over choice situations for each person. Utility from alternative j in choice situation t by person n is $U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$ with ε_{njt} being iid extreme value over time, people, and alternatives. Consider a sequence of alternatives, one for each time period, $\mathbf{i} = \{i_1, \ldots, i_T\}$. Conditional on β the probability that the person makes this sequence of choices is the product of logit formulas:

$$\mathbf{L}_{n\mathbf{i}}(\beta) = \prod_{t=1}^{T} \left[\frac{e^{\beta_n' x_{ni_t t}}}{\sum_j e^{\beta_n' x_{njt}}} \right] \tag{6.2}$$

since the ε_{njt}'s are independent over time. The unconditional probability is the integral of this product over all values of β:

$$P_{n\mathbf{i}} = \int \mathbf{L}_{n\mathbf{i}}(\beta) f(\beta) d\beta. \tag{6.3}$$

The only difference between a mixed logit with repeated choices and that with only one choice per decision-maker is that the integrand involves a *product* of logit formulas, one for each time period, rather than just one logit formula. The probability is simulated similarly to the probability with one choice period. A draw of β is taken from its distribution. The logit formula is calculated for each period, and the product of these logits is taken. This process is repeated for many draws, and the results are averaged.

Past and future exogenous variables can be added to the utility in a given period to represent lagged response and anticipatory behavior, as described in section (5.5) in relation to probit with panel data. However, unlike probit, lagged dependent variables can be added in a mixed logit model without changing the estimation procedure. Conditional on β_n, the only remaining random terms in the mixed logit are the ε_{nj}'s, which are independent over time. A lagged dependent variable entering U_{njt} is uncorrelated with these remaining error terms for period t, since these terms are independent over time. The conditional probabilities (conditional on β) are therefore the same as in equation (6.2), but with the x's including lagged dependent variables. The unconditional probability is then the integral of this conditional probability over all values of β, which is just equation (6.3). In this regard, mixed logit is more convenient than probit for representing state-dependence, since lagged dependent variables can be added to mixed logit without adjusting the probability formula or simulation method. Erdem (1996) and Johannesson and Lundin (2000) exploit this advantage to examine habit formation and variety seeking within a mixed logit that also captures random taste variation.

If choices and data are not observed from the start of the process (i.e., from the first choice situation that the person faces), the issue of initial conditions must be confronted, just as with probit. The researcher must somehow represent the probability of the first observed choice, which depends on the previous, unobserved choices. Heckman and Singer (1986) provide ways to handle this issue. However, when the researcher observes the choice process from the beginning, the initial conditions issue does not arise. In this case, the use of lagged dependent variables to capture inertia or other types of state-dependence is straightforward with mixed logit. Stated-preference data (that is, answers to a series of choice situations posed to respondents in a survey) provide a prominent example of the researcher observing the entire sequence of choices.

In the specification so far and in nearly all applications, the coefficients β_n are assumed to be constant over choice situations for a given decision-maker. This assumption is appropriate if the decision-maker's tastes are stable over the time period that spans the repeated choices. However, the coefficients associated with each person can be specified to vary over time in a variety of ways. For example, each person's tastes might be serially correlated over choice situations, such that utility is

$$\begin{aligned} U_{njt} &= \beta_{nt} x_{njt} + \varepsilon_{njt} \\ \beta_{nt} &= b + \tilde{\beta}_{nt} \\ \tilde{\beta}_{nt} &= \rho \tilde{\beta}_{nt-1} + \mu_{nt} \end{aligned}$$

where b is fixed and μ_{nt} is iid over n and t. Simulation of the probability for the sequence of choices proceeds as follows.

1. Draw μ^r_{n1} for the initial period and calculate the logit formula for this period using $\beta^r_{n1} = b + \mu^r_{n0}$.
2. Drawe μ^r_{n2} for the second period, calculate $\beta_{n2} = b + \rho\mu^r_{n1} + \mu^r_{n2}$, and then calculate the logit formula based on this β^r_{n2}.
3. Continue for all T time periods.
4. Take the product of the T logits.
5. Repeat steps 1-4 for numerous sequences of draws.

6. Average the results.

The burden placed on simulation is greater than with coefficients being constant over time for each person, requiring T times as many draws.

6.8 Case Study

As illustration, consider a mixed logit of angler's choice of fishing site (Train, 1999). The specification takes a random-coefficients form. Utility is $U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$, with coefficients β_n varying over anglers but not over trips for each angler. The probability of the sequence of sites chosen by each angler is given by equation (6.3) above.

The sample consists of 962 river trips taken in Montana by 258 anglers during the period of July 1992 through August 1993. A total of 59 possible river sites were defined based on geographical and other relevant factors. Each site contains one or more of the stream segments used in the Montana River Information System. The following variables enter as elements of x for each site:

1. Fish stock, measured in 100 fish per 1000 feet of river.

2. Aesthetics rating, measured on a scale of 0 to 3, with 3 being the highest.

3. Trip cost: cost of traveling from the angler's home to the site, including the variable cost of driving (gas, maintenance, tires, oil) and the value of time spent driving (with time valued at one-third the angler's wage.)

4. Indicator that the *Angler's Guide to Montana* lists the site as a major fishing site.

5. Number of campgrounds per US Geological Survey (USGS) block in the site.

6. Number of state recreation access areas per USGS block in the site.

7. Number of restricted species at the site.

8. Log of the size of the site, in USGS blocks.

The coefficients of variables 4 - 7 can logically take either sign; for example, some anglers might like having campgrounds while other anglers prefer the privacy that comes from not having nearby campgrounds. Each of these coefficients is given an independent normal distribution with mean and standard deviation that are estimated. The coefficients for trip cost, fish stock, and aesthetics rating of the site are expected to have the same sign for all anglers with only their magnitudes differing over anglers. These coefficients are given independent lognormal distributions. The mean and standard deviation of the log of the coefficient is estimated, and the mean and standard deviation of the coefficient itself are calculated from these estimates. Since the lognormal distribution is defined over the positive range and trip cost is expected to have a negative coefficient for all anglers, the negative of trip cost enters the model. The coefficient for the log of size is assumed to be fixed. This variable accounts for the fact that the probability of visiting a larger site is higher than that for a smaller site, all else equal. Having the coefficient of this variable vary over people, while possible, would not be particularly meaningful. A version of the model with correlated coefficients is given by Train (1998). The site choice model is part of an overall model, given by Desvousges, Waters and Train (1996), of the joint choice of trip frequency and site choice.

Simulation was performed using one thousand random draws for each sampled angler. The results are given in Table 6.1. The standard deviation of each random coefficient is highly significant, indicating that these coefficients do indeed vary in the population.

Consider first the normally distributed coefficients. The estimated means and standard deviations of these coefficients provide information on the share of the population that places a positive value on the site attribute and the share that places a negative value. The distribution of the coefficient of the indicator that the Angler's Guide to Montana lists the site as a major site obtains an estimated mean of 1.018 and estimated standard deviation of 2.195, such that 68 percent of the distribution is above zero and 32 percent below. This implies that being listed as a major site in the Angler's Guide to Montana is a positive inducement for about two-thirds of anglers and a negative factor for the other third who apparently prefer more solitude. Campgrounds are preferred by about half (53 percent) of anglers and avoided by the other half. And about one-third of anglers (31 percent) are estimated to prefer having numerous access areas, while the other

Table 6.1: Mixed Logit Model of River Fishing Site Choice

		Parameter	Std. error
Fish Stock	Mean of ln(coefficient)	-2.876	0.6066
	Std. dev. of ln(coefficient)	1.016	0.2469
Aesthetics	Mean of ln(coefficient)	-0.794	0.2287
	Std. dev. of ln(coefficient)	0.849	0.1382
Total cost (neg.)	Mean of ln(coefficient)	-2.402	0.0631
	Std. dev. of ln(coefficient)	0.801	0.0781
Guide lists as major	Mean coefficient	1.018	0.2887
	Std. dev. of coefficient	2.195	0.3518
Campgrounds	Mean coefficient	0.116	0.3233
	Std. dev. of coefficient	1.655	0.4350
Access areas	Mean coefficient	-0.950	0.3610
	Std. dev. of coefficient	1.888	0.3511
Restricted species	Mean coefficient	-0.499	0.1310
	Std. dev. of coefficient	0.899	0.1640
Log(size)	Mean coefficient	0.984	0.1077
Likelihood ratio index		0.5018	
SLL at convergence		-1932.33	

two-thirds prefer there being fewer access areas.

Consider now the lognormal coefficients. Coefficient β^k follows a lognormal if the log of β^k is normally distributed. We parameterize the lognormal distribution is terms of the underlying normal. That is, we estimate parameters b and s that represent the mean and variance of the log of the coefficient: $ln(\beta^k) \sim N(m, s)$. The mean and variance of β^k are then derived from the estimates of m and s. The median is $exp(m)$, the mean is $exp(m + s/2)$, and the variance is $exp(2m + s)[exp(s) - 1]$. The point estimates imply that the coefficients of fish stock, aesthetics, and trip cost have the following median, mean, and standard deviations.

	Median	Mean	Std. Dev
Fish stock	0.0563	0.0944	0.1270
Aesthetics	0.4519	0.6482	0.6665
Trip cost	0.0906	0.1249	0.1185

The ratio of an angler's fish stock coefficients to the trip cost coefficient is a measure of the amount that the angler is willing to pay to have additional fish in the river. Since the ratio of two independent log-normally distributed terms is also log-normally distributed, we can calculate moments for the distribution of willingness to pay. The log of the ratio of the fish stock coefficient to the trip cost coefficient has estimated mean -0.474 and standard deviation of 1.29. The ratio itself therefore has median 0.62, mean 1.44, and standard deviation 2.96. That is, the average willingness to pay to have the fish stock raised by 100 fish per 1000 feet of river is estimated to be \$1.44, and there is very wide variation in angler's willingness to pay for additional fish stock. Similarly, \$9.87 is the estimated average willingness to pay for a site that has an aesthetics rating that is higher by 1, and again the variation is fairly large.

As this application illustrates, the mixed logit provides more information than a standard logit since the mixed logit estimates the extent to which anglers differ in their preferences for site attributes. The standard deviations of the coefficients enter significantly, indicating that a mixed logit provides a significantly better representation of the choice situation than standard logit, which assumes that coefficients are the same for all anglers. The mixed logit also accounts for the fact that several trips are observed for each sampled angler and that each angler's preferences apply to each of the angler's trips.

Chapter 7

Variations on a Theme

7.1 Introduction

Simulation gives the researcher the freedom to specify models that appropriately represent the choice situations under consideration, without being unduly hampered by purely mathematical concerns. This perspective has been the overarching theme of our book. The discrete choice models that we have discussed– namely, logit, nested logit, probit and mixed logit – are used in the vast majority of applied work. However, readers should not feel themselves constrained to use these models. In the current chapter, we describe several models that are derived under somewhat different behavioral concepts. These models are variations on the ones already discussed, directed toward specific issues and data. The point is not simply to describe additional models. Rather, the discussion illustrates how the researcher might examine a choice situation and develop a model and estimation procedure that seem appropriate for that particular situation, drawing from, and yet personalizing, the standard set of models and tools.

Each section of this chapter is motivated by a type of data, representing the outcome of a particular choice process. The arena in which such data might arise is described, and the limitations of the primary models for these data are identified. In each case, a new model is described that better represents the choice situation. Often this new model is only a slight change from one of the primary models. However, the slight change will often make the standard software unuseable, such that the researcher will need to develop her own software, perhaps by modifying the codes that are available for standard models.

The ability to revise code to represent new specifications enables the researcher to utilize the freedom that the field offers.

7.2 Stated-Preference and Revealed-Preference Data

Revealed-preference data relate to people's actual choices in real-world situations. These data are called "revealed-preference" because people reveal their tastes, or preferences, though the choices they make in the world. Stated-preference data are data collected in experimental or survey situations where respondents are presented with hypothetical choice situations. The term "stated-preference" denotes the fact that the respondents state what their choices would be in the hypothetical situations. For example, in a survey, a person might be presented with three cars with different prices and other attributes. The person is asked which of the three cars he would buy if offered only these three cars in the real world. The answer the person gives is the person's stated choice. Revealed-preference data for the respondent is obtained by asking which car they bought when they last bought a car.

There are advantages and limitations to each type of data. Revealed-preference data have the advantage that they reflect actual choices. This, of course, is a very big advantage. However, revealed preference data are limited to the choice situations and attributes of alternatives that currently exist or have existed historically. Often a researcher will want to examine people's responses in situations that do not currently exist, such as the demand for a new product. Revealed-preference data are simply not available for these new situations. Even for choice situations that currently exist, there might be insufficient variation in relevant factors to allow estimation with revealed-preference data. For example, suppose the researcher wants to examine the factors that affect California households' choice of energy supplier. While residential customers have been able to choose among suppliers for many years, there has been practically no difference in price among the suppliers' offers. Customers' response to price cannot be estimated on data that contain little or no price variation. An interesting paradox arises in this regard. If customers were highly price responsive, then suppliers, knowing this, would offer prices that meet their competitors' prices; the well-known equilibrium in this situation is that all firms offer (essentially) the same price. If the data from this market were used in a choice model, the price coefficient would be found to be insignificant, since there is little price variation in the data. The researcher could erroneously conclude from this insignificance that price is unimportant to consumers. This paradox is inherent in revealed-preference data. Factors that are the most important to consumers will often exhibit the least variation due to the natural forces of market equilibrium. Their importance might therefore be difficult to detect with revealed-preference data.

Stated-preference data complement revealed-preference data. A questionnaire is designed in which the respondent is presented with one or more choice experiments. In each experiment, two or more options are described and the respondent is asked which option he/she would choose if facing the choice in the real world. For example, in the data that we examine in Chapter 11, each surveyed respondent is presented with 12 experiments. In each experiment, four hypothetical energy suppliers were described, with the price, contract terms, and other attributes given for each supplier. The respondent is asked to state which of the four suppliers he/she would choose.

The advantage of stated-preference data is that the experiments can be designed to contain as much variation in each attribute as the researcher thinks is appropriate. While there might be little price variation over suppliers in the real world, the suppliers that are described in the experiments can be given sufficiently different prices to allow precise estimation. Attributes can be varied over respondents and over experiments for each respondent. This degree of variation contrasts with market data, where often the same products are available to all customers, such that there is no variation over customers in the attributes of products. Importantly, for products that have never been offered before, or for new attributes of old products, stated-preference data allow estimation of choice models when revealed-preference data do not exist. Louviere, Hensher and Swait (2000) describe the appropriate collection and analysis of stated-preference data.

The limitations of stated-preference data are obvious: what people say they will do is often not the same as what they actually do. People might not know what they would do if a hypothetical situation were real. Or, they might not be willing to say what they would do. In fact, respondents' concept of what they would do might be influenced by factors that wouldn't arise in the real choice situations, such as their perception of what the interviewer expects or wants as answers.

By combining stated- and revealed-preference data, the advantages of each can be obtained while mitigating the limitations. The stated-preference data provide the needed variation in attributes, while the revealed-preference data ground the predicted shares in reality. To utilize these relative strengths, an estimation procedure is needed that (1) allows the ratios of coefficients (which represent the relative importance of the various attributes) to be estimated primarily from the stated-preference data (or more generally, from whatever variation in the attributes exists, which is usually from the stated-preference data), while (2) allowing the alternative-specific constants and overall scale of the parameters to be determined by the revealed preference data (since the constants and scale determine average shares in base conditions.)

Procedures for estimating discrete choice models on a combination of stated- and revealed-preference data are described by Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993) and Hensher, Louviere and Swait (1999) in the context of logit models and by Bhat and Castelar (2001) and Brownstone, Bunch and Train (2000) with mixed logit. These procedures constitute variations on the methods we have already examined. The most prevalent issue when combining stated- and revealed-preference data is that the unobserved factors are generally different for the two types of data. We describe in the following paragraphs how this issue can readily be addressed.

Let the utility that person n obtains from alternative j in situation t be specified as $U_{njt} = \beta' x_{njt} + e_{njt}$ where x_{njt} does not include alternative-specific constants and e_{njt} represents the impact of factors that are not observed by the researcher. These factors have a mean for each alternative (representing the average effect of all excluded factors on the utility of that alternative) and a distribution around this mean. The mean is captured by an alternative-specific constant, labeled c_j, and, for a standard logit model, the distribution around this mean is extreme value with variance $\lambda^2\pi^2/6$. As described in chapters 2 and 3, the scale of utility is set by normalizing the variance of the unobserved portion of utility. The utility function becomes $U_{njt} = (\beta/\lambda)' x_{njt} + (c_j/\lambda) + \varepsilon_{njt}$, where the normalized error $\varepsilon_{njt} = (e_{njt} - c_j)/\lambda$, is now iid extreme value with variance $\pi^2/6$. The choice probability is the logit formula based on $(\beta/\lambda)' x_{njt} + (c_j/\lambda)$. The parameters that are estimated are the original parameters divided by the scale factor λ.

This specification is reasonable for many kinds of data and choice situations. However, there is no reason to expect the alternative-specific constants and the scale factor to be the same for stated-preference data as for revealed-preference data. These parameters reflect the impact of unobserved factors, which are necessarily different in real choice situations than hypothetical survey situations. In real choices, a multitude of issues that affect a person but are not observed by the researcher come into play. In a stated-preference experiment, the respondent is (usually) asked to assume all alternatives to be the same on factors that are not explicitly mentioned in the experiment. If the respondent follows this instruction exactly, there would, by definition, be no unobserved factors in the stated-preference choices. Of course, respondents inevitably bring some outside concepts into the experiments, such that unobserved factors do enter. However, there is no reason to expect that these factors are the same, in mean or variance, as in real world choices.

To account for these differences, separate constants and scale parameters are specified for stated-preference choice situations as for revealed-preference situations. Let c_j^s and c_j^r represent the mean impact of unobserved factors for alternative j in stated-preference experiments and revealed-preference choices, respectively. Similarly, let λ^s and λ^r represent the scale (proportional to the standard deviation) of the distribution of unobserved factors around these means in stated- and revealed-preference situations, respectively. To set the overall scale of utility, we normalize either of the scale parameters to 1, which makes the other scale parameter be the ratio of the two original scale parameters. Let's normalize λ^r, such that λ^s reflects the variance of unobserved factors in stated-preference situations relative to that in revealed-preference situations. Utility then becomes:

$$U_{njt} = (\beta/\lambda^s)' x_{njt} + (c_j^s/\lambda^s) + \varepsilon_{njt}$$

for each t that is a stated-preference situation, and

$$U_{njt} = \beta\prime x_{njt} + c_j^r + \varepsilon_{njt}$$

for each t that is a revealed-preference situation.

The model is estimated on the data from both the revealed- and stated-preference choices. Both groups of observations are "stacked" together as input to a logit estimation routine. A separate set of alternative-specific constants is estimated for the stated-preference and revealed-preference data. Importantly, the coefficients in the model are divided by a parameter $1/\lambda^s$ for the stated-preference observations. This separate scaling is not feasible in most standard logit estimation packages. However, the researcher can easily modify available codes (or their own code) to allow for this extra parameter. Hensher and Bradley (1993) show how to estimate this model on software for nested logits.

Note that, with this set-up, the elements of β are estimated on both types of data. The estimates will necessarily reflect the amount of variation that each type of data contains for the attributes (that is, the elements of x). If there is little variance in the revealed-preference data, reflecting conditions in real-world markets, then the β's will be determined primarily by the stated-preference data, which contain whatever variation was built into the experiments. Insofar as the revealed-preference data contain useable variation, this information will be incorporated into the estimates.

The alternative-specific constants are estimated separately for the two types of data. This distinction allows the researcher to avoid many of the biases that stated-preference data might exhibit. For example, respondents often say that they will buy a product far more than they actually end up doing. The average probability of buying the product is captured in the alternative-specific constant for the product. If this bias is occurring, then the estimated constant for the stated-preference data will be greater than that for the revealed-preference data. When forecasting, the researcher can use the constant from the revealed-preference data, thereby grounding the forecast in a market-based reality. Similarly, the scale for the revealed preference data (which is normalized to 1) can be used in forecasting instead of the scale from the stated-preference data, thereby incorporating correctly the real-world variance in unobserved factors.

7.3 Ranked Data

In stated-preference experiments, respondents might be asked to rank the alternatives instead of just identifying the one alternative that they would choose. This ranking can be requested in a variety of ways. The respondents can be asked to state which alternative they would choose, and then, after they have make this choice, can be asked which of the remaining alternatives they would choose, continuing through all the alternatives. Instead, respondents can simply be asked to rank the alternatives from best to worst. In any case, the data that the researcher obtains is a ranking of the alternatives that presumably reflects the utility that the respondent obtains from each alternative.

Ranked data can be handled in a standard logit or mixed logit model using currently available software without modification. All that is required is that the input data be constructed in a particular way, which we describe below. For a probit model, the available software would need to be modified slightly to handle ranked data. However, the modification is straightforward. We consider standard and mixed logit first.

7.3.1 Standard and mixed logit

Under the assumptions for standard logit, the probability of any ranking of the alternatives from best to worst can be expressed as the product of logit formulas. Consider, for example, a respondent who was presented with four alternatives labeled A, B, C, and D. Suppose the person ranked the alternatives as follows: C, B, D, A, where C is the first choice. If the utility of each alternative is distributed iid extreme value (as for a logit model), then the probability of this ranking can be expressed as: the logit probability of choosing alternative C from the set A, B, C, D, *times* the logit probability of choosing alternative B from the remaining alternatives A, B, D, *times* the probability of choosing alternative D from the remaining alternatives A and D.

Stated more explicity, let $U_{nj} = \beta' x_{nj} + \varepsilon_{nj}$ for $j = A, \ldots, D$ with ε_{nj} iid extreme value. Then:

$$\begin{aligned} &Prob(\text{ranking } C, B, D, A) \qquad (7.1)\\ &= \frac{e^{\beta' x_{nC}}}{\sum_{j=A,B,C,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nB}}}{\sum_{j=A,B,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nD}}}{\sum_{j=A,D} e^{\beta' x_{nj}}} \end{aligned}$$

This simple expression for the ranking probability is an outcome of the particular form of the extreme value distribution, first shown by Luce and Suppes (1965). It does not apply in general; for example it does not apply with probit models.

Equation 7.1 implies that the ranking of the four alternatives can be represented as being the same as three independent choices by the respondent. These three choices are called "pseudo-observations" because each respondents' complete ranking, which constitutes an observation, is written as if it were multiple observations. In general, a ranking of J alternatives provides $J-1$ "pseudo-observations" in a standard logit model. For the first pseudo-observation, all alternatives are considered available, and the dependent variable identifies the first-ranked alternative. For the second pseudo-observation, the first ranked alternative is discarded. The remaining alternatives constitute the choice set, and the dependent variable identifies the second-ranked alternative. And so on. In creating the input file for logit estimation, the explanatory variables for each alternative are repeated $J-1$ times, making that many pseudo-observations. The dependent variable for these pseudo-observations identifies, respectively, the first-, second-, etc. ranked alternatives. For each pseudo-observation, the alternatives that are ranked above the dependent variable for that pseudo-observation are omitted (i.e., censored out.) Once the data are constructed in this way, the logit estimation proceeds as usual.

A logit model on ranked alternatives is often called an "exploded logit," since each observation is exploded into several pseudo-observations for the purposes of estimation. Prominent applications include Beggs, Cardell and Hausman (1981), Chapman and Staelin (1982), and Hausman and Ruud (1987).

A mixed logit model can be estimated on ranked data with the same explosion. Assume now that β is random with density $g(\beta \mid \theta)$ where θ are parameters of this distribution. Conditional on β, the probability of the person's ranking is a product of logits, as given in equation 7.1 above. The unconditional probability is then the integral of this product over the density of β:

$$
\begin{aligned}
&Prob(\text{ranking } C, B, A, D) \\
&= \int \left(\frac{e^{\beta' x_{nC}}}{\sum_{j=A,B,C,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nB}}}{\sum_{j=A,B,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nD}}}{\sum_{j=A,D} e^{\beta' x_{nj}}} \right) g(\beta \mid \theta) d\theta.
\end{aligned}
$$

The mixed logit model on ranked alternatives is estimated with regular mixed logit routines for panel data, using the input data set-up as described above for logit where the $J-1$ pseudo-observations for each ranking are treated as $J-1$ choices in a panel. The mixed logit incorporates the fact that each respondent has his own coefficients and, importantly, that the respondent's coefficients affect his entire ranking such that the pseudo-observations are correlated. A logit model on mixed data does not account for this correlation.

7.3.2 Probit

Ranked data can also be utilized effectively in a probit model. Let the utility of the four alternatives be as stated above for a logit except that the error terms are jointly normal: $U_{nj} = \beta' x_{nj} + \varepsilon_{nj}$ for $j = A, B, C, D$ where $\varepsilon_n = \langle \varepsilon_{nA}, \ldots, \varepsilon_{nD} \rangle'$ is distributed $N(0, \Omega)$. As before, the probability of the person's ranking is $Prob(\text{ranking } C, B, D, A) = Prob(U_{nC} > U_{nB} > U_{nD} > U_{nA})$. Decomposing this joint probability into conditionals and a marginal does not help with a probit in the way that it does with logit, since the conditional probabilities do not collapse to unconditional probabilities as they do under independent errors. Another tack is taken instead. Recall that for probit models, we found that it is very convenient to work in utility differences rather than the utilities themselves. Denote $\tilde{U}_{njk} = U_{nj} - U_{nk}$, $\tilde{x}_{njk} = x_{nj} - x_{nk}$, and $\tilde{\varepsilon}_{njk} = \varepsilon_{nj} - \varepsilon_{nk}$. The probability of the ranking can then be expressed as $Prob(\text{ranking } C, B, D, A) = Prob(U_{nC} > U_{nB} > U_{nD} > U_{nA}) = Prob(\tilde{U}_{nBC} < 0, \tilde{U}_{nDB} < 0, \tilde{U}_{nDA} < 0)$.

To express this probability, we define a transformation matrix M that takes appropriate differences. The reader might want to review section (5.6.3) on simulation of probit probabilities for one chosen alternative, which uses a similar transformation matrix. The same procedure is used for ranked data, but with a different transformation matrix.

Stack the alternatives A to D, such that utility is expressed in vector form as $U_n = V_n + \varepsilon_n$ where $\varepsilon_n \sim N(0, \Omega)$. Define the 3×4 matrix:

$$M = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

This matrix has a row for each inequality in the argument of the probability $Prob(\tilde{U}_{nBC} < 0, \tilde{U}_{nDB} < 0, \tilde{U}_{nDA} < 0)$. Each row contains a 1 and a -1, along with zeros, where the 1 and -1 identify the alternatives that are being differenced for the inequality. With this matrix, the probability of the ranked alternatives becomes

$$\begin{aligned} Prob(\text{ranking } C, B, D, A) &= Prob(\tilde{U}_{nBC} < 0, \tilde{U}_{nDB} < 0, \tilde{U}_{nDA} < 0) \\ &= Prob(MU_n < 0) \\ &= Prob(MV_n + M\varepsilon_n < 0) \\ &= Prob(M\varepsilon_n < -MV_n) \end{aligned}$$

The error differences defined by $M\varepsilon_n$ are distributed jointly normal with zero mean and covariance $M\Omega M'$. The probability that these correlated error differences fall below $-MV_n$ is simulated by GHK in the manner given in section (5.6.3). The procedure has been implemented by Hajivassiliou and Ruud (1994) and Schechter (2001).

7.4 Ordered Responses

In surveys, respondents are often asked to provide ratings of various kinds. Examples include:

How good a job do you think the President is doing? Check one:

1. very good job
2. good job
3. neither good nor bad
4. poor job
5. very poor job

How well do you like this book? Rate the book from 1 to 7, where 1 is the worst you have ever read (aside from *The Bridges of Madison County*, of course) and 7 is the best

1 2 3 4 5 6 7

How likely are you to buy a new computer this year?

1. Not likely at all
2. Somewhat likely
3. Very likely

The main characteristic of these questions, from a modeling perspective, is that the potential responses are ordered. A book rating of 6 is higher than 5 which is higher than 4; and a Presidential rating of "very poor" is worse than "poor," which is worse than "neither good nor bad". A standard logit model could be specified with each potential response as an alternative. However, the logit model's assumption of independent errors for each alternative is inconsistent with the fact that the alternatives are ordered: with ordered alternatives, one alternative is similar to those close to it and less similar to those further away. The ordered nature could be handled by specifying a nested logit, mixed logit, or probit model that accounts for the pattern of similarity and dissimilarily among the alternatives. For example, a probit model could be estimated with correlation among the alternatives, with the correlation between 2 and 3 being greater than that between 1 and 3, and the correlation between 1 and 2 also being larger than that between 1 and 3. However, such a specification, while it might provide fine results, does not actually fit the structure of the data. Recall that the traditional derivation for these models starts with a specification of the utility associated with each alternative. For the ratings question about the President's job, the derivation would assume that there are seven utilities, one for each potential response, and that the person chooses the number 1 to 5 that has the greatest utility. While it is perhaps possible to think of the decision process in this way (and the resulting model will probably provide useful results), it is not a very natural way to think about the respondent's decision.

A more natural representation of the decision process is to think of the respondent as having some level of utility or opinion associated with the object of the question and answering the question on the basis of how great this utility is. For example, on the Presidential question, the following derivation seems to better represent the decision process. Assume that the respondent has an opinion on how well the President is doing. This opinion is represented in a (unobservable) variable that we label U, where higher levels of U mean that the person thinks the President is doing a better job and lower levels mean he thinks the President is going a poorer job. In answering the question, the person is asked to express this opinion in one of five categories: "very good job," "good job," etc. That is, even though the person's opinion, U, can take many different levels representing various levels of liking or disliking the job the President is doing, the question allows only 5 possible responses. The person chooses a response on the basis of the level of his U. If U is above some cutoff, which we label k_1, the respondent chooses the answer "very good job." If U is below k_1 but above another cut-off, k_2, then he answers "good job." And so on. The decision is represented as:

- "very good job" if $U > k_1$

- "good job" if $k_1 > U > k_2$
- "neither good or bad" if $k_2 > U > k_3$
- "poor job" if $k_3 > U > k_4$
- "very poor job" if $k_4 > U$

The researcher observes some factors that relate to the respondents' opinion, such as the person's political affiliation, income, and so on. However, other factors that affect the person's opinion cannot be observed. Decompose U into unobserved and unobserved components: $U = \beta' x + \varepsilon$. As usual, the unobserved factors ε are considered random. Their distribution determines the probability for the five possible responses.

Figure 7.1 illustrates the situation. U is distributed around $\beta' x$ with the shape of the distribution following the distribution of ε. There are cut-off points for the possible responses: $k_1, \ldots, k_4$. The probability that the person answers with "very poor job" is the probability that U is less than k_4, which is the area in the left tail of the distribution. The probability that the person says "poor job" is the probability that U is above k_4, indicating that he doesn't think that the job is very poor, but is below k_3. This probability is the area between k_4 and k_3.

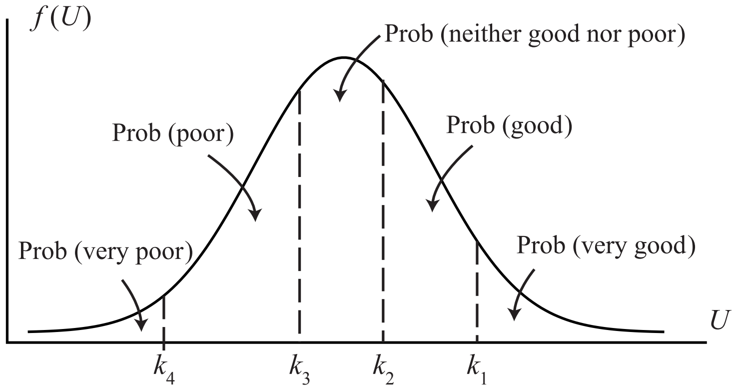

Figure 7.1: Distribution of opinion about president's job.

Once a distribution for ε is specified, the probabilities can be calculated exactly. For simplicity, assume that ε is distributed logistic, which means that the cumulative distribution of ε is $F(\varepsilon) =$

$exp(\varepsilon)/(1 + exp(\varepsilon))$. The probability of the answer "very poor job" is then:

$$\begin{aligned} Prob(\text{"very poor job"}) &= Prob(U < k_4) \\ &= Prob(\beta' x + \varepsilon < k_4) \\ &= Prob(\varepsilon < k_4 - \beta' x) \\ &= \frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}} \end{aligned}$$

The probability of "poor job" is:

$$\begin{aligned} Prob(\text{"poor job"}) &= Prob(k_4 < U < k_3) \\ &= Prob(k_4 < \beta' x + \varepsilon < k_3) \\ &= Prob(k_4 - \beta' x < \varepsilon < k_3 - \beta' x) \\ &= Prob(\varepsilon < k_3 - \beta' x) - Prob(\varepsilon < k_4 - \beta' x) \\ &= \frac{e^{k_3 - \beta' x}}{1 + e^{k_3 - \beta' x}} - \frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}} \end{aligned}$$

Probabilities for the other answers are obtained analogously. The probabilities enter the log-likelihood function as usual, and maximization of the likelihood function provides estimates of the parameters. Note that the parameters consist of β, which gives the impact of the explanatory variables on people's opinion of the President, as well as the cut-off points $k_1, \ldots, k_4$.

The model is called ordered logit, since it uses the logistic distribution on ordered alternatives. Unfortunately, nested logit models have occasionally been called ordered logits; this nomenclature causes confusion and will hopefully be avoided in the future.

Note that the probabilities in the ordered logit model incorporate the binary logit formula. This similarity to binary logit is only incidental: the traditional derivation of a binary logit specifies two alternatives with utility for each, while the ordered logit model has one utility with multiple alternatives to represent the level of that utility. The similarity in formula arises from the fact that, if two random variables are iid extreme value, then their difference follows a logistic distribution. Therefore, assuming that both utilities in a binary logit are iid extreme value is equivalent to assuming that the difference in the utilities is distributed logistic, the same as the utility in the ordered logit model.

A similar model is obtained under the assumption that ε is distributed standard normal instead of logistic (Zavoina and McElvey, 1975). The only difference arises in that the binary logit formula is replaced with the cumulative standard normal distribution. That is,

$$\begin{aligned} Prob(\text{“very poor job”}) &= Prob(\varepsilon < k_4 - \beta' x) \\ &= \Phi(e^{k_4 - \beta' x}) \end{aligned}$$

and

$$\begin{aligned} Prob(\text{“poor job”}) &= Prob(\varepsilon < k_3 - \beta' x) - Prob(\varepsilon < k_4 - \beta' x) \\ &= \Phi(e^{k_3 - \beta' x}) - \Phi(e^{k_4 - \beta' x}) \end{aligned}$$

where Φ is the standard cumulative normal function. This model is called ordered probit. Software for ordered logit and probit is available in many commercial packages.

The researcher might believe that the parameters vary randomly in the population. In that case, a mixed version of the model can be specified, as in Bhat (1999). Let the density of β be $g(\beta \mid \theta)$. Then the mixed ordered logit probabilities are simply the ordered logit probabilities integrated over the density $g(\cdot)$. For example,

$$Prob(\text{“very poor job”}) = \int \left(\frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}} \right) g(\beta \mid \theta) d\beta$$

and

$$Prob(\text{“poor job”}) = \int \left(\frac{e^{k_3 - \beta' x}}{1 + e^{k_3 - \beta' x}} - \frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}} \right) g(\beta \mid \theta) d\beta$$

and so on. These probabilities are simulated in the same way as mixed logits, by drawing values of β from $g(\cdot)$, calculating the ordered logit probability for each draw, and averaging the results. Mixed ordered probit is derived similarly.

7.4.1 Multiple ordered responses

Respondents' answers to different questions are often related. For example, a person's rating of how well the President is doing is probably related to the person's rating of how well the economy is doing. The researcher might want to incorporate into the analysis the fact that the answers are related. To be concrete, suppose that respondents are asked to rate both the President and the economy on a five point scale, like the rating given above for the President. Let U be the respondent's opinion of the job the president is doing, and let W be the respondent's assessment of the economy. Each of these assessments can be decomposed into observed and unobserved factors: $U = \beta' x + \varepsilon$ and $W = \alpha' z + \mu$. Insofar as the assessments are related due to observed factors, the same variables can be included in x and z. To allow for the possibility that the assessments are related due to unobserved factors, we specify ε and μ to be jointly normal with correlation ρ (and unit variances by normalization.) Let the cut-offs for U be denoted $k_1, \ldots, k_4$ as before and the cut-offs for W be denoted $c_1, \ldots, c_4$. We want to derive the probability of each possible combination of responses to the two questions.

The probability that the person says the President is doing a "very poor job" and also that the economy is doing "very poorly" is derived as follows:

$$
\begin{aligned}
&Prob(\text{President "very poor" and economy "very poor"})\\
&\quad = Prob(U < k_4 \text{ and } W < c_4)\\
&\quad = Prob(\varepsilon < k_4 - \beta' x \text{ and } \mu < c_4 - \alpha' z)\\
&\quad = Prob(\varepsilon < k_4 - \beta' x) \times Prob((\mu < c_4 - \alpha' z \mid \varepsilon < k_4 - \beta' x)
\end{aligned}
$$

Similarly, the probability of a rating of "very poor" for the President and "good" for the economy is:

$$
\begin{aligned}
&Prob(\text{President "very poor" and economy "good"})\\
&\quad = Prob(U < k_4 \text{ and } c_2 < W < c_1)\\
&\quad = Prob(\varepsilon < k_4 - \beta' x \text{ and } c_2 - \alpha' z < \mu < c_1 - \alpha' z)\\
&\quad = Prob((\varepsilon < k_4 - \beta' x) \times Prob((c_2 - \alpha' z < \mu < c_1 - \alpha' z \mid \varepsilon < k_4 - \beta' x)
\end{aligned}
$$

The probabilities for other combinations are derived similarly, and generalization to more than two related questions is straightforward. The model is called multivariate (or multiresponse) ordered probit. The probabilities can be simulated by GHK in a manner similar to that described in chapter 5. The explanation in Chapter 5 assumes that truncation of the joint normal is only on one side (since for a standard probit the probability that is being calculated is the probability that all utility differences are below zero, which is truncation from above), while the probabilities for multivariate ordered probit are truncated on two sides (as for the second probability above). However, the logic is the same, and interested readers can refer to Hajivassiliou and Ruud (1994) for an explicit treatment of GHK with two-sided truncation.

7.5 Contingent Valuation

In some surveys, respondents are asked to express their opinions or actions relative to a specific number that the interviewer states. For example, the interviewer might ask: "Consider a project that protected the fish in specific rivers in Montana. Would you be willing to spend \$ 50 to know that the fish in these rivers are safe?" This question is sometimes followed by another question that depends on the respondent's answer to the first question. For example, if the person said "yes" to the above question, the interviewer might follow-up by asking, "How about \$ 75? Would you be willing to pay \$ 75?" If the person answered "no" to the first question, indicating that he was not willing to pay \$ 50, the interviewer would follow-up with "Would you be willing to pay \$ 25?"

These kinds of questions are used in environmental studies where the lack of markets for environmental quality prevent valuation of resources by revelation procedures; the papers edited by Hausman (1993) provide a review and critique of the procedure, which is often called "contingent valuation." When only one question is asked, such as whether the person is willing to pay \$ 50, the method is called "single bounded," since the person's answer gives one bound on their true willingness to pay. If the person answers "yes," the researcher knows that their true willingness to pay is at least \$ 50, but she does not know how *much* more. If the person answers "no," the researcher knows that the person's willingness to pay is less than \$ 50. Examples of studies using single bounded methods are Cameron and James (1987) and Cameron (1988).

When a follow-up question is asked, the method is called "double-bounded." If the person says that he is willing to pay \$ 50 but not \$ 75, the researcher knows his true willingness to pay is between \$ 50 and \$ 75, that is, is bounded on both sides. If the person says he is not willing to pay \$ 50 but is willing to pay \$ 25, his willingness to pay is known to be between \$ 25 and \$ 50. Of course, even with a double-bounded method, some respondents' willingness to pay is only singly bounded, such as a person who says he is willing to pay \$ 50 and also willing to pay \$ 75. Examples of this approach include Hanemann, Loomis and Kanninen (1991), Cameron and Quiggin (1994), and Cai, Deilami and Train (1998).

The figure that is used as the prompt, *i.e.*, the \$ 50 in our example, is varied over respondents. The answers from a sample of people are then used to estimate the distribution of willingness to pay. The estimation procedure is closely related to that described above for ordered logits and probits, except that the cut-off points are given by the questionnaire design rather than estimated as parameters. We describe the procedure as follows.

Let W_n represent the true willingness to pay of person n. W_n varies over people with distribution $f(W \mid \theta)$ where θ are the parameters of the distribution, such as the mean and variance. The researcher's goal is to estimate these population parameters. Suppose the researcher designs a questionnaire with a single-bounded approach, giving a different prompt (or reference value) for different respondents. Denote the prompt that is given to person n as k_n. The person answers the question with a "yes" if $W_n > k_n$ and "no" otherwise. The researcher assumes that W_n is distributed normally in the population with mean $\bar{W}$ and variance σ^2.

The probability of "yes" is $Prob(W_n > k_n) = 1 - Prob(W_n < k_n) = 1 - \Phi((k_n - \bar{W})/\sigma)$ and the probability of "no" is $\Phi((k_n - \bar{W})/\sigma)$, where $\Phi(\cdot)$ is the standard cumulative normal function. The log-likelihood function is then $\sum_n y_n log(1 - \Phi((k_n - \bar{W})/\sigma)) + (1 - y_n) log(\Phi((k_n - \bar{W})/\sigma))$, where $y_n = 1$ if person n said "yes" and 0 otherwise. Maximizing this function provides estimates of $\bar{W}$ and σ.

A similar procedure is used if the researcher designs a double-bounded questionnaire. Let the prompt for the second question be k_{nu} if the person answered "yes" to the first question, where $k_{nu} > k_n$, and let k_{nl} be the second prompt if the person initially answered "no," where $k_{nl} < k_n$. There are four possible sequences of answers to the two questions. The probabilities for these sequences are illustrated in Figure 7.2 and given below:

- "no" then "no": $P = Prob(W_n < k_{nl}) = \Phi((k_{nl} - \bar{W})/\sigma)$

- "no" then "yes": $P = Prob(k_{nl} < W_n < k_n) = \Phi((k_n - \bar{W})/\sigma) - \Phi((k_{nl} - \bar{W})/\sigma)$

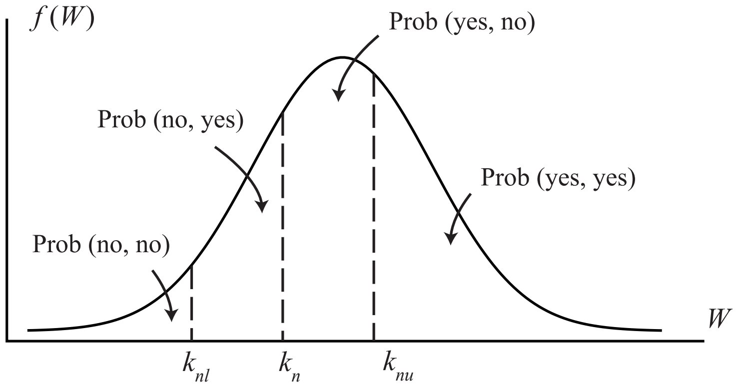

Figure 7.2: Distribution of willingness to pay.

- "yes" then "no": $P = Prob(k_n < W_n < k_{nu}) = \Phi((k_{nu} - \bar{W})/\sigma) - \Phi((k_n - \bar{W})/\sigma)$
- "yes" then "yes": $P = Prob(W_n > k_{nu}) = 1 - \Phi((k_{nu} - \bar{W})/\sigma)$

These probabilities enter the log-likelihood function, which is maximized to obtain estimates of $\bar{W}$ and σ. Other distributions can of course be used instead of normal. Lognormal is attractive if the researcher assumes that all people have a positive willingness to pay. Or the researcher might specify a distribution that has a mass at zero to represent the share of people who are not willing to pay anything, and a lognormal for the remaining share. Generalization to multiple dimensions is straightforward, to reflect, for example, that people's willingness to pay for one environmental package might also be related to their willingness to pay for another. As with multi-response ordered probit, the GHK simulator comes in handy when the multiple values are assumed to be distributed jointly normal.

7.6 Mixed Models

We have discussed mixed logit and mixed ordered logit. Of course, mixed models of all kinds can be developed using the same logic. Any model whose probabilities can be written as a function of parameters can also be mixed by allowing the parameters to be random and integrating the function over the distribution of parameters (Greene, 2001)