Chapter 21:

Discrete Choice Modeling

Wim Bernasco
Netherlands Institute
For The Study of Crime
and Law Enforcement
Amsterdam
&
Department of Spatial Economics
VU University Amsterdam,
Netherlands

Richard Block
Loyola University
Chicago, IL

Table of Contents

Chapter 21:
Discrete Choice Modeling

Introduction

This chapter describes the discrete choice framework and the two most well-known models that are part of it: the Multinomial Logit (MNL) and the Conditional Logit (CL). These techniques require a solid background in statistics and especially regression modeling. A background in economics will also be beneficial, though not necessary. Analysts wishing to use these techniques, in particular the conditional logit model, would be advised to find an expert to work with in developing applications.

The MNL and CL are two closely related statistical regression models that can be used to analyze a discrete outcome variable as a function of a set of independent variables. Discrete variables are also known as nominal or categorical variables. They can take on a finite number of unordered, mutually exclusive values. Both the MNL and the CL are generalizations of the logit model, which is used to analyze binomial (two category) outcome variables and which was discussed in Chapter 18.

Gender is an example of a binomial variable (it is either male or female). The weapon used in a robbery (gun, knife, strong arm, or other weapon) is a multinomial variable. Other examples are the mode of transport used by a rapist (car, scooter, train, bus, bike, walking) or the neighborhood in which a burglary was committed (any one of the city's neighborhoods).

Although the MNL and CL models can be used for all analytical problems where the outcome variable is discrete (nominal, categorical), in a number of disciplines the models are used to study the way that people or organizations make *choices*. Many research questions in the social and behavioral sciences, including criminology, deal with understanding and predicting discrete choices (Bernasco & Block, 2009). Political scientists aim to understand why people vote and what makes them choose a particular party (Palfrey & Poole, 1987). The party vote is a discrete variable. Sociologists want to understand what makes people decide in favor of a particular education, occupation, or marriage partner (Jepsen & Jepsen, 2002). Schools, occupations and partners are discrete choices. In marketing research, understanding and predicting consumer choice is a central concern (McFadden, 1980). Most consumer choices are discrete, such as which brand and model of car to purchase, or in which restaurant to have lunch. Transportation modelers predict why commuters choose to travel by bus, train, car or bicycle (Train, 1980). Behavioral ecological models try to find out what influences an animal's choice of where to forage, rest, or reproduce (Krebs & Davies, 1993).

Choice is also a central concern in crime analysis. What criteria does a police officer use to arrest or not arrest a juvenile? How does a robber choose a specific victim or a particular location to commit a robbery (Bernasco & Block, 2009)? This question addresses criminal location choice, which formed the major impetus to include these models in CrimeStat.

Although the MNL and CL models are both discrete choice models and share the same underlying likelihood function, they are quite different in practice. The main difference between the MNL and the CL model lies in the assumed sources of variation in choice outcomes. The MNL model assumes that variation in the characteristics of decision makers (e.g., age) determines variation in choice outcomes, whereas the CL model assumes that variation in the characteristics of the alternatives themselves (e.g., presence of a bar) determines variation in the choice outcomes.

Aggregated spatial interaction or 'gravity' models had been applied to criminal location choice and crime trips by Smith (1976) and Rengert (1981). These models bear a strong similarity in form and function to the discrete spatial choice models discussed in this chapter, but they are aggregated models of the volume of crime trips between areas. The discrete spatial choice approach was introduced in the criminological literature by Bernasco and Nieuwbeerta (2005) and has subsequently been applied in other studies (Bernasco, 2006, 2010a, 2010b; Bernasco & Block, 2009; Bernasco & Kooistra, 2010; Clare, Fernandez, & Morgan, 2009). Bernasco (2007) demonstrates how the discrete choice model can be reversed to form a tool in geographic offender profiling.

Neither the MNL nor the CL models require that the outcome variable be interpreted as a choice. In fact, the models can be used to model the outcomes of any process that results a finite number of unordered possible outcomes. For example, one study proposed a five-category typology of homicides in terms of the geographical relation between victim residence, offender residence and homicide location (Tita & Griffiths, 2005). It then used the MNL model to study the effects of various interactional, motivational and situational characteristics of the homicides on the type of the homicide. In this study it would be difficult to interpret the outcome as a decision, but the multinomial model is nevertheless useful to describe the effects of the variables on the different outcomes. Besides spatial choice, the conditional logit model has not been used very often in research on crime. An exception is a study that investigated the causes of criminal vengeance in conflicts (Phillips, 2003).

In the remainder of this chapter, the MNL and CL models are discussed in detail. First we demonstrate how the discrete choice model (encompassing both MNL and CL) is derived from random utility theory, and show the differences between the MNL and the CL models. Next we illustrate the structure of the data necessary to estimate MNL and CL models and give examples of both models.

Discrete Choice Framework

The discrete choice framework was developed in the 1970's by McFadden (1973) and others working in the field of travel demand, and the first applications of discrete choice were in the study of travel mode choice (i.e., the choice between train, bus, car, or airplane). Later the model was also applied to the choice of a travel routes and travel destinations (Ben-Akiva & Lerman 1985). This book is probably the most accessible and complete reference work on discrete choice that focuses on the conditional logit and multinomial logit model. A more advanced and more technical reference work is Train (2009), which is freely available (http://elsa.berkeley.edu/~train/).

The discrete choice framework consists of a set of assumptions regarding four elements of a choice situation (Ben-Akiva & Bierlaire 1999):

1. Decision makers. The decision maker is the person or agent that makes a choice.

2. Alternatives. The decision maker must choose one alternative from the choice set, i.e. the set of available alternatives that are mutually exclusive and collectively include all possible choices.

3. Attributes. Alternatives have attributes that make them attractive to the decision maker. The decision maker evaluates the attractiveness of all alternatives. The decision makers themselves can also have attributes.

4. Decision rule. According to economic theory, the decision maker chooses the alternative that maximizes his/her (expected) utility (net gain, profits, satisfaction).

The discussion that follows is mathematically advanced. Readers who prefer to skip the mathematical description of the models may want to continue reading at the "Data structures" section on page 21.7. We follow the notation of Train (2009).

A decision maker, labeled *n*, must make a choice among *J* alternatives. Note that the word 'alternatives' is used for the items, actions or locations that can be chosen, and the word 'choice' is used for the decision of the decision maker in selecting one of these alternatives. By convention the complete set of available alternatives is referred to as the 'choice set', although 'set of alternatives' might better describe it.

Decision maker *n* obtains a level of utility (profits, satisfaction), U_{ni}, from alternative *i* if that alterative is chosen. The principle of utility maximization asserts that the decision maker decides in favor of the alternative *i* if and only if the individual expects to derive more utility from alternative *i* than from any other available alternative. Thus, if the decision maker decides in favor of alternative *i*, then that person must expect to derive less utility from each of the other alternatives (the expression $\forall j \neq i$ means 'for all values of *j* such that *j* not equals *i*).

$$U_{ni} > U_{nj} \forall j \neq i \,. \tag{21.1}$$

The utilities are known by the decision maker, but not by the analyst. The analyst only observes the *J* alternatives, some attributes a_{ni} of the alternatives, some attributes d_n of the decision maker, and can specify a function *V*, often called *representative utility* or *systematic utility*, that links these observed attributes to the decision maker's utility:

$$V_{ni} = V(a_{ni}, d_n) \forall i \tag{21.2}$$

The analyst incompletely observes utility, so that generally $U_{ni} \neq V_{ni}$. The utility can be written as the sum of representative utility V_{ni} and a term μ_{ni} that captures the factors that determine utility but are not observed by the analyst, and that is treated as random.

$$U_{ni} = V_{ni} + \varepsilon_{ni} \tag{21.3}$$

The probability that decision maker *n* chooses alternative *i* is the probability that the utility associated with choosing *i* is greater than the utility associated with any other alternative in the choice set:

$$P_{ni} = \Pr(U_{ni} > U_{nj} \forall j \neq i) \tag{21.4}$$

$$P_{ni} = \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i) \tag{21.5}$$

$$P_{ni} = \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i) \tag{21.6}$$

This is the most general formulation of the discrete choice model, and any specific choice model that is consistent with random utility maximization can be derived from specific assumptions on the joint distribution of the unobserved utility term μ_{ni}. CrimeStat can estimate the two most basic models of this family, the multinomial logit model and the conditional logit model. There are many others, including for example nested logit, mixed logit, and multinomial probit. These are described in Train (2009).

Multinomial and Conditional Logit

If the unobserved random utility components μ_{ni} are independent and identically distributed according to an *extreme value distribution* (also referred to as a Gumbel distribution), the MNL model and the CL can be derived. Originally, the general form of both was labeled the *conditional logit model* (McFadden 1973). Today both models are usually simply referred to as 'multinomial logit model' or even 'logit model' in the discrete choice literature. CrimeStat distinguishes between the MNL and the CL models because despite their mathematical equivalence, they require a different organization of the data. In the general model that encompasses both the CL and the MNL, the choice probability, P_{ni}, the probability that decision maker *n* chooses alternative *i*, is given by:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} \tag{21.7}$$

For computational convenience, and because any function can be closely approximated by a linear function, representative utility V_{ni} is usually assumed to be linear in the parameters. The specification of observed utility V_{ni} is different in the MNL and the CL models. In the MNL model, V_{ni} depends on the characteristics of the decision maker while in the CL model, it depends on the characteristics of the alternatives.

Multinomial Logit Model

In the MNL model,

$$V_{ni} = \boldsymbol{\beta}'_{\mathbf{i}} \mathbf{X}_{\mathbf{n}} = \sum_{k=1}^{K} \beta_{ki} X_{kn} \tag{21.8}$$

In equation 21.8, *K* is the number of predictor variables in the model, X_{kn} is the value of the kth predictor variable for observational unit (e.g. decision maker) *n*, and β_{ki} is a parameter associated with the kth predictor variable and alternative *i*. Thus, as can be seen from the *k*, *i* and *n* indexes, in the MNL model, there is a separate parameter β_i for every alternative *i* in the choice set per predictor variable (including a constant). Note that the variables X_{kn} vary only across the decision makers *n*, but not across the alternatives (they have no *i* subscript). Characteristics of the alternatives do not explicitly play a role in this model (implicitly they do, as we would expect the β_{ki} to depend on characteristics of the alternatives).

Conditional Logit Model

In the CL model,

$$V_{ni} = {}^{2}\,'\,\mathbf{X_{ni}} = \sum_{k=1}^{K} \beta_k X_{kni} \tag{21.9}$$

where K is again the number of predictor variables in the model, X_{kni} is the value of the k^{th} predictor variable for observational unit (e.g. decision maker) *n* and alternative *i*, and a ${}^{2}{}_{k}$ is a parameter associated with the k^{th} predictor variable. Thus, as can be seen from the *k*, *i* and *n* indexes, in the CL model, there is only a single parameter for all alternatives in the choice set per predictor variable.

Note that the variables X_{kni} vary across the decision makers *n* and alternatives *i*. Essentially, going from equation 21.8 to equation 21.9, the *i* index (that references alternatives) moves from the parameter ^{2} to the predictor variable *X*, a manifestation of the fact that in the MNL model characteristics of alternatives are implicitly included in the estimated alternative-specific parameters, while in the CL model they are explicitly measured and their effects estimated in generic parameters.

Probabilities in the Multinomial and Conditional Logit Models

Substituting equation 21.8 into equation 21.7, the multinomial logit probability that decision maker *n* chooses alternative *i* is:

$$P_{ni} = \frac{e^{\sum_{k=1}^{K} \beta_{ki} X_{kn}}}{\sum_{j=1}^{J} e^{\sum_{k=1}^{K} \beta_{kj} X_{kn}}} \tag{21.10}$$

Note that in equation 21.10, the predictor variables vary across decision makers *n* but not across alternatives *i*. Analogously, substituting equation 21.9 into equation 21.7, the conditional logit model asserts that the probability that decision maker *n* chooses alternative *i* is:

$$P_{ni} = \frac{e^{\sum_{k=1}^{K} \beta_{k} X_{kni}}}{\sum_{j=1}^{J} e^{\sum_{k=1}^{K} \beta_{k} X_{knj}}} \tag{21.11}$$

Note that in equation 21.11, it is impossible to estimate effects of attributes of the decision maker that do not vary across alternatives (such as age or gender), because such variables (and their parameters) automatically cancel out of the equation; the characteristic of the decision maker cannot affect which alternative is chosen because those characteristics do not vary across the alternatives. This feature differentiates the conditional logit model from the multinomial logit model.

It is possible, though, to estimate the interaction of characteristics of decision makers and characteristics of alternatives by creating or measuring variables that vary across both alternatives and decision makers. Such variables must have the n and the i subscript so that they do not cancel out in the equation.

An example of a measured interaction is the experience that the decision maker has with each of the alternatives. A given decision maker has more experience with one alternative than others, and, therefore, is more or less likely to choose the alternative. Repeat and near repeat victimization may be examples of this. Another example of a measured interaction is the distance between decision makers and alternatives.

An example of a created interaction is the multiplication of a characteristic of decision makers (e.g. gender S_n) with a characteristic of alternatives (e.g. location L_i) resulting in SL_{ni} . The resulting variable varies across decision makers (as for a given alternative i its value is different for males and females) and across alternatives (because for a given decision maker n it varies across locations).

Data Structures

Although the same mathematical model underlies the MNL model and the CL model, the estimation of the CL model requires the data to be organized differently than the estimation of the MNL model. This section considers the data structures that hold the information that is required to estimate either model.

The MNL model applies to a $n \times k$ matrix (where n refers to cases and k refers to variables that vary across cases), while the CL model applies to a $n \times i \times k$ matrix, where n refers to cases, i refers to alternatives, and k refers to variables that vary across cases and across alternatives. The distinctions between these two data structures are explained below.

The Multinomial Logit Model

The MNL model is estimated on a data set that is similar to the data structure of most other regression models and many incident spreadsheets. Each row (record) represents an observational unit *n* (a case, sometimes a decision maker) and each column represents a variable (a characteristic of the unit). The dependent variable is nominal and indicates which alternative from a set of alternatives was chosen. The variation in outcomes is explained by variation in the characteristics of the observational units (decision makers). Table 21.1 shows a simple example that describes the first 5 incidents of a larger data file. For each case we know the area where the offender lived (*Origin*), the area where the crime was committed (*Destination*), the offender's age (*Age*), the type of crime committed (*CrimeType*), and the time of day it was committed (*Time*). There is also a variable that uniquely identifies cases (*ID*).

The first record indicates that at 3AM (*Time*) a burglary (*CrimeType*) was committed in zone P (*Destination*) by an 18 year old (*Age*) offender who lived in zone P (*Origin*). Case 2 is a robbery committed at 7 PM in zone P by an offender aged 23 living in zone Q. The third record shows that someone aged 42 living in zone R purchased an illicit drug in zone S at 2pm.

Table 21.1:

Case file Describing 5 Incidents

ID	*Origin*	*Destination*	*Age*	*CrimeType*	*Time*
1	P	P	18	Burglary	3am
2	Q	P	23	Robbery	7pm
3	R	S	42	Illicit drug	2pm
4	R	Q	32	Robbery	1pm
5	S	R	19	Burglary	6am

In principle, any variable (except ID) in the case file can be analyzed as representing a choice outcome (an alternative being chosen) although for some variables a choice interpretation is more natural than for others. *Destination* represents the choice outcome of the decision of where to commit the crime, *CrimeType* would be the choice outcome of the decision which type of crime to commit, and *Time* would be the choice outcome of the decision of when to commit the crime. *Origin* could also be a choice outcome, the outcome of the decision of where to live. *Age* can be a seen as the outcome of the choice at what age to commit the offence.

If the decision to be analyzed is where to commit the offence, the first record in Table 21.1 indicates that the offender offended in zone P rather than in zones Q, R or S. If the decision to be analyzed is which type of crime to commit, the first record in Table 21.1 shows that the offender decided to commit a burglary rather than commit a robbery or purchase an illicit drug. If the decision to be analyzed is when to commit the offence, the first record in Table 21.1 indicates that the offender offended at 3am rather than at 1pm, 3pm or any other time of the day.

Let us assume that in Table 21.1, the outcome variable is *Destination* (i.e. the area in which the offender committed the crime). Note that we assume that each offender was able to choose any of the alternatives (zones P=1, Q=2, R=3, and S=4 for four alternatives), and also note that the data do not contain attributes of the alternatives (e.g. whether the areas are affluent, have mixed land use, etc.). A MNL model could be used to assess the relation between linear combinations of *Time* (T) and *Age* (A) with the choice of a *Destination* (D) zone. In this case, equation 21.12 becomes:

$$P_n(D=i)=\frac{e^{\beta_i+\beta_{Ti}T_n+\beta_{Ai}A_n}}{\sum_{j=1}^{4}e^{\beta_j+\beta_{Tj}T_n+\beta_{Aj}A_n}} \tag{21.12}$$

where $P_n(D=i)$ is the probability that in the in n^{th} case the *Destination* chosen is i, T_n is the *Time* of the n^{th} case, A_n is the *Age* of the n^{th} case, and β_{Ti} is the parameter that represents the effect of *Time* on the probability that *Destination i* is chosen, β_{Ai} is the effect of *Age* on the probability that *Destination i* is chosen, and β_i is an alternative-specific constant, representing the average attractiveness of alternative *i* in the sample. Note that if *Destination* has four categories, the multinomial logit model involves the following four categorical equations.

$$P_n(D=1)=\frac{e^{\beta_1+\beta_{T1}T_n+\beta_{A1}A_n}}{\sum_{j=1}^{4}e^{\beta_j+\beta_{Tj}T_n+\beta_{Aj}A_n}} \tag{21.13}$$

$$P_n(D=2)=\frac{e^{\beta_2+\beta_{T2}T_n+\beta_{A2}A_n}}{\sum_{j=1}^{4}e^{\beta_j+\beta_{Tj}T_n+\beta_{Aj}A_n}} \tag{21.14}$$

$$P_n(D=3)=\frac{e^{\beta_3+\beta_{T3}T_n+\beta_{A3}A_n}}{\sum_{j=1}^{4}e^{\beta_j+\beta_{Tj}T_n+\beta_{Aj}A_n}} \tag{21.15}$$

$$P_n(D=4)=\frac{e^{\beta_4+\beta_{T4}T_n+\beta_{A4}A_n}}{\sum_{j=1}^{4}e^{\beta_j+\beta_{Tj}T_n+\beta_{Aj}A_n}} \tag{21.16}$$

All four equations are linked by having the same denominator and by

$$P_n(D=1)+P_n(D=2)+P_n(D=3)+P_n(D=4)=1 \tag{21.17}$$

Altogether, 12 parameters are estimated, 4 alternative-specific constants (β_1, β_2, β_3, β_4), 4 for the Time predictor variable (β_{T1}, β_{T2}, β_{T3}, β_{T4}), and 4 for the Age predictor variable (β_{A1}, β_{A2},

$^{2}{}_{A3}$, $^{2}{}_{A4}$). However, because the effects apply to the differences between the alternatives, the parameters for one of the J alternatives must be fixed, and the remaining effects are expressed in relation to this fixed 'reference' alternative. Like other programs, CrimeStat fixes these parameters of the reference alternative to 0 (the user can choose which alternative to use as the reference alternative. By default the most frequent alternative is the reference alternative.

Example 1: Modeling Choice of Premises in Chicago Non-street Robberies with the Multinomial Logit Model

In 1997, there were 1,587 robbery incidents in Chicago that did not occur on the street, in which a specific type of premises was robbed, and for which at least one offender was arrested. In 1998, there were 1,441 such incidents. In this example, characteristics of offenders and incidents will be used to describe differences in the type of premises victimized. The statistics used to differentiate models will be explained and the premise pattern of robberies in 1998 will be predicted using the robbery patterns of 1997.

Figure 21.1 maps the premises type of non street robberies in 1997. In 1997, 48.6% of these robberies were residential, 11.3% were in parking lots and garages, 23.8% were commercial, 2.1% were at banks or currency exchanges, 5.5% were in schools and school yards, 5.1% were in parks, and 3.5% were in public transit or stations. Although parks are amenities, not premises, they will be subsumed under 'premises ' here.

Some areas of the city are nearly free of non-street robberies. Unsurprisingly, commercial, and bank robberies are concentrated on main streets. Residential robberies are widespread over large sections of the west and south sides. The remainder of this brief will look at crime and offender characteristics that differentiate residential robberies from each of the other premises types using the multinomial logit model in CrimeStat IV.

In Table 21.2, 1,587 non-street robberies in 1997 are analyzed using the multinomial logit model. Residential robberies are compared to 6 other robbery premises. In the summary section of the table, the log likelihood ratio (LLR), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC/SC) are measures of the differences between a model that includes offender and crime characteristics and the naïve (null) hypothesis that only includes the frequency of the various premises types.

The best use for these statistics is in comparing models. The most negative log likelihood ratio, and the smallest positive AIC or BIC are best. Unlike the LLR, the AIC and BIC correct for the number of explanatory (independent) variables. This is important because a model with

Figure 21.1:

Distribution of Chicago Non Street Robberies in 1997

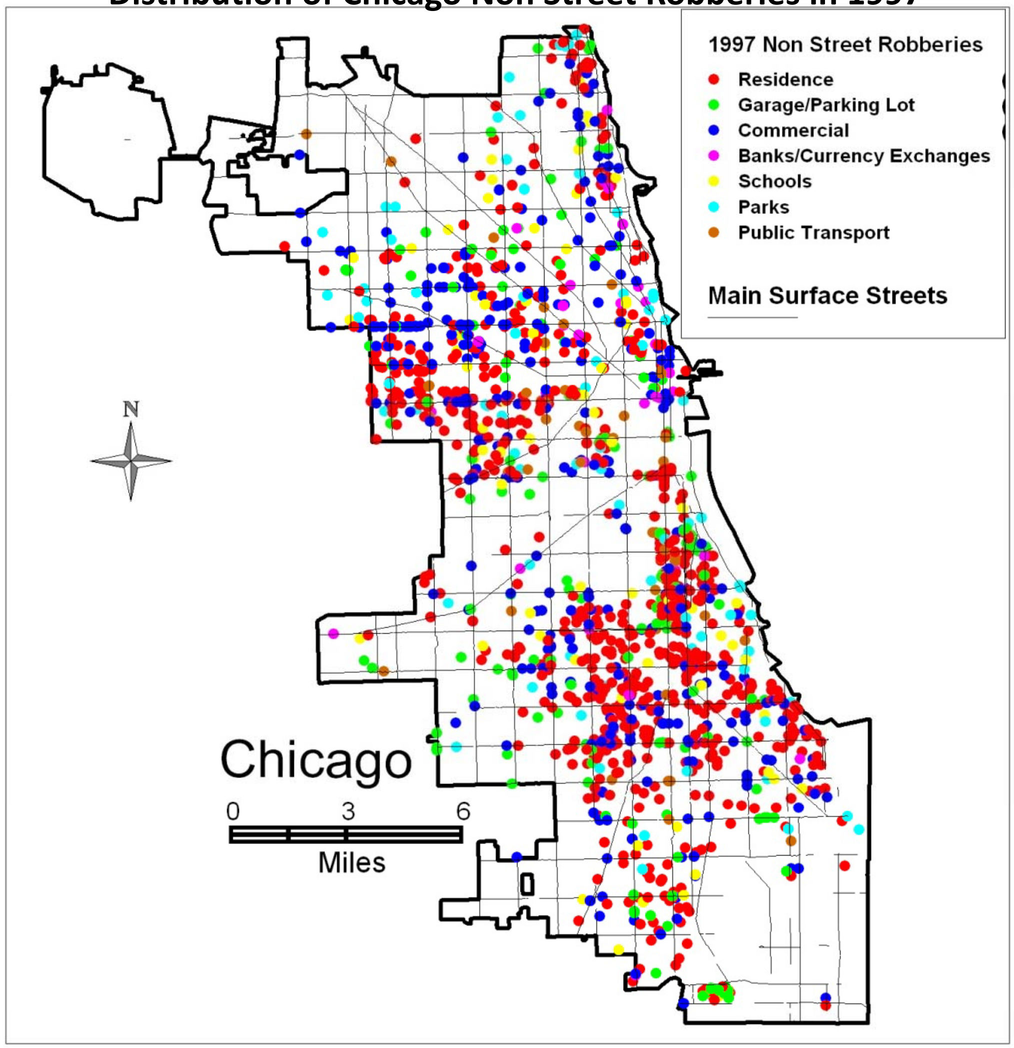


Table 21.2:

Multinomial Logit Model of Crime Premises: Non-Street Robbery 1997

Model result:	
Data file:	1997 CHICAGO NON-STREET ROBBERIES
DepVar:	TYPE OF PREMISES
N:	1,587
Df:	1,580
Type of choice model:	Multinomial logit model
Number of Alternatives:	7
Method of estimation:	MLE
Likelihood statistics	
Log Likelihood:	-1,963.1
Per case:	-1.2
AIC:	3,996.3
Per case:	2.5
BIC/SC:	4,184.2
Per case:	2.6
Model error estimates	
Mean absolute deviation:	0.2
Mean squared predicted error:	0.1

REFERENCE CHOICE: 2 RESIDENTIAL

Predictor	Coefficient	Stand Error	t-value	p-value	Odds Ratio
3 GARAGES AND PARKING LOTS					
Alternative N=180					
Constant	-1.3156	0.250	-5.27	0.001	0.27
GUNCRIME	0.2768	0.187	1.48	n.s.	1.32
EVENING	0.4431	0.193	2.30	0.05	1.56
LATENIGHT	-0.3754	0.235	-1.60	n.s.	0.69
TRAVEL DIST	0.0059	0.001	4.43	0.001	1.01
OFFAGE	-0.0016	0.001	-1.88	n.s.	1.00
OFFBLACK	-0.4792	0.237	-2.02	0.05	0.62
4 COMMERCIAL					
Alternative N=378					
Constant	-0.6512	0.197	-3.31	0.001	0.52
GUNCRIME	1.3900	0.137	10.12	0.001	4.01
EVENING	-0.0614	0.163	-0.38	n.s.	0.94
LATENIGHT	-0.5360	0.180	-2.97	0.01	0.59
TRAVEL DIST	0.0049	0.001	4.23	0.001	1.00
OFFAGE	-0.0018	0.001	-2.66	0.01	1.00
OFFBLACK	-0.6909	0.186	-3.71	0.001	0.50

Table 21.2: (continued)

Predictor	Coefficient	Stand Error	t-value	p-value	Odds Ratio
5 BANKS AND CURRENCY EXCHANGES					
Alternative N=34					
Constant	-2.2444	0.411	-5.46	0.001	0.11
GUNCRIME	1.3043	0.362	3.60	0.001	3.69
EVENING	-1.5034	0.620	-2.43	0.05	0.22
LATENIGHT	-1.9124	0.742	-2.58	0.01	0.15
TRAVEL DIST	0.0083	0.002	3.51	0.001	1.01
OFFAGE	-0.0022	0.002	-1.09	n.s.	1.00
OFFBLACK	-1.4572	0.395	-3.69	0.001	0.24
6 SCHOOLS					
Alternative N=88					
Constant	6.6012	0.783	8.43	0.001	735.99
GUNCRIME	-2.0686	0.612	-3.38	0.001	0.13
EVENING	-2.1280	0.480	-4.44	0.001	0.12
LATENIGHT	-2.2797	0.750	-3.09	0.01	0.10
TRAVEL DIST	0.0065	0.003	2.30	0.05	1.01
OFFAGE	-0.3692	0.040	-9.17	0.001	0.69
OFFBLACK	-1.1614	0.382	-3.04	0.01	0.31
7 PARKS					
Alternative N=81					
Constant	3.1511	0.573	5.50	0.001	23.36
GUNCRIME	-0.6429	0.328	-1.96	0.05	0.53
EVENING	0.0558	0.282	0.20	n.s.	1.06
LATENIGHT	-1.1378	0.459	-2.48	0.05	0.32
TRAVEL DIST	0.0070	0.002	3.14	0.01	1.01
OFFAGE	-0.1966	0.025	-7.90	0.001	0.82
OFFBLACK	-1.2570	0.311	-4.05	0.001	0.28
8 PUBLIC TRANSPORT					
Alternative N=55					
Constant	-3.4423	0.617	-5.58	0.001	0.03
GUNCRIME	-0.6316	0.395	-1.60	n.s.	0.53
EVENING	-0.4296	0.409	-1.05	n.s.	0.65
LATENIGHT	-0.0134	0.337	-0.04	n.s.	0.99
TRAVEL DIST	0.0091	0.002	5.21	0.001	1.01
OFFAGE	-0.0010	0.001	-0.88	n.s.	1.00
OFFBLACK	0.6914	0.608	1.14	n.s.	2.00

Reference Alternative: 2 Residential

Multicollinearity statistics

Predictor	Pseudo-Tolerance
GUNCRIME	0.98
EVENING	0.93
LATENIGHT	0.93
TRAVEL DIST	0.99
OFFAGE	1.00
OFFBLACK	1.00

many explanatory variables is likely to have the most negative log likelihood ratio (LLR), but part of the size of the LLR results from the large number of variables used in the explanation.

In the second section of Table 21.2, each of the other types of premises is compared to the reference type (residential units), which is the type that is chosen most frequently. Note that the coefficients will differ for each of the alternatives. This is because the variables predicting each alternative are unique and will differ in their weights. For some alternatives, an independent variable may have a significant positive effect while for other alternatives it may have a significant negative effect.

For even other alternatives, the variable may not have a significant effect. For example, the use of a gun in a robbery (GUNCRIME) is positively associated with bank robberies but negatively associated with school robberies. For robberies in parks, the use of a gun is not related to the type of robbery. Note, also, that these are relative to the reference alternative, which in this case are residential robberies.

The numbers in the far right column, the Odds Ratios, are useful for substantive interpretation of the model. They indicate the odds increase or decrease associated with the variable that the robbery took place on the specific premise compared to the reference alternative (residential premises). They are measured as the relative change when the corresponding predictor variable increases by one unit. Odds ratios above 1 indicate that the odds increase as the predictor variable increases, odds ratios between 0 and 1 indicate they decrease as the predictor variable increases.

The p value indicates whether the odds ratio were likely to have occurred by chance if there was no relationship in the unit of the explanatory variable. For example, in the top panel on ‘Garages and parking lots’, the value of 1.56 indicates that-- if a robbery occurs in the evening (1) it is 1.56 times more likely (or, in other words, 56% more likely) to be in a garage or parking lot than at a residence. The p value indicates whether the odds ratio could have occurred by chance if there was no relationship. Thus, the p-value of 0.05 in the output demonstrates that the above odds ratio of 1.56 could have occurred by chance 5% of the time if there was no time-of-day difference in the probability that robberies take place in residences or in parking lots.

Commercial robberies are 4 times as likely to be committed with a gun than residential robberies, and a difference this large could occur only .1% of the time. Robberies occurring in and around schools are significantly different from residential robberies on all six explanatory variables. They are much less likely to involve guns or be committed by black offenders and are slightly further away from the offender’s home. Unsurprisingly, they are all less likely to occur in the evening or late night and the offenders are younger than in residential robberies.

The coefficients column in section two are similar to those in any regression equation. Coefficients are created for each choice (here premises type) including the reference category. They are particularly useful for prediction of choice with a new data set(see below).

The third section indicates to what extent the explanatory variables vary together (multicollinearity). A pseudo tolerance below .90 indicates that this may be a problem in the model. If this is so, delete the variable with the lowest pseudo tolerance and run the model again. In this model all pseudo tolerances are above .9. Multicollinearity is not a problem.

Adding another variable to the 1997 model

Using the Log Likelihood, AIC, and BIC/SC statistics, it is possible to compare one decision making model to another. The decision making model in table 21.2 included six explanatory variables. Table 21.3 below adds the variable, number of offenders, to the model. Perhaps residential robbers are more likely to solo offenders than school yard robbers? However, adding the number of offenders to the model has little effect.

The more explanatory variables, the fewer degrees of freedom (df) and the more complex the model. The log likelihood decreases from -1963 to -1957 (more negative is better). The AIC declines slightly from 3996 to 3994, but the more comprehensive BIC/SC increases from 4184 to 4209 (closer to 0 is better for both). In other words, adding number of offenders to the model does not improve the differentiation of residential premises from other premises. For every premises type, the number of offenders is not significantly differentiated from residential robberies.

Predicting non-street Robberies in 1998 based on the 1997 model

Once a multinomial logit model is estimated, the parameter estimates can be used to predict a dependent variable in other data. The model developed in predicting the premises of robberies in 1997 can be used to predict the premises of robberies in 1998. This is done by saving the coefficients and applying them to the 1998 robbery data. The results can show how well the 1997 model estimated predicted 1998 robberies.

Table 21.3:

Multinomial Logit Model of Crime Premises: Non-Street Robbery 1997

Number of Offenders Added

```
Model result:
 Data file:                                                        1997 CHICAGO NON-STREET
ROBBERIES
 DepVar:                            TYPE OF PREMISES
 N:                                 1,587
 Df:                                1,579
 Type of choice model:              Multinomial logit model
 Number of Alternatives:            7
 Method of estimation:              MLE

      Likelihood statistics
 Log Likelihood:                    -1957.2
 AIC:                               3,994.3
 BIC/SC:                            4,209.1

      Model error estimates
 Mean absolute deviation:           0.2
 Mean squared predicted error:      0.1
```

REFERENCE CHOICE: 2 RESIDENTIAL

Predictor	Coefficient	Stand Error	t-value	p-value	Odds Ratio
3 GARAGES AND PARKING LOTS					
Alternative N=180					
Constant	-1.2044	0.302	-3.98	0.001	0.30
GUNCRIME	0.3039	0.191	1.59	n.s.	1.36
EVENING	0.4481	0.193	2.32	0.05	1.57
LATENIGHT	-0.3608	0.236	-1.53	n.s.	0.70
TRAVEL DIST	0.0059	0.001	4.46	0.001	1.01
OFFAGE	-0.0016	0.001	-1.90	n.s.	1.00
OFFBLACK	-0.4807	0.237	-2.02	0.050	0.62
NUM OFF	-0.1011	0.155	-0.65	n.s.	0.90
4 COMMERCIAL					
Alternative N=378					
Constant	-0.7871	0.221	-3.55	0.001	0.46
GUNCRIME	1.3501	0.140	9.61	0.001	3.86
EVENING	-0.0709	0.163	-0.43	n.s.	0.93
LATENIGHT	-0.5702	0.183	-3.12	0.01	0.57
TRAVEL DIST	0.0048	0.001	4.13	0.001	1.00
OFFAGE	-0.0017	0.001	-2.62	0.01	1.00
OFFBLACK	-0.6908	0.186	-3.71	0.001	0.50
NUM OFF	0.1261	0.093	1.36	n.s.	1.13

Table 21.3: (continued)

Predictor	Coefficient	Stand Error	t-value	p-value	Odds Ratio
5 BANKS AND CURRENCY EXCHANGES					
Alternative N=34					
Constant	-1.2864	0.715	-1.80	n.s.	0.28
GUNCRIME	1.4098	0.365	3.86	0.001	4.09
EVENING	-1.4180	0.621	-2.28	0.05	0.24
LATENIGHT	-1.7920	0.743	-2.41	0.05	0.17
TRAVEL DIST	0.0085	0.002	3.56	0.001	1.01
OFFAGE	-0.0023	0.002	-1.12	n.s.	1.00
OFFBLACK	-1.4442	0.396	-3.64	0.001	0.24
NUM OFF	-0.9035	0.560	-1.61	n.s.	0.41
6 SCHOOLS					
Alternative N=88					
Constant	6.8782	0.873	7.87	0.001	970.84
GUNCRIME	-2.0414	0.614	-3.32	0.001	0.13
EVENING	-2.1276	0.479	-4.44	0.001	0.12
LATENIGHT	-2.2976	0.752	-3.06	0.01	0.10
TRAVEL DIST	0.0066	0.003	2.35	0.05	1.01
OFFAGE	-0.3716	0.040	-9.19	0.001	0.70
OFFBLACK	-1.1790	0.381	-3.09	0.01	0.31
NUM OFF	-0.1896	0.279	-0.68	n.s.	0.83
7 PARKS					
Alternative N=81					
Constant	2.7911	0.618	4.52	0.001	16.30
GUNCRIME	-0.7293	0.337	-2.17	0.05	0.48
EVENING	0.0657	0.283	0.23	n.s.	1.07
LATENIGHT	-1.1658	0.461	-2.53	0.05	0.31
TRAVEL DIST	0.0068	0.002	3.05	0.01	1.01
OFFAGE	-0.1946	0.025	-7.83	0.001	0.82
OFFBLACK	-1.2409	0.312	-3.98	0.001	0.30
NUM OFF	0.2586	0.177	1.46	n.s.	1.30
8 PUBLIC TRANSPORT					
Alternative N=55					
Constant	-3.0490	0.721	-4.23	0.001	0.05
GUNCRIME	-0.5504	0.400	-1.38	n.s.	0.58
EVENING	-0.4107	0.409	-1.00	n.s.	0.66
LATENIGHT	0.0074	0.338	0.02	n.s.	1.01
TRAVEL DIST	0.0092	0.002	5.28	0.001	1.01
OFFAGE	-0.0010	0.001	-0.90	n.s.	1.00
OFFBLACK	0.6919	0.608	1.14	n.s.	2.00
NUM OFF	-0.3629	0.348	-1.04	n.s.	0.70

Reference Alternative: **2 RESIDENTIAL**

Table 21.3: (continued)

```
---------------------------------------------------------------------------
Multicollinearity statistics
Predictor           Pseudo-Tolerance
GUNCRIME            0.94
EVENING             0.92
LATENIGHT           0.92
TRAVEL DIST         0.98
OFFAGE              0.99
OFFBLACK            1.00
NUM OFF             0.94
```

In Table 21.4, the percentage distribution of the 7 premises types is compared for 1997 and 1998 with the 1998 percentage correctly predicted for each type of premises using the MNL equation developed for 1997 robberies. Overall, not much has changed between the two years.

Table 21.4:

Non-Street Robberies in 1997 & 1998
1998 Predicted by the 1997 Multinomial Logit Model

Type of Premises	**1997**	**1998**	**Percent Correctly Predicted for 1998**
Residential	48.6%	47.8%	53.0%
Garages/Parking	11.3%	10.5%	12.6%
Commercial	23.8%	25.4%	33.2%
Banks/CurrEx	2.1%	3.7%	4.8%
Schools	5.5%	5.3%	39.2%
Parks	5.1%	4.0%	11.7%
Public Transit	3.5%	3.3%	5.0%
Number of Robberies	1587	1441	

In order to be an improvement on the naïve assumption that the percentage of incidents at each premises type is no better than the overall distribution of premises in 1998, the multinomial logit model based on 1997 (Column 3) should predict the premises of incidents better than the marginal percentage distribution of incidents in 1998 (Column 2). It does for all premise types. A few examples:

1. 47.8% of incidents were residential with the model correctly predicting 53.0% percent of them.
2. 25.4% of incidents were commercial with the model correctly predicting 33.2% percent.

3. 5.3% of incidents were in schools or school yards and the model correctly predicted 39.2% percent.
4. Garages and parking lots were only slightly better predicted by the model than by the 1998 percentage distribution. 10.5% of incidents were in garages or parking lots, and the model correctly predicted 12.6% of these.

Example 1 Conclusion

When an offender chooses a type of premise to commit a robbery, the choice is not random. Personal characteristics such as age and racial group make a difference, but so do decisions that the offenders make when coming into the incident such as gun availability, distance from home, and time of day. This example demonstrates how Multinomial Logit models can be used to clarify the offender's choice by the type of premise. The example also demonstrates that a model based on robbery choices made in one year can be useful in prediction of robberies in another year.

Another example of the Multinomial Logit model is presented in the attachment where Levine, Robertson and Fosberg analyze the type of weapon used in Houston robberies.

The Conditional Logit Model

The CL model is estimated on a different data structure. It is a matrix where each row (record) represents a combination of an observational unit *n* (a case, often a decision maker) with an alternative in the choice set *i*, and where each column represents a variable (a characteristic of the observational unit and/or the alternative). In this case, each record represents a possible alternative that the case (or decision maker) is presented with. The dependent variable is a binomial variable and indicates which alternative *i* from a set of alternatives was chosen by observational unit *n*.

For example, a community is divided into twenty neighborhoods (alternatives). Each of these is classified according to number of businesses, wealth, racial makeup and population size (5 variables). For each case, an offender 'selects' a neighborhood where the crime is committed (choice).[1] For 100 cases and twenty alternatives, a matrix of 2,000 records and a minimum of six variables would be necessary. The sixth variable identifies the chosen alternative. The number of records can grow quickly. The following is a simplified example. The variation in outcomes is explained by variation in the characteristics of the alternatives. CrimeStat is able to construct such a file by combining a 'case file' and an 'alternatives file'. Below we present a simplified description of the process.

---

[1] The offender may not do this rationally, of course, and may simply be at that location (a routine activity). Nevertheless, the model assumes that the offender has made a utility calculation to commit the crime at the location. To that extent, it is a decision among many alternatives.

Destination Choice

We start with the case file shown in Table 21.1, that is the file that is used for estimating a multinomial logit model, and by a model of the *Destination* (i.e.. in which area, P, Q, R or S, did the offender commit the crime?). Whereas a MNL model is used to assess whether linear combinations of characteristics associated with the cases (e.g.,*Origin, Age, Time* and *CrimeType*) predict which zones the offender selects to commit the crime, a CL model is used to assess whether characteristics of the alternatives predict the zone chosen. The alternatives are the zones themselves and, obviously, additional information is needed on the alternatives.

Table 21.5 shows part of an example file, containing the four alternative destination areas P, Q, R and S for the decision on where to offend. The variables include an identifier (Zone), the percentage of the household below a poverty threshold (*Poverty*) and the percentage of the non-residential land use (*Non-Residential*) in the zone.

Table 21.5:
Zone File Describing 4 Alternative Zones

Zone	*Poverty*	*Non-Residential*
P	2	40
Q	2	16
R	4	23
S	5	12

The data structure required for estimation of the CL model represents all possible combinations of the rows in the case file and the rows in the zone file, including the variables in both files. It also includes for each decision maker a binomial variable indicating the alternative that was chosen by the decision maker. For example, if there are 200 cases (decision makers) and 7 alternatives that are available, there will be 1,400 records (200 x 7) in the data set. Each decision maker will be represented 7 times, representing each of the 7 alternatives that the decision maker is confronted with. However, the decision maker will have selected only one of these alternatives. For that record, the value of the binomial choice variable will be 1; for the other six records, the value of the binomial choice variable will be 0.

To go back to the example, Table 21.6 displays the combination of Tables 21.1 and 21.5. Note that the columns 1-6 of Table 21.6 are a copy of Table 21.1 with each row repeated four times (the first original row in rows 1-4). Also verify that columns 7-9 are copies of Table 21.5, with each row repeated five times (the first original row in rows 1, 5, 9, 13 and 17, the second original row in rows 2, 6, 10, 14 and 18, etc.). Finally note that the indicator variable,

Table 21.6:
Case-alternative File Describing 20 Case-Alternative Combinations

ID	Org	Dest	Age	CrimeTyp	Time	Zone	Pov	NonRes	Chosen	Home
1	P	P	18	Burglary	3am	P	2	40	1	1
1	P	P	18	Burglary	3am	Q	2	16	0	0
1	P	P	18	Burglary	3am	R	4	23	0	0
1	P	P	18	Burglary	3am	S	5	62	0	0
2	Q	P	23	Robbery	7pm	P	2	40	1	0
2	Q	P	23	Robbery	7pm	Q	2	16	0	1
2	Q	P	23	Robbery	7pm	R	4	23	0	0
2	Q	P	23	Robbery	7pm	S	5	62	0	0
3	R	S	42	Illicit drug	2pm	P	2	40	0	0
3	R	S	42	Illicit drug	2pm	Q	2	16	0	0
3	R	S	42	Illicit drug	2pm	R	4	23	0	1
3	R	S	42	Illicit drug	2pm	S	5	62	1	0
4	R	Q	32	Robbery	1pm	P	2	40	0	0
4	R	Q	32	Robbery	1pm	Q	2	16	1	0
4	R	Q	32	Robbery	1pm	R	4	23	0	1
4	R	Q	32	Robbery	1pm	S	5	62	0	0
5	S	R	19	Burglary	6am	P	2	40	0	0
5	S	R	19	Burglary	6am	Q	2	16	0	0
5	S	R	19	Burglary	6am	R	4	23	1	0
5	S	R	19	Burglary	6am	S	5	62	0	1

Chosen, is set to 1 if the value in variable *Destination* matches the value in variable *Zone*. The variable *Home* will be discussed below.

Note that in Table 21.6, the zone characteristics *Pov* and *Nonres* only vary across alternatives but not across cases (decision makers): the values of these two variables are just repeated in every case. Quite often, however, the model includes variables that vary across alternatives and across cases as well. The last column in Table 21.6 contains a binomial variable, *Home*, that indicated whether an alternative zone is the zone of residence of the offender. Thus, it has value 1 if *Origin=Zone*, and 0 otherwise. This variable varies both across alternatives (e.g. for a given ID, one alternative equals 1 and the other equal 0) and across cases (for a given Zone, say A, it equals 1 for case 1, but 0 for cases 2-5). In a similar fashion (but more difficult to verify by just inspecting the table), we could define a new variable that represents the distance between the alternative zone and the zone of the offender's residence.

Note also that the indexing in the independent variables in the MNL and CL models reflects the data structures used for estimation. In the MLN model, $V_{in} = {}^{2}{}_{i}X_{n}$, where the index n in the term X_n indicates that the variables only vary between cases (decision makers), so that we only need one row per case. In the CL model, $V_{in} = {}^{2}X_{ni}$, where the indices n and i in the term X_{ni} reflect that the variables can vary between cases (decision makers) and between alternatives, so that we need multiple rows case (equation 21.11 demonstrated that in the CL model, the variables must vary across alternatives and may vary across decision makers, but cannot be estimated when they vary across decision makers only).

Crime Type Choice

Now let us consider another type of choice: the choice of a crime type. An offender urgently needing money may have to choose a criminal activity that generates the required amount as easily and with as little risk as possible. If we assume that burglary, robbery and illicit drug dealing are the available alternatives, an alternatives file could look like Table 21.7. The variables in this file represent attributes that may differentiate between the crime types: *Expected Profits*, *Detection Risk* and *Time needed to search and attack a target* and that may affect the attractiveness of these offences to the offenders.

Table 21.7:

Crime Type File

(Alternative Crime Types)

Crime type	*Expected Profits*	*Detection Risk*	*Sanction Severity*	*Time Needed*
Burglary	200	.07	3	60
Robbery	50	.15	5	20
Illicit drug	20	.02	2	40

Analogously to the case of destination zone choice, the data structure required for estimation of the CL model represents all possible combinations of the rows in the case file (Table 21.1) and the rows in the alternatives file (Table 21.6), including the variables in either file, and also including for each decision maker a binomial variable indicating the alternative (which crime type) that was chosen by the decision maker.

Table 21.8 displays the combination of Tables 21.1 and 21.7. The first six columns of Table 21.8 are a copy of Table 21.1 with each row repeated three times. Also verify that column 7-9 are copies of Table 21.3, with each row repeated four times (the first original row in rows 1,4,7, 10, and 13, the second original row in rows 2, 5, 8, 11 and 14, etc.). Finally note that the indicator variable *Chosen* is set to 1 if the value in variable *Type* matches the value in variable *CrimeType*.

Table 21.8:
Case-alternative File Describing 15 Case-alternative Combinations

ID	*Org*	*Dest*	*Age*	*CrimeType*	*Time*	*Type*	*Profit*	*Risk*	*Sanc*	*Time*	*Chosen*
1	P	P	18	burglary	3am	burglary	200	.07	3	60	1
1	P	P	18	burglary	3am	robbery	50	.15	5	20	0
1	P	P	18	burglary	3am	il. drug	20	.02	2	40	0
2	Q	P	23	robbery	7pm	burglary	200	.07	3	60	0
2	Q	P	23	robbery	7pm	robbery	50	.15	5	20	1
2	Q	P	23	robbery	7pm	il. drug	20	.02	2	40	0
3	R	S	42	il. drug	2pm	burglary	200	.07	3	60	0
3	R	S	42	il. drug	2pm	robbery	50	.15	5	20	0
3	R	S	42	il. drug	2pm	il. drug	20	.02	2	40	1
4	R	Q	32	robbery	1pm	burglary	200	.07	3	60	0
4	R	Q	32	robbery	1pm	robbery	50	.15	5	20	1
4	R	Q	32	robbery	1pm	il. drug	20	.02	2	40	0
5	S	R	19	burglary	6am	burglary	200	.07	3	60	1
5	S	R	19	burglary	6am	robbery	50	.15	5	20	0
5	S	R	19	burglary	6am	il. drug	20	.02	2	40	0

Example 2: Modeling Choice of Neighborhood for Residential Burglaries in The Hague with the Conditional Logit Model

The discrete spatial choice approach was first applied to criminal location choices by Bernasco & Nieuwbeerta (2005). This example uses CrimeStat to replicate their analysis of 548 cleared burglaries committed in the years 1996-2001 in the city of The Hague, the Netherlands, by solitary offenders (i.e., offenders who perpetrated the burglary without known accomplishes).

The discrete spatial choice model of burglary integrated journey-to-crime research (that focuses on distance traveled without considering other aspects of criminal location choice) and ecological research (that addresses variation in opportunities and target attractiveness, but ignores the distance offenders have to travel to reach the targets).

Bernasco & Nieuwbeerta distinguished 89 neighborhoods in The Hague, which served as the spatial units of analysis. They argued that neighborhoods would be attractive for burglary if they (1) were affluent, (2) had a large proportion of single-family dwellings, (3) had high population turnover (4) had high ethnic heterogeneity, (5) had large numbers of households, (6) were situated relatively close to the city center and (7) were located relatively close to the offender's residence. Note that the first six criteria are the attributes of the 89 alternative neighborhoods (independently of any attributes of the burglar), while the last criterion (proximity to offender's home) depends on the locations of both the offender and the potential target neighborhoods.

In Table 21.9, the 548 The Hague burglaries are analyzed with the conditional logit model, using the following 7 variables as predictors of the burglars' selection of a target neighborhood:

1. PROPVAL. Average value of residential properties, in 100,000 euro
2. SINGFAM. Percentage of units that are single-family dwellings, in 10% units
3. RESMOBIL. Percentage of residents that moved during past year, in 10% units
4. ETNHETERO. (Ethnic Heterogeneity). Blau / Herfindahl index (x 10)
5. PROXIMITY. Negative distance between offender neighborhood and potential target neighborhood, in kilometers. The authors used negative distance instead of distance because this yielded a model in which all expected parameters were positive.
6. PROXCITY.Negative distance between city center and potential target neighborhood, in kilometers
7. RESUNITS. Number of residential properties in the neighborhood, in 1,000 properties

The results in Table 21.9 replicate the findings reported by Bernasco & Nieuwbeerta (2005, p. 308).[2] The summary section of the output reports general information about the model and the estimation procedure, including the names of the data file and the dependent variable. The output shows that the number of records is 48,772, which is 548 × 89 (i.e. the number of offenders multiplied by the number of The Hague neighborhoods). The number of degrees of freedom is 541 (the number of offenders -548, minus the number of estimated parameters -7). As discussed in the multinomial logit example, the likelihood statistics (Log Likelihood, Akaike Information Criterion, AIC ; and Bayesian Information Criterion, BIC) indicate how well the model fits the data (lower values indicate better fit). These statistics are only used to compare different models, and have no useful interpretation for a single model.

The coefficient section reports the results for each predictor variable and include the estimated coefficients, their standard errors, t-values, and p-values. The odds ratios column is the most useful statistic for substantive interpretation of the outcome. The odds ratio (which equals e^{β}) represents the factor by which the odds that a neighborhood is chosen for a burglary increases or decreases when the value of the predictor increases by one unit. An odds ratio greater than 1 indicates that the odds increase while an odds ratio between 0 and 1 indicate that the odds decrease.

For example, the odds ratio of 1.05 for variable PROPVAL indicates that as the average value of properties in the neighborhood increases by 100,000 euro, the odds that it is selected by a burglar increase by a factor 1.05 (i.e. by approximately 4.5 percent). Another example: the

[2] The standard errors reported here are slightly smaller than those reported by Bernasco & Nieuwbeerta (2005), a difference due to their correction of the standard errors for the possible interdependence among the burglaries (the 548 burglaries were committed by 290 unique persons; thus, some of them committed multiple burglaries).

estimated value of 1.67 for proximity means that if a neighborhood is located one kilometer closer to the offender's home, the odds that it will be selected by this burglar increase by a factor 1.67 (i.e. by approximately 67 percent).

Table 21.9:
Conditional Logit Model of Burglary Neighborhood Choice

```
Model result:
Data file:                         TheHagueBurglary.dbf
DepVar:                            CHOSEN
N:                                 48,772
Df:                                541
Type of choice model:              Conditional logit model
Number of Alternatives:            89
Method of estimation:              MLE

     Likelihood statistics
Log Likelihood:                    -2,203.3
AIC:                               4,420.6
BIC/SC:                            4,450.7

     Model error estimates
Mean absolute deviation:           0.02
Mean squared predicted error:      0.01
```

Predictor	Coefficient	Stand Error	Pseudo-Tolerance	t-value	p-value	Odds Ratio
PROPVAL	0.0445	0.112	0.33	0.40	n.s.	1.05
SINGFAM	0.1239	0.042	0.43	2.96	0.01	1.13
RESMOBIL	-0.0285	0.046	0.48	-0.62	n.s.	0.97
ETNHETERO	0.1380	0.032	0.37	4.37	0.001	1.15
PROXIMITY	0.5140	0.034	0.74	15.22	0.001	1.67
PROXCITY	-0.0812	0.049	0.37	-1.66	n.s.	0.92
RESUNITS	0.3039	0.029	0.80	10.61	0.001	1.36

Average Predicted Probability:

CHOSEN	Mean Probability	N	StdDev
0	0.011	48224	0.012
1	0.028	548	0.024
Total	0.011	48772	0.013

The section also includes the pseudo-tolerances of the indicator variable (see Chapters 15 and 17 for discussion of this statistic). If the tolerance of a variable is low, this indicates that the variable is strongly correlated with linear combinations of the other predictor variables in the equation, and that it therefore does not add much unique variability to the prediction of the dependent variable. This situation is called multicollinearity and is usually solved by removing the variable with the lowest tolerance from the equation. Note that three of the variables are not significant and several have low tolerances and that a simplified model can be produced by dropping them without much loss of generality (not shown).

The last section also lists average predicted probabilities for neighborhoods that were chosen (.028) and those that were not chosen (.011). Note that the average predicted probability multiplied by the total number of records yields the total number of burglary cases in the file.[3]

Conclusion

In discrete choice modeling, the dependent variable is made up of mutually exclusive and exhaustive categories. The category that is chosen is based upon characteristics of the decision maker (in the multinomial logit model), the characteristics of the alternatives (in the conditional logit model), or the interaction of the two (also in the conditional logit model). Interpretations of discrete choice models can be closely linked to the economic theory of utility maximization. Of all possible alternatives, the alternative is selected that maximizes gain and minimizes cost.

The CrimeStat discrete choice module is designed for regression when the dependent variable consists of unordered categories such as type of weapon or neighborhood where a crime is committed. This is in contrast to more traditional regression that is mainly concerned with dependent variables that are continuous or quasi-continuous, such as rates or counts. The Discrete Choice module is a multinomial extension of binomial logistic regression, discussed in Chapter 18, which allows for only two categories of the dependent variable.

The Discrete Choice module provides for two different models, the multinomial and the conditional logit model. Which one is used must be based upon the availability and relevance of data that reflect attributes of the categories and attributes of the cases (usually offenders, or crimes). To some extent it also depends upon the number of categories of the dependent variable since the tractability of the multinomial model decreases as the number of categories grows.

The conditional logit model is most appropriate if the outcome is assumed to be based on characteristics of the alternatives or their interaction with characteristics of the situation or the decision-maker. The CL data structure duplicates every possible alternative for each case and designates one as chosen. The results summarize the difference between the chosen selection and all others. For example, Chicago has seventy-seven neighborhoods that vary in terms of wealth, number of businesses, level of drug crime, and population. They also vary in distance from an offender's home. Each offender's decision about in which neighborhood to commit the

3 To do this accurately, one needs more than 3 decimal places. The CrimeStat output includes six decimal places. We have reduced the number of decimal places in the table to make it clearer.

crime is based upon a comparison of the characteristics of the 77 neighborhoods. The data file has one record for each alternative that the decision maker faces. If 1,000 offenders are analyzed, the resultant file would have 77,000 records.

The multinomial model may be appropriate if the choice has fewer categories and is dependent mainly on characteristics of the offender and the particular incident. A separate equation is constructed that compares a reference category with every other category of the dependent variable. For example, if weapon choice is dependent upon the victim's age and gender, type of target, and time of day, then a separate equation is constructed comparing gun incidents, the most frequent category, to knives, other weapons and strong armed. The data file contains one record for each offender.

In Chapter 22, we discuss the use of the CrimeStat discrete choice module routines to estimate these two models. Two additional routines are included in the discrete choice module. First, as discussed above the Conditional logit requires data organization that combines characteristics of the incident and all possible choices. CrimeStat will build this file for you. Second, both discrete and conditional models allow for prediction of dependent variables in one data set from the relationships found in another. Thus, in the example of Chicago robberies (above) 1998 robbery locations are predicted based on MNL coefficients of 1997 robberies.

If an analyst wants to consider the 'who', 'where', or 'why' of choice among multiple mutually exclusive possibilities and has a model in which criminals maximize the utility of their choices, then the either Conditional logit or Multinomial logit in the discrete choice module are appropriate techniques.

References

Ben-Akiva, M. E., & Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In R. W. Hall (Ed.), *Handbook of Transportation Science* (pp. 5-34). Norwell, MA: Kluwer.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Bernasco, W. (2010a). Modeling Micro-Level Crime Location Choice: Application of the Discrete Choice Framework to Crime at Places. *Journal of Quantitative Criminology, 26*(1), 113-138.

Bernasco, W. (2010b). A Sentimental Journey to Crime; Effects of Residential History on Crime Location Choice. *Criminology, 48*, 389-416.

Bernasco, W. (2007). The usefulness of measuring spatial opportunity structures for tracking down offenders: A theoretical analysis of geographic offender profiling using simulation studies. *Psychology, Crime & Law*, *13*, 155-171.

Bernasco, W. (2006). Co-Offending and the Choice of Target Areas in Burglary. *Journal of Investigative Psychology and Offender Profiling, 3*, 139-155.

Bernasco, W., & Block, R. (2009). Where Offenders Choose to Attack: A Discrete Choice Model of Robberies in Chicago. *Criminology*, *47*(1), 93-130.

Bernasco, W., & Kooistra, T. (2010). Effects of Residential History on Commercial Robbers' Crime Location Choices. *European Journal of Criminology*, *7*(4), 251-265.

Bernasco, W., & Nieuwbeerta, P. (2005). How Do Residential Burglars Select Target Areas? A New Approach to the Analysis of Criminal Location Choice. *British Journal of Criminology*, *45*, 296-315.

Clare, J., Fernandez, J., & Morgan, F. (2009). Formal Evaluation of the Impact of Barriers and Connectors on Residential Burglars' Macro-Level Offending Location Choices. *Australian and New Zealand Journal of Criminology*, *42*, 139-158.

Jepsen, L., & Jepsen, C. (2002). An empirical analysis of the matching patterns of same-sex and opposite-sex couples. *Demography, 39*(3), 435-453.

References (continued)

Krebs, J. R., & Davies, N. B. (1993). *An Introduction to Behavioural Ecology* (3th ed.). Oxford: Blackwell.

McFadden, D. (1980). Econometric Models for Probabilistic Choice Among Products. *The Journal of Business*, *53*(3), S13-S29.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

Palfrey, T. R., & Poole, K. T. (1987). The Relationship between Information, Ideology, and Voting Behavior. *American Journal of Political Science, 31*(3), 511-530.

Phillips, S. (2003). The Social Structure of Vengeance: A Test of Black's Model. *Criminology, 41*(3), 673-708.

Rengert, G. F. (1981). Burglary in Philadelphia: A Critique of an Opportunity Structure Model. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental Criminology* (pp. 189-202). Beverly Hills, CA: Sage.

Smith, T. S. (1976). Inverse Distance Variations for the Flow of Crime in Urban Areas. *Social Forces, 54*(4), 802-815.

Tita, G., & Griffiths, E. (2005). Traveling to Violence: The Case for a Mobility-Based Spatial Typology of Homicide. *Journal of Research in Crime and Delinquency, 42*, 275-308.

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). New York: Cambridge University Press.

Train, K. E. (1980). A Structured Logit Model of Auto Ownership and Mode Choice. *The Review of Economic Studies, 47*(2), 357-370.

Attachment A

Modeling Correlates of Weapon Use in Houston Robberies With the Multinomial Logit Model

Ned Levine
Ned Levine & Associates
Houston, TX

Alan Robertson
Houston Police Department
Houston, TX

Barry Fosberg
Houston Police Department
Houston, TX

Introduction

We made an analysis of weapon use in Houston robberies. Because the type of weapon used rarely changes, the Multinomial Logit model was an appropriate modeling tool. Between 2007 and 2009, there were 33,419 robberies that occurred within the City of Houston. Of these, suspect information was obtained for 3,709 of these offenses. Using the suspect information for these 3,709 offenses, we modeled predictors of weapon use.

Figure 21A.1 shows the distribution of weapons within the area covered by the Houston Police Department. Of the weapons used, 1,744 (or 47%) involved firearms, 272 (or 7%) involved knives, 1,184 (or 32%) involved bodily force, 192 (or 5%) involved threat, and 317 (or 9%) involved another weapon. Using the 'other weapon' as the reference category, we related weapon choice to 11 variables grouped into five categories: 1) Offender characteristics (age, gender, being of Hispanic ethnicity, being of African-American ethnicity); 2) Presence of co-offenders (the number of suspects); 3) Whether the robbery occurred on a commercial premise or not; 4) Time period (night, afternoon, morning), and 5) Crime location characteristics (median household income of the block group at the crime location, distance from the offenders residence to the crime location).

Method

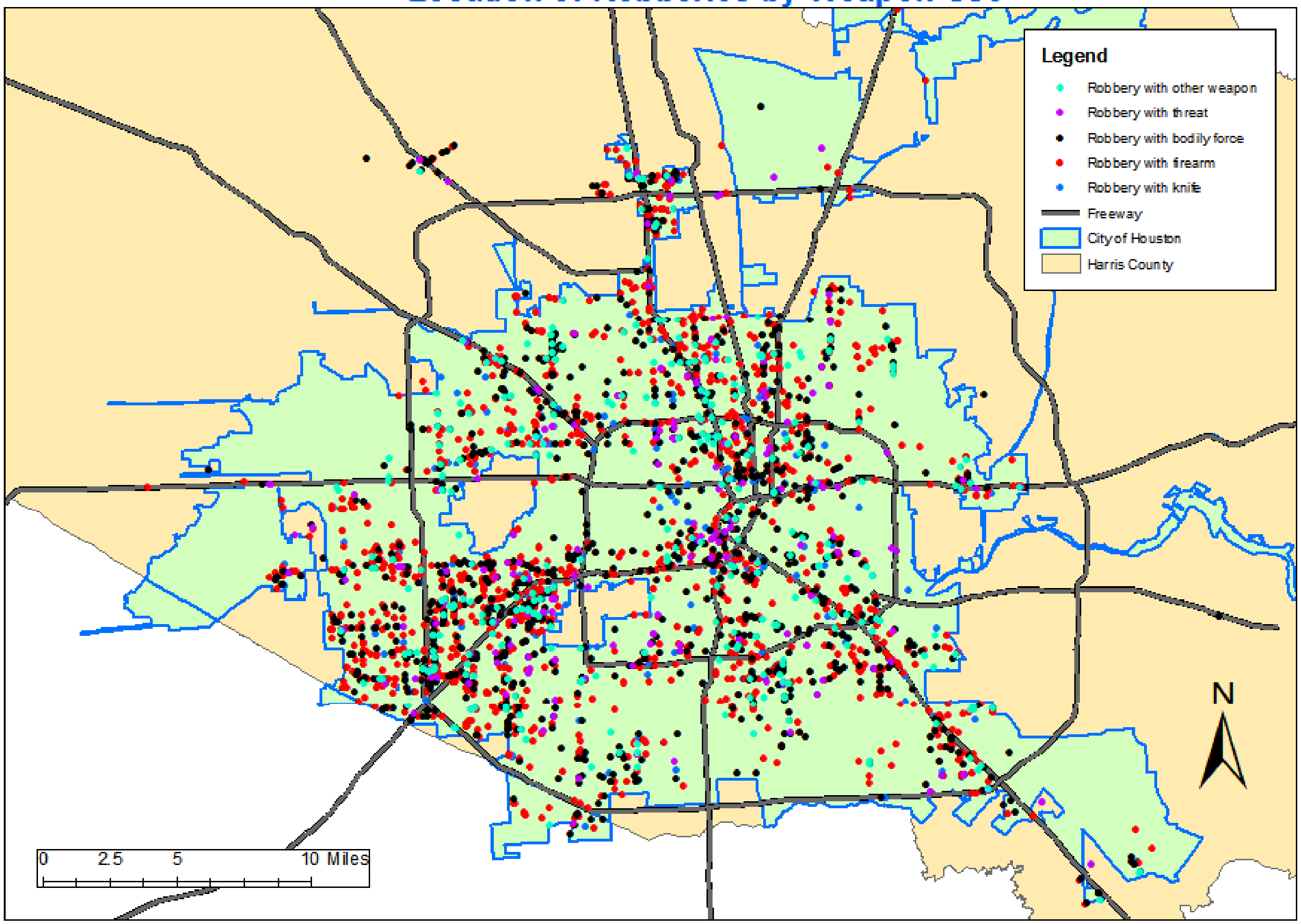
The multinomial logit model was used to estimate the effect of the coefficients on weapon choice. The choice probability, P_{ni}, that the offender, *i*, chooses a particular weapon, *j*, is estimated by an exponentiated linear combination of independent predictors associated with the offender, *k*, proportional to the choice probabilities for all weapons:

$$P_{ij} = \frac{Observed\ utility\ of\ weapon\ j}{Observed\ utility\ of\ all\ weapons} = \frac{e^{\beta_{0j}+\Sigma_1^K(\beta_{jk}X_{ijk})}}{\Sigma_1^J e^{\beta_{0j}+\Sigma_1^K(\beta_{jk}X_{ijk})}}$$

Figure 21A.1:

Houston Robberies: 2007 to 2009

Location of Robberies by Weapon Use



That is, the probability of the offender choosing any one weapon is estimated by an exponentiated linear combination of observed variables associated with the offenders divided by the sum of the exponentiated linear combination for all weapon choices. The coefficients are estimated across offenders but the probabilities are calculated for each offender separately.

Table 21A.1 presents the results of the model and Table 21A.2 summarizes the initial frequencies and the average predicted probabilities. Compared to the use of another weapon (the reference group), firearm use was associated with younger Hispanic or African-American males, with more accomplices, and was more likely to be committed on commercial premises in higher income locations at night or in the early morning. Crime travel distance was farther.

Table 21A.1:

Multinomial Logit Predictors of Weapon Use in Houston Robberies: 2007-09

Model result:

DepVar:	**WEAPON**
N:	**3709**
Df:	**3696**
Likelihood statistics	
Log Likelihood:	**-4432.1**
AIC:	**8936.3**
BIC/SC:	**9160.2**
Model error statistics	
Mean absolute deviation:	**0.27**
Mean squared predicted error:	**0.14**

Weapon: Firearm

Predictor	Coefficient	Stand Error	t-value	p-value	Odds ratio
Constant	**0.4959**	**0.005**	**91.43**	**0.001**	**1.64**
	Offender characteristics				
AGE	**-0.0279**	**0.003**	**-9.79**	**0.001**	**0.97**
FEMALE	**-0.9308**	**0.005**	**-171.60**	**0.001**	**0.39**
HISPANIC	**1.0317**	**0.005**	**190.54**	**0.001**	**2.81**
AFRICAN-AMERICAN	**1.3980**	**0.005**	**258.29**	**0.001**	**4.05**
	Co-offenders				
NUMSUSPCTS	**0.1774**	**0.005**	**33.08**	**0.001**	**1.19**
	Type of premise				
COMMERCIAL	**0.6431**	**0.005**	**118.73**	**0.001**	**1.90**
	Time period				
NIGHT	**0.3927**	**0.005**	**72.47**	**0.001**	**1.48**
AFTERNOON	**-0.2119**	**0.005**	**-39.11**	**0.001**	**0.81**
MORNING	**0.1672**	**0.005**	**30.84**	**0.001**	**1.18**
	Crime location				
MED HH INC	**0.00001**	**0.000003**	**1.97**	**0.05**	**1.00**
TRAVEL DISTANCE	**0.0289**	**0.004**	**6.65**	**0.001**	**1.03**

Table 21A.1: (continued)

Weapon: Knife

Predictor	Coefficient	Stand Error	t-value	p-value	Odds ratio
Constant	-0.6883	0.005	-126.82	0.001	0.50
Offender characteristics					
AGE	0.0205	0.004	5.82	0.001	1.02
FEMALE	-0.1706	0.005	-31.43	0.001	0.84
HISPANIC	0.4801	0.005	88.52	0.001	1.62
AFRICAN-AMERICAN	-0.1024	0.005	-18.88	0.001	0.90
Co-offenders					
NUMSUSPCTS	-0.1842	0.005	-34.03	0.001	0.83
Type of premise					
COMMERCIAL	0.0914	0.005	16.84	0.001	1.10
Time period					
NIGHT	0.2574	0.005	47.45	0.001	1.29
AFTERNOON	-0.0165	0.005	-3.05	0.01	0.98
MORNING	0.1056	0.005	19.46	0.001	1.11
Crime location					
MED HH INC	0.000004	0.000003	1.123	n.s.	1.00
TRAVEL DISTANCE	-0.0300	0.005	-5.91	0.001	0.97

Weapon: Bodily force

Predictor	Coefficient	Stand Error	t-value	p-value	Odds ratio
Constant	0.5384	0.005	99.27	0.001	1.71
Offender characteristics					
AGE	-0.0012	0.003	-0.42	n.s.	1.00
FEMALE	0.0542	0.005	9.99	0.001	1.06
HISPANIC	0.2362	0.005	43.61	0.001	1.27
AFRICAN-AMERICAN	0.5802	0.005	107.17	0.001	1.79
Co-offenders					
NUMSUSPCTS	-0.1342	0.005	-24.96	0.001	0.87
Type of premise					
COMMERCIAL	0.2963	0.005	54.70	0.001	1.34
Time period					
NIGHT	0.0861	0.005	15.88	0.001	1.09
AFTERNOON	0.4638	0.005	85.64	0.001	1.59
MORNING	0.3381	0.005	62.35	0.001	1.40
Crime location					
MED HH INC	0.00001	0.000003	4.14	0.001	1.00
TRAVEL DISTANCE	-0.0227	0.005	-5.02	0.001	0.98

Table 21A.1: (continued)

Weapon: Threat

Predictor	Coefficient	Stand Error	t-value	p-value	Odds ratio
Constant	-1.9169	0.005	-353.16	0.001	0.15
Offender characteristics					
AGE	0.0187	0.004	5.17	0.001	1.02
FEMALE	-1.2088	0.005	-222.67	0.001	0.30
HISPANIC	0.2279	0.005	42.00	0.001	1.26
AFRICAN-AMERICAN	0.6623	0.005	122.08	0.001	1.94
Co-offenders					
NUMSUSPCTS	-0.3061	0.005	-56.47	0.001	0.74
Type of premise					
COMMERCIAL	0.7707	0.005	142.04	0.001	2.16
Time period					
NIGHT	-0.1765	0.005	-32.52	0.001	0.84
AFTERNOON	-0.0113	0.005	-2.08	0.05	0.99
MORNING	0.4941	0.005	91.049	0.001	1.64
Crime location					
MED HH INC	0.00002	0.000003	4.25	0.001	1.00
TRAVEL DISTANCE	0.0193	0.005	3.86	0.001	1.02

Reference choice: Other weapon

On the other hand, knife use was associated with older, Hispanic males with few accomplices. The robberies were more likely to be committed on commercial premises at night or early morning. Crime travel distance was shorter.

Bodily force was associated with Hispanic or African-American females and with few accomplices. The robberies were more likely to be committed in higher income locations on commercial premises in the afternoon, morning or, to a lesser extent, late at night. The crime travel distance was shorter.

Finally, threats were associated with older Hispanic or African-American males with no or few accomplices. The robberies were more likely to be committed on commercial premises in the morning in higher income locations. The crime travel distance was farther.

Table 21A.2 shows that the average predicted probabilities for weapon use across all robbers exactly predicted the actual distribution of weapon use.

Conclusion

The most distinguishing variable is the number of suspects. More co-offenders lead to a greater use of firearms, suggesting the involvement of gangs. Other consistent predictors are ethnicity - Hispanic or African-Americans are more likely to use weapons than non-Hispanic White or Asian suspects, and gender - males are more likely to use firearms, knives or threats than females, who in turn are more likely to use bodily force. Commercial properties tend to be

Table 21A.2:

Summary of Predictions

Weapon	**Frequency of Weapon Use**	**(%)**	**Average Predicted Probability**
Firearm	1,744	(47%)	0.47
Knife	272	(7%)	0.07
Bodily force	1,184	(32%)	0.32
Threat	192	(5%)	0.05
Other weapon	317	(9%)	0.09
TOTAL	**3,709**	**(100%)**	**1.00**

disproportionately associated with weapons of all sorts primarily because they are the most common location for robberies in general. There are subtle differences in the time period and in the travel distance in predicting the type of weapon used.