***	n/a
Pre-Succession Performance	zero	0.0063***	n/a
Constant	zero	n/a	n/a
Y=1 : Routine/Outsider CEO			
Succession Year Performance	-0.003	0.0003	0.9965
Pre-Succession Performance	-0.025*	-0.0008	0.9755
Constant	-2.02*	n/a	n/a
Y=2 : Dismissal/Insider CEO			
Succession Year Performance	-0.030***	-0.0041***	0.9707
Pre-Succession Performance	-0.017**	-0.0013	0.9832
Constant	0.67***	n/a	n/a
Y=3 : Dismissal/Outsider CEO			
Succession Year Performance	-0.022***	-0.0018**	0.9785
Pre-Succession Performance	-0.041***	-0.0042	0.9602
Constant	-0.70***	n/a	n/a
<u>Model Information</u>			
Log-likelihood	-175.55		
χ^2 value (6 dof)	65.16		
p-value	0.0000		
Pseudo R^2	0.156		
% Correctly Classified	65.43%		
Observations	188		

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

[a] Since Y=0 (routine succession and insider CEO replacement) is the base choice all coefficients for choice Y = 0 are restricted to equal zero.

TABLE 5

Tabulation of Actual versus Predicted Choices for CEO Succession

CEO Succession Type		Predicted Choice				Totals
		Routine/ Insider	Routine/ Outsider	Dismissal/ Insider	Dismissal/ Outsider	
Actual Choice	Routine/Insider	100	0	2	3	105
	Routine/Outsider	7	0	2	0	9
	Dismissal/Insider	23	0	12	7	42
	Dismissal/Outsider	11	0	10	11	32
Totals		141	0	26	21	188

TABLE 6

Ordered Logit Model Predicting Succession Type

Variable	Model Coefficients	Marginal Effects (at variable means)				Odds Effects
		Y=0 Routine/ Insider CEO	Y=1 Routine/ Outsider CEO	Y=2 Dismissal/ Insider CEO	Y=3 Dismissal/ Outsider CEO	
Succession Year Performance	-0.019***	0.0046***	-0.0003*	-0.0024***	-0.0019***	1.019
Pre-Succession Performance	-0.024***	0.0058***	-0.0003*	-0.0030***	-0.0024***	1.024
Cut-point 1 (t_1)	-0.1447					
Cut-point 2 (t_2)	0.1150					
Cut-point 3 (t_3)	1.5562					

Model Information

Log likelihood	-182.333
χ^2 value (2 dof)	51.58
p-value	0.0000
Pseudo R^2	0.124
% Correctly Classified	62.23%
Observations	188

* $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

TABLE 7

Summary of Issues for Multinomial Models

	Multinomial Logit	Multinomial Probit	Ordered Logit	Ordered Probit
When to Use	▪ Values of dependent variable are discrete and unordered	▪ Values of dependent variable are discrete and unordered	▪ Values of dependent variable are discrete and rank ordered	▪ Values of dependent variable are discrete and rank ordered
Key Assumptions	▪ Model error has standard Logistic distribution ▪ Error variance homoscedastic ▪ Independence of Irrelevant Alternatives (can test this)	▪ Model error has standard Multivariate Normal distribution ▪ Error variance homoscedastic	▪ Model error has standard Logistic distribution ▪ Error variance homoscedastic ▪ Odds across choices are proportional (can test this)	▪ Model errors have standard Normal distribution ▪ Error variance homoscedastic
What to Report	▪ Model log-likelihood ▪ Percent correct predictions ▪ Pseudo-r-square ▪ Likelihood ratio test of full model against model with only a constant term	▪ Model log-likelihood ▪ Percent correct predictions ▪ Pseudo-r-square ▪ Likelihood ratio test of full model against model with only a constant term	▪ Model log-likelihood ▪ Pseudo-r-square ▪ Likelihood ratio test of full model against model with only a constant term	▪ Model log-likelihood ▪ Pseudo-r-square ▪ Likelihood ratio test of full model against model with only a constant term
Interpreting Coefficients	▪ Must compute marginal effects. ▪ Sign and size of coefficient does NOT indicate direction and size of effect on probability of Y = j. ▪ Exponential of a coefficient indicates proportional change in odds in favor of Y = j versus base choice due one unit change in x variable.	▪ Must compute marginal effects ▪ Sign and size of coefficient does NOT indicate direction and size of effect on probability of Y = j.	▪ Must compute marginal effects ▪ Sign and size of coefficient does NOT indicate direction and size of effect on probability of making choice Y = j.	▪ Must compute marginal effects ▪ Sign and size of coefficient does NOT indicate direction and size of effect on probability of making choice Y = j.
Marginal Effects	▪ Depend on values of all model variables and coefficients ▪ Compute using derivative expression or as discrete change in probability. Must use discrete change if dummy variable ▪ All other variables held fixed, usually at their mean values	▪ Depend on values of all variables and coefficients ▪ Compute using derivative expression or as discrete change in probability. Must use discrete change if dummy variable ▪ All other variables held fixed, usually at their mean values	▪ Difficult to interpret. ▪ Depend on values of all variables and coefficients ▪ Compute using derivative expression or as discrete change in probability. Must use discrete change if dummy variable ▪ All other variables held fixed, usually at their mean values	▪ Difficult to interpret ▪ Depend on values of all variables and coefficients ▪ Compute using derivative expression or as discrete change in probability. Must use discrete change if dummy variable ▪ All other variables held fixed, usually at their mean values
Issues	▪ Test of a variable's significance involves a joint test on all coefficients for that variable across equations ▪ Violation of Independence of Irrelevant Alternatives invalidates procedure. ▪ If IIA violated, can use Nested Logit or Multinomial Probit	▪ Violation of normality assumption invalidates procedure	▪ Model invalid if odds not proportional across categories	▪ Violation of normality assumption invalidates procedure

TABLE 8

A Censored Dependent Variable: TOBIT and OLS Estimates for Predicting Firm Diversification

Variable	Level of Firm Diversification			
	Tobit Results		OLS Results	
	Estimates	Marginal Effects[a]	Full Sample	Diversified Firms Only
Import Penetration	-0.076***	-0.030***	-0.027***	-0.021**
Core Business Profitability	-0.055**	-0.022**	-0.022***	0.036**
Firm Size	0.548***	0.219***	0.231***	0.147***
Firm Performance	0.084***	0.034***	0.015***	-0.047***
Industry Growth	-0.112***	-0.045***	-0.035***	-0.012
Industry Profitability	-0.047**	-0.019**	-0.011***	-0.006
Industry Concentration	-0.116***	-0.046***	-0.034***	-0.033***
Industry R&D Intensity	-0.119***	-0.048***	-0.023***	0.001
Industry Capital Intensity	-0.092***	-0.037***	-0.024***	-0.036***
Industry Export Intensity	-0.056**	-0.022**	-0.009	0.023***
Intercept	-0.038		0.414***	0.791***
TD86[b]	-0.027		-0.023	-0.021
TD87	-0.031		-0.027	-0.026
TD88	-0.106**		-0.061***	-0.044
TD89	-0.214***		-0.106***	-0.068***
TD90	-0.178***		-0.095***	-0.068***
TD91	-0.258***		-0.127***	-0.09***
TD92	-0.219***		-0.116***	-0.087***
TD93	-0.294***		-0.147***	-0.086***
TD94	-0.316***		-0.154***	-0.074***
Log Likelihood	-6742		N/A	N/A
R^2 in percent (pseudo-R^2 for Tobit)	15.1		27.9	13.66
Chi-square or F-statistic for model significance [c]	2376***		182.29***	29.73***
Observations	8961		8961	3587

p < .05, ** *p* < .01, *** *p* < .001.

[a] Computed as the estimated Tobit coefficient times the proportion of diversified firms in the sample (= 2857/8961)
[b] Each "TD" variable is a time dummy for the indicated year.
[c] For Tobit, test of the model against the model that includes only the intercept and the time dummies; For OLS, test of the model against the model that includes only the intercept.

Source: Adapted from Bowen, H. and Wiersema, M. (2003).

FIGURE 1

Selecting Statistical Techniques for Discrete Dependent Variables