

## APPENDIX D

### **Questions of Statistical Analysis and Discrete Choice Models**

In discrete choice models, the dependent variable assumes categorical values. The models are binary if the dependent variable assumes only two values. Framing deterrence outcomes, for example, in terms of success and failure provides a typical case. In this situation, a simple binomial logit or probit model is required for estimations. Most studies of international conflict use this simple approach by examining whether war is likely to occur or not, or whether international disputes are likely to escalate or not, and so forth. The examples are too numerous to cite them all. In this book, too, tables 5.3 and 7.1 present the models of the onset of extended-immediate deterrence, where the dependent variable was coded one (1) if general deterrence failure led to the rise of extended-immediate deterrence crisis, and zero (0) if it did not. For this reason, binomial logit models were the appropriate method for estimations.

If the outcome of deterrence is diversified to include multiple possibilities, as it is in this book (i.e., either side's acquiescence, compromise, or war), then the polychotomous nature of the dependent variable requires more complex models for estimation. Since foreign policy choices often involve more than two options or outcomes, the polychotomous dependent variable should be an obvious choice for a number of studies. Still, we rarely find it utilized in the international relations literature partly due to the fact that the discrete choice models are still in the developing stage. Nevertheless, econometric research has progressed in the last two decades to give us several options when deciding on what models to use if our dependent variable takes on more than two discrete values.<sup>1</sup> In this appendix, I provide a brief survey of such models to facilitate a better understanding of the model estimations used in the book. The survey may also help clarify some methodological issues involved in modeling decisions, which could be useful for discrete choice models in future works.

*Ordered Probit.* Multinomial logit is the method of choice for this book since the outcome of deterrence is a polychotomous discrete variable comprising four possible categories (AcqDef, AcqCh, Compromise, War). If these choice categories are assumed to be ordinal outcomes, ranking from low to high according to some of their dimensions, then ordered probit or logit would be a convenient method. Ordered probit is the most frequently used method in international conflict studies when the dependent variable is easily measured on an ordinal scale. For example, an analysis of conflict escalation from low intensity to high intensity requires an ordinal probit or logit analysis (e.g., Rousseau et al. 1996). If we measure deterrence outcomes in a somewhat simplified manner, such as that utilized by Huth and Russett (1988) with their distinction between success, stalemate, and failure, the easily discernible ordinal scale for the outcomes would also warrant the use of ordered probit.

For this analysis, however, it is difficult to rank the four outcomes according to a single criterion. As an illustration, if the degree of a disputant's resolve, the Challenger's for example, is selected as the measure for ordering the outcomes, then the ordered values that range from low (0) to high (3) reflect increasing levels of Challenger's resolve. It is fairly easy to group AcqCh and Compromise at the lower end of such a scale as both outcomes indicate a less resolved Challenger, though compromise clearly indicates a relatively stronger resolve. However, the rank order between AcqDef and War can be disputed since it is a function of both the Challenger's and the Defender's resolve. A highly resolute Challenger can face an equally resolute Defender, in which case War is assigned the highest ordinal value (i.e., AcqCh = 0, Compromise = 1, AcqDef = 2, War = 3). On the other hand, a highly resolute Challenger can face a less resolved Defender, in which case AcqDef is assigned the highest ordinal value (i.e., AcqCh = 0, Compromise = 1, War = 2, AcqDef = 3). As this example shows, assigning an order to the dependent variable has theoretical as well as methodological implications.

For exploratory purposes, I ran ordered probit estimations based on both ordinal scales for all models in each chapter. The results showed a significant sensitivity to different orderings of outcomes. This suggests that the empirical results should be interpreted as a function of both the explanatory variables and the particular ordinal measures used for the dependent variable. Namely, if the ordered probit test shows statistical significance for the rank ordering of choice cate-

gories, interpretation of the independent variables is also conditional upon the assumptions implicit in the rank ordering of the dependent variable.<sup>2</sup> For this reason, the choice of one ordinal scale over another when measuring the same dependent variable across different theoretical models needs to be theoretically justified. In the models used in this book, there is no theoretical rationale for a particular ordering of the deterrence outcomes.

It is clear, therefore, that ordered probit analysis is not the optimal choice for the analysis here. The theoretical argument in this book focuses only on the *independent* effects of explanatory variables on deterrence, not assuming any intrinsic ordered nature of deterrence outcomes. If only one explanatory model had been used, which also implied the assumptions about the ordered nature of outcomes, the use of ordered probit would not be questionable. This book explores multiple explanatory frameworks across different chapters, which do not imply any issue related to the ordered nature of deterrence outcomes. Therefore, only the modeling options that treat deterrence outcomes as unordered discrete categories (i.e., nominal variables) would be appropriate. In this respect, there are several estimation options, and multinomial logit (MNL) was deemed to be the most appropriate choice. After presenting the MNL model, I will discuss the basis for its selection over other alternative models for unordered discrete choices.

*Multinomial Logit.* In a multinomial logit estimation, the coefficients indicate the predicted marginal effects of each explanatory variable ( $x_i$ ) on the log-odds ratio between two outcomes in each possible pair of deterrence outcomes. The log-odds ratio is computed as follows:

$$\log \left[ \frac{\text{Prob}(y = j)}{\text{Prob}(y = 0)} \right] = \sum_{k=1}^J \beta_k x_i$$

This expression means that the MNL coefficients indicate the effects of an explanatory variable ( $x_i$ ) on the log-odds ratio of the probabilities of choosing alternative  $J$  relative to the baseline alternative (outcome)  $K$ .

Note that the MNL model estimates the effects of the independent variables on the likelihood of having one outcome (e.g., compromise) versus another outcome (e.g., war), requiring separate estimates for each pair of outcomes. As I examine four outcomes (AcqDef, AcqCh, Compromise, War), the coefficients are estimated for six possible pairs

of outcomes. For example, table 4.3 in chapter 4 gives the log-odds coefficients for the effects of the power variables for all possible pairs of outcomes: AcqCh vs. AcqDef, Compromise vs. AcqDef, War vs. AcqDef, Compromise vs. AcqCh, War vs. AcqCh, Compromise vs. War.

Since they indicate the log-odds ratios, these coefficients are useful for indicating the sign and statistical significance of parameter estimates, but they are not intuitively appealing for estimating the substantive significance of parameter estimates. In fact, all discrete choice models, whether ordinal probit, nested logit, or any other, share the property that their parameter estimates cannot be interpreted straightforwardly in terms of choice probabilities. Instead, additional computations are required.<sup>3</sup> In the case of the MNL estimates, to assess their substantive significance, the probability of having a deterrence outcome  $J$  rather than outcome  $K$  (a base category), given a set of values in the explanatory variables ( $x_i$ ), is calculated as follows:

$$\Pr(Y = J) = \frac{\exp(\beta_J \chi_i)}{1 + \sum_{k=1}^J \exp(\beta_k \chi_i)} \quad (1)$$

Given  $J + 1$  outcomes,  $J$  number of equations are estimated to show the effects of the explanatory variables on producing outcome  $J$  rather than the baseline outcome  $K$ . The probability for choosing the base case is

$$\Pr(Y = 0) = \frac{1}{1 + \sum_{k=1}^J \exp(\beta_k \chi_i)} \quad (2)$$

The sum of predicted probabilities for all four deterrence outcomes should sum to a total of 100 percent. In table 4.4, for instance, the relative probability  $p$  for each outcome occurring, given the same set of values for the independent variable (e.g., power parity), is as follows:  $\Pr(\text{AcqDef}) = 18.31$  percent,  $\Pr(\text{AcqCh}) = 51.94$  percent,  $\Pr(\text{Compromise}) = 18.40$  percent,  $\Pr(\text{War}) = 11.35$  percent. The sum of these four probabilities is 100 percent. These MNL estimating procedures were used for the empirical analyses in other chapters as well.

There are several advantages to the MNL model, and three in par-

ticular are relevant for this analysis. First, like all other discrete choice estimations, MNL is a nonlinear probability model, based on maximum likelihood (ML) estimation. Unlike linear probability models, most commonly estimated by ordinary least squares (OLS), discrete choice models do not assume that each unit change in the independent variable produces a fixed change in the dependent variable. For example, the effect of the power variable on deterrence outcomes may be greater if two sides are relatively equal rather than vastly unequal. Second, besides this common nonlinearity assumption of all discrete choice models, MNL has the unique property of allowing that the relationship between the independent variables and each deterrence outcome does not necessarily have the same functional form. For instance, the use of MNL allowed us to see that the change in the distribution of military capabilities or GNP from the Defender's advantage to the Challenger's superiority, transiting through parity, does not yield uniform results. The effects are positive on the likelihood of AcqDef or War, negative on the likelihood of AcqCh, and there is a curvilinear relationship with the probability of compromise. Most other discrete choice models would not provide such easily discernible patterns of nonuniform functional form for the relationship between the explanatory factors and each of the four outcomes. The final advantage of the multinomial model is that it does not require that any assumptions be made about the ordered nature of deterrence outcomes. As already discussed, such an assumption should be handled with care, especially if the sensitivity to different rank orderings is observed when analyzing a variety of explanatory models. The results would be less biased and more consistent across different theoretical models if no such assumption is made (see also Stam 1996).

The only disadvantage of the MNL model lies in its property known as the independence of irrelevant alternatives. The independence of irrelevant alternatives (IIA) problem results from the MNL assumption that the relative probability of choosing between two alternatives is unaffected by the presence of additional alternatives. However, the IIA property does not always have to bias the parameter estimates. Whether it does simply depends on the nature of the data at hand as well as on the theoretical model. The Hausman test, as proposed by Hausman and McFadden (1984), is the standard procedure to test whether the IIA property of the MNL model is problematic for a particular empirical analysis (see Greene 1997; Long 1997; Kennedy 1998). It compares the estimators of the same parameters in two slightly different models. One model is unrestricted, comprising all

choice categories of the dependent variable (i.e., alternatives); the other model is restricted, one or more alternatives being excluded from the original unrestricted model. If the IIA assumption is not significantly violated in the data, then excluding one or more outcome choices from the model (or, alternatively, including them in the model) should not lead to a substantial change in the parameter estimates.

To test for the IIA restriction of the MNL models in this book, the Hausman specification test was run for all models and all possible combinations of alternative eliminations. The results consistently showed that IIA was *not* violated, and the inferences made from the MNL models should not be suspect.<sup>4</sup> In other words, these results indicate that none of the four deterrence outcomes can be reasonably substituted with another outcome since they are all dissimilar. As Amemiya suggests in his influential study concerning the use of MNL model, “so far as the alternatives are dissimilar, it [multinomial logit] is the most useful multi-response model, as well as being the most frequently used one” (1981, 1517). We can safely conclude, then, that the Hausman test provides strong evidence that the use of MNL estimation was not problematic, and was even more appropriate, in this analysis. Although the MNL model is fairly frequently used in econometrics, unfortunately we find it only rarely used in the international relations literature (e.g., Stam 1996; Bennett and Stam 1998). Since it is reasonable to frame most foreign policy choices as nominal (i.e., unordered discrete) variables, such a slim record of using MNL or other polychotomous discrete choice estimations is quite alarming.

If the Hausman tests had failed, however, the only solution would have been to choose one of the alternative polychotomous discrete choice models that do not contain the IIA property. To understand why these models are not a suitable choice for the analysis in this book, besides the fact that the Hausman tests did not fail to warrant their use as modeling alternatives to MNL, I will briefly outline their basic principles, advantages, and shortcomings.

*Conditional Logit.* The conditional logit model also assumes polychotomous unordered choices. Nevertheless, it is inappropriate for this analysis since it is a choice-specific model. Namely, explanatory variables are considered to vary for different choices, unlike MNL where they are invariant to the choice categories. As an illustration, in multinomial logit, the power variable does not vary according to the properties of the alternatives available to decision makers (e.g., one’s acquiescence or compromise). In the conditional logit model, however, the explanatory variables,

such as relative power and other factors, are assumed to be choice-specific, taking on different values dependent on whether the decision makers consider to acquiesce, compromise, or fight.

Since the conditional logit also maintains the IIA assumption, the selection between multinomial and conditional logit models should depend primarily on whether explanatory variables are framed as choice- or individual-specific in the theoretical model. Conditional logit would be a convenient method if we were to explore, for instance, how much the costliness of actions in terms of the involved use of force affects a decision maker's choice between different courses of actions. On the other hand, it is clear that the factors of interest in this analysis, such as relative power, regional interests, or domestic regime type, vary only across states, but not across different choices that states have to make. The use of conditional logit is, therefore, inappropriate for the models tested here.

*Nested Logit.* If a multinomial logit model fails the independence of irrelevant alternatives test, one must use an alternative model that permits the *interdependence* of alternatives. Multinomial probit and nested logit are considered as main alternatives to MNL when the IIA property is violated in the data.<sup>5</sup> To specify the nested logit model, originally developed by McFadden (1981), we need to partition the choice set into branches, assuming a similarity of choices within each branch or subset of choices. Most commonly, the partitioning is sequential, dividing the choice set into branches as if the outcome categories represented different stages in the decision-making process. In the context of this analysis, deterrence outcomes such as compromise or one side's acquiescence do not necessarily represent different stages in a sequence of decisions. Thus, the use of nested logit should be ruled out. Moreover, a number of studies report that different specifications of the tree structure, holding all other factors equal, lead to different results. It is then critical to point out that the decision of how choices are assumed to be nested can bias the estimators. To date, there is no rigorous procedure for testing the statistical validity of tree structures, further compounding the problem of incorporating the sequential nature of outcomes into the analysis.

*Multinomial Probit.* The multinomial probit model allows for IIA to be violated in data, but its computational complexity often limits its use. The model requires an estimation of multiple integrals across separate bivariate normal distributions, which are difficult to compute,

especially when there are more than three choices (Greene 1997, 912). Also, although the MNP eliminates the IIA problem, it is not free from related restrictive specifications. Namely, it assumes that the errors across choices can be correlated, but computational problems arise if these correlations are left completely unrestricted. The computation of multiple integrals can, in turn, become unfeasible due to the convergence problems. For this reason, it is expected that the values for the correlations between error terms be specified in order to make multinomial probit computation possible. Note that in the multinomial logit model, the correlation between error terms is assumed to be zero. It is precisely this restriction that gives rise to the IIA problem. In the multinomial probit estimation, we need to specify the values (other than zero) for the correlations in order to incorporate them in the model. Substantively, this means that we need to have a theory that will inform us concerning the numerical correlation between the outcome categories. With the example at hand, this requirement expects us to answer the question of how each deterrence outcome is related to another outcome. We also must decide if the correlations between each pair of outcomes are equal and express them numerically. Such decisions are assumed to be driven by theory, even though such theory rarely exists. The imposition of the restrictions on correlations, therefore, is most likely to bias the resulting estimators. In other words, multinomial probit avoids the assumption that the correlations across terms have zero value, but still requires the researcher to specify their value in order to make the estimations feasible, which is ultimately an arbitrary choice. This again brings us back to the conclusion that the multinomial logit model is the most appropriate for testing the theory in this book.