Forthcoming in *Strategic Management Journal*

THE USE OF LIMITED DEPENDENT VARIABLE TECHNIQUES IN STRATEGY RESEARCH: ISSUES AND METHODS

By

MARGARETHE F. WIERSEMA
(please send all correspondence to this author)
Graduate School of Management
University of California
Irvine, Ca. 92697-3125
+1 949.824.5839
Email: mfwierse@uci.edu

and

HARRY P. BOWEN
McColl School of Business
Queens University of Charlotte
Charlotte, NC 28274
+1 704 688-2707
Email: bowenh@queens.edu

Last Revised, December 2007

Keywords: empirical methods, limited dependent variable, strategy research

THE USE OF LIMITED DEPENDENT VARIABLE TECHNIQUES IN STRATEGY RESEARCH: ISSUES AND METHODS

ABSTRACT

Strategy researchers are increasingly turning their attention from examining the impact of strategic choices on firm performance to examining the factors that determine strategic choices at the firm level. This shift of research orientation has meant that researchers are increasingly faced with a *limited dependent variable* (LDV) that takes a limited number of usually discrete values, for which LDV methods such as Logit or Probit are required. Despite their growing popularity, there appears to widespread problems in the use of LDV methods. This paper complements recent papers that offer general guidelines by presenting and illustrating the practical steps needed to implement the methods essential for analyzing and interpreting the results from LDV models.

THE USE OF LIMITED DEPENDENT VARIABLE TECHNIQUES IN STRATEGIC MANAGEMENT RESEARCH: ISSUES AND METHODS

The statistical techniques used in strategic management research are becoming more sophisticated and more complex. While ordinary least squares (OLS) regression remains predominant,[1] the array of statistical techniques used has expanded significantly. New methods of modeling and estimation have renewed attention on many strategy research topics for which past empirical studies yielded contradictory or confounding results. However, since many of the newer techniques involve less familiar methods of estimation, analysis, and interpretation, strategy researchers are being required to become increasingly sophisticated in terms of their understanding of statistical methods and applications (Shook, Ketchen, Cycytoa, & Crockett, 2003; Shook, Ketchen, Hult, & Kacmar, 2004). In this regard, systematic assessments of the use in strategy research of some of the newer analytical techniques have found major shortcomings regarding implementation and interpretation of results, and in response have offered general guidelines aimed at improving awareness of key issues and problem areas in an effort to improve current practice (e.g., Bowen & Wiersema, 2004; Hoetker, 2007; Shook et al., 2004)[2].

In addition to studies that seek to improve current practice regarding specific techniques, another growing body of work addresses key methodological issues. These include the use of cross-sectional data (Bergh, 1995; Bowen & Wiersema, 1999), the issue of endogeneity (Hamilton & Nickerson, 2003), and the validity of measures (Chatterjee & Bloucher, 1992; Davis & Duhaime, 1992; Hoskisson, Johnson, & Moesel, 1993; Lubatkin, Merchant, & Srinivasan, 1993; Robins & Wiersema, 2003). These studies all highlight systematic problems

[1] Scandura and Williams (2000) report that 42% of recent management research studies used OLS.
[2] *Research Methodology in Strategy and Management*, D. Bergh and D. J. Ketchen, Jr., Series Co-Editors, Elsevier Press, 2004 as well as articles in this journal have addressed a host of specific research design and methodological issues.

with prior empirical strategy research that may have led to biased results and invalid inferences and therefore uncertainty regarding the meaning of prior research findings. Growing recognition of a number of methodological failings means that, despite a multitude of empirical studies, the findings regarding many key theoretical linkages remain equivocal.

This paper contributes to these streams of methodological inquiry by examining a set of statistical issues likely to occur in the analysis of firm level data, the most common level of analysis in strategy research. In this regard, researchers often model strategic choices, such as acquisitions, market exit, and joint ventures, as a limited dependent variable (LDV); for example, as a binomial (yes/no) decision. Since OLS is an inappropriate estimation method in such cases, researchers have turned to LDV methods such as Logit to estimate their models. However, many researchers may be less familiar with such specialized techniques and may feel uneasy about the appropriate methods for analyzing and interpreting the results from such techniques.

Despite the growing use of LDV methods in strategy research (Shook et al., 2003), only recently has attention be given to assessing whether LDV methods are being appropriately and consistently used in strategy research. Recent inquiries (e.g., Bowen & Wiersema, 2004; Hoetker, 2007) have indeed found widespread problems and in response have offered general guidelines regarding key issues and problem areas. While general guidelines are imminently useful, they may be too broad to be fully understood and hence adopted by researchers who lack practical knowledge of LDV models. Given the growing use of LDV models in strategy research, and evidence of widespread problems regarding the analysis and interpretation of such models, it seems warranted to augment general guidelines with a more pragmatic approach.

Toward this goal, this paper first considers the research design issue of when a LDV model is appropriate. It then presents and illustrates the essential methods for analyzing and interpreting the results from any LDV model via simple examples that use the binary Logit model - the most common LDV model used in the strategy literature. The specific STATA commands used to conduct the analyses that we illustrate are presented in the Appendix to this paper. In taking a pragmatic approach, this paper complements and contributes to the growing stream of articles in the strategy literature that highlight general methodological and statistical issues and, more specifically, recent papers that have raised awareness of key problems and offered general guidelines to foster the correct use of LDV methods.

RESEARCH DESIGN

A study's research design refers to the systematic plan to be undertaken to arrive at an answer to a particular research question. This includes formulating the research question, operationalizing theoretical constructs, identifying and collecting data, selection of statistical method, analysis, interpretation, and reporting of results. An important part of a research design is the choice and operationalization of a model's dependent variable, which in general will depend on the phenomenon of interest and the unit of analysis. Historically, empirical strategy research has predominantly focused on organizational outcomes such as firm performance or diversification that could be operationalized by a continuous variable (e.g., ROA) which permitted the use of ordinary least squares (OLS). However, strategy researchers are increasingly examining phenomenon characterized by a discrete set choices or organizational outcomes that cannot be operationalized by a continuous dependent variable. Examples include international expansion operationalized as either start-up or acquisition (Vermeulen & Barkema, 2002); strategic alliances categorized into various organizational forms (Colombo, 2003); the existence or absence of certain conditions such as managerial turnover (Bloom & Michel, 2002; Shen & Cannella, 2002); the existence of a COO position (Hambrick & Canella, 2004);

CEO/Chair duality (Nelson, 2003); and whether or not a new CEO differs from his predecessor (Zajac & Westphal, 1996). As researchers increasingly examine phenomena that cannot be operationalized by a continuous variable they have come to rely on LDV methods to model the relationship of interest. But while the use of LDV models is growing, there has been scant discussion of whether and when such models are appropriate.

The choice of measure (and therefore model) should be dictated first and foremost by the research question of interest. For example, if one's interest is to understand whether a firm's R&D activity influences its decision to expand abroad then a discrete variable indicating whether or not the firm is international is appropriate. If one's interest is instead to understand how a firm's R&D activity impacts the geographic diversity of its activities then the phenomenon of interest is instead the extent of a firm's international operations, and a continuous measure of a firm's foreign activities is appropriate. Utilizing instead a discrete measure, such as a count of the number of countries in which a firm operates, will fail to capture the extent of a firm's foreign presence; a firm with sales offices in multiple countries would appear highly internationally diversified but in reality may have minimal foreign presence. In this example, the country count measure not only fails to capture the full range of variation of the phenomenon of interest, it also fails to adequately reflect the real meaning of the construct under consideration, that is, it lacks content validity. One should not impose a discrete categorization on a strategic phenomenon if it fails to represent the actual set of managerial choices or organizational outcomes since information on the full range of variation in the dependent measure is then ignored. Generally, a continuous measure is appropriate if the phenomenon is the extent of something; a discrete measure is appropriate if the phenomenon is instead a set of discrete choices or outcomes. Recent studies note that strategy researchers rarely articulate a rationale for their choice of measures, and that measures are often adopted without systematic consideration of their content validity (Acar and Sankaran, 1999; Robins and Wiersema, 2003). The result is a research design that fails to fully explain the research question of interest.

ANALYZING AND INTREPETING LDV MODELS

If the phenomenon of interest is best operationalized by a discrete measure then the use of a LDV model becomes appropriate.[3] Recent reviews (Bowen & Wiersema, 2004; Hoetker, 2007) of the use of LDV models in the strategy literature indicate considerable variation in the accuracy with which the results from LDV models are analyzed and interpreted. To a large extent, this variation arises because many researchers are not aware that supplementary analysis is required to correctly analyze the results from a LDV model. This need for additional effort, and often a lack of understanding of the special characteristics of LDV models, undermines the ability of many researchers to interpret their findings and to therefore make a research contribution.

Hoetker (2007) recently echoed such concerns and in response offered general guidelines regarding analysis and interpretation of the most commonly used LDV models in strategy research. Although general guidelines are imminently useful, for many researchers they may be too broad to provide sufficient understanding of how to implement the recommendations. This paper therefore provides a practical contribution in understanding LDV models by providing straightforward examples that illustrate the essential methods for analyzing and interpreting hypotheses.

[3] Such models include all forms of the Logit (e.g., binary, multinomial, ordered) and Probit models, as well as models such as the TOBIT.

Why LDV Models Are Different

To correctly analyze and interpret any LDV model, it is important to understand two fundamental differences between LDV and OLS type models. First, LDV models are *intrinsically* nonlinear, which means the relationship to be estimated cannot be written as a summation of terms, where each term is a model coefficient times a model variable.[4] The *intrinsic* nonlinearity of LDV models has two major methodological ramifications. First, an explanatory variable's *marginal effect* — the effect of a unit change in an explanatory variable on the dependent variable — does not equal the variable's model coefficient. Second, the value of this *marginal effect* varies over the value of all model variables. These facts imply that *one cannot infer the nature of the true relationship between an explanatory variable and the dependent variable based solely on the estimated coefficient in a LDV model.*

The second fundamental difference is that most LDV models are estimated using the method of maximum likelihood which, unlike the method of least squares, is not based on minimizing error variance. This means there is no measure of model "fit" directly comparable to the R-square in OLS and, as a result, model assessment is largely restricted to testing the joint significance of all model variables as is done in OLS using an F-test of overall model significance.

These differences imply that it is potentially misleading, and often not correct, to analyze and interpret LDV models using the methods commonly used for OLS type models. In particular, information presented in the output of most statistical packages regarding the sign, magnitude and statistical significance of a variable's estimated coefficient is rarely sufficient to

[4] For an intrinsically linear model, a variable can be a function of itself (e.g., the square or logarithm of its value) or a function that includes other model variables (e.g., an interaction variable). Key is that relationship to be estimated is linear in model coefficients (e.g., Greene, 1997: 450).

detect the nature of the true relationship between an explanatory variable and the dependent variable, and hence also to test a hypothesis about the nature of their relationship. Instead, for LDV models, the focus of analysis is on the value and statistical significance of an explanatory variable's marginal effect, which requires analysis beyond simply estimating one's model. The following sections present examples that illustrate this additional analysis and provide recommendations for writing up results; the specific STATA commands used to conduct this analysis are presented in the Appendix to this paper.

Model Estimation and Assessment

For our illustrations, we model the probability that a CEO's succession is either a dismissal or a routine succession using a binary Logit specification. While the binary Logit model is in some respects a special case, the methods illustrated are general and applicable to any LDV model. Our dependent variable, "dismissal," is coded as a binary variable; it equals 1 if the CEO was dismissed and zero for routine succession. Our dataset comprises 199 large public firms that have undergone a CEO succession over a specific period of time; for 76 (38%) of the observations the CEO was dismissed. For simplicity, there are two "generic" explanatory variables *X* and *Z*. Given this, our model for that the probability that a CEO succession event will be a dismissal can be written:

(1) $$\Pr(dismissal = 1 \mid X, Z; \beta_0, \beta_X, \beta_Z) = \Pi(\mathbf{V\beta}') = \frac{e^{\mathbf{V\beta}'}}{1 + e^{\mathbf{V\beta}'}} = \frac{e^{\beta_0 + \beta_x X + \beta_z Z}}{1 + e^{\beta_0 + \beta_X X + \beta_Z Z}}$$

In the above, $\Pi(\mathbf{V\beta}')$ is the probability that a CEO succession will be a dismissal, $\boldsymbol{\beta}$ is the (row) vector of model coefficients β_0, β_X and β_Z, and **V** is the (row) vector of model variables (including a "1" for the intercept).

Our model is estimated using STATA's *logit* command. As shown, in Table 1, variables *X* and *Z* are each significant (p < .004), where significance is based on the value of a normal z-statistic rather than a t-statistic since the statistical theory underlying maximum likelihood estimation refers to large sample (asymptotic) properties of the estimates (Greene, 1997). Model significance is indicated by the significance (p < 0.001) of the Likelihood Ratio Chi-square statistic which tests our model against an intercept-only model; this test is analogous to the overall F-test of model significance in OLS estimation. The McFadden pseudo R-square reported in Table 1 is one of the many "goodness-of-fit" measures proposed for LDV models (e.g., see Hoetker, 2007).

Table 1. Results from Stata's *logit* command estimating probability of CEO dismissal

Logistic regression	Number of obs	=	199
	LR chi2(2)	=	49.03
	Prob > chi2	=	0.0000
Log likelihood = -107.81781	Pseudo R2	=	0.1853

dismissal	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
X	-4.196446	.7772556	-5.40	0.000	-5.719839	-2.673053
Z	1.106789	.3815787	2.90	0.004	.3589082	1.854669
_cons	-3.145352	.8648826	-3.64	0.000	-4.840491	-1.450213

Hypothesis Testing – Direct Effects

Testing a hypothesis about the nature of the direct relationship between an explanatory variable and the dependent variable requires the researcher to estimate their relationship and assess its statistical significance. In strategy research, such a hypothesis most often concerns the sign (e.g., positive or negative) of the relationship. For OLS type models estimated using the method of least squares, such a directional hypothesis is tested by observing the sign and statistical significance of the explanatory variable's estimated model coefficient. However, in

LDV models, the direct relationship between an explanatory variable and the dependent variable is not given by the explanatory variable's model coefficient but instead by the variable's marginal effect, which will vary with the value of all model variables. Hence, a directional hypothesis in LDV models is tested by examining the sign (positive or negative) and statistical significance of the values of an explanatory variable's marginal effect over all values of the model variables.[5]

Consider then testing the hypothesis that the relationship between variable *X* and the probability that a CEO succession will be a dismissal is negative. This will require analysis of *X*'s marginal effect. The first step is to determine if *X* is significant in the model, since otherwise the theoretical value of its marginal effect is zero. Since *X* is significant (see Table 1), the next step is to determine the equation for its marginal effect. Since *X* is a continuous variable, this is found by differentiating equation (1) with respect to *X*.[6] The result is:

$$\text{(2)} \qquad \text{Marginal effect of } X = \frac{\partial \Pr(dismissal = 1 \mid \mathbf{V}, \boldsymbol{\beta})}{\partial X} = \frac{\partial \Pi(\mathbf{V}\boldsymbol{\beta}')}{\partial X} = \pi(\mathbf{V}\boldsymbol{\beta}')\,\beta_X$$

This indicates that the marginal effect of *X* is proportional to its model coefficient β_X. Since $\pi(\mathbf{V}\boldsymbol{\beta}')$ is always positive,[7] *X*'s marginal effect has the same sign as its model coefficient (β_X). Hence, in this model, the nature of the true relationship between *X* and the probability that

---

[5] Analysis of the individual values of a marginal effect is the most general approach (for additional approaches see Long, 1997). Less general are "representative value" methods, of which two are common. The first computes the value of the marginal effect at the means of all model variables and assesses its significance; the second method computes the average of the individual marginal effect values and assesses its significance. In STATA, the *mfx* command performs the first method while the *margeff* command performs the second method. However, we caution that these commands assume a conventional model; one where all model variables are measured in natural units (e.g., not in logarithms) and each variable appears only once in the model and in level form. This excludes models that, for example, include the square of a variable or an interaction variable.

[6] For a discrete explanatory variable its marginal effect is the change in the dependent variable when the explanatory variable is incremented by one unit. For example, for a dummy (0/1) explanatory variable "D," its marginal effect is equation (1) when D = 1 minus equation (1) when D = 0 (e.g., see Long, 1997).

[7] The term $\pi(\mathbf{V}\boldsymbol{\beta}') = d\Pi(\mathbf{V}\boldsymbol{\beta}')/d\,\mathbf{V}\boldsymbol{\beta}' = e^{\beta_0+\beta_x X+\beta_z Z}\Big/\left(1+e^{\beta_0+\beta_x X+\beta_z Z}\right)^2 = \Pi(\mathbf{V}\boldsymbol{\beta}')\left(1-\Pi(\mathbf{V}\boldsymbol{\beta}')\right)$ is the derivative of the Logit cumulative distribution function. By definition, $\pi(\mathbf{V}\boldsymbol{\beta}')$ is the Logit probability density function.

a CEO succession will be a dismissal can be directly inferred from the sign of the estimated model coefficient (in this case negative).

We now need to determine if this negative relationship is statistically significant. Due the presence of the term $\pi(\mathbf{V}\beta')$ in equation (2), the value and significance of *X*'s marginal effect is not given by the value and significance of its estimated model coefficient. Instead, there are many values of the marginal effect, and each has its own standard error. We must therefore compute, at each observation, the value of the marginal effect, its standard error, and implied z-statistic value (ratio of the marginal effect value to its standard error) to test the significance of each value. To analyze these values and their significance, a graphical analysis is useful. The Appendix presents the STATA commands that performed these calculations and generated the graphical analysis for our Logit model of CEO dismissal.

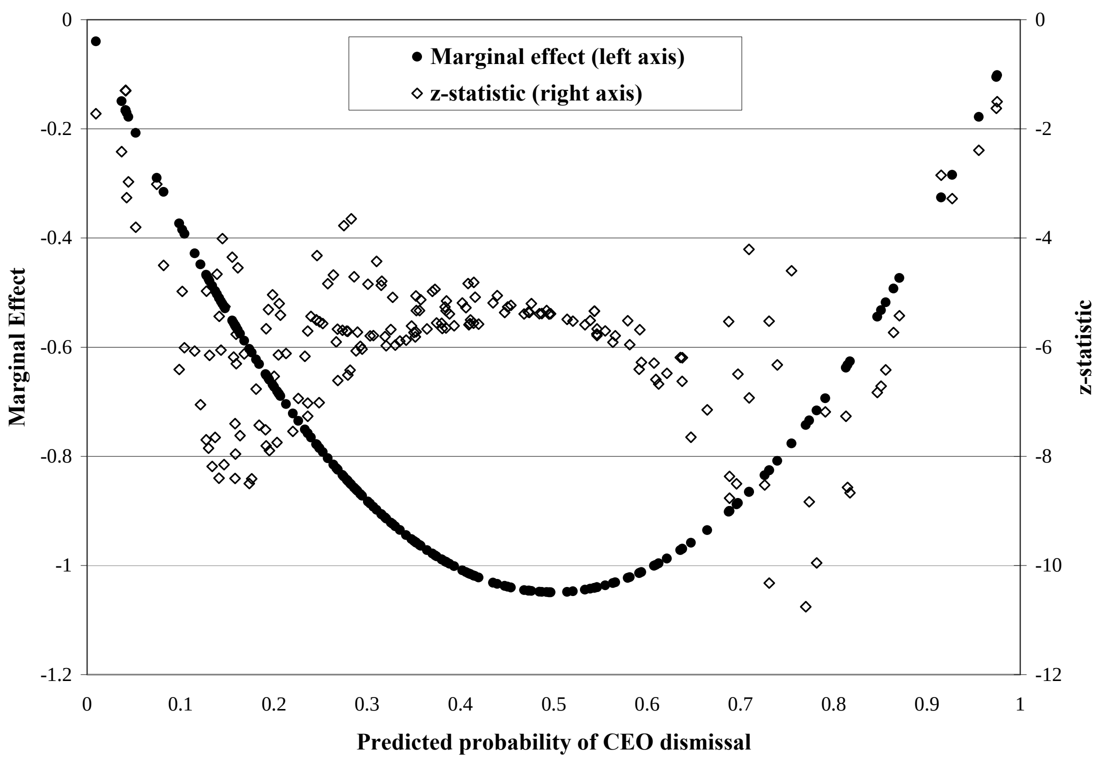
Since our sample comprises 199 observations there are 199 values of the marginal effect given by equation (2). Each value and its associated z-statistic value are plotted in Figure 1; the solid symbols indicate values of the marginal effect (recorded on the left axis) while the diamond shaped symbols indicate z-statistic values (recorded on the right axis). As expected, all values of the marginal effect are negative;[8] the values range from -1.0491 to -0.0397 and the mean or "average marginal effect" value is -0.7658. The z-statistic values range from -10.75508 to -1.298058, but as indicated in Figure 1, the z-statistic value associated with any given marginal effect value exceeds 1.96 in absolute value except at very high and low probabilities of CEO dismissal.[9] We conclude from this analysis that, as hypothesized, the relationship between *X* and

[8] The values of the marginal effect follow a U-shape since the term $\pi(\mathbf{V}\beta')$ in equation (2) is simply the "bell-shaped" Logit probability density function (cf. footnote 7). In Figure 1, this function is inverted since the estimated model coefficient for *X* is negative.

[9] That values of the marginal effect at extreme probability values are not significant is to be expected. It simply reflects that the slope of the Logit cumulative distribution function approaches zero at the extreme ends of the distribution.

the probability that a CEO succession will be a dismissal is negative and statistically significant ($p < 0.05$).

Figure 1. Marginal effect analysis of X1 on the Probability of CEO dismissal



To summarize, testing a hypothesis about the nature of the relationship between an explanatory variable and the dependent variable in a LDV model requires a supplementary analysis that examines the value and significance of the explanatory variable's marginal effect. This analysis is only undertaken if the explanatory variable's estimated model coefficient is significant, since otherwise its marginal effect is theoretically equal to zero. One then analyzes the sign (positive or negative) and statistical significance of the value of the marginal effect at each observation to determine if the hypothesized relationship between the explanatory and dependent variable is accepted or rejected. For presenting results, the mean and range of the marginal effect and z-statistic values should be reported and, as done in Figure 1, a plot of these values against the predicted values of the dependent variable is recommended.

Hypotheses Testing – Moderating Effects

In the strategy literature, it is common for researchers to postulate that one or more variables moderate the relationship between an explanatory variable and the dependent variable. The methods for analysis and testing of moderating (interaction) hypotheses in the OLS framework are both well documented (e.g., Jaccard, Turrisi, & Wan, 1990) and familiar to most researchers. It is therefore not surprising that researchers have often adopted OLS procedures to guide them in interpreting an interaction hypothesis (Hoetker, 2007) in a LDV model, despite that these procedures are not correct in the context of a nonlinear LDV model. In particular, in a LDV model, the influence of a moderator variable on the relationship between an explanatory variable and the dependent variable is not indicated by the sign and significance of the estimated coefficient on the interaction variable in the model. Instead, a moderating effect is itself a marginal effect, and hence all the issues regarding a marginal effect discussed in the previous section apply. In particular, the equation for the moderating effect will be nonlinear, its value will depend on the values taken by all model variables and, as noted, it will not equal the coefficient on the model's interaction variable. Hence, a moderator hypothesis in a LDV model is tested by examining the sign (positive or negative) and statistical significance of the values of the moderator variable's marginal effect on the relationship between the explanatory variable and the dependent variable over all values of the model variables.

To illustrate, we examine the hypothesis that the relationship between X and the probability that a CEO succession will be a dismissal is positively moderated by variable Z.

Since the relationship between *X* and CEO dismissal is expected to be negative, this moderating hypothesis means that the relationship between *X* and CEO dismissal is expected to become less negative at higher values of *Z*. To test this hypothesis we add the interaction variable (*X***Z*) to our Logit model. Given this, our model can now be written:

$$(3) \qquad \Pr(dismissal = 1 \mid \mathbf{V}, \boldsymbol{\beta}) = \Pi\left(\mathbf{V}\boldsymbol{\beta}'\right) = \frac{e^{\beta_0+\beta_X X+\beta_Z Z+\beta_{XZ}(X*Z)}}{1+e^{\beta_0+\beta_X X+\beta_Z Z+\beta_{XZ}(X*Z)}}$$

To test our moderator hypothesis, we must first determine if the interaction variable is statistically significant. Since this is indeed the case (see Table 3), the next step is to determine the equation for the marginal effect of the moderator variable on the relationship between the explanatory variable and dependent variable; this marginal effect is called the *true interaction effect* (Ai and Norton, 2003).

Table 3. Results from Stata's *logit* command estimating probability of CEO dismissal

```
Logistic regression                                 Number of obs   =        199
                                                    LR chi2(3)      =      55.17
                                                    Prob > chi2     =     0.0000
Log likelihood = -104.74792                         Pseudo R2       =     0.2085

------------------------------------------------------------------------------
    dismissal |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
          X   |  -16.57889   5.298664   -3.13   0.002    -26.96408   -6.193697
          Z   |   1.319382   0.410072    3.22   0.001     0.515654    2.123109
         XZ   |   5.502511   2.264677    2.43   0.015     1.063826    9.941195
      _cons   |  -3.736800   0.961833   -3.89   0.000     -5.62196   -1.851641
------------------------------------------------------------------------------
```

Since *X* and *Z* are continuous variables, the equation for the true interaction effect is given by the cross-partial derivative of equation (3), first with respect to *X* and then with respect to *Z*. The derivative of equation (3) with respect to *X* yields the following equation for the marginal effect of *X* in this expanded model:

$$(4) \qquad \text{Marginal effect of } X = \frac{\partial \Pi\left(\mathbf{V}\boldsymbol{\beta}'\right)}{\partial X} = \pi\left(\mathbf{V}\boldsymbol{\beta}'\right)\left(\beta_X + \beta_{XZ} Z\right)$$

Differentiating equation (4) with respect to Z then yields the equation for the *true interaction effect* of moderator variable Z:[10]

(5) True interaction effect $= \frac{\partial \Pi(\mathbf{V\beta}')}{\partial X \partial Z} = \frac{\partial (\text{Marginal effect of } X)}{\partial Z}$

$$= \Pi(\mathbf{V\beta}')\left(1-\Pi(\mathbf{V\beta}')\right)\left[\beta_{XZ} + \left(1-2\Pi(\mathbf{V\beta}')\right)\left(\beta_X + \beta_{XZ} Z\right)\left(\beta_Z + \beta_{XZ} X\right)\right]$$

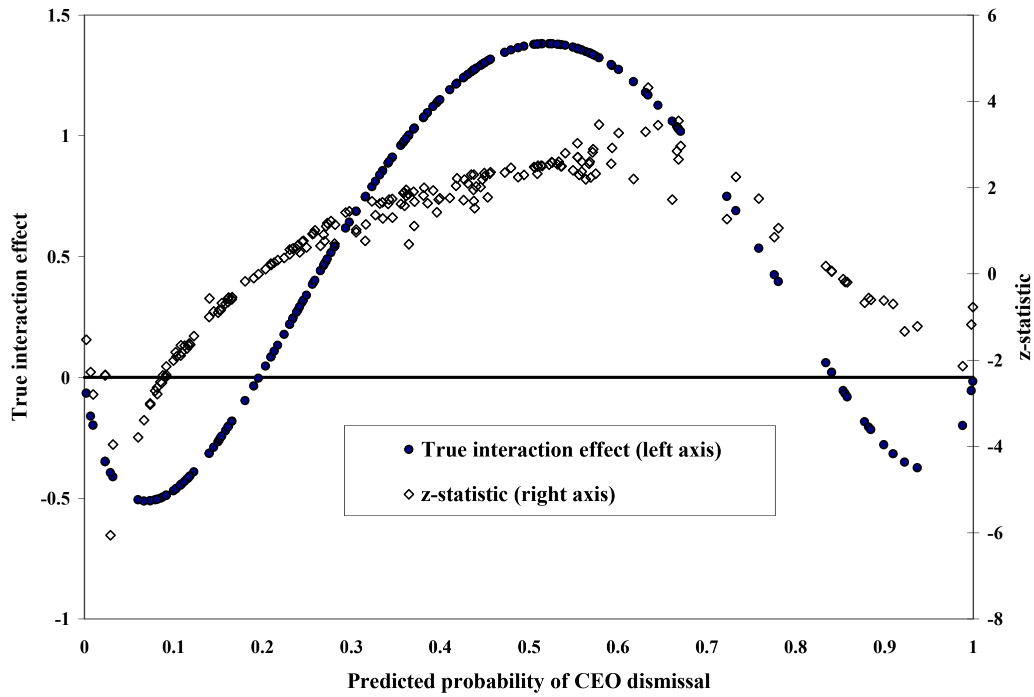
In the above, the term $\Pi(\mathbf{V\beta}')$ is now given by equation (3). Clearly, the value and significance of the true interaction effect is not given by the value and significance of the interaction variable coefficient (β_{XZ}). Instead, there are many values of the true interaction effect, and each has its own standard error. As in the previous section, to assess the nature and significance of the marginal effect given by equation (5), we must compute its value, and the implied z-statistic value, at each observation and then examine these values graphically as presented in Figure 2.

In Figure 2, the solid symbols indicate values of the true interaction effect (recorded on the left axis) while the diamond shaped symbols indicate z-statistic values (recorded on the right axis). As seen, the value, sign, and significance of the true interaction effect differs over its range of variation. The values of the true interaction effect range from -0.5116 to 1.3819, with a mean value of 0.5309. The z-statistic values range from -6.0614 to 4.3106, so some values of the true interaction effect are not significant. As seen in Figure 2, the true interaction effect is positive and significant only when the probability that a CEO succession will be a dismissal is in the range from about 45% to 65%, and its value is negative and significant when the probability that a CEO succession will be a dismissal is less than about 10%. These findings indicate the complexity of analyzing the results from nonlinear LDV models, and that clear cut results are not

[10] cf. footnote 6. If Z were a dummy 0/1 variable, the true interaction effect would be the difference between equation (4) when Z =1 and equation (4) when Z = 0. The value of this difference equation would then be computed at each observation.

always to be expected. The results also underscore that examining only the sign and statistical significance of the average of the interaction effect values would not provide an accurate assessment of the moderating effect. Based on the graphical analysis, we conclude that there is limited support for the hypothesis that variable *Z* has a positive and significant moderating effect.

Figure. 2 True interaction effect of *Z* on relationship between *X* and probability of CEO dismissal



Although the nature and significance of the true interaction effect differs over its range of variation, the significance of variable Z and of the interaction variable in the model means one can still examine the impact of different values of Z for the value and significance of the marginal effect of *X*. For this analysis, the value of all model variables except Z (i.e., *X*) must be held fixed.

To illustrate, we compute the value of equation (4) at a low, mean, and high value of *Z* while always keeping the value of *X* at its sample mean value.[11] By convention, the low and high value of variable *Z* is one standard deviation below and above its mean. Each calculated value of equation (4) has a standard error and implied z-statistic value. The STATA commands that performed this analysis are given in the Appendix. Table 5 reports the results.

Table 5. Moderating effect of Z for marginal effect of X on the probability of CEO dismissal

Value of moderator Z	Marginal effect of X	z-statistic
Low	-1.37074	-3.78
Mean	-1.10287	-5.51
High	-0.48969	-1.76

As shown in Table 5, the relationship between *X* and the probability that CEO succession will be a dismissal is less negative at higher values of *Z*, indicative of a generally positive moderating effect of *Z*. At the high value of *Z*, this positive effect is sufficient to render the (negative) marginal effect of *X* statistically insignificant ($p > .05$). These results indicate that, holding fixed the value of *X*, higher values of *Z* reduce the impact that *X* has on the probability of CEO dismissal.[12]

Finally, it is common in an OLS analysis to graphically illustrate the influence of a moderating variable by a pair of straight lines, where the slope of each line is the marginal effect of the explanatory variable at a low and high value of the moderator variable. A visual comparison of the slopes of the two lines then indicates how the explanatory variable's marginal

[11] A more general analysis would compute equation (4) at each sample value of *Z* while setting the value of all other model variables (i.e., *X*) equal to their sample mean value. The values of equation (4) and associated z-statistic values can then be plotted against the values of variable *Z* to observe the entire range of variation in the value and significance of equation (4).

[12] We emphasize that this analysis does not test for the significance of the moderating effect; that analysis instead requires analysis of the true interaction effect as presented above.

effect differs at the two values of the moderator. Although this type of graphical analysis can be done in the context of a LDV model, we believe it is seriously misleading, not least because it assumes that the value of the explanatory variable's marginal effect is constant over values of the explanatory variable, which directly contradicts that a variable's marginal effect in a LDV model is not a constant value. At best, this type of graphical analysis is only suggestive of the moderating influence of a moderating variable, and cannot be used for statistical inference.[13]

To summarize, the method for testing a moderator hypothesis in a LDV model differs substantially from that used in OLS. If the estimated model coefficient on the interaction variable is significant, the values of the *true interaction effect* are then computed and analyzed to assess the sign (positive or negative) and statistical significance of a moderating effect. For presenting results, the mean and range of the true interaction effect and z-statistic values computed at each observation should be reported, and a plot of these values against the predicted values of the dependent variable, as done in Figure 2, is recommended. Finally, analysis of the value and statistical significance of the explanatory variable's marginal effect at different values of the moderator variable, as shown in Table 5, can be undertaken to gauge the extent of influence of the moderator variable for the value and significance of the explanatory variable's marginal effect.

CONCLUSION

This paper has provided and illustrated the most general methods for analyzing and interpreting the results from LDV models. In taking a pragmatic approach, it is hoped that our detailed presentation of the essential methods for making statistical inferences in LDV models will enable strategy researchers to feel more confident when using LDV models, and can help

[13] If such a graphical analysis is desired, we recommend graphing, for each at specific value of the moderator (e.g., low, mean, and high), the values of equation (3) against the values of the explanatory variable, with all other model variables always set equal to their sample mean value.

prevent them from perpetuating past mistakes when analyzing and reporting of the results from such models. In this respect, this paper has contributed to the ongoing stream of methodological inquiry in strategy research by providing greater clarity on an important set of statistical issues that can arise in the analysis of firm level data.

REFERENCES

Acar W, Sanakaran K. 1999. The myth of the unique decomposability: specializing the Herfindahl and entropy measures? *Strategic Management Journal* 20(10): 969-975.

Agresti A. 2002. *Categorical Data Analysis*. John Wiley & Sons: New Jersey

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

Bergh DD. 1995. Problems with repeated measures analysis: demonstration with a study of the diversification and performance relationship. *Academy of Management Journal* **38**(6): 1692–1708.

Bergh D, Ketchen Jr., DJ. (editors) 2004. *Research methodology in strategy and management*. Elsevier Press.

Bloom M, Michel JG. 2002. The relationships among organizational context, pay dispersion, and among managerial turnover. *Academy of Management Journal* 45(1): 33-42.

Bowen HP, Wiersema MF. 1999. Matching method to paradigm in strategy research: Limitations of cross-sectional analysis and some. *Strategic Management Journal* 20(7): 625-636.

Bowen HP, Wiersema MF. 2004. Modeling limited dependent variables: guidelines for researchers of strategic management" in *Research Methodology in Strategy and Management*, D. Ketchen, Jr. and D. Bergh Series Co-Editors, Elsevier Press.

Chatterjee S, Bloucher JD. 1992. Measurement of firm diversification: Is it robust? *Academy of Management Journal* **35**(4): 874-888.

Chatterjee S, Harrison JS, Bergh DD. 2003. Failed takeover attempts, corporate governance and refocusing. *Strategic Management Journal* 24(1): 87-97.

Colombo MG. 2003. Alliance form: A test of the contractual and competence perspectives. *Strategic Management Journal* 24(12):1209-1229.

Davis R, Duhaime I. 1992. Diversification, vertical integration, and industry analysis: New perspectives and measurement. *Strategic Management Journal* 12: 511-524

Greene WH. 1997. *Econometric analysis* (3rd Edition). Prentice Hall: New Jersey.

Hambrick A, Cannella AAJ. 2004. CEOs who have COOs: Contingency analysis of an unexplored structural form. *Strategic Management Journal* 25(10): 959-979.

Hamilton B, Nickerson J. (2003) 'Correcting for endogeneity bias in strategic management research', *Strategic Organization*, I (1): 51-78.

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

Hoskisson RE, Hitt MA, Johnson RA, Moesel DD. 1993. Construct validity of an objective (entropy) categorical measure of diversification strategy. *Strategic Management Journal* 14: 215-235.

Jaccard J, Wan CK, Turrisi R. 1990. The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research* 25(4): 467-479.

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

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

Nelson T. 2003. The persistence of founder influence: Management, ownership, and performance effects at initial public offering. *Strategic Management Journal* 24(8): 707-724.

Powers D, Xie Yu. 1999. *Statistical Methods for Categorical Data Analysis*, Academic Press, Inc.

Robins J, Wiersema MF. 2003. The measurement of corporate portfolio strategy: Analysis of the content validity of related diversification indexes. *Strategic Management Journal* 24(1): 39-59.

Scandura TA, Williams EA. 2000. Research methodology in management: Current practices, trends, and implications of future research. *Academy of Management Journal* 43(6): 1248-1264.

Shen W, Cannella AAJ. 2002. Power dynamics within top management and their impacts on CEO dismissal followed by inside succession. *Academy of Management Journal* 45(6): 1195-1206.

Shook CL, Ketchen DJJ, Cycyota CS, Crockett D. 2003. Data analytic trends and training in Strategic Management. *Strategic Management Journal* 24: 1231-1237.

Shook CL, Ketchen DJJ, Hult GTM, Kacmar KM. 2004. An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal* 25(4): 397-404.

Vermeulen F, Barkema H. 2002. Pace, rhythm, and scope: Process dependence in building a profitable multinational corporation. *Strategic Management Journal* 23(7): 637-653.

Zajac EJ, Westphal JD. 1996. Director reputation, CEO/board power, and the dynamics of board interlocks. *Administrative Science Quarterly* 41: 507-529.

Appendix

Hypothesis Testing – Direct Effects

The following STATA commands compute at each observation the value of the marginal effect given by equation (2), its standard error, and implied z-statistic value. The mean and range of the marginal effect and z-statistic values are computed, and the marginal effect and z-statistic values are then plotted as in Figure 1.

```
* Estimate Logit model
    logit dismissal X Z
* Predict probability of CEO dismissal, store values in variable pprob
   predict   pprob
* Define expression for marginal effect to use in predictnl command
* Store marginal effect values in new variable "me"
* Store standard errors values in new variable "me_se"
    local vb  _b[_cons] + _b[X]*X +_b[Z]*Z
    local phat (exp(`vb')/(1+exp(`vb'))^2)
    predictnl me = `phat'*(1-`phat')*_b[X], se(me_se)
* Compute z-statistic values and store in new variable "z_stat";
    gen z_stat = me/me_se
* print mean, minimum and maximum value of marginal effect and z-statistic values
    tabstat me z_stat, stats(mean min max)
* Graph marginal effect and z-statistic values (Figure 1)
    graph twoway (scatter me pprob)|| ///
                (scatter z_stat pprob, yaxis(2)yline(-1.96 1.96, axis(2)))
```

Hypothesis Testing – Moderating Effects

The following STATA commands compute at each observation the value of the true interaction effect given by equation (5), its standard error, and implied z-statistic value. The mean and range of the interaction effect and z-statistic values are computed, and the interaction effect and z-statistic values are then plotted as in Figure 2.

```
* Create interaction variable
     gen XZ = X*Z
* Estimate Logit model
     logit dismissal X Z XZ
* Predict probability of CEO dismissal, store values in variable pprob
     predict pprob
* Define expression for interaction effect to use in predictnl command.
* Store interaction effect values in new variable "ie"
* Store standard errors values in new variable "ie_se"
    local vb  _b[X]*X+_b[Z]*Z+_b[XZ]*XZ+_b[_cons]
    local phat (exp(`vb')/(1+exp(`vb')))
    local term1 (`phat'*(1-`phat'))
    local term2 (1-2*`phat')
    local coef1 (_b[X]+_b[XZ]*Z)
```

```
   local coef2 (_b[Z]+_b[XZ]*X)
   predictnl ie = `term1'*(_b[XZ]+`term2'*`coef1'*`coef2'), se(ie_se)
* Compute z-statistic values and store in new variable “z_stat”;
   gen z_stat = ie/ie_se
* Print mean, minimum and maximum of interaction effect and z-statistic values
     tabstat ie z_stat, stats(mean min max)
* Graph marginal effect and z-statistic values (Figure 2)
     graph twoway (scatter ie pprob)|| ///
              (scatter z_stat pprob, yaxis(2)yline(-1.96 1.96, axis(2)))
```

Marginal Effect of *X* in Interaction Model

The following STATA commands compute and report significance of the value of the marginal effect of *X* given by equation (4) and at a low, mean and high value of variable *Z*.

```
* Compute and save sample means of variables X and Z
     egen meanX = mean(X)
     egen meanZ = mean(Z)
* Compute and save standard deviation of moderator variable Z
     egen sdZ = sd(Z)
* Compute and save high and low values of moderator variable Z
     gen Zhigh = meanZ + sdZ
     gen Zlow  = meanZ - sdZ
* Estimate Logit model in equation (3)
     logit dismissal X Z XZ
* Compute and test significance of marginal effect equation (4)
*  when Z is at its low value; X is at its mean value
     local vb (_b[_cons] + _b[X]*meanX + _b[Z]*Zlow + _b[XZ]*meanX*Zlow)
     nlcom mean_low: ((exp(`vb'))/(1+exp(`vb'))^2)*(_b[X]+_b[XZ]*Zlow)
* Compute and test significance of marginal effect equation (4)
*  when Z is at its mean value; X is at its mean value
     local vb (_b[_cons]+_b[X]*meanX+_b[Z]*meanZ+_b[XZ]*meanX*meanZ)
     nlcom mean_mean: ((exp(`vb'))/(1+exp(`vb'))^2)*(_b[X]+_b[XZ]*meanZ)
* Compute and test significance of marginal effect equation (4)
*  when Z is at its high value; X is at its mean value
     local vb (_b[_cons]+_b[X]*meanX+_b[Z]*Zhigh+_b[XZ]*meanX*Zhigh)
     nlcom mean_high: ((exp(`vb'))/(1+exp(`vb'))^2)*(_b[X]+_b[XZ]*Zhigh)
```