

# Discrete Choice Methods

Garrett Glasgow and R. Michael Alvarez

## 1 Introduction

A great many political science research questions involve the choices of political agents. These questions arise in most substantive areas of political science. For instance, does a citizen choose to vote in an election or abstain? Does a legislator choose to seek reelection, run for higher office, or retire? Does a country choose to go to war or stay at peace? These kinds of choice situations are known as *discrete choice* problems, so named because political agents are making choices from among a finite set of distinct options. In this chapter we examine the statistical models commonly used by political scientists to study discrete choices.

The first part of this chapter focuses on the “basic” discrete choice models, and is organized around the theoretical choice situations that lead to these models. The first choice situation we consider is the *ordered choice* situation, which assumes that the political agent’s choice is determined by some continuous underlying rating of a political item or a propensity to engage in some behavior known as a *latent variable*. We do not observe this latent variable, instead we observe some discrete choice categories related to the latent variable. We begin with the special case of two ordered categories, and show how this leads to the *probit* and *logit* models. For instance, a country might have an underlying propensity to fight with its neighbors, but we only observe two discrete choice categories — war or peace. We then extend this to the case of multiple ordered choice categories, and show how this leads to the *ordered probit* and *ordered logit* models. For instance, an individual might have some underlying level of agreement with a policy position, but we only observe five ordered discrete choice categories offered by a survey question — agree strongly, agree, neither agree nor disagree, disagree, and disagree strongly.

The second choice situation we consider is the *unordered choice* situation. This choice problem generally involves a political agent selecting a political item from among a discrete set of items that have no intrinsic ordering (the choice set). Again, we begin with a special case, in this instance selecting a political item from a choice set of two items, and show that this again leads to the *probit* and *logit* models. An individual choosing which political candidate to vote for in a two candidate election is an example of this type of choice problem. We then extend this to the case of selecting a political item from a larger choice set, and show how this leads to the *multinomial logit* and *conditional logit* models. An example of a multinomial choice situation is a legislator choosing whether to seek reelection, run for higher office, or retire.

After introducing the basic discrete choice models we consider the estimation and interpretation of these models. We then examine two extensions of the basic discrete choice models commonly seen in political science research — models allowing for heteroskedasticity in the choices made across political agents (such as the *heteroskedastic probit*), and models that estimate substitution patterns across choice alternatives (such as the *multinomial probit* and *mixed logit*). In our conclusion we offer suggestions for further reading.

## 2 Ordered Choices

### 2.1 Ordered Choices with Two Categories

We begin by deriving a discrete choice model for the special case of two ordered choices. Examples of this kind of choice problem are an individual choosing to vote for or against a particular political candidate, a country choosing to go to war or stay at peace, and so on. We think of these binary choice categories as ordered because they represent different degrees of feeling (increasing approval leads the individual to vote for the candidate) or propensities to undertake political action (increasing bellicosity leads the country to go to war).

In these cases we generally assume that a political agent's propensity to undertake some action versus do nothing, provide a favorable versus unfavorable evaluation of some political candidate, or some other ordered binary choice is determined by a latent variable  $y^*$ , with a threshold dividing this latent variable into two choice categories. We do not observe the agent's position on this latent variable, instead observing  $y$ , indicating the choice category the political agent selected. We observe  $y_i = 0$  for agents who are at or below  $\tau$ , the threshold between choice categories, and  $y_i = 1$  for agents above the threshold.

We derive a binary discrete choice model in this case by assuming the latent variable  $y_i^*$  is a linear function of a set of some observed variables and coefficients that influence the positions of political agents on the latent variable ( $X_i\beta$ ) and some unobserved stochastic influences ( $\varepsilon_i$ ). The latent choice preference is given by:

$$y_i^* = X_i\beta + \varepsilon_i \tag{1}$$

Obviously, since  $y_i^*$  is unobserved we cannot estimate this equation with OLS. However, we do observe some information about  $y_i^*$  that will allow us to determine how the observed characteristics  $X_i$  influence choice behavior, because we observe  $y_i$ , the choice that agents make, and thus whether they are above the threshold  $\tau$ . From this we estimate the probability that political agents with particular values of  $X_i$  will fall into particular choice categories.

Since  $y_i^*$  is latent, we require identifying assumptions in order to estimate the probability that a political agent makes a particular choice. We begin by noting that we cannot obtain a unique estimate for both  $\tau$  and any constant term ( $\beta_0$ ) in the model (since  $y_i^*$  is unobserved, we could add an arbitrary constant to both  $\tau$  and  $\beta_0$  and still generate the same probability of an observed outcome). A standard identifying assumption is to assume  $\tau = 0$ . This identifying assumption is arbitrary, and does not influence the estimate of  $\beta$ . With this assumption the probability that agents select the choice category represented by  $y_i = 1$

given the observed variables  $X_i$  is:

$$Pr(y_i = 1|X_i) = Pr(y_i^* > 0|X_i) \quad (2)$$

or, realizing that  $y_i^* = X_i\beta + \varepsilon_i$ , we can rewrite Equation 2 as:

$$Pr(y_i = 1|X_i) = Pr(X_i\beta + \varepsilon_i > 0|X_i) \quad (3)$$

It is clear from Equation 3 that the probability that political agents select the choice category denoted by  $y_i = 1$  will depend in part on the as-yet undefined error term  $\varepsilon_i$ . Thus, in order to estimate this probability, we must make some additional assumptions about the distribution of  $\varepsilon_i$ . Start by rearranging Equation 3 to read:

$$Pr(y_i = 1|X_i) = Pr(\varepsilon_i > -X_i\beta|X_i) \quad (4)$$

If we assume that the distribution of  $\varepsilon_i$  is symmetric, we can further rearrange this equation to read:

$$Pr(y_i = 1|X_i) = Pr(\varepsilon_i \leq X_i\beta|X_i) \quad (5)$$

This equation is simply the evaluation of the cumulative density function (CDF) of  $\varepsilon_i$  at  $X_i\beta$ , which we can express as:

$$Pr(y_i = 1|X_i) = F(X_i\beta) \quad (6)$$

We can now calculate the probability that we will observe  $y_i = 1$  given the values of  $X_i$  by making some assumptions about the distribution of  $\varepsilon_i$ . One possibility is to assume that  $\varepsilon_i$  are drawn from an independently and identically distributed (IID) normal distribution.

This results in the binary *probit* model:

$$Pr(y_i = 1|X_i) = \frac{1}{\sqrt{2\pi}} \int_{\varepsilon=-\infty}^{X_i\beta/\sigma} \phi(\varepsilon)\partial\varepsilon = \Phi\left(\frac{X_i\beta}{\sigma}\right) \quad (7)$$

where  $\phi$  is the PDF of the normal distribution,  $\Phi$  is the CDF of the normal distribution, and  $\sigma$  is the standard deviation of the normal distribution of  $\varepsilon_i$ . As  $y_i^*$  is unobserved, we cannot estimate the variance of  $\varepsilon_i$  as we might in the OLS case. Thus, a standard assumption for binary probit models is that  $\sigma = 1$ .

Another possible assumption about the distribution of  $\varepsilon_i$  is that it follows an IID logistic distribution. This results in the binary *logit* model:

$$Pr(y_i = 1|X_i) = \frac{\exp(\mu X_i\beta)}{1 + \exp(\mu X_i\beta)} \quad (8)$$

where  $\mu$  is a positive scale parameter. As with the binary probit model, we must make an assumption about the variance of this CDF — in this case, we assume  $\mu = 1$ , which is equivalent to assuming that the variance of  $\varepsilon_i$  is  $\frac{\pi^2}{6}$ .

Note that estimating a binary probit model involves estimating an integral, while the binary logit is closed form, requiring no integration to solve. Given the computational capabilities of modern computers, this makes little practical difference in the binary choice case. However, differences in ease of estimation between models based on the normal and logistic distributions do emerge in unordered choice settings, as we explain below.

Estimation of a binary logit model will produce estimates of  $\beta$  that are approximately  $\sqrt{\pi^2/6}$  times bigger than the corresponding estimates from a binary probit model — however, this difference is due solely to the different assumptions made about the variance of  $\varepsilon_i$  in the two models. In practice there is almost never a substantive difference between the binary logit and binary probit models, and either choice is valid in binary choice cases.

## 2.2 Ordered Choice with More than Two Categories

Many political science problems involve political agents making choices from more than two ordered choice categories. An example of this kind of choice problem would be asking an individual to rate their ideology on an ordered scale that runs from liberal to conservative. Again, we can think of choices like these as ordered because they represent different degrees of feeling (increasing conservatism leads the individual to select more conservative points on the survey question scale).

Extending the ordered choice models for two categories to ordered choice settings with more than two choice categories is straightforward. Again, we assume that the intensity of feeling or propensity to undertake some action by political agents is determined by a latent variable  $y_i^*$ . However, in these cases we assume there is more than one threshold on the latent variable, dividing the latent variable into three or more ordered choice categories. We do not observe a political agent's position on this latent variable, but we do observe  $y_i$ , which indicates which choice category a political agent selects. Specifically, we observe:

$$y_i = j \text{ if } \tau_{j-1} \leq y_i^* < \tau_j \text{ for } j = 1 \text{ to } J \quad (9)$$

where  $\tau_0 = -\infty$  and  $\tau_J = \infty$ .

We can derive an ordered discrete choice model in this case in a way similar to the derivation of the ordered binary choice model, by assuming the latent variable is a linear function of a set of some observed variables and coefficients that influence the positions of agents on the latent variable and some unobserved stochastic influences:

$$y_i^* = X_i\beta + \varepsilon_i \quad (10)$$

As in the binary case,  $y_i^*$  is unobserved, so we must make some identifying assumptions in

order to estimate the probability that a political agent makes a particular choice. We cannot obtain unique estimates for all of the thresholds and any constant term ( $\beta_0$ ) in the model (since  $y_i^*$  is unobserved, we could add an arbitrary constant to all of these terms and still generate the same probability of an observed outcome). Standard identifying assumptions include setting  $\tau_1 = 0$  or setting  $\beta_0 = 0$  — as with the binary case, these assumptions are arbitrary, and do not influence the estimate of  $\beta$ . With one of these identifying assumptions we can estimate the probability that political agents with particular values of  $X_i$  will fall into a particular choice category. Under the assumption that  $\beta_0 = 0$ , the probability that agents select the choice category represented by  $y_i = j$  given the observed variables  $X_i$  is given by:

$$Pr(y_i = j|X_i) = Pr(\tau_{j-1} \leq y_i^* < \tau_j|X_i) \quad (11)$$

or substituting  $y_i^* = X_i\beta + \varepsilon_i$  and rearranging,

$$Pr(y_i = j|X_i) = Pr(\tau_{j-1} \leq X_i\beta + \varepsilon_i < \tau_j|X_i) = Pr(\tau_{j-1} - X_i\beta \leq \varepsilon_i < \tau_j - X_i\beta|X_i) \quad (12)$$

As the probability that a random variable will fall between two values is the difference between the CDF at those two values, we can rewrite Equation 12 as:

$$\begin{aligned} Pr(y_i = j|X_i) &= \\ Pr(\varepsilon_i < \tau_j - X_i\beta|X_i) - Pr(\varepsilon_i < \tau_{j-1} - X_i\beta|X_i) &= \\ F(\tau_j - X_i\beta|X_i) - F(\tau_{j-1} - X_i\beta|X_i) & \end{aligned} \quad (13)$$

where  $F$  represents the CDF of  $\varepsilon_i$ . These probabilities can be simplified for the lowest and highest choice categories by realizing that  $F(\tau_0 - X_i\beta) = F(-\infty - X_i\beta) = 0$ , and that  $F(\tau_J - X_i\beta) = F(\infty - X_i\beta) = 1$ . As with the binary discrete choice models, we can now calculate the probability that  $y_i = j$  by making some assumptions about the distribution of

$\varepsilon_i$ . If we make the same assumptions as those made for a binary probit model ( $\varepsilon_i$  is an IID normal distribution with mean zero and unit variance) this leads to an *ordered probit* model. If we make the same assumptions as those made for a binary logit model ( $\varepsilon_i$  is an IID logistic distribution with mean zero and variance  $\frac{\pi^2}{3}$ ), this leads to an *ordered logit* model. As is the case with the binary choice models, there is little to distinguish between the ordered logit and ordered probit models in either ease of estimation or substantive results.

While there is no theoretical maximum to the number of categories in the dependent variable, there is certainly a practical limit beyond which researchers might consider treating the dependent variable as continuous. For instance, the commonly used “thermometer score” survey question is actually an ordered scale with 101 categories, but estimating an ordered logit or probit in this case would be impractical (requiring estimating 100 separate points on the latent variable), and most statistical software places a limit on the number of categories allowed in the dependent variable (for instance, the maximum number of categories allowed in Stata is 50). We are unaware of any agreed upon rule of thumb for a maximum number of ordered categories, although it is rare in practice to see an ordered logit or probit with more than 7-10 categories in the dependent variable.

## 3 Unordered Choices

### 3.1 Unordered Choices for Two Categories

As with our treatment of ordered choices, we begin by deriving a discrete choice model for the special case where there are only two unordered choices. Examples of this kind of choice problem would be an individual choosing which of two political candidates to vote for, a political candidate choosing which of two cities to use as a site for a major campaign speech, and so on. We can think of these binary choice categories as unordered because they represent two distinct items with no intrinsic ordering, in contrast to the binary ordered

choice categories considered earlier (which could be considered two different rating categories for the *same* item).

In the unordered choice case we generally assume that political agents assign some kind of relative rating known as a *utility score* to each political item in the choice set, and then agents select the choice category with the highest utility. This theory of choice behavior is known as *random utility maximization (RUM)*, so named because political agents are selecting the choice category that gives them the highest utility, and these utility scores are unobserved, and thus appear random to us as researchers.

Due to the fact that political agents are considering more than one choice, we must keep track of both the characteristics of the political agent and the characteristics of the different choice alternatives (for instance, the issue positions of the various candidates an individual might choose in an election). This makes the derivation of a binary choice model more complicated than in the ordered choice case. We derive a binary discrete choice model in this case by assuming that an agent's utility scores are composed of a linear function of a set of observed variables and coefficients specific to the political item  $j$  as perceived by agent  $i$  ( $C_{ij}\lambda$ ), a set of observed variables and coefficients specific to the political agent  $i$  ( $X_i\beta_j$ ), and unobserved stochastic influences ( $e_{ij}$ ). For example, the utility yielded by the choice category denoted by  $y = 1$  for agents is given by:

$$U_{i1} = X_i\beta_1 + C_{i1}\lambda + e_{i1} \tag{14}$$

Note that the coefficients for  $X_i$  are subscripted by choice category, which allows the political agent's characteristics to influence their utility for different alternatives in different ways — for instance, an individual's partisan identification might increase the utility for a candidate of the same party, but decrease the utility for a candidate of an opposing party.

We do not observe an agent's utility scores, but we do observe  $y_{ij}$ , which indicates which

choice category the political agent selected. The probability that we will observe agents selecting the choice category represented by  $y_i = 1$  over the choice category represented by  $y_i = 0$  is given by:

$$Pr(y_i = 1|X_i, C_{ij}) = Pr(U_{i1} > U_{i0}) \quad (15)$$

or, realizing that  $U_{ij} = X_i\beta_j + C_{ij}\lambda + e_{ij}$ , we can rewrite Equation 15 as:

$$Pr(y_i = 1|X_i, C_{ij}) = Pr(X_i\beta_1 + C_{i1}\lambda + e_{i1} > X_i\beta_0 + C_{i0}\lambda + e_{i0}) \quad (16)$$

It is clear from Equation 16 that we will not be able to obtain unique estimates for both  $\beta_0$  and  $\beta_1$ , since we could add an arbitrary constant to both of these sets of coefficients and still generate the same probability of an observed outcome. A standard identifying assumption is to set  $\beta_0 = 0$ .

As with our derivation of the binary choice model for ordered alternatives, it is now clear that some assumptions will need to be made about the distribution of the stochastic portion of utility  $e_{ij}$  in order to estimate the probability that agents select the choice category denoted by  $y_i = 1$ . Setting  $\beta_0 = 0$ , we can rearrange Equation 16 to read:

$$Pr(y_i = 1|X_i, C_{ij}) = Pr(X_i\beta_1 + (C_{i1} - C_{i0})\lambda > e_{i0} - e_{i1}) \quad (17)$$

This equation is simply the evaluation of the cumulative density function (CDF) of  $e_{i0} - e_{i1}$  at  $X_i\beta_1 + (C_{i1} - C_{i0})\lambda$ , which we can express as:

$$Pr(y_i = 1|X_i) = F(X_i\beta_1 + (C_{i1} - C_{i0})\lambda) \quad (18)$$

Just as in the ordered choice case, we can now calculate the probability that we will observe  $y_i = 1$  given the values of  $X_i$  and  $C_{ij}$  by making assumptions about the distribution of

$e_{ij}$ . One possibility is to assume that  $e_{ij}$  is an IID normal distribution — the difference of two normal distributions is also a normal distribution, so this assumption (and the assumption that  $\sigma = 1$ ) leads to a probit model similar to that presented in Equation 7:

$$Pr(y_i = 1|X_i) = \Phi(X_i\beta_1 + (C_{i1} - C_{i0})\lambda) \quad (19)$$

Another possibility is to assume that  $e_{ij}$  follows an IID type I extreme value distribution — the difference of two IID type I extreme value distributions is IID logistic, so this assumption (and the assumption that  $\mu = 1$ ) leads to a logit model similar to that presented in Equation 8:

$$Pr(y_i = 1|X_i) = \frac{\exp(X_i\beta + (C_{i1} - C_{i0})\lambda)}{1 + \exp(X_i\beta + (C_{i1} - C_{i0})\lambda)} = \frac{\exp(X_i\beta + C_{i1}\lambda)}{\exp(C_{i0}\lambda) + \exp(X_i\beta + C_{i1}\lambda)} \quad (20)$$

Note that if there are no variables that measure the characteristics of the alternatives in the choice set, the probit and logit models presented here are identical to those presented for binary ordered choices above. Including variables that measure the characteristics of alternatives in both the probit and logit models does not require the researcher to do anything beyond calculating  $C_{i1} - C_{i0}$  and including this term as an independent variable.

### 3.2 Unordered Choice with More than Two Categories

Many political science problems also involve political agents making choices from more than two unordered choice categories. Examples of this kind of choice problem would be an individual choosing which of several political parties to vote for, a political candidate choosing which of several cities to use as a site for a major campaign speech, and so on. Again, we can think of these choice categories as unordered because they represent distinct items with no intrinsic ordering.

Extending the unordered choice models for two categories to unordered choice over multiple categories is straightforward. Again, assume that political agents assign utility scores to each political item in the choice set, and select the choice category that yields the highest utility. We do not observe an agent's utility scores, instead observing  $y_i$ , which indicates which of the  $J$  choice categories the political agent selected. The probability that we observe agents selecting the choice category represented by  $y_i = j$  over all other choice categories is:

$$Pr(y_i = j|X_i, C_{ij}) = Pr(U_{ij} > U_{ik}) \forall k \neq j \quad (21)$$

or, substituting in  $U_{ij} = X_i\beta_j + C_{ij}\lambda + e_{ij}$  and rearranging:

$$Pr(y_i = j|X_i, C_{ij}) = Pr(X_i\beta_j + C_{ij}\lambda - X_i\beta_k - C_{ik}\lambda > e_{ik} - e_{ij}) \forall k \neq j \quad (22)$$

As in the binary choice case, we will not be able to obtain unique estimates for each set of  $\beta$ 's in the model, since we could add an arbitrary constant to these sets of coefficients and still generate the same probability of an observed outcome. A standard identifying assumption is to pick one choice category as a *baseline category*, and set  $\beta = 0$  for that category (the choice of baseline categories is arbitrary and does not affect the substantive interpretation of the model).

Once again, we must make some assumptions about the distribution of the stochastic portion of utility  $e_{ij}$  in order to estimate the probability that agents select the choice category denoted by  $y_i = j$ . Assuming that  $e_{ij}$  is an IID type I extreme-value distribution (and that  $\mu = 1$ ) leads to a *multinomial logit* model:

$$Pr(y_i = j|X_i) = \frac{\exp(X_i\beta_j + C_{ij}\lambda)}{\sum_{k=1}^J \exp(X_i\beta_k + C_{ik}\lambda)} \quad (23)$$

Many political scientists make a distinction between the multinomial logit and the *con-*

*ditional logit*. For these researchers the multinomial logit does not include variables that measure the characteristics of the choice alternatives, while conditional logit does contain these variables. We note here that the only difference between these models is in the kinds of variables included in the model, and many research traditions simply refer to both models as multinomial logit.

Assuming that  $e_{ij}$  follows an IID normal distribution leads to a special case of the *multinomial probit* (MNP) model. The multinomial probit is a much more difficult model to estimate than a multinomial logit — while the multinomial logit has the closed-form expression presented in Equation 23, estimating a multinomial probit involves integrating over a multivariate normal distribution of dimension  $J - 1$ . Thus, the multinomial probit is generally only used in cases where we want to relax the assumption of an IID error term, as we describe below.

## 4 Estimation

The discrete choice models discussed above are not typically estimated via OLS because of the nonlinearity of these choice models. Instead, these models are usually estimated through maximum likelihood estimation (MLE). This is done by first creating a likelihood function, and then taking the natural log of this function to create a log likelihood function (we take the natural log of the likelihood function because it is easier to work with).

For instance, the log likelihood of the probit model is:

$$\ln L = \sum_i [y_i \ln \Phi(X_i\beta)] + [(1 - y_i) \ln [1 - \Phi(X_i\beta)]] \quad (24)$$

Maximizing this log likelihood means maximizing the sum of the predicted probabilities assigned to the choices that political agents in fact made. This is done by estimating  $\beta$  to

have values such that  $X_i\beta$  tends to be large when  $y_i = 1$ , and small when  $y_i = 0$ .

The log likelihood functions for all of the models derived above are globally concave and thus relatively easy to maximize, with most current statistical software packages reaching the solution in seconds. However, note that the consistency, normality, and efficiency of MLE depends on asymptotic arguments — the small sample properties of MLE are largely unknown. Thus, researchers should be cautious when using MLE in small samples.

Many common statistical software packages allow researchers to estimate all of the discrete choice models discussed earlier, with the exception of the conditional logit — to our knowledge only LIMDEP, SAS, and Stata offer this option.

## 5 Interpretation

For all of the discrete choice models discussed above, the relationship between  $X_i\beta$  and the choice probabilities depends on a nonlinear function. Thus, the effect of a 1-unit change in one independent variable on the choice probabilities will depend on the values of the other independent variables in the model. This means that we cannot interpret the  $\beta$ 's in discrete choice models as we would in an OLS setting.

The most common method for interpreting parameter estimates in discrete choice models in political science is through the use of a hypothetical observation. The researcher begins by selecting values for the independent variables in the model that are representative of some case of interest in the study. Setting variables to their mean or modal values is common, although this is not the only possibility — often the researcher will select values for the independent variables that reflect some case of substantive interest. Then, using the estimated values of  $\beta$  and the selected values of  $X$  a baseline probability for  $y = 1$  is calculated. For

instance, we would interpret the coefficients in a probit model by calculating:

$$\hat{Pr}(y = 1|X) = \Phi(\bar{X}\hat{\beta}) \tag{25}$$

where  $\bar{X}$  represents the selected values for the independent variables. The influence of a particular independent variable on the probability that  $y = 1$  can then be examined by changing the value of this variable while holding all other independent variables constant at  $\bar{X}$  and recalculating  $\hat{Pr}(y = 1|X)$ .

These predicted probabilities can be also calculated fairly easily using some built-in commands in most statistical software packages, such as the “mfx compute” command in Stata, and can even be calculated with a hand calculator in the case of logit probabilities. CLARIFY (King, Tomz and Wittenberg 2000), is another option, with the added benefit of the ability to calculate standard errors on these predicted probabilities.

## 6 Common Extensions of the Basic Discrete Choice Models

Many extensions to the basic discrete choice models discussed above have been developed for political science research. Below we examine the two most commonly seen extensions in political science research — models to examine heteroskedasticity in choices across political agents, and models to estimate more realistic substitution patterns across choice alternatives.

### 6.1 Heteroskedasticity

In some cases there are reasons to suspect there is heteroskedasticity in choices made across political agents. For instance, Alvarez and Brehm (1995) hypothesized that individuals who were ambivalent about allowing abortion would feel a greater sense of internal conflict when

making discrete choices related to abortion policy, leading their policy preferences to be more susceptible to external forces that could vary from moment to moment, such as priming effects. In an ideal setting we would simply include these external forces as independent variables in the discrete choice model — however, in this case these forces were unobserved (and perhaps unobservable) in the survey available. Thus, Alvarez and Brehm modeled the effect of these external forces as heteroskedasticity across individuals, with greater variance expected in the error term for ambivalent individuals (those who held conflicting opinions about allowing abortion) as compared to unambivalent individuals (those who held one-sided opinions about abortion).

Accounting for heteroskedasticity is important methodologically as well as substantively. Unlike the OLS case, where violations of the assumption of an IID error term lead to inefficient but unbiased parameter estimates, IID violations in discrete choice models can lead to inconsistent parameter estimates.

Since the error term  $\varepsilon_i$  is unobserved, we cannot directly estimate heteroskedasticity in discrete choice models. Instead, we infer heteroskedasticity through the relationship between  $X_i\beta$  and observed choice behavior — as the error variance increases, political agents will be more likely to make choices that differ from those predicted by  $X_i\beta$ . To see how heteroskedasticity is modeled in a discrete choice setting, consider the binary probit, where the probability that we observe  $y_i = 1$  given the values of  $X_i$  is:

$$Pr(y_i = 1|X_i) = \Phi\left(\frac{X_i\beta}{\sigma}\right) \tag{26}$$

where  $\Phi$  is the CDF of the normal distribution and  $\sigma$  is the standard deviation of the normally distributed error term. As we do not observe the error term, we must make an identifying assumption about  $\sigma$  — standard probit models generally assume  $\sigma = 1$ .

The heteroskedastic probit model relaxes this assumption, allowing the variance of the

error term to differ across political agents by specifying the variance of the error term as  $\sigma_i = \exp(Z_i\gamma)$ , where  $Z_i$  is a set of observed variables thought to determine the variance of the error term and  $\gamma$  is a set of coefficients to be estimated (the exponential ensures the estimated variance of the error term remains positive). As long as  $Z_i$  does not contain a constant term this model will be identified. The probability that  $y_i = 1$  is then given by:

$$Pr(y_i = 1|X_i, Z_i) = \Phi\left(\frac{X_i\beta}{\exp(Z_i\gamma)}\right) \quad (27)$$

As the term  $\exp(Z_i\gamma)$  increases, it pushes the predicted choice probabilities closer to 0.5 for either alternative, indicating that political agents are more likely to make choices that differ from those predicted by  $X_i\beta$  — this is what would expect to see if  $\exp(Z_i\gamma)$  measures the variance of the error term. Thus, we interpret a positive coefficient on a variable in  $Z$  as indicating that increasing the value of this variable increases the error variance, while interpreting a negative coefficient as indicating that increasing the value of this variable decreases the error variance.

This heteroskedastic probit model is a straightforward generalization of a simple probit, and can be estimated and interpreted in the same way as the basic discrete choice models. Both Stata and LIMDEP include simple commands that allow for the estimation of heteroskedastic probits. This heteroskedastic probit model can easily be generalized to a heteroskedastic ordered probit model (Alvarez and Brehm 1998). Although less commonly seen, heteroskedastic logits and heteroskedastic ordered logits are specified in much the same way as the heteroskedastic probit models by specifying the logit scale parameter as  $\mu = \frac{1}{\exp(Z_i\gamma)}$  rather than  $\mu = 1$  (see Equation 8). Models of this type also exist for multiple unordered choices (Dubin and Zeng 1991), although heteroskedasticity in these cases is usually addressed using the approach outlined in the next section.

## 6.2 Substitution Patterns

In many unordered choice situations researchers are interested in *substitution patterns*, or how political agents might substitute one choice alternative for another as choice options are added to or removed from the choice set. For instance, in a study of voting behavior in the 1992 U.S. Presidential election Alvarez and Nagler (1995) examined the possibility that voters viewed Perot as a substitute for Bush. While ideally we could observe all of the characteristics of the candidates that lead voters to view them as similar and include them as independent variables in the discrete choice model, in many cases at least some of these characteristics will be unobserved. Thus, in order to estimate more realistic substitution patterns Alvarez and Nagler modeled these common unobserved attributes as correlations between the unobserved portions of utility for each candidate.

As multinomial logit assumes the unobserved portions of utility are IID, it cannot estimate substitution patterns that take this kind of unobserved similarity between choice alternatives into account. Specifically, MNL (and all other unordered discrete choice models that assume the unobserved portions of utility are IID) imposes the *independence of irrelevant alternatives* (IIA) property on the estimated substitution patterns. IIA requires that the relative odds that a political agent selects one choice alternative over another do not change no matter which other alternatives are added to or deleted from the choice set. A simple Hausman test can be used to determine the validity of the IIA assumption in a multinomial logit model (Hausman and McFadden 1984).

It is easy to imagine political choice settings where this is an unrealistic assumption — any choice situation where political agents might view some choice alternatives as substitutes is likely to violate the IIA assumption. In addition to the substantive motivation, addressing correlations in the error term across choice alternatives is also important methodologically, as violations of the assumption of an IID error term can lead to inconsistent parameter estimates (although some suggest that the inconsistency might be slight in many political

choice settings (Dow and Endersby 2004; Quinn, Martin, and Whitford 1999)).

Violations of the IIA assumption are most commonly addressed in political science research through discrete choice models that allow the unobserved portion of utility to be correlated across choice alternatives. Specifically, we assume  $e_{ij} = Z_{ij}\eta_i + \nu_{ij}$ , where  $Z_{ij}$  is an identity matrix of the same dimension as the number of alternatives in the choice set,  $\nu_{ij}$  is an error term that is IID across political agents and alternatives, and  $\eta_i$  is a vector of random terms with mean 0 that varies over political agents and can be correlated across choice alternatives (models that assume IIA implicitly assume  $Z_{ij} = 0$ ). Different assumptions about the distributions of  $\eta_i$  and  $\nu_{ij}$  will lead to different discrete choice models that can estimate realistic substitution patterns.

Assuming that both  $\eta_i$  and  $\nu_{ij}$  are normally distributed leads to the *multinomial probit* model. As the sum of two normal distributions is also normal, the unobserved portion of utility will be normally distributed, leading to the MNP mentioned above. However, as  $\eta_i$  can be correlated across choice alternatives the unobserved portions of utility for each choice alternative will follow a multivariate normal distribution with a general covariance matrix  $\Sigma$ . Estimating the off-diagonal terms in  $\Sigma$  allows the researcher to examine how the unobserved portions of utility are correlated, and thus which choice alternatives are viewed as substitutes by political agents for unobserved reasons. The probability that political agent  $i$  selects choice alternative  $j$  is given by integrating over the multivariate normal distribution of the  $J - 1$  differences in the unobserved portions of utility (where  $J$  is the number of choice alternatives). For instance, if we define  $\varepsilon_{kj} = e_{ik} - e_{ij}$  and  $\bar{U}_j = X_i\beta_j + C_{ij}\lambda$ , the probability that political agent  $i$  selects alternative 1 from a choice set of three alternatives is given by:

$$Pr(y_i = 1 | X_i, C_{ij}) = \int_{\varepsilon_{21}=-\infty}^{\frac{\bar{U}_1 - \bar{U}_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}} \int_{\varepsilon_{31}=-\infty}^{\frac{\bar{U}_1 - \bar{U}_3}{\sqrt{\sigma_1^2 + \sigma_3^2 - 2\sigma_{13}}} \phi(\varepsilon_{21}, \varepsilon_{31}) \partial\varepsilon_{21} \partial\varepsilon_{31} \quad (28)$$

where  $\sigma_j^2$  denotes the variance of the unobserved portion of utility for alternative  $j$  and  $\sigma_{jk}$

denotes the covariance between the unobserved portions of utility for alternatives  $j$  and  $k$ . Note that one element in  $\Sigma$  must be set to a constant to identify the model, which means that only  $[J(J - 1)]/2$  elements in  $\Sigma$  can be estimated.

Assuming that  $\nu_{ij}$  is IID extreme value and that  $\eta_i$  follows a distribution  $g$  specified by the researcher leads to the *mixed logit* (MXL) model. Any number of random elements that follow any distribution can be included in  $\eta_i$ , making mixed logit a more general discrete choice model than MNP — in fact, McFadden and Train (2000) demonstrate that mixed logits can be specified to approximate any discrete choice model derived from random utility maximization with the appropriate choices of  $g$  and  $Z$ . The probability that political agent  $i$  selects choice alternative  $j$  is given by integrating over the general multivariate distribution of the unobserved portions of utility  $g$ :

$$Pr(y_i = j | X_i, C_{ij}) = \int_{\eta} \left( \frac{\exp(X_i \beta_j + C_{ij} \lambda + Z_{ij} \eta_i)}{\sum_{k=1}^J \exp(X_i \beta_k + C_{ik} \lambda + Z_{ik} \eta_i)} \right) g(\eta) \partial \eta \quad (29)$$

This equation shows that the choice probability is a mixture of MNL probabilities, with the weight of each particular MNL probability determined by the mixing distribution  $g$  (hence the term “mixed logit”). As with the multinomial probit, the unobserved portions of utility in a mixed logit can be specified to follow a multivariate distribution with a general covariance matrix, and estimating the off-diagonal terms in this covariance matrix (subject to the same identification restrictions as the MNP) will allow for more realistic substitution patterns across choice alternatives.

Multinomial probits and mixed logits can also be specified to estimate random coefficients. For instance, Glasgow (2001) hypothesized that the influence of working class membership on vote choice in the 1987 British general election would vary across individuals for unobserved reasons (such as the level of social mobility within an individual’s geographic region), and thus specified a mixed logit model for vote choice with random coefficients for the variables

indicating working class membership. Random coefficients in these models are specified by including some variables in both  $X_i$  and  $Z_{ij}$  — in this case the coefficients estimated on the variables in  $X_i$  represent the mean of the random coefficient, while the estimated variances on the variables in  $Z_{ij}$  represent the variances of the random coefficients. MNPs are restricted to  $J - 1$  normally distributed random coefficients, while MXLs can have any number of random coefficients that follow any distribution.

Unlike the closed form of the MNL, estimation of the MNP and MXL models involves the evaluation of multidimensional integrals. Thus, these models are generally estimated using some kind of simulated maximum likelihood technique, which approximates the multidimensional error distribution through a Monte Carlo technique. LIMDEP, SAS, and Stata allow for estimation of multinomial probits, while SAS and LIMDEP allow for estimation of mixed logits.

It should be noted that multinomial probits and mixed logits are not the only way to relax the IIA assumption and estimate more realistic substitution patterns across choice alternatives. *Generalized extreme value* (GEV) models also allow for correlation across choice alternatives. The most common type of GEV model is the *nested logit*, which divides the choice alternatives into subsets, with IIA holding within each subset, but not across subsets. Although more tractable than either the MNP or MXL, and potentially useful in cases where the researcher has some theory or prior knowledge that allows some choice alternatives to be grouped together, GEV models have not seen many applications in political science.

## 7 Further Reading

Obviously this chapter only provides an overview of the use of discrete choice models in political science, and the interested reader should consult other sources for more details and examples of applications. Long (1997) offers a fairly accessible introduction to the

basic discrete choice models discussed here, while more advanced treatments can be found in Ben-Akiva and Lerman (1985) and Train (2003). Jones and Westerland (2006) offer a detailed discussion of the analysis of choices from ordered categories, and present several models not discussed here. For applications of heteroskedastic discrete choice models in political science see the work of Alvarez and Brehm (1995, 1997, 1998). For applications of the multinomial probit to political science problems see Alvarez and Nagler (1995, 1998a, 1998b), Lacy and Burden (1999, 2000), and Schofield, Martin, Quinn, and Whitford (1998). For applications of the mixed logit to political science problems see Glasgow (2001) and Glasgow and Alvarez (2005). Finally, note that there are numerous extensions to the discrete choice methods discussed here that were not covered in this chapter, such as techniques for panel and time-series cross-section (TSCS) data (Hsiao 2003, chapter 7), event history data (Box-Steffensmeier and Jones 2004), censored and truncated dependent variables (Sigelman and Zeng 2000), selection bias (Boehmke 2003), endogeneity in choice models (Alvarez and Glasgow 2000), event counts (King 1989) and models based on distributions other than the normal or logistic (Nagler 1994).

## References

- Alvarez, R. Michael, and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science* 39:1055-1082.
- Alvarez, R. Michael, and John Brehm. 1997. "Are Americans Ambivalent Towards Racial Policies?" *American Journal of Political Science* 41:345-374.
- Alvarez, R. Michael, and John Brehm. 1998. "Speaking in Two Voices: American Equivocation About the Internal Revenue Service." *American Journal of Political Science* 42:418-452.
- Alvarez, R. Michael, and Jonathan Nagler. 1995. "Economics, Issues, and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39:714-144.
- Alvarez, R. Michael, and Jonathan Nagler. 1998a. "Economics, Entitlements, and Social Is-

sues: Voter Choice in the 1996 Presidential Election.” *American Journal of Political Science* 42:1349-1363.

Alvarez, R. Michael, and Jonathan Nagler. 1998b. “When Politics and Models Collide: Estimating Models of Multicandidate Elections.” *American Journal of Political Science* 42:55-96.

Alvarez, R. Michael and Garrett Glasgow, 2000. “Two-Stage Estimation of Non-Recursive Choice Models.” *Political Analysis* 8(2): 147-165.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application*. Cambridge, MA: MIT Press.

Boehmke, Frederick J. 2003. “Using Auxiliary Data to Estimate Selection Bias Models, With an Application to Interest Groups’ Use of the Direct Initiative Process.” *Political Analysis* 11, 234-254.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. New York: Cambridge University Press.

Dow, Jay, and James Endersby. 2004. “Multinomial Logit and Multinomial Probit: A Comparison of Choice Models for Voting Research.” *Electoral Studies* 23:107-122.

Dubin, Jeffrey A., and Langche Zeng. 1991. “The Heterogeneous Logit Model.” Caltech Social Science Working Paper #759.

Glasgow, Garrett. 2001. “Mixed Logit Models for Multiparty Elections.” *Political Analysis* 9:116-136.

Glasgow, Garrett, and R. Michael Alvarez. 2005. “Voting Behavior and the Electoral Context of Government Formation.” *Electoral Studies* 24:245-264.

Hausman, Jerry, and Daniel McFadden. 1984. “Regression-Based Specification Tests for the Multinomial Logit Model.” *Econometrica* 52, 1219-1240.

Hsiao, Cheng. 2003. *Analysis of Panel Data*. New York: Cambridge University Press, second edition.

Jones, Bradford S. and Chad Westerland. 2006. “Order Matters (?): Alternatives to Conventional Practices for Ordinal Categorical Response Variables.” Paper presented at the Annual Meetings of the Midwest Political Science Association.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press, 1989.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. “Making the Most of Statistical

Analyses: Improving Interpretation and Presentation.” *American Journal of Political Science* 44:341-355.

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

McFadden, Daniel, and Kenneth Train. 2000. “Mixed MNL Models for Discrete Response.” *Journal of Applied Econometrics* 15(5), 447-470.

Nagler, Jonathan. 1994. “Scobit: An Alternative Estimator to Logit and Probit.” *American Journal of Political Science* 38: 230-255.

Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. 1999. “Voter Choice in a Multi-Party Democracy: A Test of Competing Theories and Models.” *American Journal of Political Science* 43:1231-1247.

Schofield, Norman, Andrew D. Martin, Kevin M. Quinn, and Andrew B. Whitford. 1998. “Multiparty Electoral Competition in The Netherlands and Germany: A Model Based on Multinomial Probit.” *Public Choice* 97:257-293.

Sigelman, Lee and Langche Zeng, 2000. “Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models.” *Political Analysis* 8(2), 167-182.

Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.