Article

Organizational Research Methods
1-26

DOI: 10.1177/1094428114560024
orm.sagepub.com
SAGE

Interpreting Results From the Multinomial Logit Model: Demonstrated by Foreign Market Entry

Jesper N. Wulff[1]

Abstract

This article provides guidelines and illustrates practical steps necessary for an analysis of results from the multinomial logit model (MLM). The MLM is a popular model in the strategy literature because it allows researchers to examine strategic choices with multiple outcomes. However, there seem to be systematic issues with regard to how researchers interpret their results when using the MLM. In this study, I present a set of guidelines critical to analyzing and interpreting results from the MLM. The procedure involves intuitive graphical representations of predicted probabilities and marginal effects suitable for both interpretation and communication of results. The practical steps are illustrated through an application of the MLM to the choice of foreign market entry mode.

Keywords

multinomial logit, empirical methods, marginal effects, predicted probabilities, graphical interpretation, entry mode

Researchers in business strategy often wish to draw inferences about factors underlying the strategic choices that firms make. Because such choices are rarely binary, they regularly apply discrete choice models allowing for multiple outcomes such as multinomial logit/probit. However, when interpreting such models, scholars are faced with a set of different challenges than in the binary dependent variable case. One very critical difference is that the sign of the estimated model coefficients does not determine the direction of the relationship between an independent variable and the probability of choosing a specific alternative (Bowen & Wiersema, 2004). Instead, to be able to draw valid conclusions about relationships, scholars must rely on other interpretational devices such as predicted probabilities and marginal effects.

[1]Aarhus University, Aarhus, Denmark

Corresponding Author:
Jesper N. Wulff, Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.
Email: jwulff@econ.au.dk

Despite the importance of this difference in interpretation, assessments of the use of limited dependent variable models in strategic research have found that many studies have severe limitations in the implementation and interpretation of results (Wiersema & Bowen, 2009). Hoetker (2007) identifies coefficient interpretation as one of four critical issues, and Bowen and Wiersema (2004) find that researchers routinely assume that the sign and significance of MLM model coefficients can be used to evaluate their hypotheses. This tendency has motivated authors such as Wiersema and Bowen, Hoetker, and Zelner (2009) to present important guidelines for the implementation and interpretation of limited dependent variable models.

This article contributes to the methodological literature that draws attention to systematic problems with the empirical strategies currently used in strategy research, specifically studies that have illustrated guidelines essential to the use of limited dependent variables models. In this study, I focus on the multinomial logit model (MLM). The MLM is the most commonly applied model when researchers examine multiple unordered strategic choices. I describe how and why this model complicates the interpretation of results. Because many researchers may not be familiar with the special care that must be taken when interpreting results from an MLM in comparison to regular binary models, I believe that it is valuable to further introduce them to this important topic. As a part of this introduction, I present a set of guidelines to follow while mentioning some of the consequences of choosing not to do so. Therefore, the contribution of this study lies in highlighting central challenges and illustrating the procedure essential for using and interpreting results from the MLM illustrated by an example from the entry mode literature.

To achieve this goal, I review the use of discrete choice models with multiple outcome alternatives within the field of foreign market entry. Because managers choose from a variety of different options when they enter a foreign market, entry mode researchers commonly use the MLM (Canabal & White, 2008). Next, I consider the complexity of analyzing relationships in MLMs and present and illustrate the necessary techniques for analyzing results from an MLM. In particular, I recommend specific graphical representations that are powerful not only for interpretation but also for presentation of results. The Stata code for generating the computed quantities and graphical representations are available in Appendix sections A.1 and A.2.

Multiple Outcome Models and Entry Mode

In the literature, the conceptualization of entry mode choice (EMC) has been defined in a variety of ways. Canabal and White's (2008) review shows that researchers most often view mode choice as a choice between wholly owned subsidiary (WOS) and joint venture (JV) (e.g., Chang, Chung, & Moon, 2013). Other popular ways to conceptualize EMC include acquisition versus greenfield (e.g., Brouthers & Dikova, 2010), equity versus non-equity (e.g., Schwens, Eiche, & Kabst, 2011), licensing versus exporting (e.g., Martin & Salomon, 2003), or non-equity partnership versus equity JV (e.g., Sampson, 2004). Methodologically, such studies all specify entry mode choice as a binary variable. Most commonly, this leads to an analysis utilizing a limited dependent variable model, often logistic regression (Canabal & White, 2008).

When researchers reduce the choice to a binary variable and perform a logistic regression, it often involves either pooling choices together or deleting observations. Martin (2013) argues that even researchers who acknowledge a distinction between multiple modes of entry (e.g., acquisition, greenfield, and JV) reduce the dependent variable to a binary choice. However, as the author suggests, researchers can avoid having theory and interpretation of their results depend on the exclusion of some modes of entry by conducting a multinomial logistic regression analysis. In the entry mode literature, the method is used in studies focusing on a dependent variable involving more than two types of entry modes (Agarwal & Ramaswami, 1992; Anand & Delios, 1997; Kim & Hwang, 1992).

Considering the wide range of entry modes a firm can choose from, it is surprising that statistical methods that allow a dependent variable to have more than two outcomes are not dominating the entry mode literature. One reason for this might be that modeling EMC as multiple discrete alternatives confronts the researcher with a much more complex set of challenges than a binary approach. First, the model includes several parameters depending on the number of choice possibilities, which may make both understanding and communicating results seem overwhelming (Long & Freese, 2006). Second, the interpretation of coefficients becomes more complex in multiple outcome models than in binary ones (Long, 1997). For instance, the sign and size of the coefficients indicate neither the direction nor the size of the marginal effect on the probability that an alternative is chosen (Cameron & Trivedi, 2005; Greene, 2003).

Still, limited dependent variable models with multiple outcomes are the second most popular choice for entry mode researchers (Canabal & White, 2008). Generally, studies using such models can be classified into two different categories. One line of work relies mainly on coefficient interpretation when assessing the direction and significance of a predictor. However, these studies often fail to consider that the coefficient does not indicate the direction of the relationship of interest. Examples include Gatignon and Anderson (1988), who find that their MLM has some value in explaining higher control modes while being less successful explaining lower control choices. Distinguishing between transaction costs and costs arising from cultural factors, Kogut and Singh (1988) also apply an MLM and interpret their results as both supporting and contradicting transaction cost theory. In later studies, authors have expanded the traditional transaction cost-framework when using an MLM for analysis. For instance, Klein, Frazier, and Roth (1990) compare market, intermediate, and hierarchical entry options; Kim and Hwang (1992) find support for their expanded transaction cost-model; and Meyer (2001) establishes a link between institutional theory and transaction cost theory. In more recent studies, authors have continued the praxis on interpreting model coefficients but on new theoretical frameworks. For example, Brouthers, Brouthers, and Werner (2008a) examine predictions from real options theory when analyzing the choice between a variety of entry modes focusing mainly on JVs. Based on institutional theory, Uhlenbruck (2004) studies model coefficients to determine the relationship between host country corruption and different types of non-equity and equity modes.

Basing the analysis on coefficient interpretation has some unfortunate downsides. First, it greatly limits interpretation and complicates communication of results. For instance, researchers limit themselves to making statements about how a predictor is related to the probability of one choice outcome relative to the base category. This neglects that outcomes are inexorably linked and makes it very difficult to see the implications for each category. Second, it increases the risk of misinterpretation. The reason is that the sign on a single coefficient only tells us about the contrast among two categories. If researchers use a model coefficient as support for a hypothesis about the effect that a single predictor has on the probability of a single outcome, it risks leading to invalid inference and creates uncertainty of what to make of the results of their empirical work.

In another line of studies, on the other hand, scholars recognize the limitations of coefficient interpretations and supply their analysis with important additional computations. Most commonly, they draw on marginal effects or predicted probabilities as methods of interpretation. However, they rarely take into account that there are many different marginal effects and that these may change sign depending on the observations in the sample. For instance, Chen and Dimou (2005) perform an ordered logistic regression analysis adding computations of predicted probabilities in the interpretation of the results. By computing and reporting marginal effects, Meyer, Estrin, Bhaumik, and Peng (2009) investigate the influence of market-supporting institutions on the choice between greenfield, acquisition, and JVs, finding evidence suggesting that JVs are important in weaker institutional contexts. A prime example is found in a paper by Li and Li (2010), who report single measures for predicted probabilities and marginal effects while also including odds ratios interpretations to analyze the relationship between options-based variables and different types of equity entries. Although still including interpretations of coefficients, Wei, Liu, and Liu (2005) find a positive marginal effect of host country intensity on the probability of choosing an equity joint venture.

Even though reporting summary measures of marginal effects can be a good way to sum up results, it may not be sufficient when using an MLM. The econometric and organizational research methodology literature warns that the marginal effects are not constant across the range of the specific predictor. In discrete choice models with multiple outcomes, this has the consequence that the marginal effects may be positive for some values of the predictor and negative for others (Greene, 2003). When authors limit the interpretation to summary measures, they may miss important variations across the range of the predictor. For instance, the marginal effects may be significant for some values while insignificant for others or even change from negative to positive. This makes an analysis relying solely on a single summary measure of the marginal effects incomplete, leaving out important information about the nature of the marginal effects.

In sum, as in the general strategic literature (Bowen & Wiersema, 2004), the EMC literature shows systematic issues with the interpretation of the results from MLMs. Issues of this kind are concerning as they may lead to invalid inferences creating uncertainty about the conclusions to be drawn from past research findings (Wiersema & Bowen, 2009). Moreover, previous research may be missing important findings about key theoretical relationships that are very difficult to discover and present without the use of a graphical approach.

Interpreting Results From the Multinomial Logit Model

In particular, two issues complicate the interpretation of the coefficients in an MLM.[1] First, the outcomes represent contrasts among the categories, making it difficult to see the implications for each category from the coefficients. Further complicating the issue is the fact that unlike binary models a positive sign on a coefficient in an MLM does not necessarily mean that an increase in the independent variable corresponds to an increase in the probability of choosing a particular mode of entry (Long, 1997; Long & Freese, 2006). Second, the relationship between the explanatory variables and the probability of a given choice outcome is nonlinear and may even change sign across the distribution of a single predictor. Consequently, it is necessary to use other means of interpretation than we are used to in linear models. In the following, I describe two core devices[2] that researchers may use when drawing conclusions about the direction, magnitude, and significance of model variables: predicted probabilities and marginal effects.

Predicted Probabilities

One way of interpreting the relationship between a predictor and the dependent variable in an MLM is by computing and plotting predicted probabilities.

The dependent variable in this study can take the values 0 (exports), 1 (JV), and 2 (WOS). In the MLM the predicted probabilities can be calculated as

$$p_{ij} = \Pr(y_i = j|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{\sum_{j=0}^{2}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}, \tag{1}$$

which is the probability that the i^{th} firm will choose alternative j (j = 0, 1, 2), $\mathbf{x}_i$ are case-specific regressors thought to explain entry mode choice, $\boldsymbol{\beta}_j$ is the coefficient vector and contains the intercept β_{0j} and the slope coefficients β_{kj}. Thus, there is one set of coefficients for each choice alternative. In this example, the model in Equation 1 has 3 (J) equations of which only 2 (J – 1) can be estimated. Therefore, to guarantee identification, $\boldsymbol{\beta}_j$ is set to zero for one of the categories. This

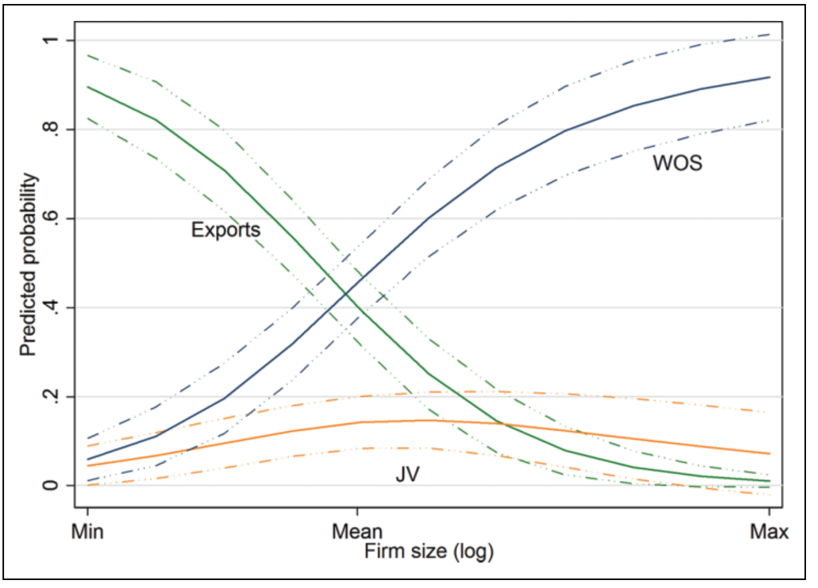

Figure 1. Analysis of the predicted probabilities of entry mode choice. Dashed lines signify 95% confidence intervals.

category is the base category, and coefficients are interpreted with respect to that category. Setting $\boldsymbol{\beta}_0 = 0$ and computing the predicted probabilities yields

$$p_{ij} = \Pr(y_i = j|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{\exp\left(\mathbf{x}_i'0\right) + \sum_{j=1}^{2}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)} \tag{2}$$

$$= \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{1 + \sum_{j=1}^{2}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}. \tag{3}$$

And for the baseline category, we have

$$p_{ij} = \Pr(y = 0|\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i'0)}{\exp\left(\mathbf{x}_i'0\right) + \sum_{j=1}^{2}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)} \tag{4}$$

$$= \frac{1}{1 + \sum_{j=1}^{2}\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}. \tag{5}$$

With Equations 3 and 5 in our hands, we can compute predicted probabilities in order to assess the relationship between a predictor and each outcome. Plotting the predicted probabilities provides a quick and informative way of presenting the relationship between a selected predictor and the predicted probabilities of the different alternatives. Figure 1 provides an example of this in the empirical example that follows. Because the predicted probabilities are point estimates, it is recommended to compute a confidence interval to take sampling variability into account. There are several procedures that can achieve this. As I show in Appendix sections A.1 and A.2, Stata allows for straightforward computation of standard errors using a variety of methods, for example, the delta method.

Marginal Effects

While predicted probabilities provide us with very informative graphical information about the direction and magnitude of the relationship, it may be difficult to precisely determine whether a relationship can really be established, especially at places where the curve is flat. To further make sense of our results, we may rely on another powerful interpretative device: marginal effects. The marginal effects are defined as the slope of the prediction function at a given value of the explanatory variable and thus inform us about the change in predicted probabilities due to a change in a particular predictor. This has made authors argue that if one wishes to draw valid conclusions about the direction and magnitude of the relation between an independent and dependent variable in an MLM, one must calculate marginal effects (Bowen & Wiersema, 2004).

Even though marginal effects for a multinomial model may be complicated to derive (Wooldridge, 2010), they have a quite distinctive and simple form (Greene, 2003). For a continuous[3] independent variable, the marginal effects are

$$\mathrm{ME}_{ij} = \frac{\partial p_{ij}}{\partial x_{ik}} = \frac{\partial \Pr(y = j|\mathbf{x}_i)}{\partial x_{ik}} = p_{ij}(\beta_{kj} - \bar{\beta}_i), \tag{6}$$

where $\bar{\beta}_i = \sum_{m=1}^{2} \beta_{km}\Pr(y = m|\mathbf{x_i})$ is a probability weighted average of the coefficients for different choice combinations, β_{km}. Equation 6 shows that the marginal effects are nonlinear and vary across values of all the variables in the model. It is highly noteworthy that the value of the marginal effect depends on several factors, counting the probabilities of other alternatives and the effect of x_{ik} on the same probabilities. This means that not only do the values of the marginal effect change as the model variables $\mathbf{x}_i$ change, the marginal effect may be positive $(\beta_{kj} > \bar{\beta}_i)$ for some values of x_{ik} and negative $(\beta_{kj} < \bar{\beta}_i)$ for others. In other words, the sign of the marginal effect may change across the range of the predictor.

The result in Equation 6 has some important implications. First, testing whether a specific coefficient is equal to zero and/or interpreting its sign makes little sense if one wishes to draw valid conclusions about the direction, significance, and/or magnitude of the relationship between the dependent variable and a given predictor (Bowen & Wiersema, 2004; Cameron & Trivedi, 2005). Indeed, there is no guarantee that the marginal effects share the sign of the coefficients in the model. Instead, the coefficient of a predictor with regard to a specific category tells us about how that predictor relates to the probability of observing a particular category relative to the base category. This means that if we insist on using a binary logit interpretation of the coefficients, we must restrict our comparisons to the base category. As I have argued, this is quite different from making statements about the relationship between a predictor and the probability of a specific outcome. More details about this are available in Appendix section A.3.

Second, because the values of the marginal effects may change sign across the predictor range, it would be very valuable to have a way of observing how the values of the marginal effects change as x_{ik} changes. Here, a graphical representation proves to be a powerful way to interpret the changes in sign and significance of the marginal effects (Hoetker, 2007). First, the marginal effects are assessed by computing the values of the marginal effects for a given predictor from its smallest to its largest sample value while holding other model variables at their mean.[4] Then, the computed marginal effects and their confidence intervals are plotted against the corresponding predictor to show how the marginal effects change from low to high predictor values. An example is provided in Figure 2 in the empirical example given later.

Summary Measures

After a graphical analysis has been performed, results may be summarized through marginal effects computed while setting the values of the model variables. There are two main ways in which results

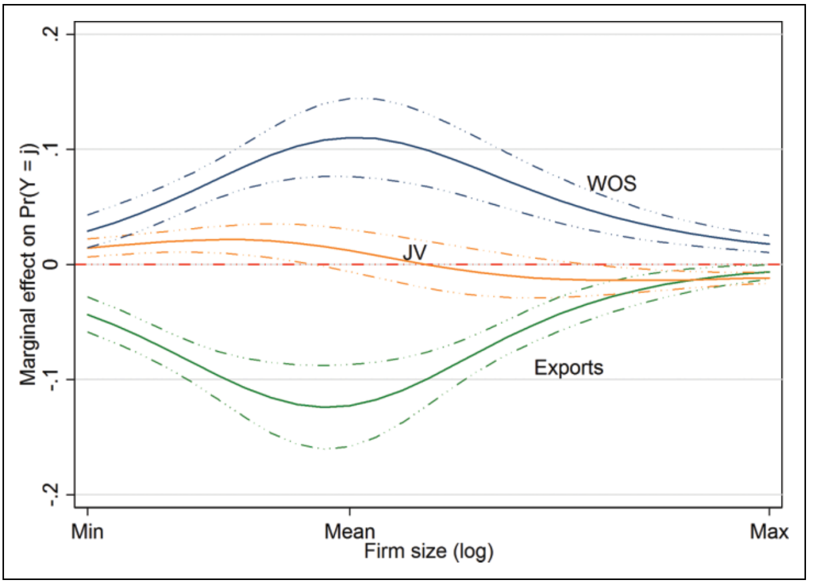

Figure 2. Analysis of the marginal effects of firm size on entry mode choice. Dashed lines signify 95% confidence intervals.

are usually summarized. The first is to set all of the predictors to their mean values resulting in marginal effects at the mean (MEM):

$$\mathrm{MEM} = \bar{p}_j(\beta_{kj} - \bar{\beta}_i), \tag{7}$$

where $\bar{p}_j$ is now calculated by holding $\mathbf{x}_i$ at their mean values. Obviously, centering or standardizing all independent variables considerably simplifies the computation. One downside of MEMs is that it is unlikely that there is a unit in the sample that is average on all model variables. To avoid this, another approach called average marginal effects (AME) relies on actual values of the independent variables. First, the marginal effects for each unit are produced and then averaged:

$$\mathrm{AME} = \frac{1}{n}\sum_{i=1}^{n} p_{ij}(\beta_{kj} - \bar{\beta}_i). \tag{8}$$

Because the MEM and AME may produce different estimates (Bartus, 2005) and because there is no agreement as to which of the two is the most representative (Greene, 2003), both can be included to provide the most informative summary of the marginal effects.

The main issue with the two approaches of summarizing results described previously is that they only produce a single estimate of the marginal effects. This means that no matter how we decide to average the effects, we may be unintentionally obscuring important differences across the range of the predictor (Williams, 2012). This makes it preferable to compute the MEM and AME at relevant values (e.g., low, mean, and high) of the predictor, especially if the graphical analysis reveals that the marginal effects change sign across the range of the predictor. In the literature, this procedure is occasionally referred to as marginal effects at representative values (MERs). For the AME and MEM summary measures, computing MERs involves choosing values representative to how the marginal effects change. These values are commonly one standard deviation above and below the mean but may be expanded if graphical representation reveals it to be necessary. An example is provided in Table 4, and the Stata implementation is shown in Appendix A.2.

It is worth repeating that the real analysis lies in the graphical representation described previously and illustrated in the following, which should therefore always be reported. If space limitations prevent including plots, the author should as a minimum comment on changes in the sign and significance of the marginal effect across the range of the predictor.

In order to determine the statistical significance of the marginal effects, one cannot rely on the model coefficient of the relevant predictor. Because of the complex nature of the marginal effects, it is necessary to compute the values of the marginal effect from the smallest to the largest sample value of the predictor, its standard error, and z-statistic. As with the predicted probabilities, the standard errors can be obtained using the delta method. The standard errors are used to compute the confidence intervals, which are included together with the values of the marginal effects in the graphical representation.

Summarized, the sign of an estimated coefficient in an MLM cannot be used to determine the relation between an independent variable and EMC. Instead, the relationship should be analyzed by computing marginal effects and their standard errors. In the following section, I adapt and build on the approach suggested by Wiersema and Bowen (2009). First, I plot predicted probabilities for the entry mode options against each predictor. Then, I plot the computed marginal effect surrounded by confidence intervals against the values of the predictor. Finally, I report the MEMs and AMEs to supply and summarize the results provided in the graphical representation.

Data and Variables

Data

The empirical example in this study is based on a sample of Danish, Swedish, and Norwegian firms focusing on the entry mode choice (exporting/JV/WOS) made by these companies. A questionnaire was used to collect the data except for the data on the cultural distance measure. The firms were asked for their most recent foreign entry. Consequently, about 90% of all entries in the sample were within the time span from 2007 to 2012. This ensured that institutional environmental changes that happen over time were relatively fixed (Brouthers et al., 2008b). Moreover, potential problems caused by recall bias were minimized by only asking with regard to the most recent entry (Brouthers & Brouthers, 2003).

The sample was drawn from the Userneeds database. A random sample was drawn from the database, and online questionnaires were sent to managers at the corporate level. In all, 876 Danish (DK), 1014 Swedish (S), and 1,739 Norwegian (N) firms were contacted, yielding a total of 3,629 firms. After follow-up rounds, 2,107 (656 from DK, 609 from S, and 842 from N) had responded to the survey. A total of 1,420 indicated that they had no international activity, and 332 selected not to participate leaving 355. The final sample used in the empirical example includes 246 firms (82 DK, 86 S, and 78 N). Subtracting firms with no international activity, the overall response rate was 16%, and the usable response rate was about 11.1%[5]. Respondents in the sample had a median of 113 and a mean of 12,335 employees, and the majority was service firms (72% of the sample). In the sample, about 43% of the entries were WOS, 16% were JVs, and 41% were exporting. These firms made their latest entries in 91 different countries of which the most commonly entered were Sweden (19), China (18), Germany (14), and the United States (12).

Variables

Entry mode type was obtained from respondent firms. In the questionnaire, the respondents were asked which entry mode was chosen in their most recent foreign market entry. The respondents could choose between greenfield, acquisition, JV, contractual agreements, and independent exporting. I conceptualized the mode of entry as a categorical variable: (0) independent exporting, (1) JV, and (2) WOS (including greenfield and acquisition). Independent exporting modes are defined as non-equity market–based modes where the firm uses entities in the host country to either provide or produce their product or service. JVs are modes where the entrant shares equity ownership of the host country operations with a local partner. WOS are operations where the investing firms hold an equity share of 95% or more (Brouthers et al., 2008a).

The predictor variable used as an example in the empirical demonstration is *firm size*. In the entry mode literature, firm size is one of the most commonly used predictors found in 40 studies in the period 1980 to 2006 (Canabal & White, 2008). Scholars suggest that larger firms tend to use more integrated entry modes. Commonly, it is argued that larger firms possess more resources than smaller firms (Combs & Ketchen, 1999) and that firm size is an indication of a firm's potential to meet resource requirements (Buckley & Casson, 1998). However, literature reviews report inconsistent empirical findings (see e.g., Brouthers & Hennart, 2007). Some studies find no significant difference in the choice between equity and non-equity modes (Brouthers & Nakos, 2004; Nakos & Brouthers, 2002) depending on the size of the entrant, while others have found evidence to suggest that firm size has different impacts on exporting, JV, and WOS modes (Shrader, 2001). In the empirical example in this study, I measure firm size as the log of the number of employees worldwide (Brouthers & Brouthers, 2000).

Several control variables were included. I controlled for the degree of *international experience* measured using two different items (alpha = 0.75) from Brouthers et al. (2008a) and Brouthers and Dikova (2010). I used the log-transformed value of the intensity of firm experience (number of years of general international experience) as well as the log-transformed value of the diversity of firm experience (number of countries where the firm had operations). Furthermore, I controlled for potential influences from industry differences (Erramilli & Rao, 1990). Following previous research (Brouthers, 2002; Brouthers & Brouthers, 2003; Brouthers & Nakos, 2004), I created a dummy variable, *service*, based on the respondents' answers to whether the organization had established a manufacturing (value of 0) or a service operation (value of 1). I controlled for potential home country differences (Brouthers et al., 2008b). A Danish and a Norwegian dummy variable were each coded 1 if the respondent was from the specified home country and 0 if otherwise.

Finally, I included variables based on transaction cost theory (TCT). In TCT terminology, specific assets have less value outside the transaction in which they are tailored to be utilized (Williamson, 1985). To gauge the specificity of assets, I followed Dikova and van Witteloostuijn (2007) and asked respondents how much money as a percentage of annual sales was spent on R&D (*technological intensity*). Internal uncertainty was measured using Kogut and Singh's (1988) cultural distance index. This index uses the differences in the scores on Hofstede's (1980, 2001) cultural indices between the foreign country and the home country of the entering firm, which in the case of this study was either Denmark, Norway, or Sweden. The higher the score on the cultural index, the higher the level of cultural distance. Finally, *external uncertainty* was measured using a set of nine 7-point scale Likert type questions taken from Brouthers and Dikova (2010) and Brouthers (2002). Confirmatory factor analysis showed acceptable fit (Comparative Fit Index [CFI] = 0.91, root mean square error of approximation [RMSEA] = 0.08, standardized root mean square residual [SRMR] = 0.05). Low values of the composite index represent low uncertainty, and high values represent high uncertainty. Because the measures of the control variables are on different scales, they are standardized before the analysis.

Empirical Example

In the following empirical example, I present the recommended procedure for an analysis of the results from an MLM. Specifically, I concentrate on the association between firm size and EMC.

Table 1. Correlation and Descriptive Statistics.

Variables	Mean	SD	1	2	3	4	5	6	7	8
1. Exports (0) versus joint venture (JV) (1) and wholly owned subsidiary (WOS) (2)	1.03	0.92								
2. Firm size (log)	5.27	3.09	0.392***							
3. International experience (log)	2.48	1.00	–0.041	0.494***						
4. Service	0.72	0.45	0.059	–0.138*	–0.164*					
5. Denmark	0.33	0.47	0.053	0.041	0.067	0.045				
6. Norway	0.32	0.47	0.084	0.000	–0.076	–0.054	–0.482***			
7. Technological intensity	0.15	0.21	0.020	–0.052	–0.172**	0.023	–0.005	0.056		
8. Cultural distance	2.81	1.90	–0.132*	0.056	0.343***	–0.089	0.056	–0.184*	–0.090	
9. External uncertainty	3.7	1.38	–0.250***	0.055	0.249***	0.044	0.002	0.004	–0.126*	0.298***

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 2. Results of Multinomial Regression.

Variables	Model A (1) Exports Versus Joint Venture	(2) Exports Versus Wholly Owned Subsidiary	(3) Joint Venture Versus Wholly Owned Subsidiary	Model B (4) Exports Versus Joint Venture	(5) Exports Versus Wholly Owned Subsidiary	(6) Joint Venture Versus Wholly Owned Subsidiary
International experience	0.782***	0.143	−0.639**	0.140	−0.706**	−0.846**
	(0.230)	(0.162)	(0.232)	(0.293)	(0.225)	(0.283)
Service	−0.317	0.424	0.741†	−0.151	0.695†	0.846†
	(0.422)	(0.332)	(0.436)	(0.447)	(0.393)	(0.447)
Denmark	0.452	0.574	0.122	0.496	0.683	0.187
	(0.468)	(0.358)	(0.479)	(0.491)	(0.416)	(0.490)
Norway	0.039	0.505	0.466	0.270	0.872*	0.602
	(0.525)	(0.353)	(0.538)	(0.560)	(0.424)	(0.542)
Technological intensity	−0.180	0.018	0.198	−0.243	−0.059	0.184
	(0.279)	(0.140)	(0.279)	(0.290)	(0.166)	(0.288)
Cultural distance	0.430†	−0.071	−0.501*	0.536*	0.017	−0.520*
	(0.225)	(0.163)	(0.229)	(0.243)	(0.194)	(0.243)
External uncertainty	−0.245	−0.564***	−0.319	−0.297	−0.709***	−0.412†
	(0.227)	(0.160)	(0.233)	(0.238)	(0.191)	(0.249)
Firm size				**0.369*****	**0.541*****	**0.171**†
				(0.093)	(0.081)	(0.089)
Constant	−3.220***	−1.027†	2.193**	−3.486***	−1.968**	1.518†
	(0.809)	(0.552)	(0.834)	(0.847)	(0.663)	(0.855)
R^2 Nagelkerke	0.214	0.214	0.214	0.437	0.437	0.437
Akaike Information Criterion	2.018	2.018	2.018	1.781	1.781	1.781
χ^2	51.96***	51.96***	51.96***	117.8***	117.8***	117.8***
Correctly classified	56.75	56.75	56.75	67.07	67.07	67.07
N	246	246	246	246	246	246
Change in χ^2 Model A				65.86***	65.86***	65.86***

Note: Standard errors in parentheses.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

First, I consider hypothesis testing and model fit. Then, I analyze the association between firm size and EMC through graphical representations, first using predicted probabilities and then marginal effects. Finally, I summarize the results using the summary measures presented previously. The Stata code that generates the figures and measures is available in Appendix sections A.1 and A.2.

Table 1 reports the means, standard deviations, and correlations of all variables. Table 2 presents the coefficients and model fit statistics of the multinomial logistic regression analysis[6] with EMC as an unordered categorical dependent variable.

Hypothesis Testing and Model Fit

Model A (Panels 1, 2, and 3) contains only the control variables. Model B (Panels 4, 5, and 6) adds the predictor variable firm size to the control variables. Table 2 contains the model coefficients for

Table 3. The Marginal Effect of Firm Size on the Probability of Choosing Exports, Joint Venture, or Wholly Owned Subsidiary.

Marginal effect on the probability of	Marginal Effect at Variable Means (MEM)	Average Marginal Effect (AME)
Exports	–0.1202***	–0.0829***
	(0.0174)	(0.0462)
JV	0.0099	0.0094
	(0.0096)	(0.0073)
WOS	0.1102***	0.0735***
	(0.0175)	(0.0086)

Note: Delta-method standard errors in parentheses.
***$p < .001$.

Table 4. The Marginal Effect of Firm Size on the Probability of Choosing Joint Venture.

Value of Predictor	Marginal Effect at Variable Means (MEM)	Average Marginal Effect (AME)
Low (1 *SD* below)	0.0213***	0.0223***
	(0.0054)	(0.0042)
Mean	0.0100	0.0174*
	(0.0096)	(0.0087)
High (1 *SD* above)	–0.0110	–0.0022
	(0.0091)	(0.0094)

Note: Delta-method standard errors in parentheses.
*$p < .05$. ***$p < .001$.

the different combinations of outcomes. Because there is no natural base case, the three possible combinations of choices are computed by varying the base case with the last mentioned category being the base. For instance, in the choice equation JV versus WOS, WOS is the base case. Thus, for studies with the same number of alternatives and a natural base case, the output may be simplified to two columns per model.

Looking at the model fit statistics, we observe that the likelihood ratio (LR) test in Model A is significant ($p < .001$, $R^2 = 0.214$), meaning that at least a subset of the predictors have non-zero effects. Model B including firm size exhibits a significant increase in chi-square ($p < .001$) and a substantial increase in R^2 from 0.214 to 0.437. Thus, Model B demonstrates increased explanatory power over Model A. Model B's Akaike Information Criterion (AIC) is lower (1.781), which is an indication that the model fit is improved enough to compensate for the fact that the model grows in complexity (Raftery, 1996). Overall, these results indicate a good model fit with the predictor firm size. Note that the procedure for assessing model significance and the interpretation of model fit statistics is the same as in binary models.

The model fit statistics form an important basis for the whole analysis and should always be reported and commented on as it is custom in binary models. If the overall model is significant, we can start examining the hypotheses about individual predictors. Several hypotheses may be tested in an MLM. One of particular interest may be to test the null hypothesis that firm size is independent of the choice between the three entry modes. This can be tested using an LR test or a Wald test.[7] This type of hypothesis testing is different from binary models because we have more than one coefficient for each variable. Thus, we require a test of the overall significance of a predictor. Because we already used the LR test when comparing the models with and without the predictor previously, I use the Wald test to test this hypothesis. Letting β_{1k} be the coefficient for firm size, the Wald test returns a value of 45.14 ($p < .000$), and we can thus reject Hypothesis 0: $\beta_{1k} = 0$, where k = 0, 1, 2. Put differently, we can reject the hypothesis that firm size is independent of EMC. This fits the information we gain from observing the coefficients in Table 2: Firm size is statistically significant and positively related to choosing JV over exports ($p < .001$) and WOS over exports ($p < .001$). It is worth mentioning that if the overall test is significant, it is useful to analyze all three outcomes in the predicted probabilities and marginal effect analyses even though a coefficient is nonsignificant in the choice equation (Bowen & Wiersema, 2004).

Analysis of Predicted Probabilities

As explained previously, the directional relationship between a predictor and the probability of a choice depends on all variables and their estimated coefficients across the choice alternatives. Therefore, we cannot rely on the estimated coefficients when evaluating the direction of the relationship between firm size and the probability of choosing a specific entry mode. Instead, we make sense of the results by computing and plotting the predicted probabilities. A plot of the predicted probabilities for the three different entry mode categories and firm size is shown in Figure 1.

Figure 1 shows that the probability of entering through exports drops dramatically as firm size increases. Thus, there seems to be a clear negative relationship between exporting entries and firm size. Inversely, the probability of entering through a WOS increases from being miniscule for small firms to very large for large firms, indicating a clear-cut positive relationship. According to the figure there seems to be a cut-off point around the mean where the WOS and exports curves are steepest. After this point, firms have a higher probability of entering through WOS than exports.

JV entries, on the other hand, seem to have a less straightforward link to firm size: Over the range of the predictor, the probability of entering through a JV tends to first increase and then decrease. This is a clear example of the complexities of analyzing relationships using an MLM. For small firms, the probability of a JV entry seems to increase as firms grow larger. However, as firms grow larger, the positive impact on the probability becomes less before the probability of a JV choice finally starts dropping.

The computation and plotting of predicted probabilities has several advantages. First, the graphical representation in Figure 1 provides a clear and intuitive way of interpreting and communicating the relationship between a predictor and the dependent variable. A similar interpretation is very difficult to reach by observing the model coefficients in Table 2. Indeed, the positive and significant coefficient on the choice between WOS and exports tells us that an increase in firm size is associated with an increase in the probability of WOS relative to exports, which is consistent with the plot in Figure 1. But it is much less apparent how this result is related to the positive coefficient between JV and exports and the positive coefficient between WOS and JV without having first looked at the plot in Figure 1. Note that a coefficient interpretation also has the disadvantage that it magnifies in complexity as a function of how many values the dependent variable takes. With no natural base case, a dependent variable with four values would have us evaluate six different choice combinations, while it would just mean one extra line in the predicted probabilities plot. These are clear examples of the interpretative power of plotting predicted probabilities. Not only are we able to gain a more complete understanding of model relationships, we are also much better suited to communicate and explain the results to others.

Second, we can make precise statements about the predicted probabilities given specific values of the predictors. Small firms (5th percentile) have a predicted probability of 0.8378 of entering through exports while having just 0.0992 of choosing WOS as entry vehicle. Conversely, large firms (95th percentile) have a predicted probability of 0.0464 of exports and 0.8446 of choosing WOS. Such assessments cannot be made based on the model coefficients.

Third, we can get a clear idea of the precision of the estimates across the range of the predictor. As shown in Figure 1, the width of the confidence intervals varies greatly for different values of firm size. Due to sampling variability, the interval for the WOS line is considerably narrower around the middle than for large values of firm size.

In sum, a predicted probability plot as the one in Figure 1 provides invaluable information about the relationship between a predictor and the dependent variable. Although very informative, it is limited in the sense that it can be difficult to determine whether an increase in a predictor is significantly associated with an increase/decrease in predicted probabilities. Whereas there is no doubt that an increase in firm size is positively associated with an increase in the predicted probability of WOS, it is not evident whether the slope on the JV curve is steep enough to warrant a similar statement. To investigate claims about how the predicted probabilities change when the predictor changes, we turn to the marginal effects.

Analysis of Marginal Effects

Figure 2 graphically illustrates the estimated marginal effects surrounded by 95% confidence intervals. With marginal effects, we draw the attention to the curvature of the relationship rather than the relationship itself. The fact that marginal effects are second-order relationships makes them harder to interpret than predicted probability curves.[8] But what we lose in intuition we gain in information as we have now become able to precisely assess the magnitude and significance of the relationship between a predictor and the choice outcomes. When interpreting Figure 2, it is helpful to keep in mind that the marginal effects represent the slope of the curve for given predictor values.

The marginal effects in this example are calculated using the MEM approach, although an AME approach could prove equally valid. With Figure 2 and our computations of the marginal effects in hand, we can make concrete statements about how changes in firm size are related to changes in the predicted probabilities. We observe that the values of the marginal effect of firm size on the predicted probability of WOS grow increasingly positive, slow down, and start dropping for higher values of the predictor. Note that the drop does not mean that the probability of WOS is getting lower but simply that the rate of increasing probability is slowing down. The values of the marginal effects range from 0.0178 to 0.1104 and are always positive. The smallest firms (5th percentile) have a value around 0.0289, mean-sized firms have a peak value around 0.1104, and the largest firms (95th percentile) have a marginal effect value around 0.0178. Intuitively, this makes sense as the smallest/largest firms already have a very low/high probability of entering through a WOS: An equal increase in firm size is much more likely to affect firms that are equally likely to choose a WOS over another entry mode than a firm that is already very likely to enter through a WOS.

A clear advantage of marginal effects is that they provide us with rich and intuitively meaningful information not available through interpretation of coefficients. For mean-sized firms, a 1% increase in firm size is associated with an increase of 0.1104 in the predicted probability of entering through WOS. On the other hand, an equal increase in firm size for small firms is associated with an increase of just 0.0178 in the predicted probability. Note that this interpretation is completely consistent with the information gained from Figure 1 where it can be seen that the slope of the WOS curve is steepest around the mean. By adding the computation and plotting of the marginal effect to our analysis, we can make formal statements about how much the predicted probabilities for a given outcome changes when we change firm size at specific values.

During the analysis of Figure 1 previously, it became evident that the relationship between the probability of JV entry and firm size was more complicated to interpret. The JV curve in Figure 2 confirms our notion that the marginal effects shift from being positive for small firms and negative for large firms with a range spanning from 0.0216 to –0.0138. Observing the JV curve in Figure 2, another benefit of the marginal effects graph becomes clear: Where it was difficult to tell from Figure 1 whether the increase and decrease in the probabilities for small and large firms, respectively, were significantly different from zero, Figure 2 clearly informs us about the uncertainty around the estimates and how it depends on the predictor level. We observe that for small firms the confidence intervals do not include zero until it nears the mean of firm size where the marginal effects become insignificant. Further, we observe that for firms near the maximum level of firm size, the marginal effects become significantly negative. The z-statistic values reflect this observation, as they span from 4.27 to –4.65 over the range of firm size. Note that not only is this important information unobtainable through an analysis of the coefficients, but it may also be difficult to deduct from an analysis of the predicted probabilities alone.

The results from the previous analysis are consistent with the distinguishability test performed in the beginning: The marginal effects of firm size exhibit very different behavior depending on the outcome category. This supports the conclusion made earlier that the JV and WOS categories should not be collapsed into a single equity category.

When presenting the results, I recommend graphing the marginal effects for the three outcomes together as it is done for the predicted probabilities.[9] This provides a clear picture of how the distribution shifts from one category to another across the levels of the predictor. If authors choose to graph the outcomes separately, they should be cautious about exaggerating trivial effects due to differences in scaling and careful about encouraging that the relationships are treated as if they were independent.

A graphical analysis such as the one in Figures 1 and 2 offers the most complete assessment of the relationship between a given predictor and the dependent variable in an MLM and should always be reported. Furthermore, the visualizations make it considerably easier to communicate the results. If space limitations or other restrictions prevent their inclusion, summaries of the marginal effects may be reported (see the following) with comments about the range of the marginal effects, the uncertainty of the estimates (e.g., their z-statistic values), and a description of whether the values of the marginal effects change sign over the range of prediction.

Summarizing Results

As described earlier, marginal effects can be summarized after the graphical interpretation. However, in an MLM, great care must be taken to avoid valuable information to be lost to too simplistic conclusions. For instance, when computing the MEM and AME for firm size with regard to the three outcomes, we get the quantities presented in Table 3.

When comparing the results to information gained from Figures 1 and 2, the measures are clearly ignoring important information about relationships, most severely with regard to the JV category. According to the MEM measure, holding all variables, even firm size, at their mean value and increasing firm size by 1% is associated with an insignificant increase in the probability of a JV entry of 0.0099. Indeed, this result is consistent with what we observed in Figure 2: Around the mean, the marginal effect of firm size is insignificant. However, this completely ignores the significant marginal effects for small and large firms. The AME measure does not do any better. The reason is that the averaging of all marginal effects makes the positive and negative values cancel each other out arriving at a value near zero.

Instead, when summing up the marginal effect, I recommend computing the summary measures at a low, mean, and high value of the predictor. For low and high values, I choose one standard deviation below and above the mean. Table 4 shows the recommended presentation format for the JV outcome.

Table 4 more precisely sums up the nature of the marginal effect across the range of the predictor. At low values of the predictor, the marginal effect of firm size is significantly positive. The AME measure tells us that when averaging the marginal effects computed for each sample unit and holding the predictor at one standard deviation below the mean, a 1% increase in firm size is significantly associated with a 0.0223 increase in the probability of a JV entry. This is consistent with what we learned from Figures 1 and 2.

Two further comments are worth making about Table 4. First, the AME actually reports a significantly positive marginal effect at the predictor mean for the JV outcome. This is an example of the differences that may arise when not fixing the model variables at their mean but at their actual sample values. Despite of this difference, the measures still summarize the tendency we observed in the figures. Second, the sign of the summary measures for firms with a size one standard deviation above the mean is negative but not significant. While this does not suggest a negative marginal effect for large firms, I caution that the summary measures are neither a test nor a complete description of the marginal effect in an MLM. As we saw in Figure 2, the marginal effects do not become significantly negative until near the maximum value. In fact, summary measures with firm size fixed at the 95th percentile return significantly negative values. However, after a thorough analysis of the relationship between the predictor and the dependent variable as performed previously, the format in Table 4 can be a good way to summarize the marginal effects.

Conclusion

To summarize, extra care needs to be taken when analyzing and testing hypotheses about relationships in MLMs compared to binary models. The interpretational methods most commonly used in praxis and the solutions proposed in this study are summed up in Table 5.

In this article, I demonstrate how the most complete analysis of the relationship between a predictor and the dependent variable can be achieved utilizing the following approach. First, the entire set of predictors is tested using an LR or Wald test procedure. If a specific predictor does not have a zero effect in the population and improves model fit statistics to a satisfactory amount over the base model, the predicted probabilities and marginal effects are calculated and analyzed to evaluate the sign, magnitude, and statistical significance of the marginal effects.

Interpretations based on the model coefficients should be made with great care as these do not represent the relationship between a predictor and the predicted probability of a specific outcome. Instead, a full presentation of the relationship between the predictor and the dependent variable is achieved through a graphical representation. In the first graph, the predicted probabilities (including confidence intervals) are plotted against the predictor to gain an intuitive understanding of the relationship as in Figure 1. In the second graph, the marginal effects (including confidence intervals) are plotted to assert the variation in sign, magnitude, and statistical significance across the range of the predictor as in Figure 2. If the marginal effects do not change sign or significance across their range of variation, measures such as MEM and AME may be reported to summarize the relationship. If the marginal effects are too complicated to sum up in single measure, for example, due to changes in sign, significance, or major changes in magnitude, MEM and AME

Table 5. Summary of Dominating Approaches and Proposed Solution.

Standard Interpretational Device	Drawbacks of the Standard Method	Gains From a Graphical Approach	Losses From a Graphical Approach
Coefficient-based	Restricts the analysis to statements about one category relative to the base. Complicates communication of results. Risks misinterpretation by confusing the sign and size of a coefficient with the direction and size of the effect on the probability of a single outcome.	Information from the coefficient-based interpretation is retained and supplemented by information about magnitude and reliability. Results are more easily and intuitively communicated by graphing predicted probabilities. Risks of misinterpretation are greatly reduced.	Graphs of predicted probabilities may occasionally make it difficult to assess the reliability of a relationship.
Single summary measures of marginal effects	Ignore that marginal effects may change sign and size across the range of the predictor. Ignore that marginal effects are nonlinear across the range of the predictor. Risk that the analysis may be incomplete because information about change in magnitude, direction, and reliability is reduced to a single or a few summary estimates.	Clearly shows the full behavior of the marginal effects across the range of the predictor. Acknowledges and presents the complexity of the marginal effects in an intuitive fashion. Provides a complete presentation of the nature of the marginal effects.	Authors should take care not to exaggerate trivial effects through careful scaling of axes.

may be reported with great care for relevant values (e.g., low, mean, and high) of the predictor as shown in Table 4.

In this study, I proposed that researchers base their interpretations of the results from the MLM less on coefficients and single summary measures and more on complete graphical representations. I provided practical guidelines about what researchers concretely may do and used a practical example to illustrate the gains and losses of using a graphical approach. If we are interested in inferring the true nature of the relationship between a predictor and the dependent variable in an MLM, we must acknowledge that coefficients and sometimes even single summary measures are potentially misleading. Instead, by shifting to full graphical representations we leave little in the dark and provide our readers with full and relatively intuitive information about the intrinsically nonlinear and complex nature of the relationship.

The issues I highlight and the suggestions I make in this study are not restricted to the MLMs alone but are applicable to other discrete models with multiple outcomes. The flexibility of the proposed approach is relevant because of one particular major hurdle when applying the MLM to strategic choices: the assumption of independence of irrelevant alternatives. This assumption implies that the choice between any two alternative pairs is a binary logit model. When using the MLM to model strategic decision making, it is often questionable whether this assumption is justified. However, it is not clear whether the violation is severe enough to warrant using richer models for unordered choices that relax this assumption. Future research should investigate in which areas in organizational research we may make substantial progress by applying less restrictive models as, for example, the nested logit or multinomial probit when we investigate discrete choices with multiple outcomes. If we are indeed in possession of the alternative-specific data sets demanded in these models, we can avoid having to impose potentially unrealistic restrictions on individual and strategic choices.

Appendix

A.1. Stata Code for Hypothesis Tests and Producing Figure 1

The following Stata code performs the hypothesis tests and generates the model fit statistics used in the empirical example. The annotated commands appear in the following.

```
1.  /*Hypothesis Testing and Model Fit*/
2.  mlogit exjvwo $indvars
3.  fitstat, saving(m1)
4.  mlogit exjvwo $indvars firmsize
5.  fitstat, using(m1)
6.
7.  /*Individual hypothesis testing*/
8.  test firmsize // Wald test of firm size
```

The commands are largely self-explanatory. Line 2 runs the multinomial logit command for the model including only the control variables contained in the global variable `indvars` and saves the model fit statistics in "m1." In lines 4 and 5, the model is run including the variable of interest, in this case firm size, and the model fit characteristics are compared to those in the baseline model (m1). In line 8 the individual hypothesis of firm size is run.

The next piece of code shows how to generate and graph the predicted probabilities as done in Figure 2. After running the `mlogit` command, this can be accomplished using the following command contained in the Stata add-on SPost (Long & Freese, 2006):

```
prgen firmsize, gen(pr) rest(mean) ci
```

The command has several options. The ones used here generate three sets of new variables with the prefix "pr," one set for each outcome in the dependent variable. The `rest(mean)` option sets the rest of the model variables to the desired relevant values, in this case the mean. The `ci` option generates variables with the suffices "lb" and "ub" containing the upper and lower bound confidence intervals, respectively, for each outcome. The variables can then easily be plotted using the conventional `twoway` command. The code that generated Figure 1 appears in the following.

```
/*Analysis of Predicted Probabilities*/
prgen firmsize, gen(pr) rest(mean) ci
/*Generate graph*/
twoway (line prp0 prx, lcolor(green) lpattern(solid)) ///
(line prp0lb prx, lcolor(green) lpattern(dash_3dot) lwidth(thin)) ///
(line prp0ub prx, lcolor(green) lpattern(dash_3dot) lwidth(thin)) ///
///
(line prp1 prx, lcolor(orange) lpattern(solid)) ///
(line prp1lb prx, lcolor(orange) lpattern(dash_3dot) lwidth(thin)) ///
(line prp1ub prx, lcolor(orange) lpattern(dash_3dot) lwidth(thin)) ///
///
(line prp2 prx, lcolor(navy) lpattern(solid)) ///
(line prp2lb prx, lcolor(navy) lpattern(dash_3dot) lwidth(thin)) ///
(line prp2ub prx, lcolor(navy) lpattern(dash_3dot) lwidth(thin)) ///
, scheme(s1color) ///
ytitle("Predicted probability", color(gs1)) ///
ylabel(, grid) ///
xtitle("Firm size (log)", color(gs1)) ///
text(0.05 6 "JV", place(se)) text(0.75 11 "WOS", place(se)) ///
text(0.6 2 "Exports", place(se)) ///
legend(off) ///
xlabel(`minValFirm' "Min" `meanValFirm' "Mean" `maxValFirm' "Max")
```

Predicted values for specific values of the predictor can be generated using the `prvalue` command. The examples that follow show the predictions used in the empirical example in this study for firm size set at the 5th and 95th percentile, here represented by local variables. The `delta` option specifies that the standard errors are calculated using the delta method:

```
prvalue, x(firmsize=`5thPercentile’) rest(mean) delta

prvalue, x(firmsize=`95thPercentile’) rest(mean) delta.
```

A.2. Stata Code for Producing Figure 2

To compute the marginal effects, the very useful `margins` command can be used. However, in order to generate Figure 2, some additional programming is necessary. While computing and generating the marginal effects for each outcome separately is easily achieved through the `marginsplot` command, combining the marginal effects across the predictor range for each outcome in one graph requires some additional programming. The reason is that the `margins` command is limited in the sense that it only simulates one model at a time. This makes it a little convoluted to use for estimations of the multinomial logit model with different outcomes. Luckily, using a nested loop we can create a not too overly complicated procedure that saves the marginal effects data for each outcome:

```
forvalues j=0/2 {
 qui mlogit exjvwo $indvars firmsize
 qui margins, dydx(firmsize) at(firmsize=(0(1)13)) predict(outcome(`j')) ///
 atmeans post
 mat x=J(13,3,.)
 mat z=(1\2\3\4\5\6\7\8\9\10\11\12\13)
 forvalues i=1/13 {
   mat x[`i',1] = _b[`i'._at] // get marginal effect estimates
   mat x[`i',2] = _b[`i'._at] - 1.96*_se[`i'._at] // compute lower limit
   mat x[`i',3] = _b[`i'._at] + 1.96*_se[`i'._at] // compute upper limi
  }

 mat x=x,z
 mat colnames x = marg_`j' lci_`j' uci_`j' at_`j'
 svmat x, names(col)
}
```

Inside the first loop the `mlogit` and `margins` commands are run for each outcome J. Thus, in this study the `‘‘j’’` is replaced by first 0, 1, and finally 2 in the `margins` command. The output is suppressed using the `quietly` prefix. Running the `margins` command with the option `dydx(firmsize) atmeans` computes the marginal effects at the mean. Dropping the `atmeans` option instead computes the average marginal effects. The marginal effects are calculated for the specified range of the predictor inside the `at()` option. Note that this at() option may also be used to compute the marginal effects at representative values (MERs). In the previous piece of code, increments of 1 are used to save space while the values in the empirical example are calculated using increments of 0.5. Finally, the `post` argument is used to save the results of the estimation in matrix form.

In line 5, an empty matrix “x” consisting of 13 rows and 3 columns is created. This will later be filled with the marginal effects data. Line 6 creates an empty vector “z” explicitly containing the `_at` values 1 through 13.

In lines 7 through 11, we iterate 13 times and each time store the marginal effects data in the “x” matrix. Line 8 stores the marginal effect estimates and lines 9 and 10 store the lower and upper bounds of the 95% confidence intervals. In line 13, the “z” vector is appended to the “x” matrix after which the column names can be changed in lines 14 and 15.

```
/*Generate graph*/
twoway (line marg_0 at_0, lcolor(green) lpattern(solid)) /// J = Exports
(line lci_0 at_0, lcolor(green) lpattern(dash_3dot) lwidth(thin)) ///
(line uci_0 at_0, lcolor(green) lpattern(dash_3dot) lwidth(thin)) ///
///
(line marg_1 at_1, lcolor(orange) lpattern(solid)) /// J = JV
(line lci_1 at_1, lcolor(orange) lpattern(dash_3dot) lwidth(thin)) ///
(line uci_1 at_1, lcolor(orange) lpattern(dash_3dot) lwidth(thin)) ///
///
(line marg_2 at_2, lcolor(navy) lpattern(solid)) /// J = WOS
(line lci_2 at_2, lcolor(navy) lpattern(dash_3dot) lwidth(thin)) ///
(line uci_2 at_2, lcolor(navy) lpattern(dash_3dot) lwidth(thin)) ///
, legend(off) ///
scheme(s1mono) ///
ylabel(, grid) ///
ytitle("Marginal effect on Pr(Y = j)", color(gs1)) ///
xtitle("Firm size (log)", color(gs1)) ///
title("") ///
xlabel(`MinVal’ "Min" `MeanVal’ "Mean" `MaxVal’ "Max") ///
text(0.02 13 "JV", place(se)) ///
text(0.08 20 "WOS", place(se)) ///
text(-0.08 18 "Exports", place(se))
```

The previous piece of code is simply the `twoway` command used to generate Figure 2. It uses the values computed in the previous loops to generate one line for each outcome, each surrounded by a confidence interval. Of course, this can be moderated to change the looks of the graphs. Note that the previous procedure can easily be adapted to be used for other discrete models with multiple outcomes as ordered logit, multinomial probit, and so on. For more on how to use the margins command to compute and graph marginal effects and adjusted predictions, I recommend consulting Williams (2012) or Gauvin (2012).

A.3. Coefficient Interpretation for Base Case Comparison

It is possible to express the coefficients in the MLM as binary logit models. Given that we restrict the model to two choices j and h, it can be shown that

$$\begin{aligned}\pi_{ij} &= \Pr(y_i = j \mid y_i = j \text{ or } h)\\ &= \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{\exp\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right) + \exp\left(\mathbf{x}_i'\boldsymbol{\beta}_\mathbf{h}\right)}\\ &= \frac{\exp\left(\mathbf{x}_i'(\boldsymbol{\beta}_j - \boldsymbol{\beta}_h)\right)}{1 + \exp\left(\boldsymbol{\beta}_j - \boldsymbol{\beta}_h\right)},\end{aligned} \tag{A1}$$

where the result in Equation A1 is a logit model with the coefficient $(\boldsymbol{\beta}_j - \boldsymbol{\beta}_h)$. The result shows the conditional probability of observing outcome j given that outcome j or h is observed. Setting one of the categories to the base category by restricting $\boldsymbol{\beta}_h = 0$, we get

$$\begin{aligned}\pi_{ij} &= \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{1 + \exp\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right)}\\ &= \Lambda\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right),\end{aligned} \tag{A2}$$

where $\Lambda(^\circ)$ is the standard logistic cumulative distribution function with $\boldsymbol{\beta}_j$ as the coefficient in the regular binary logit model with a choice between j and the base category. Thus, conditioned on the choice being between j or h, the probability of observing an outcome that is equal to j follows a standard logistic model. To compute the marginal effects, we differentiate with respect to the predictor x_{ik} and get the familiar result from the regular logit model:

$$\begin{aligned}ME_k &= \frac{\partial \pi_{ij}}{\partial x_{ik}}\\ &= \frac{\partial \Lambda\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right)}{\partial x_{ik}}\\ &= \lambda\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right)\beta_{jk},\end{aligned} \tag{A3}$$

where $\lambda(°)$ is logistic probability density function. Since $\lambda\left(\mathbf{x}_i'\boldsymbol{\beta}_j\right) > 0$, the sign of the marginal effect is determined by the sign of β_{jk}. This shows that if we focus solely on two alternatives in the MLM, we can read the sign of the marginal effect from the coefficient. The coefficients can therefore be given an interpretation similar to the one in the logit model, *if* we restrict the choice to be between two categories. Note that this does not mean that the coefficient tells us about the direction of the relationship between a predictor and the probability of a specific choice. Instead, if we wish to interpret our results in this manner, we need to drop the imposed initial restriction. This operation brings us back to the result in Equation 6 where the sign of an estimated coefficient does not tell us about the direction of the relationship between a predictor and the probability of a specific choice. For a more detailed explanation, see Cameron and Trivedi (2005), Wooldridge (2010), or Green (2003).

Author Note

This paper has also benefited from suggestions from Morten Berg Jensen, Linda Ejlskov Jeppesen, Anders Ryom Villadsen, Jesper Rosenberg Hansen, Marianne Simonsen, Niels Peter Mols, and participants at the 2014 AOM conference.

Acknowledgments

I am grateful for the financial support funded by Ferdinand Sallings Mindefond and Solar Fonden used for collecting parts of the data used in the study. I would like to give a special thanks to associate editor Brian Boyd and the anonymous reviewers for their constructive feedback.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. I am grateful to an anonymous reviewer for suggesting this intuitive summary of the issues complicating coefficient interpretation in a multinomial logit model (MLM).
2. Another way of analyzing results from an MLM is the odds ratio, namely, the ratio of the choice probabilities for two distinct alternatives. Odds ratios have nice properties, for instance they neither depend on the level of the predictor nor do they depend on any other predictor in the model. However, their interpretation is less intuitive, and many people have difficulties grasping the concept of odds. A discussion of the use of odds ratios can be found in, for example, Long and Freese (2006).
3. Changes for discrete independent variables are conventionally calculated by letting the predictor vary between two values while holding all of the remaining predictors constant at their mean values. For a dummy variable x_k, the discrete change in predicted probabilities is given by $\frac{\Delta p_{ij}}{\Delta \mathbf{x}_{ik}} = \Pr(y = m|\bar{\mathbf{x}}, x_{ik} = 1) - \Pr(y = m|\bar{\mathbf{x}}, x_{ik} = 0)$. Long and Freese (2006) suggest computing an average absolute discrete change measure that summarizes the J discrete change coefficients for a predictor $x_k : \bar{\Delta} = \frac{1}{J}\sum_{j=1}^{J}\frac{\Delta \Pr(y=j|\bar{\mathbf{x}})}{\Delta x_{ik}}$
4. Alternatively, it may sometimes make more sense to hold other model variables at other values. For instance, dummy variables may be held at the mode while other variables may be held at their median value.
5. Usable response rates in past survey-based entry mode studies are typically of this size. A couple of examples include 7.5 % (Dikova & van Witteloostuijn, 2007), 8.9 % (Schwens et al., 2011) and 13.3 % (Brouthers, Brouthers, & Werner, 2003).

6. The multinomial logit model relies on the assumption of independence of irrelevant alternatives (IIA). In the terms of the model, this means that $\frac{\Pr(y=m|x_i)}{\Pr(y=n|x_i)} = \exp\{\mathbf{x}_i(\beta_{m|b} - \beta_{n|b})\}$,where the odds are not dependent on other available alternatives. Therefore, the odds of one choice versus an alternative choice are not dependent on the number of choice alternatives included. To test the IIA assumption, one may perform the Hausman-Mcfadden test (1984) and the Small-Hsiao test (1985). The Small-Hsiao test randomly divides the data into subsamples why the results will change with each computation. To ensure that results can be replicated, the seed of the random-number generator must be set (in this example the seed is set to 1000). These tests have shown poor properties in small samples and often show conflicting results under certain data structures (Cheng & Long, 2007), but they are still the most common tests available to test the assumption of IIA and recommended in parts of the literature to help choose between the MLM and the multinomial probit model (Bowen & Wiersema, 2004). The mentioned tests compare the estimated coefficients from the null model to the estimated coefficients from a restricted model where one or more of the alternatives are excluded. Thus, a significant test statistic rejects the assumption of IIA. In the example used in this study, both tests were insignificant, thus the assumption of IIA could not be rejected.
7. The individual z-statistic and p-value for an independent variable in a given choice equation can be used to assess whether that particular variable is significant in determining the probability of that particular choice. However, an LR test of an individual variable tests the significance of that variable as a whole by comparing the maximum likelihood value of the full model l_0 to the maximum likelihood value of the model excluding that particular variable l_1. The LR test statistic follows a χ2-distribution with J degrees of freedom: $LR = 2(l_1 - l_0) \sim \chi^2$.
8. I am grateful to an anonymous reviewer for noticing the potential for the graphs to be misinterpreted.
9. I am indebted to two anonymous reviewers for the valuable suggestion of graphing the marginal effects collectively.

References

Agarwal, S., & Ramaswami, S. N. (1992). Choice of foreign-market entry mode—impact of ownership, location and internalization factors. *Journal of International Business Studies*, *23*(1), 1-27.

Anand, J., & Delios, A. (1997). Location specificity and the transferability of downstream assets to foreign subsidiaries. *Journal of International Business Studies*, *28*(3), 579-603.

Bartus, T. (2005). Estimation of marginal effects using margeff. *Stata Journal*, *5*(3), 309-329.

Bowen, H. P., & Wiersema, M. F. (2004). Modeling limited dependent variables: Guidelines for researchers of strategic management. In D. J. Ketchen & D. Bergh (Eds.), *Research methodology in strategy and management* (Vol. *1*, pp. 87-134). Oxford, UK: JAI/Elsevier Press.

Brouthers, K. D. (2002). Institutional, cultural and transaction cost influences on entry mode choice and performance. *Journal of International Business Studies*, *33*(2), 203-221.

Brouthers, K. D., & Brouthers, L. E. (2000). Acquisition or greenfield start-up? Institutional, cultural and transaction cost influences. *Strategic Management Journal*, *21*(1), 89-97.

Brouthers, K. D., & Brouthers, L. E. (2003). Why service and manufacturing entry mode choices differ: The influence of transaction cost factors, risk and trust. *Journal of Management Studies*, *40*(5), 1179-1204.

Brouthers, K. D., Brouthers, L. E., & Werner, S. (2003). Transaction cost-enhanced entry mode choices and firm performance. *Strategic Management Journal*, *24*(12), 1239-1248.

Brouthers, K. D., Brouthers, L. E., & Werner, S. (2008a). Real options, international entry mode choice and performance. *Journal of Management Studies*, *45*(5), 936-960.

Brouthers, K. D., Brouthers, L. E., & Werner, S. (2008b). Resource-based advantages in an international context. *Journal of Management*, *34*(2), 189-217.

Brouthers, K. D., & Dikova, D. (2010). Acquisitions and real options: The greenfield alternative. *Journal of Management Studies*, *47*(6), 1048-1071.

Brouthers, K. D., & Hennart, J. (2007). Boundaries of the firm: Insights from international entry mode research. *Journal of Management*, *33*(3), 395-425.

Brouthers, K. D., & Nakos, G. (2004). SME entry mode choice and performance: A transaction cost perspective. *Entrepreneurship-Theory and Practice*, *28*(3), 229-247.

Buckley, P. J., & Casson, M. C. (1998). Models of the multinational enterprise. *Journal of International Business Studies*, *29*(1), 21-44.

Cameron, A. C., & Trivedi, K. P. (2005). *Microeconometrics: Methods and applications*. New York, NY: Cambridge University Press.

Canabal, A., & White, G. O. (2008). Entry mode research: Past and future. *International Business Review*, *17*(3), 267-284.

Chang, S., Chung, J., & Moon, J. J. (2013). When do wholly owned subsidiaries perform better than joint ventures? *Strategic Management Journal*, *34*(3), 317-337.

Chen, J. J., & Dimou, I. (2005). Expansion strategy of international hotel firms. *Journal of Business Research*, *58*(12), 1730-1740.

Cheng, S., & Long, J. S. (2007). Testing for IIA in the multinomial logit model. *Sociological Methods & Research*, *35*(4), 583-600.

Combs, J. G., Ketchen, J., & David, J. (1999). Explaining interfirm cooperation and performance: Toward a reconciliation of predictions from the resource-based view and organizational economics. *Strategic Management Journal*, *20*(9), 867-888.

Dikova, D., & van Witteloostuijn, A. (2007). Foreign direct investment mode choice: Entry and establishment modes in transition economies. *Journal of International Business Studies*, *38*(6), 1013-1033.

Erramilli, M. K., & Rao, C. P. (1990). Choice of foreign market entry modes by service firms: Role of market knowledge. *Management International Review*, *30*(2), 135-150.

Gatignon, H., & Anderson, E. (1988). The multinational corporation's degree of control over foreign subsidiaries: An empirical test of a transaction cost explanation. *Journal of Law, Economics, & Organization*, *4*(2), 305-336.

Gauvin, J. (2012). *A quick look at the margins command*. Unpublished manuscript.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hausman, J., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, *52*(5), 1219-1240.

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

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Kim, W. C., & Hwang, P. (1992). Global strategy and multinationals' entry mode choice. *Journal of International Business Studies*, *23*(1), 29-53.

Klein, S., Frazier, G. L., & Roth, V. J. (1990). A transaction cost analysis model of channel integration in international markets. *Journal of Marketing Research*, *27*(2), 196-208.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, *19*(3),

Li, J., & Li, Y. (2010). Flexibility versus commitment: MNEs' ownership strategy in china. *Journal of International Business Studies*, *41*(9), 1550-1571.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables (advanced quantitative techniques in the social sciences)*. Thousand Oaks, CA: Sage.

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using stata* (2nd ed.). College Station, TX: Stata Press.

Martin, X. (2013). Solving theoretical and empirical conundrums in international strategy research: Linking foreign entry mode choices and performance. *Journal of International Business Studies*, *44*(1), 28-41.
Martin, X., & Salomon, R. (2003). Knowledge transfer capacity and its implications for the theory of the multinational corporation. *Journal of International Business Studies*, *34*(4), 356-373.
Meyer, K. E. (2001). Institutions, transaction costs, and entry mode choice in Eastern Europe. *Journal of International Business Studies*, *32*(2), 357-367.
Meyer, K. E., Estrin, S., Bhaumik, S. K., & Peng, M. W. (2009). Institutions, resources, and entry strategies in emerging economies. *Strategic Management Journal*, *30*(1), 61-80.
Nakos, G., & Brouthers, K. D. (2002). Entry mode choice of SMEs in central and Eastern Europe. *Entrepreneurship Theory and Practice*, *27*(1), 47-63.
Raftery, A. E. (1996). Baysian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111-163). Oxford, UK: Blackwell.
Sampson, R. C. (2004). The cost of misaligned governance in R&D alliances. *Journal of Law, Economics, & Organization*, *20*(2), 484-526.
Schwens, C., Eiche, J., & Kabst, R. (2011). The moderating impact of informal institutional distance and formal institutional risk on SME entry mode choice. *Journal of Management Studies*, *48*(2), 330-351.
Shrader, R. (2001). Collaboration and performance in foreign markets: The case of young high-technology manufacturing firms. *Academy of Management Journal*, *44*(1), 45-60.
Small, K. A., & Hsiao, C. (1985). Multinomial logit specification tests. *International Economic Review*, *26*(3), 619-627.
Uhlenbruck, K. (2004). Developing acquired foreign subsidiaries: The experience of MNEs in transition economies. *Journal of International Business Studies*, *35*(2), 109-123.
Wei, Y., Liu, B., & Liu, X. (2005). Entry modes of foreign direct investment in china: A multinomial logit approach. *Journal of Business Research*, *58*(11), 1495-1505.
Wiersema, M. F., & Bowen, H. P. (2009). The use of limited dependent variable techniques in strategy research: Issues and methods. *Strategic Management Journal*, *30*(6), 679-692.
Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, *12*(2), 308-331.
Williamson, O. E. (1985). *The economic institutions of capitalism*. New York, NY: Free Press.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge MA: The MIT Press.
Zelner, B. A. (2009). Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal*, *30*(12), 1335-1348.

Author Biography

Jesper N. Wulff is an assistant professor of management science at the Department of Economics and Business, Aarhus University. His research interests include discrete choice modeling, applied econometrics, and foreign market entry. He has delivered presentations at and acted as a reviewer for the Annual Academy of Management meetings.