1-1-1995

The Aggregated Spatial Logit Model: Theory, Estimation And Application

Mark Richard Ferguson

THE AGGREGATED SPATIAL LOGIT MODEL:

THEORY, ESTIMATION AND APPLICATION

By

MARK RICHARD FERGUSON, B.E.S., M.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

1995

THE AGGREGATED SPATIAL LOGIT MODEL

DOCTOR OF PHILOSOPHY (1995) McMASTER UNIVERSITY
(Geography) Hamilton, Ontario

TITLE: The Aggregated Spatial Logit Model: Theory, Estimation And Application

AUTHOR: Mark Richard Ferguson, B.E.S. (University of Waterloo, 1989)

M.A. (University of Waterloo, 1991)

SUPERVISOR: Dr. P.S. Kanaroglou

NUMBER OF PAGES: xii, 250

Abstract

In problems of spatial choice, the choice set is often more aggregated than the one considered by decision-makers, typically because choice data are available only at the aggregate level. These aggregate choice units will exhibit heterogeneity in utility and in size. To be consistent with utility maximization, a choice model must estimate choice probabilities on the basis of the maximum utility within heterogeneous aggregates. The ordinary multinomial logit model (OMNL) applied to aggregate choice units fails this criterion as it is estimated on the basis of average utility. In this thesis, the aggregated spatial logit model, which utilizes the theory underlying the nested logit model to estimate the appropriate maximum utilities of aggregates, is derived and discussed. Initially, the theoretical basis for the model is made clear and an asymptotic version of the model is derived. Secondly, the model is tested in a simulated environment to demonstrate that the OMNL model lacks the generality of the aggregated model in the presence of heterogeneous aggregates. Thirdly, full endogenous estimation of the aggregated model is studied with a view toward finding the best optimization algorithm. Finally, with all the elements in place, the model is tested in an application of migration from the Canadian Atlantic Provinces.

Acknowledgements

Initially, I would like to thank my supervisor, Dr. Pavlos Kanaroglou who has put a lot of effort into the task of having me mature as a researcher and has always provided sage advice. It has been a pleasure and a learning experience to watch Pavlos, over the past four years, roll up his sleeves and attack the research problems with which he is confronted. Thank you Pavlos, I have valued your guidance highly. Now it's up to me! I wish to thank the other two members of my supervisory committee: Dr. Bill Anderson and Dr. Barry Boots who I think have also provided me with excellent advice. I, along with other graduate students, have always appreciated Bill's open-door policy and the fact that he is never too busy to give you a moment of his time. I have enjoyed discussing matters with Barry Boots and having him on my committee was a pleasure. Thank you Barry for putting up with the inconvenience of this "out-of-town" assignment. I give my thanks also to all graduate students and staff in the Geography department here at McMaster with whom I have had the pleasure of interacting over the past four years.

To all members of my family both here in Canada and in Scotland I say hello and thank you for all the encouragement and positive reinforcement I have received from you over the years. Finally, I owe a special debt to Sosy whose positive outlook on life and unwavering support has made the last two years go a lot more smoothly for me. A special note for Andrew and Alison: now that this thesis is done, let's go out and play baseball tonight!

Table of Contents

List Of Tables

List Of Figures

Preface

It should be noted that Chapters 3 to 6 in this thesis are papers which were prepared with the objective of having them published in journals. The papers follow a logical sequence though as they all represent results of this one body of research. It is an unavoidable fact, however, that there is some degree of repetition throughout the chapters, particularly in introductory sections. In some respects, this might be of benefit since the same mathematical arguments are discussed from slightly differing perspectives. In any case, the ordering of the papers is as follows:

Chapter 3: Kanaroglou, P.S. and M.R. Ferguson (1995). "Discrete Spatial Choice Models for Aggregate Destinations." Accepted for publication in the *Journal of Regional Science*.

Chapter 4: Ferguson M.R. and P.S. Kanaroglou (1995). "Utility Variability Within Aggregate Spatial Units and its Relevance to Discrete Models of Destination Choice", in (R. Florax and L. Anselin, eds.) *New Directions in Spatial Econometrics*, Amsterdam: Springer-Verlag. (Available in the autumn of 1995).

Chapter 5: Ferguson M.R. and P.S. Kanaroglou (1995). "The Aggregated Logit Model: A Comparative Analysis of Estimation Methodologies", Accepted for publication in *Geographical Systems*.

Chapter 6: Ferguson M.R. and P.S. Kanaroglou (1995). "An Empirical Evaluation of the Aggregated Spatial Choice Model", Submitted to *International Regional Science Review*.

Notwithstanding the exception which is discussed below, the dissertation author has undertaken the role of primary investigator in all matters relating to the theory and implementation of the aggregated spatial logit model, with the supervisor acting in a supporting role. With regard to the papers which appear in this thesis, and which are co-authored with the supervisor, the dissertation author has undertaken the primary role with the supervisor contributing through critique of initial drafts and editorial advice.

The exception to these statements occurs in portions of Chapter 3 where the supervisor bears primary responsibility for the content. In particular, Section 3 of this Chapter and the proof which appears in the appendix are primarily due to the efforts of Pavlos Kanaroglou. In response to our initial submission of this paper to the Journal of Regional Science, it was requested that we incorporate a section on the asymptotic properties of the aggregated logit model. In accommodating this request, section 3 of Chapter 3 came about. It is important to emphasize, however, that the primary thrust of this thesis has been the evaluation of this model when the assumptions underlying the asymptotic version cannot be met, which is the case for most problems. An example is the application in Chapter 6.

Chapter 1

An Introduction to Discrete Choice Models

1.1 Introduction

The purpose of this first chapter is to outline the theory underlying discrete choice models. The various forms are each derived and the logic associated with the derivations is discussed in an intuitively appealing way. The exposition is primarily aimed at readers with little experience in discrete choice modelling and hence may seem too deliberate for experts in the field. Many geographers and environmental scientists have embraced a purely statistical approach to choice modelling (Wrigley, 1985) which is not based on a utility framework. It is hoped that this discussion will benefit that group of researchers and overall that readers will appreciate the major trade-offs and differences between the main model types. In general, this chapter forms the basis for what follows, including the aggregated spatial logit model, the main topic of this thesis. While the thesis is ultimately about spatial choice, the discussion in this chapter is generic to all choice scenarios. It is only in later chapters that the move to spatial choice situations is made after this initial discussion of the probit, ordinary multinomial logit (OMNL), nested logit and aggregated models.

1.2 The Discrete Choice Model Family

1.2.1 Theory Common to the Models

The models outlined below will be taken to be of the *revealed preference* type. In essence, this implies that the preferences of people are revealed through the choices that they are observed to make. Hence, provided with a data set containing choice information, the assumption is that we should be able to infer something about the underlying preferences that motivated these choices. This assumption has been challenged in some quarters (Pirie, 1976; Sheppard, 1980) on the basis that, among other things, the actions of individuals are often motivated by forces beyond their control and hence are not really indicative of the true set of preferences. The alternative is the *stated preference* type of model popularized by Louviere (1983) which attempts to identify peoples preferences in a controlled environment thus allowing for the development of more complex utility functions.

Discrete choice models are premised on the tenets of random utility theory which has been formalized by Manski (1977). This framework provides a theoretical basis for the fact that choice-makers are not always observed to choose the alternatives predicted by the analyst. In particular, it is assumed that the utility U_{in} for choice-maker n relating to alternative i is measured on an ordinal scale. The assumption is that U_{in} is known and deterministic from the perspective of the choice-maker who is able to perfectly assess the utilities associated with alternatives in their choice set C_n. It is further assumed that the researcher views U_{in} as a random variable which hence cannot be predicted with certainty. Manski describes four main sources of error which contribute to this uncertainty: 1) a missing attribute which either varies only across alternatives or across both choice-makers and alternatives; 2) a missing socio-economic variable associated with taste variation which varies only across choice-makers; 3) imperfect measurements of included variables; 4) the use of proxy variables in cases where the true variable cannot be obtained.

As a starting point for a model, the total utility U_{in} can be written as:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

where the new terms are the systematic utility V_{in}, a deterministic quantity estimated by the researcher, and ε_{in} a random error component which accounts for the difference between V_{in} and U_{in}. The systematic utility V_{in} is a function of choice-maker characteristics, attributes of alternatives and variables measuring distance, cost or time. The rationale behind qualitative choice models is that, in the face of uncertainty, the best that can be hoped is to make probabilistic assessments of an individual's choice behaviour. If V_{in} exceeds V_{jn}, then intuitively it makes sense that alternative i should be associated with a higher choice probability than alternative j. Such a probabilistic assessment of their relative attractiveness, however, does not preclude the possibility that alternative j is in reality the chosen option. In the social sciences, this flexibility is very useful.

The starting point for the actual estimation of probabilities is the following rule:

$$P_n(i) = \Pr(U_{in} > U_{jn},\ for\ all\ j \in C_n,\ j \neq i) \tag{2}$$

which translates verbally into the likelihood that alternative i has a higher utility than any other alternative in the choice set and without ties being allowed. Substituting 1 in 2 we obtain:

$$P_n(i) = \Pr(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn},\ for\ all\ j \in C_n,\ j \neq i)$$

which can be written:

$$P_n(i) = \Pr(\varepsilon_{jn} < V_{in} - V_{jn} + \varepsilon_{in},\ for\ all\ j \in C_n,\ j \neq i)$$

To further clarify, assume that for some choice-maker n, $V_{in} - V_{jn} = 1$, implying that alternative i should be associated with the higher probability in a binary choice situation. In that case, ε_{jn} could be portrayed graphically as a function of ε_{in} as in *Figure 1.1*. For an area defined by any arbitrary circle centred on the

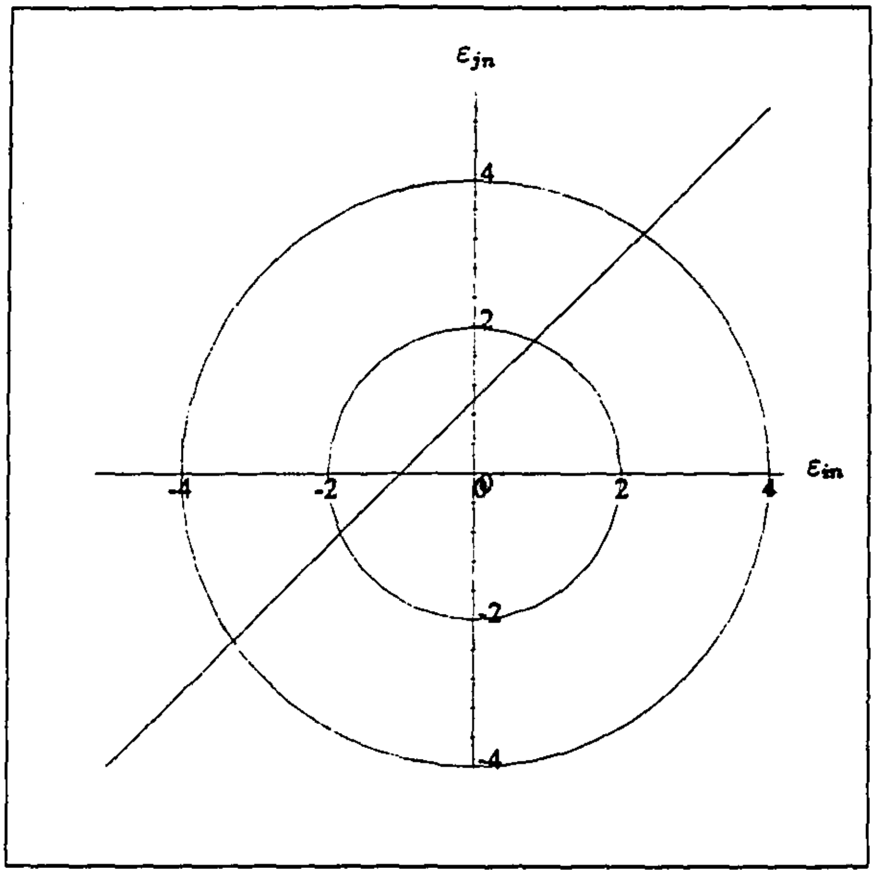


Figure 1.1: A Choice Probability Demarcation Line

origin, note that the proportion below the line exceeds the proportion above the line since there are many more possibilities for $\varepsilon_{in} + 1 > \varepsilon_{jn}$.

The relative areas above and below this line are a large part of what defines the choice probabilities. The other part relates to the joint density function which defines the relationship between the error terms. If it was assumed that both errors had some uniform (constant) positive density then it is not difficult to derive the probabilities based on the volumes on either side of the demarcation line.

It is when more general assumptions about the joint distribution of the error terms are made that the widely utilized qualitative choice models emerge. We can start by making the assumption that $E(\varepsilon_{jn})$ and $E(\varepsilon_{in}) = 0$ under the belief that our estimates V_{in} and V_{jn} of the true utilities U_{in} and U_{jn} are located centrally on the distributions of the true utilities. So long as constants can be added to the systematic utilities, which is the case, then this is not a restrictive assumption (Ben-Akiva and Lerman, 1985, p. 64). Now intuitively speaking, it seems that the best choice model would not treat the error terms with uniform densities. Over repeated observations, we would expect to make frequent small errors in the prediction of the true utilities and large errors with less frequency. Hence it is reasonable that $\varepsilon_{in} = 0$ and $\varepsilon_{jn} = 0$ be associated with the largest densities and that we should have a bell-shaped joint distribution. As a result, for $V_{in} - V_{jn} = 1$, $P_n(i)$ would be much larger with a bell-shaped density than with a uniform density because relatively more weight would be given to the volume defined below the demarcation line in *Figure 1.1.*

1.2.2 The Probit Model

A natural choice for the bell-shaped joint density of random errors, and one which leads to the probit model, is the multivariate normal distribution. Such an assumption is the most general of those underlying the family of discrete choice models. As we consider models which result from more restrictive

Figure 1.2a: Normal Distribution of Uncorrelated Random Errors

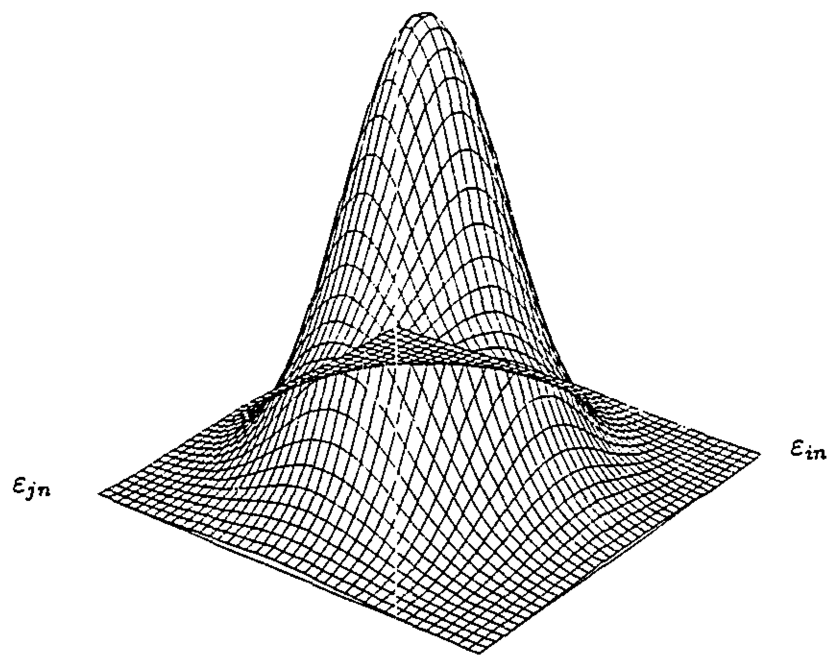
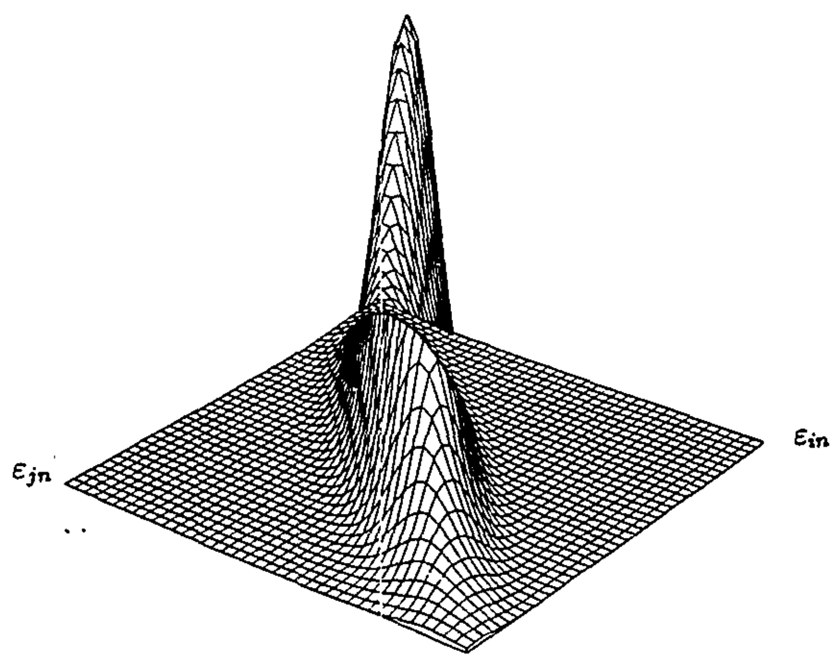


Figure 1.2b: Normal Distribution of Highly Correlated Random Errors



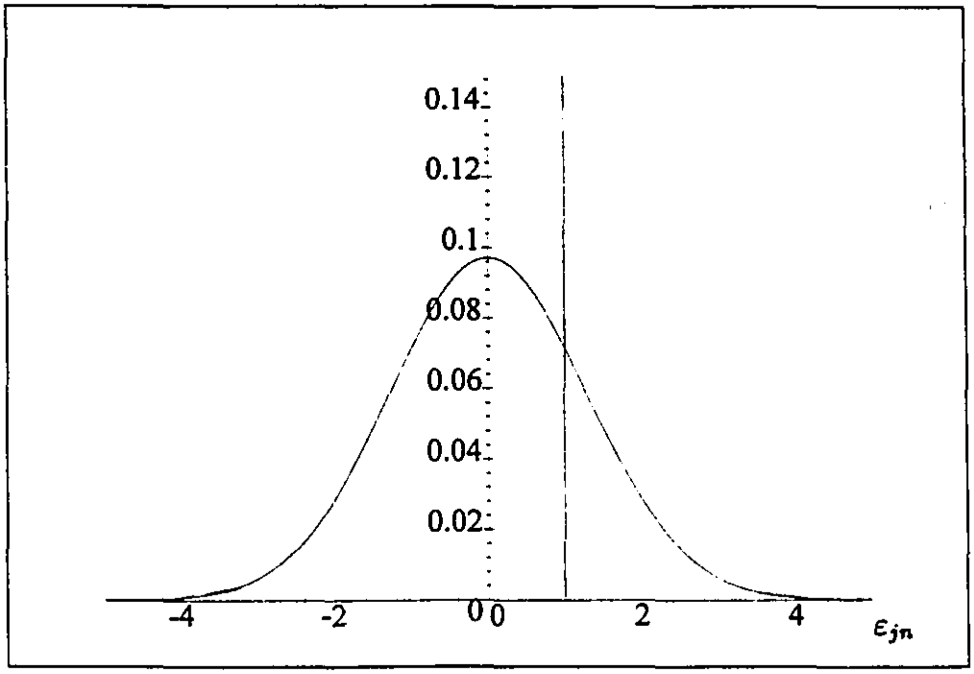


Figure 1.3a: Normal distribution of ε_{jn} conditional on $\varepsilon_{in} = 0$

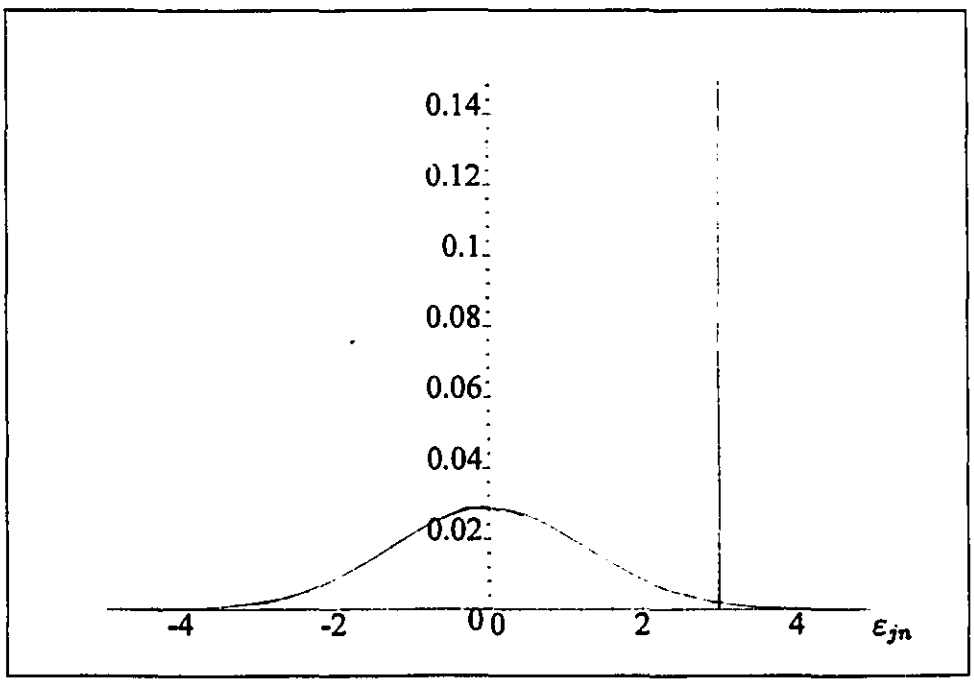


Figure 1.3b: Normal distribution of ε_{jn} conditional on $\varepsilon_{in} = 2$

assumptions, we will see how such models only approximate the capabilities of the probit model. The multivariate normal distribution has the form:

$$\phi(\varepsilon_n) = (2\pi)^{-L/2}\,(\det(\Omega_n))^{-1/2}\exp(-(\varepsilon_n'\Omega_n^{-1}\varepsilon_n)/2) \tag{3}$$

where ε_n is a column vector of error terms which apply over the set of alternatives for choice-maker n, L indicates the number of alternatives, and Ω_n is the variance-covariance matrix of the error terms.

In *Figure 1.2* we see how two bivariate densities appear. The upper *Figure 1.2a* illustrates the density function when there is zero covariance between the errors while *Figure 1.2b* depicts a 0.9 positive correlation. Given that ε_{in} attains a certain value, the distribution of ε_{jn} is under much more of a 'straightjacket' in the lower figure since the high level of correlation discourages a substantial deviation from ε_{in}. This is not the case in the upper figure where the value of ε_{jn} is unrelated to that of ε_{in}. An attractive aspect of the probit is that the pairwise relationships between the error terms of various alternatives can be represented in different ways with a distribution perhaps qualitatively like *Figure 1.2a* applying to one pairing while *Figure 1.2b* simultaneously applies to a different pairing. Hence choice probabilities will not respond in the same relative way to an exogenous shift in the utility of one alternative.

Figure 1.1 still applies in showing how a binary probit probability would be obtained given an exogenous difference $V_{in} - V_{jn} = 1$. If *Figure 1.1* is super-imposed on *Figure 1.2a*, for example, the demarcation line in the former figure becomes a plane which bisects the normal density surface. The choice probability emerges directly from the relative magnitudes of the two volumes defined by the plane. In *Figure 1.3*, we see two representations of the ε_{jn} density in the joint distribution conditional on $\varepsilon_{in} = 0$ (*Figure 1.3a*) and on $\varepsilon_{in} = 2$ (*Figure 1.3b*) with the vertical lines in each indicating the position of the demarcation line given by $V_{in} - V_{jn} = 1$[1]. On an absolute basis, the value of $P_n(i)$ will be increased more

[1]The variances of the random error components are assumed to be $\pi^2/6$ which is consistent with the variances assumed for the

when $\varepsilon_{in} = 0$ since the area to the left of the demarcation line is much larger than it is when $\varepsilon_{in} = 2$. On a relative basis though, it is clear that when $\varepsilon_{in} = 2$ the contribution to $P_n(i)$ is overwhelmingly larger than the contribution to $P_n(j)$.

Mathematically, the problem is one of evaluating a double integral. An infinite number of density surface cross-sections are taken, each one conditional on some value of ε_{in} as in *Figure 1.3*, and for each the area to the left of the demarcation line is considered through integration along the ε_{jn} axis. Integration along the ε_{in} axis accumulates these conditional areas and provides the total volume lying below the demarcation line in *Figure 1.1*. The area defined by some exogenously fixed ε_{in} is $\int_{\varepsilon_{jn}=-\infty}^{\varepsilon_{in}+V_{in}-V_{jn}} \phi(\varepsilon_n)$ $d\varepsilon_{jn}$, where the upper limit on the integral of this example is $\varepsilon_{in} + 1$. In order to obtain the result for $P_n(i)^2$, this integral is evaluated for every conceivable value of ε_{in} to yield:

$$P_n(i) = \int_{\varepsilon_{in}=-\infty}^{\infty} \int_{\varepsilon_{jn}=-\infty}^{\varepsilon_{in}+V_{in}-V_{jn}} \phi(\varepsilon_n) d\varepsilon_{jn}\, d\varepsilon_{in} \tag{4}$$

The extension of this expression to three or more alternatives is straightforward. In the calculation of $P_n(i)$ for L alternatives, each additional alternative is treated as was alternative j in the bivariate case. Hence 4 is generalized into the form:

$$P_n(i) = \int_{\varepsilon_{in}=-\infty}^{\infty} \int_{\varepsilon_{1n}=-\infty}^{\varepsilon_{in}+V_{in}-V_{1n}} \int_{\varepsilon_{2n}=-\infty}^{\varepsilon_{in}+V_{in}-V_{2n}} \cdots \int_{\varepsilon_{Ln}=-\infty}^{\varepsilon_{in}+V_{in}-V_{Ln}} \phi(\varepsilon_n) d\varepsilon_{Ln} \ldots d\varepsilon_{2n} d\varepsilon_{1n}\, d\varepsilon_{in} \tag{5}$$

where it can be seen that calculation of a single probability when there are L alternatives will require an

error components in the logit model to be discussed later. Any assumption about the variance is arbitrary however and has no effect on choice probabilities since systematic utilities will be scaled accordingly.

[2] In practice, calculation of binary probit probabilities is reduced to an expression with one integral. Since the difference between error terms is also normally distributed, the probabilities can be estimated directly on the basis of the difference in systematic utilities. The expressions for the probabilities are presented as they are here so that the generalization to three or more alternatives is easily grasped. Another important note about the binary probit is that choice probabilities are independent of any imposed $\mathbf{\Omega}_n$ variance-covariance matrix of the error disturbances. There must be a minimum of three alternatives for $\mathbf{\Omega}_n$ to have an impact in the same way that there must be at least three alternatives for a nested logit structure to make any sense.

integration along L dimensions. The fact that computation becomes burdensome with a relatively small number of alternatives is the main disadvantage of the probit model and has led researchers to consider other models.

Nevertheless, the probit model has many interesting features. The expressions in equations 4 and 5 indicate that the multivariate normal distribution in 3 forms the basis for the model's mathematical form. The important component in equation 3 is Ω_n, the variance-covariance matrix of the error terms for choice-maker n. In theory, this matrix can represent all patterns of dependence or independence exhibited in the unobservable components of utility. In practice, there are limitations but clearly the researcher has considerable freedom to model the impacts resulting from exogenous shifts in utility. Moreover, it is possible to do this for individual choice-makers, thereby accounting for taste variations in populations. How this is done is beyond the scope of this chapter but is covered elsewhere (Train, 1986; DaGanzo, 1979). The main point though is that the probit probability functional form does not depend only on the specification of the systematic utilities (V_{in}) which define the upper boundaries of all the integrations. The hypothesized structure of Ω_n, which can be generalized to account for unobserved taste variation, correlations in random errors and differences in the variances of the random errors, is also very important.

1.2.3 The Logit model

While the ordinary multinomial logit model (OMNL) is based on a more restrictive set of assumptions than the probit model, it has gained wide popularity because it is computationally much less demanding, especially for choice problems with many alternatives. At this point, the derivation of the OMNL is presented in detail to illustrate how the model's underlying assumptions are restrictive compared to the probit.

The seeds for the OMNL had been around for several decades, having first appeared in the psychology literature as Luce's (1959) strict utility model. Luce's model however, was a constant utility model in which the true utilities for all alternatives were assumed fixed and where choice-makers were not assumed to necessarily choose the option with highest utility Significantly though, the model featured the Independence from Irrelevant Alternatives (IIA) property which is discussed in more detail below. In a widely quoted paper, McFadden (1974) derived a model similar to Luce's in a form consistent with random utility theory. Having given the OMNL model an attractive theoretical justification, McFadden had set the stage for the model's application to a wide range of choice problems in numerous disciplines.

The derivation for the OMNL proceeds along similar lines as that of the probit. The end result is a model without the generality of the probit but with a closed form which requires no computationally intensive integrations. The major difference between the probit and the logit is that for the former, we explicitly assume a **joint** multivariate normal distribution, while for the latter, no joint distribution is explicitly included anywhere in the derivation. Instead, McFadden (1974) ingeniously employs a series of independently and identically distributed Gumbel distributions which are used to effectively approximate a joint multivariate normal distribution. First, the logic behind McFadden's approach is described verbally and graphically, and then the derivation is outlined mathematically.

Consider again the calculation of probabilities in a binary situation where $V_{in} - V_{jn} = 1$ as in *Figure 1.1.* Assume initially that for any exogenous ε_{in} the ε_{jn} are Gumbel distributed. If this is the case, then the counterpart to *Figure 1.3a* and *Figure 1.3b* would appear as in *Figure 1.4* where we see a Gumbel distribution with variance $\frac{\pi^2}{6}$. Instantly, it is clear that the distribution in *Figure 1.4* does not serve in this capacity very well since the area under the curve is 1.0 while it is obviously much less in the cross-sections of *Figure 1.3.* What is missed in *Figure 1.4* is that the density of ε_{jn} is very much dependent on the level at which ε_{in} is fixed. If $\varepsilon_{in} = 0$, then the total area under the ε_{jn} is its largest, albeit much less than 1.0. The further that ε_{in} deviates from zero, the smaller the area under the ε_{jn} density.

The solution to the problem is simply to weight each ε_{jn} density cross-section with the density of the associated ε_{in} as taken from a Gumbel distribution with variance $\frac{\pi^2}{6}$. For example, when $\varepsilon_{in} = 0$,

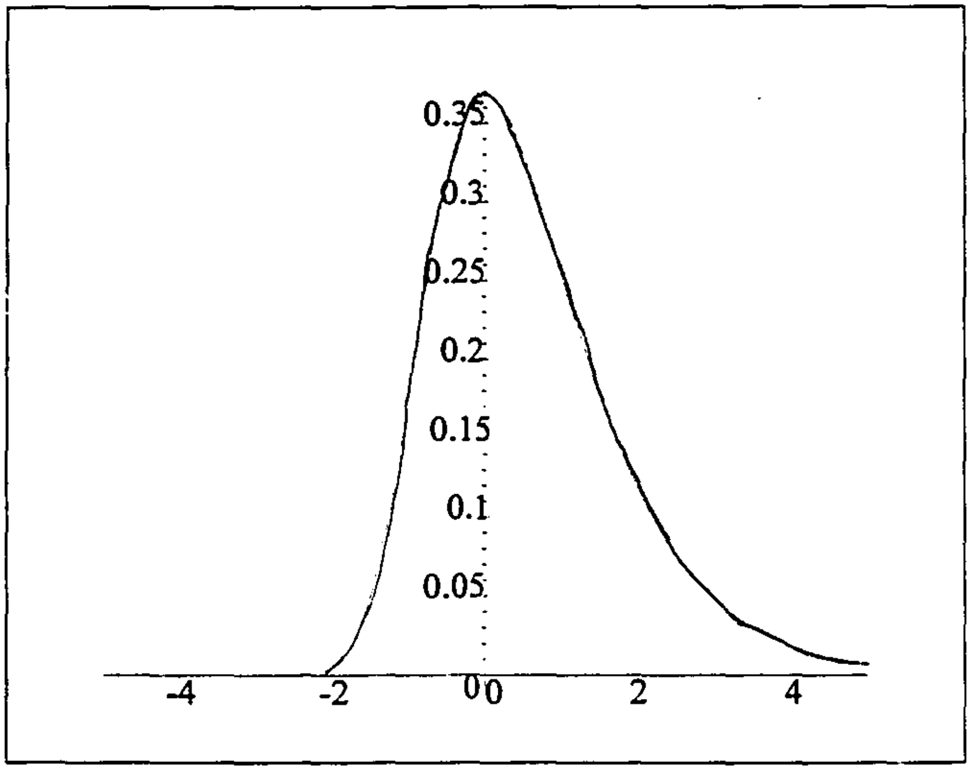


Figure 1.4: The Gumbel Distribution with mode zero and variance $\pi^2/6$

the associated Gumbel density is approximately 0.37 while it is approximately 0.11 when $\varepsilon_{in} = 2$. Now if the Gumbel density in *Figure 1.4* is multiplied by these scalar values, we now have the logit approximations to the distributions of ε_{jn} at $\varepsilon_{in} = 0$ and $\varepsilon_{in} = 2$ respectively. These are displayed in *Figure 1.5* where it is seen that the results closely resemble those of *Figure 1.3* with similar probabilities for $P_n(i)$ ultimately resulting when the volumes under the full set of cross-sections are calculated.

Mathematically, the derivation of the logit model proceeds as follows. As before:

$$P_n(i) = \Pr(\varepsilon_{jn} < V_{in} - V_{jn} + \varepsilon_{in},\ for\ all\ j \in C_n,\ j \neq i)$$

We choose some exogenous ε_{in} and attempt to define the density of ε_{jn} given this exogenous value. To this end, the Gumbel distribution has density function:

$$f(\varepsilon) = \mu \exp(-\mu(\varepsilon - \eta)) \exp(-\exp(-\mu(\varepsilon - \eta))) \quad (6)$$

and cumulative density:

$$F(\varepsilon) = \exp(-\exp(-\mu(\varepsilon - \eta))) \quad (7)$$

where μ is inversely related to the variance of the distribution through the expression $\frac{\pi^2}{6\mu^2}$ and η defines the mode of the probability distribution. These can be made notationally simpler by assuming that $\mu = 1$ so that the variance will be $\frac{\pi^2}{6}$ as above and by assuming that $\eta = 0$ so that the maximum value of the density determines the point where ε_{in} and $\varepsilon_{jn} = 0$. For a fixed ε_{in}, it is necessary to integrate the ε_{jn} density in *Figure 1.4* from $-\infty$ to $(\varepsilon_{in} + V_{in} - V_{jn})$. Since, by definition, the integral of a density function from $-\infty$ is simply the cumulative distribution evaluated at the upper bound of the integral, we substitute this bound into 7 so that the area under ε_{jn} up to the demarcation line would be:

$$\exp(-\exp(-(\varepsilon_{in} + V_{in} - V_{jn})))$$

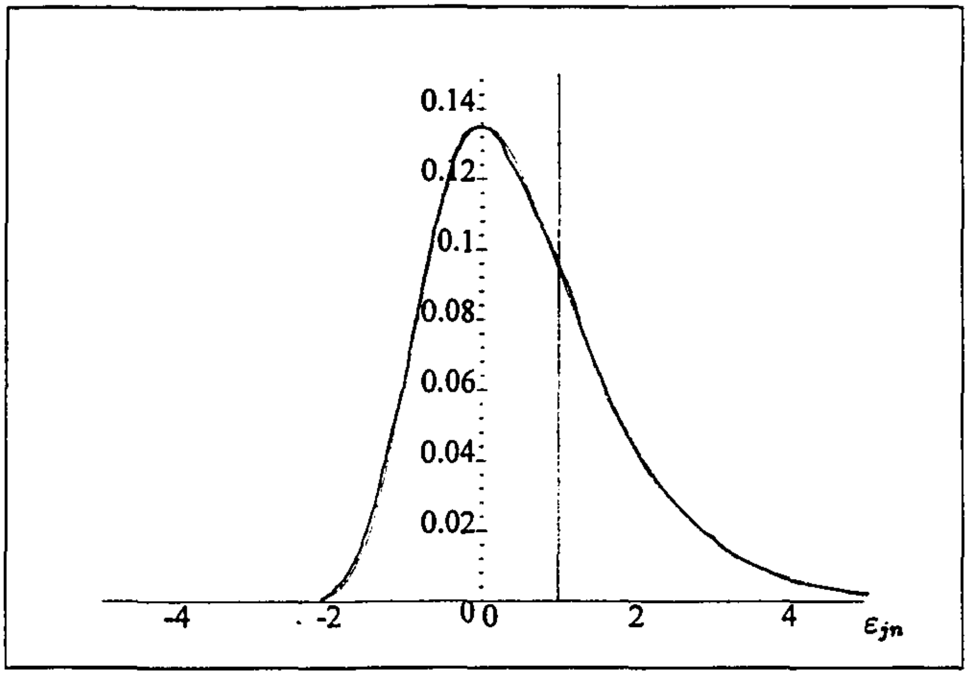


Figure 1.5a: Gumbel distribution of ε_{jn} conditional on $\varepsilon_{in} = 0$

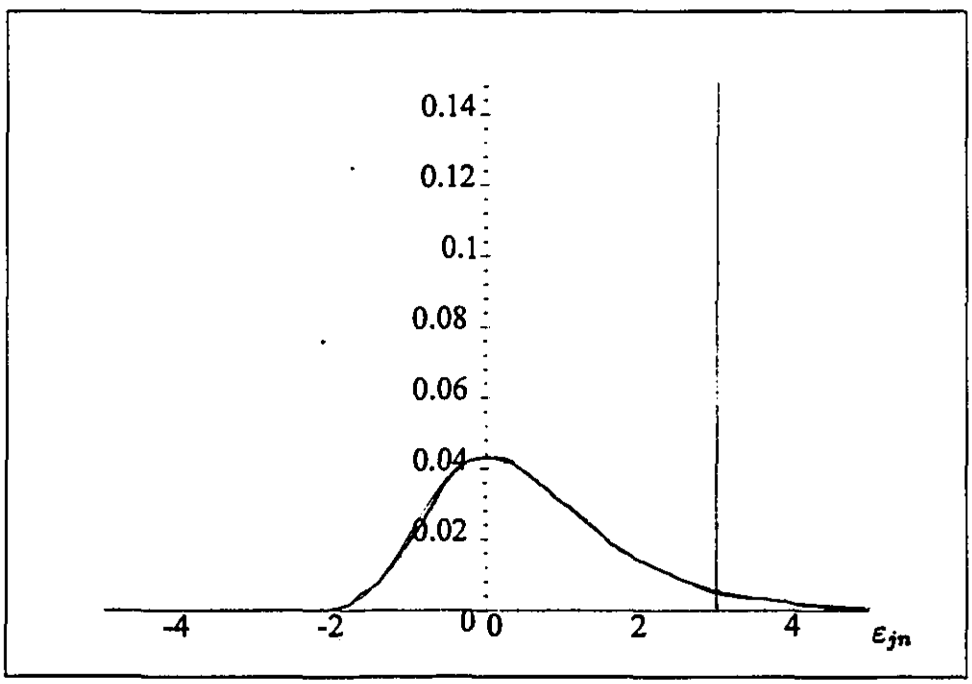


Figure 1.5b: Gumbel distribution of ε_{jn} conditional on $\varepsilon_{in} = 2$

For $\varepsilon_{in} = 0$, this area would be $\exp(-\exp(-1)) = 0.692$ and for $\varepsilon_{in} = 2$, it would be $\exp(-\exp(-3)) = 0.951$. These results are proportionally consistent with *Figure 1.5*, however to make them correct in the absolute sense, it is necessary to weight the distributions by the appropriate density of ε_{in}. We substitute ε_{in} into 6 giving:

$$\exp(-\varepsilon_{in})\exp(-\exp(-\varepsilon_{in}))$$

This provides a density at $\varepsilon_{in} = 0$ of 0.368 and a density at $\varepsilon_{in} = 2$ of 0.118. The weighted areas are 0.368×0.692 = 0.254 and 0.118×0.951=0.112 which are the areas seen to the left of the demarcation lines in *Figure 1.5*. In general, these areas are obtained as the product of the Gumbel density function evaluated at ε_{in} and the cumulative Gumbel function evaluated at $(\varepsilon_{in} + V_{in} - V_{jn})$ giving:

$$[\exp(-\varepsilon_{in})\exp(-\exp(-\varepsilon_{in}))]\,[\exp(-\exp(-(\varepsilon_{in} + V_{in} - V_{jn})))]$$

This expression is readily generalized if there are three or more alternatives. The ε_{in} term remains fixed and the alternatives other than j are assessed as j was, with the ultimate result being the product of all the individual pairings of i with each alternative. Hence:

$$\exp(-\varepsilon_{in})\exp(-\exp(-\varepsilon_{in}))\prod_{\substack{j\in C_n\\ j\neq i}}[\exp(-\exp(-(\varepsilon_{in} + V_{in} - V_{jn})))] \tag{8}$$

Note that the density expression to the left of the product sign in 8 contains the term:

$$\exp(-\exp(-\varepsilon_{in}))$$

which can be taken as the cumulative density of $(\varepsilon_{in} + V_{in} - V_{in})$. Since this term is a special case of what is seen on the right side of equation 8, it can be shifted so that the expression becomes:

$$\exp(-\varepsilon_{in})\prod_{j\in C_n}[\exp(-\exp(-(\varepsilon_{in} + V_{in} - V_{jn})))]$$

For a problem with any number of alternatives, $P_n(i)$ can now be evaluated for some fixed value of ε_{in}. To assess the total probability, it is necessary to integrate over all values of ε_{in} so that all possible ε_{jn} cross-sections are considered simultaneously. When this is done, the overall choice probability is:

$$P_n(i) = \int_{\varepsilon_{in}=-\infty}^{\infty} \exp(-\varepsilon_{in}) \prod_{j\in C_n} [\exp(-\exp(-(\varepsilon_{in} + V_{in} - V_{jn})))] \; d\varepsilon_{in} \tag{9}$$

The reasoning underlying the logit derivation has now been outlined along with its similarity to that of the probit model. The transformation of the expression in equation 9 into the well-known closed form logit probability:

$$P_n(i) = \frac{\exp[V_{in}]}{\sum_{j\in C_n} \exp[V_{jn}]}$$

is a multi-step exercise in integration which is not reproduced here since the main objective is to make clear the conceptual framework underlying the model. Details of the integration can be found in Train (1986, p. 54). An alternative and more straightforward way of deriving the logit model is shown in Ben-Akiva and Lerman (1985, p. 106) where they make explicit use of properties of independent and identically distributed Gumbel variates to reduce multinomial choice situations to a binary problem. At that point, the calculation of the choice probability can be captured by a logistic distribution in one dimension where the error terms are portrayed as differences rather than absolutes. While that approach gives a better intuitive feel of the fact that utility is relative in choice models, the method adopted here indicates better the relationship of the OMNL model with the probit.

1.2.4 The Problem of Alternatives with Correlated Utilities

A valid question is: under what circumstances does the use of i.i.d. Gumbel variates provide a poor approximation to the more general multivariate normal distribution? As it turns out, the answer to this question provides the rationale for the development of the nested logit approach. Consider a sample problem which illustrates the potential shortcomings of the logit model.

Assume that there is a situation with three alternatives i, j and k and that the latter two alternatives are perceived by choice-makers as being quite similar. We could assume that alternative i represents a brand of beer while alternatives k and j are competing soft drink brands. The correlation matrix of the associated utilities might appear as:

	U_{in}	U_{jn}	U_{kn}
U_{in}	1.0	0.0	0.0
U_{jn}	0.0	1.0	0.9
U_{kn}	0.0	0.9	1.0

in which high positive correlations among the random errors of the soft drinks are exhibited along with zero correlations in the pairings involving the beer option. The soft drink utilities would tend to share the same unobservable components while the beer utility would have its own distinct set of unobservable components. For some exogenous value of ε_{in}, the joint density of ε_{jn} and ε_{kn} ideally would appear qualitatively as in *Figure 1.2b*[3]. The other pair-wise distributions would ideally not have the distinct 'cigar-shaped' appearance that is associated with strongly correlated normal variates. The distributions of ε_{in} and ε_{jn} conditional on ε_{kn} and that of ε_{in} and ε_{kn} conditional on ε_{jn} would ideally appear qualitatively as in *Figure 1.2a.*

Assume that the price of the soft drinks is inversely related to the utility associated with them. How would the choice probabilities be affected through a 25% decrease in the price of alternative k's soft drink? Such an increase would result in a shift of the demarcation plane in both of the conditional distributions featuring ε_{kn} as a random variable since V_{kn} would increase. It is clear though that a fixed shift in the plane will have far more impact on the relative choice probabilities when there is a high

[3] Of course if $\varepsilon_{in} = 0$, the total volume under the surface would exceed that which would be the case if $\varepsilon_{in} = 3$ since the former outcome is far more likely. Nevertheless, the same qualitative characteristics would apply to the set of conditional distributions.

correlation in the relevant random errors as in *Figure 1.2b*. With highly correlated errors, as is the case between ε_{in} and ε_{kn}, it is comparatively unlikely that $V_{jn} + \varepsilon_{jn}$ will be sufficiently large to overcome the newly increased value of $V_{kn} + \varepsilon_{kn}$ since $\varepsilon_{kn} - \varepsilon_{jn}$ should be small. With uncorrelated random errors, as is the case between ε_{in} and ε_{kn}, there is a far greater likelihood of the ε_{in} being able to compensate for the increase in V_{kn}.

The net result for the behaviour of the probit model would be as follows. Since there will be no change in the systematic utilities governing the relative attractiveness of alternatives i and j, the model will predict no difference in the **ratio** of their choice probabilities. Since alternative k is made more attractive but the random errors are uncorrelated with the beer alternative i, then the ratio $P_n(i)/P_n(k)$ will decrease only slightly. However the ratio $P_n(j)/P_n(k)$, between the two soft drinks, will be seen to decrease significantly since those alternatives were viewed as virtual substitutes before the price change. Since alternative k has been made more attractive than j, it is apparent that j should suffer a much larger loss of market share than alternative i.

While the probit model is obviously general in the random error effects it can capture, the logit model is far more restrictive. **All** pairings of errors are characterized by the Gumbel approximation to the type of joint distribution seen in *Figure 1.2a*. An increase in V_{kn} will reduce the ratios $P_n(j)/P_n(k)$ and $P_n(i)/P_n(k)$ by the same amount which we know to be an incorrect assessment of the changes which would take place. Fortunately, a compromise solution, which preserves some of the computational tractability of the OMNL along with some of the realism provided by the probit model, is available in the form of the nested logit model.

1.2.5 The Nested Logit Model

The nested logit model, like the OMNL, utilizes i.i.d. Gumbel variates but does so in a more elaborate manner. In particular, alternatives which are similar are grouped into nests in recognition of the fact that an increase in the utility of one alternative will decrease the relative attractiveness of close substitutes much more than the relative attractiveness of alternatives which are unrelated. The restriction still applies that distributions qualitatively like *Figure 1.2a* must be used. However, there is a way in which utility dependencies can be accommodated within these constraints.

Assume that options j and k form a composite alternative encompassing the two brands of soft drinks while alternative i forms its own distinct beer composite. In essence, an upper level choice between beer and soft drinks and a lower level choice between soft drink brands is assumed. The choice between having a beer or soft drink will still be modelled fundamentally as an OMNL choice process as will the choice between the two soft drinks. However to ensure that an increase in the systematic utility of the soft drink maker only slightly reduces $P_n(i)/P_n(k)$, there must be some mechanism in place to dampen the increase in the composite soft drink utility.

This is accomplished through an important property governing the maximization of i.i.d. Gumbel variates (see Ben-Akiva and Lerman, Chapter 5). First define M^J as the number of alternatives in nest J. Now when each of the options $j \in \{1, 2, ..., M^J\}$ within a nest is associated with a Gumbel variate having mode V_j and variance $\frac{\pi^2}{6\mu^2}$,the distribution of the maximum utility derived from the nest is characterized by mode $\frac{1}{\mu} \ln \sum_{j=1}^{M^J} \exp(\mu V_j)$and variance $\frac{\pi^2}{6\mu^2}$. As it turns out, the discounting mechanism is dependent on the μ parameter. For example, assume that the systematic utilities for both soft drinks are 0.0 for some choice-maker. If $\mu = 2$ then the mode of the distribution of the maximum would be 0.34, while if $\mu = 1$, the mode of the distribution of the maximum would be 0.69.

The various distributions are depicted in *Figure 1.6.* In *Figure 1.6a* the display represents, from left to right, the distribution of the two elemental utilities (which are identical) for the soft drinks and secondly the distribution of their maximum with $\mu = 2$. *Figure 1.6b* represents the same thing except the variances of the elemental utilities are larger since $\mu = 1$. When μ is large, the expected maximum which can result from the two random variables is substantially lower than is the case when $\mu = 1$.

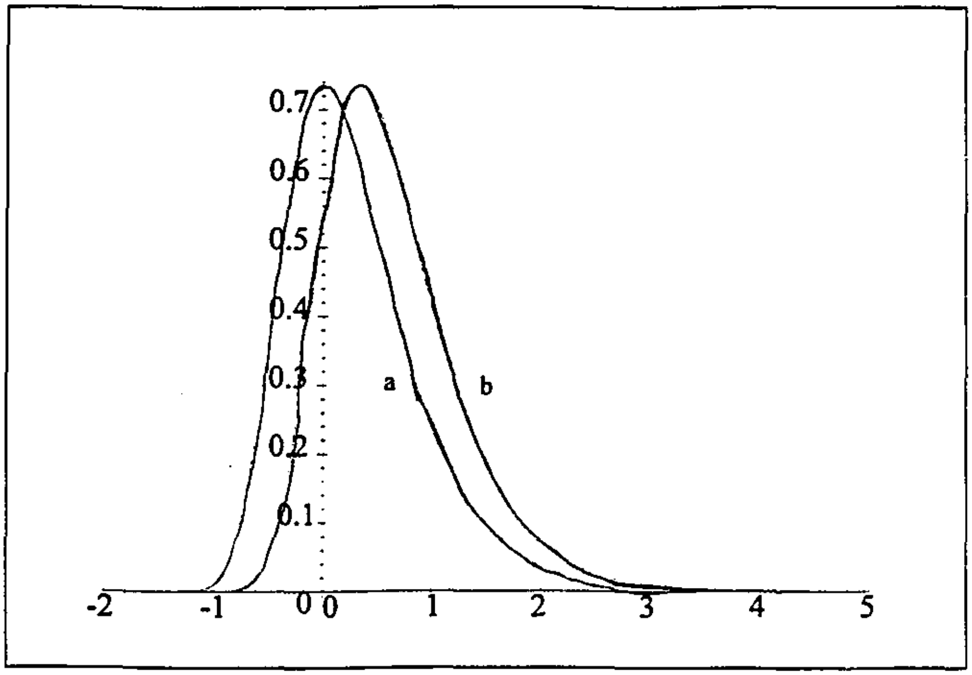


Figure 1.6a: Gumbel distributions of a) elemental and b) maximum utility with $\mu = 2$

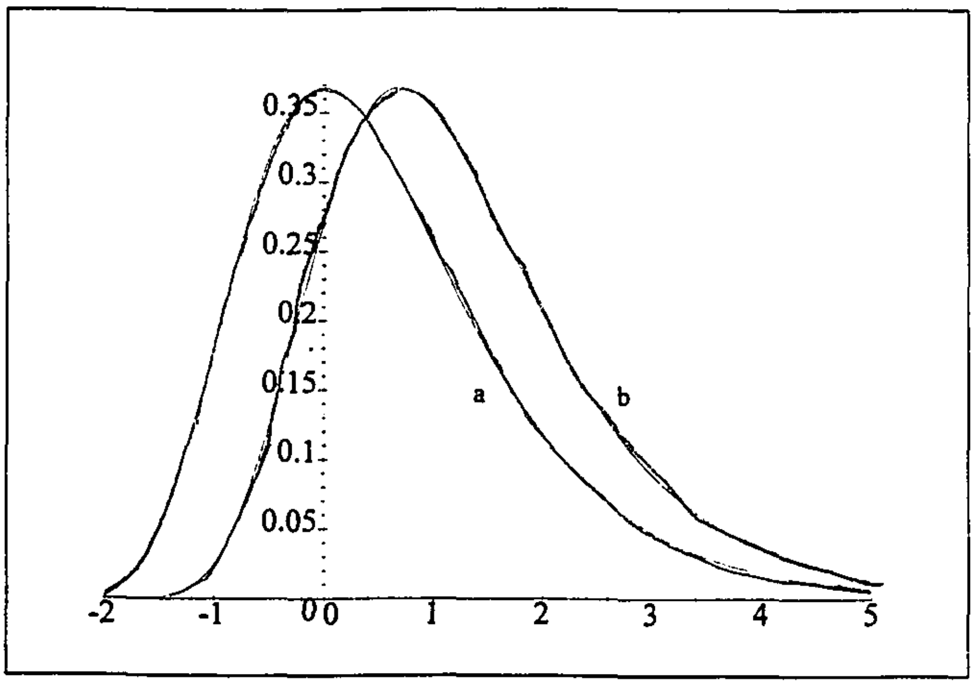


Figure 1.6b: Gumbel distributions of a) elemental and b) maximum utility with $\mu = 1$

There is clearly a dampening effect with $\mu = 2$ which would tend to decrease the impact of a soft drink price reduction at the upper level of the model. Moreover, despite the fact that the joint distribution of the elemental errors appear qualitatively as in *Figure 1.2a*, shrinking the variance of these random variables ensures that a given change in elemental systematic utilities has a dramatic effect on the relative soft drink probabilities. The overall conclusion is that when there is a high degree of substitutability between subsets of alternatives, the nested logit model has the capability to reduce the strictly elemental variance in utilities so that the impacts of exogenous changes are felt mainly within the relevant nest. As such, the nested logit model mimics the probit model while utilizing nothing other than i.i.d. Gumbel variates.

Consider now how the upper level marginal probabilities of the nested logit model would arise. The utility of the elemental alternative put forth to represent aggregate J would be given by:

$$U'^J_n = V^J_n + \varepsilon^{*J}_n$$

where $V^J_n = \frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} \exp\left(\mu^J V^J_{jn}\right)$ and ε^{*J}_n is a composite random error component with variance $\frac{\pi^2}{6}$. For the utility of a given elemental alternative, there are two sources of variance: one from the random error component ε^J_n relating equally to all elemental alternatives in the nest J, and the other from the random error component ε^J_{jn} which is unique to all elemental alternatives in the nest J. All of the ε^J_{jn} errors, which are assumed Gumbel distributed, are taken to be independent among themselves and in relation to the ε^J_n errors which are distributed in such a way that ε^{*J}_{jn} is Gumbel distributed. Hence the total variance for any elemental utility, including the representative one depicted by U^J_n, is $Var(\varepsilon_{jn} + \varepsilon^J_n)$ which equals $Var(\varepsilon_{jn}) + Var(\varepsilon^J_n)$ since ε_{jn} and ε^J_n are independent. Also, $Var(\varepsilon^{*J}_{jn}) = Var(\varepsilon_{jn}) + Var(\varepsilon^J_n) = \frac{\pi^2}{6}$. Since the upper level random error always has the same variance, the ultimate effect on aggregate utilities of a nest of highly substitutable elemental alternatives is transmitted through the impact of μ^J on the estimate of V^J_n. Under these circumstances, an upper level marginal probability would be given by the rule:

$$P_n(J) = \Pr(V_n^J + \varepsilon_n^{*J} \geq V_n^Q + \varepsilon_n^{*Q}, Q \in \{1,2,...,L\}, Q \neq J)$$

which is a very similar expression to that of the OMNL. Hence:

$$P_n(J) = \frac{\exp\left[V_n^J\right]}{\sum_{Q=1}^{L} \exp\left[V_n^Q\right]} = \frac{\exp\left[\frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} e^{\mu^J V_{jn}^J}\right]}{\sum_{Q=1}^{L} \exp\left[\frac{1}{\mu^Q} \ln \sum_{q=1}^{M^Q} e^{\mu^Q V_{qn}^Q}\right]} \tag{10}$$

Conditional probabilities are given by the expression:

$$P_n(j \mid J) = \Pr(V_{jn}^J + \varepsilon_n^J + \varepsilon_{jn} \geq V_{qn}^J + \varepsilon_n^J + \varepsilon_{qn},\ q \in \{1,2,...,M^J\},\ q \neq j)$$

Since ε_n^J will be constant between any pair of elemental alternatives in the nest and hence will not affect the conditional probability:

$$P_n(j \mid J) = \Pr(V_{jn}^J + \varepsilon_{jn} \geq V_{qn}^J + \varepsilon_{qn},\ q \in \{1,2,...,M^J\},\ q \neq j)$$

The result will be a standard OMNL probability:

$$P_n(j \mid J) = \frac{\exp\left[\mu^J V_{jn}^J\right]}{\sum_{q=1}^{M^J} \exp\left[\mu^J V_{qn}^J\right]} \tag{11}$$

It should be noted that the upper level scale parameter, associated with ε_n^{*J}, is exogenously fixed to 1.0 and hence does not appear in equation 10. However, the conditional probabilities in 11 are very much dependent on the endogenously determined scale parameter μ^J governing the variance of the ε_{jn} error terms. Clearly this cannot be assumed away if the desire is to properly represent the conditional probabilities and associated parameter estimates. The nesting concept is generalizable to any number of nests containing any number of alternatives and to any number of levels although estimation would become unwieldy beyond three levels.

Let us examine in more detail the pattern of dependency in utilities that will be permitted by the nested logit structure. Clearly, the only source of correlation between elemental alternatives within the same nest is the fact that they share the same random error component ε_n^J, the variance of which is inversely related to μ^J. As the variance of the random error common to all elemental alternatives in the nest increases, the covariance between elemental utilities increases and vice versa. It can be shown that there is a functional relationship between the covariance applying to all pairs of elemental alternatives in a nest and the variance of the random error. The relationship is:

$$\begin{aligned} Cov\left(U_{jn}^J, U_{kn}^J\right) &= Var(\varepsilon_n^J) \\ &= \frac{\pi^2}{6} - \frac{\pi^2}{6(\mu^J)^2} \\ &= \frac{\pi^2}{6}(1 - \frac{1}{(\mu^J)^2}) \end{aligned}$$

Armed with this result, the variance-covariance matrix of the utilities for the beer/soft drink problem would be:

$$\begin{array}{cccc} & U_{1n}^{(1)} & U_{1n}^{(2)} & U_{2n}^{(2)} \\ U_{1n}^{(1)} & \frac{\pi^2}{6} & 0 & 0 \\ U_{1n}^{(2)} & 0 & \frac{\pi^2}{6} & \frac{\pi^2}{6}(1 - \frac{1}{(\mu^J)^2}) \\ U_{2n}^{(2)} & 0 & \frac{\pi^2}{6}(1 - \frac{1}{(\mu^J)^2}) & \frac{\pi^2}{6} \end{array}$$

If this is standardized into a correlation matrix, the result is:

$$\begin{array}{cccc} & U_{1n}^{(1)} & U_{1n}^{(2)} & U_{2n}^{(2)} \\ U_{1n}^{(1)} & 1 & 0 & 0 \\ U_{1n}^{(2)} & 0 & 1 & 1 - \frac{1}{(\mu^{(2)})^2} \\ U_{2n}^{(2)} & 0 & 1 - \frac{1}{(\mu^{(2)})^2} & 1 \end{array}$$

Compare now the correlation matrices which result when $\mu^{(2)} = 1$ (on top) compared with the case when

$\mu^{(2)} = 2$ (on the bottom):

	$U_{1n}^{(1)}$	$U_{1n}^{(2)}$	$U_{2n}^{(2)}$
$U_{1n}^{(1)}$	1	0	0
$U_{1n}^{(2)}$	0	1	0
$U_{2n}^{(2)}$	0	0	1

	$U_{1n}^{(1)}$	$U_{1n}^{(2)}$	$U_{2n}^{(2)}$
$U_{1n}^{(1)}$	1	0	0
$U_{1n}^{(2)}$	0	1	0.75
$U_{2n}^{(2)}$	0	0.75	1

As it turns out, a value of $\mu^{(2)} = 3$ would come very close to generating the correlations of 0.9 originally assumed for the problem. Note that the correlations approach their limit of 1.0 for μ^J values which are quite small even though the upper value of $\mu^J$is ∞. Another observation is that when $\mu^J = 1$ and zero correlations result, the final model is identical to that of the OMNL. In that case, the two soft drink alternatives would behave as independent entities at the upper level with no dampening of the composite soft drink utility taking place.

The nested logit model has been demonstrated here in a bottom-up context in which the basic building blocks have been utilized to develop an overall model structure. McFadden (1978) is responsible for showing that the nested logit model, as well as the OMNL, can be derived in a top-down manner whereby they emerge as a special cases of the Generalized Extreme Value (GEV) model. Interested readers are referred to McFadden's paper or Train (1986).

1.2.6 The Aggregated Logit Model

Most choice scenarios do not generate nests of alternatives with more than two or three members. For example, in the beverage example, the two soft drink brands formed a nest of two. An exception is found in the realm of spatial choice where the choice units are available at many different levels of aggregation. A province, for example, could be sub-divided quite easily down to the level of census sub-divisions which would imply a provincial 'nest' containing hundreds of elemental alternatives. In the beverage example, it is quite likely that a full set of choice information can be obtained: that is the analyst will have complete knowledge of what elemental alternative was chosen by a given person. In the spatial choice context, it is quite unlikely that a full set of choice information can be obtained. Micro-data may reveal, for example, the province of choice but not the census sub-division of choice implying a discrepancy between the spatial resolution of the choice data and that of the choice process. The aggregated logit model, which is what this thesis is fundamentally about, is a probably the best way of reconciling this discrepancy.

The aggregated model can best be understood in light of the nested logit model since the two share the same underlying theory. The difference between the two is really related more to the means of implementation. Clearly, with the nested logit, the focus is on the correct representation of elemental probabilities and parameters associated with the employed variables. Hence a complete model utilizes both marginal and conditional probabilities to arrive at estimates of the overall elemental probabilities. With the aggregated model, the focus is on the correct representation of aggregate level marginal probabilities and the model parameters.

The final form of the aggregated model is given by equation 10 which is also the expression for the marginal probability of the nested logit model. Note that the aggregate level probabilities are portrayed as a function of the elemental systematic utilities along with the characteristics of the random errors distribution about these elemental systematic utilities as determined by the μ parameters. Fundamentally, the aggregate level utilities are the outcome of a maximization process applied to the utilities of the constituent elemental alternatives in a given nest. The conditions determining the mode and variance of the distribution of the maximum are the same as is the case with the nested logit model. These matters are covered in much more detail throughout the chapters of the thesis, particularly Chapter 3.

One expansion on the nested logit discussion, which will be repeated throughout the thesis, and which is particularly useful for the aggregated logit model, is if we rewrite the expression for V_n^J according to the following:

$$
\begin{aligned}
V_n^J &= \frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} \exp\left(\mu^J V_{jn}^J\right) \\
&= \overline{V_n}^J + \frac{1}{\mu^J} \ln \frac{1}{M^J} \sum_{j=1}^{M^J} \exp\left(\mu^J (V_{jn}^J - \overline{V_n}^J)\right) + \frac{1}{\mu^J} \ln M^J
\end{aligned}
\qquad (12)
$$

This reveals that the expected maximum utility is given by the sum of a mean utility term, a heterogeneity term which captures the intra-aggregate variation in systematic utility, and a size term which depends on the number of elemental alternatives. In addition, these corrective terms depend on the pattern of dependency between elemental utilities. The individual properties of these corrective terms are discussed throughout the thesis. These concepts are mentioned here simply as a preliminary overview.

1.3 Contents of Thesis

Having provided a background to the theory of discrete choice models, the main task is to outline the progression that can be expected through the following chapters and to discuss the reasoning for the order of inquiry. As has been mentioned, the main focus of this thesis is the aggregated spatial logit model, an approach which has not been systematically addressed in the past. While much of the underlying theory is due to the transport demand research group (such as McFadden, 1978) who established links with the nested logit model, researchers have never utilized the model in its complete form. The objectives of this thesis are to fully describe, test, and estimate the aggregated model using theory, simulations and an application. Since the issues of spatial aggregation of alternatives and similarity among spatial choice units are of so much importance in spatial choice modelling, it is felt that attaining these objectives will constitute a worthy contribution to the field. Using a brief precis of each chapter to follow, an overview is now provided of how these objectives are to be met.

Chapter 2 to follow is essentially a review paper. While Chapter 1 examined various discrete choice models in some detail, but not within the context of spatial choice, Chapter 2 is about spatial choice. In light of some of the key concepts outlined in this Chapter 1, this chapter seeks to assess the usefulness of the various discrete choice models in the spatial choice context. The particular problems of interest are: the substitutability of spatial units, aggregation of spatial units and choice set definition. The latter concept is emphasized less than the previous two. As well as outlining the implications of these issues for spatial choice models, the work of others in the spatial choice context is reviewed and critiques of this work are provided.

Chapter 3 contains a detailed theoretical outline of the workings of the aggregated spatial choice model. The model is derived formally from first principles and its properties are described. A related topic of interest in this paper are the asymptotic properties of the aggregated logit model. It is shown that if the number of elemental alternatives is sufficiently large, it is not necessary to assume that the elemental random error components are Gumbel distributed for the distribution of the aggregate utilities to be Gumbel. Something which is not resolved in this thesis is whether these asymptotic insights are useful in the empirical context. In practice, the number of elemental spatial units in all aggregates would need to be very large for implementation of the more general aggregated model to be feasible. This is a matter of degree though and is not examined further in this thesis. While the theory of the aggregated model has always been oriented toward asymptotic properties (McFadden, 1978), we consider the functioning of the model when such properties clearly cannot apply. In that context, it is necessary to assume Gumbel distributed elemental errors.

Chapter 4 provides the results of a simulation experiment comparing the aggregated logit model with the OMNL model. The main objective of this chapter is to show that one may be taking a big risk in applying the OMNL model at the aggregate level in order to predict the choice of aggregate spatial units. A Monte Carlo experiment is conducted in which the observed aggregate choices are assigned based on randomly generated elemental utilities. In different trials, this procedure is done in such a way that the variance of elemental systematic utilities is gradually increased. As this occurs, the importance of the heterogeneity term in equation12 increases. The fact that the OMNL does not account for the heterogeneity term but instead estimates the expected aggregate utility on a mean utility basis is shown to be a shortcoming which gets progressively worse in heterogeneous aggregates. One restrictive assumption enforced in Chapter 4 is that the scale parameter at the elemental level is fixed at $\mu = 1$ implying that aggregates contain uncorrelated elemental utilities. If anything, this assumption would portray the OMNL in a more favourable light against the aggregated model than would be the case if μ was large. The reason for this conclusion is that the heterogeneity effect, which was the focus of the simulation, is maximized and thus its importance magnified when there are highly correlated utilities.

The work in Chapter 5 provides a basis for a relaxation of the restrictive assumption about μ. Clearly, the estimation problem with endogenous μ is unlike that of the nested logit model and the OMNL. While simultaneous estimation of the nested logit model is complex, difficulties have been bypassed in the past through sequential estimation procedures. Such an option does not exist with the aggregated model since the elemental choice information needed to estimate conditional choice probabilities is, by definition, absent. As a result, this chapter focuses on the different algorithms available to estimate the aggregated model and conducts tests of their performance. It is found that the specification of the μ scale parameters has a lot to do with the speed and reliability of convergence. Most estimation difficulties occur when the number of endogenous scale parameters is equivalent to the number of aggregates. When the number of endogenous scale parameters is reduced, either through equality constraints or exogenous fixing, algorithm performance shows marked improvement.

Having outlined the theory of the aggregated model, tested it in a simulated environment and settled on estimation procedures, the remaining task for this thesis, the testing of the model with real-world data, is the subject of Chapter 6. In particular, the test is undertaken in the context of modelling migration from Atlantic Canada in the 1990-1991 time period. The micro-data used for the analysis is characterized by having detailed information about individual choice-makers but having relatively poor spatial resolution regarding migration choices. While the data are such that it is difficult to make inferences about choices at the sub-provincial level, destination attribute information from other sources is available at as fine a level of aggregation as is desired. As a result, the data overall are a very good example of a problem suited to the aggregated model: detailed elemental attribute information but relatively coarse choice information. A series of models is estimated with the results indicating considerable potential for the aggregated model. Various specification techniques are explored and models which accommodate choice-maker heterogeneity are contrasted with models which emphasize only spatial heterogeneity and do not differentiate individuals in their systematic utilities.

The brief Chapter 7 of this thesis is included to provide a concise summary of the important findings of this body of work and also to comment on directions for future research.

Chapter 2

The Use of Discrete Choice Models in the Spatial Choice Context

2.1 Introduction

The goal of this chapter is to provide an overview of the application of discrete choice models to spatial choice problems. There are many spatial choice contexts such as recreational choice, tourism, industrial location, migration and shopping. All of these have in common the fact that the choice units considered by individuals are spatial units. The spatial choice context is subject to many complicating issues. Some of these are: that spatial units are characterized by differential levels of substitutability based on the fact that different patterns of similarity are perceived, that the issue of aggregation is prevalent, unlike in most other problems, and that the definition of choice sets is probably much more uncertain than in other contexts.

With regard to substitutability, it is clear that spatial units cannot be considered as independent choice entities. The utility of a given spatial unit is likely to be larger if it is located next to a particularly attractive spatial unit, something which would be unlikely if the two units were independent. Unfortunately, the ordinary multinomial logit (OMNL) model is consistent with the Independence from Irrelevant Alternatives (IIA) property which essentially states that pairwise relative relationships between the probabilities of alternatives will be unaffected by exogenous changes. The fact this property is inconsistent with anything other than independent alternatives has been the main motivator of criticism directed toward the OMNL and has contributed towards the testing of approaches such as the nested logit model.

The second main issue is aggregation. Spatial units can be sub-divided in an infinite number of ways which may or may not correspond to the fundamental aggregation level of the choice process. In a problem of inter-regional migration, for example, it would be ludicrous to model housing units as the choice set members. In the inter-regional context it is far more plausible to assume that choice-makers make their major evaluations at the level of urban centres and that the choice of the actual house is an incidental by-product of the fact that the city has been chosen. At the other extreme, there are studies of inter-regional migration (Liaw, 1990; Newbold, 1994) where the choices of provinces or regions are modelled without acknowledgment of the distinct utilities associated with metropolitan centres within aggregates. Ideally, we would want our data to be at the same level of resolution as the choice process. While attribute data are typically available at any level, the fact that choice data may be too coarse is a problem in need of consideration. The desire for greater generality in the modelling of aggregate spatial units has led to the implementation, in this thesis, of the aggregated spatial choice model.

The essential structure of the logit family of models is such that aggregation cannot be separated from substitutability. In the nested logit model and the aggregated model, nests are determined by virtue of their characteristics of substitutability and the nests themselves form the basis for how spatial units should be aggregated. As was illustrated in Chapter 1, a significant nest is defined in situations where, in the event of an exogenous change, we would expect the probability of an increasingly attractive elemental alternative to gain substantially at the expense of another within the nest but not significantly at the expense of others outside the nest. If such substitutability patterns are not observed, then there are clear grounds for questioning the given cluster of elemental alternatives which form an aggregate. Patterns of substitutability over space then, form the basis for how spatial units should be partitioned.

A third issue of importance in spatial choice modelling is that of choice set definition. Thill (1992) provides an excellent overview of this field and leaves the distinct impression that despite the best efforts of researchers, the endogenous modelling of choice sets is very much work in progress. The basic problem in the definition of choice sets, it is thought, is that each choice-maker should be modelled only with respect to spatial units that are in the person's actual choice set. In spatial choice it is normally argued that the full set of alternatives which are available over space are too many to be considered by the typical choice-maker (Fotheringham, 1986) and hence that some means is required to make the choice sets realistic. The matter of choice set definition is not reviewed in detail in this chapter since the literature in this field embodies an approach which is quite distinct from the focus of this thesis. Nevertheless, a few comments are offered on how the framework designed to accommodate aggregation and similarity of alternatives is also relevant to choice set definition.

The outline of this chapter is as follows: the first major section reviews substitutability issues while the second considers work on aggregation, particularly as relevant to the aggregated spatial choice model. In a concluding section it is noted that the theory underlying the nested logit model and the aggregated model is such that it can accommodate many of the modelling complexities outlined in the chapter while suggested alternative approaches typically leave something to be desired.

2.2 Substitutability

This first major section is itself divided in three. Initially, a chronicle of the spatial choice debate on substitutability is presented. It is noted that the gravity model approach was dominant in earlier years but that discrete choice approaches, particularly the logit, have recently begun to dominate. Secondly, an examination of the Independence from Irrelevant Alternatives (IIA) property and its implications is offered. Spatial choice researchers have tended to diminish the applicability of the OMNL since it is consistent with this property. Finally, a review is undertaken of approaches such as the probit and the nested logit model which are well-suited to non-IIA choice processes.

2.2.1 The Spatial Choice Debate

The issue of substitutability has been debated extensively in the spatial choice literature. The roots of the debate can be traced back to research done on the gravity model. While spatial choice was heavily biased towards the use of the gravity model in the 1970's after the contributions of Wilson (1970), the trend in the field beginning in the early 1980's was toward greater prominence for discrete choice models. Researchers such as Fotheringham (1986), who had earlier done work on gravity models, began to see limitations in the underlying theory of that approach and attempted to adapt ideas to the discrete choice framework. The shift in emphasis was encouraged by researchers such as Anas (1983) who demonstrated that while the gravity and logit approaches emerge from different theoretical backgrounds, their mathematical forms are identical for the purposes of implementation[4]. Of the two, the logit approach can be more widely applied since it is formulated at the micro-level and hence is adaptable, without loss of generality, to the macro-level problems for which the gravity model is designed.

At first, the essence of the gravity model was not well grasped. There was, for example, no direct discussion of the fact that the gravity model was consistent with the IIA property which is central to the issue of substitutability and hence to the understanding of the model. The main topic of interest was the behaviour of the impedance parameter and whether it was independent of spatial structure (Curry, 1972; Curry et al. 1975, Johnston, 1973; Griffith and Jones, 1980). Clearly much of the evidence in support of the fact that the two are not independent was explainable through the IIA property: the models of the time

[4] In particular, Anas showed that the OMNL arises directly out of the constrained optimization problem by which the production-constrained gravity model can be derived. Also, he illustrated that the doubly constrained gravity model is analogous to a joint logit model of origin and destination choice. The term 'gravity model', when seen in the text, can be taken to refer to the production-constrained model since this is the counterpart to the OMNL. The outflows from origins are taken as exogenous in both cases

were treating spatial units as independent entities when in reality they were not. Another problem was that the attractiveness of a given spatial unit was not conceptualized as a random variable. Hence, there was no mathematical basis for commenting on dependency among spatial units. A paper by Sheppard (1978) was one of the first, however, to consider the gravity model within a utility framework which at least acknowledged the presence of destination interdependence.

In the mean time, the properties of the logit model were becoming increasingly well understood and a separate group of researchers, primarily oriented toward transport demand problems, began applying the model to intra-urban problems such as housing choice (Lerman, 1975). Soon after, the step was being made in the logit literature toward the generalization of the model to account for varying patterns of substitutability, through the nested logit model, and also to account for aggregation (McFadden, 1978).

While the random utility framework underlying the OMNL was being generalized for more complicated problems, the gravity modelling group took the course of attempting to accommodate spatial structure effects without generalizing the theory of the model. The most noteworthy proponent of this approach was Fotheringham (1981,1983,1985) and his competing destinations hypothesis. His improvement to the basic gravity model was essentially to include a variable which measured the accessibility of a given destination to surrounding destinations that were 'competing' for flows. When the new accessibility variable was introduced, it was found that the systematic spatial trend in origin-specific gravity model parameters was eliminated. In the logit context, Fotheringham's approach was to increase the chances that the OMNL type model would have independent random error components by incorporating as much relevant information as possible in the systematic utilities. The model was not new but the variables in it were. Fotheringham's claim that he had developed a new theory would be equivalent to a practitioner of the OMNL model saying the same because he had done a good job of specifying the systematic utilities.

It can only be concluded that the general acclaim which greeted the competing destinations model was attributable to the unimaginative and restrictive specifications which had plagued the gravity model to that point. These specifications had a profound over-reliance on distance as an explanatory variable. Routine logit specification techniques, such as alternative-specific constants, were typically ignored in gravity models, causing the biased probability estimates to which Fotheringham (1983) alludes. The paper of Black (1983), for example, details techniques that had been known to discrete choice modellers for some time. Recent work by Lowe and Sen (1994) is noteworthy in that it tests the performance of alternative-specific constants in the gravity context.

Fotheringham's work subsequently was subject to much criticism. Some took issue with the airline data that Fotheringham had used to test his model (Ewing, 1986). Gordon (1985) was one of the first to criticize the competing destinations model on the grounds that it attempted to explain the substitutability phenomena in purely spatial terms. He rightly points out that models of spatial interaction involve choices made by individual agents or consumers. However, the aggregate nature of the gravity model promoted a tendency to personify origins as if they were autonomous decision-making units. As Webber (1980) says however, origins do not behave. Recently, Lo (1990, 1991a, 1991b, 1992) has expanded upon earlier themes. She joins Gordon in noting a preoccupation in the spatial choice literature to seek spatial explanations for interaction patterns, when economic and other non-locational relationships will often be at the core of the explanation. She notes after all, that spatial interaction is a derived demand; that is, it is a means and not an end and that there is an important distinction between locational substitutability and economic substitutability. She states that there is room for both in explaining spatial choice behaviour but that few processes are either one or the other.

The second half of the 1980's marked the beginning of a clear trend in which discrete choice modelling approaches tended to supplant the gravity model as the preferred model for spatial choice. Gravity modellers such as Fotheringham (1986), who even in his earlier papers was using terms like 'utility' and 'hierarchical choice', began to use logit models. Applications in spatial choice were reported such as intra-urban migration (Boots and Kanaroglou, 1988), housing choice (Fischer and Aufhauser, 1988), industrial location (Hansen, 1987); recreational demand (Caulkins et al, 1986) and inter-regional migration (Liaw and Ledent, 1988; Kanaroglou et al., 1986). The mainstream discourse on spatial choice modelling issues shifted unmistakably toward the discrete choice approach (Borgers and Timmermans, 1987; Haynes et al., 1988; Haynes and Fotheringham, 1990; Thill, 1992). There is no question that much of this shift in emphasis had to do with the micro-perspective of the logit approach which was suited to the increasing availability of micro-data. The prospect of being able to model alternatives possessing differential levels of substitutability through the nested logit model was also an explanation.

2.2.2 The IIA property

In the spatial choice literature, there is a tendency to take the IIA property very literally (Haynes et. al, 1988; Haynes and Fotheringham, 1990; Fotheringham, 1986) with many researchers tending to dismiss the OMNL model out of hand. It is reasoned that since spatial alternatives are not viewed by choice-makers as independent entities, the IIA property does not apply and the OMNL model is inappropriate. There are, for example, statements to the effect that, with substitutable subsets of alternatives, the OMNL will invariably result in pathological predictions of choice probabilities. Such conclusions are overstated. For this reason, an assessment of the IIA property is undertaken here.

2.2.2.1 Implications of the IIA Property

The IIA property was first proposed by Luce (1959) as an axiom underlying choice processes. Essentially the property implies that the relative attractiveness of any pair of alternatives should not depend on the presence of other alternatives in the choice set. It is important to understand that the IIA property applies in a **relative** sense only. It certainly does not imply that **absolute** choice probabilities are independent of choice set size and composition. Another important point is that an IIA process is one which holds generally for individual choice-makers, not only the aggregate market shares. This is a particularly important point in spatial choice where choice-makers are likely to have entirely different conceptualizations of spatial units based on the location of their origin.

Some choice scenarios are clearly not IIA-consistent choice processes. In the much-quoted red bus/blue bus problem, it seems clear that the addition of the blue bus option to a choice set containing red bus and automobile would tend to result in a new set of probabilities in violation of the IIA property. The ratio of the red bus/automobile probabilities would not be maintained in the before and after scenarios since the new blue bus option would claim its market share from those who before used to ride the red bus and not from those who drove. Expressed alternatively, if there was some increase in utility of blue buses (perhaps their comfort level was improved) then the red bus probability would suffer much more than the automobile probability. If it is the automobile alternative which is removed from the choice set, though, then there is a good chance that the relative attractiveness between bus types would be unaffected. Hence IIA clearly does not apply when the relative attractiveness of auto and one bus type is compared but is a much more reasonable assumption when the ratio of the two bus probabilities is considered. The clear implication is that it is often too simplistic to say whether an entire set of alternatives is consistent with the IIA property. The more likely answer is that IIA will apply within some subsets of the choice set but not others. This of course is the basis for the nested logit model.

2.2.2.2 The IIA Property and the Logit Model

The fact that the OMNL possesses the IIA property has been the basis for much of the criticism it has received in the spatial choice literature. The property is reflected by the mathematical structure of the OMNL in that the ratio of two probabilities for any two alternatives will be independent of the presence of other alternatives. Hence:

$$\frac{P_n(i)}{P_n(k)} = \frac{\dfrac{\exp[V_{in}]}{\sum_{j=1}^{L}\exp[V_{jn}]}}{\dfrac{\exp[V_{kn}]}{\sum_{j=1}^{L}\exp[V_{jn}]}} = \frac{\exp[V_{in}]}{\exp[V_{kn}]} \tag{1}$$

where it can be clearly seen that the final expression for the ratio of the probabilities for alternatives i and k relate only to the strict utilities of those two alternatives.

An important point to make, however, and one which is typically overlooked in the spatial choice literature, is that an IIA choice process is not necessary to apply the OMNL (Train, 1986, pp. 21-24). Any choice scenario that can be expressed in choice probabilities can be represented in the form of an OMNL model. Such flexibility exists so long as alternative-specific constants are included in the specifications of utility for the various options. Constants are used to account for all systematic variation that is uncaptured by other elements of an alternative's systematic utility and their inclusion allows reliable estimates of the other parameters in a model. The magnitude of an alternative-specific constant, by definition, is dependent on the presence of all alternatives in a model since it responds to the proportions of choice-makers choosing each of the alternatives. Hence V_{in} and V_{kn} themselves are functionally dependent on the systematic utilities of **all** alternatives. This result would seem to contradict, for example, the assertions by Haynes et al. (1988) that in the OMNL, probabilities for similar alternatives will necessarily be biased upwards and those of dissimilar alternatives will be biased downwards. By the same token, it supports the findings of Borgers and Timmermans (1987) that the OMNL was quite robust in their set of simulations.

The fact that one is not really **modelling** substitutability per se when the OMNL with constants is applied is a limitation. Hence when the OMNL model is applied to a non-IIA process, a re-estimation would typically be required any time there is a change in circumstances, such as the inclusion of a new alternative or an increase in the utility of an alternative. Avoidance of re-estimations under such circumstances has typically been rare, even when it was thought that the IIA property stood a good chance of holding (Train, 1978). The nested logit and probit models respond better to such changes since they are generalized to incorporate the pattern of similarities across subsets of alternatives and hence are not dependent on a set of calibrated constants. It would be a useful contribution if the constants in the OMNL could systematically be altered to respond better to exogenous changes. Perhaps the literature on choice-based sampling (Thill and Horowitz, 1991; Kanaroglou, 1994), in which adjustments to the constants are an integral part of the theory, might be of assistance.

The OMNL can be made less dependent on constants by replacing them with attributes of alternatives. Such an approach attempts to model how the constants (or at least that portion of systematic utility which does not vary across choice-makers) would be affected in the event of an exogenous change. From the point of view that a more complete accounting of the choice process is implemented, this is a laudable action, but there are two difficulties. One is that since the role of the constants is only being approximated, there is a risk of substantially biased probability estimates and adverse effects on other parameters. Largely this is because attributes of alternatives are not specified in an alternative-specific manner; hence an attribute may represent the utility of one alternative very well but not the others. Secondly, the inclusion of several attributes of alternatives will often lead to collinearity problems.

Overall, the OMNL is a more robust model than tends to be portrayed in the spatial choice literature. If the goal of an analysis is to provide a good overall explanation of choices from existing data, then the OMNL model will do an adequate job and with small computational cost. This will be true irrespective of whether the choice process is consistent with IIA. Typically, applications utilizing the OMNL and other models such as the nested logit and the probit find that the fit of the more sophisticated models might be only marginally better but at much greater costs in terms of time and effort and that conclusions often are not significantly altered.

2.2.2.3 Variability of IIA Across Origins

In spatial choice problems, the patterns of substitutability between subsets of alternatives may well vary across choice-makers implying that spatial alternatives are classified or grouped in different ways. In applications, however, there has been a tendency to assume that the perceived patterns of similarity will be repeated across all choice-makers. The origin of the choice-maker, however, will have considerable bearing on how alternatives are grouped and therefore what the patterns of substitutability between alternatives are. In the migration context, for example, most residents of Ontario would tend to cluster potential destinations in the Eastern townships of Quebec, thus perceiving any one as being a perfect substitute for another. A resident of Quebec, having more intimate knowledge of the area, might well be able to make strong distinctions between alternatives and would not perceive them as being substitutes.

As a solution, the technique of origin-specific models (Fotheringham, 1983) can be employed. In this case, a separate destination choice model is estimated for each origin in the data and a unique set of parameter estimates obtained in each case. Of course, different aggregation schemes can be employed across origins in acknowledgment of differing perceptions. It is desirable but complex to accommodate such phenomena within a single model. While probit models do so since they are capable of capturing taste variations in both the systematic and random error components of utility, these models do not handle spatial aggregation as easily as the logit family.

2.2.3 Approaches to Modelling Substitutability in Spatial Choice

In this section, some of the alternatives to the straightforward application of the OMNL model are considered. The potential approaches are: 1) extensions of the basic OMNL formula and the use of models which are not derived from random utility theory, 2) use of the probit model, 3) use of the nested logit model. These are discussed in turn with comments offered on past applications of such techniques. While the previous section has pointed out that application of the OMNL model to spatial choice problems will result in good estimates of probabilities, the alternatives below may offer more insight into the implications of substitutability.

2.2.3.1 Logit Extensions and Models Not Based on Random Utility Theory

Several researchers have proposed extensions to the basic logit formula as a means of accommodating substitutability (Meyer and Eagle, 1982; Borgers and Timmermans, 1987; Fotheringham, 1983). In general, the extensions of the standard OMNL formula include an additional term which measures the average degree of similarity between one alternative and the others in the choice set. The general form is:

$$P_n(i) = \frac{R_i \exp[V_{in}]}{\sum_{j=1}^{L} R_j \exp[V_{jn}]} \tag{2}$$

where R_i is the measure of the average degree of similarity for alternative i. A variety of ad hoc expressions have been used for R_i in an attempt to capture similarity effects. One suggested by Borgers and Timmermans (1987) is of the form:

$$R_j = \prod_{m=1}^{M} \left[\frac{1}{L-1} \sum_{k} |x_{jm} - x_{km}| \right]^{\frac{\phi_m}{M}}$$

where x_{jm} is the value of attribute m on alternative j, $m = 1, 2, ..., M$, L is the total number of alternatives and ϕ_m are 'substitution' parameters to be estimated. The R_j term is clearly a weighted measure of how the attributes of one alternative differ from the attributes of all other alternatives. The term should be of small magnitude for alternatives which are clustered together in terms of attribute values. The substitution parameter should be zero for attributes which contribute nothing to assessing substitutability. Another approach to modelling R_j was suggested by Haynes and Fotheringham (1990) whereby the accessibility measure from Fotheringham's competing destinations approach:

$$R_j = \left[\frac{1}{L-1} \sum_{k \neq j} \frac{W_k}{d_{jk}} \right]^{\phi}$$

would be utilized where ϕ is a parameter to be estimated.. This accessibility term is essentially a measure of the pattern of alternate spatial opportunities in which W_k is a measure of the attractiveness of the set of alternatives other then j and d_{jk} is some measure of the distance between the jth alternative and an alternate alternative. Unlike the R_j measure of Borgers and Timmermans which is small for alternatives perceived as 'similar', this version of R_j will get larger when there are several large spatial substitutes near alternative j. Clearly, this measure is inherently spatial, unlike the one of Borgers and Timmermans.

In assessing this type of approach, it is important not to interpret such 'extensions' as generalizations. First, it should be remembered that the OMNL is reasonably robust itself in handling cases where the IIA property does not apply. Secondly, these extensions do not proceed from first principles and are thus quite arbitrary in nature. It has been extensively outlined in Chapter 1 that discrete choice models result from assumptions made about the joint distribution of the random error components. It was shown that the OMNL model arises from quite restrictive assumptions about the random error components: namely that they are uncorrelated and homoscedastic. Under these circumstances, the mathematical form of the model emerges and turns out to be consistent with the IIA property. Typically, in the papers which propose extensions to the basic logit formula, the properties of the random error components and their implications for the final form of the model are never extensively examined. It is never mentioned that any generalization of the OMNL model will arise from less restrictive assumptions about the variances and covariances of utilities relating to the set of alternatives.

In the end, the proposed extensions are simply alternative specifications of the OMNL model. None of the expressions for R_j above lead to different assumptions about the random error components. Since the same assumptions still apply, the outcome of the derivation would still be the OMNL as first outlined by McFadden (1974). In fact, the R_j terms are simply elaborate variable specifications which enter into the expressions for systematic utilities in the same manner as any other variable. This can be seen more clearly by writing equation 2 as:

$$P_n(i) = \frac{\exp\left[V_{in} + \ln R_i\right]}{\sum_{j=1}^{L} \exp\left[V_{jn} + \ln R_j\right]}$$

where it can be seen that the essential OMNL structure is unaltered and that the additional term is simply incorporated into the systematic utility of each alternative.

Another criticism of these models is that they assume all similarity or dissimilarity to be assessed on the basis of destination attributes which vary only across space and not choice-makers. As Ben-Akiva and Lerman (1985) point out, the incorporation of taste variation among the different population segments is an important component in model specification to ensure that the risks of serious violations of the IIA property are minimized. Such precautions relate to the fact that the IIA property should be obeyed for each choice-maker in the sample, not for the aggregate market shares. Since the R_j terms do an inadequate job of capturing taste variation, they are an inadequate overall representation of substitutability patterns.

There are other potential spatial choice models to consider which are not necessarily from random utility theory. The dogit model (Gaudry and Daganeis, 1979) is not considered as a prime candidate for spatial choice problems. First of all, in the original paper, there is no mention of random utility theory and whether the model was derived from such principles. McFadden (1981) is critical of the dogit model on the basis that patterns of substitutability are not general. Nevertheless, according to McFadden, the model is consistent with random utility theory. Regardless, the model has not gained favour with researchers and is not applied in spatial choice problems.

Another model advocated for spatial choice is the translog model described by Lo (1990). This model is not related to random utility theory although it arises from the transcendental logarithmic indirect utility function. It is interesting because it allows general patterns of substitutability between all pairwise arrangements of alternatives. Any two alternatives can be perfect substitutes, perfect complements or independent. Using simulations, Lo illustrates that a system of close substitutes has the effect of increasing the friction-of-distance since choice-makers have less incentive to overcome it, while a system of perfect complements has the effect of shrinking space since consumers are encouraged to visit multiple destinations. Lo (1991a) mentions that the modelling of complementarity is a matter of relevance for multi-purpose interaction behaviour while the issue of substitutability is within the domain of single purpose interactions. Her criticisms of discrete choice modelling are unfair in this context since these models are clearly designed for single discrete choices.

Lo's account of the translog model in general is lacking because there are no empirical examples to back up the workings of the model and to show that it can be estimated. While the model seems quite general with respect to the patterns of substitutability, it appears limited in the extent to which a wide range of explanatory variable can be used to characterize spatial interaction. For example, it is mentioned that the friction-of-distance is not explicitly estimated. Distance enters only indirectly in the calculation of effective prices. Overall, spatial interaction is predicted on the basis of substitution parameters, household incomes and price relationships. The extent to which flexibility in specification is permitted, given these constraints, is not elaborated upon.

2.2.3.2 The Probit Model

There is no question that the probit model, in theory, has many capabilities that are desirable in a spatial choice model. A completely general pattern of substitutability among pairings of alternatives can be accommodated and these relationships can be specified to vary across choice-makers or origins. In this sense, the probit model certainly possesses the generality that Lo claims for the translog model. The difficulty with the probit model is one of implementation as was outlined in Chapter 1. For computational reasons, it is extremely difficult to accommodate a large number of alternatives and the process of obtaining general patterns of similarity across alternatives is typically inefficient in the sense that 'nuisance' parameters proliferate. Ben-Akiva and Bolduc (1991) suggest a partial solution to this problem in a paper discussed in more detail below.

In spatial choice contexts, researchers typically talk about the model's favourable theoretical properties (Haynes et al, 1988) but rarely apply it for the reasons mentioned above. A notable exception is the study by Miller and Lerman (1981) on the choice of shopping district within Boston although this model did not have a large number of spatial alternatives. Overall, the realm of the probit has been restricted to problems like mode choice with small numbers of alternatives. Spatial analysts have had more use for the probit in the modelling of binary response variables which are spatially referenced, which of course is a distinct problem from that of spatial choice. An example is the paper of McMillen (1992) where a binary high/low crime response variable is used in a probit which takes account of spatial autocorrelation and heteroscedasticity in the random errors. The approach is econometric and not associated with random utility theory. The autologistic model outlined in Haining (1991) is an example of a logit-type approach with a categorical response variable.

The paper by Ben-Akiva and Bolduc (1991) represents an early initiative in the use of spatial statistical concepts for choice problems and is discussed here in some detail since some of the paper's features are relevant to spatial choice. In particular, they model random error components through a Generalized autoregressive process of the form:

$$\varepsilon_n = \rho \mathbf{W}_n \varepsilon_n + \xi_n$$

where ε_n is an $L \times 1$ vector, $\mathbf{W}_n$ is an $L \times L$ weight matrix and ξ_n is an $L \times L$ diagonal matrix containing standard normal variates on the diagonal. Additional sources of generality are the facts that they permit a random coefficients structure for the parameters associated with explanatory variables and that they include an i.i.d. Gumbel error structure in addition to the normally distributed errors so that the logit model is obtainable as a special case. All of this is accommodated within a probit model structure in which probabilities are estimated through simulation techniques. While their application, the choice of different telephone service options, is aspatial, they mention how their technique is applicable to problems with large choice sets since the simulation technique is computationally much less burdensome.

The other important aspect is the use of a binary weight matrix to determine which pairings of alternatives are to have mutual dependencies. Such an approach greatly reduces the number of 'nuisance' parameters which are associated with the variance-covariance matrix Ω_n of the random errors. With the autoregressive approach, the number of parameters increases only linearly with the number of alternatives whereas in less effective specifications, a geometric relationship would tend to apply. In particular, the structure of the Ω_n matrix they hypothesize is a function of only five parameters: the ρ autocorrelation parameter, which is a measure of the strength of the autoregressive process, and four variance parameters.

While the model is impressive, the applicability of the approach to large-scale spatial choice problems remains in doubt. Innovations such as simulated probabilities and parsimonious Ω_n specifications have greatly assisted in the applicability of the probit model but there is no indication that problems with upwards of 100 elemental spatial units can be adequately handled. In a later paper, Bolduc (1994) applies the technique to a mode choice problem of nine alternatives but as yet the effectiveness of the technique for an extremely large number of alternatives is unproven.

2.2.3.3 The Nested Logit Model

The nested logit model has been applied more frequently than the probit in the spatial choice literature, mainly because it is perceived as the best compromise between computational tractability and theoretical generality. Spatial choice applications have included: industrial location (Hansen, 1987), housing choice (Fischer and Aufhauser, 1988; Thill and Van de Vyvre, 1989), migration (Anderson and Papageorgiou, 1994; Newbold, 1994) and recreational choice (Parsons and Needleman, 1992) to name a few. The capability of the model to accommodate differential levels of substitutability between subsets of alternatives is a major reason for the model's application to problems where a perceived violation of the IIA property exists.

One criticism of the nested logit is that its estimation tends to be more involved than that of the OMNL. The sequential estimation approach has gained considerable popularity since it is implemented as a series of OMNL models linked by inclusive value terms which measure the expected maximum utility to be derived from lower level nests. Unfortunately, statistical inference is likely to be compromised in the case of parameters estimated at the upper level. Brownstone and Small (1989) discuss simulations where the standard error of the μ parameter associated with the inclusive value is substantially downward biased and in need of the complex corrections advocated by Amemiya (1978) and McFadden (1981). The standard error bias stems form the fact that the inclusive value term, which is a random variable, is passed as a fixed term to the upper level when ideally all aspects of the model should be endogenously and simultaneously determined. Brownstone and Small go to the extent of saying that with sequential estimation, statistical inference should not be attempted without correcting the μ standard error.

Clearly the problem of statistical inference with the μ parameter is troubling in the spatial choice context since this is an important theoretical parameter. However, this is a problem with the estimation procedure much more than the model. Yet spatial choice applications, to the best of my knowledge, have always utilized the sequential procedure. The clear solution is the widespread adoption of simultaneous full-information estimation which in reality is not any more burdensome than the process of going through sequential estimation followed by the correction. Also, as mentioned by Train (1986, pp. 75-76) simultaneous estimation results in far more parsimonious models since there is no need to specify the same variables separately at different levels of the model. The types of algorithms which are available for simultaneous estimation are discussed in Chapter 5 and software for the simultaneous estimation of nested logit models is described in the Appendix.

A destructive aspect of sequential estimation has been its effect on how researchers interpret the choice process. Since the model is estimated sequentially there is a tendency for some to assume that this mirrors exactly how the choice is made. Fotheringham (1986), for example, states at the beginning of his paper: 'Consider an individual i who has made a decision to move but has yet to choose a particular destination j'. Such an interpretation ignores the fact that individual lower nest alternatives in a nested logit are not precluded from affecting how an upper level choice is made. Use of the nested logit model does not automatically assume a hierarchical choice process since the OMNL is a special case when elemental utilities are uncorrelated.

The nested logit model has certainly not gone without criticism in how it handles substitutability. One of the major criticisms in the spatial choice context is that exogenous determination of a tree structure is too arbitrary and liable not to correspond to the actual patterns of similarity (Lo, 1991b; Fotheringham, 1986; Borgers and Timmermans, 1987). In a non-spatial choice problem with a relatively small number of alternatives, the best grouping is usually intuitively obvious. The number of spatial alternatives, however, is typically large, thereby resulting in several possible patterns of similarity. The problem of an accurate clustering of spatial units, however, is not one that is unique to the nested logit model. Such clustering is the only means to capture the patterns by which the IIA property applies and is an issue which underlies the choice process in general. The fact that the nested logit model provides a systematic means to test for these patterns is an asset more than a liability. Another asset is that unlike the competing destinations model, for example, it is not necessary to make an a priori assumption that substitutability must occur in purely spatial terms based on contiguity. A process which occurs based on an economic hierarchy can also be accommodated.

The nested logit model also receives criticisms about the specific patterns of similarities that it can represent (Lo, 1991b, 1992). In assessing such criticisms it is important to remember that a model is by definition an abstraction of reality which is intended to capture the essential processes. As seen in Chapter 1, a nested logit treats the correlation between all pairings of elemental alternatives within an aggregate as equal and positive while alternatives within different aggregates will be represented as having uncorrelated utilities. By downweighting the utilities of correlated alternatives, it can be argued that the model is capturing the essential mechanism at work when people assess a set of spatial units. The fact that the relationships governing similarity can be differentiated across nests or be made more realistic by adding another choice level is also not to be neglected and can result in quite sophisticated patterns of similarity. The translog model suggested by Lo (1990) on the other hand provides no mechanism for the phenomena of clustering and is guilty of incorporating too much detail since parameters are estimated for all pairwise relationships.

2.3 Aggregation

While the spatial choice debate on substitutability has been quite lively, with several papers written on the topic. The same cannot be said for the issue of spatial aggregation in choice models. This should become clear in the discussion below. This section on aggregation has three components. Initially, a background on some of the theory underlying spatial aggregation in choice modelling is offered. Secondly, the implications of this theory for choice set modelling are briefly examined. Finally, a closer look at empirical applications central to the subject of this thesis is undertaken.

2.3.1 Theoretical Background

In spatial analysis, aggregation has received much more attention in the context of the modifiable areal unit problem (MAUP) which has to do with how various statistical measures can give vastly different results based on the level of spatial aggregation at which they are applied (Openshaw, 1984). Other researchers (Arbia, 1989) have devised theoretical constructs to assess whether such aggregation effects (more popularly known as 'scale effects' in the MAUP literature) can systematically be predicted. Still others (Amrhein, 1995; Amrhein and Flowerdew, 1992) have been testing econometric models at differing levels of aggregation in search of patterns. Aggregation in this body of literature, typically concerned with modelling the variation of some attribute over space, is not to be confused with aggregation in the spatial choice literature. Nevertheless, the work of Jones and Bullen (1994), on the modelling of housing prices over space, bears an interesting resemblance to the aggregated logit model in the sense that their model incorporates random error components which function at different spatial levels of resolution. Still, what they consider is not a choice process.

2.3.1.1 Spatial Aggregation versus Choice-maker Aggregation

Even within the choice literature there are distinctions to be made in terms of the types of aggregation. On the one hand, there is aggregation across choice-makers. Each choice-maker in a given origin will have a different choice probability for some destination. The question is, if variables can be collected at the level of the individual, how can this information be utilized to yield one choice probability to act as a representative for all members of that origin? There is an extensive body of theory (see Ben-Akiva and Lerman, 1985, Ch. 6) which says that obtaining a simple average of the set of relevant attributes across individuals and then using these representative variables to estimate choice probabilities is not the best approach.

The second type of aggregation relevant to spatial choice models, and the one which is the main topic of this thesis, has to do with aggregation of spatial alternatives. Assuming that we are able to obtain a representative probability for a given origin by aggregating across choice-makers, it is important to note that the spatial interaction probabilities for prospective destinations will depend on how these destinations are partitioned. It is often the case that the partitioning will be such that the spatial units involved will be aggregations of the fundamental spatial units or *elemental alternatives* considered by choice-makers. In a problem of inter-regional migration, for example, a choice involving a move of several thousand kilometres might well be motivated by the attractiveness of a particular urban agglomeration. The best available choice data, however, might say only which province was chosen.

Most modelling efforts in the past would accept this limitation and apply the model in such a way that all attributes would be collected at the same level as the choice information. In this respect, aggregation in choice modelling at present is viewed as spatial autocorrelation was in the early 1970's (Gould, 1971). It is seen more as a nuisance to be assumed away than an opportunity on which to define and distinguish spatial choice analysis.

2.3.1.2 The Representation of Aggregate Utility

The aggregated model assumes that the utility of a spatial aggregate is the outcome of the maximization of those random utilities which apply at the true level of the choice process. The modelling question essentially is: Given the distribution over space of explanatory attributes at the level of the choice process, what pattern of elemental utilities would be most likely to give rise to the observed pattern of aggregate level choice behaviour? In other words, we attempt to assess the attractiveness of elemental alternatives even though we are unsure of those that are chosen. Since we have means to assess which destination attributes are important in molding the pattern of aggregate level choices, we can use elemental attribute information to infer things about the attractiveness of elemental alternatives. Such a modelling approach is far more ambitious than that of simply applying the OMNL model at the aggregate level as has typically been the practice.

The roots of this approach lie in the work of Lerman (1975) in the context of intra-urban housing choice where researchers are aware of the aggregate level neighbourhoods which are chosen but not the particular dwelling units. Using a first-order Taylor expansion, and assuming Gumbel distributed elemental utilities, the form he deduced for the systematic utility of aggregate J from the perspective of choice-maker n was:

$$V_n^J \simeq \overline{V}_n^J + \ln M^J \tag{3}$$

which corresponds to12 in Chapter 1 except for: the omission of the heterogeneity term, the omission of the scale parameters, and the fact that the expression is an approximation. Lerman noted that this form had limitations when dealing with heterogeneous aggregates but states that in general the size effect would be of more prominence than any missing heterogeneity effect. Two ways in which the importance of the heterogeneity effect are minimized are, on the one hand if the elemental systematic utilities in a given aggregate are equal to $\overline{V}_n^J$, and on the other if the heterogeneity effect does not vary significantly from aggregate to aggregate. This contribution by Lerman was the first theoretical recognition that it was necessary to correct utility for aggregation, however, the approach did not go so far as to model elemental utilities.

The result of the second-order Taylor series expansion of Kitamura et al. (1979), again assuming Gumbel elemental utilities, revealed that the systematic utility of an aggregate alternative could be even more accurately approximated as:

$$V_n^J \simeq \overline{V}_n^J + \frac{1}{2}Var(V_{jn}) + \ln M^J \tag{4}$$

the essential improvement over 3 being the addition of a term measuring the variance of the elemental systematic utilities within the aggregate. It is important to emphasize that both 3 and 4 assume that all elemental random error components in the choice set are independent and that there is no random error which is unique to groupings of elemental alternatives so that degrees of similarity can be accommodated as in the nested logit model. The paper of Kitamura et al. is also interesting in that the implications of assuming other distributions for the elemental utilities are assessed. It is shown that a model similar in spirit to the aggregated model can be derived by assuming normally distributed elemental utilities. The resulting approximation for the aggregate systematic utility is:

$$V_n^J \simeq \overline{V}_n^J + \frac{1}{2}Var(V_{jn})\sqrt{2\ln M^J} + \sqrt{2\ln M^J}$$

where the most interesting differences are that the approximation to the heterogeneity effect is interactive with the magnitude of the size effect and that the size effect in relation to 3 and 4 will be smaller, particularly for large M^J. In addition it is shown that if elemental utilities are normally distributed, then the distribution of the maximum utility converges weakly, after a normalization, to the Gumbel distribution. Unlike the case with Gumbel distributed elemental utilities, where the variance of the distribution of the maximum never changes, with normally distributed utilities the distribution of the maximum has shrinking variance as the number of elemental alternatives increases. While this property is desirable, the resulting choice model does not have a closed form and is therefore much less amenable to application than the Gumbel model.

The contribution by McFadden (1978) was important because the expression for the aggregate systematic utility:

$$V_n^J = \overline{V_n}^J + \frac{1}{\mu^J} \ln \frac{1}{M^J} \sum_{j=1}^{M^J} \exp\left(\mu^J (V_{jn}^J - \overline{V_n}^J)\right) + \frac{1}{\mu^J} \ln M^J$$

first shown in Chapter 1 and repeated frequently elsewhere in this thesis, is obtained as an equality instead of an approximation and because the theory which accommodates aggregation is shown to mirror that of the nested logit model. The Taylor series approximations of Lerman (1975) and Kitamura et al (1979) had certainly made no links to the nested logit model since no similarity among elemental alternatives could be accommodated. These issues are discussed in much more detail in later chapters. Suffice it to say that the capability to accommodate differing levels of similarity among subsets of alternatives allows for an elegant interplay between the size and the heterogeneity effects and a rich theoretical interpretation.

2.3.1.3 Asymptotics and Spatial Aggregation

The literature on spatial aggregation in discrete choice models has put a lot of emphasis on asymptotic properties which apply when the number of elemental alternatives in aggregates becomes very large. In particular, asymptotics have been shown to be an issue with respect to the assumptions required for elemental random error components and in regard to simplification of the heterogeneity term. Asymptotic properties have been emphasized based largely on the nature of intra-urban housing choice, the field for which choice models of spatial aggregation were first developed. Since elemental alternatives in that context are individual dwelling units, the typical aggregate alternative in such a problem would contain an extremely large number of elemental alternatives.

With regard to the random error components, research in the statistics of extremes (Galambos, 1978) indicates that random variables do not need to be Gumbel distributed to give rise to a Gumbel distribution under maximization. Much of this thesis is premised on the notion that elemental utilities are Gumbel distributed. Mainly this is because the types of choice problems considered, unlike McFadden (1978) and Lerman (1975), are not guaranteed to have a sufficiently large number of elemental alternatives. It is shown in Chapter 3 that the asymptotic result, which does not depend on assuming Gumbel elemental errors, yields the same model as the scenario we consider. In other words, the end result is the same but the asymptotic model has less restrictive assumptions about the elemental errors underlying it.

Turning to asymptotics and the heterogeneity term , if it can be assumed that the systematic utilities of elemental alternatives are normally distributed around $\overline{V}_n^J$, then as the number of elemental alternatives becomes large the heterogeneity effect approaches $\frac{1}{2}Var(V_{jn})$ which of course is consistent with the additional term arising from the second-order Taylor series expansion. According to McFadden (1978), this asymptotic property is advantageous because it allows the use of standard logit estimation software. Information detailing the relationships between attributes which enter V_{jn} can be used in conjunction with estimated parameters to obtain the variance of the systematic utilities. While this may be a useful result for some problems, in this thesis the approach has been taken that there is no need to approximate the heterogeneity term since use of a non-standard logit estimation will yield the exact results. Also, the numbers of elemental alternatives in some contexts will be inadequate and also the assumption of normally distributed systematic utilities within aggregates might be overly restrictive.

To conclude this section, while the literature on spatial aggregation in choice models is not voluminous, it is very insightful and is supported by a rich body of theory. Much of the original work has been done by researchers who are not primarily spatial analysts but found themselves having to deal with the intra-urban housing choice problem. When spatial aggregation has been considered in the core spatial choice literature, it has been in the context of the gravity model based on entropy theory and not utility maximization (Webber, 1980; Batty and Sikdar, 1982abcd; Schwab and Smith, 1985, Putman and Chung, 1989). Like the wider comparison between discrete choice approaches and gravity model approaches, the discrete choice approach to spatial aggregation is the more likely to have a prominent future.

2.3.2 Choice Set Definition and Aggregation

A substantial amount of attention has been given to the issue of choice set definition in spatial choice modelling (see Thill, 1992 for a review). There have been questions about the extent to which choice-makers are capable of processing all the information associated with a large number of spatial alternatives. The general consensus is that if some alternative is not considered by an individual, then it should be removed from the choice set to avoid biased estimates of parameters and probabilities. The body of work is oriented toward modelling the likelihood that individual elemental alternatives are in the choice set. In the spatial choice context, however, there is a case to be made that choice-maker's do not consciously omit spatial units from their choice set but instead cluster spatial units, thereby considering the clusters as homogeneous entities.

The spatial choice problem can be conceptualized in these terms. It can be assumed that choice-makers perceive a continuous surface of utility over the study area and that they select their alternative based on where this surface is highest. This conceptualization is consistent with the so-called continuous spatial choice model (Ben-Akiva and Watanadada, 1981; Ben-Akiva et al., 1985) which is derived as a limiting case as the number of elemental alternatives in an aggregate becomes large (see Chapter 3 for further discussion).While it is unclear at this point, largely for reasons of data availability, whether the continuous model itself is implementable, the idea of a surface of utility unfolding over space is intuitively appealing. Nevertheless, using conventional choice models, the surface can be approximated at a number of discrete locations in keeping with the way that spatial attribute data are made available to researchers. Essentially, research on choice set composition is saying that we need to find means to separate the small discrete units that are in a person's choice set from those that are not. Thus at selected zones in space the surface presumably would be undefined.

In a mode choice problem, if some mode is unavailable to a given choice-maker then there is little doubt that the given mode should be removed from the individual's choice set. This is a clear-cut problem with a clear-cut solution. In spatial choice contexts, it is a far more risky proposition to ascertain through some means what spatial units are considered, especially on an individualized basis. Perhaps a better approach is to make the assumption that choice-makers have a tendency to cluster that which they do not know much about. Most likely, the spatial region of uncertainty will encompass a cluster of several elemental alternatives; thus while the choice-maker might know little about the constituent elemental alternatives, the aggregate zone itself will have some meaning.

The theory of the aggregated model contains a mechanism for clusters of indistinguishable elemental alternatives to be represented properly. In particular, the expression for the systematic utility of aggregate J is given by:

$$V_n^J = \frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} \exp\left(\mu^J V_{jn}^J\right)$$

which can be re-written as a function of a mean utility term and size and heterogeneity effects. As $\mu^J \to \infty$, the expression has the interesting property that $V_n^J = \max_{j \in \{1,2,\ldots,M^J\}} V_{jn}^J$ or namely that the aggregate systematic utility will correspond with the largest systematic utility from among the M^J elemental alternatives in aggregate J. The implication is that the elemental alternative with largest systematic utility will act as a representative for all within that aggregate. There is no chance of some other elemental alternative offering a higher level of utility because random error components are perfectly correlated. The size effect goes to zero implying that the number of elemental alternatives is irrelevant to the predicted level of aggregate utility. The effect of all this is that individual elemental alternatives within that cluster, many of which will not be individually considered by the choice-maker, will have no effect on the aggregate level probabilities. On the other hand, if individual elemental alternatives are meaningful to the aggregate level probabilities, then $\mu^J \rightarrow 1$, which will make the size term quite relevant and allow the individual elemental alternatives to behave as independent entities.

This mechanism in theory should work well in the context of the aggregated model because the interest is in the correct representation of aggregate probabilities and the model is estimated on the basis of how well it conforms to the aggregate level choices. The clustering of lowly regarded elemental alternatives mirrors the type of process we would see in reality where the elemental members are not perceived individually but choice-makers are aware that a cluster of opportunities is available. Where problems occur is when the researcher tries to assess the set of conditional probabilities within a cluster of poorly understood elemental alternatives. Clearly, choice probabilities and parameter estimates will be very much affected if too many irrelevant alternatives are modelled. Hence the results of a full nested logit model, which operates at the elemental level, will be compromised.

It might, however, be argued that it is unwise to incorporate spatially detailed choice information into a model where the vast majority of choice-makers perceive the individual options only as a member of a cluster. When only one resident of British Columbia out of 10000 is observed to choose Cornerbrook, Newfoundland in a migration data set, the wisdom of explicitly modelling the choices by B.C. residents of all elemental alternatives within Newfoundland has to be questioned. In that context, only choice information at the level of Newfoundland should be included (although perhaps St. John's might be modelled as separate from smaller urban centres in the province). The aggregated model will estimate elemental utilities and the μ parameter should confirm a lack of independence among the set of elemental alternatives. Overall, it is felt that the spatial choice set definition literature could benefit from these types of conceptualizations. The issue may not be so much whether an elemental alternative is considered explicitly in a choice set as it a matter of how some insignificant spatial unit is clustered with other spatial units.

2.3.3 Spatial Choice Applications Accounting for Aggregation

The aggregated logit model as proposed in this thesis has never been applied. There are examples where the approach of Lerman (1975) has been adopted in an empirical context such as in Lerman (1979), Horowitz (1986) and Tay and McCarthy (1994) but in each of these cases, choices at the aggregate level are not modelled through the heterogeneity term. Instead the assumption is made that elemental systematic utilities within an aggregate are equal so that only a correction for size is necessary. This frees the researcher from having to explicitly define elemental systematic utilities. It is shown in Chapter 4 that this is a potentially risky practice in heterogeneous aggregates. In this section, recent studies in recreation choice, one of the few spatial choice fields where aggregation is being acknowledged, are assessed in some detail. In particular, several researchers (Parsons and Needleman, 1992; Parsons and Kealy, 1992 and Tay and McCarthy, 1994) have taken advantage of the fact that there exist rich data on the destination choice behaviour of American recreation enthusiasts. The elemental alternatives for such problems are typically the numerous inland lakes and other small bodies of water that are found within the much larger administrative zones that would constitute the aggregate alternatives.

Consider initially the study by Tay and McCarthy. They assess the behaviour of Indiana fishermen in terms of trips made to the 366 fishing zones administered in the United States. Since only 66 of these aggregates are actually chosen in their data, and since a model with 366 aggregates is considered too many for estimation purposes, a stratified importance sampling of alternatives procedure is undertaken to reduce the choice set for each person to five aggregates. The five alternatives are the result of sampling from three strata and in the end are composed of the origin zone, two alternatives from nearby aggregates, and two alternatives from far off aggregates. Of course McFadden (1978) had shown that such a procedure is justified for any set of alternatives governed by an Independence from Irrelevant Alternatives process so long as the appropriate correction to the utilities is applied.

Having reduced the choice set to five aggregates per person, Tay and McCarthy note that these spatial units themselves contain a massive number of elemental alternatives and hence allude to the need for corrections in the form of the size and heterogeneity terms. Largely for the reason that they do not have access to software which will implement the heterogeneity term, they make the restrictive assumption that the elemental systematic utilities of all the lakes in each aggregate are equal. This eliminates the need to collect attribute data, such as pollution levels, at the elemental level.

The size term for each aggregate is not implemented directly as the number of eligible bodies of water but rather as the total surface water area within each aggregate. Since their size term does not explicitly consider the number of choice units within aggregates, the question must be posed as to whether their estimate of the μ parameter is subject to bias. In general, the problem of Tay and McCarthy is similar in spirit to initiatives undertaken in this thesis. however their decision to abandon all attribute information relating to elemental alternatives removes a great deal of realism from their model.

The papers by Parsons and Needleman (1992) and Parsons and Kealy (1992) are essentially two of a series and represent empirical tests on some of the more innovative discrete choice concepts. Unlike Tay and McCarthy, these researchers have access to data which have a high degree of spatial resolution in the choice information. In particular, over 1000 residents of Wisconsin in 1978 were asked about their use of over 1000 Wisconsin lakes in the previous year. Clearly, the individual lakes represent the elemental alternatives in the analysis while aggregates are defined at two different levels: counties and larger Wisconsin sub-regions. With such rich data available, a series of model types is applied:

- a) The aggregate level OMNL with no corrections for size and heterogeneity effects; no use is made of elemental choice information

- b) The aggregate level OMNL with a correction for the size of aggregates; again no use is made of elemental choice information

- c) A full nested logit model (estimated sequentially) which makes use of elemental choice information and which calculates size and heterogeneity effects from the lower level estimations before obtaining μ estimates at the upper level

- d) An OMNL applied at the elemental level which employs a sampling technique to reduce the number of elemental alternatives to a manageable amount. This is a non-nested model in every sense since aggregate alternatives are not assumed.

In essence, the authors take the results from the latter two models in this list as being those providing the most accurate set of parameters with the other two models being judged in comparison. Of course, these comparisons are done using the two different definitions of aggregate alternatives (county and regional) with the authors' preconceived hypothesis clearly being that the greater the degree of spatial aggregation, the more unreliable the estimates of parameters. Indeed, the results indicate this hypothesis to be valid in that the parameters obtained from the two aggregate level OMNL models (a and b) bear little resemblance to the other two models when the aggregates are regions within Wisconsin. When the county definition of aggregates is employed, the performance of the aggregate level models improves.

There are grounds for criticism of this work. First of all, the aggregated model as it has been described in this thesis, does not appear anywhere in the analysis. Rather this paper employs extremely restrictive special cases of the aggregated model and then is highly critical of the results in relation to the models which have full information at the elemental level (c and d). These are unfair comparisons. A general application of the aggregated model requires the estimation of sub-aggregate utilities so that the heterogeneity of aggregates is properly accounted for. There is no justification given for the assumption in this paper that elemental systematic utilities are equal. Certainly they do estimate the heterogeneity effect, but only in the context of the sequential estimation of the full nested logit model where the elemental choice information is of course available. This shortcoming of the analysis is most likely due to the fact that their software estimates only the conventional OMNL. Such a limitation in and of itself should reduce confidence in the results.

The poor performance of the aggregate level models is also not surprising given their restrictive definition of the size term (model b). In the most general case, there will be a single μ parameter associated with the size of each aggregate. The arbitrary assumption that the μ are equal across aggregates is likely to have a profound impact on the other parameters, especially when aggregates are large. The size effect in such models, like alternative-specific constants, plays the important role of absorbing all systematic variation not captured by other variables in the model. The assumption is that this unobserved systematic variation can be explained by the size of the associated aggregate. The fact that these models do not at least test for significant variation in the μ parameter across aggregates is thus a serious oversight. A further potential problem with their representation of the size effect is the use of proxy variables such as surface area and depth of lakes. The use of proxies is bound to affect the interpretation of μ and may have much to do with cases in the analysis where μ strays outside the theoretical range.

The full nested logit and the elemental level OMNL (models c and d) applied in the paper are also not without shortcomings. First of all, direct comparisons of parameters are made between the former and the latter despite the fact that one model is explicitly assuming an IIA process as the justification for the sampling while the nested logit, by definition, assumes no such thing. Hence, there really is some doubt about which model type provides the 'true' set of parameters to act as the benchmark. The sequential estimation of the nested logit model is subject to the same set of problems outlined by Brownstone and Small (1989) with the situation being exacerbated by the fact that the lower level conditional parameters are obtained using samples of 12 lakes from the much larger number which would otherwise define each nest. Actual calculation of the inclusive values however includes all lakes in each nest. Again, the restriction that the μ parameter is equivalent across nests is arbitrarily imposed.

While the paper by Parsons and Needelman has its focus the issue of aggregation biases, it is worth noting that the paper of Parsons and Kealy utilizes the same data set but has as its focus the sampling of alternatives as advocated by McFadden (1978). The results from this paper form the basis for the estimation of the fourth model type in the aggregation paper: the elemental level OMNL model employing sampling of spatial units. Overall, the interest of the authors is in the best way to model the choices among spatial units given that the total number of alternatives is too large to facilitate within a model. The clear theme running through the papers is that a sampling approach is to be much preferred over an approach which would reduce the computational burden through aggregation. While there are obvious flaws in the implementation of their comparison, this conclusion is reasonable since it is undesirable to literally throw away valuable elemental choice information. In cases where the elemental choice information is unavailable however, as is assumed in this thesis, these papers do little to discredit the approach advocated in the aggregated model.

2.4 Conclusions

The essential goal of this paper has been to advocate the use of the logit family of models for spatial choice problems. Logits have a clear advantage over gravity model formulations since they are derived at the level of the individual choice-maker. While logits can always be easily adapted to a gravity model type of problem that contains both aggregation across choice-makers and across space, the gravity model can be derived from the micro-level only under unrealistic assumptions.

It has been pointed out that the ordinary multinomial logit model is not as sensitive to violations of the Independence from Irrelevant Alternatives property as is commonly thought by many spatial choice researchers. While the model does not explicitly attempt to model the patterns of similarities among alternatives, an OMNL with constants typically sacrifices very little in terms of fit and will yield parameter estimates for explanatory variables which will often be similar to those of more sophisticated models. Statements that the OMNL will necessarily lead to substantially biased estimates of probabilities in a non-IIA choice process are simply not true.

On the subject of the IIA, it has been mentioned that it is particularly important in spatial choice models to allow for its variation across origins. Residents of one origin are likely not to cluster spatial alternatives in the same way as those at another. The use of origin-specific models to capture differences in how alternatives are grouped represents an interesting avenue for future research.

In the context of models that explicitly capture unequal substitutability among spatial units, it is suggested that the nested logit and the aggregated models are by far the best approaches for spatial choice problems. The framework inherent to both is the only one which easily accommodates the challenges posed by aggregation of spatial units and by the patterns of similarity across spatial units. The closest challenger, the probit model, remains computationally intractable for all but the smallest spatial choice problems and the underlying theory to accommodate spatial aggregation has never been implemented. It is felt that the nested logit model has been hampered by the sequential estimation procedure with which it has always been associated but that more widespread application of simultaneous estimation will result in cleaner specifications and less reluctance to experiment with different decision trees for a given problem. Also, the statistical reliability of the important substitutability parameters will be greatly improved.

Finally, it has been discussed that the issue of spatial aggregation has not received adequate treatment in the spatial choice literature, with there being a tendency for it to be assumed away. Other researchers warn of the dangers of aggregation and attempt to discredit the effectiveness of the framework for aggregation in choice models. However, it has been pointed out that their simulation experiments are not conceived sufficiently well to be making such claims and were not conducted with full implementations of the aggregated logit model.

Overall, there is a tendency in the spatial choice literature to advocate alternative models, some of them quite arbitrary, before the conventional models have been completely understood. The logit family of models are based on a rich body of theory which provides an excellent medium through which spatial choice processes can be understood. Certainly, there needs to be more widespread recognition that spatial aggregation and the substitutability of spatial units work hand in hand. Both of these issues, which are central to spatial choice, can be modelled simultaneously in the logit framework.

Chapter 3

Discrete Spatial Choice Models for Aggregate Destinations

3.1 Introduction

Choice models are useful in theoretical and empirical analysis in assessing the behaviour of a group of individuals (hereafter, decision or choice-makers) faced with a set of alternatives. Such models are derived from first principles of consumer choice behaviour based on utility maximization. When the choice set of alternatives consists of spatial units, the associated models are referred to as spatial choice models. As discussed in Tardiff (1979) elements of a spatial choice set are not readily or uniquely identifiable as in other choice problems. In most instances the choice set is dictated by the available data with the spatial units being, more often than not, larger than what decision-makers perceive. A related problem is the large number of spatial alternatives that in practice necessitate either aggregation into larger spatial units or the sampling of alternatives. The problem of aggregate spatial units is pervasive in studies of spatial interaction such as intraurban or interregional migration and the analysis of shopping trips. Yet, choice models are routinely applied to problems of spatial choice with little or no consideration of spatial aggregation effects.

Because choice models are firmly rooted in behavioural theory, accounting for aggregation must result in models consistent with theory. Such an adaptation of the theory requires that one go beyond the strict problem of spatial aggregation and consider the wider problem of human choice behaviour in space. That this has not been done is perhaps the main reason why essential progress in capturing spatial effects in choice models has been modest. Researchers have tended to see the problem as stemming from the substitutability of spatial alternatives, a matter poorly handled by the family of choice models sharing the independence from irrelevant alternatives (IIA) property (Lo, 1991). Solutions to this problem in the form of ad hoc changes to the logit formula have been proposed by Borgers and Timmermans (1987) and Haynes and Fotheringham (1990).

Closer to the approach taken in capturing spatial effects in the standard linear model, Boots and Kanaroglou (1988) proposed a correction for spatial structure be included directly in the specification of the systematic utilities. Although this approach can be effective, it is not satisfactory because it provides no link between individual behaviour and the corrective term introduced. Thus spatial structure is viewed as a nuisance in need of correction rather than an integral part of the choice process. This philosophy is also shared by the model of Bolduc et. al. (1989). An additional difficulty with the latter model is that the functional form utilized (logit) implies i.i.d. extreme value error structure for the utilities, while the errors are subsequently assumed to be spatially correlated and normally distributed. This inconsistency in assumptions is alleviated in the comprehensive probit model proposed by Ben Akiva and Bolduc (1991) and Bolduc (1994) at the expense of computational tractability in parameter estimation, given the typically large number of alternatives in realistic spatial choice problems. Similar in spirit are the probit models used by Case (1992) and McMillen (1992) that try to capture heteroscedasticity with a block covariance matrix structure.

The models discussed in this paper have theoretical appeal, like the probit, but are much better suited to large problems with many spatial alternatives. They are capable of accommodating a block covariance matrix structure while permitting estimation with conventional means as opposed to the elaborate simulation approach advocated by Bolduc (1994) for the probit. The formative ideas for these models are in the work of Lerman (1975), McFadden (1978) and Kitamura et. al. (1979). A useful summary of findings in this research is provided in Ben-Akiva and Lerman (1985, ch. 9). However, the theory underlying the models has never been fully developed and neither have such models been applied in their complete form.

The purpose of our paper is to synthesize and clearly outline the theoretical underpinnings of these models. Furthermore, we go beyond the existing theory by relaxing the conventional logit assumption of Gumbel distributed error terms. This is accomplished by making use of the asymptotic theory of extremes. Following this introduction we derive the aggregated spatial logit model. Finally, the asymptotic theory of extremes and its relationship to the aggregated model is discussed.

3.2 The Aggregated Spatial Logit Model

The model in this section is discussed in the context of migration (intraurban or interregional), although it applies to any scenario where decision-makers, originating from some point in space, decide to locate, shop, or pursue leisure at some other point in space. We assume the use of data that are disaggregate with respect to choice-makers but aggregated across the spatial choice units. Studies which utilize this type of data include Liaw and Ledent (1988) and Liaw (1990). Others have used micro-level data but the elements of the choice set were not spatial units (e.g. Meyer and Speare, 1985). In such models, the choice set typically consists of two alternatives: stay or move. There are an abundance of studies which utilize data aggregated across both individuals and spatial units such as Shaw (1985) and Anderson and Papageorgiou (1994).

3.2.1 Theoretical Framework

We postulate that decision-makers (individuals, family, households) perceive space in the form of zones rather than points. Furthermore, we assume that all decision-makers identify the same zonal system on average[5]. We call the zones in this zonal system *elemental*[6]. In intraurban migration, an elemental choice unit could be a city neighborhood while in other forms of migration it might be a single spatial labour market as discussed by Shaw (1985, p.202) and Kerr (1950). Any given decision-maker associates a level of utility with each elemental zone. Thus, each decision-maker perceives a unique utility surface over the study area which is discontinuous at the zone boundaries.

Consider a study area which is divided into M elemental zones. Following Kanaroglou et al. (1986), we assume a two level model. In the first level, the departure sub-model, the probability of a potential migrant leaving a zone is expressed as a function of the difference between the utility of the decision-maker's origin and the expected maximum utility available from prospective destinations. In the second level, termed a destination choice model, the conditional probability of a migrant selecting one of the $M-1$ potential destinations is expressed as a function of the expected utilities of those destinations. The product of the two probabilities is the probability that a given decision-maker will move from their origin to a destination. This paper focuses on the sub-model of destination choice which is affected by zonal aggregation.

Consider that the M elemental spatial units are combined into L aggregate units, each of which contains $M^J, J \in \{1, 2, ..., L\}$ elemental units. Thus:

$$\sum_{J=1}^{L} M^J = M \tag{1}$$

From the behavioural point of view, one may postulate that individual decision makers process spatial information hierarchically (Fotheringham, 1986). The potential for a hierarchical selection process can be captured through the nested logit model. The model presented in this section has its roots in the theory of the nested logit. Detailed presentation of this theory is to be found in Train (1986, chapter 4) and Ben

[5] Many authors have pointed out that in certain spatial choice contexts this assumption may lead to model misspecification. For a summary of detailed arguments see Thill (1992).

[6] This term was first utilized by Lerman (1975) in the context of intraurban housing choice.

Akiva and Lerman (1985, chapter 10).

Let U_j^J represent the utility that decision-maker n perceives at elemental destination j, which belongs to aggregate destination J [7]. It can be written:

$$U_j^J = V_j^J + \epsilon^J + \epsilon_j^J \tag{2}$$

where:

- V_j^J is determined by attributes which relate to the aggregate alternative J and/or the elemental alternative $j \in \{1, 2, ..., M^J\}$ and characteristics of individual n.
- ϵ^J captures unobserved variation that relates to the aggregate spatial unit J.
- ϵ_j^J captures unobserved variation that relates to the elemental spatial unit $j \in \{1, 2, ..., M^J\}$.

A variety of variables are used to define V_j^J. These include attributes which relate only to aggregate alternatives such as federal transfer payments to states or provinces, or they may be attributes, such as housing prices, which differ among elemental alternatives. The former group of attributes uniformly affects the choices of all elemental alternatives in an aggregate. Other variables included are characteristics of the decision-maker such as age or income. Variables which measure the physical or social distance between i and j are also utilized in V_j^J.

From the perspective of decision-maker n at origin i, U_j^J is a known quantity which is used to evaluate the attractiveness of destinations. From the perspective of the researcher however, U_j^J is a random variable. Thus, each elemental alternative is associated with a random utility variable. Across space, there is a level of dependency among the unknown utilities U_j^J which is influenced by relative proximity. This dependency is consistent with conceptualizations in spatial statistics. The systematic utility V_j^J can be considered a potential realization of the true utility U_j^J. The assumptions underlying the model discussed

[7] The more proper expression $U_{j|in}^J$ is not used in order to maintain notational simplicity. The same applies for the systematic utility V and the elemental error.

here are much more restrictive than this conceptualization.

Assumption 1: The random error components $\epsilon^J$and ϵ_j^J are independent for all $J \in \{1, 2, ..., L\}$ and $j \in \{1, 2, ..., M\}$.

Assumption 1 implies three things: firstly, unobserved random components relating strictly to aggregate alternatives are independent of each other; secondly, unobserved random components relating strictly to elemental alternatives are independent of each other; thirdly, for any aggregate J, its unobserved variation is independent of the unobserved variation of any elemental alternative, either within or outside J.

Assumption 2: The random error components ϵ_j^J, $j \in \{1, 2, ..., M^J\}$ are identically Gumbel distributed with parameters $(0, \mu^J)$,where $\mu^J > 0$, for every $J \in \{1, 2, ..., L\}$.

The utilities of any two elemental alternatives are identically distributed only if both belong to the same aggregate. The mode of all distributions is zero and μ^J determines the variance as discussed in detail below. The assumption of a mode equal to 0, is not restrictive since a non-zero locational parameter is absorbed into the systematic component V_j^J.

Assumption 3: For any Gumbel variate ϵ with parameters $(0, \mu^J)$, $\epsilon^J$is distributed in such a way that $\epsilon^J + \epsilon$ is also a Gumbel variate with parameters $(0, \mu)$ for all $J \in \{1, 2, ..., L\}$.
The purpose of assumption 3 is to ensure that the random components of aggregate alternatives are identically distributed Gumbel variates.

It is very important to understand at this stage that assumptions 1 to 3 **do not** imply that the total utilities U_j^J, $j \in \{1, 2, ..., M^J\}$, associated with elemental alternatives of the same aggregate, are independent. In fact, these assumptions imply a positive covariance of total utilities in the same aggregate since each total elemental utility shares a random component defined by unobserved variation operating at the aggregate level. Actually, given equation 2 and assumption 1, for any $J \in \{1, 2, ..., L\}$ and

$j, k \in \{1, 2, ..., M^J\}$:

$$cov(U_j^J, U_k^J) = cov(\epsilon_j^J + \epsilon^J, \epsilon_k^J + \epsilon^J) = Var(\epsilon^J) \quad (3)$$

While the utilities of any two elemental alternatives that belong to the same aggregate have the same covariance, elemental utilities that belong to different aggregates are uncorrelated. To put it differently, the variance-covariance matrix has blocks of non-zero covariances along the diagonal. The correlation coefficient of any two elemental utilities within an aggregate is related to the ratio $\frac{\mu}{\mu^J}$. Since ϵ_j^J is Gumbel distributed with scale parameter $\mu^J > 0$ (assumption 2), then $Var(\epsilon_j^J) = \pi^2/6(\mu^J)^2$. Similarly, by assumption 3, $\pi^2/6\mu^2 = Var(\epsilon^J + \epsilon_j^J) = Var(\epsilon^J) + Var(\epsilon_j^J)$. The last equality is true because of the independence of the error terms (assumption 1).

Therefore:

$$\begin{aligned}\frac{\mu}{\mu^J} &= \sqrt{\frac{Var(\epsilon_j^J)}{Var(\epsilon^J) + Var(\epsilon_j^J)}} = \sqrt{1 - \frac{Var(\epsilon^J)}{Var(\epsilon_j^J + \epsilon^J)}} \\ &= \sqrt{1 - \frac{cov(U_j^J, U_k^J)}{\sqrt{Var(U_j^J)}\sqrt{Var(U_k^J)}}} = \sqrt{1 - corr(U_j^J, U_k^J)}\end{aligned} \quad (4)$$

It is important to elaborate on the range of potential values for the ratio $\frac{\mu}{\mu^J}$. The above discussion suggests that the variance of the total utility for any elemental alternative has two components:

(1) the variability of the random error component ($Var(\epsilon^J)$) relating strictly to the aggregate alternative in which the elemental alternative is located
(2) the variability of the random error component relating strictly to the elemental alternative ($Var(\epsilon_j^J)$).

The first equality in 4 suggests that $\frac{\mu}{\mu^J}$ depends on the relative magnitudes of uncertainty at the elemental and aggregate levels and since the denominator is larger than the numerator $0 \leq \frac{\mu}{\mu^J} \leq 1$[8].

It is of interest to discuss the two extreme cases. The largest possible value of 1 is approached when $Var(\epsilon^J)$ is insignificant relative to $Var(\epsilon_j^J)$. In this circumstance, most unobserved variation relates to the elemental alternative, not the aggregate. Conversely, the smallest possible ratio of 0, is approached when $Var(\epsilon_j^J)$ is insignificant relative to $Var(\epsilon^J)$. Unobserved variation relating to the elemental alternative is strongly outweighed by aggregate variation. In this case, elemental utilities within an aggregate are highly (and positively) correlated, sharing the unobserved variation of the aggregate alternative. In practical applications, if assumptions 1 to 3 are valid, it is expected that $\frac{\mu}{\mu^J}$ will be between 0 and 1. Note that this framework precludes negative correlations between elemental utilities since for any given aggregate J the random components ϵ_j^J, $j \in \{1, 2, ..., M^J\}$ which might well contain such effects, are considered independent by Assumption 1.

3.2.2 The Model

We now introduce the behavioural mechanism by which individual decision-makers select a spatial unit as a destination:

Assumption 4: Decision-makers aim to maximize their utility.

The model derived below assumes that decision-makers choose the single elemental alternative in the system which is associated with the highest level of utility. Even if the available data contain aggregate choice units, it does not change the fact that choice-makers may well select on the basis of elemental spatial units. The strength of the resulting model is its ability to correctly represent probabilities at the aggregate level for choice processes which potentially occur at the elemental level. The proof of the proposition that follows is an adaptation of the nested logit model derivation. It is reproduced here because it makes clear the necessity of the four assumptions introduced in this section.

---

[8]In the strict sense, $\frac{\mu^J}{\mu} > 0$ because $\mu^J, \mu > 0$. This ratio however, can approach zero if $\mu^J >> \mu$.

Proposition 1: For any $J \in \{1, 2, ..., L\}$:

$$P(J) = \frac{\exp\left[\frac{\mu}{\mu^J} \ln \sum_{j=1}^{M^J} e^{\mu^J V_j^J}\right]}{\sum_{Q=1}^{L} \exp\left[\frac{\mu}{\mu^Q} \ln \sum_{q=1}^{M^Q} e^{\mu^Q V_q^Q}\right]} \tag{5}$$

Proof:

The point of highest utility in aggregate alternative J is given by $\max_{j\in\{1,2,...,M^J\}} U_j^J$. There are L such 'utility peaks' in the system, one for each aggregate alternative. By assumption 4, a choice-maker will select the elemental alternative corresponding to the highest utility peak, thereby selecting the aggregate alternative that contains it. Thus, the choice probability for aggregate J is equivalent to the probability that J contains the highest utility peak:

$$P(J) = \Pr(\max_{j\in\{1,2,...,M^J\}} U_j^J \geq \max_{j\in\{1,2,...,M^{J'}\}} U_j^{J'}, J' \in \{1, 2, ..., L\}, J' \neq J) \tag{6}$$

Substituting 2 in 6:

$$P(J) = \Pr(\epsilon^J + \max_{j\in\{1,2,...,M^J\}} (V_j^J + \epsilon_j^J) \geq$$

$$\epsilon^{J'} + \max_{j\in\{1,2,...,M^{J'}\}} (V_j^{J'} + \epsilon_j^{J'}), J' \in \{1, 2, ..., L\}, J' \neq J) \tag{7}$$

Because of assumptions 2, $V_j^J + \epsilon_j^J$ is Gumbel distributed with parameters (V_j^J, μ^J)[9]. Also, $\max_{j\in\{1,2,...,M^J\}} (V_j^J + \epsilon_j^J)$ is a random variable which is Gumbel distributed with parameters $(\frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} e^{\mu^J V_j^J}, \mu^J)$[10]. Finally,

[9]Property of Gumbel: if ϵ is Gumbel with parameters (η, μ), and b is a scalar constant, then $\epsilon + b$ is Gumbel with parameters $(\eta + b, \mu)$.

[10]Property of Gumbel: if $\epsilon_1, \epsilon_2, ..., \epsilon_k$ are Gumbel with parameters (η_1, μ), (η_2, μ), ..., (η_k, μ) respectively, then $\max(\epsilon_1, \epsilon_2, ..., \epsilon_k)$ is Gumbel with parameters $(\frac{1}{\mu} \ln \sum_{j=1}^{k} e^{\mu \eta_j}, \mu)$.

due to assumption 3, $\epsilon^J + \max_{j\in\{1,2,...,M^J\}}(V_j^J + \epsilon_j^J)$ is Gumbel with parameters $(\frac{1}{\mu^J}\ln\sum_{j=1}^{M^J} e^{\mu^J V_j^J}, \mu)$. Thus, 7 can be written as:

$$P(J) = \Pr(V^J + \epsilon^{oJ} \geq V^{J'} + \epsilon^{oJ'}, J' \in \{1,2,...,L\}, J' \neq J)$$

where $V^J = \frac{1}{\mu^J}\ln\sum_{j=1}^{M^J} e^{\mu^J V_j^J}$ and ϵ^{oJ} is Gumbel with parameters $(0,\mu)$. By the theory of the multinomial logit model then:

$$P(J) = \frac{\exp\left[\mu V^J\right]}{\sum_{Q=1}^{L}\exp\left[\mu V^Q\right]} = \frac{\exp\left[\frac{\mu}{\mu^J}\ln\sum_{j=1}^{M^J} e^{\mu^J V_i^J}\right]}{\sum_{Q=1}^{L}\exp\left[\frac{\mu}{\mu^Q}\ln\sum_{q=1}^{M^Q} e^{\mu^Q V_q^Q}\right]} \square$$

Note that equation 5 is equivalent to the marginal probability model of a two level nested logit. Were we to employ the full nested logit form, we would complement 5 with the conditional probability model which provides a choice probability for an elemental alternative conditional on the aggregate to which it belongs. Although the conditional probability model is not utilized, it is important to note that 5 provides the choice probability of an aggregate as a function of attributes relating *both* to aggregate and elemental alternatives.

In applications of spatial choice problems, researchers have not conceptualized the utility of an aggregate as the utility associated with the point of maximum utility within. As a result, there is the tendency to assume that decision-makers consider the *mean* utility of an aggregate as opposed to its *maximum* utility. This tendency is equivalent to assuming that aggregate spatial units, such as the Canadian provinces, are elemental zones, a grossly inaccurate assumption. Under the assumption that elemental alternatives are relatively homogeneous with respect to size, let: $\overline{V}^J = \frac{1}{M^J}\sum_{j=1}^{M^J} V_j^J$, $J \in \{1,2,...,L\}$, be the systematic utility of an aggregate alternative when decision-makers are assumed to consider only the mean utility of an aggregate. If elemental utilities are not homogeneous then a weighted average is usually considered.

In model 5 the systematic utility of an aggregate alternative J is provided by $V^J = \frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} e^{\mu^J V_j^J}$. The following proposition manipulates the expression for V^J in order to establish its relationship with $\overline{V}^J$. A general description of this derivation is provided in Ben Akiva and Lerman (1985, chapter 9).

Proposition 2:

$$V^J = \overline{V}^J + \frac{1}{\mu^J} \ln[\frac{1}{M^J} \sum_{j=1}^{M^J} e^{\mu^J (V_j^J - \overline{V}^J)}] + \frac{1}{\mu^J} \ln M^J \tag{8}$$

Proof: It is sufficient to show that the R.H.S. of 8 equals the L.H.S. :

$$\begin{aligned} RHS &= \overline{V}^J + \frac{1}{\mu^J} \ln[\frac{1}{e^{\mu^J \overline{V}^J}} \frac{1}{M^J} \textstyle\sum_{j=1}^{M^J} e^{\mu^J V_j^J}] + \frac{1}{\mu^J} \ln M^J \\ &= \overline{V}^J + \frac{1}{\mu^J} \ln \frac{1}{e^{\mu^J \overline{V}^J}} + \frac{1}{\mu^J} \ln \frac{1}{M^J} + \frac{1}{\mu^J} \ln \textstyle\sum_{j=1}^{M^J} e^{\mu^J V_j^J} + \frac{1}{\mu^J} \ln M^J \\ &= \frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} e^{\mu^J V_j^J} = LHS \square \end{aligned}$$

Intuitively, because V^J is the expected value[11] of the maximum utility while $\overline{V}^J$ is the mean utility, one would expect $V^J \geq \overline{V}^J$. Equation 8 confirms this to be true. The term $\frac{1}{\mu} \ln M^J$, which can be interpreted as correcting for the size of an aggregate alternative, is necessarily non-negative since $M^J \geq 1$. It can also be shown that $\ln[\frac{1}{M^J} \sum_{j=1}^{M^J} e^{\mu (V_j^J - \overline{V}^J)}] \geq 0$ (Ben-Akiva and Lerman, 1985, p. 257). Hence the estimate of maximum utility must exceed the estimate of mean utility. The relationship $V^J = \overline{V}^J$ is possible if $M^J = 1$. In this case, J is an elemental alternative itself and therefore there is no aggregation effect.

The term $\frac{1}{\mu^J} \ln[\frac{1}{M^J} \sum_{j=1}^{M^J} e^{\mu^J (V_j^J - \overline{V}^J)}]$ can be interpreted as measuring the variability of utility within aggregate alternative J. Since this heterogeneity term is non-negative, it increases the systematic utility of aggregate alternatives which have a greater variation in opportunity. The heterogeneity term is

[11]Strictly speaking the expected value is $V^J + \frac{\gamma}{\mu}$, where γ is Euler's constant, but $\frac{\gamma}{\mu}$ is dropped with no effect on differences in utility and hence the choice probabilities.

appealing if we consider several aggregate alternatives and their distributions of utility. Other things being equal, we expect that the alternative whose surface of systematic utility has the largest assortment of peaks and valleys is the alternative most likely to contain the elemental alternative of maximum utility. From the behavioural point of view, choice-makers seek the elemental zone of maximum utility. The heterogeneity term contributes toward the identification of the aggregate zone associated with this utility. A similar interpretation is true for the size term. The larger an aggregate alternative the more likely that it will contain the elemental alternative of highest utility. Thus, size tends to increase the systematic utility of an aggregate alternative thereby increasing the choice probability associated with it.

3.2.3 The Uncertainty Parameters

Assumptions 1 to 3 imply that $Var(\epsilon_j^J) = \pi^2/6(\mu^J)^2$ and $Var(\epsilon^J + \epsilon_j^J) = \pi^2/6\mu^2$. Because $Var(\epsilon^J + \epsilon_j^J) > Var(\epsilon_j^J)$ for all J, we have $0 < \mu \leq \min_{J \in \{1,2,...,L\}} \mu^J$. The case when $\mu \rightarrow 0$ as μ^J remains fixed at some positive value is of theoretical interest. It is equivalent to $Var(\epsilon^J) \rightarrow \infty$ which makes the utilities of elemental alternatives that belong in J highly correlated. The systematic utilities contribute no information in predicting the probabilities of choosing an aggregate. Hence, in the absence of any information all aggregate alternatives are assigned equal probabilities as can be seen from model 5, which reduces to $P(J) = \frac{1}{L}, J \in \{1, 2, ..., L\}$.

Parameter μ, as in the ordinary multinomial logit model, is not identifiable during estimation. With a linear-in-parameters systematic utility form, the parameters estimated are all multiplied by the implicit but unidentified value of μ. For operationalization purposes, analysts restrict μ arbitrarily to a convenient value, usually 1. Such an action sets the scale of measurement for the systematic utilities and allows identification of the relevant ratios $\frac{\mu}{\mu^J}$, $J \in \{1, 2, ..., L\}$. The total variance of model 5 is fixed to $Var(\epsilon^J) + Var(\epsilon_j^J) = \pi^2/6$ and $1 = \mu < \mu^J < \infty$ for all $J \in \{1, 2, ..., L\}$.

Before engaging in further discussion it is useful to examine the relationship between elemental and aggregate systematic utilities. It is possible to re-arrange the systematic utility of an aggregate alternative J as:

$$V^J = V_k^J + \frac{1}{\mu^J} \ln(1 + \sum_{j \neq k}^{M^J} e^{\mu^J (V_j^J - V_k^J)}) \tag{9}$$

where $V_k^J = \max_{j \in \{1,2,\ldots,M^J\}} V_j^J$. A compatible form for the heterogeneity term H is:

$$H = V_k^J - \overline{V}^J + \frac{1}{\mu^J} \ln[\frac{1}{M^J}(1 + \sum_{j \neq k}^{M^J} e^{\mu^J (V_j^J - V_k^J)})] \tag{10}$$

One can demonstrate that the second term in the right hand side of equation 9 is non-negative. This implies that $V^J \geq V_k^J$. The estimated systematic utility of an aggregate, therefore, is at least equal to the highest elemental systematic utility within it. On the other hand, the third term in the right hand side of 10 is non-positive, implying $H \leq V_k^J - \overline{V}^J$. The heterogeneity term of an aggregate, therefore, is at most equal to the highest individual discrepancy in elemental systematic utility from the mean systematic utility. The particular outcomes are dependent on μ^J.

If $\mu = 1$, parameters μ^J, $J \in \{1, 2..., L\}$ are directly identifiable during estimation, and their values provide the researcher with important information regarding the employed spatial aggregation scheme. Therefore, it is of theoretical and empirical interest to examine the model for the two extreme values of μ^J :

(1) $\mu^J \to \infty$

The variance at the elemental level approaches zero asymptotically: $\lim_{\mu^J \to \infty} Var(\epsilon_j^J) = 0$. At the same time $\lim_{\mu^J \to \infty} Var(\epsilon^J) = \pi^2/6$. The elemental alternatives in J will be highly correlated since for any two such alternatives j and j' $\lim_{\mu^J \to \infty} corr(U_j^J . U_{j'}^J) = \lim_{\mu^J \to \infty} (1 - \frac{1}{(\mu^J)^2}) = 1$. At the limit then, the utility of an elemental alternative j in aggregate J will be $U_j^J = V_j^J + \epsilon^J$ since ϵ_j^J is now a constant, absorbed into V_j^J. The elemental alternative k of the highest systematic utility in J can be identified with complete certainty to be associated with maximum utility: $\max_{j \in \{1,2,\ldots,M^J\}} U_j^J = \max_{j \in \{1,2,\ldots,M^J\}} V_j^J + \epsilon^J = V_k^J + \epsilon^J$.

The systematic utility of an aggregate J attains its lowest value V_k^J. At the same time, the size term in this case approaches its lowest value which is zero, while the heterogeneity term approaches its highest value $H = V_k^J - \overline{V}^J$. Because of the absence of randomness at the elemental level and for a given aggregate, the elemental alternative associated with the maximum systematic utility can always be identified no matter how many elemental alternatives the aggregate contains. This is why the size term becomes irrelevant and disappears. On the other hand, as μ^J gets smaller and uncertainty ϵ_j^J at the elemental level increases, the maximum utility within J, identified by $\max_{j\in\{1,2,\ldots,M^J\}}\{V_j^J + \epsilon_j^J\}$, is likely larger than V_k^J, which is now only the central tendency of the random variable $V_k^J + \epsilon_k^J$.

(2) $\mu^J = 1$.

In this case $Var(\epsilon_j^J) = \pi^2/6$ and thus $Var(\epsilon^J) = 0$. Also, the elemental alternatives within J are perceived by decision maker n as uncorrelated since, for any pair j and j' of such alternatives, equation 4 now becomes $corr(U_j^J, U_{j'}^J) = 1 - (\frac{\mu}{\mu^J})^2 = 0$. Consider the event of decision-maker n selecting one elemental alternative out of the M^J elemental alternatives in aggregate J. We have M^J such events, one per elemental alternative in J, that are pair wise independent. This adaptation of equation 5 reflects the situation:

$$P(J) = \frac{\sum_{j=1}^{M^J} \exp\left[V_j^J\right]}{\sum_{j=1}^{M^J} \exp\left[V_j^J\right] + \sum_{Q\neq J}^{L} \exp\left[\frac{1}{\mu^Q} \ln \sum_{q=1}^{M^Q} e^{\mu^Q V_q^Q}\right]}$$

This model form implies that, if the data allow, elemental units in J can be grouped into aggregates in any desirable way. A further implication is that for varying aggregations of the M^J elemental alternatives, the same set of estimated parameters could be used.

Further insight can be gained by examining expression 8 for the utility of an aggregate, which in this case attains its highest value. Since μ^J attains its minimum theoretical value, the size term becomes its maximum and equal to $\ln M^J$. On the other hand the heterogeneity term attains its minimum value, which according to equation 10 is the maximum discrepancy in systematic elemental utility augmented by the non-positive value $\ln[\frac{1}{M^J}(1+\sum_{j\neq k}^{M^J} e^{(V_j^J - V_k^J)})]$. Intuitively, the size term becomes so important in this case because of the large amount of randomness unique to individual elemental alternatives. The identification of the maximum utility in this case becomes an uncertain process since many elemental alternatives in J could provide the maximum. Also, heterogeneity in systematic utility becomes rather unimportant since the systematic utility is not a good representation of total utility.

Having discussed the theoretically extreme values of μ^J for a given aggregate J, an empirically valid question is: what is a large value and how could it be attained in practice? From equation 4 $Var(\epsilon^J)/\ Var(\epsilon_j^J) = (\mu^J)^2 - 1$. Thus, the ratio of aggregate to elemental variance changes with the square of μ^J. A value of $\mu^J = 10$ implies $Var(\epsilon^J) = 99\ Var(\epsilon_j^J)$. Thus, aggregate variance becomes very large relative to elemental variance, which in this case is the almost negligible value $\pi^2/600$. The variances of ϵ^J and ϵ_j^J are controlled by the independent variables in V_j^J for aggregate J and by the imposed grouping of elemental alternatives. As is the case with the nested logit model, it would be desirable to have $\mu^J = 10$ since it implies that we are accounting for sub-aggregate systematic variation very well and that the grouping of elemental alternatives is relevant. The results for μ^J reflect the possible need for an improved specification or perhaps are recognition of the fact that the imposed set of aggregate choice units are distorting representation of the elemental choice process.

Another case of interest is when the maximum systematic utility V_k^J is so much larger than all other elemental systematic utilities within aggregate J that no matter how large $Var(\epsilon_j^J)$, $\max_{j\in\{1,2,\ldots,M^J\}} U_j^J$ is always provided by elemental option k. In such a case, the estimated value of μ^J will tend to be larger than the true value and hence upward biased.

3.3 Asymptotic Spatial Logit Models

Assumptions 1 to 4 constitute the foundation of the aggregated logit model. In this section we demonstrate that when the number of elemental alternatives within aggregates becomes large, then the same model can be derived after replacing assumption 2 (Gumbel distributed elemental errors) with a milder, behaviourally sound, assumption. This idea was first discussed in Leonardi and Papageorgiou (1992), in the context of the multinomial logit model. We first explore the nature of the new assumption and then we derive the aggregated logit model under the new set of assumptions.

3.3.1 The Distribution of Elemental Errors

Gumbel distributed errors in utility is a standard assumption in the derivation of the logit model. Besides its mathematical convenience, Gumbel is justified as an approximation to the normal. No behavioural justification is offered, however, as to why the utility error terms should be normally or Gumbel distributed. The asymptotic theory of extremes provides the possibility of replacing assumption 2 with another which is less restrictive. We shall think of errors as arbitrarily large. Beyond a certain boundary C all errors obey the following:

Assumption 2.1: For all ϵ_j^J, $j \in \{1, 2, ...M^J\}$:

$$\Pr(\epsilon_j^J \geq x + c \mid \epsilon_j^J \geq c) = \Pr(\epsilon_j^J \geq x) \quad \text{for all } x > C \text{ and } c \geq 0 \tag{11}$$

Because C can be arbitrarily large, condition 11 can be thought of as referring to the right tail of the distribution of errors[12]. Thus, condition 11 implies that if an error term is shifted by adding to it an arbitrary non-negative number, the distribution of its right tail remains unaffected. Equation 2 implies that a constant added to ϵ_j^J can be absorbed into the systematic component of utility V_j^J. In light of this, assumption 2.1 requires that adding a constant to the systematic component of utility does not affect the shape of the right tail of the corresponding error. To put it differently, no matter at what point on the line

[12] The idea of the memoriless property holding for the right tail of the error distribution was suggested by an anonymous referee.

of real numbers the analyst starts measuring systematic utility, it is required that the distribution of right tail of errors remains unchanged.

Equation 11 is known as the lack of memory property. For a random variable $X \geq 0$ that represents the life length of an object or being, we say that X does not age (or lacks memory) if, given that the object lived c time units, the probability of its lasting for another x time units is the same for any c. Equation 11 is equivalent to:

$$\frac{\Pr(\epsilon_j^J \geq x + c)}{\Pr(\epsilon_j^J \geq c)} = \Pr(\epsilon_j^J \geq x)$$

or

$$\Pr(\epsilon_j^J \geq x + c) = \Pr(\epsilon_j^J \geq x) \Pr(\epsilon_j^J \geq c)$$

or

$$1 - F(x + c) = [1 - F(x)][1 - F(c)] \tag{12}$$

The exponential distribution, which is a special case of the Γ distribution, has the lack of memory property. For λ> 0 it is defined as:

$$F(x) = \Pr(X < x) = \ 1 - e^{-\lambda x} \quad for \quad x > C \tag{13}$$

It is straightforward to verify that 13 satisfies 12. It is also possible to prove the inverse, that is, any distribution that satisfies 12 is an exponential distribution. Here we simply state the proposition, a proof of which is provided in Galambos (1978, pp.31-32).

Proposition 3: If the random variable $X \geq 0$ has a continuous, nondegenerate distribution function $F(x)$ that satisfies equation 12 for any $x \geq 0$ and $c \geq 0$, then F is exponential as in 13 with $\lambda > 0$.

The above discussion suggests that assumption 2.1 is a necessary and sufficient condition for the right tail of ϵ_j^J, $j \in \{1, 2, ...M^J\}$ to be exponentially distributed as in 13 for any aggregate J. Random variables $(V_j^J + \epsilon_j^J)$, $j \in \{1, 2, ...M^J\}$ share the same property and their distribution is as follows:

$$F_j(x) = \Pr(V_j^J + \epsilon_j^J < x) = \Pr(\epsilon_j^J < x - V_j^J) = \ 1 - e^{-\lambda^J(x - V_j^J)}; \quad x > V_j^J + C \tag{14}$$

Notice that for every elemental alternative $j \in \{1, 2, ...M^J\}$ within aggregate J there is a different distribution F_j. All distributions have arbitrary shapes, except that they are identical as far as the shape of their right tail is concerned. Furthermore, this shape is negative exponential. The fact that discrete choice theory emphasizes the distribution of maximum utility, highlights the interest placed here on the right tail of the distribution of errors.

3.3.2 The Relaxed Aggregated Spatial Logit Model

Assumption 4 requires that decision-makers compare aggregate alternatives on the basis of the peak of elemental utility in them. For alternative J the peak is expressed as $\max_{j \in \{1,2,...M^J\}} U_j^J = \max_{j \in \{1,2,...M^J\}} (V_j^J + \epsilon_j^J + \epsilon^J) = \epsilon^J + \max_{j \in \{1,2,...M^J\}} (V_j^J + \epsilon_j^J)$. We shall demonstrate that under assumptions 1 and 2.1 for any aggregate J the normalized random variable $\max_{j \in \{1,2,...M^J\}} (V_j^J + \epsilon_j^J)$ tends to become Gumbel as M^J becomes large.

Let $Z_{M^J} = \max_{j \in \{1,2,...M^J\}} (V_j^J + \epsilon_j^J)$. Normalization of Z_{M^J} entails the formation of a random variable $(Z_{M^J} - a_{M^J})/b_{M^J}$, where a_{M^J} and b_{M^J} are sequences of constants dependent on M^J. In mathematical terms, we wish to demonstrate that $\lim_{M^J \to \infty} (Z_{M^J} - a_{M^J})/b_{M^J}$ is a Gumbel variate. Because F_j, $j \in \{1, 2, ...M^J\}$ are independent:

$$\Pr(\frac{Z_{M^J} - a_{M^J}}{b_{M^J}} < x) = \Pr(Z_{M^J} < a_{M^J} + b_{M^J} x) = \prod_{j=1}^{M^J} F_j(a_{M^J} + b_{M^J} x) \tag{15}$$

In light of equation 15, our objective is to identify the sequences a_{M^J} and b_{M^J}, and to demonstrate that for those sequences the right hand side of 15 tends to the double exponential function as M^J goes to infinity. This task can be accomplished with the help of the following proposition.

Proposition 4:

$$e^{-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})}} - 4(\sum_{j=1}^{M^J}(e^{-\lambda^J(x-V_j^J)})^2)\prod_{j=1}^{M^J} F_j(x)$$

$$< \prod_{j=1}^{M^J} F_j(x)$$

$$< e^{-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})}}$$

The proof of this proposition is rather elaborate and is provided in the appendix to this chapter. Obviously, the double exponential, which is the functional form of the Gumbel distribution, provides an upper limit for the distribution of the maximum. A close examination of the inequalities in proposition 4, not only provides a solution to the problem at hand, but it also suggests ways for the specification of a_{M^J} and b_{M^J}. If one specifies a_{M^J} and b_{M^J} in a way that the limit of $-\lambda^J(a_{M^J}+b_{M^J}x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})$ exists and is finite, while at the same time $\sum_{j=1}^{M^J}(e^{-\lambda^J(a_{M^J}+b_{M^J}x-V_j^J)})^2$ tends to zero as M^J goes to infinity, then the asymptotic distribution of the maximum is indeed Gumbel.

Obvious selections for a_{M^J} and b_{M^J} are:

$$a_{M^J} = \frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J} \text{ and } b_{M^J} = \frac{1}{\lambda^J} \quad (16)$$

For these selections we have $-\lambda^J(a_{M^J}+b_{M^J}x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J}) = -x$ that does not vary with M^J. We now demonstrate that $\sum_{j=1}^{M^J}(e^{-\lambda^J(a_{M^J}+b_{M^J}x-V_j^J)})^2$ becomes arbitrarily small as M^J becomes large. Observe that $a_{M^J} > V_j^J$ for all j. Thus, $a_{M^J} > V_k^J = \max_{j\in\{1,2,..M^J\}} V_j^J$. Also note that a_{M^J} increases with M^J since new positive terms are added to its summation as J is split into increasingly smaller elemental alternatives. For a suitable value of x then $e^{-\lambda^J(a_{M^J}+b_{M^J}x-V_k^J)} < \frac{1}{M^J}$. For this value of x this inequality holds for all j. Therefore:

$$0 \leq \sum_{j=1}^{M^J} (e^{-\lambda^J (a_{M^J} + b_{M^J} x - V_j^J)})^2 < \sum_{j=1}^{M^J} (\frac{1}{M^J})^2 = \frac{1}{M^J}$$

These inequalities make the point. To gain further insight consider a_{M^J} at the limit. Elemental alternatives at the limit become points with coordinates (y, z). Let A^J be the collection of all the points (y, z) that make up J. Then $V^J(y, z)$ for all $(y, z) \in A^J$ represents the surface of systematic utility over J. We now have:

$$\lim_{M^J \to \infty} a_{M^J} = \frac{1}{\lambda^J} \int \int_{(y,z) \in A^J} e^{\lambda^J V^J(y,z)} dy dz$$

The definite integral in the right hand side represents the volume between A^J on the (y, z) plane and the functional surface represented by $e^{\lambda^J V^J(y,z)}$. Clearly at the limit a_{M^J} becomes so large relative to V_k^J, the point of maximum utility, that $e^{-\lambda^J (a_{M^J} + b_{M^J} x - V_k^J)} < \frac{1}{M^J}$ is true for any x.

What we have shown is that for large M^J :

$$\Pr(\frac{Z_{M^J} - a_{M^J}}{b_{M^J}} < x) \approx e^{-e^{-x}}$$

This equation means that for large M^J, $(Z_{M^J} - a_{M^J})/b_{M^J}$ is approximately Gumbel distributed with parameters (0,1). Let us assume that M^J is large enough so that $(Z_{M^J} - a_{M^J})/b_{M^J}$, can be considered, for all practical purposes, Gumbel. According to the property of Gumbel (footnote 9) Z_{M^J} is also Gumbel with parameters $(a_{M^J}, 1/b_{M^J})$. This is equivalent to stating that for a suitably large M^J $\max_{j \in \{1,2,\dots M^J\}} (V_j^J + \epsilon_j^J)$ is Gumbel with parameters $(\frac{1}{\lambda^J} \ln \sum_{j=1}^{M^J} e^{\lambda^J V_j^J}, \lambda^J)$. Recall from the previous section that when ϵ_j^J, for all j, are Gumbel distributed with parameters $(0, \mu^J)$ then $\max_{j \in \{1,2,\dots M^J\}} (V_j^J + \epsilon_j^J)$ is always Gumbel with parameters $(\frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} e^{\mu^J V_j^J}, \mu^J)$ irrespective of the size of M^J. The similarity between λ^J and μ^J is also obvious. Note that assumption 3 still holds so that asymptotically the total error for an aggregate is Gumbel. Furthermore, all the properties of the aggregated logit model discussed in section 3.2 hold for arbitrarily large M^J.

3.3.3 Other Asymptotic Models

Relaxing the assumptions of the aggregated model has clear theoretical significance and allows one to gain further insight into the model. The question arises whether the relaxation of assumption 2 has any practical significance. The asymptotic theory of extremes is rich in results. Different combinations of conditions for distribution F and specifications for the normalizing parameters a_{M^J} and b_{M^J} lead to different types of models that can be useful in practice. An example of this is the set of models discussed by Kitamura et. al. (1979), one of the very few papers that deals with spatial aggregation in the context of discrete choice models. It is worth explaining further the contribution of this paper, using the mathematical notation introduced here.

Kitamura et. al. (1979) first deal with the aggregated logit model except that $Var(\epsilon^J) = 0$ for all J. This is the equivalent case of $\mu^J = 1$. Thus, elemental utilities are independent even if they belong to the same aggregate alternative. The error terms of elemental utilities are assumed to be normally distributed, while the normalizing parameters are defined as:

$$a_{M^J} = \overline{V}^J + \frac{1}{2} Var(V_j^J)(2\ln M^J)^{\frac{1}{2}} + (2\ln M^J)^{\frac{1}{2}}$$

$$b_{M^J} = \frac{(2\ln M^J)^{\frac{1}{2}}}{1 + \frac{1}{2} Var(V_j^J)}$$

where $Var(V_j^J)$ is the variance of the systematic elemental utilities within aggregate J. Under these circumstances one can prove that the maximum of elemental utilities within an aggregate J is an extreme value distributed random variable with parameters (a_{M^J}, b_{M^J}) (Galambos, 1978, pp.65).

In this case a_{M^J}, being the mode of the distribution, is the representative utility for aggregate J. The equivalent form in the aggregated logit model is $V^J = \ln \sum_{j=1}^{M^J} e^{V_j^J}$. The relationship between a_{M^J} and V^J can be seen if one expands the latter into a second order Taylor series around the mean $\overline{V}^J$ as

McFadden (1978) has done:

$$V^J \approx \overline{V}^J + \frac{1}{2}Var[V_j^J] + \ln M^J \tag{17}$$

In fact, McFadden (1978) has demonstrated that if elemental systematic utilities within aggregates are normally distributed then the right hand side expression of 17 is an asymptotic representation of V^J. By direct comparison of 17 with 8 the above statement implies that the heterogeneity term is asymptotically equal to $\frac{1}{2}Var[V_j^J]$. Thus, the heterogeneity term is directly linked to the variance of the systematic utilities in an aggregate.

Comparison of 17 to the expression for a_{M^J} clearly reveals the differences in the way utility is calculated for the two models. The most important difference, however, is in the scale parameters of the two models. In the case of the Gumbel elemental alternatives, the scale parameter μ remains unaffected with maximization, meaning that the aggregate utility has the same scale parameter as the elemental utilities. Increasing the number of elemental alternatives has no effect on the scale parameter. In contrast, the model with normal elemental utilities is associated with aggregate utilities that have scale parameter b_{M^J} which increases with the size of aggregate J. But we know that the variance of a Gumbel variate is inversely proportional to the square of its scale parameter. Therefore, in this case the variance of an aggregate utility decreases with the size of the aggregate.

Despite this nice asymptotic property of the aggregate utility, the model assumes no aggregate variance and thus zero covariance between elemental utilities that belong to the same aggregate. Attempts to extend the model in this direction by Kitamura et. al. (1979) led to the probit model with its known computational problems during estimation.

The aggregated logit model was first introduced by Lerman (1975). He actually used a first order Taylor series expansion of V^J around its mean $\overline{V}^J$:

$$V^J \approx \overline{V}^J + \ln M^J$$

Through his analysis he acknowledged the existence of the heterogeneity term but he found the size term to be more important in practical applications. McFadden (1978) is responsible for deriving the form of the model with the heterogeneity term.

Leonardi and Papageorgiou (1992) used the asymptotic theory of extremes in a different context. They contend that the assumption of the Gumbel distributed error terms in a multinomial logit model of location choice has no behavioural significance. Instead variations of the multinomial logit model can be derived by assuming that individuals associate a particular consumption bundle with every location. Individuals arrive at a consumption bundle through a large number of trials each of which is associated with an error obeying an assumption such as 2.1. Thus, their derivation bears some similarity to ours.

3.4 Conclusions

This paper has proposed the notion that a fundamental spatial effect in models of destination choice is a direct result of zonal aggregation. Such an effect results from the fact that the potential destinations in a choice set are invariably aggregates of the zones perceived by choice-makers. Derivation of the aggregated spatial logit model makes use of the theory of the nested logit model. Furthermore, we demonstrated by invoking the asymptotic theory of extremes that when the number of elemental alternatives in aggregate zones becomes large then the assumption of Gumbel distributed error terms, inherent in any logit model, can be replaced with a more behaviourally sound assumption.

The crucial assumption that is necessary for the derivation of the model is the independence of error terms. The formulation of the model allows error terms associated with elemental alternatives that belong to the same aggregate to be correlated and the level of correlation to vary between aggregates. This is in contrast to the multinomial logit model, where all error terms are uncorrelated. To put it differently, the multinomial logit model has a diagonal variance-covariance matrix of error terms while the aggregated logit model allows certain blocks of the matrix corresponding to elemental zones within the same aggregate to have non-zero covariance.

With respect to the structure of the variance-covariance matrix, the aggregated logit model is more restrictive than the model proposed by Ben-Akiva and Bolduc (1991). This loss in statistical generality is balanced by the gain in theoretical insight. The μ^J, $J \in \{1, 2, ...L\}$ parameters that were extensively discussed in this paper are similar to the autocorrelation coefficients in autoregressive models. Their significance, however, as demonstrated in this paper, is more important than merely conveying information on the spatial dependency of utility. Furthermore, the Ben-Akiva/Bolduc formulation leads to the probit model with its known computational problems, given the typically large choice sets encountered in practical spatial choice problems. Thus, the aggregated spatial choice model combines a theoretically rich formulation with computational tractability. In terms of the structure of its error term variance-covariance matrix, it strikes the middle ground between the multinomial logit model and the Ben-Akiva/Bolduc model.

An estimation procedure for the aggregated logit model is discussed in Chapter 5. Although the performance of the aggregated model relative to the multinomial logit model is assessed extensively with simulated data in Chapter 4, there is a need to evaluate the model in real world applications. An initial attempt in this direction is taken in Chapter 6. More empirical work with Canadian interprovincial migration data is presently under way. Also, the asymptotic model derived here is viewed as an initial attempt to derive spatial choice models under different sets of assumption. More specifically, a possible development is the derivation of asymptotic models where the assumption of independence of errors between aggregates is relaxed.

3.5 Appendix to Chapter 3: Proof of Proposition 4

The proof of proposition 4 follows the ideas developed in lemma 1.3.1., theorem 1.3.1. and corollary 1.3.1. of Galambos (1978, pp. 10-15). Those propositions are general for the maximum of independent and identically distributed random variables. In proposition 4 we deal specifically with the maximum of exponential distributions which are independent and identically distributed but which differ with respect to the location parameter as shown in equation 14.

Proposition 4:

$$e^{-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})}} - 4\left(\sum_{j=1}^{M^J}(e^{-\lambda^J(x-V_j^J)})^2\right)\prod_{j=1}^{M^J} F_j(x)$$

$$< \prod_{j=1}^{M^J} F_j(x)$$

$$< e^{-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})}}$$

Proof:

For the right hand side inequality we need to establish:

$$\prod_{j=1}^{M^J}(1-e^{-\lambda^J(x-V_j^J)}) = e^{\sum_{j=1}^{M^J}\ln(1-e^{-\lambda^J(x-V_j^J)})} < e^{-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\mu V_j^J})}}$$

But this is equivalent to:

$$\sum_{j=1}^{M^J}\ln(1-e^{-\lambda^J(x-V_j^J)}) < -e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\mu V_j^J})} = -\sum_{j=1}^{M^J} e^{-\lambda^J(x-V_j^J)}$$

which will be true if:

$$\ln(1-e^{-\lambda^J(x-V_j^J)}) < -e^{-\lambda^J(x-V_j^J)} \text{ for all } j \in \{1,2,...M^J\}$$

But the inequality $\ln(1-z) < -z$ is true for every $0 < z < 1$. It is then sufficient to observe that $\lambda^J > 0$ and $x > V_j^J$ which leads to $0 < e^{-\lambda^J(x-V_j^J)} < 1$.

For the left hand side, let us start from the inequality $z\ln(z) > z-1$, which is true for $0 < z < 1$. Set $z = 1 - e^{-\lambda^J(x-V_j^J)}$. Then, for every $j \in \{1, 2, ...M^J\}$:

$$(1 - e^{-\lambda^J(x-V_j^J)})\ln(1 - e^{-\lambda^J(x-V_j^J)}) > e^{-\lambda^J(x-V_j^J)}$$

Thus:

$$\sum_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})\ln(1 - e^{-\lambda^J(x-V_j^J)}) > -\sum_{j=1}^{M^J} e^{-\lambda^J(x-V_j^J)} = -e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})}$$

Or:

$$\prod_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})\prod_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})^{-e^{-\lambda^J(x-V_j^J)}} > e^{-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})}}$$

This inequality can be re-arranged as:

$$-e^{-\lambda^J(x-\frac{1}{\lambda^J}\ln\sum_{j=1}^{M^J} e^{\lambda^J V_j^J})} - \prod_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})\{\prod_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})^{-e^{-\lambda^J(x-V_j^J)}} - 1\}$$

$$< \prod_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})$$

Observe now that:

$$\prod_{j=1}^{M^J}(1 - e^{-\lambda^J(x-V_j^J)})^{-e^{-\lambda^J(x-V_j^J)}} = e^{\sum_{j=1}^{M^J}(-e^{-\lambda^J(x-V_j^J)}\ln(1-e^{-\lambda^J(x-V_j^J)}))} < e^{2\sum_{j=1}^{M^J} e^{-2\lambda^J(x-V_j^J)}}$$

This is because $-\ln(1 - e^{-\lambda^J(x-V_j^J)}) < 2e^{-\lambda^J(x-V_j^J)}$. The result follows from the observation that:

$$e^{2\sum_{j=1}^{M^J} e^{-2\lambda^J(x-V_j^J)}} - 1 < 4\sum_{j=1}^{M^J} e^{-2\lambda^J(x-V_j^J)} \square$$

Chapter 4

Utility Variability Within Aggregate Spatial Units and its Relevance to Discrete Models of Destination Choice

4.1 Introduction

In this chapter, we examine a logit model (the aggregated spatial logit) designed specifically for spatial choice among aggregate destinations. Typically, the logit and gravity models are applied to problems of spatial interaction without due consideration for the aggregation scale of the data. There is of course a general awareness that models estimated at different levels of aggregation will yield different parameter estimates but most researchers proceed directly with their analysis at the aggregate level and make no adjustments. Research on aggregation issues has been intertwined with work on the 'modifiable areal unit problem' (MAUP). Research on MAUP examines how spatial statistical models and diagnostics are affected by the aggregation scale and spatial configuration of spatial units. **Choice processes** per se, as are present in all problems of spatial interaction, are not specifically addressed in this body of research. Rather the focus is on the univariate or bivariate spatial processes associated with one or two spatially-referenced attributes (Arbia, 1989; Openshaw, 1984). Aggregation issues are relevant in both scenarios but as will be illustrated in this paper, the theoretical framework to accommodate aggregation in a choice context is quite different from that considered in analysis of the modifiable areal unit problem.

From the perspective of geographers and regional scientists, the study of aggregation in the choice context has been confined mostly to work on the gravity model (Batty and Sikdar, 1982a, 1982b, 1982c,1982d; Webber, 1980; Schwab and Smith, 1985; Putman and Chung, 1989; Amrhein and Flowerdew, 1992). Most research on aggregation in discrete choice modelling has been done by those outside these fields (Kitamura, 1979; McFadden, 1978) and is not widely applied by practitioners in the spatial sciences. Given the work by Anas (1983) on the strong similarity between the gravity and logit models, the insights gained from the work outside regional science and geography need to be understood and further developed in the spatial context.

It is interesting that most contemporary work in spatial interaction modelling has tended to focus on spatial substitutability and structure effects as the primary contributors to model misspecification (Griffith and Jones, 1980; Fotheringham, 1981, 1983; Haynes and Fotheringham, 1990; Borgers and Timmermans, 1987 ; Lo, 1991), often not recognizing the fact that the choice process is typically not at the same aggregation level as the data. Problems of substitutability among alternatives can often be solved with alternate aggregation schemes but aggregation biases cannot be eliminated through accounting for substitutability. As was pointed in Chapter 2, however, it is debatable whether much of this contemporary work has really gotten to the core of the substitutability problem. Fortunately, the framework described here is capable of addressing each of these issues.

In the practical terms of a choice problem, interaction data are often available only at the level of aggregate spatial units and in other cases, the interaction data are so disaggregated that the number of spatial units involved would be prohibitively large. In either situation, the researcher is presented with a problem whereby the spatial choice units of the model are more aggregated than those likely evaluated in reality by choice-makers. Considering the spatial aggregation of the data to be exogenous then, and not necessarily optimal, how should a researcher correctly go about representing the utility of aggregates and obtaining appropriate parameter estimates?

The central premise of this chapter is that to succeed in this objective it is necessary to account for utility variability within aggregates. In a study of inter-regional migration, for example, it is unlikely that choice-makers value each point within a region equally. However, when the ordinary multinomial logit model is applied to a regional choice set, this is exactly what is assumed. All variables employed in the model are some form of regional average implying **average** regional utilities. Alternatively, it is possible to employ variables measured at the sub-aggregate level and then represent utility at a finer level of resolution. Using such an approach, a theoretically appealing aggregated model can be derived which postulates the utility of an aggregate to equal the point of **maximum** utility within the aggregate. While the choice data for spatial interaction are often aggregated over space and across decision-makers, there is no need to compound the problem by utilizing spatially aggregated explanatory variables.

It is interesting to note that work in spatial econometrics has not found common ground with that of spatial choice analysis, particularly when aggregation is an issue. Typically, discrete choice models have found favour with spatial econometricians in studies utilizing a categorical spatially-referenced dependent variable. The problem is normally binary, with the alternatives being non-spatial since they usually represent some attribute which can be measured over space, and the model is typically statistical rather than being a manifestation of choice theory. Such a study is that of McMillen (1992) who employs a probit model in the examination of crime rates as a dependent variable. In a similar vein, Bolduc et al. (1989) examine issues such as spatial autocorrelation and heteroskedasticity but in the context of mode choice where the decisions of travellers are seen to be dependent on location. These examples are to be contrasted with the model in this chapter, which is clearly a choice model whose alternatives are multiple spatial units. Just as spatial econometric techniques apply to the work of McMillen and Bolduc et al., they are of interest in the context of spatial choice but it must be recognized that such analysis is made more complex by the fact that the number of alternatives is large, that a heterogeneous population is choosing among these spatial units, and that issues of aggregation are present. In this chapter, we present a model which addresses the effects of aggregation elegantly but is less effective in accounting for complex spatial dependency.

This chapter is divided into five main sections. In the second section to follow, we summarize the theoretical basis for the aggregated model and provide necessary background information. In the third section, a basic estimation procedure is outlined and differences with the estimation of the ordinary multinomial logit model (OMNL) are stressed. In the fourth section, the specification of the model when data are highly aggregated across choice-makers is contrasted to micro-data specifications. These specifications are then tested in a controlled environment. More specifically, the performance of the average utility model is compared with that of the maximum utility aggregated model when the utility heterogeneity of aggregates is systematically altered. In the concluding section, some synthesizing remarks are offered with a view toward future work.

4.2 Theoretical Background

In problems of spatial choice, the study area is typically divided into zones necessitated by the available interaction data or by practical considerations which dictate that the aggregation level must be manageable. The resulting aggregate choice set does not necessarily correspond to the zones perceived by individual choice-makers in the spatial choice process. Usually, decision-makers will 'perceive' a more disaggregate zonal system composed of alternatives which are termed *elemental*. Although in reality the perception of elemental alternatives and hence the definition of individual choice sets may change with the characteristics and the origin of the decision-maker (Thill, 1992; Tardiff, 1979), it is assumed here that all choice-makers perceive the same zonal system on average. In practical models of spatial choice then, the problem of aggregation is always present where zonal systems are imposed. The particular model of spatial choice examined below is the multinomial logit (OMNL). Theory indicates that the use of aggregate choice units necessitates the introduction of corrective size and heterogeneity terms which modify utility. 'Size' refers to the number of elemental alternatives within an aggregate and 'heterogeneity' is a measure of the variation of elemental utility within that aggregate.

The aggregated model in its most basic form dates to the work of Lerman (1975). In a problem of intra-urban housing choice, Lerman showed that the utility of an aggregate spatial unit depends on its size. Simulations illustrated that his model was unsatisfactory when aggregates were heterogeneous, suggesting the incorporation of a variance term in the utility of the aggregate. McFadden (1978) devised the improved theoretical framework which provides the necessary size and variance effects. Other things being equal, a heterogeneous aggregate alternative is more likely to contain a utility peak than a homogeneous aggregate and therefore will be more attractive to choice-makers. The heterogeneity term captures this effect and upgrades the estimated utility of heterogeneous aggregates, unlike the OMNL.

With this background in mind, we now briefly outline the theory behind the aggregated model in the context of inter-regional migration. A more comprehensive theoretical discussion was provided in Chapter 3. The focus below is on means to model the process of destination choice where potential migrants have already made the decision to move. Consider that the study area is divided into M elemental alternatives. These are the actual spatial units which the typical migrant considers as potential destinations. Assume that the M spatial units are combined into L aggregates, each of which contains $M^J, J \in \{1, 2, ..., L\}$ elemental units. Thus:

$$\sum_{J=1}^{L} M^J = M \tag{1}$$

Given that individual n has decided to depart origin i, their utility for prospective elemental destination unit j within aggregate J is:

$$U_{ijn}^J = V_{ijn}^J + \epsilon_{in}^J + \epsilon_{ijn} \tag{2}$$

where:

- V_{ijn}^J is a deterministic component expressed as a function of attributes which relate to the aggregate alternative J, the elemental alternative $j \in \{1, 2, ..., M^J\}$ and characteristics of the choice-maker. Physical and social distance variables between i and j are also utilized.
- ϵ_{in}^J is a Gumbel distributed random component that captures unobserved variation relating to the aggregate spatial unit J.
- ϵ_{ijn} is a Gumbel distributed random component with parameters $(0, \mu^J)$ that captures unobserved variation relating to the elemental spatial unit $j \in \{1, 2, ..., M^J\}$.

The ϵ_{ijn} terms are independent of each other within and between aggregates and are also independent of the ϵ_{in}^J terms. The ϵ_{in}^J terms are themselves assumed independent of each other. Furthermore, it is assumed that $\epsilon_{ijn} + \epsilon_{in}^J$ are Gumbel variates with parameters $(0, \mu)$. In essence, these assumptions imply that the total error ϵ_{ijn}^J for an elemental alternative is the sum of two independent component errors. All utilities of elemental units within aggregate J share the same component ϵ_{in}^J implying that the utilities of any two elemental units within the same aggregate are not independent. As is the case with many spatial econometric models, contiguity implies similarity. Of course in this context, contiguity is defined by an elemental alternative having a neighbour within the same aggregate, and the similarity evolves from a spatial process which determines utility.While there is a positive correlation between intra-aggregate elemental utilities, inter-aggregate elemental utilities are uncorrelated in this model.

In addition, the assumptions imply that elemental errors within the same aggregate are homoskedastistic, which is potentially restrictive. Meanwhile $Var(\epsilon_{ijn})$ is allowed to vary across aggregates, hence accommodating heteroskedasticity to some extent. In a context other than spatial choice, McMillen (1992) has examined the issue of heteroskedasticity over space in probit models.

Since choice-makers are hypothesized to seek the elemental zone within an aggregate which provides maximum utility, this zone's estimate of utility in effect determines the utility of the aggregate:

$$U_{in}^{J} = \epsilon_{in}^{J} + \max_{j \in \{1,2,\ldots,M^J\}} (V_{ijn}^{J} + \epsilon_{ijn}) \tag{3}$$

where the maximization term is a Gumbel variate with parameters

$$\left(\frac{1}{\mu^J} \ln \sum_{j=1}^{M^J} \exp(\mu^J V_{ijn}^J), \mu^J \right).$$

The first expression defines the mode of the distribution (which is interpreted as the aggregate systematic utility) and the second term is a scale parameter which is inversely proportional to the variance of the utility distribution. We can represent the choice probability of an aggregate in the general format of an ordinary multinomial logit which is also the expression for the aggregated model:

$$P_n(J) = \frac{\exp\left[\frac{\mu}{\mu^J} \ln \sum_{j=1}^{M^J} \exp\left(\mu^J V_{ijn}^J\right)\right]}{\sum_{K=1}^{L} \exp\left[\frac{\mu}{\mu^K} \ln \sum_{k=1}^{M^K} \exp\left(\mu^K V_{ikn}^K\right)\right]} \tag{4}$$

In this expression, μ^J is related to $Var(\epsilon_{ijn})$ while μ is related to $Var(\epsilon_{ijn}+\epsilon_{in}^J)$, the combined uncaptured variation of strictly aggregate and elemental utilities. Note that μ^J is allowed to vary across aggregate alternatives, possibly being unique for each aggregate or unique for a subset of aggregates. The model is thus sufficiently general that the distribution of unobserved elemental variation is permitted to differ across aggregates. As a by-product of estimating maximum utilities, the proposed model yields estimates of elemental utilities within the aggregates.

The ratio μ/μ^J is critical to the theoretical interpretation of the aggregated model and can be viewed as an overall measure of the correlation of elemental utilities within aggregate alternatives. We assume that elemental utilities within the same aggregate are positively correlated but that those between aggregates are uncorrelated. The correlation is assumed to result from common unobserved variation (ϵ_{in}^J) of the aggregate which they share, not from correlation among ϵ_{ijn}, $j \in \{1, 2, ..., M^J\}$. It is clear that these assumptions will prove problematic on any occasion when there is high correlation in utility between two elemental spatial units belonging to **different** aggregates. Similarly, the utility dependency between pairs of elemental alternatives within the same aggregate may vary. Such violations of assumptions may require redefinition of aggregate units if possible.

Irrespective of the definition of the aggregates, it is possible to show that:

$$\frac{\mu}{\mu^J} = \frac{Var(\epsilon_{ijn})}{Var(\epsilon_{ijn} + \epsilon_{in}^J)} = \sqrt{1 - corr(U_{ijn}^J, U_{ikn}^J)} \tag{5}$$

One value applies to all pairs of elemental alternatives which belong to the same aggregate. If the correlation tends to become high then $\mu/\mu^J \rightarrow 0$, while $\mu/\mu^J \rightarrow 1$ as the correlation becomes low. Estimation procedures allow for the identification of the ratio μ/μ^J, $J \in \{1, 2, ..., L\}$, but not the true values of its components. Without loss of generality though, the convention $\mu = 1$ is typically adopted. If the unobserved elemental utility variation within aggregate J is small relative to the unobserved variation shared by all elemental alternatives within J, then μ^J is large relative to μ, implying highly correlated total utilities among the elemental alternatives in J. In this case, the estimate of the expected maximum elemental utility corresponds closely with the largest of the elemental utilities predicted by the model. Conversely, if unobserved elemental variation is relatively large implying uncorrelated elemental utilities, then there is a lot of uncertainty with respect to the true levels of strictly elemental utilities. Given the individual expected values of Gumbel variates, the expected value of their maximum will be larger the greater the variance of the individual Gumbel variates. It is this principle which upgrades the estimate of systematic utility for an aggregate which contains a set of elemental alternatives with uncertain utilities.

In the context of the definition of aggregate alternatives, it is desirable that $\mu/\mu^J \rightarrow 0$ as this implies that there is little uncertainty at the sub-aggregate level and that similar elemental alternatives form the aggregate. The factors affecting the choice of an aggregate are also a good representation of the choice behaviour at the sub-aggregate level. If the ratio is close to one, the implication is that the individual elemental utilities are not being modelled very well within that aggregate when only the aggregate choice behaviour is used to estimate the model.

If econometric estimation yields a value of μ/μ^J outside the range $(0, 1]$ then the result indicates a violation of assumptions or a poorly specified model. In practice, the ratio is often found to be greater than one indicating that elemental utilities are assigned a very large variance. Such a violation of theory has the effect of upgrading the utility of that aggregate and reducing the variance of elemental systematic utilities. It makes no theoretical sense however to increase the elemental variance beyond the total variance established by μ and thus an improved model specification is desirable. Similar conclusions apply for an estimated negative ratio, although such an event is rare in practice.

Clearly, the behaviour of the ratio of scale parameters is critical to the interpretation of the model. The estimates of other explanatory variables as well as the goodness-of-fit of the model is very much related to the estimates of these ratios across aggregates. These issues are sufficiently important that their detailed discussion and analysis are reserved for future work. In this paper, the focus is on contrasting the performance of OMNL with that of model 4 when μ/μ^J is given exogenously.

It is important to understand the composition of the expected maximum utility for an aggregate alternative. Through algebraic manipulation, it is possible to show that the initial model in 4 can be written as:

$$\begin{aligned} P_n(J) &= \frac{\exp\left[V_{in}^J\right]}{\sum_{K=1}^{L} \exp\left[V_{in}^K\right]} \\ V_{in}^J &= \mu\,\bar{V}_{in}^J + \frac{\mu}{\mu^J}\ln\left[\frac{1}{M^J}\sum_{j=1}^{M^J}\exp\left(\mu^J(V_{ijn} - \bar{V}_{in}^J)\right)\right] + \frac{\mu}{\mu^J}\ln M^J \end{aligned} \tag{6}$$

With $\mu = 1$, note that the expected maximum elemental utility from the set of elemental alternatives in an aggregate (i.e. the systematic utility of the aggregate) can be broken into a mean utility effect: $\bar{V}_{in}^{J}$, a heterogeneity effect: $(1/\mu^J)\ln\left[(1/M^J)\sum_{j=1}^{M^J}\exp\left(\mu^J(V_{ijn}-\bar{V}_{in}^{J})\right)\right]$, and a size effect: $(1/\mu^J)\ln M^J$. In contrast, the OMNL utilizes only the mean utility term. However, the predicted mean utility with the OMNL will be different from that predicted by the proposed model since the parameters estimated will be dependent on the presence of the additional corrective terms. The size effect is constant across decision-makers and is self-explanatory. The heterogeneity effect varies across choice-makers and aggregates and, as mentioned earlier, is a measure of the variability of elemental systematic utilities. Typically, the correction from the size effect will be larger than that of the heterogeneity effect (Lerman, 1975).

A condensed form of 6 is also intuitively appealing:

$$P_n(J) = \frac{\exp\left[\bar{V}_{in}^{J} + \frac{1}{\mu^J}\ln\sum_{j=1}^{M^J}\exp\left(\mu^J(V_{ijn}-\bar{V}_{in}^{J})\right)\right]}{\sum_{K=1}^{L}\exp\left[\bar{V}_{in}^{K} + \frac{1}{\mu^K}\ln\sum_{k=1}^{M^K}\exp\left(\mu^K(V_{ikn}-\bar{V}_{in}^{K})\right)\right]} \tag{7}$$

where the convention $\mu = 1$ has been used in this expression. The size and heterogeneity terms from 6 are combined to form a term which measures the expected maximum deviation from the mean utility of aggregate alternative J. From the latter term it is easy to extract the heterogeneity and size effects for each choice-maker and aggregate alternative since the size effect does not vary across choice-makers. Equations 4,6 and 7 are simply different ways of expressing the aggregated model, each of which makes different aspects of the model clear. Since in this chapter we are particularly interested in the heterogeneity of aggregate spatial units, the aggregated model as represented by these equations will be identified with the acronym HETRO.

It is important to emphasize at this point that the mean utility term in 6 and 7 differs from mean utility as typically represented in applications of the OMNL. The $\bar{V}_{in}^{J}$term is assumed to be the simple *unweighted* average of the constituent elemental utilities. By extension, this implies that the attributes of aggregates are taken as the unweighted averages of the attributes measured at the elemental level. In practice, the attributes measured at the aggregate level and applied to represent the utility of aggregates in the OMNL are *weighted* averages of the contained elemental attributes. Typically, the aggregate averages in this context are weighted by the populations associated with individual elemental alternatives as would be the case with economic variables such as wages. With the exception of a perfectly even population distribution, the unweighted average of attributes relevant to 6 and the weighted average of the OMNL will not be the same. Of course for variables which cannot be related to choice-makers, such as climatic attributes, the concept of a weighted average makes little sense. This discussion highlights another potential difficulty with the OMNL, namely that the elemental alternatives which contribute most to some weighted average of appropriate variable types may not be the ones which provide the likely point of maximum utility for a given choice-maker.

To conclude this section , it should be mentioned that to the knowledge of these authors, no empirical study has utilized the heterogeneity and size effects in the manner advocated above[13]. In the context of a full nested logit model of housing choice, Fischer and Aufhauser (1988) do acknowledge the presence of a heterogeneity effect in the choice of dwelling units within census tracts. They do not acknowledge though that the same idea also applies to the higher levels of their model where the spatial units are more aggregated. For that matter, it does not seem widely known that the theory behind HETRO transcends the field of intra-urban housing choice and that there is a heterogeneity effect to be considered in any spatial or non-spatial choice problem where the alternatives are aggregations of perceived elemental alternatives. The approach seems particularly applicable in problems of inter-regional choice such as human migration or industrial location.

[13]Since the paper on which this chapter is based was first written, papers have been found, such as Parsons and Needleman (1992) which come closer to the form of the aggregated model. These are discussed in detail in Chapter 2. Nevertheless, the statement made here, to the best of our knowledge, remains valid.

4.3 Estimation of the Maximum Utility Model

In this section the estimation procedure for HETRO is outlined. Also, distinctions with the estimation normally used for the OMNL are highlighted. The procedure established in this section is utilized in the simulations which follow.

The likelihood function for L aggregate zones and N observed choice-makers is:

$$\mathcal{L}(\beta, \mu) = \prod_{n=1}^{N} \prod_{J=1}^{L} P_n(J)^{Y_{Jn}} \tag{8}$$

where:

- β is a vector of parameters that enter the systematic utility V_{ijn}^{J}
- μ is a vector of parameters representing $(\mu^1, \mu^2, ..., \mu^L)$
- Y_{Jn} is a dummy variable defined as 1 when decision-maker n is observed to choose aggregate J and 0 if not

The log-likelihood function is:

$$\mathcal{L}^*(\beta, \mu) = \sum_{n=1}^{N} \sum_{J=1}^{L} Y_{Jn} \ln P_n(J) \tag{9}$$

A suitable estimation procedure must identify the values of β and μ that maximize $\mathcal{L}^*(\beta, \mu)$. Endogenous estimation of μ complicates the estimation procedure considerably. Since the focus of this paper is not on estimation, it is accepted that the values of μ are specified exogenously. Discussion on their endogenous estimation is reserved for Chapter 5. Bearing this in mind, the log-likelihood function has a maximum where:

$$\frac{\partial \mathcal{L}^*(\beta)}{\partial \beta} = 0 \tag{10}$$

The first derivative of $\mathcal{L}^*(\beta)$ with respect to the kth parameter is:

$$\frac{\partial \mathcal{L}^*(\beta)}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{J=1}^{L} \frac{Y_n^J}{P_n(J)} \frac{\partial P_n(J)}{\partial \beta_k} \tag{11}$$

In general, the partial derivative of a logit probability with respect to β_k is:

$$\frac{\partial P_n(J)}{\partial \beta_k} = P_n(J)\left[\frac{\partial V_{in}^J}{\partial \beta_k} - \sum_{Q=1}^{L} \frac{\partial V_{in}^Q}{\partial \beta_k} P_n(Q)\right] \tag{12}$$

Now for OMNL, the systematic utility V_{in}^J is given by the mean utility term:

$$V_{in}^J = \sum_{k=1}^{K} \beta_k \bar{X}_{ink}^J \tag{13}$$

which has the partial derivative:

$$\frac{\partial V_{in}^J}{\partial \beta_k} = \bar{X}_{ink}^J \tag{14}$$

Substituting 14 into 12, the partial derivative of the OMNL probability is obtained:

$$\frac{\partial P_n(J)}{\partial \beta_k} = P_n(J)\left[\bar{X}_{ink}^J - \sum_{Q=1}^{L} \bar{X}_{ink}^Q P_n(Q)\right] \tag{15}$$

The equivalent of 14 for HETRO[14] is:

$$\frac{\partial V_{in}^J}{\partial \beta_k} = \bar{X}_{ink}^J + \frac{\sum_{j=1}^{M^J} (X_{ijnk} - \bar{X}_{ink}^J) \exp\left[\mu^J \sum_{l=1} \beta_l (X_{ijnl} - \bar{X}_{inl}^J)\right]}{\sum_{j=1}^{M^J} \exp\left[\mu^J \sum_{l=1} \beta_l (X_{ijnl} - \bar{X}_{inl}^J)\right]} \tag{16}$$

The derivative $\partial P_n(J)/\partial \beta_k$ for HETRO is obtained through the substitution of 16 into 12.

Given the partial derivatives of the HETRO probabilities, the iterative solution is no different from that of OMNL. Cramer (1991) advocates the use of the method of scoring, a gradient technique similar to the Newton -Raphson approach. Like other methods in this family, the gradient vector $\mathbf{q} = \partial \mathcal{L}^*(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$, also known as a score vector, and a direction matrix must be calculated at each iteration. The non-linear system of equations defined by setting $\mathbf{q}$ equal to zero must be solved iteratively. A Taylor series approximation of $\mathbf{q}(\boldsymbol{\beta})$ around $\boldsymbol{\beta}'$ in the vicinity of $\boldsymbol{\beta}$ is utilized:

$$\mathbf{q}(\boldsymbol{\beta}) \approx \mathbf{q}(\boldsymbol{\beta}') + \mathbf{H}(\boldsymbol{\beta}')(\boldsymbol{\beta} - \boldsymbol{\beta}') \tag{17}$$

[14]Here the derivative of the aggregate systematic utility of the aggregated model is taken with respect to the expression in 7. In Appendix 2, it is shown that a more concise expression is obtained if the derivative is taken with respect to the systematic utility in 4.

where $H_{kl} = \partial^2 \mathcal{L}^*(\beta)/(\partial\beta_k \partial\beta_l)$ represents an element in the Hessian matrix $\mathbf{H}$ of second derivatives. Since $\mathbf{q}(\beta) = 0$, 17 can be rearranged to yield:

$$\beta \approx \beta' - \mathbf{H}(\beta')^{-1}\mathbf{q}(\beta') \tag{18}$$

When this approximation is applied repeatedly in the vicinity of the optimum, with β being substituted for β', convergence should occur. This forms the basis for the Newton-Raphson algorithm expressed as:

$$\beta_{t+1} \approx \beta_t - \mathbf{H}(\beta_t)^{-1}\mathbf{q}(\beta_t) \tag{19}$$

With respect to either OMNL or HETRO, the most convenient expression for each element of $\mathbf{q}$ is:

$$\frac{\partial \mathcal{L}^*(\beta)}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{J=1}^{L} [Y_n^J - P_n(J)] \frac{\partial V_{in}^J}{\partial \beta_k} \tag{20}$$

which is obtained when 12 is substituted into 11 and the result simplified.

The method of scoring differs from the Newton-Raphson approach in that the information matrix replaces the Hessian. Since the asymptotic variance-covariance matrix of β is directly related to the information matrix, this replacement is useful. The expression for any element of the information matrix is:

$$Q_{kl} = \sum_{n=1}^{N} \sum_{J=1}^{L} \frac{1}{P_n(J)} \frac{\partial P_n(J)}{\partial \beta_k} \frac{\partial F_n(J)}{\partial \beta_l} \tag{21}$$

where 12 is substituted into 21. Since 16 for HETRO is quite different from 14 for OMNL, the results for Q_{kl} will also differ. Because the information matrix is given by $-E[\mathbf{H}]$, the negative expectation of the Hessian, the Newton-Raphson sequence of 19 is revised:

$$\beta_{t+1} \approx \beta_t + (-E[\mathbf{H}(\beta_t)])^{-1}\, \mathbf{q}(\beta_t) \tag{22}$$

McFadden (1974) has shown that for OMNL, the Hessian matrix will be negative definite under all but extreme circumstances. This implies that the information matrix should be positive definite. After extensive experimentation with simulated data we conclude that the convergence success rate is almost as high for HETRO with μ exogenous as it is for OMNL. As will be shown in future work, this success rate is considerably lower when μ is endogenous. A modified estimation procedure is needed in such circumstances.

An important issue affecting the convergence of HETRO is the possibility of high correlation among the different destination attributes. Such attributes, which are typically used in spatial choice models to specify utilities, and which do not vary across individuals, can potentially create dependencies in the information matrix. In the cases where convergence does not occur, it is likely that the information matrix is near singular. Clearly, it is critical to thoroughly screen destination attributes, and any interaction terms related to them, beforehand. It is suggested that a technique such as principal components analysis might be applied in situations where many generic destination variables are used. By utilizing the significant principal components, most of the variance is accounted for without risk of multicollinearity.

4.4 Model Specification

To an extent, specification of the aggregated model depends on whether the interaction data are aggregated or disaggregated across choice-makers which is a problem distinct from spatial aggregation. Typically, the variables used to specify utilities are classified as: destination attributes, measures of distance and socio-economic individual characteristics. The latter category is prevalent with data that are totally disaggregated across choice-makers while the first two categories are relevant to both aggregation levels. In this section, we consider differences in models which are aggregated spatially but may or may not be aggregated across decision-makers

4.4.1 Specifications with Data Aggregated Across Choice-makers

Several studies of human migration utilize logit models with interaction data aggregated across both choice-makers and destinations (e.g. Anderson and Papageorgiou (1994), Day (1992)). While there is no question that the true utilities choice-makers have for a particular destination are highly variable, the use of aggregate data precludes the differentiation of systematic utilities among choice-makers originating from the same zone. In essence, such a model works on the premise that the 'average' individual can be used as representative of all individuals from a given origin. Distance is one variable type which does introduce variation in systematic utilities across decision-makers since it can be used to segment the population based on the zone of origin[15]. A similar approach is the use of so-called 'social distance' variables.

An issue of some importance is the use of alternative-specific versus generic specification techniques. The former implies that for a given variable, the marginal effect on utility of a fixed change in this variable differs across alternatives, while the latter implies that the marginal effect is constant. Alternative-specific specifications of distance variables and destination variables can be problematic. With respect to distance, there is the issue of whether it is theoretically valid to allow the friction-of-distance to vary across destinations. Are differences in parameters a product of the fact that distance interacts with destination attributes or does the friction-of-distance vary? Some argue that there is no single valid measure of the friction-of-distance (Lo, 1993). Alternative-specific specifications of destination attributes can cause problems in that each parameter associated with this variable type implies one fewer alternative-specific constant that can be used (see Chapter 2). If there are more than $L-1$ such parameters, a model becomes indeterminate since it is no longer possible to measure utility in relative terms.

With respect to HETRO, distance variables must be measured at the sub-aggregate level. In so doing, the systematic utility of a prospective aggregate destination from a given origin will tend to depend on the closest elemental zone within the aggregate, not the average distance to all elemental zones. In a completely homogeneous study area, we would expect people to choose the closest elemental alternative

[15] In a model of destination choice, it is possible to specify origin characteristics in the alternative-specific manner associated with socio-economic individual variables. The main problem would be that in a model with 25 alternatives, up to 24 parameters could in theory be associated with a given origin characteristic. Clearly, this approach does not result in parsimonious models and often it is impossible to develop reasonable hypotheses about which subsets of alternatives should be constrained to the same parameter.

within the aggregate. That people do not do this in reality reflects the fact that other factors are at work which ought not be confused with the friction-of-distance.

Destination attributes can be measured either at the aggregate level or the sub-aggregate level. Variables associated with, for example, climate or employment are really sub-aggregate phenomena which can never be properly captured by the OMNL. Their values fluctuate within aggregates and hence should introduce variability into utilities. Other attributes such as regional tax rates are constant across a given aggregate. Since there is no intra-aggregate utility variability introduced by such variables, they are specified as they would be in OMNL.

4.4.2 Specification Issues With Micro-data

Observations in micro-data are at the individual choice-maker level. Such data usually incorporate variables describing the socio-economic characteristics of decision-makers and are thus richer in information content than aggregate data sets. With the help of such variables, it is possible to differentiate the utilities of individual choice-makers or population segments and capture the heterogeneity of tastes in the population. As in the case of aggregate data, however, destination attributes do not vary across decision-makers and act as substitutes for alternative-specific constants.

Socio-economic individual characteristics available in micro-data are specified in an alternative-specific manner since such variables will typically affect the utilities associated with different destinations in different ways. The marginal contribution to utility of a decision-maker's income, for example, might be positive for one destination but negative in another. However, since utility is a relative concept in the logit model, some spatial unit must act as a zero utility reference against which the utilities of other alternatives are compared. The option is available to specify both OMNL and HETRO in this manner although the number of parameters may proliferate to unacceptable levels if there are many spatial alternatives.

A difficulty in using this approach for HETRO is that the socio-economic variables can be used to differentiate utilities only on an aggregate alternative basis. One might reasonably expect that a socio-economic variable will affect the perceived utility of elemental alternatives within the same aggregate quite differently. However, since the choice behaviour is unobserved at the elemental level and since the number of elemental alternatives is likely unmanageably large, alternative-specific specifications at the sub-aggregate level are not feasible.

An alternative means of introducing variability into the elemental utilities associated with socio-economic characteristics is to specify these variables interactively with destination attributes. This approach is appealing in the sense that utilities are modelled as a matching of decision-maker characteristics with those of potential destinations. It has been utilized by Liaw and Ledent (1988) and Liaw (1990) in studies which interacted **categorical** socio-economic variables with destination attributes. There is, however, no practical or theoretical reason that this cannot be done with continuous socio-economic variables (see Thill and Van de Vyvere, 1989). A variable of this type is defined as:

$$Z_{ijn}^{J} = X_{in} X_{j}^{J} \tag{23}$$

where X_{in} is the observed value on the socioeconomic variable for individual n at origin i, and X_j^J is the value of the destination attribute at elemental alternative j in aggregate J.

An interaction variable such as Z_{ijn}^J can be specified as generic in the context of one aggregate alternative since the same parameter will apply to all elemental alternatives within this aggregate. However it can be specified as alternative-specific across aggregates meaning that elemental alternatives in different aggregates will be associated with different parameters. By allowing this alternative-specific specification of a generic interaction term, one captures essentially the same effects as would be obtained by specifying the socio-economic variable alone in an alternative-specific manner at the aggregate level.

The usefulness of this interaction form, however, is that different choice-makers can be associated with different levels of utility heterogeneity within aggregates. Given a single aggregate and with all things equal, the choice-maker in the sample with the minimum value of X_{in} will necessarily be associated with less heterogeneous values of Z^J_{ijn} in 23, and therefore V^J_{ijn}, than the choice-maker with the largest value. Note that it might be necessary to transform the socio-economic variable in some circumstances since choice-makers associated with the smallest value, not the largest, might perceive the greatest amount of utility heterogeneity.

4.5 Simulations on the Effect of Spatial Utility Heterogeneity

4.5.1 Description of the Simulation Method

The purpose of this section is to illustrate through simulations the model specifications just discussed and to demonstrate that as the heterogeneity of aggregate choice units increases, so does the potential bias of parameters estimated through the OMNL. In parallel, it will be seen that HETRO is not subject to similar problems. To address these points, a Monte Carlo procedure is utilized. It should become evident that the mean utility of an aggregate, on which the OMNL bases its predictions, can become a minor determining factor in how decision-makers select aggregates.

There are four types of simulations under study which can be classified on the basis of two criteria. The **first** is whether the variables are associated with a micro-data model or a model in which choice-makers are aggregated. The simulations with aggregate data are generated with a distance variable and a destination attribute as is typical in such a model of migration. The simulations with micro-data are generated on the basis of an interaction between a socio-economic individual characteristic, unique to each choice-maker, and a destination attribute. The **second** classification criterion is based on whether the alternative whose utility variability is altered has the highest level of mean utility or not. It turns out that the biases in OMNL differ if a heterogeneous aggregate has **lower** mean utility than other aggregates as compared to situations where the heterogeneous aggregate has the **highest** mean utility among all aggregates.

Some other general points to be noted about the simulations are as follows:

- A bottom-up approach is used. That is, we generate true utilities at the **elemental** level and assume that the actual utility of an aggregate for that decision-maker is given by the largest of the contained elemental utilities. It is important to note that the aggregated model is not directly used to generate the true utilities at the aggregate level.

- The effects of mean utility and aggregate alternative size are controlled by maintaining the same set of mean utilities for all levels of heterogeneity and by assigning the same number of elemental alternatives (four) to each aggregate.

- It is assumed that the ratio μ/μ^J is exogenously set to one for all aggregates. Hence the random error components of the elemental utilities within and between aggregates are independently generated and therefore uncorrelated.

- It is assumed that the attribute value for an aggregate alternative is simply the mean of the constituent elemental attributes.

- Note that even though we generate data which are 'aggregate' across decision-makers in the sense that systematic utilities are not differentiated on the basis of socio-economic variables, the true utilities which determine the choice shares are generated one decision-maker at a time.

A precise description of the data generation procedure is as follows:

- 1) A data set of attribute values is generated for a problem consisting of three aggregate alternatives each containing four elemental alternatives within. The data set is assumed to contain 50 choice-makers positioned at five different origins. None of the five corresponds to any potential destination. The data consist of a destination attribute, a socio-economic individual variable, the interaction between the two and a measure of distance to the given potential destination. The destination attribute values, regardless of the prevailing level of heterogeneity, are adjusted to conform with predefined mean attribute values in the aggregates.

- 2) For the appropriate specification, systematic elemental utilities are defined as a linear combination of the attribute values and **known** parameters.

- 3) Actual utilities for a given choice-maker, are obtained with the addition to the systematic utilities of a Gumbel distributed error component having variance parameter $\mu = 1$ and hence a variance of $\pi^2/6$.

- 4) In accordance with theory, the chosen elemental alternative for a given individual is taken as the one which provides the maximum utility of all elemental alternatives in the system. Obviously, this defines the chosen aggregate alternative also.

- 5) Steps 3 and 4 are repeated N times to correspond with the number of choice-makers in a given trial. The observed aggregate alternative choice behaviour and the known variable and attribute values provide the necessary information to estimate the unknown parameters of the two model types (OMNL and HETRO) for that trial.

- 6) Steps 3, 4 and 5 are repeated appropriately to generate choice data for 100 trials at a given level of heterogeneity. Preliminary work revealed that little was to be gained by employing a larger number of trials.

- 7) In order to alter the utility variance within a single aggregate in relation to the other two aggregates, the elemental destination attribute values in that aggregate are multiplied by some common factor and the result translated to maintain the same mean utilities across aggregates. The appropriate variances are obtained through the identity $Var(ax) = a^2Var(x)$ where $Var(x)$ is the initial variance of the destination attribute in the aggregate alternative and a is the necessary constant used to obtain the new variance. Multiplying the destination attribute by some factor also affects the interaction term. Steps 1 to 6 are then completed for the new level of heterogeneity.

- 8) For each new level of heterogeneity (the variances tested ranged from 2 to 50) or new specification, step 7 is repeated.

4.5.2 Results

4.5.2.1 Case #1: Aggregate data specification; aggregate with variable heterogeneity does not have largest mean utility

In this case, data are aggregated across choice-makers with the exception of distance which introduces some variability into systematic utilities. The known parameter associated with distance is -0.1 and the parameter for the destination attribute is 0.3. The latter implies that spatial units with high values on that attribute are perceived to have more utility than those with low values. The variance levels for the destination attribute in aggregate 1 range from 2, which is smaller than the corresponding variances in the other aggregates, to 50 which makes aggregate 1 far more heterogeneous than all other aggregates with respect to this attribute. All values in Table 4.1 represent mean results from the 100 trials on each model and at each level of variance. For a given variance level, the top row contains results for HETRO while the bottom row has results for OMNL. The average parameter estimates, t-scores, predicted standard errors

Table 4.1: Aggregate data; Variable aggregate does not have highest mean utility

(VAR)		β_1	β_2	t_1	t_2	se_1	se_2	ρ^2	shares		
2	*HETRO*	0.35	-0.10	1.98	-2.48	0.17	0.04	0.12	16.9	23.1	9.9
	OMNL	0.29	-0.08	2.26	-1.85	0.12	0.04	0.11			
5	*HETRO*	0.31	-0.10	2.10	-2.42	0.14	0.04	0.11	18.5	21.5	9.9
	OMNL	0.26	-0.06	2.08	-1.58	0.12	0.04	0.09			
10	*HETRO*	0.31	-0.11	2.85	-2.59	0.11	0.04	0.13	20.8	20.1	9.1
	OMNL	0.26	-0.06	2.09	-1.39	0.12	0.04	0.08			
15	*HETRO*	0.29	-0.11	3.38	-2.54	0.09	0.04	0.13	22.4	18.8	8.8
	OMNL	0.25	-0.04	1.99	-1.08	0.12	0.04	0.07			
25	*HETRO*	0.30	-0.09	4.67	-2.18	0.06	0.04	0.17	26.8	16.0	7.2
	OMNL	0.26	-0.00	2.03	-0.04	0.13	0.04	0.06			
50	*HETRO*	0.31	-0.11	6.54	-2.06	0.05	0.05	0.33	34.7	10.4	4.9
	OMNL	0.26	0.06	1.90	1.50	0.13	0.04	0.07			

Table 4. 2: Aggregate data; Variable aggregate has highest mean utility

(VAR)		β_1	β_2	t_1	t.	se_1	se_2	ρ^2	shares		
2	*HETRO*	0.30	-0.10	2.52	-2.39	0.11	0.04	0.13	23.0	18.5	8.4
	OMNL	0.36	-0.08	2.88	-1.98	0.12	0.04	0.12			
5	*HETRO*	0.29	-0.10	3.07	-2.27	0.09	0.04	0.15	25.0	16.9	8.1
	OMNL	0.39	-0.07	2.97	-1.76	0.13	0.04	0.13			
10	*HETRO*	0.30	-0.10	4.04	-2.33	0.07	0.04	0.19	27.3	15.8	6.9
	OMNL	0.46	-0.07	3.28	-1.65	0.14	0.04	0.16			
15	*HETRO*	0.31	-0.10	4.85	-2.22	0.06	0.05	0.23	30.1	13.6	6.3
	OMNL	0.51	-0.05	3.41	-1.26	0.15	0.04	0.18			
25	*HETRO*	0.31	-0.11	5.83	-2.17	0.05	0.05	0.31	33.6	11.5	4.8
	OMNL	0.64	-0.04	3.65	-0.98	0.17	0.04	0.24			
50	*HETRO*	0.31	-0.10	6.70	-1.48	0.05	0.06	0.48	40.3	6.3	3.4
	OMNL	0.91	0.01	3.64	0.15	0.25	0.05	0.37			

and rho-squared (ρ^2) goodness-of-fit statistic are displayed. The latter is calculated as:

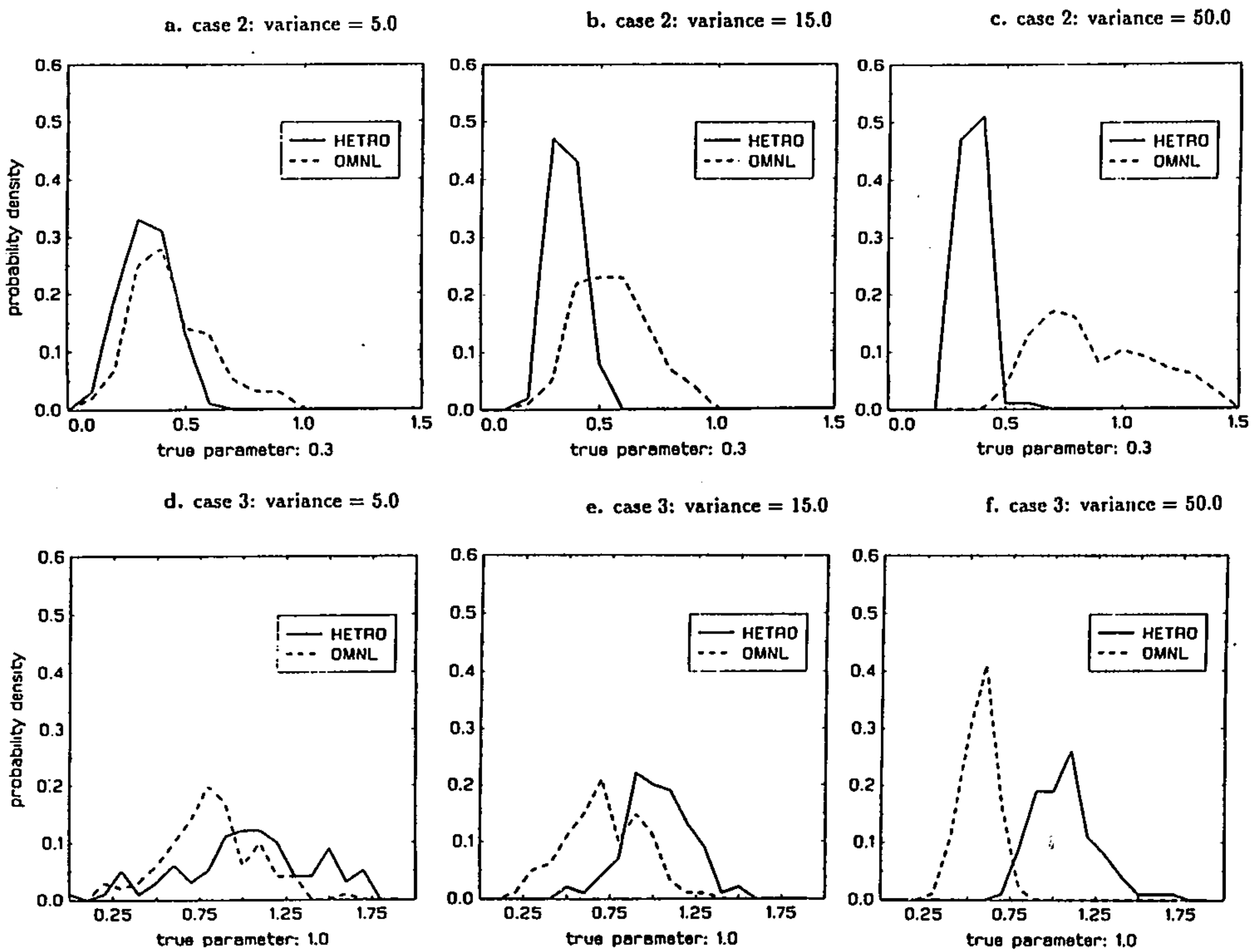
$$\rho^2 = 1 - \frac{\mathcal{L}^*(\beta)}{\mathcal{L}^*(0)}$$

where $\mathcal{L}^*(\beta)$ defines the value of the log-likelihood function at convergence and $\mathcal{L}^*(0)$ defines its value with a naive model of zero utilities. The column labeled 'share' represents the average number of choice-makers out of the group of fifty, who choose aggregate 1. The 'true' standard errors, which are given by the parameter standard deviations from 100 trials, are not displayed because they are almost identical, in this case, to the average standard errors derived from the variance-covariance matrix.

Initially, the first aggregate has slightly lower mean utility than aggregate 2 which is also more heterogeneous. As the level of variance in the destination attribute of aggregate 1 increases, with unchanging mean utilities, it becomes progressively more likely that alternative 1 is chosen as reflected in the share column of *Table 4.1*. It is more likely to contain an attractive elemental alternative with regard to the destination attribute and, therefore, also contain the point of maximum utility in the system. On a mean utility basis the other aggregates are more attractive than 1, but this fact becomes increasingly irrelevant to the utility maximizing choice-maker.

At all but the smallest level of variance, there is a consistent downward bias in the OMNL parameters. At higher levels of variance, the OMNL model is unable to reconcile the patterns implied by the mean utility with that of the choice shares. On a mean utility basis, it makes no sense that aggregate 1 is more often chosen, hence the parameters of variables used to predict the mean utilities become less significant. Even though there are no huge biases in parameters, the fact that OMNL does not include the corrective terms is very damaging to its goodness-of-fit. It is noteworthy also that while the changing variance affects only the destination attribute directly, with all distances being fixed, there are indirect effects which reduce the significance of the distance parameter. In addition, OMNL consistently

Figure 4.1: Simulation Results: Sampling Distributions



underestimates the significance of the two variables and, for the destination attribute, is generally less efficient in relation to HETRO.

4.5.2.2 Case #2: Aggregate data specification; aggregate with variable heterogeneity has largest mean utility

This simulation differs from Case #1 in that aggregate 1 has the largest mean utility with regard to the destination attribute. Hence, for all variance levels, aggregate 1 is more often chosen than the other alternatives. In *Table 4.2* we witness the opposite bias in the OMNL. With increasing variance, the OMNL destination attribute parameter begins to display a substantial upward bias. The sampling distributions for the destination attribute are displayed graphically in *Figure 4.1 (a-c)*. At higher variances, the extra choices which accrue to aggregate 1 are on the basis of an increased heterogeneity effect which in turn increases the likelihood that alternative 1 contains the point of maximum utility. From the perspective of OMNL, however, the prevailing choice shares indicate that the destination attribute under consideration is very important, implying that too much significance is associated with the corresponding parameter and therefore the impact of mean utility. HETRO properly separates the heterogeneity effect from the mean utility effect and consistently produces unbiased and efficient estimates of the two parameters. Apparently, OMNL so overestimates the effect of the destination attribute that the impact of distance is judged to be minimal, resulting in a severe mis-diagnosis of its significance. Again, HETRO results in better fitting models.

4.5.2.3 Case #3: Micro-data specification; aggregate with variable heterogeneity does not have largest mean utility

This scenario differs from those of Case #1 and #2 since the use of micro-data is assumed in which information relating to specific choice-makers is available. It is like Case #1 in that aggregate 1 has lower mean utility than aggregate 2. While destination attributes and distance variables could be included in the specification of *Table 4.3*, a one variable model in which a destination attribute interacts with a

Table 4.3: Micro-data; Variable aggregate does not have highest mean utility

(VAR)		β_1	t_1	se_{pred}	se_{true}	ρ^2	shares		
2	*HETRO*	1.11	2.29	0.46	0.49	0.10	20.6	22.0	7.4
	OMNL	0.36	2.66	0.32	0.33	0.09			
5	*HETRO*	1.05	2.68	0.38	0.43	0.11	23.8	19.3	6.9
	OMNL	0.76	2.46	0.30	0.26	0.07			
10	*HETRO*	1.05	4.13	0.25	0.25	0.17	27.7	16.6	5.7
	OMNL	0.73	2.39	0.30	0.21	0.06			
15	*HETRO*	0.98	4.90	0.20	0.19	0.21	30.4	14.2	5.3
	OMNL	0.67	2.24	0.30	0.22	0.06			
25	*HETRO*	1.02	6.19	0.16	0.16	0.35	35.9	10.2	3.9
	OMNL	0.59	2.03	0.29	0.16	0.04			
50	*HETRO*	1.02	6.53	0.16	0.18	0.57	41.9	5.6	2.5
	OMNL	0.52	1.85	0.28	0.09	0.03			

Table 4.4: Micro-data; Variable aggregate has highest mean utility

(VAR)		β_1	t_1	se_{pred}	se_{true}	ρ^2	shares		
2	*HETRO*	1.03	3.41	0.30	0.31	0.22	30.7	14.5	4.9
	OMNL	1.29	3.65	0.35	0.39	0.21			
5	*HETRO*	0.99	4.33	0.23	0.19	0.26	33.1	12.5	4.4
	OMNL	1.50	3.86	0.38	0.38	0.25			
10	*HETRO*	1.03	5.39	0.19	0.21	0.36	36.4	9.7	3.9
	OMNL	1.94	4.06	0.47	0.69	0.32			
15	*HETRO*	1.02	5.89	0.17	0.17	0.43	38.5	8.1	3.4
	OMNL	2.31	4.21	0.54	0.70	0.38			
25	*HETRO*	1.03	6.20	0.17	0.17	0.55	41.4	5.9	2.7
	OMNL	3.06	4.37	0.69	0.86	0.48			
50	*HETRO*	1.06	5.86	0.19	0.22	0.71	44.9	3.3	1.7
	OMNL	5.08	4.36	1.20	2.13	0.64			

socio-economic characteristic is utilized in order to focus on the behaviour of this type of variable. This interaction variable has a true parameter of 1.0. Therefore, a choice-maker who has a large value on the socio-economic variable will perceive more heterogeneity in intra-aggregate utility than someone with a small score. The positive true parameter implies that this interaction adds to utility, regardless of the level of the socio-economic variable, rather than acting as a disutility.

The results in *Table 4.3* indicate that OMNL provides biased estimates at any level of variance in aggregate 1 although the performance becomes progressively worse as this variance increases. The bias in the interaction parameter behaves qualitatively the same as did the destination attribute parameter in *Table 4.1*: when the choice shares are not explainable with the mean utilities, the OMNL parameter will be biased downward. In *Table 4.3* the true standard errors exhibited in 100 trials are compared with the average of the predicted standard errors because there are some substantial differences. In particular, the predicted standard errors of OMNL are consistently larger than the true ones indicating that there is a tendency to underestimate the significance of this interaction parameter. Interestingly, as can be seen in *Figure 4.1 (d-f)*, OMNL generally provides a more efficient estimate than HETRO albeit a substantially biased one. Nevertheless, HETRO by far produces superior fitting models[16].

4.5.2.4 Case #4: Micro-data specification; aggregate with variable heterogeneity has largest mean utility

This case employs the same specification as Case #3 but the mean utility in aggregate 1 is always larger than in the other two aggregates. As the 'share' column in *Table 4.4* indicates, the number of choices going to aggregate 1 is larger than is the case in *Table 4.3* because of the differing mean utility

[16]The fit of the OMNL would be improved if a full set of alternative-specific constants were included. Constants are intended to absorb all systematic variation not captured by variables in a model. Since this is a controlled environment, however, we know that there is no other systematic variation to capture. Hence there is no sound theoretical argument to be made for the inclusion of constants in the OMNL specification.

effects. At high levels of variance in aggregate 1, the biases in the OMNL estimate of the interaction parameter become quite drastic. Again, as was the case in Case #2, the bias is upward since OMNL is over-estimating the impact of the mean utility effect while the inflated standard errors are indicative of considerable inefficiency. Moreover, the predicted standard errors for OMNL are biased downward resulting generally in over-estimates of significance. In contrast to Case #3, where the biases were not so severe, the poor parameter estimates do not compromise the overall goodness-of-fit very much. It can be concluded that the general pattern of biases in the interaction parameter estimate is similar to the type occurring with a destination attribute. In comparison, the performance of HETRO is good in all aspects and at all levels of variance.

4.6 Conclusions

The main contribution of this study has been to illustrate the potential importance of a utility heterogeneity effect in properly characterizing the utility of an aggregate spatial unit. The notion that sub-aggregate attribute information should be applied when the available interaction data are at the aggregate level is a new perspective also. While it has long been known that there are size effects which need to be associated with aggregates, this paper represents an initial attempt at illustrating to regional scientists the usefulness of a heterogeneity term. The proposed theoretical framework also provides the customary size effect but one which does not emerge in the ad hoc manner common to many empirical studies. The theoretical principle under which these corrective terms are derived is simply that decision-makers seek the point within an aggregate which provides them the maximum utility. With the ordinary multinomial logit model, parameters are incorrectly estimated on the basis of some measure of *average* utility within aggregates rather than through *maximum* utility.

The simulations here have shown that this theory of choice among aggregates performs very well. The proposed model has excelled in a variety of circumstances whereas with the same data, the ordinary multinomial logit model suffers from potentially severe biases and inefficiencies, especially in a heterogeneous system. In particular, the parameters associated with destination attributes and the interactions of destination attributes with socio-economic variables are prone to downward bias when the most attractive aggregates on a mean utility basis are less heterogeneous than other aggregates. The parameters are biased upward when the heterogeneity of aggregates tends to complement the differences in mean utilities. Also, the goodness-of-fit of the ordinary multinomial logit model is often seriously compromised in relation to the proposed model.

Considerable work lies ahead in this area of research. Much of it relates to the estimation and interpretation of the ratio μ/μ^J which is an index of the degree to which elemental alternatives within some aggregate are perceived as similar and as such it can be described as a spatial autocorrelation parameter. The endogenous estimation of this ratio is a considerably more complicated optimization problem than that of the exogenous treatment applied in this paper (see Chapter 5). Ultimately, it will be necessary to examine the sampling distributions of μ/μ^J under different circumstances to assess its reliability in remaining within the theoretically defined $(0, 1]$ interval.

Chapter 5

The Aggregated Spatial Choice Model: A Comparative Analysis of Estimation Methodologies

5.1 Introduction

In spatial choice problems, destinations are often aggregations of elemental zones considered by decision-makers. For example, in a scenario of inter-regional migration, researchers may attempt to model movements between states or provinces while ignoring the fact that cities within regions are the actual units considered as potential destinations by migrants. Typically, employment of this approach results from the available choice data lacking sufficient spatial resolution. This problem can be resolved through application of the proposed *aggregated spatial logit model.* The superiority of this model over the Ordinary Multinomial Logit model (OMNL) or the gravity model stems from the fact that aggregate attractiveness is measured by an estimate of its internal maximum utility and not some measure of average utility. Hence the aggregated model, unlike the OMNL, is consistent with utility maximization theory and is an approach which acknowledges sub-aggregate variation in the attributes that determine utility. It is important to emphasize that the attributes used to explain a spatial choice process do not need to be at the same spatial aggregation level as the choice data. The aggregated spatial choice model exploits this fact while practitioners using the OMNL model have overlooked it.

Important theoretical issues surrounding the aggregated model have been covered elsewhere (Lerman, 1975; McFadden, 1978; Chapter 3). This chapter, however, focuses on the practical issue of its estimation which is decidedly more complex than is the case with the routinely applied OMNL model. The added complexity is due to parameters which determine the level of perceived independence of utilities at the sub-aggregate level. To some extent the past dominance of the OMNL and the basic gravity model in the spatial choice context is attributable to their relative simplicity and ease of application, with software being widely available for each. Also, these models have less onerous data requirements than the aggregated model which requires collection of attribute data and measurement of distances at the sub-aggregate level.

The outline of the chapter is as follows. The next section presents a brief overview of the theoretical background for the aggregated logit model, outlines the nature of the estimation problem, and compares it to estimation of the nested logit model with which it has similarities. The following section provides an examination of estimation techniques suitable for problems of unconstrained non-linear optimization while a final section evaluates these techniques in the context of the problem at hand using Monte Carlo simulations. A concluding section summarizes these results and their implications in assessing the best approach for estimating the parameters of an aggregated spatial logit model.

5.2 Background

5.2.1 The Theory of the Aggregated Model

The theory surrounding the aggregated spatial logit model has been discussed previously but to facilitate discussion in later sections, the main points are outlined here. Initially we focus on the conventional OMNL model before illustrating how the aggregated model is an important generalization suitable for problems with aggregate spatial units.

Consider that the study area is divided into L aggregate zones. Each aggregate zone $J \in \{1, 2, ...L\}$ is in turn divided into M^J elemental zones, with $\sum_{J=1}^{L} M^J = M$ being the total number of elemental zones. The latter are the theoretical units considered by decision-makers in their spatial selection process. For the OMNL model a decision-maker n originating in zone i is assumed to associate with aggregate J a utility level U_{in}^J. This utility, although deterministic from the decision-maker's viewpoint, is stochastic for the analyst. It is therefore written as the sum of a systematic component V_{in}^J and a random component ϵ_{in}^J: $U_{in}^J = V_{in}^J + \epsilon_{in}^J$. Under the assumptions that decision-makers will select the aggregate $J \in \{1, 2, ...L\}$ that is associated with the maximum U_{in}^J, and that the error terms ϵ_{in}^J, $J \in \{1, 2, ...L\}$ are identically Gumbel distributed with parameters $(0, \lambda)$, the OMNL is derived as:

$$P_n(J) = \frac{\exp\left[\frac{V_{in}^J}{\lambda}\right]}{\sum_{K=1}^{L} \exp\left[\frac{V_{in}^K}{\lambda}\right]} \tag{1}$$

where $P_n(J)$ is the probability that decision-maker n selects aggregate alternative J. Parameter λ is related to the variance of the error terms as $Var(\epsilon_{in}^J) = (\pi\lambda)^2/6$[17]. This parameter, which must be fixed prior to estimation, defines the scale of the utilities. In practice, it is set equal to the arbitrary value $\lambda = 1$. This implies $Var(\epsilon_{in}^J) = \pi^2/6$. The OMNL is hence characterized by error terms which are spatially homoscedastic.

In the case of the aggregated logit model, utilities are defined for each elemental alternative j within aggregate J as: $U_{ijn}^J = V_{ijn}^J + \epsilon_{in}^J + \epsilon_{ijn}^J$. The utility of an aggregate is represented by the maximum **elemental** utility within it: $U_{in}^J = \max_{j \in \{1,2,...M^J\}} U_{ijn}^J = \max_{j \in \{1,2,...M^J\}} (V_{ijn}^J + \epsilon_{ijn}^J) + \epsilon_{in}^J$. Thus, the deterministic utility V_{in}^J of the OMNL is replaced with the stochastic expression $\max_{j \in \{1,2,...M^J\}} (V_{ijn}^J + \epsilon_{ijn}^J)$.

[17] In Chapter 5, the term λ is utilized for the first time. It is simply the inverse of μ as defined in Chapter 3. The term λ^J is equivalent to $\frac{\mu}{\mu^J}$, also first used in Chapter 3, and is simply a more compact way of representing the ratio when μ is fixed to 1.0

The more general errors: $\epsilon_{in}^J + \epsilon_{ijn}^J$ for all $j \in \{1, 2, ...M^J\}$ and $J \in \{1, 2, ...L\}$, are assumed to be identically Gumbel distributed with parameters $(0, 1)$. As in the OMNL, this assumption implies $\lambda = 1$. Note that ϵ_{in}^J applies to all elemental alternatives in aggregate J and is assumed to be independent of the errors ϵ_{ijn}^J, $j \in \{1, 2, ...M^J\}$ associated with them. Thus, the variance of an elemental utility is $Var(\epsilon_{in}^J + \epsilon_{ijn}^J) = Var(\epsilon_{in}^J) + Var(\epsilon_{ijn}^J) = \pi^2/6$.

Within an aggregate J, error terms ϵ_{ijn}^J are independent and identically Gumbel distributed with parameters $(0, \lambda^J)$. This implies that $Var(\epsilon_{ijn}^J) = (\pi\lambda^J)^2/6$ and hence $Var(\epsilon_{in}^J) = \pi^2/6 - (\pi\lambda^J)^2/6$. Furthermore, the total utilities U_{ijn}^J of elemental alternatives within aggregate J share the same unobserved component ϵ_{in}^J and are not independent. One can show that for any two elemental alternatives j, k in J : $Cov(U_{ijn}^J, U_{ikn}^J) = Var(\epsilon_{in}^J)$. Elemental alternatives that belong to different aggregates are, however, independent. The correlation coefficient between elemental utilities that belong to the same aggregate can be identified as $corr(U_{ijn}^J, U_{ikn}^J) = 1 - (\lambda^J)^2$. If j and k belong to different aggregates then this correlation is zero.

Under these circumstances the aggregated model is expressed exactly as in equation 1. The distinction is that the systematic utility V_{in}^J of an aggregate J is now taken as a function of the systematic utilities of the elemental alternatives in it:

$$V_{in}^J = \lambda^J \ln \sum_{j=1}^{M^J} \exp[\frac{V_{ijn}^J}{\lambda^J}] \tag{2}$$

The difference between the two models can be seen more clearly if one rewrites 2 as:

$$V_{in}^J = \bar{V}_{in}^J + \lambda^J \ln \frac{1}{M^J} \sum_{j=1}^{M^J} \exp[\frac{V_{ijn}^J - \bar{V}_{in}^J}{\lambda^J}] + \lambda^J \ln M^J \tag{3}$$

where $\bar{V}_{in}^J = \frac{1}{M^J} \sum_{j=1}^{M^J} V_{ijn}^J$ is the mean systematic utility within aggregate J. Note that only in the absence of aggregation, with $M^J = 1$ for all $J \in \{1, 2, ...L\}$,does the expression for V_{in}^J reduce to the mean utility $\bar{V}_{in}^J$. In this case, V_{in}^J is equivalent to the systematic utility given by the OMNL.

In 3 it can be seen that accounting for aggregation introduces two additional terms into the systematic utility of an aggregate. The second term on the RHS of 3 is non-negative and measures the heterogeneity or variability of systematic utility within an aggregate. Intuitively, a heterogeneous aggregate is more likely to contain the most attractive destination and hence should be assigned a higher utility, all things being equal. The third term on the RHS of equation 3 is also non-negative and measures the size of aggregate J in terms of the number of choice units. Predictably, larger aggregates are associated with a higher level of systematic utility although the extent to which this is true depends on λ^J .

The uncertainty parameters $\lambda^J, J \in \{1, 2, ...L\}$ are identifiable through estimation and are important because they directly determine $Var(\epsilon^J_{in})$ and $Var(\epsilon^J_{ijn})$ and have much to do with the relative magnitudes of the size and heterogeneity terms. The allocation of variance in utility between ϵ^J_{ijn} and ϵ^J_{in} is a direct measure of the level of correlation, or degree of perceived similarity, between any pair of elemental total utilities within an aggregate. In general, the utility of an aggregate is decreased if it contains highly correlated elemental utilities. Since by definition, we have that $Var(\epsilon^J_{in} + \epsilon^J_{ijn}) \geq Var(\epsilon^J_{ijn})$ and hence that $\pi^2/6 \geq (\pi\lambda^J)^2/6$, and since λ^J cannot be negative, as a scale parameter of a Gumbel, it must be that $0 < \lambda^J \leq 1$.

It is instructive to examine the two polar cases for λ^J. As $\lambda^J \rightarrow 0$, $corr(U^J_{ijn}, U^J_{ikn}) \rightarrow 1$, elemental uncertainty vanishes because $Var(\epsilon^J_{ijn}) \rightarrow 0$ and aggregate uncertainty attains its maximum value: $Var(\epsilon^J_{in}) \rightarrow \pi^2/6$. For aggregate J the expected level of utility corresponds **exactly** to the largest of the contained elemental systematic utilities. Since the identity of the aggregate's 'representative' elemental alternative is known with certainty, the size term in equation 3 vanishes as M^J is immaterial, while the heterogeneity term reaches its maximum. If $\lambda^J \rightarrow 1$, then for any pair of elemental alternatives j and k within aggregate J we have $corr(U^J_{ijn}, U^J_{ikn}) \rightarrow 0$. This is equivalent to the aggregate error ϵ^J_{in} vanishing, while the elemental variation in utility becomes large: $Var(\epsilon^J_{ijn}) \rightarrow \pi^2/6$. Other things being equal, the size term in this case attains its highest value and the heterogeneity term its lowest. The former occurs because with ϵ_{ijn}^{J} at its most unpredictable, many of the M^J elemental alternatives could provide the true maximum utility. As a result, the systematic utility of the aggregate is upgraded.

The differences between the widely used OMNL model and the aggregated logit model are not only theoretical. Specification of V_{ijn}^{J} obviously requires collection of independent variables at the elemental level. While the aggregated model is thus more data-intensive than the OMNL, it makes use of micro-geographical data that the OMNL ignores. Also, in applications of the two models, identical aggregate choice information and model specifications will, in general, produce different parameter estimates, different relative levels of systematic utility V_{in}^{J}, and different predicted choice probabilities. In addition, the aggregated model permits estimation of elemental utilities V_{ijn}^{J} and elemental choice probabilities. More importantly, estimation of the aggregated model will yield estimates for λ^J, $J \in \{1, 2, ...L\}$, which as discussed in this section are parameters of primary importance for the process under study. The price to pay for these benefits is the need for a more sophisticated estimation procedure.

5.2.2 The Aggregated Model Versus the Nested Logit Model

While we have highlighted the superiority of the aggregated model over the OMNL for problems with aggregate spatial units, the actual estimation of the aggregated model, in terms of difficulty, is often more closely related to full-information estimation of the nested logit model. To undertake a full application of the nested logit model, we would need choice information at the elemental level since the nested approach is intended to model elemental choices. When such data are not available, the aggregated logit model must be used. Hence, in this section we compare estimation of the nested logit to that of the aggregated model.

As is the case with other discrete choice models, parameters for the aggregated model are obtained by maximizing the log-likelihood function:

$$\mathcal{L}^*(\beta, \lambda) = \sum_{n=1}^{N} \sum_{J=1}^{L} Y_{in}^{J} \ln P_n(J) \tag{4}$$

where:

- β is a vector of parameters that enter the systematic utilities V^J_{ijn}, $j \in \{1, 2, ..., M^J\}$, $J \in \{1, 2, ..., L\}$
- λ is a vector of parameters representing $(\lambda^1, \lambda^2, ..., \lambda^L)$
- Y^J_{in} is a binary dummy variable which is 1 if decision-maker n selects aggregate J and 0 otherwise.

The corresponding function for the two level nested logit model is dependent on the same variables:

$$\begin{aligned} \mathcal{L}^*(\beta, \lambda) &= \sum_{n=1}^{N} \sum_{J=1}^{L} \sum_{j=1}^{M^J} Y^J_{ijn} \ln P^J_n(j) \\ &= \sum_{n=1}^{N} \sum_{J=1}^{L} Y^J_{in} \ln P_n(J) + \sum_{n=1}^{N} \sum_{J=1}^{L} \sum_{j=1}^{M^J} Y^J_{ijn} \ln P_n(j \mid J) \end{aligned} \quad (5)$$

where the newly introduced elements are:

- $P^J_n(j)$, the overall probability that elemental alternative j in aggregate J is chosen.
- $P_n(j \mid J)$, the conditional choice probability of selecting j in J, provided J has been selected
- Y^J_{ijn} which is 1 for the single elemental alternative in the study area that is chosen and 0 otherwise. Because $\sum_{j=1}^{M^J} Y^J_{ijn} = Y^J_{in}$, the occurrence of $Y^J_{ijn} = 1$ defines the aggregate for which $Y^J_{in} = 1$.

The first expression in 5 illustrates the ultimate objective of the nested logit: prediction of elemental probabilities. Since $P^J_n(j) = P_n(J)P_n(j/J)$, the log-likelihood function can be split into two components to allow meaningful comparison with the aggregated model. Comparing 5 with 4 we observe that the nested logit shares the marginal component of the log-likelihood function with the aggregated model but has an additional component given by the conditional model. The conditional segment of course cannot be implemented in 4 because Y^J_{ikn} is unknown when the interaction data are at the aggregate level.

The expression for $P_n(J)$, given in 1 and 2 for the aggregated model, also applies to the nested logit model although specific results will differ since parameter estimation will clearly be affected by the presence of the conditional model. The complete set of conditional probabilities is:

$$P_n(j \mid J) = \frac{\exp\left[\frac{V^J_{ijn}}{\lambda^J}\right]}{\sum_{q=1}^{M^J} \exp\left[\frac{V^J_{iqn}}{\lambda^J}\right]}, \qquad j \in \{1, 2, ..., M^J\}; J \in \{1, 2, ..., L\} \tag{6}$$

Although the aggregated model does not actively employ $P_n(j \mid J)$ in the optimization, alternative estimates of these probabilities can be obtained after estimation using the elemental systematic utilities.

A prominent issue affecting the estimation of both models is the manner in which λ are specified. The importance of this issue for the aggregated model is reflected in *Figures 5.1* and *5.2* where graphical representations of the log-likelihood surface are shown for a a single sample problem consisting of fictitious choices among two aggregate spatial units. The two aggregates contain 3 and 6 elemental alternatives respectively. The same choice data were used to generate each of the log-likelihood surfaces with distance affecting the systematic utilities of all spatial units as a generic independent variable.

Figure 5.1 depicts a log-likelihood surface for which λ of the two alternatives are constrained to equality implying that the function is dependent on two parameters: β and λ. The greater concavity of the surface with respect to the friction-of-distance parameter (β) is consistent with the much smaller standard error obtained for this parameter in the estimation of the model. As is expected, there is a pronounced discontinuity on the surface in the region where $\lambda \to 0$ and in fact, the surface is not defined in this region. Note however that the surface is defined for values of λ well outside the unit interval both in terms of values larger than 1.0 and less than 0.0. Overall, obtaining the optimal values of $\beta = -0.39$ and $\lambda = 1.31$ for this problem is not difficult providing the algorithm is not started near the discontinuity. In our experience, this conclusion applies in general to applications of the aggregated model with a single λ, irrespective of the number of elements in β.

Figure 5.2 illustrates a conditional surface where λ^1 and λ^2 are unequal and β is fixed at -0.35. Hence the data are the same as in *Figure 5.1* but the specification differs. The optimal result for this model

FIGURE 5.1: AGGREGATED MODEL WITH CONSTRAINED λ

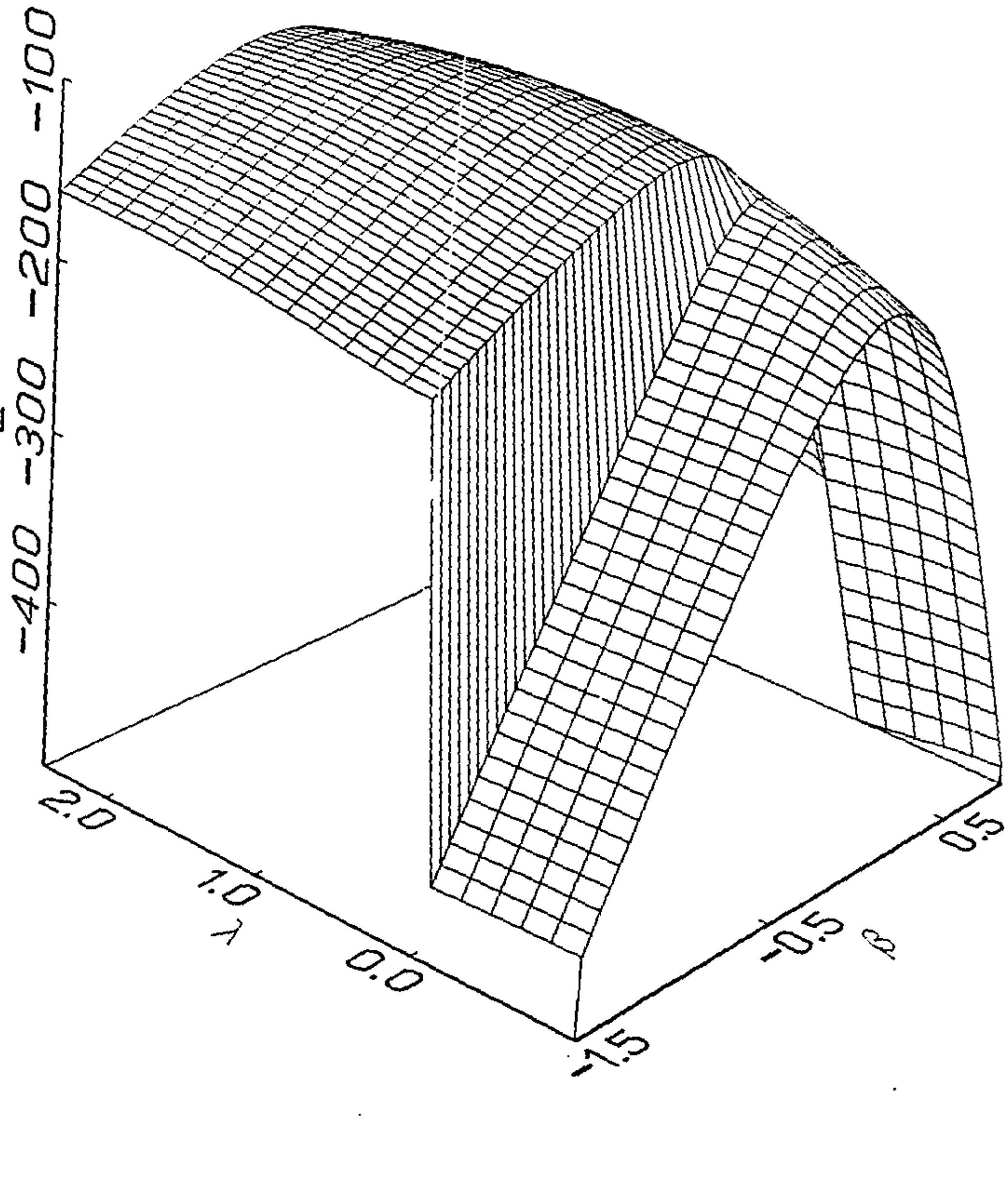
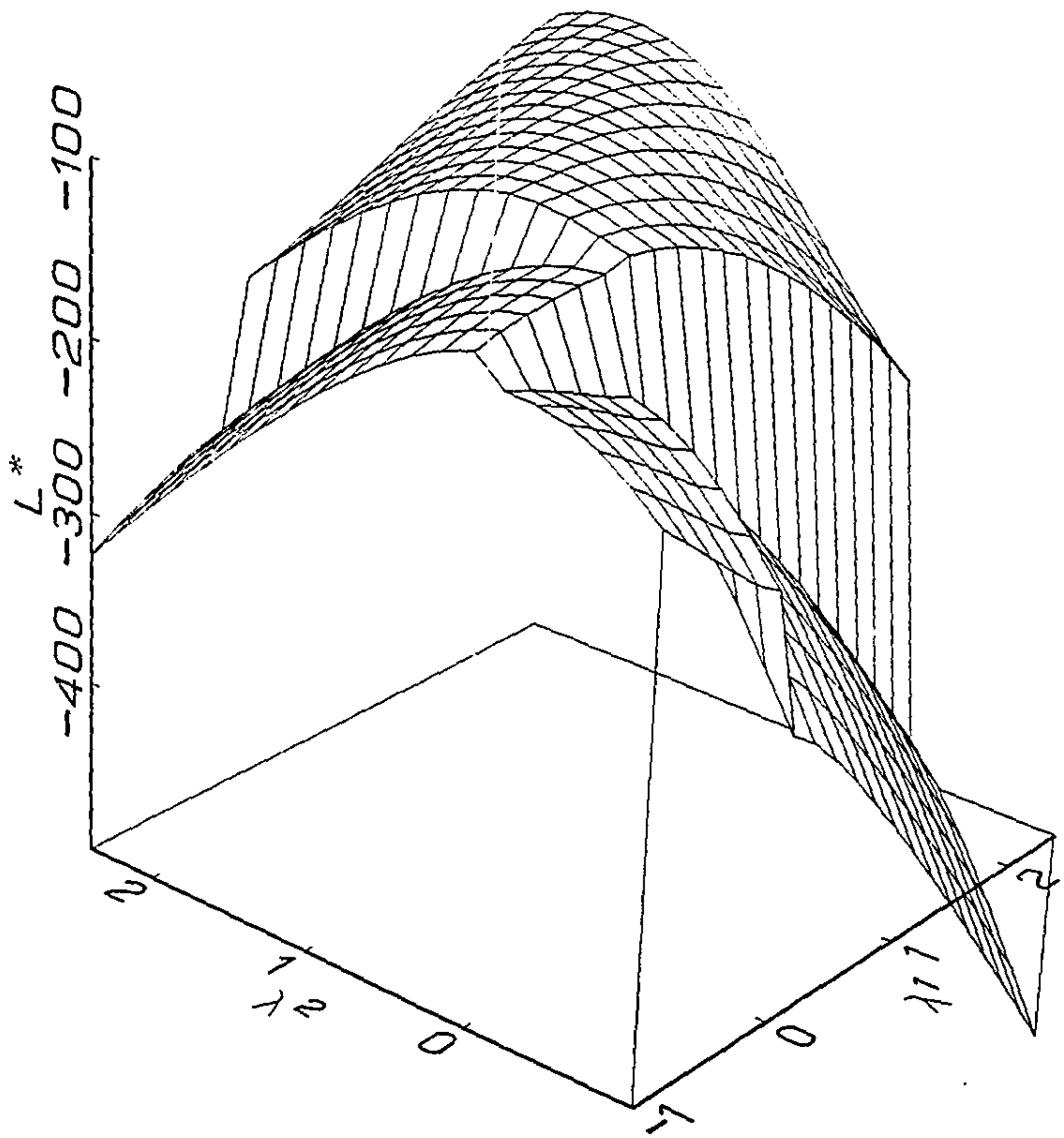


FIGURE 5.2: AGGREGATED MODEL WITH UNCONSTRAINED λ



is $\lambda^1 = 0.79$ and $\lambda^2 = 1.02$. It is not difficult to see that estimation of this model is more complex since the surface now contains discontinuities along two dimensions. In the following section, we describe optimization algorithms available to find the maximum on such surfaces. That discussion will be useful in generating insights to further assess *Figures 5.1* and *5.2*.

5.3 Estimation Techniques

5.3.1 Optimization Algorithms in General

The common theme among the different unconstrained optimization methods is that any twice differentiable function should be reasonably well-approximated in the vicinity of an optimum by a quadratic function. If this is true, then convergence is likely. Each of the approaches discussed below possess the quadratic termination property which imply that a quadratic function in k variables will converge to the optimal solution in no more than k steps (Avriel, 1976). Unfortunately, the aggregated logit model and the nested logit model do not result in a likelihood function that is quadratic, implying that quadratic termination cannot hold. A viable alternative though is the approximation of the likelihood function by a quadratic through the use of a second order Taylor series expansion.

Assume that the expansion takes place about the point defined by the current vector of parameter estimates β_t which we take to include the parameters associated with explanatory variables as well as the uncertainty parameters. For β_{t+1} in the vicinity of β_t :

$$\mathcal{L}^*(\beta_{t+1}) \simeq \mathcal{L}^*(\beta_t) + (\beta_{t+1} - \beta_t)' \frac{\partial \mathcal{L}^*(\beta_t)}{\partial \beta_t} + \frac{1}{2}(\beta_{t+1} - \beta_t)' \frac{\partial^2 \mathcal{L}^*(\beta_t)}{(\partial \beta_t)^2}(\beta_{t+1} - \beta_t) \qquad (7)$$

The objective is to find the values of the unknown parameters such that:

$$\frac{\partial \mathcal{L}^*(\beta_{t+1})}{\partial \beta_{t+1}} \simeq \frac{\partial \mathcal{L}^*(\beta_t)}{\partial \beta_t} + \frac{\partial^2 \mathcal{L}^*(\beta_t)}{(\partial \beta_t)^2}(\beta_{t+1} - \beta_t) = 0 \tag{8}$$

which is simply a linear first order Taylor approximation of the true gradient at the point β_{t+1}. Solving for β_{t+1}:

$$\beta_{t+1} = \beta_t - \left[\frac{\partial^2 \mathcal{L}^*(\beta_t)}{(\partial \beta_t)^2}\right]^{-1} \frac{\partial \mathcal{L}^*(\beta_t)}{\partial \beta_t}$$

which for notational simplicity can be written:

$$\beta_{t+1} = \beta_t - [\mathbf{H}_t]^{-1}\, \mathbf{q}_t \tag{9}$$

where β_t is a vector of k parameters, $\mathbf{H}_t$ is the $k \times k$ Hessian and $\mathbf{q}_t$ is the $k \times 1$ gradient vector. The step in 9 defines the well known Newton-Raphson method in which both the Hessian and gradient vectors are recalculated at every iteration t. This technique has gained considerable prominence in discrete choice modelling since it is straightforward and is extremely reliable in the context of the OMNL likelihood function which McFadden (1974) has demonstrated to be globally concave. However, the Newton-Raphson method is considerably less effective in problems which do not share this property. An improvement is the introduction of some scalar ξ to find the maximum improvement in the likelihood function along the direction defined by $[\mathbf{H}_t]^{-1}\, \mathbf{q}_t$. The result is that 9 is written:

$$\beta_{t+1} = \beta_t - \xi_t\, [\mathbf{H}_t]^{-1}\, \mathbf{q}_t \tag{10}$$

With or without the line search invoked by ξ, the Newton-Raphson method is recognized as being one of the most demanding algorithms computationally. Furthermore, it requires that the starting solution be in the vicinity of the optimum (Greene, 1990). Hence, this algorithm would not be the first choice for the problem in this paper.

The gradient methods, of which Newton-Raphson is a special case, are defined when $\mathbf{d}_t = [\mathbf{Q}_t]^{-1}\, \mathbf{q}_t$ represents the direction matrix, where the form of $\mathbf{Q}_t$ depends on the particular technique. The simplest form is the method of steepest ascent in which $\mathbf{Q_t} = \mathbf{I}$. A useful attribute of this method is that the optimal line search for each iteration is known beforehand as: $\xi_t = \frac{-\mathbf{q}_t'\mathbf{q}_t}{\mathbf{q}_t'\mathbf{H}_t\mathbf{q}_t}$. The algorithm will tend to perform better than Newton-Raphson at points far away from the optimum on a surface which is not globally concave but will converge inefficiently in the region of the optimum since it does not possess the quadratic termination property. Overall, this method would rank as less attractive than the Newton-Raphson algorithm for the problem in this paper although it might be useful for an initial set of iterations preceding a switch to some other method.

5.3.2 Methods Using Conjugate Search Directions

A critical matter in algorithms more refined than Newton-Raphson is the concept of mutually conjugate directions. Assume that we seek to optimize a quadratic function in k dimensions (or in the case of aggregated logit model some form of quadratic approximation to the true log-likelihood function). Any quadratic function can be written in the simple matrix form:

$$f(\mathbf{x}) = a + \mathbf{b}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}$$

where $\mathbf{A}$ is positive definite and symmetric. If two vectors $\mathbf{y}$ and $\mathbf{z}$ are found such that $\mathbf{y}'\mathbf{A}\mathbf{z} = 0$, it can be said that $\mathbf{y}$ and $\mathbf{z}$ are mutually conjugate with respect to $\mathbf{A}$. In other words, $\mathbf{y}$ is orthogonal to $\mathbf{A}\mathbf{z}$ which is the more general case of the well-known situation where $\mathbf{A} = \mathbf{I}$ and $\mathbf{y}$ is orthogonal to $\mathbf{z}$ itself. This is a matter of great importance since any unconstrained optimization algorithm which uses mutually conjugate directions of ascent in relation to $\mathbf{A}$ is quadratically convergent (Wismer and Chattergy. 1978) implying that the optimum for a quadratic function will be found in no more than k steps. Once k mutually conjugate vectors are established, and if the function is quadratic, it is theoretically possible to find the optimal point through a sequence of line searches along these vectors and then taking the best result.

The concept of conjugate directions manifests itself directly in a family of algorithms. Powell's

(1964) method uses properties of conjugate vectors to develop a means by which conjugate directions are established without need for any derivatives and where repeated line searches ultimately yield the optimum. The conjugate gradient method accomplishes the same thing as Powell's method in a manner which requires the calculation only of functional values and gradients and hence is well-suited to automated implementation. In particular, the Polak-Ribiere conjugate gradient method works on the basis of generating successive mutually conjugate search directions until convergence occurs, something which is likely to take more than k iterations when the actual function is not quadratic. The algorithm is such that the direction vector at each iteration is defined as: $\mathbf{d}_{t+1} = -\mathbf{q}_t + \mathbf{B}_k\mathbf{d}_t$ where $\mathbf{B}_k = \frac{(\mathbf{q}_t - \mathbf{q}_{t-1})'\mathbf{q}_t}{\mathbf{q}'_{k-1}\mathbf{q}_{k-1}}$. With the direction at each iteration t defined in this manner, a line search is carried out in this direction to find the largest possible increase in the function. Information from previous iterations is a necessity in establishing a direction which is mutually conjugate to those previous. Overall, this method has obvious computational advantages over Newton's method because the obstacle of calculating and storing a Hessian matrix is removed. While the conjugate gradient methods perform better than the steepest ascent method, they are most likely less effective than the family of Quasi-Newton methods to be outlined now.

5.3.3 The Quasi-Newton Methods

A Quasi-Newton or variable-metric method is also characterized by the search direction $\mathbf{d}_t = [\mathbf{Q}_t]^{-1}\mathbf{q}_t$ where $\mathbf{Q}_t$ is a $k \times k$ matrix which varies at each iteration and which may not be negative definite or symmetric, especially in early iterations (Avriel, 1976). Typically, $\mathbf{Q}_0$ might be set to the identity matrix so that the first step would mimic that of a steepest ascent iteration. With repeated steps an updating formula is applied which should eventually approach the Hessian matrix near the optimum. Hence, at a point near convergence, a Quasi-Newton method should behave in a manner similar to the Newton-Raphson algorithm since both are essentially using the Hessian to calculate the direction. Quasi-Newton methods, however, operate on the basis of defining directions which are mutually conjugate and hence are less likely to be led astray in regions far from an optimum unlike the Newton-Raphson method.

The Quasi-Newton method known as Broyden-Fletcher-Goldfarb-Shanno (BFGS) is examined in this paper and is generally recognized to be one of the best performers. It is characterized by the updating formula:

$$\mathbf{Q_k} = \mathbf{Q_{k-1}} + \left(1 + \frac{\gamma_t' \mathbf{Q_{k-1}} \gamma_k}{\mathbf{p_k'} \gamma_k}\right) \frac{\mathbf{p_k p_k'}}{\mathbf{p_k'} \gamma_k} - \frac{\mathbf{p_k} \gamma_k' \mathbf{Q_{k-1}}}{\mathbf{p_k'} \gamma_k} - \frac{\mathbf{Q_{k-1}} \gamma_k \mathbf{p_k'}}{\mathbf{p_k'} \gamma_k}$$

where $\mathbf{p_k} = \beta_k - \beta_{k-1}$ and $\gamma_k = q_k - q_{k-1}$. The derivations of the updating formulae in the approximation of the Hessian are quite complex. Moreover, they have many variations, although Huang (1970) developed a general expression which encompasses all variable-metric possibilities as special cases. Such involved discussion is, however, beyond the scope of this paper. It is enough to appreciate that the Quasi-Newton methods incorporate the strong points of the Newton-Raphson and conjugate directions approach and are generally the most powerful search algorithms available. While it is true that they may take many iterations to solve non-quadratic problems, the computational burden of each iteration is much less than that of a single Newton-Raphson step.

5.3.4 Past Optimization of the Nested Logit Model

Different gradient methods have been used in full-information maximum likelihood (FIML) estimation of the nested logit model. The Newton-Raphson algorithm is out of favour since there is the possibility of a Hessian which is not negative definite. One alternative route is the quadratic hill climbing method of Goldfeldt et al (1966) in which a transformed Hessian is utilized: $\mathbf{H}_\alpha = \mathbf{H} - \alpha\mathbf{I}$, where α is computed sufficiently large to ensure that $\mathbf{H}_\alpha$ is negative definite. Avriel (1976) likens this method to restricting the search within a circular region in which the function is guaranteed to be concave.

The most popular method to ensure a negative definite Hessian replacement, at least in the context of the nested logit model (Brownstone and Small, 1989), is to utilize the negative expectation of the outer product of the gradient vector across decision-makers, which assumes the form:

$$\mathbf{Q}_t = \sum_{n=1} \sum_{j=1} P_n^J(j) \left(\frac{\partial \ln P_n^J(j)}{\partial \beta_t} \right) \left(\frac{\partial \ln P_n^J(j)}{\partial \beta_t} \right)' \tag{11}$$

where $P_n^J(j)$ is the product of the marginal and conditional choice probabilities applying to each elemental alternative at the given parameter values. This approach is essentially a modified method of scoring which implies that an estimate of the information matrix is used in place of the Hessian. This is expedient since the information matrix is key to obtaining parameter standard errors. Cramer (1991) provides further discussion on the distinction between the method of scoring and the Newton-Raphson approach. The outer product form first proposed by Berndt et al. (1974) was designed for all maximum likelihood problems and has the more general form:

$$\mathbf{Q}_t = \sum_{n} \left(\frac{\partial \ln \mathcal{L}^*(\beta_t)}{\partial \beta_t} \right) \left(\frac{\partial \ln \mathcal{L}^*(\beta_t)}{\partial \beta_t} \right)' \tag{12}$$

in which no expectations have been taken. The outer product approach was first advocated for the nested logit by McFadden (1981) and has been the favored method for the nested logit. Daly (1987) further advocates its use for a version of the aggregated logit model which contains only mean utility and size terms. The more advanced Quasi-Newton methods have, to the best of our knowledge, not been applied for the nested logit or the aggregated model.

5.4 Estimation Trials

5.4.1 Preliminary Observations

Clearly, the important aspect affecting the performance of these optimization algorithms is the nature of the log-likelihood surface. Typically algorithms are dependent on a direct calculation or approximation of the Hessian matrix which in turn is directly affected by concavity characteristics of the surface.

When in regions which do not possess desirable concave characteristics, it is clear that the Hessian may be near-singular and hence difficult to invert. Of course, the ultimate result of inverting a near-singular matrix will be an algorithm which takes a long time to converge, if indeed it does ultimately succeed. Also, in succesful convergences, parameter correlations will tend to be high and estimated standard errors may be inflated. In short, estimation can be inefficient both from the algorithmic and statistical points of view.

Preliminary experimentation has shown that the aggregated model is potentially vulnerable to such problems and that the difficulties are closely related to the specification of the λ parameters[2]. In general, the aggregated model becomes easier to estimate as the number of endogenous λ are reduced either through equality constraints or by exogenous fixing. This is shown, for example, by the more complex nature of *Figure 5.2* in relation to *Figure 5.1*. For an explanation, note that the influence of λ^J are particularly strong in the size terms: $\lambda^J \ln M^J$, $J \in \{1, 2, ..., L\}$. Like all logit probabilities, those of the aggregated model are determined by differences in utility. Hence, estimation becomes more efficient if one or more size terms are fixed through λ in some manner so that the others can be determined in relation. Convergence is still possible if all λ are endogenous, however it is often the case that all λ can get uniformly much larger or smaller, often well outside the theoretical range, without inducing appreciable change in the log-likelihood function. This contributes to estimation inefficiency. Fixing or equating λ results in a surface with a more clearly identifiable maximum, and hence provides a clearly defined algorithm direction in order to obtain appropriate relative utilities. Also, the chances of a theoretically valid outcome are increased since λ tends to be more stable. These observations are confirmed in experiments undertaken and described below.

5.4.2 Experiment Setup and Rationale

In this section, using a simulated environment, we examine algorithm performance in estimating

[2]From now on in this chapter, λ is used to denote a scale parameter specific to an aggregate in the aggregated model instead of λ^J.

the aggregated logit model. The results are obtained using the GAUSS maximum likelihood module which implements a variety of optimization techniques. The module is activated in concert with a GAUSS shell which calculates the necessary functional values and gradients. Note, that the analytical gradients of the aggregated model are derived in an appendix.

Extensive preliminary experimentation has been done in estimations by varying the number of aggregates and elemental alternatives as well as the number of explanatory variables. These tests have revealed that problems with multiple parameters in both β and λ are manageable. The objective in this paper though, is to compare algorithms using the simplest experiment that captures the essence of the estimation problem.

In this experiment then, ten elemental alternatives are assumed, distributed as 3, 4 and 3 among three aggregates. A Monte Carlo procedure is applied in which choice realizations are generated on the basis of known parameters. Subsequently, the aggregated logit model is estimated using the generated realization and the known underlying parameters as a starting point for each estimation. Each realization consists of 1000 decision-makers, assumed to be departing from distinct points in space outside the three aggregate alternatives. The systematic utility of each decision-maker is a linear function of distance. Furthermore, distance is specified as a generic variable, implying that its marginal effect on utility is the same for all destinations.

While the primary theme of interest in this paper is identification of a superior algorithm for estimation of the aggregated model, there are important secondary themes. In particular, we examine the effectiveness of constraining uncertainty parameters to equality as well as the possibility of fixing some exogenously. The latter approach is similar in spirit to the specification of alternative-specific constants in which the constant associated with the reference alternative is arbitrarily fixed at zero while the remaining endogenous constants are estimated in relation. Another theme to be considered is the effectiveness of algorithm switching as it is anticipated that some algorithms, such as the method of steepest ascent and the Newton algorithm, may be complimentary.

5.4.3 Results

Table 5.1 presents a sampling of the conducted trials, summarizing the essential findings of this analysis. Each trial consists of 100 estimations. The central tendency of the distribution approximated by the 100 estimated parameter values is reported (column three), as well as the standard deviation (in parenthesis) as a measure of the standard error for all the parameters. The number of estimations out of the 100 that actually converged in less than 50 iterations are reported in column four. Note that a reasonably specified OMNL model converges in four to six iterations. Tests indicated that if 50 iterations are exceeded then the algorithm is searching far from the optimum with little hope of identifying the solution. The number of times out of the 100 that the 50 iterations limit was exceeded for each case is reported in column five. The following column six provides the number of times the algorithm crashed. Finally, for the cases of convergence in less than 50 iterations, the average number of iterations is presented in column seven.

Note in *Table 5.1* that the true values utilized for λ in the simulations are generally 1.0. Exceptions to this are experiments 2 and 9 to 12. Even for these the true λ are not less than 0.6. In general it was found that with smaller λ, it was difficult to run a Monte Carlo experiment since algorithm failure was frequent. The closer the true value of λ to 0, the greater the chance of failure. This is consistent with McFadden's (1981) observation relating to the nested logit model that the surface becomes highly non-linear as $\lambda \rightarrow 0$. A possible solution to the problem of algorithm failure is the estimation of $\mu = 1/\lambda$ in place of λ. This approach greatly reduces the chance of algorithm failure, since the nature of the log-likelihood surface is altered, however it has been unclear in experimentations that the convergence rates are improved. As a result, we focus on λ despite the problems which occur when the true value is small.

TABLE 5.1: ALGORITHM PERFORMANCE UNDER DIFFERENT SPECIFICATIONS

ALGORITHM	SPECIFICATION OF UNCERTAINTY	MEAN PARAMETERS AND STANDARD ERRORS						CONVERGENCE	EXCESSIVE ITERATIONS	FAILURE	MEAN ITERATIONS
		β	λ1	λ2	λ3	λ4	λ5				
1. BFGS	λ1≠ λ2≠ λ3	-0.21	1.02	1.01	0.93	-	-	76	22	2	20.09
	(1.0)(1.0)(1.0) a	(0.02)	(0.45)	(0.35)	(0.54)						
2. BFGS	λ1≠ λ2≠ λ3	-0.20	0.69	0.89	1.11	-	-	73	22	5	18.96
	(0.6)(0.8)(1.0)	(0.03)	(0.51)	(0.39)	(0.54)						
3. STEEP-NEWT	λ1≠ λ2≠ λ3	-0.20	1.28	1.22	1.26	-	-	45	37	18	35.42
	(1.0)(1.0)(1.0)	(0.02)	(0.89)	(0.71)	(0.92)						
4. BFGS-NEWT	λ1≠ λ2≠ λ3	-0.21	24.44	19.56	24.40	-	-	94	5	1	24.79
	(1.0)(1.0)(1.0)	(0.03)	(52.44)	(41.55)	(52.51)						
5. PRCG-NEWT	λ1≠ λ2≠ λ3	-0.20	1.70	1.55	1.71	-	-	44	35	21	36.68
	(1.0)(1.0)(1.0)	(0.02)	(1.58)	(1.25)	(1.63)						
6. BFGS	λ1≠ λ2; λ3=1 b	-0.20	0.98	0.99	-	-	-	100	0	0	8.20
	(1.0)(1.0)	(0.02)	(0.10)	(0.07)							
7. BFGS	λ1= λ2= λ3	-0.20	0.99	-	-	-	-	100	0	0	6.54
	(1.0)(1.0)(1.0)	(0.02)	(0.16)								
8. BFGS	λ1≠ λ2= λ3=1	-0.20	1.00	-	-	-	-	100	0	0	5.78
	(1.0)	(0.04)	(0.17)								
9. BFGS	λ1≠ λ2; λ3=1	-0.20	0.61	0.80	-	-	-	100	0	0	8.60
	(0.6)(0.8)	(0.02)	(0.12)	(0.07)							
10. BFGS	λ1= λ2= λ3	-0.23	0.67	-	-	-	-	100	0	0	7.01
	(0.6)(0.8)(1.0)	(0.02)	(0.16)								
11. BFGS	λ1≠ λ2≠ λ3≠ λ4= λ5	-0.20	0.82	1.02	1.25	1.23	-	90	3	7	25.91
	(0.6)(0.8)(1)(1)(1)	(0.02)	(0.52)	(0.42)	(0.47)	(0.43)					
12. BFGS	λ1≠ λ2≠ λ3≠ λ4≠ λ5	-0.20	0.77	0.99	1.16	1.17	1.16	87	5	8	32.67
	(0.6)(0.8)(1)(1)(1)	(0.02)	(0.65)	(0.58)	(0.62)	(0.56)	(0.59)				

a - numbers in parentheses represent true values associated with endogenously estimated λ
b- λ3, for example, is exogenously fixed to 1

5.4.3.1 Performance of Individual Algorithms

As *Table 5.1* readily reveals, there is no question that the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is the most successful in a stand alone fashion, providing the best combination of speed and stability. This judgment is made on the basis of trials 1 and 2 of *Table 1*. In these, the λ are not constrained to equality and a single λ is estimated for each aggregate. As mentioned earlier, such a model is more complicated to estimate than one with constraints on the λ and hence should act as a good test for algorithm effectiveness. While BFGS converged 76 and 73 per cent of the time under these scenarios, the performance of the other algorithms by themselves was very poor. In particular, the Newton-Raphson (NEWT), Polak-Ribiere conjugate gradient method (PRCG), and Berndt-Hall-Hall-Hausman (BHHH) technique failed to record any convergences within 50 iterations from their 100 trials. By themselves, NEWT, PRCG and BHHH seem to perform quite poorly in the early iterations and therefore are never in a good position to converge. BFGS though is quite flexible in the early iterations as it mimics the steepest ascent algorithm (STEEP) and then imitates the Newton-Raphson algorithm near the point of convergence. The superior performance of the BFGS algorithm makes the absence of the Quasi-Newton algorithms in the nested logit estimation literature (McFadden, 1981; Hensher, 1986; Brownstone and Small, 1989) surprising since these authors mention estimation difficulties. Equally surprising is the bad performance of the BHHH algorithm, which is promoted by this literature.

5.4.3.2 The Use of Algorithm Switching

While the Newton-Raphson method was quite unsuccessful in a stand-alone fashion and apparently had problems getting into the region of an optimum, it is reputed to be the fastest of the algorithms in actually reaching convergence when the optimum is nearby. As a result, it was paired in turn with STEEP, BFGS and PRCG in trials 3 to 5 to see if the overall rates of convergence could be improved with an algorithm switch to NEWT taking place after 30 iterations. The combinations of STEEP-NEWT and PRCG-NEWT converged some of the time but proved to be substantially less stable than other means. The best convergence rates are provided by BFGS-NEWT where the algorithm switch can lead to a better convergence rate than with BFGS alone. In general it was noted that if the NEWTON portion of the switched algorithm does not converge within a few iterations, it will tend to cycle infinitely in the same sub-optimal zone.

The comparatively poor performance of algorithms applied in experiments 1 to 5, where there is one endogenously determined λ for each aggregate zone, is in need of explanation since even the BFGS algorithm was frequently unsuccessful. A frequent symptom of such difficulties is extremely high correlations exhibited in the parameter correlation matrices between λ parameters. These reflect other problems relating to statistical inference, algorithm stability and violations of the model's theory. Certainly, the results for the standard errors and mean parameters in trial 4 dramatically illustrate potential effects on statistical inference. It is clear in this case that convergences are taking place in the context of a model with underlying specification problems. These results confirm earlier statements that estimation of λ tends to be more inefficient and less stable when there are no restrictions on these parameters. This appears the case even when the underlying true λ are unequal as is the case in trials 2 and 12. When restrictions on λ are imposed however, statistical efficiency and estimation success improves since the near-singularity is removed, the direction matrix has more desirable characteristics and aggregate utilities can efficiently be adjusted in relative terms.

5.4.3.3 The Use of Exogenous Parameters and Equality Constraints

The convergence performance in experiments 6 to 10 is the most successful. It is clear that estimation is greatly simplified in the case of the aggregated model, both in terms of speed and convergence proportions, by imposing equality constraints on the λ parameters or through fixing one or two of them exogenously. The BFGS algorithm was used for these trials mainly because it proved itself as generally the most successful. The performance of other algorithms however, such as NEWTON and PRCG also improves under constrained λ estimation.

The results for trial 10, in which one λ is estimated for all three aggregates, despite the fact that the true λ values are unequal, is an interesting case. Note that this is the only incidence where the parameter associated with distance has any substantial bias. What this result shows is that, while it may be expedient in an estimation sense to equate λ, the researcher must be cautious or biases will result. While difficulty in attaining a convergence is probably a reflection of a misspecification, there are cases, as trial 10 shows, where specification problems can be obscured by a relatively smooth convergence, with statistical efficiency being gained at the expense of introducing biases into estimation.

Clearly statistical tests are needed to identify cases where two or more parameters are not significantly different. Such tests can be based on the parameter variance-covariance matrix. Given potential estimation and statistical difficulties, it is important for a researcher to estimate a variety of exploratory models before resorting to a specific one. Such an approach is analogous to the tests developed for assessing whether logit model parameters should be specified generically (Ben-Akiva and Lerman, 1985) or in assessing the IIA property of the logit model (McFadden, Tye and Train, 1977) in which various models with different subsets of alternatives need to be estimated before diagnostics can be performed.

An important technique in the model exploration phase, is the fixing of individual λ exogenously, especially if estimation seems poorly behaved and if there is no basis for equating various λ. This action will stabilize estimation while the exogenous λ can be incrementally changed to determine which conditional optimization provides the maximum. Once in the vicinity of the optimum, a complete estimation stands a better chance of converging.

5.5 Conclusions

This paper has revealed that the aggregated logit model is amenable to current optimization algorithms although the log-likelihood surface may be less well-behaved than that of the nested logit model. In particular, the Quasi-Newton family of methods, notably the Broyden-Fletcher-Goldfarb-Shanno algorithm, represent powerful means to obtain parameter estimates under the most difficult circumstances. Other algorithms, such as the method of steepest ascent, can be effective as initial methods preceding an algorithm switch. In particular, the Newton-Raphson technique behaved well in obtaining final convergence once in the region of the optimum, and hence is often a good algorithm to which a switch can be made. This same algorithm, however, was found to be inadequate for the aggregated model in a stand-alone fashion. As far as the closely related nested logit model is concerned, it is thought that full-information estimation, itself considered a difficult problem, would likely benefit from some of the methods in this paper. In particular, the Quasi-Newton approaches have been overlooked in the literature.

Results indicate that specification of uncertainty parameters can have considerable bearing on whether an optimal solution is found. Also, difficulties in attaining convergence can be exacerbated by highly correlated parameters, often leading to predicted uncertainty parameters outside the theoretical range. Such problems can be combated by appropriately constraining uncertainty parameters to equality or specifying some exogenously. Overall, researchers must remain aware that difficulty in obtaining solutions may not be because the problem is intractable but rather that some aspect of the specification is poor. Given a well-specified model, the Quasi-Newton methods seem well-equipped to successfully carry out optimization.

5.6 Appendix to Chapter 5: Log-likelihood Derivatives for the Aggregated Model

Estimation techniques discussed in this paper require the first order derivatives of the aggregated logit log-likelihood function. Unless analytical formulas for the calculation of the derivatives are provided by the user, the GAUSS maximum likelihood module, for example, calculates derivatives numerically, a process that slows down estimation considerably. This appendix provides the derivative formulas used in this paper.

Equation 4 gives:

$$\frac{\partial \mathcal{L}^{*}(\beta,\mu)}{\partial \zeta} = \sum_{n=1}^{N}\sum_{J=1}^{L} \frac{Y_n^J}{P_n(J)} \frac{\partial P_n(J)}{\partial \zeta} \tag{13}$$

where ζ is any parameter in β or μ. For any logit probability formula, such as 1, it is true that:

$$\frac{\partial P_n(J)}{\partial \zeta} = P_n(J)\left[\frac{\partial V_{in}^J}{\partial \zeta} - \sum_{Q=1}^{L} \frac{\partial V_{in}^Q}{\partial \zeta} P_n(Q)\right] \tag{14}$$

A linear-in-parameters elemental systematic utility is given by:

$$V_{ijn}^J = \sum_{k=1}^{K} \beta_k X_{ijnk}^J \tag{15}$$

where X_{ijnk}^J is the k'th of K independent variables. For a parameter β_k in β equation 2 gives:

$$\frac{\partial V_{in}^J}{\partial \beta_k} = \frac{\sum_{j=1}^{M^J} X_{ijnk}^J \exp\left[\frac{V_{ijn}^J}{\lambda^J}\right]}{\sum_{j=1}^{M^J} \exp\left[\frac{V_{ijn}^J}{\lambda^J}\right]} \tag{16}$$

$$= \sum_{j=1}^{M^J} X_{ijnk}^J P_n(j \mid J)$$

which applies only to the parameters associated with the explanatory variables, not the uncertainty parameters . Substituting 16 into 14:

$$\frac{\partial P_n(J)}{\partial \beta_k} = P_n(J)\left[\sum_{j=1}^{M^J} X_{ijnk}^J P_n(j \mid J) - \sum_{Q=1}^{L}\sum_{j=1}^{M^Q} X_{ijnk}^J P_n(j \mid Q) P_n(Q)\right] \tag{17}$$

Similarly, for a λ^J in λ:

$$\begin{aligned}\frac{\partial V_{in}^J}{\partial \lambda^J} &= \ln \sum_{j=1}^{M^J} \exp\left[\frac{V_{ijn}}{\lambda^J}\right] - \frac{\frac{1}{\lambda^J}\sum_{j=1}^{M^J} V_{ijn}\exp\left[\frac{V_{ijn}}{\lambda^J}\right]}{\sum_{j=1}^{M^J}\exp\left[\frac{V_{ijn}}{\lambda^J}\right]} \\ &= \frac{1}{\lambda^J}\left[V_{in}^J - \sum_{j=1}^{M^J} V_{ijn} P_n(j \mid J)\right]\end{aligned} \tag{18}$$

from which the reader can gather $\frac{\partial P_n(J)}{\partial \lambda^J}$, by substituting in 14.

In the event that the Newton-Raphson method is being used, having the expressions for the gradient vector alone is not sufficient. It is necessary to calculate the Hessian matrix of second derivatives which is employed directly in the procedure unlike the Quasi-Newton methods in which it is approximated. The expression for the element of the Hessian matrix corresponding to parameters k and l is:

$$Q_{kl} = -\sum_{n=1}^{N}\sum_{J=1}^{L} \frac{1}{P_n(J)} \frac{\partial P_n(J)}{\zeta_k} \frac{\partial P_n(J)}{\zeta_l} \tag{19}$$

This equation is combined with 14 to calculate elements of the Hessian matrix.

Chapter 6

An Empirical Evaluation of the Aggregated Spatial Choice Model

6.1 Introduction

This chapter constitutes an initial attempt at empirically testing a spatial choice model designed for situations where the choice destinations in the available data are aggregates of the actual spatial units considered by choice-makers. For example, we may have data on the choice of regions made by migrants but need a model to acknowledge that the choices are typically motivated at the level of metropolitan areas within each region. This proposed approach is henceforth referred to as *the aggregated spatial choice model*, a technique which explicitly links spatially aggregate levels of utility to the spatially disaggregate utility components and which raises the possibility that disaggregate attribute data can be used in the same model as the aggregate choice data imposed on the researcher. Theoretically, the model is logically premised on the notion that the utility of an aggregate for a given choice-maker is represented by the contained elemental alternative of maximum utility. The issues examined in this paper differ from those considered by the modifiable areal unit problem (MAUP) and should not be confused. The MAUP evaluates statistical properties of spatially referenced attributes in the presence of aggregation (Arbia, 1989). Here we are dealing with a choice process as is also the case with the gravity model.

The theory of the aggregated spatial choice model is closely linked with that of the nested logit model with much of the seminal work on both being done by McFadden (1978) and Lerman (1975) in the context of intra-urban housing choice. The mathematical form of the aggregated model differs since, by definition, it cannot exploit the sub-aggregate choice information available for the nested logit. In applications with aggregate choice data, the most popular tool has been the Ordinary Multinomial Logit Model (OMNL) but it has the shortcoming of not representing sub-aggregate variation in utility (Chapter 3). This misrepresentation can lead to erroneous parameter estimates in heterogeneous aggregates, particularly if a corrective set of alternative-specific constants is not employed. Fortunately, the aggregated model is more general in this respect and can accommodate a detailed surface of predicted utility.

The purposes of this paper are essentially two-fold. Previous work in Chapters 4 and 5 has assessed the aggregated model in a simulated environment addressing matters such as whether unbiased and efficient parameters can be obtained and whether there is an optimization algorithm which is well suited to the problem. Hence the first purpose is to show that the model provides meaningful results with real-world data. The particular data in question relate to the aggregate destination choice patterns of Canadian Maritime migrants in the 1990-1991 period. The second purpose is to highlight advantages and disadvantages of the aggregated model in relation to existing approaches.

The outline of this paper is as follows. In the section to follow, a brief theoretical exposition of the aggregated model is provided, highlighting its major features. The third section introduces the Maritime migration data we use, placing it in context and discussing its manipulation. The focus of the fourth section is on specification of the aggregated model using variables defined for this analysis. Also matters relating to estimation and model diagnostics are covered. In the fifth section, the results of the application are discussed.

6.2 Model Background

In this section, only a summary of the aggregated model's theory is presented since these matters have been previously covered in detail in Chapter 3. Assume that in some destination choice context, there are L aggregate spatial units which form a comprehensive choice set for a sample of choice-makers. In the interest of simplicity, we will assume that the zone of origin for the choice-makers is not itself eligible as a destination although this would not be a major complication. Each of the L aggregates, $J \in \{1, 2, ..L\}$ can be sub-divided into M^J elemental zones with $\sum_{J=1}^{L} M^J = M$ defining the total number of elemental zones in the system. We assume that the set of elemental alternatives defines the finest level of resolution on which the choice-makers base their selection but that we, as researchers, have knowledge of the spatial choices only at the aggregate level.

Each choice-maker selects the elemental destination associated with the highest utility, thereby selecting the aggregate alternative in which the chosen elemental alternative belongs. Thus, an aggregate's true utility equals the true utility of the most attractive contained elemental alternative: $U_n^J = \max_{j \in \{1,2,...,M^J\}} U_{jn}^J$. Our estimate of U_n^J is V_n^J, which can be thought of as $E[U_n^J]$, while that of U_{jn}^J is V_{jn}^J. The relationship between the latter two is through $U_{jn}^J = V_{jn}^J + \epsilon_{jn}^J + \epsilon_n^J$, where $\epsilon_{jn}^J, j \in \{1, 2, ..., M^J\}$ and $\epsilon_n^J, J \in \{1, 2, ..., L\}$ are independently and identically Gumbel-distributed errors with $E[\epsilon_{jn}^J] = 0$ and $E[\epsilon_n^J] = 0$. The systematic utilities $V_{jn}^J, j \in \{1, 2, ..., M^J\}$ are a function of attributes of places, characteristics of choice-makers and/or distance variables.

It is assumed that $Var[\epsilon_{jn}^J + \epsilon_n^J] = Var[\epsilon_{jn}^J] + Var[\epsilon_n^J] = \pi^2/6$. The particular allocation of the total variance between ϵ_{jn}^J and ϵ_n^J may differ between aggregates and is determined by an aggregate specific parameter λ^J. One can show that $\lambda^J = \sqrt{Var[\epsilon_{jn}^J]/Var[\epsilon_{jn}^J + \epsilon_n^J]}$, which restricts the theoretical range of λ^J to: $0 \leq \lambda^J \leq 1$. When the elemental error tends to occupy all the possible variance: $Var[\epsilon_{jn}^J] \rightarrow \pi^2/6$, $Var[\epsilon_n^J] \rightarrow 0$ and $\lambda^J \rightarrow 1$. In this case the utilities of elemental alternatives within aggregate J are uncorrelated. The polar opposite is $Var[\epsilon_{jn}^J] \rightarrow 0$, $Var[\epsilon_n^J] \rightarrow \pi^2/6$ and $\lambda^J \rightarrow 0$, resulting in highly correlated utilities.

Under these theoretical circumstances the choice probabilities for aggregate $J \in \{1, 2, ..., L\}$ can be shown as:

$$P_n(J) = \frac{\exp\left[V_n^J\right]}{\sum_{K=1}^{L} \exp\left[V_n^K\right]} \tag{1}$$

where:

$$\begin{aligned} V_n^J &= \lambda^J \ln \sum_{j=1}^{M^J} \exp\left[\frac{V_{jn}^J}{\lambda^J}\right] \\ &= \bar{V}_n^J + \lambda^J \ln \frac{1}{M^J} \sum_{j=1}^{M^J} \exp\left[\frac{(V_{jn}^J - \bar{V}_n^J)}{\lambda^J}\right] + \lambda^J \ln M^J \end{aligned} \tag{2}$$

The systematic utility of an aggregate is decomposed into a mean systematic utility term $\bar{V}_n^J$, a positive size term $\lambda^J \ln M^J$, and the remaining term which can be shown to be positive and measures the variability of systematic utility within aggregate J. The latter is appropriately referred to as the heterogeneity term. It is important to note that parameter λ^J can be estimated empirically from model 1 and 2. Beyond its theoretical significance as a measure of utility autocorrelation within aggregate J, λ^J is also linked theoretically to the size and heterogeneity terms. One can show that when $\lambda^J \rightarrow 1$ the size effect is maximized while the heterogeneity is minimized. Intuitively, this is because the large, uncorrelated elemental errors within aggregate J will make the identification of the elemental alternative of maximum utility a highly uncertain process. The opposite is true when $\lambda^J \rightarrow 0$.The behavioural rationale for the heterogeneity term is that the higher the variability in systematic utility within an aggregate, the more likely that this aggregate contains the highest elemental utility peak. The size term can be intuitively understood as measuring the extent to which several elemental alternatives are candidates to provide the maximum utility. With $\lambda^J \rightarrow 0$, for example, size becomes irrelevant since the identity of the most attractive elemental alternative is precisely identified by the model.

The OMNL applied at the aggregate level is a special case of the aggregated logit model in that it takes account only of the average utility in an aggregate, omitting the heterogeneity and size terms. From the behavioural perspective the OMNL postulates that choice-makers select the aggregate destination of highest mean utility, while the aggregated spatial choice model postulates that they select the aggregate containing the highest elemental peak in utility. The two rules applied to the same case may lead to different selections of aggregate alternatives.

6.3 The Data and Their Context

6.3.1 Migration from the Canadian Maritimes

The Maritime or Atlantic region of Canada, composed of the four provinces: Newfoundland, Prince Edward Island, Nova Scotia and New Brunswick, has for many years been seen as the classic 'have not' region within the country. As such it is characterized by an industrial sector which is relatively underdeveloped compared to more prosperous regions such as Ontario and by a population which is comparatively dependent on social assistance. Resource sectors, which drive much of the Canadian economy and can be the salvation of otherwise struggling regions, are less well-developed in comparison to Western Canada. The single dominant resource over the past decades has been the Atlantic fishery on which the economy of Newfoundland in particular has been highly dependent. Unfortunately, poor fishery management practices have led to a ban on cod fishing and further stress on the Maritime economy. New Brunswick has been the one Maritime province to formulate progressive policies and in recent years has invested massively in human capital and technological infrastructure in an effort to eliminate dependency on resource industries and social assistance.

For many there remain clear economic disincentives associated with living in the Maritimes although it has been speculated that the costs are cushioned by the generous social safety net (Courchene, 1994). The propensity of different Maritime population segments to migrate or stay within the region is an interesting matter, however, in this analysis we restrict ourselves to considering the destination choice patterns of those who actually change regions. Among this sub-population there are two dominant tendencies. One is that Ontario is the location of choice for migrants and typically captures more than half of Maritime outmigrants among the various population segments. Clearly, easterners perceive Ontario as the province that drives the national economy and hence is associated with the largest proportion of opportunities. A second dominant aspect is the glaring unattractiveness of Quebec to the vast majority of migrants; this despite the fact that Quebec is the nearest alternative region to the Maritimes. It is avoided for reasons of cultural and language dissimilarity and uncertainties about Quebec's place in the Canadian Federation. These aspects and several others will be highlighted through the models which follow.

6.3.2 The Data

All the data for this analysis ultimately originate from the 1991 Canadian census. Data relating to the personal attributes of specific migrants have been taken from a special micro-data set: the public use sample of individuals. From the public use sample, we are able to assess the migrations of Canadians among the ten provinces and the two territories but are unable to obtain comprehensive information on sub-aggregate choices. Hence we are presented with real-world data possessing the exact limitation which the aggregated model seeks to address.

6.3.2.1 Manipulation of the Micro-data

A representative sample of over 800 000 persons from the Canadian population was compiled to form the public use sample data. This is almost a 3% sample of the population. Of the over 800 000 individuals in the original sample, the overwhelming majority are removed for this analysis. Initially, the sample is reduced by nearly two-thirds through the fact that we deal only with individuals who identify themselves as the primary household maintainer. It is assumed that such persons are the ones who can best be associated with the migration decision. Secondly, all persons who do not live in the Maritimes at the beginning of the time period are removed, thereby reducing the sample to 23 995. Of these persons, only 287 were observed to change regions between June 1990 and June 1991, the one-year reference time period for this analysis. This represents a departure rate in the sample of 1.2%. In earlier censuses, the researcher was obligated to use a five year time interval, something which is risky since many movements are hidden with this low level of temporal resolution. It is thought that the one year interval is a considerable improvement in this regard. The down side of course is that the sample size of migrants is substantially reduced.

6.3.2.2 Definition of the Spatial Aggregation Scheme

The Yukon and the Northwest Territories were eliminated as potential destinations for Maritimers since the flows involved are tiny. This leaves six provinces which could act as potential aggregate destinations. It was deemed that Manitoba and Saskatchewan be combined since they would tend to be perceived as similar entities from the perspective of a Maritimer and because the associated flows are small. The four provinces of the Maritimes themselves of course are aggregated to form the Atlantic Canada region. In the end, six regions are employed in the model to define the set of aggregates: the Maritimes, Quebec, Ontario, the Prairies, Alberta and British Columbia. At the disaggregate level, elemental alternatives are based on urban agglomerations. It is felt that in an inter-regional context, census agglomerations (CA's) and census metropolitan areas (CMA's) represent the most realistic elemental choice entities both from behavioural and computational points of view. The distribution of these elemental alternatives by aggregate is as follows: Quebec (28), Ontario (35), the Prairies (10), Alberta (10), British Columbia (21) which together result in a choice set with 104 elemental alternatives.

The destination attributes used for the analysis are spatially referenced census data which Statistics Canada derives from 20% samples of each census enumeration area. The results from these samples can be obtained at any desired level of aggregation supported by the census. We are interested in the CA/CMA data. In order to ensure that the spatial information is comprehensive, a residual elemental alternative is defined in all regions to capture the influence of rural areas and towns outside the CA/CMA boundaries. For destination attributes which are rates or averages, the estimated values for the residual elemental alternative in each region are calculated as:

$$\overline{x}_{res} = \frac{N_{reg}\overline{x}_{reg} - \sum_j N_j \overline{x}_j}{N_{res}}$$

where j indexes the non-residual elemental alternatives in a region, $\overline{x}$ relates to the particular variable, and N refers to population which is used as a weight. Destination attributes which are simple 'count' variables are easily obtained for the residual alternatives as: $x_{res} = x_{reg} - \sum_j x_j$.

6.4 The Variables and Their Specification

This section's objective is to highlight the migration variables used in the analysis and to identify the dominant specification techniques, some of which are not widely used with the OMNL, but become of more relevance when applied to the aggregated model. While it is possible to construct models which utilize all specification techniques at once, in this paper we wish to maximally differentiate the techniques through separate specifications. We wish to emphasize at the outset that all combinations of the techniques discussed below could be utilized within the context of a single specification.

The range of variables tested in this analysis has by no means been comprehensive and does not pretend to constitute a thorough migration analysis. Their selection is based partially on the results of other research and partially on preliminary analysis. The descriptions are listed in *Table 6.1*. Clearly the

Table 6.1: Variables Employed in the Analysis

<u>Destination Attributes</u>

popsize: natural logarithm of population
lang1: proportion of population at that point speaking english only
dist1: natural logarithm of elemental alternative distance from Maritime centroid
temp: mean annual temperature in degrees Celsius
precip: mean annual precipitation in thousands of millimetres

<u>Choice-maker characteristics</u>

native1: 1 if in 1990, the migrant resided in province of birth; 0 if a non-native
english: 1 if migrant speaks english only; 0 if migrant is bilingual
nfld: 1 if migrant moves from Newfoundland; 0 if not
ns: 1 if migrant moves from Nova Scotia; 0 if not (note: nfld=0 and ns=0 defines migrants from NB and PEI)
hskill: 1 if migrant is in a high skill occupation; 0 if not
mskill: 1 if migrant is in a medium skill occupation; 0 if not
lskill: 1 if migrant is in a low skill occupation; 0 if not
(note: hskill=0, mskill=0, and lskill=0 defines migrants who are not in the labour force)

<u>Interactions</u>

dist2: natural logarithm of distances from individual Maritime provinces
(i.e. province of origin interacted with appropriate set of distances)
lang2: Interaction of "english" with appropriate language proportions over space
native2: Interaction of "native1" with 5 year in-migration totals

types of variables which affect destination choice, the model type of this analysis, may differ considerably from those which affect the initial decision to move. Age for example is strongly associated with the likelihood of moving but it is less effective in predicting the choice of destination.

6.4.1 Specifications Emphasizing Spatial Heterogeneity

At one extreme, patterns of destination choice can be modelled as a function of spatially referenced attributes, implying that choice-makers will not be differentiated. Hence, $V_{jn}^{J} = f(Z_{kj}^{J})$, $k \in \{1, 2..., K\}$ where Z_{kj}^{J} defines the value of the kth destination attribute and K indicates the total number of attributes. The goal of such a model is to determine the main destination factors which account for aggregate choice behaviour recognizing that this characterization may not capture how individual segments of the population behave.

The destination attributes used in this analysis are essentially self-explanatory and are defined in *Table 6.1*. The population size variable (POPSIZE), mean annual temperature (TEMP) and total annual precipitation (PRECIP) variables are included for obvious reasons. With respect to distance, DIST1 is relevant to the model type of this section, which employs no systematic utility variation across choice-makers. It is assumed that all choice-makers originate from one central point in the Maritimes which acts as the base for calculating the great circle distances over the earth's surface to each of the 104 elemental alternatives. These results are subsequently transformed by the natural logarithm since the friction-of-distance is reputed in the literature to have a diminishing marginal impact. The language variable, LANG1 is simply the proportion of the population at each elemental alternative which speaks English without knowledge of a second language. It is expected that this variable will help capture the unattractiveness of Quebec to Maritimers.

In specifying destination attributes for the aggregated model, it is important to keep in mind that they bear strong resemblance to alternative-specific constants in that both are used to explain the aggregate or market shares exhibited in the choices. Constants are described by Train (1986) as being used to absorb all systematic variation not captured by a model's variables. Since a full set of constants will result in $L - 1$ parameters to be estimated, it can be a rather pointless exercise to utilize this approach, especially if there are many choice alternatives. A prudent approach is the specification of as few as possible destination attributes in the place of constants which efficiently explain the aggregate choice shares while providing insights that constants do not. Hence, the number of estimated parameters is greatly reduced because destination attributes are applied in a generic manner, with one parameter applying to all spatial alternatives for a given attribute. In our context, with the five aggregates that Maritimers can choose, the bigger problem in employing constants is the associated lack of insight rather than needing too many parameters, but the former is a problem we wish to address in the analysis. With constants a maximum of $L - 1$ parameters can be employed to differentiate the L aggregate utilities. When generic destination attributes are used instead, if $K \rightarrow L$, near or even partial collinearity among the attributes will typically lead to estimation difficulties although in some cases a model may converge with $K = L$.

It is noteworthy that in 2 an aggregate level constant is essentially built into the model through the size term. Hence it is possible to set $V_{jn}^J = 0$ for all j and J, implying that $\bar{V}_n^J = 0$ and:

$$\lambda^J \ln \frac{1}{M^J} \sum_{j=1}^{M^J} \exp[\frac{(V_{jn}^J - \bar{V}_n^J)}{\lambda^J}] = 0$$

and yet capture the market share probabilities through the size effect. Since many relevant variables might be omitted from such a model, theoretical violations with λ well outside the unit interval would tend to occur. If λ^J, $J \in \{1, 2, ..., L\}$ are constrained to equality, however, explanation can be regained by specifying important destination attributes to complement the generic size effect. Such an approach will reduce the likelihood of theoretical violation. It should be remembered though that, as above, the number of parameters should not exceed the number of aggregates.

6.4.2 Specifications Emphasizing Heterogeneity Across Choice-makers

It is possible to apply the aggregated model even without incorporating sub-aggregate variation in utility. One popular specification method for the OMNL in the discrete choice literature (Ben-Akiva and Lerman, 1985) relates to the use of characteristics of choice-makers in an alternative-specific manner. Such characteristics might be categorical, as in gender, or continuous, such as income. For example, in a three alternative OMNL model which is fully specified with respect to a single choice-maker characteristic (X_n) we would have: $V_n^1 = \beta_1 X_n, V_n^2 = \beta_2 X_n$ while $V_n^3 = 0$. Hence there would be two parameters estimated (β_1 and β_2) to account for the two differences in utility needed to discriminate among three alternatives. This variable type is important in that it allows segmentation so that choice-makers are not modelled as perceiving things in the same way. Indeed, it is through population segmentation that improvements on the fits of market share models are obtained.

With respect to the aggregated model, such specification means are possible but of course these variables cannot generate sub-aggregate variation in utility since sub-aggregate choice information is unavailable. These variables can, however, introduce aggregate spatial variation but only through several parameters. In this context, the heterogeneity effect will be zero while $\bar{V}_n^J$ may be non-zero. The λ^J , $J \in \{1, 2, ..., L\}$ will be directly associated with the size effect and can be used to assess the nature of uncertainty and hence the need for potential re-specifications. In contrast, the OMNL uses alternative-specific constants to capture remaining systematic variation, but there is no theoretical interpretation associated with the resulting parameters.

The particular choice-maker characteristics used for the models of this section are found in *Table 6.1* in the form of variables on nativity, language, province of origin and occupational skill level. The language variable (ENGLISH) contrasts the behaviour of bilingual Maritimers with those who speak only English while the province of origin variables (NFLD, NS) add a spatial dimension to the classification of choice-makers. Note that Prince Edward Island migrants and New Brunswick migrants are joint members of the reference category with no effort being made to contrast the two as there are only six PEI migrants in the sample.

Other variables in *Table 6.1* include nativity which has been a prominent factor affecting destination choices. In particular, natives who move (primary migrants), have been shown more responsive to choosing on the basis of economic opportunity than migrants who are non-natives (Liaw, 1990). Note that the definition for this variable (NATIVE1) does not preclude the possibility that the person was already a return migrant at the beginning of the period. The skill level of the migrant's occupation (HSKILL, MSKILL, LSKILL) is an interesting way to distinguish the abilities of migrants and its effect on migrations. The high skill occupations include professionals and managers, the middle skills are associated with skilled trades and foreman, while the lower skills are associated with many service and retail occupations. Obviously, this is a gross simplification of the categories which apply at the most disaggregate level however the inclusion of more categories is beyond the scope of this analysis.

6.4.3 Specifications with Full Heterogeneity Across Space and Choice-makers

The previous two approaches have been limited to some extent in that they omit important sources of variation either across choice-makers or across space. The OMNL is capable only of simulating spatial variation at the aggregate level although complete population segmentation is possible. From the behavioural point of view, the ideal model should combine detailed population segmentation with high spatial resolution of utility. The model should simultaneously assess what it is about choice-makers which causes them to select certain elemental destinations and what it is about these destinations which makes them attractive to some population segments. This can be accomplished through three general types of variables.

6.4.3.1 Physical Distance

A widely used variable in spatial choice models is the distance between the choice-maker at the origin and the set of potential destinations. The spatial variation in utility induced by distance has a systematic linear spatial trend which can be made non-linear by employing the log of distance. A non-linear effect is assumed in DIST2 which differs from DIST1 in that three potential origins in the Maritimes are considered as opposed to one, which leads to three sets of distances and utilities. The centroids of Newfoundland, New Brunswick and Nova Scotia are utilized to form DIST2. The physical distance variable in general is particularly well-suited to the aggregated model since distances to sub-aggregate locales are easily determined. Interestingly, the use of sub-aggregate distances provides a measure of the manner in which an aggregate's shape and orientation affects its utility, something which the OMNL is incapable of doing.

6.4.3.2 Social Distance

In the Canadian migration context, the language barrier between the predominantly French-speaking province of Quebec and the remaining provinces, which are English-speaking, is often noted. In interactions involving Quebec, this barrier manifests itself in migration flows which seem unrelated to the friction-of-distance. This phenomenon has led migration researchers to define social distance variables which capture such effects (e.g. Anderson and Papageorgiou,1994). The form of the social distance variable is dependent on whether the data are aggregate across choice-makers or if they are microdata. For the former, given an origin zone i, a destination zone j and some social characteristic such as language, the population can be divided into l categories based on the proportional shares S_{ik} and S_{jk}, $k \in \{1, 2, ..., l\}$ of each language group. The variable is defined as $d_{ij}^{(1)} = \sum_{k=1}^{l} |S_{ik} - S_{jk}|$. The larger the cumulative absolute differences in the shares for the given social variable, the greater the social distance between zones i and j. A variable similar in spirit is devised by Sen and Soot (1991). It has the form $d_{ij}^{(2)} = \log\left[\sum_{k=1}^{l} \sqrt{S_{ik}}\sqrt{S_{jk}}\right]$ which has the property that $\exp[d_{ij}^{(2)}] = 0$ when there is a non-existent match between the proportional shares and $\exp[d_{ij}^{(2)}] = 1$ when the shares correspond perfectly.

With micro-data, composite variables such as these make little sense since a migrant of one group most likely will not factor the relative shares of many language groups in making an evaluation. Presumably, the interest is more in assessing the relative concentration of a person's own group as a proportion of the total population at each destination. In that case, following Newbold (1994) the social distance is simply S_{jk} for origin i migrant of group k. Unlike physical distance, we expect a positive associated parameter since migrants are more likely to choose those zones in which $S_{jk} \rightarrow 1$. One problem with this form is that it systematically induces less spatial utility heterogeneity for those groups which everywhere constitute a small proportion of the population. This approach is the one used to define LANG2 (*Table 6.1*) in which the utility of English-speaking people is related to the English proportions at each elemental alternative while the utility of bilingual people is related to the concentrations of French. Hence, by this definition it is essentially assumed that bilinguals would prefer Quebec, all things being equal.

In an alternative variable form, a person might evaluate the proportion that each elemental alternative contains of the country's total population in their group. For a French-speaking person, since Quebec contains the bulk of Canada's french people, it is the highest utility province. The same logic could apply at the sub-provincial level and form the basis for a social distance variable. A model with many such variables essentially contains multiple size effects, each measuring the distribution of opportunities for various population segments. Conceptually, these do not clash with the size term in 3 which expresses size as the number of behaviourally-based choice units. The size or density associated with each of these choice units is a different matter but one which must be assessed, through population or social distance variables, if plausible levels of elemental utility are to be estimated.

6.4.3.3 Interactions

An interaction variable in this analysis is defined by the product of a binary choice-maker characteristic and a destination attribute. The purpose of such a variable is simply to recognize that some population segments will be influenced by a given destination attribute while other segments will not. For example, a reason for the poor performance of employment variables in many aggregate migration studies is that they are of concern mostly to unemployed people rather than those with established careers (DaVanzo, 1978). This type of effect can be accommodated in the aggregated model by an interaction variable which, for example, will generate a heterogeneous utility surface for the unemployed and a flat surface for others. The variable NATIVE2 is one of this type in that it is assumed that natives will be more responsive to recent five year in-migration totals, a surrogate for economic opportunity, than non-natives.

6.5 Estimation and Results

6.5.1 The Optimization Problem and Model Diagnostics

All parameters in this analysis are obtained through the maximum likelihood technique, the goal of which is to find those parameters β, λ which maximize the log-likelihood function:

$$\mathcal{L}^*(\beta, \lambda) = \sum_{n=1}^{N} \sum_{J=1}^{L} Y_n^J \ln P_n(J) \tag{3}$$

where:

- β is a vector of parameters that enter the systematic utilities V_{jn}^J, $j \in \{1, 2, ..., M^J\}, J \in \{1, 2, ..., L\}$ which in turn determine V_n^J and $P_n(J)$ which were defined earlier
- Y_n^J is 1 if choice-maker n chooses aggregate J and 0 otherwise
- N is the total number of choice-makers.

In estimating the aggregated logit model, it has been reported in Chapter 4 that the ordinary Newton-Raphson algorithm may not be successful in obtaining parameters, especially when confronted by regions on the log-likelihood surface which are not concave. Considerable success has been achieved through the use of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) technique, which is a Quasi-Newton algorithm, and the one employed in this analysis.

In devising goodness-of-fit statistics for the aggregated model, there are two useful benchmarks. The log-likelihood for a market share model, which does not differentiate choice-makers, is given by

$$\mathcal{L}^*(c) = \sum_{J=1}^{L} Y^J \ln\left[\frac{Y^J}{N}\right]$$

where Y^J defines the total selections of J. The second benchmark is the naive log-likelihood given by $\mathcal{L}^*(0) = Y \ln\left[\frac{1}{L}\right]$ which corresponds to assuming zero utilities for all aggregates.

Goodness-of-fit is typically assessed on the basis of how well a model improves on these log-likelihood benchmarks. In particular, the ρ^2 statistics are given as $\rho_0^2 = 1 - \frac{\mathcal{L}^*(\beta, \lambda)}{\mathcal{L}^*(0)}$ and $\rho_c^2 = 1 - \frac{\mathcal{L}^*(\beta, \lambda)}{\mathcal{L}^*(c)}$. These measures fall within the unit interval. In a model with population segmentation, it is important to report ρ_c^2 since this will illustrate the extent to which segmentation is improving the fit over the model with no segmentation. If ρ_0^2 alone is reported, a high value is possible without significant contribution from the variables which create the segmentation, particularly when the market shares differ substantially from the naive shares. In general $\rho_c^2 \geq 0.2$ is considered a good fit.

6.5.2 Results and Discussion

6.5.2.1 Specifications Emphasizing Spatial Heterogeneity

The results for this model type, which hypothesizes that destination attributes alone can be useful in describing the essential nature of Maritime migration, are in *Table 6.2*. Interestingly the ρ_0^2 results in this table are almost as good as those of the more complex models in Tables 3 and 4. The true aggregate probabilities are displayed in *Table 6.2* along with the aggregate probabilities predicted by each model. Overwhelmingly, Ontario is the destination of choice for Maritimers while Quebec is least attractive despite the fact that it is the physically closest destination.

At the aggregate level, we would expect that the population size associated with each elemental alternative would be a fairly good predictor, being a surrogate for the density of opportunities available over spacc. This does in fact turn out to be the case and overall, POPSIZE is the strongest and most significant variable. The other main significant result is obtained by controlling for the effects of language (LANG1), with the sign being positive as expected. Beyond these variables, it is difficult to find others which improve the model's explanatory power. This is related to issues raised in the earlier discussion where the difficulty of explaining four aggregate differences in utility with more than three or four generic variables was highlighted. In particular, climate variables such as mean annual temperature (TEMP) and levels of annual precipitation (PRECIP) do not contribute much although the signs at least are as would be expected. With language being controlled for, the DIST1 variable is also employed but provides an insignificant result. We would tend to conclude that distance is not a large consideration for Maritime migrants.

The λ values are consistently high indicating that the size effect is being maximized and the heterogeneity effect minimized. It seems that size, in terms of the number of choice units in an aggregate and in terms of the population in each of these choice units, is quite important and that unobserved variation most likely tends to be unique to individual elemental alternatives. An example of how λ can be used as a theoretical check is given by model 4 which clearly has a specification problem. Note that the role of POPSIZE is minimized and that randomness unique to elemental alternatives is over-exploited in differentiating the aggregates as reflected by the large λ. Model 3 corrects the specification problem by controlling for language with the proof being that λ becomes theoretically valid.

Results from model 3, which is the best in terms of fit and in conforming to expectations, are graphically displayed in *Figures 1* and *6.2*. For prospective Maritime migrants, this is a representation of their probability surface over Canada (*Figure 6.1*) and Canada's most populous region: the corridor defined

Table 6.2: Models Emphasizing Spatial Heterogeneity

		(1)	(2)	(3)	(4)	(5)
Variable	popsize	1.299 (6.766)[c]	1.571 (5.134)	1.321 (5.897)	0.625 (2.224)	1.277 (6.421)
	dist1	-	0.381 (1.092)	-	-	-
	temp	-	-	-	-	0.049 (1.441)
	precip	-	-	-0.092 (-0.27)	-5.780 (5.763)	-
	lang1	1.592 (4.944)	1.283 (3.184)	1.583 (4.826)	-	-
	λ	1.000 (2.801)	1.241 (3.476)	0.994 (2.279)	3.234 (8.314)	1.162 (7.497)
Probs	Que (0.098)[a]	0.098[b]	0.098	0.098	0.101	0.255
	Ont (0.551)	0.555	0.550	0.556	0.554	0.437
	Prairies (0.059)	0.088	0.075	0.089	0.106	0.064
	Alta (0.146)	0.113	0.112	0.115	0.110	0.081
	B.C. (0.146)	0.147	0.166	0.142	0.129	0.162
$\mathcal{L}^*$		-371.76	-371.16	-371.73	-374.27	-397.55
ρ_0^2		0.195	0.196	0.195	0.190	0.155

a - observed Quebec aggregate probability
b - predicted Quebec aggregate probability
c - t-statistics in parentheses

Figure 6.1: The Attractiveness of Canadian Cities to Maritimers

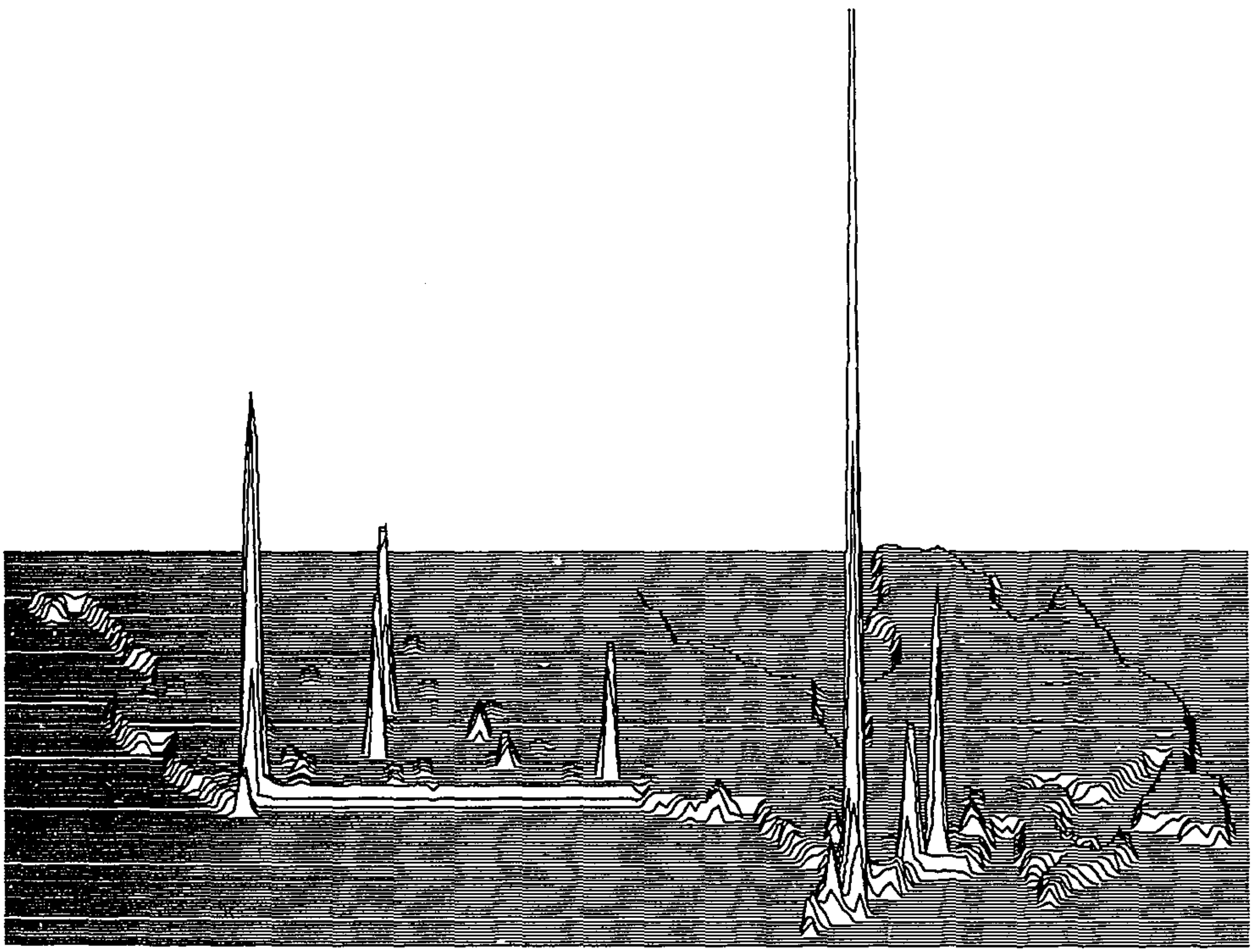
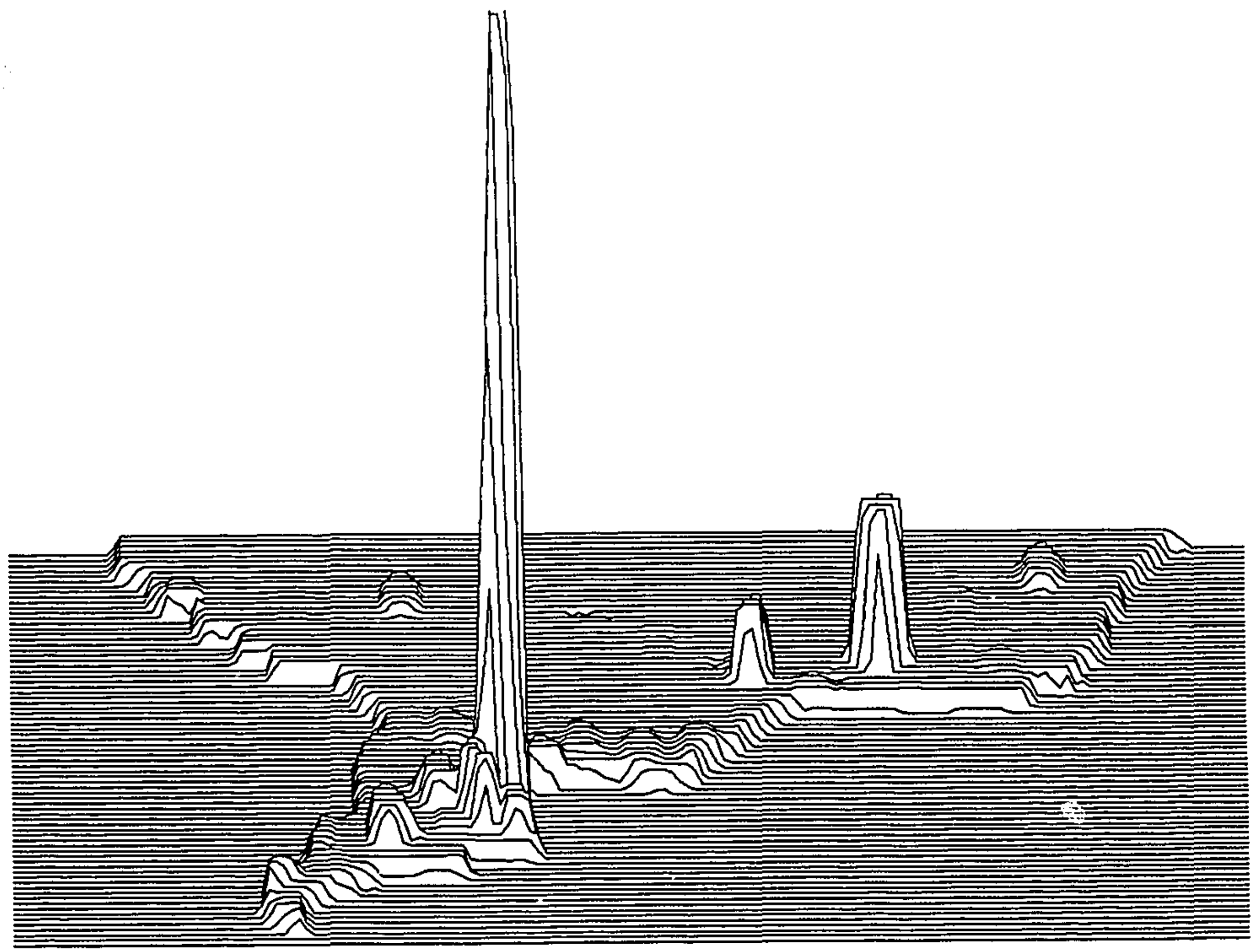


Figure 6.2: The Attractiveness of Ontario/Quebec Cities to Maritimers



by southern Ontario and Quebec (*Figure 6.2*). The figures are meant to give an intuitive feel for the relative attractiveness of the various elemental alternatives although problems with areal definition mean that the surfaces should not be over-interpreted. They are important because they clearly show that despite the total absence of sub-provincial **choice** data, we are able to differentiate elemental alternatives using sub-provincial **attribute** data and associated parameters. The dominance of population in model 3 explains why only a few of the approximately 50 census agglomerations covered in *Figure 6.1* have a substantial probability of being chosen and why Toronto towers above the rest. Montreal is represented by a much lower secondary peak even though its population is only marginally less than Toronto. Since it is in Quebec though, the impact of the language variable drastically reduces Montreal's comparative attractiveness.

6.5.2.2 Specifications Emphasizing Heterogeneity Across Choice makers

A selection of the results for this type of model is presented in *Table 6.3*. There are separate models for each of the main variables: nativity, language, province of origin and occupational skill level. For all variables, Quebec is used as the reference geographic aggregate. The reference categories for the variables are made clear in *Table 6.1*. The overwhelming dominance of positive parameters in the *Table 6.3* is indicative of the fact that being a native as opposed to a non-native, speaking English only as opposed to knowing other languages, moving from either Newfoundland or Nova Scotia or being an employed worker with a skill all increase the chances of moving to a region other than Quebec. Clearly, the glaring unattractiveness of Quebec is the dominant theme in these results with the province tending to be attractive mostly to bilinguals and non-natives who are most likely return migrants. The fact that those not in the labour force are more likely to choose Quebec may explain the actions of such return migrants. It is important to note that the vast majority of the parameters are significantly different from Quebec's reference utility of zero.

Table 6.3: Models Emphasizing Population Segmentation

Variable	*specific to:*	(1)	(2)	(3)	(4)	(5)
native[b]	ontario	0.997 (2.281)[a]				0.706 (8.209)
	prairies	0.981 (1.603)				1.016 (12.09)
	alberta	0.811 (1.590)				0.283 (2.775)
	bc	1.194 (2.328)				0.846 (14.10)
english[b]	ontario		2.469 (5.414)			2.414 (10.23)
	prairies		1.974 (3.018)			2.095 (13.52)
	alberta		4.094 (4.921)			3.997 (10.89)
	bc		3.100 (5.245)			2.828 (9.752)
nfld[b]	ontario			1.253 (2.238)		0.632 (5.643)
	prairies			-0.693 (-0.61)		-1.303 (-6.12)
	alberta			1.163 (1.715)		0.433 (4.203)
	bc			1.030 (1.471)		0.244 (2.416)
ns[b]	ontario			1.041 (2.324)		0.503 (3.445)
	prairies			0.693 (1.072)		0.214 (2.326)
	alberta			1.569 (2.889)		0.955 (7.520)
	bc			1.609 (2.996)		0.958 (7.484)
hskill[b]	ontario				1.743 (3.047)	2.312 (11.86)
	prairies				2.329 (2.151)	2.798 (13.99)
	alberta				1.338 (1.758)	1.990 (15.79)
	bc				1.743 (1.965)	2.293 (14.79)
mskill[b]	ontario				1.040 (1.567)	1.168 (15.12)
	prairies				0.469 (0.331)	0.564 (1.488)
	alberta				1.163 (1.381)	1.289 (7.538)
	bc				2.416 (2.661)	2.462 (19.54)
lskill[b]	ontario				1.478 (2.678)	1.493 (9.449)
	prairies				1.791 (1.639)	1.808 (7.175)
	alberta				2.037 (2.798)	2.123 (26.54)
	bc				2.079 (2.475)	2.054 (11.16)
λ	quebec	0.830 (1.277)	1.199 (12.62)	0.900 (11.69)	1.131 (8.378)	1.768 (23.89)
	ontario	1.165 (25.89)	1.242 (15.53)	1.152 (20.95)	1.197 (10.32)	1.216 (20.61)
	prairies	0.833 (8.010)	1.111 (6.816)	1.001 (8.556)	0.735 (2.543)	1.004 (12.24)
	alberta	1.259 (15.54)	0.714 (3.188)	1.098 (10.17)	1.211 (6.340)	0.605 (8.643)
	bc	0.892 (12.93)	0.841 (5.964)	0.830 (8.737)	0.783 (3.575)	0.631 (10.52)
$\mathcal{L}^*$		-365.85	-344.44	-359.71	-355.24	-324.33
ρ_0^2		0.208	0.254	0.221	0.231	0.298
ρ_c^2		0.008	0.066	0.025	0.037	0.121

a - t-statistics are in parentheses

b - quebec is acting as a reference alternative

Secondary themes revolve around how the other regions fare relative to one other, something which can be assessed through the alternative-specific results for a given variable. A rigorous analysis was not taken out to assess the extent to which some alternative-specific parameters could be constrained to equality, with this being partly to show how parameters can proliferate under the technique of this section. Regarding nativity, the results are generally uninteresting with the main priority of natives being to avoid Quebec. The language results are more interesting with central Canada being more attractive to bilinguals. The province of origin variable is noteworthy for the only negative sign in the table, indicating that Newfoundlanders, being residents of the most easterly isolated province, find the Prairies to be less attractive than Quebec. The other interesting thing is the fact that Nova Scotians, having lived in the Maritime economic heartland, perceive more opportunity in the Western provinces than do those in Newfoundland and New Brunswick. Perhaps Nova Scotians were more aware of the extent to which recession was ravaging Ontario's economy at this time. If there is a general trend in the skill level variable, it is that high skill workers (i.e. professionals and managers) find central Canada more attractive while the lower and medium skill workers, related more to service and infrastructure jobs, are more oriented toward western Canada. Certainly the economies of central Canada were much later in their economic cycles than western Canada and thus perhaps less opportunity existed in lower level occupations.

The results for λ are important for the aggregated model of this section because these are the parameters which distinguish the model from the OMNL whose alternative-specific constants have no theoretical connotation. The results are consistently high reflecting the fact that fit is best obtained by letting uncertainty relate uniquely to individual elemental alternatives in the system. Utilities within aggregates have very low levels of correlation. By definition this version of the aggregated model does not allow intra-aggregate heterogeneity in systematic utility, hence if heterogeneity in utility is necessary to optimize fit, it is left to sets of independent elemental random errors to impose it. For the most part, the theoretical violations with $\lambda > 1$ are minor with the exception of model 5 where the restriction on the nature of systematic utility has its largest effect. The λ are directly associated with the size terms, since there are no heterogeneity terms, and the size terms are being forced to over-compensate. Perhaps the specifications might benefit from the judicious addition of relevant destination attributes, as employed in *Table 6.2*, to relieve the burden on the size terms.

Note that these models provide fairly good fits although it takes a large number of parameters to achieve these results. Note especially that the ρ_c^2 value describing the improvement in fit over the market share model is quite respectable. A study of the model 5 predicted choice probabilities revealed considerable success in describing differential behaviour across population segments. The combination of variables used seems to work very well with there being a substantial reduction in the standard errors when this more thorough appraisal of the migration process is utilized rather than the partial specifications of models 1 to 4.

6.5.2.3 Specifications with Full Heterogeneity

This third type of model attempts to simultaneously capture the important differences between population segments and also the manner in which the utility surfaces for these segments unfold variably over space. As we have seen, it is mathematically possible to specify destination attributes and socio-economic variables separately but the most intuitively appealing method is to interact choice-maker characteristics with destination attributes. Such an approach results in more parsimonious models and is in keeping with the assumption that migrants match their characteristics with certain key attributes of the destinations they consider. In a GIS then, we could produce visual representations of elemental attractiveness as in *Figure 6.1*, except different surfaces or thematic maps could be used for the individual population segments.

Table 6.4: Models with Full Heterogeneity

Variable	*specific to:*	(1)	(2)	(3)	(4)	(5)
lang2	quebec	2.882 (2.335)[a]		9.200 (1.559)	1.212 (6.659)	1.558 (6.183)
	ontario	0.954 (1.963)		1.269 (0.859)	"	"
	prairies	0.443 (1.280)		0.831 (0.546)	"	"
	alberta	2.569 (2.613)		2.941 (1.922)	"	"
	bc	1.570 (3.369)		1.951 (1.273)	"	"
dist2	quebec		-0.553 (-1.304)	-0.362 (-0.49)	0.088 (0.633)	0.314 (0.509)
	ontario		-0.221 (-0.580)	1.083 (1.078)	"	"
	prairies		-0.493 (-1.494)	0.777 (0.879)	"	"
	alberta		-0.359 (-1.122)	0.641 (0.740)	"	"
	bc		-0.353 (-1.110)	0.715 (0.823)	"	"
native1[b]	ontario					0.419 (1.383)
	prairies					0.487 (1.068)
	alberta					0.144 (0.391)
	bc					0.533 (1.464)
native2	quebec				0.034 (3.400)	
	ontario				"	
	prairies				"	
	alberta				"	
	bc				"	
hskill[b]	ontario					1.369 (3.315)
	prairies					1.852 (3.144)
	alberta					0.520 (1.018)
	bc					0.576 (1.125)
λ	quebec	0.488 (1.762)	0.026 (0.377)	0.282 (0.547)	1.499 (7.848)	1.016 (4.032)
	ontario	1.366 (7.806)	"	"	"	1.188 (9.504)
	prairies	1.307 (3.913)	"	"	"	0.647 (4.793)
	alberta	0.908 (10.20)	"	"	"	1.297 (10.38)
	bc	0.988 (4.117)	"	"	"	0.897 (6.847)
$\mathcal{L}^*$		-344.44	-366.81	-342.63	-376.45	-340.50
ρ_0^2		0.254	0.206	0.258	0.185	0.263
ρ_c^2		0.066	0.006	0.071	-	0.077

a - t-statistics are in parentheses
b - quebec is acting as a reference alternative
" indicates identical to parameter above (i.e. constrained to equality across aggregates)

The results in *Table 6.4* show again that language considerations (LANG2) are dominant in the migration of Maritimers. In some specifications this social distance variable is specified generically and in others it is alternative-specific but in all cases the result is positive. This indicates that migrants are drawn to places that feature a high proportion of the population speaking the language(s) with which they are familiar. The alternative-specific specifications indicate that this variable is particularly strong for Quebec and the Western provinces, the former being a deterrent for those who are not bilingual and the latter attracting a large proportion of unilingual english.

The DIST2 variable is employed because, unlike DIST1, it differentiates the distances of the individual Maritime provinces. Overall, this variable contributes much less to discriminating among choice-makers than does language. Again, generic and alternative-specific specifications of this variable are attempted and if anything, the results indicate that a friction-of-distance effect is absent since the results are generally not significant and since the sign of the distance parameter shows volatility. The parameter tends to be negative when specified alone but turns positive in the presence of the language variable. Overall, it appears that little is gained through utilizing different distances depending on the province of origin and that treating all the Maritime provinces as one origin is equally effective.

The NATIVE2 variable is an example of an interaction variable as opposed to a social or physical distance variable. To form NATIVE2, the NATIVE1 variable is interacted with the inter-provincial in-migration total from the previous five years. The five year migration total (1986-1991) is intended to capture those places which would be perceived by 1990-1991 migrants as having fast-growing economies. The interaction tests whether natives are more responsive to pure economic opportunity than non-natives. The parameter is positive and significant as expected (model 4) indicating that places with recent high in-migration totals are of more utility to natives than non-natives.

The final specification (model 5) is included to show that choice-maker characteristics can be employed as they were in the previous section along with social and physical distance variables to produce a model which can be plausibly interpreted. Such an approach can be used to enhance population differentiation while acknowledging spatial heterogeneity in utility. Overall the results seem consistent with those noted in previous models in terms of the interpretations.

In general, the λ values in *Table 6.4* do not stray far outside the theoretical range. When this does happen, most likely the size effect is required to capture variation not accounted for explicitly by other variables and that heterogeneity in elemental utility is inferred to exist in the ϵ_{jn} terms. In an improperly specified model, since the size term may be required to upgrade utility considerably, its associated parameter may be unfeasibly large. In general, Lerman (1975) showed that the size effect would tend to outweigh the heterogeneity effect. Logically , this implies that λ is typically influenced more by the size effect. Certainly, this seems to be the case in this analysis. Theoretical violations can be accommodated, however, through constraints on the λ across alternatives and the inclusion of some relevant destination attributes.

In terms of goodness-of-fit, it appears that the aggregated spatial choice model will not perform better than the OMNL. Comparing the results of *Table 6.4* with those of *Table 6.3* bears this out in the sense that the models of the latter, with no sub-aggregate variation in systematic utility, have close resemblance to the OMNL . While the aggregated model incorporates sub-aggregate information, fit is determined by the ability to accurately forecast aggregate level probabilities, something that the OMNL does very well. However the aggregated model goes about predicting the aggregate probabilities in a manner which captures the underlying elemental processes, something that the OMNL ignores altogether. While it is clear that considerable care needs to be taken in devising social distance and interaction variables for the aggregated model, it is felt that a thorough analysis would culminate in fits comparable to the OMNL.

6.6 Conclusions

We are in a position now to reflect on the objectives of this paper as outlined at the beginning. First of all, there is little question that the aggregated model provides results that are reasonable and that would conform to preconceived hypotheses. The Maritime migrants were assessed using this model and it was found, for example, that language considerations dominate to the point that the generally reliable friction-of-distance measure is of only minor importance in discriminating the behaviour of migrants from different Maritime provinces. Examination of sub-aggregate estimated utilities conform also with what we would expect. Typically the large metropolitan areas tend to be associated with utility peaks, the levels of which however are moderated by the characteristics of the choice-makers. For example, Montreal is the acknowledged peak utility location within Quebec but its level of utility is very much dependent on the languages spoken by the choice-maker.

The doorway to greater realism in choice model specifications is being opened through application of this model. The use of social distance variables and interactions of choice-maker characteristics with destination attributes recognizes that spatial choice processes represent matching of what a choice-maker desires or has to offer with what is available at potential destinations. It is hence acknowledged that some destination attributes are of relevance only to certain segments of the population and that only a few are of broad importance to choice-makers. The greater spatial resolution of this model provides incentive to focus more on these types of variables than would typically be the case with the OMNL. This latter model represents no sub-aggregate variation in utility so there is a tendency to apply choice-maker characteristics which are specified very easily and result in good fit but which provide coarse spatial representations of utility. Such a model answers what it is about choice-makers that motivate choices but leaves the researcher to speculate about what makes destinations attractive.

The advantages and the disadvantages of the aggregated model in relation to the OMNL can be quickly summarized. The main advantage is that the aggregated model provides a more realistic overall representation of the choice process and one which takes better account of all factors. It is a model which is doubtlessly more complex to specify and which leads to greater thought being needed to capture the process. It is a model whose high level of spatial resolution of utility is quite amenable to applications with Geographical Information Systems both in the model-building and final display phases. The main disadvantages are that specifications take longer to implement and that estimation is less reliable and predictable than is the case with the OMNL. Typically though, estimation problems are usually indicative of specification problems. This is an indicator which is not available in applications of the OMNL where even the most poorly specified model is almost guaranteed to converge.

Chapter 7
Conclusions

Discrete choice models have risen to prominence in many choice contexts with the spatial choice context being no exception. The models are based on random utility theory which provides researchers an inherently probabilistic framework with which to explain the choices of individuals. This thesis has focused on a particular problem which is relevant in the spatial choice context and thus to spatial choice models. Namely, what approach should be employed when it is suspected that the choice process for the phenomena under study operates at a more spatially disaggregate level than the available spatial interaction data? A solution has been presented in the form of the *aggregated spatial logit* model, which is well-suited to the problem of utilizing available data in a theoretically plausible way. It is important to emphasize that the model is associated with spatial aggregation as opposed to the problem of aggregation across choice-makers.

Spatial choice problems are unique in the number of potential complications they offer for the application of choice models. It is suggested in this thesis that the aggregated model and its close cousin, the nested logit model, are the best able to strike a compromise between theoretical generality and a manageable computational burden. It was discussed in Chapter 2 how these approaches address the three main spatial choice complications: the non-independence of spatial units, the definition of spatial choice sets, and spatial aggregation of alternatives. The ordinary multinomial logit model (OMNL), since it is consistent with the independence from irrelevant alternatives property, is not capable of explicitly modelling dependency among spatial units. Nevertheless, it is has been the preferred technique in spatial choice problems. The applicability of the probit model, a very general construct, is questionable since for reasons of computational burden, it has yet to be applied to a large scale spatial choice application with many alternatives.

7.1 Findings

7.1.1 Theoretical Insights

From the theoretical perspective, this thesis has provided the most comprehensive available account of the aggregated spatial logit model. In so doing, the thesis has built on the seminal works of Lerman (1975) and McFadden (1978) who provided a theoretical basis for this research but never meaningfully implemented the ideas or discussed model properties in detail. The major theoretical contributions of the research in this thesis are: the formalizing of the properties of the basic form of the aggregated model, the evaluation of these properties, and the derivation of a relaxed asymptotic version of the model based on a less restrictive set of assumptions. These aspects are reviewed here in turn.

This thesis makes it plain that in order to assess the utility of aggregate spatial units, it is important to make a distinction between **aggregate** alternatives and **elemental** alternatives. It is argued that elemental alternatives are the fundamental, spatially disaggregate spatial units potentially considered by choice-makers in the decision process while aggregate alternatives are often defined out of necessity given the aggregate nature of choice data. Since we wish to represent aggregate probabilities and associated parameters in the correct way, it is essential to consider elemental utilities since a proper model should have the capability to acknowledge an elemental choice process when it exists.

Under the assumption that choice-makers pursue the elemental zone of maximum utility within an aggregate alternative, the utility of that aggregate can be shown to be a function of its mean utility, its size in terms of the number of contained elemental choice units, and the variability in utility which occurs within the aggregate. Theoretical discussion in this thesis, unlike previous work, has put a lot of emphasis on how the size and heterogeneity components of aggregate systematic utility are linked to the behaviour of the random error components of utility. Results demonstrated to be true in theory have the added virtue of being intuitively sensible. For example, it is shown that the importance of an aggregate's size in terms of the number of contained elemental choice units is inversely related to their degree of perceived similarity. Similarity is assessed based on the extent to which elemental alternatives have correlated random error components. The heterogeneity of systematic utility within aggregates is shown to be the largest when uncertainty about individual elemental alternatives is its smallest. Essentially, there is an elegant mechanism in place which is capable of assessing whether choice-makers cluster potential destinations in their decision process. Hence, the model is sufficiently general to identify which aggregates are associated with an elemental choice process and which are not.

While the importance of the heterogeneity effect is dependent on the nature of the random error components, it was also shown to be dependent on the spatial variability of the important destination attributes which affect utility. Results of simulations (Chapter 4) are of interest as it was demonstrated how the heterogeneity of spatial attributes could seriously undermine the performance of the OMNL which makes no distinction between elemental and aggregate alternatives. In particular, it was clearly shown that the focus of the OMNL model on mean utility could cause it to miss the overwhelming attractiveness of a single elemental alternative and hence the essential nature of the choice process. This would also lead to severely biased parameter estimates. On the other hand, the aggregated model was shown to perform well under all circumstances.

The asymptotic results examined in this thesis are of theoretical interest because they allow the aggregated model to be based on a less restrictive set of assumptions. In particular, the assumption that elemental utilities are Gumbel distributed can be relaxed if the number of elemental alternatives is sufficiently large and yet the distribution of the maximum utility will nevertheless be Gumbel. Since we are dealing with a maximization process in our assessment of aggregate utility, it is necessary only to make an assumption about the **right tail** of the distribution of an elemental error, or that part of the distribution which could affect the maximum. It is shown that the 'lack of memory' property must be obeyed where this property states that the conditional probability of a random error being x units larger than some given value c is independent of c. Since the exponential distribution is the only one which obeys this property, the assumption is made by extension that the right tail of the random error distribution is exponentially distributed.

7.1.2 Practical Insights

On a practical level, the aggregated model offers researchers new opportunities to make best use of the data which are at their disposal. There are cases where the data containing the observed choices are at a more aggregate spatial level than the choice process and the attribute data which could be used to describe the process. This was demonstrated in the context of Canadian inter-provincial migration in Chapter 6. The Public Use Sample of Individuals, on which that set of trials was based, is notorious for having quite detailed information on choice-makers with respect to most aspects except their geography. Hence, in the case of migrants, we are liable to know nothing more than what province the person has chosen, or in some cases, what census metropolitan area. On the other hand, census data contain useful attribute information about potential destinations at just about any level of spatial aggregation level we desire. It was shown empirically (and graphically) that the aggregated logit model offers a unique way to acknowledge, through elemental utilities, the potential for a sub-provincial choice process even while being restricted to provincial choice data.

A clear practical contribution of this research is that a suitable algorithm for the estimation of the aggregated model has been chosen and for the first time, software for complete implementation of the model has been developed (Appendix 1). In the past, researchers may have recognized the merit of the basic theory, but never was the model implemented in its complete form. Typically, the assumption was made that **systematic** sub-aggregate variation in utility was non-existent and hence that elemental alternatives could not be differentiated from the perspective of the researcher when the available choice information operated at the aggregate level. Even McFadden (1978) sought ways to simplify the model through an asymptotic approximation of the heterogeneity term and ultimately make estimation possible with conventional OMNL software. In Chapter 5 it was shown that the aggregated model in its **complete** form, with choice information being available only at the aggregate level and systematic utilities being differentiated at the elemental level, could be operationalized.

It was concluded that in estimation of the aggregated model through the maximum likelihood technique, the Quasi-Newton family of optimization algorithms offered the best performance. In particular, the Broyden-Fletcher-Goldfarb-Shanno algorithm, which avoids the use of second derivatives altogether but uses first derivatives to obtain an increasingly accurate approximation of the Hessian matrix, was advocated as a superiour algorithm. It was noted that any instability in obtaining parameter estimates would typically be due to the endogenous estimation of the scale parameters which control the relative impacts of the size and heterogeneity terms. Various options were explored to reduce the complexity of the problem such as imposing equality constraints on the scale parameters across aggregates or through the exogenous fixing of one or more scale parameters. Such adjustments were shown to greatly increase the likelihood of convergence and were advocated as excellent intermediate measures in the course of obtaining a final model.

7.2 Directions for Future Research

7.2.1 The Aggregated Model and GIS

An important argument in favour of the aggregated model is its potential when used in combination with a Geographical Information System (GIS). Since one is typically modelling a large number of elemental alternatives in applications of the aggregated model, it can be very awkward to review results without use of graphics. GIS provides a graphical interface and maintains the spatial relationships between choice units. While a small amount of work was done in this thesis on linking the aggregated model to GIS (Chapter 6), it is fair to say that the full potential is much larger than has been shown. The capability of a GIS to quickly display multiple coverages for a given study area lends itself well to assessing the impacts of different variables on utility and on doing comparisons over various population segments. It is expected that use of GIS will quickly show that the ordinary logit model applied to aggregate spatial units provides fairly simplistic results, with very coarse and simplistic representations in utility. In contrast, with the aggregated model, one should be able to appreciate the full spatial diversity associated with choice units.

While many software packages are capable of doing 'basic mapping',a true GIS is one which facilitates more complex analysis by, for example, performing mathematical operations simultaneously on multiple overlays to produce a composite coverage. It is with such capabilities that a researcher can produce a map which shows, for example, all places where a population segment has utility or choice probability above a certain level, where two population segments simultaneously have utilities below a certain level, or where three spatial attributes have particularly large values. In short, if the aggregated model is well-integrated with a powerful GIS, the model will be 'brought to life.' Clear benefits will be realized in the model specification phase, where the researcher will visually be able to see the spatial impacts of different variables and certainly in the phase where results are reported to others. Without a GIS interface, it is questionable whether the aggregated model will gain favour in spatial analysis since its capabilities may not be fully appreciated.

7.2.2 Other Topics

The process of compiling this thesis has generated several ideas that are worthy of examination. One is that the aggregated model provides interesting theoretical results with respect to physical distance variables which measure the spatial separation between origin and destination. Typically, the distance between some origin and an aggregate destination is given by the centroid of the aggregate, thereby treating the aggregate as a dimensionless point. By measuring the individual distances to elemental alternatives within the aggregate however, the aggregated model is capable of capturing the shape and orientation of potential aggregate destinations and how this affects their choice probabilities. This is a fundamentally spatial issue which deserves further attention.

One avenue which has not been explored in this thesis is the idea of hybrid models which combine the nested logit with the aggregated model. In the Canadian Public Use Sample Tape mentioned earlier, researchers have access to choice information about **some** sub-provincial units, namely selected census metropolitan areas, but not others. It is therefore possible to form groupings based on whether the choice information is available. In cases where it is available, elemental utilities will be estimated as they would with a nested logit model; in cases where it is not, the aggregated model can be used. All of these aspects can be accommodated within a single model structure which utilizes all the available information. Another example of a hybrid model is in the migration context, where the move/stay component of the migration decision, in which the required choice information is known, is operationalized at one level while the aggregated model is applied to the destination choice portion of the model.

It is apparent that more extensive comparisons between the nested logit model and the aggregated model are needed. In this thesis, the comparison with the OMNL model has been stressed because it has been the dominant spatial choice model. There is no question that the aggregated model gives a much richer interpretation of the choice process than the OMNL. However, as micro-data sets become more highly developed, with more elemental choice information becoming available, we have greater freedom to apply the nested logit model. It is hence important to understand what is given up if all available choice information is not utilized and the aggregated model is applied. Preliminary examination of the issue, which has not been reported here, indicates that for identical specifications, the aggregated model actually compares quite well to the nested logit.

In a related topic, the sampling of elemental alternatives, as opposed to their aggregation, is a possibility to consider when more elemental choice information is available than can be accommodated. This approach is clearly motivated by a concern for computational burden. While the aggregated model is computationally intensive in accounting for all elemental utilities, it should be noted that the burden is less than is the case with a full application of the nested logit model where every elemental unit is **explicitly** considered as a model alternative.

Insights gained from the aggregated model are of relevance in applications of the nested logit model. In particular, from the estimation perspective, there is the possibility that the Quasi-Newton methods will perform well in full-information estimation of the nested logit. Typically, researchers are still utilizing sequential estimation procedures for the nested logit, with simultaneous estimation being considered a difficult process. There is every indication that the Quasi-Newton methods might reduce the level of difficulty.

Finally, it is important in the future that the aggregated model be applied in a wide range of spatial choice contexts such as industrial location, tourism or retailing. The extent to which this will succeed depends on the validity of the maximum utility hypothesis which states that the utility of an aggregate is equivalent to the utility of the most attractive elemental alternative. Generally, this hypothesis is most reasonable when it is applied in a context which will have long-term implications for the choice-maker. Migration is an example where this is true. On the other hand, a shopping mall might be conceptualized as an aggregate alternative, however it doubtful that the associated short-term choices are made on the basis of the maximum utility available within the mall. An additional complication is that more than one elemental alternative is likely to be chosen in a given trip. Clearly, for this type of complex choice process, an alternative theoretical construct is necessary.

Appendix 1: Software for the Specification and Estimation of Logit Models

1 Introduction

The purpose of this document is to provide a review of software designed by me to make possible the specification and estimation of a wide range of logit models. In particular, the ordinary multinomial logit model (OMNL), the two-level nested logit model and the aggregated logit model can each be implemented. It is a characteristic of GAUSS, the language of this software, that problems of an imposing size can be easily manipulated and estimated with a high degree of computational efficiency. This characteristic lends itself well to discrete choice problems, particularly those with a large number of observations, and those, such as the aggregated model, which are likely to have a substantial number of spatial alternatives. It should be emphasized that while this software has been developed with spatial choice applications in mind, it is readily applicable to aspatial choice problems also such as mode choice or the selection of consumer products, for example. The software code is not included in this manual since it is quite lengthy.

2 Overview

For this overview the reader will find it useful to consult the flow chart in *Figure A* which shows the components of this system: program files, data files, specification files and result files as well as the manner in which they are related. Note that in general, as we move further down the chart, the use of the associated components is likely to decrease. The data and variable creation program **SETDATA** will be used less frequently than the specification program **SPEC** which in turn should be used less frequently than **MODULE**. The reasons for this are to be expanded upon.

2.1 The Programs

It is safe to say that all models, whether the OMNL, the nested logit or the aggregated models, are implemented in two phases: specification and estimation. This software contains a specification program named **SPEC**. This program is used to correctly manipulate choice-maker, attribute and distance data into a form which is consistent with the techniques of discrete choice model specification. Having set up a problem, it is then necessary to obtain the best set of parameters through the maximum likelihood technique.

There are two programs in this software used for parameter estimation: **MODULE** and **NEWTON** , but only one of the two is required for any given attempt to attain convergence. **MODULE** presents a powerful assortment of algorithms for the estimation of complex problems, with the algorithm diversity being due to the maximum likelihood module available in GAUSS. In **MODULE** the particular functional values and analytical results needed to estimate the three general types of models are calculated and then provided for use by an already existing GAUSS maximum likelihood estimation module. Its main advantage is its ability to handle a wide range of difficult problems through its rich algorithmic diversity. Its main disadvantage is that the program, being based on a general maximum likelihood module, is relatively slow. On the other hand, **NEWTON** is designed specifically for speedy application of the Newton-Raphson algorithm and will typically converge many times faster than **MODULE**. This latter program was designed independently of any pre-fabricated GAUSS modules. The main disadvantage of **NEWTON** is that a lack of sophistication in optimization algorithms leads to situations where convergences cannot be obtained. Overall, **MODULE** is best-suited to difficult estimations of the nested logit and aggregated model while **NEWTON** will excel in estimation of the OMNL and simple specifications of the more involved models. In the discussion below, when mention is made of the usage of **MODULE**, it can be assumed, unless otherwise specified, that the comments apply also to **NEWTON** since the two programs are essentially doing the same thing.

There is one other program in the software. It is known as **SETDATA** and is used to convert ASCII data files provided by the user into files in the GAUSS data set format used by the specification and estimation programs. Such conversion is important because storage is very efficient and because any portion of the data can instantly be accessed and obtained. Another source of importance is the fact that the names of variables are stored in these GAUSS data sets and are utilized in **SPEC**. This program is also responsible for controlling which of the GAUSS data sets are considered active by **SPEC** and which are inactive. Finally, this program has the capability of generating new variables on the basis of logical or mathematical statements provided by the user.

Now that the basic objectives of the software are clear, the question might arise as to why the programs have not been combined in some integrated fashion so that specification and estimation can be carried out in a single, seamless step. There are two main reasons.

The most obvious is that the different components are likely to be used at quite different frequencies. The first program **SETDATA**, may only be run once in a given session as it is responsible for creating GAUSS-data which can be accessed according to many combinations of specifications. It determines the family of variables which can be accessed. The program **SPEC** will be run quite frequently, whenever different combinations of β parameters are to be estimated. For the OMNL, **SPEC** and **MODULE** will be run equivalent amounts but for the aggregated model and the nested logit, the latter will be run more frequently as it is responsible for specification of the λ scale parameters. These parameters are typically subject to extensive experimentation for a given specification of β.

Secondly, it is important to separate the programs to allow for thorough checking of the previous actions. The process of incorporating raw data into a specification and estimating the model is quite involved and there are things that can go wrong. It is important for the user to check that data has been set up as intended and that models have been specified as desired before any potentially time-consuming parameter estimation proceeds. Separating the entire process into these three distinct stages facilitates this type of validation procedure.

2.2 Data Files

The data for discrete choice models comes in two basic types: variables referring to choice-makers (or population segments in the case of data aggregated across individuals) or destination attributes which are characteristics of the potential spatial alternatives. A third type, the distance variable, is essentially a hybrid of the two primary types. The data needed to implement any specification will come from at most three distinct GAUSS data set files. The most important of these files is known as **SOCIO.DAT** and is the file which will contain all choice behaviour and all characteristics of choice-making units. No matter what the model or specification, this file must be accessible since it provides access to the choice information against which a model is judged. The file **DEST.DAT**, which contains destination attributes relevant to the spatial alternatives (or non-spatial attributes of alternatives in the case of an aspatial model), need be present only if destination attributes are part of the current specification in some manner. The same goes for **DIST.DAT** which will contain all distance variables potentially relevant to a specification. Clearly there are particular details associated with the structure of each of these files, however such discussion is not relevant to this initial brief overview. It should be emphasized that the user will be working with these main file **types** but will be working under their own file names. The names **SOCIO.DAT**, **DIST.DAT** and **DEST.DAT** are reserved names for active GAUSS data sets. The discussion below will utilize these reserved file names in explanation of the main file types and will also define the concept of an 'active' GAUSS data set.

2.3 Specification Files

These files tell **SPEC** and **MODULE** how the data files should be manipulated to set up and subsequently estimate the model. These files are ASCII and constitute the tools at the user's disposal to make the programs function as desired. There are two specification files: **DESIGN.SPC** and **DEFINE.DAT**. Once the structure of these files, and in particular **DESIGN.SPC**, has been mastered, then use of the programs is mastered also.

DESIGN.SPC is responsible for providing specification instructions to **SPEC**. It contains fundamental information such as names of specified variables, how they are utilized and how elemental alternatives are distributed among aggregates. The reason for the existence of **DEFINE.DAT** is that **SPEC** cannot properly implement the instructions of **DESIGN.SPC** unless the set of alternatives available to each choice-making unit is known. For instance, in a mode choice example, the option of bicycle is presumably not open to someone who does not own a bike. In the spatial choice context, choice sets are frequently constrained in destination choice models since known migrants clearly are not considering their origin as a potential destination. For this reason, the file **DEFINE.DAT** must be specified to differentiate a given choice-maker's choice set from the universal choice set.

3 Detailed Outlines of the Data File Formats

One question that might be answered has to do with why so many data files are needed to encompass the different forms of data. In the past, when software has been developed for the OMNL, it has been possible to develop one data file which will contain all variables no matter what their type. Of course such storage is inefficient in the case of a destination attribute, for example, since such a variable does not vary across choice-makers. Hence in a file with 10 000 observations on choice-makers, the same destination attribute value will repeat itself 10 000 times. Meanwhile, incorporating distance variables, which vary across both alternatives and choice-makers, into a data file adds a lot of confusion to matters. Such variables will be associated with many columns (alternatives) while other variables such as choice-maker traits will be associated with only a single column.

It was felt that such structures for permanent data sets was not sensible for applications of the aggregated model where confusion and inefficiencies would be multiplied many times over. Consider that the utilities for a massive number of spatial units must be represented as opposed to the few typically modelled with the OMNL. Also, the distinction of the different data types through their separation by file allows a better understanding of specification techniques and what measures are appropriate for each variable. For these reasons, the data types are accommodated in the three files discussed below.

Other general comments regarding ASCII data format need to be made. One is that columns of numeric data should not contain any text headers. All relevant variable titles are entered interactively by the user during the running of **SETDATA**. Any text headers in the ASCII data will simply be translated into nonsensical numbers and will throw off the dimensions of matrices to subsequently be formed, hence triggering an error message. At the time of running **SETDATA**, the user should know the ASCII data and be able to appropriately name the columns. The second general comment is that the user should feel free to use comma delimiters in the ASCII data as the program is capable of interpreting these as well as data utilizing spaces instead of delimiters. The sample data in the appendices utilize spaces to separate fields.

3.1 SOCIO.DAT

The only compulsory file of the three main data file types is **SOCIO.DAT** because it is the one which contains choice information. In *Appendix 2* we see a sample of this file's setup which contains the data used later in the examples section of the manual. The sample represents the 1981 to 1986 migration behaviour of in excess of 175 000 persons. The number of rows in **SOCIO.DAT** depends on the number of choice-makers or population segments which are being modelled. In the event that we are using individual level data then each row in **SOCIO.DAT** will relate to a single choice-maker. If grouped data are employed, in which choice-makers who are homogeneous in certain characteristics are aggregated into population segments, then it could be that each row in **SOCIO.DAT** will be relevant to thousands of choice-makers. In *Appendix 2* it is apparent that we are using grouped data.

Notice that **SOCIO.DAT** has a distinct block structure to it. The block on the left contains choice information while each column of the block on the right contains a choice-maker characteristic. Multiple columns are needed to capture all the choice information since all choices typically will not go to one alternative. This is seen in *Appendix 2* where the first six columns are allocated to capturing the choices. Column 1 shows the number of people in each segment who did not change regions in the five year period while the latter five columns capture the destination choices of those who did. On the other hand, in a micro-data model with six alternatives, the sum of each row from columns 1 to 6 would equal 1 since a given choice-maker will choose only one alternative.

The storage of choice information is only one aspect of **SOCIO.DAT**. The other important aspect is the delineation of the particular characteristics associated with a choice-maker or population segment. This is depicted in columns 7 to 17 of the sample **SOCIO.DAT** where variables which identify population segments based on education level, age, nature of household and province of origin are employed. In column 7 for example, all population segments coded as one indicate that the members have post-secondary education and hence are considered well-educated. While all characteristics are binary in our example from *Appendix 2* since grouped data are being used, it is quite possible to have choice-maker characteristics with micro-data which are non-binary such as the observed income for an individual.

One thing that the user should consider at this stage is what names are to be used for the choice-maker characteristics featured in **SOCIO.DAT**. When the translation program **SETDATA** is run, the user will be prompted in turn for the names of these characteristics. These are the names which the user will need to recall in setting up the specification file **DESIGN.SPC**.

3.2 DEST.DAT

The simplest data file in structural terms is **DEST.DAT**. This file contains information on the destination attributes available for specification. Certainly, for a spatial choice model, these attributes will be used to introduce much of a model's spatial variation in utility. The number of rows in this file corresponds with the total number of elemental alternatives in the model while each column is related to a different destination attribute. Note that while **SOCIO.DAT** is completely dependent on data for specific individuals or population segments, the data in **DEST.DAT** is unrelated to choice-makers. The setting up of **DEST.DAT** is essentially straightforward but there is one other matter which is of concern to spatial choice models. We will leave **DEST.DAT** for the moment until this matter is discussed and then point out its relevance for the file.

In many spatial choice models, each choice-maker or population segment has the option of choosing their origin. As a result, an alternative is defined as 'stay' which attempts to capture the probability of not moving while the other alternatives are destination choice options chosen by those who do decide to move. Notice how the diagonal pattern of zeros in the choice information block of our *Appendix 2* example reflect the fact that one's own origin is ineligible as a **destination** choice possibility but is available in the stay option cf column 1.

The key point in this discussion is that a variable associated with the stay option should be modelled as choice-maker characteristic rather than as a destination attribute. Mainly this is because the particular spatial unit associated with the stay option depends on the origin of the choice-maker. If the average provincial income is hypothesized to affect the utility of staying, than this is a characteristic which varies over choice-makers based on their origin. On the other hand, if the provincial income is hypothesized to affect the utility of migrants who consider a given province in a destination choice context, then the provincial income will not vary across choice-makers. Hence provincial income could be considered as a destination attribute or a choice-maker (origin) characteristic depending on the context.

For modelling purposes, the stay option in any spatial choice model can be considered a non-spatial alternative in the sense that it does not make sense to associate spatial utility variation over the surface of the origin. It is considered a non-spatial, point entity. The relevance of all this for **DEST.DAT** is that its number of rows should correspond to the number of elemental alternatives associated with the model structure, whether they are spatial or non-spatial. This issue is illustrated in *Appendix 3* where **DEST.DAT** files used in later examples are presented. The data relating to the stay option is contained in row 1 where it can be seen that all destination attributes are forced, by a row of zeros, to have no effect on the stay elemental alternative. Hence, even non-spatial alternatives can be represented in the otherwise spatially-oriented file **DEST.DAT**.

3.3 DIST.DAT

This file is designed to accommodate all variables which are simultaneously dependent on the origin of the choice-maker and the destination that is being considered. Such variables would include distance, travel time and travel cost, for example. The previous two files contained data that were dependent either on characteristics of the choice-maker or attributes of destinations but did not link the two in any way. While the first two files had distinct columnar structures, **DIST.DAT** is composed of a series of horizontally concatenated blocks of data. Like **SOCIO.DAT**, the number of rows in this file is determined by the number of choice-makers or population segments in the model. The number of columns however is dependent on the number of elemental alternatives being modelled (spatial and non-spatial) and the number of distance variables in the data.

The **DIST.DAT** files used in sample estimations are shown in *Appendix 4*. Unfortunately, there is insufficient room on the page to show how **DIST.DAT** looks for our sample application of the aggregated spatial choice model. The first **DIST.DAT** has 60 rows which is consistent with the model assessing 60 population segments. Also notice that there are two distinct blocks in the data. The first block presents the distances in thousands of kilometers from the associated origin to each destination. The second block simply contains the natural logarithm of the actual distances in kilometers. Note how the sequence of spatial alternatives repeats itself in the header row according to the number of distance variables. Note also that the same logic for the handling of non-spatial alternatives is applied in **DIST.DAT** as was done in **DEST.DAT**. This is reflected by the two columns associated with the stay option all being zeros in keeping with the fact that staying is not linked to measures of distance in any way.

We should note how **DIST.DAT** would differ in the event that an aggregated logit model was being applied. In that case, we would be using sub-aggregate distance measures to differentiate intra-aggregate utility levels. In the first data set of *Appendix 4*, each column corresponds to one OMNL aggregate. If specifying the aggregated model, the Jth column would be replaced with $M^J, J \in \{1, 2..., 6\}$ columns of distance information since elemental alternatives within the J aggregates are all associated with unique distances. For all except the stay option ($J = 1$), M^J typically will exceed 1. In the sample application of the aggregated logit model, the distribution of M^J is $(1, 17, 28, 35, 20, 21)$ while for the OMNL it is simply six ones. For a spatially disaggregate model, it is clear that **DIST.DAT** can have a massive number of columns. It is critical to keep in mind though that the number of columns is given directly by the product of the number of distance variables in the data and the total number of elemental alternatives being modelled.

The opportunity will be taken now to show how **DIST.DAT** would be set up in a non-spatial choice problem, not to mention an example which employs micro-data as opposed to grouped data. In particular, we examine a small mode choice example which is outlined in Ben-Akiva and Lerman (1985, 87-92) where the choice between public transit and automobile is modelled based on the respective travel times associated with each. Clearly in this case the alternatives are the two modes: transit and car in contrast to our other examples where the alternatives are spatial units.

The main variable, travel time, is considered to be in the general category of 'distance' variable since it is neither strictly an attribute of the two modes nor a choice-maker characteristic. The particular travel time incurred depends both on the mode and the spatial relationship between the residential origin and the destination of the workplace. The latter are factors which depend on the person involved. In the second data set of *Appendix 4*, the one relevant to the Ben-Akiva problem, we see two columns of data which form a single block. This is in contrast to our previous example which had two blocks of distance variables. This single block with two columns corresponds to the two 'aggregate' alternatives in the model. Unlike space, these cannot be sub-divided into elemental alternatives but for the sake of consistency, they can be considered as two aggregates with each containing one elemental alternative.

The number of rows in this version of **DIST.DAT** corresponds simply to the number of choice-makers in the micro-data. There are 21 people in this example with each row indicating the travel time that would be associated with the two modes given the person's circumstances.

3.4 DEFINE.DAT

This file is responsible for instructing **SPEC** as to which alternatives are available to a given choice-maker or population segment. Alternatives which are not an option must be omitted from the calculation of choice probabilities. Note that it is not enough to restrict an unavailable alternative to zero utility since an alternative which has zero utility can very easily be associated with a non-zero choice probability in a logit model. Technically this file should be considered as a specification file as it is providing **SPEC** a set of instructions. Nevertheless, its format is discussed in this section concerned with data sets.

The setup of **DEFINE.DAT** is straightforward and is illustrated along with the **DESIGN.SPC** files for our sample applications of *Appendix 5*. As with **SOCIO.DAT** and **DIST.DAT**, the number of rows in this file of binary elements will correspond to the number of choice-makers or population segments. The number of columns in the file will equal the total number of elemental alternatives (spatial and non-spatial) being modelled. Note that the sample file contains columns of zeros which form a diagonal pattern across the associated population segments. These zeros are associated with choice-makers choosing their own origin in the **destination choice** component of the model, which of course is not possible. As a result, the choice-maker's origin must be coded as zero in **DEFINE.DAT**.

It should be noted that the user will be relieved from the responsibility of preparing an ASCII version of **DEFINE.DAT** if a model is fitted in which all alternatives are available to all choice-makers. The user will be prompted for whether this is the case by running **SETDATA**. If yes, then a **DEFINE.DAT** composed entirely of ones will automatically be created. This is what happens with the Ben-Akiva example where both modes are presumed available to all 21 choice-makers.

4 Running SETDATA

This program is an important one for undertaking many of the 'housekeeping' tasks which are necessary. It provides four main functions which are not necessarily related but are nevertheless all included in this single program. In general, the purpose of this program is to manipulate the original ASCII data in such a way that the universe of potential variables is readily accessible through the running of **SPEC**. It is important to note that **SETDATA** is not intended for the creation of huge GAUSS data sets. For this purpose, it is probably better to use the ATOG utility available with the GAUSS software and the **data transformation** package which allows for the creation of data sets with proper variable names. Even if this route is taken some of the four main functions discussed below will need to be employed such as the activation of GAUSS data sets.

4.1 The conversion of ASCII data into GAUSS data sets

A major function of this program is to turn the user's raw ASCII data into GAUSS data which can be easily accessed and manipulated. The program is responsible for the creation of **SOCIO.DAT**, **DEST.DAT**, **DIST.DAT** and **DEFINE.DAT** with all, or perhaps only one of them, being created in a single run. If distance variables for example are irrelevant in some choice problem, then there is no need to create **DIST.DAT**. Dummy versions of unneeded data sets though are automatically created since **SPEC** scans for each of the main variable types and needs to find data sets with the four names above.

In running the program, the user will first be prompted for basic, but important information about the characteristics of the model: 1) the number of observations (or population segments) in the data and 2) the number of elemental alternatives being modelled. In the creation of all files, this information will be checked against the number of elements detected in the ASCII raw data for inconsistencies. If the user has set up the ASCII data in the formats described earlier, then there should be no error messages. For each form of data file, the user will be asked if its creation is desired. If so, then the user will be prompted for the name of the text file containing the needed raw data.

4.1.1 Defining SOCIO.DAT

After entering the file name of the ASCII data, the user is asked to describe the data by column. As described earlier, the left column block (see *Appendix 2*) will be associated with choice information while the right column block will be associated with variables. The respective widths of the choice information block and the variables block then, is information that will be required of the user.

Having defined the size of each block, the user is asked to provide the variable names associated with the columns in the variables block. This is done interactively and sequentially with the request being made to keep variable names to eight or fewer characters.

The user is then asked if the raw ASCII data contain any non-integer values. If the data are only integers, they can be stored much more efficiently than if they contain real numbers and hence will take up less disk space. Finally, it is asked that a file name be provided (without the three letter extension) to store the GAUSS data. The result will be the creation of two files: one with the .DAT extension and the other with a .DHT extension. If one is accidentally erased, then the other will not be operational.

4.1.2 Defining DEST.DAT

If the user chooses to define **DEST.DAT** then the only necessary information is the number of columns or destination attributes that are contained in the file. As with the creation of **SOCIO.DAT**, the user will be asked for the names associated with the attributes, whether the data contain non-integer values and what name should be used for the GAUSS data set.

4.1.3 Defining DIST.DAT

If the user chooses to define **DIST.DAT** then the number of distance variables contained in the data must be indicated. Then prompts will follow for their names.whether the data are integer or not and the name of the GAUSS data set..

4.1.4 Defining DEFINE.DAT

The key aspect in the definition of this file will be whether all model alternatives are available to all choice-makers. If so, then **DEFINE.DAT** will be a matrix of ones and will be automatically created based on information that the program already knows. In this case, it would be unnecessary for the user to have previously defined an ASCII text file since it would go unused anyway. If all options are not available to all choice-makers however, then the text file of binary numbers must be accessible. The user will have created it on the basis of the format previously outlined. Note that **DEFINE.DAT** will be created automatically as an integer file for obvious reasons.

4.1.5 Processing large ASCII data sets

In the event of massive raw ASCII data sets, the user might find that it is possible only to create one of the GAUSS data sets during a single run due to memory constraints. The program is run, one GAUSS data set is created, the memory is cleared, and then the program is run again to create more GAUSS data (Note that in GAUSS the workspace can be cleared by typing 'new' from the command prompt). With huge data, as in the case of choice-maker data with many thousands of observations, it might be necessary to invoke the GAUSS virtual memory capability so that the hard disk mimics RAM memory. This action will slow matters down considerably but will eventually lead to the desired result. This can be activated through the GAUSS run-time module by typing 'VMI' at the command prompt and then following the instructions. It is only in **SETDATA** that there is risk of having to resort to virtual memory. All other components in the system make use of the capabilities of GAUSS data to ensure that the data are broken into chunks which are easily managed by the available RAM. The size of chunks to be processed is controlled interactively when the specification and estimation programs are run. Alternatively, if GAUSS is activated when WINDOWS is running (either define a GAUSS icon or exit to DOS from WINDOWS) then the WINDOWS virtual memory capability should free up a large workspace.

4.2 Determining the ACTIVE GAUSS data sets

By definition, the active GAUSS data for any running of **SPEC** and subsequently **MODULE** will be those data sets named: **SOCIO.DAT**, **DEST.DAT**, **DIST.DAT** and **DEFINE.DAT**. These names are reserved for this purpose. It was felt that running **SPEC** would be tedious if the user needed to key in nearly a half dozen different file names everytime the same model was run with perhaps only a subtle difference in specification. Hence it made sense to designate certain GAUSS data sets as being active and hence perceived by **SPEC** as the data to be acted upon until advised otherwise. An important function of **SETDATA** then is the control over which data are active.

Most likely, the user will have developed a library of different GAUSS data sets, relating to different model types, from undertaking the function described in the previous sub-section. If at any time it is desired to do one or more specifications of a particular model type then the associated GAUSS data must be activated. To do this the user will be prompted for the names assigned by the user to these data and then the data files will be copied automatically to **SOCIO.DAT** and whichever of the other main GAUSS data sets are to be activated. The user should see on the screen whether the copying has been successful based on the normal DOS messages which should be displayed. Whenever a different choice-maker file, for example, is to be activated, the current **SOCIO.DAT** file is erased and replaced by a new **SOCIO.DAT** containing the information in the current data.

4.3 The Definition of new Choice-maker variables

Often, the choice-maker or population segment variables which were in existence with an original ASCII data file will differ to some extent from the variables in the final model specified. There is always the desire to create new variables or design choice-maker interactions from the original set. Both the addition and deletion of variables is easily undertaken with **SETDATA.**

To add a variable, the user is first asked the name of the GAUSS data set to be modified and then is asked to confirm that a new variable is being created. Now the name for this variable to be created is required. At this point the instruction is given to enter some logical or mathematical expression involving current variables available in the user-specified data. Some samples of acceptable statements are as follows where the user should substitute the applicable variable names for VAR1, VAR2 etc.

- VAR1 .* VAR2 - the new variable will be the product of VAR1 and VAR2
- VAR1 .* VAR2 .* VAR3 - the new variable will be the product of three variables
- VAR1 .gt 2 - a binary variable will be formed which is 1 when VAR1 exceeds 2 and 0 otherwise.
- VAR1 .ge 2 - the new binary variable is 1 when VAR1 is greater than or equal to 2.
- VAR1 ./ VAR2 - the new variable is results from the division of VAR1 by VAR 2.
- VAR1^2 - the new variable is the square of VAR1
- ln(VAR1) - the new variable is the natural logarithm of VAR1
- VAR1 + VAR2 + VAR3 - the new variable is the sum of three variables
- exp(VAR1) - the new variable is the result of VAR1 to the base *e*.
- VAR1 - VAR2 - the new variable is the difference between VAR1 and VAR2

- VAR1 .gt 2 .and VAR2 .lt 3 - a binary variable is formed which is 1 if VAR1 exceeds 2 while VAR2 is simultaneously less than 3.

- VAR1 .gt 2 .or VAR2 .lt 3 - a binary variable is formed which is 1 either if VAR1 exceeds 2 or if VAR2 is less than 2.

- VAR1 .eq 3 - a binary variable is formed which is 1 if VAR1 .eq 3

- VAR1 .ne 3 - a binary variable is formed which is 1 if VAR1 is unequal to 3.

There are probably other types of expressions which are possible but these are most of the important ones. There are no limits on the number of terms in a statement other than everything should fit within a line. The syntax in the statement must adhere to the rules of the GAUSS language which, as the user can see above, are quite intuitive. The dot operators in the form of .*, ./, .and, .gt,.le,.or and so on are indicative of the fact that the given operation must be carried out on an element-by-element basis for the entire vector. If the variable creation is successful then the user will be asked if another action, either addition or deletion of a variable, is to be undertaken.

The deletion of a variable is quite simple. After saying a deletion is required, the user is asked to name the variable to be removed from the GAUSS data set. Once named, the action is carried out.

4.4 The Definition and Checking of INTERDAT

Another important function of **SETDATA** has to do with the specification of interactions between choice-maker characteristics and destination attributes. A specification file known as **INTERDAT** is defined here according to the instructions of the user. This file says essentially what combinations of choice-maker characteristics and destination attributes are utilized when the interaction is actually formed. The two matters of concern for this specification file are on the one hand how to create it and on the other how to check its contents.

4.4.1 Defining INTERDAT

INTER.DAT is a specification file and not a data file, hence there is no chance of it causing memory problems. As mentioned, it gives instructions on how interaction variables are to be formed. The underlying logic of this variable type is that different population segments are assumed to assess different destination attributes over space in deriving their levels of spatial utility. There are two main types of interaction variables accommodated: similarity indexes which affect all population segments for a given characteristic (e.g. ethnicity) and simple interactions for which the given characteristic is relevant for only a few or one of the population segments.

As an example of the former interaction type, perhaps we are postulating that migrants seek places which have high concentrations of their own ethnic group. We might have a series of three binary choice-maker characteristics which indicate whether or not the given migrant is a member of some group or not. Also, we might have destination attribute information which details the proportions of the three ethnic groups at various spatial units. By matching choice-makers with the relevant destination attribute, we have a means for assessing in general how ethnic similarity affects locational choice.

An example of the latter form of interaction might be the fact that unemployment over space is not really a factor to those people who have a claim on jobs but may be a strong factor for those who are unemployed. Hence an interaction could be specified in such a way that those who are unemployed are the only ones to be affected by this destination attribute. For the employed people, we would need to define the 'zeroes' variable which is added automatically to the data in **DEST.DAT.** This is simply a column which is zero across all elemental alternatives. In this example, the employed would be interacted with that column. This is in contrast to the previous similarity type variable where all population segments were modelled as being affected by some destination attribute.

This explains conceptually what these types of variables are about however it remains to mention how the actual specification of **INTER.DAT** works. After saying that we desire to form such a file the question will be asked as to whether the newly specified variables are to be appended to a previous

INTER.DAT or are to form an entirely new **INTER.DAT**. If the latter is chosen, then any existing **INTER.DAT** files will be overwritten unless the .DAT and .DHT files are saved to an alias. Then the number of new interaction variables must be indicated. The user will then define the first interaction variable by indicating the total number of socio-economic categories, the names of the choice-maker binary variables, the names of the matching destination attributes, and finally the name of the new interaction variable itself. The same process follows for the remainder of the interaction variables to be formed. The resulting GAUSS data file **INTER.DAT** contains all the relevant column numbers of the variables from **SOCIO.DAT** and **DEST.DAT** which are needed.

Essentially **INTER.DAT** is storing numerically the column locations of needed variables as they were found in the **SOCIO.DAT** and **DEST.DAT** files active at the time. Clearly if radically different **SOCIO.DAT** and **DEST.DAT** files using similar variable names are active during the running of **SPEC**, then misspecifications will occur since different variables will occupy the same columns. The user must guard against the possibility of the wrong columns in the **SOCIO.DAT** and **DEST.DAT** files being used. However, as long as distinct variable names are used for the variables specified in **INTER.DAT**, then it is possible to store the specifications of interaction variables for many different model types within the same **INTER.DAT** file.

4.4.2 Checking INTER.DAT

To avoid mix ups, the manner in which **INTER.DAT** is interpreting the active **SOCIO.DAT** and **DEST.DAT** files must be monitored. Upon the request of the user, **SETDATA** has the ability to check how **INTER.DAT** is relating to the current active data sets. In turn the status of each variable defined in **INTER.DAT** will be reviewed. For each variable, the number of categories and the manner in which these categories is formed is indicated along with how destination attributes are matched to the defined categories. Presumably, the user will know which variables in **SOCIO.DAT** and **DEST.DAT** are required to form a given variable. If the user does this checking procedure and finds that unexpected variables will be used to create the interaction, then it is clear that the wrong data sets are active or that the specification of some variable in **INTERDAT** is obsolete. If the former, then activating different data sets should solve the problem.

5 The Specification of DESIGN.SPC

Having set up data in the correct ASCII format, created GAUSS data sets and determined how interaction variables are formed (if needed), the stage is set to actually specify models. Integral to this process is **DESIGN.SPC**, the specification file responsible for getting **SPEC** and ultimately **MODULE** to operate as the user intends. It is a user-developed ASCII file containing both numeric and character data. Its format is relatively straightforward and is outlined here in terms of its important components. Since several different versions of this file will likely co-exist as many specifications are typically tested, the user will need to provide unique names for each of these files. These names will need to be remembered when it comes time to run **SPEC** and it should be emphasized that the three letter extension .SPC (indicating specification file) is compulsory. In the discussion to follow, the generic name 'DESIGN' is used.

5.1 The distribution of elemental alternatives

To this point, the data for any specification has been defined at the most spatially disaggregate level however there is no information about how these elemental alternatives should be grouped together, if at all. As a result, in the first row of **DESIGN.SPC** the user must define a $(1 \times alts)$ numeric row vector referred to a *elem* where *alts* defines the number of **aggregate** alternatives in the model to be specified. The number of columns in *elem* is determined by the number of aggregates in the model while the elements of the vector correspond to the number of elemental alternatives in the associated aggregate. It is important to understand the meaning of *alts* and the meaning of an aggregate in the context of the different model types. For the OMNL, we consider that each alternative is an aggregate despite the fact that each contains only one elemental alternative. Hence, *elem* would simply be a ($1 \times alts$) vector of ones. The aggregated model differs from the OMNL in that some of the aggregates will contain multiple elemental alternatives but the spatial resolution of choice information will be the same: namely at the aggregate level. For the nested logit model, the spatial resolution of aggregate nests will not correspond to the spatial resolution of choice information. We might, for example, have knowledge of sub-provincial choices even though the aggregate alternatives are the provinces themselves. These definitions will be made clearer in examples to follow.

5.2 The naming of alternatives

Having delineated alternatives, the next task is to indicate their names through individual strings of up to eight characters in length. The number of columns in the string vector to be defined in the second row of **DESIGN.SPC** will exactly correspond in dimension to *elem* and is used simply to name the aggregates indicated in *elem*. In the case of the nested logit model, there will be a more disaggregate level of alternatives at which choice information is available and at which the complete range of specification options can be exercised. Hence in addition to naming aggregate alternatives, in the third row of **DESIGN.SPC** the names of the nested model's elemental alternatives should be indicated. To review, the naming of alternatives is done in one row for the OMNL and aggregated models but will take up two rows for the nested logit.

5.3 Parameter-specific Information

The remaining block of information in **DESIGN.SPC** relates to how individual variables are to be used to obtain associated parameters. The format is such that rows relate to parameters while columns relate to alternatives. Hence a tabular specification format is employed in the form popularized by Ben-Akiva and Lerman (1985). One difference is that the tables for this software are transposes of the tables in that work. Mainly this is to take advantage of any text editor's capability to quickly cut and paste and hence rapidly respecify a new set of parameters for a model with a fixed set of alternatives. A related question might ask why the whole table specification process does not take place interactively rather than in the form of the chosen external ASCII file approach. Again, an external file can take advantage of the features of text editors and can far more easily accommodate specifications with a large number of alternatives or parameters. In general, the approach seems to be far less tedious. We now examine the components of the block containing parameter-specific information.

5.3.1 The parameter starting values

Define the number of rows in the block of data as kk, which represents the total number of parameters in the given specification. The first column in the block provides the desired parameter starting values for all β parameters, that is, those associated with independent variables. Starting values for the λ scale parameters are not an issue in **DESIGN.SPC**. Starting values are important because the process of obtaining parameters through maximum likelihood is an iterative one. In many cases, initializing all parameters at zero is **not** a good idea. For example, the alternative-specific constant associated with the 'stay' option in migration models is often much larger than zero since the vast majority of people are non-migrants.

5.3.2 The variable names associated with parameters

For each parameter to be estimated, there is one associated variable. The name of the variable in question should be included in the second column of the block in the appropriate row. This is the name which will appear in any output of the software. If some variable is specified in an alternative-specific fashion, then the same variable name will be associated with more than one parameter. The output provided in the **SUMMARY.OUT** file, however, will indicate to which alternatives a given specification of a variable is relating.

5.3.3 Forming the design matrix

The actions taken to this point will have defined the sequence of variables that will ultimately be observed in the specification matrix, however the results are not refined in the sense that no mention has yet been made of the spatial alternatives to which the variables and their associated parameters are related. For this reason, it is necessary to compose a binary design matrix. Entries of 1 indicate that the given variable is specific to the associated alternative while entries of zero indicate that it is not.

The matrix *design* is ($kk \times alts$) for the OMNL and the aggregated logit model but is ($kk \times sum(elem)$) for the nested logit model. The reason for this difference is the fact that with the nested logit model, alternative-specific specifications at the sub-aggregate level are possible while for the other two models, this is not the case. Hence, for the aggregated model, an element from *design* will affect all elemental alternatives within the associated aggregate in the same way.

Those used to seeing specifications like those in Ben-Akiva and Lerman should note that variable names are not included in the design matrix, only binary numbers which indicate whether the variable is relevant to the associated alternative. The column adjacent to the design matrix provides the variable and applies it to the alternatives indicated by the design. The data sets for this software are set up in such a way that there is no chance of more than one variable being associated with a given row in the design matrix. In the Ben-Akiva example discussed earlier, auto travel time and transit travel time would conventionally be considered as two variables which are generically specified and hence associated with a single parameter. In this software, they are classified as a single distance variable: travel time.

6 Running SPEC

Most of the work relating to this program will have been done in setting up the ASCII specification file **DESIGN.SPC**. Nevertheless, the user will be prompted for a few things. Initially, there will be a prompt to indicate the particular name assigned to the **DESIGN.SPC** text file. Secondly the user will be requested to indicate what general type of model is being estimated: a nested logit on the one hand or the OMNL/aggregated model on the other hand. The distinction, as mentioned, relates to the respective spatial resolutions of the choice information versus the specification. It is necessary also to ask for the number of alternatives (i.e. *alts*) so that the program knows how many elements to expect when it reads in the informatio.. on *elem*. The user should always remember that the number of alternatives equals the number of rows in *elem*.

At this stage, all that is needed to interpret **DESIGN.SPC** has been provided and the program will set about the task of finding all the variables that are indicated in the specification. The only names which should not be found are those assigned to alternative-specific constants since such data are not stored in any data set but rather created as the program runs. If the program indicates that a variable name **other** than a constant is absent from each of the data sets, then this indicates a problem. Either the user has typed the variable name wrong in **DESIGN.SPC** or the data do not exist in the indicated data sets. If this is the case, then the user should answer the posed question accordingly and the program will stop so that the problem can be sought.

Having confirmed that all the needed data can indeed be accessed, the remaining issue is whether the specification process should be split into portions because the enormity of a given problem overcomes the available RAM. If the user indicates that this may be a problem, then the program will ask for the number of observations (population segments) to be processed at once. For large problems, some experimentation on the part of the user might be required to find the right amount to process at one time. If too much is demanded by the user, and the GAUSS virtual memory capability is not active, then the program will terminate with an error message. It is strongly recommended that use of the virtual memory capability be avoided at all costs. It is by definition heavily dependent on cumbersome interaction with the hard disk and will slow matters down much more than the processing by blocks. At this point all issues will have been addressed and the implementation of the specification will commence. The user will be advised of the number of choice-makers or population segments that have been processed and a message will indicate whether the required output files have been generated.

Two output files are created by the program. One is **SPEC.DAT** (SPEC.DHT) which is the major new GAUSS data set created by the user's instructions and is the information acted on by the estimation program. The user should not be concerned with the contents of this data set since it is quite complex. Also, depending on the size of the problem, this file can be quite large and may take up considerable space on the hard disk.

The other file is **SUMMARY.OUT** which should be checked by the user before running **MODULE**. This file will illustrate which portions of the design matrix are associated with which variable types. It will also show how the variables (named earlier in the running of **SETDATA**) have been utilized. If there are any surprises in this file, then the user should go back and respecify the **DESIGN.SPC** file until the desired result is obtained. Because these are still early days in the development of this software, it is possible that **SPEC** might run successfully despite the fact that some specification mistake has been made. For this reason, it is critical to carefully scrutinize the **SUMMARY.TXT** file for something out of the ordinary.

7 Running MODULE

The major task for the user in running this program is the specification of the λ scale parameters. Of course if the model is the OMNL, there is nothing to specify and using **MODULE** will pose no challenge. Other than specification concerns, the only issue to address is the number of choice-makers (population segments) which can be processed at once. The need to also do this in **MODULE** is because its memory requirements are more demanding than is the case in **SPEC**, hence fewer observations can be processed at once. Again some experimentation on the part of the user may be required since generally, the more observations which are done at once, the faster a solution will be obtained.

Assuming that we are estimating the nested logit model or the aggregated model, the main matter of concern will be the λ parameters. In particular, the user will have to indicate which alternatives have exogenous λ and which have λ which are to be endogenously estimated. Of those that are endogenous, the user must identify which of the parameters is constrained across aggregates and what the pattern of contraints is. Implementing this information is not as complicated as it sounds. Essentially, the program will determine the number of λ required based on *alts* and then prompt the user by providing the aggregate alternative name and then requesting that some information on the λ for that aggregate be provided. If the associated λ is to be exogenously fixed, then the user simply indicates the value between 0 and 1 which should be forced to apply. Values outside this range of course would be a violation of theory and are not acceptable. Information on endogenous λ is entered by using multiples of ten so as to be kept quite distinct from exogenous λ. The first endogenous λ should be indicated with 10, the second with 20, the third with 30 and so on. If the second endogenous λ for example is being forced to apply to alternatives 3 and 6, then 20 should be entered when the λ information for those alternatives is requested. For all endogenous λ, starting values will automatically be set to 1.0 which is typically the best place to begin in such models. Overall, the λ specification process is quite simple and the user should have no problems in grasping it.

As a brief aside, it is important to note that in the case of a nest of alternatives which contains only one elemental alternative (also known in the discrete choice literature as a 'degenerate node'), the associated λ should always be set exogenously to 1.0.

At this point, the estimation should commence and the user will be able to monitor its progress. Note that the parameter names specified in **DESIGN.SPC** will be shown next to the current parameter estimates. They may not be displayed in the same order as specified in **DESIGN.SPC** because sorting of the specification by variable type took place in the running of **SPEC** however results are easily followed by noting the variable names during and after the estimation. At this point, either there will be a successful convergence, termination because of too many iterations (the maximum is 100), or termination from a floating point exception. The latter is liable to happen with the aggregated model or the nested logit model if an endogenous λ strays too close to zero. In the first two cases, the main output of the program will be created: **RESULT.OUT**. This file will contain the final parameter estimates, information about their significance and various diagnostics which we assume here that the user will be able to understand. After the table of parameters, there is an important row vector displayed which indicates the current values of λ which are applying in sequence to the model's aggregates. Those λ which have been fixed or endogenous should be readily apparent and associated with the proper aggregate. In addition, λ which have been constrained across aggregates should be readily apparent also.

A potentially useful feature of **MODULE** is that you can interactively retrieve information on the final set of choice probabilities. The final set of choice probabilities at the aggregate level can be obtained by displaying the contents of the matrix **PROB**. The probabilities at the most disaggregate level can be viewed by typing the contents of the matrix **PROBEL**. In the following section, there is a short discussion on displaying matrices interactively.

8 Other Issues

8.1 The GAUSS Environment

In the package received by the user, the four programs in this software: **SETDATA, SPEC, MODULE** and **NEWTON**, are in complied form with the extension .GCG. When the user invokes GAUSS with the command: GAUSSI from DOS (unfortunately, the software is required to run even the compiled programs), the command prompt: >>, should be visible. By using the RUN command from this prompt and indicating the program name, the given program should start running. Also, when the programs have run successfully, the command prompt should again be visible.

There are actions that the user can undertake interactively from the command prompt which might prove useful. A matrix in memory can be instantly printed to the screen by typing the name of the matrix. If the first column of matrix A is to be viewed, the user would type $A[.,1]$, the first row would be $A[1,.]$, the block of 16 elements in the upper-left corner of the matrix could be viewed with $A[1:4,1:4]$. Clearly there are an infinite number of ways to view a large matrix.

8.2 Troubleshooting

This software still has much work to be done on it before it is bug free. No doubt there will be situations where the program will terminate rudely and perhaps without an adequate explanation of why. Here is a list of some of the more likely difficulties and what should be done to get around them.

- 1) *A persistent 'file already open' error* - This GAUSS error occurs when an attempt is made to open a GAUSS data set which is already open. Often this error will fall on the heels of another error which prevented closing of the data sets. There are two solutions: a) type CLOSEALL from the command prompt and attempt to run the program again, b) if that fails, then type NEW from the command prompt, which will clear the workspace, and then run the program again.

- 2) *Persistent 'file not found' errors* - Probably these types of errors result when you type in an incorrect file name or else if you include the three letter DOS extension. This extension is unnecessary. The solution of course are simply to run the programs again in the hopes that you have not set the stage for error#1 to occur.

- 3) *A crashing estimation program* - There are a few things which can cause this problem. One is that data are not scaled very well. If you include a variable measuring annual income **in dollars** rather than in **thousands of dollars** in the same specification as some binary variable, then you are asking for trouble. Some crashes result from the λ parameters in the more complex models straying too close to zero. The solution in that case is a respecification. Difficulties will also occur if the starting values for a specification are arbitrary (e.g. all zeros) or if the Newton-Raphson algorithm is used to estimate models with complex surfaces. Remember, **MODULE** is based on the more flexible BFGS algorithm while **NEWTON** is based on the Newton-Raphson approach.

- 4) *The first time you attempt to run* **SETDATA**, *you receive a message saying that the file TEMP.PRG was not found.* TEMP.PRG is actually a program which is created by **SETDATA** and then executed while **SETDATA** is active. This program must be in existence at the outset though or **SETDATA** is not executed. If you do a directory and find that it is not there, then create a trivial one-line ASCII file: " "; (i.e. two quote marks and a semi-colon) and then save the file under the name TEMP.PRG.

Now **SETDATA** should run.

- 5) *You do not have much success in getting the program to recognize variable names in your GAUSS data sets.* Outright typographical errors are an obvious culprit but more likely it is because GAUSS data sets store variable names in a case-sensitive way. When you specify variables in an .SPC file, you will need to be sure that it is typed in the same way as when you first defined the variables in the associated data set.

- 6) *Insufficient memory errors* - You are attempting too process too many observations at once. Make use of the ability of the software to sequentially process blocks of observations or, if you have some spare time, run the programs using virtual memory.

- 7) MODULE will not run - This program requires information from SPEC that is kept in active memory, not written to file. Hence if the session starts with the running of MODULE, the program will not run because certain elements will be undefined. The solution is to always run SPEC before running MODULE unless you are simply respecifying a set of λ parameters that you just tried.

9 Tutorial

In this tutorial, the raw data which are specified for you on disk represent the starting point as you are taken step-by-step through everything that must be done to specify and ultimately estimate two models: a nested logit model and an OMNL model. The files **SOCIO.TXT**, **DEST.TXT**, **DIST.TXT** and **DEFINE.TXT** are used to test the specifications **OMNLSPC** and **NEST1.SPC** which appear in *Appendix 4*. Please consult the data in the appendices and also examine the actual files in a text editor.

9.1 Running SETDATA

9.1.1 Creating the GAUSS data sets

Note that each of the bullet items below correspond with a single step. The user will need to hit the 'enter' button once for each of these steps although this is not explicitly mentioned below. It is not essential that each data set be created in the same running of **SETDATA**.

- From the DOS prompt, type 'gaussi'

- From the GAUSS command prompt, type 'run setdata'.
- You must now provide preliminary information about your ASCII data. In response to the query about the number of choice-makers (population segments), type '60', to the second question about the number of elemental alternatives, type '6'.

Creation of choice-maker GAUSS data set

- Type '1' to indicate that we do want to create the choice-maker GAUSS data set.
- Type 'SOCIO' to indicate that the raw data are called **SOCIO.TXT**.
- Type '6' to indicate that 6 of the 17 columns in **SOCIO.TXT** are allocated to choice information.
- Type '11' to indicate that the remaining columns are allocated to choice-maker variables.
- You are now prompted for the names of the 11 variables. You should name them (in lower case) as: 1. educ, 2. young, 3. middle, 4. single, 5. east, 6. que, 7. ont, 8. pray, 9. bc, 10. notque, 11. noteduc.
- Type in '0' to indicate that there are no non-integer elements in **SOCIO.TXT**.
- Type 'SOCIAL' to indicate that the new GAUSS data set will be called SOCIAL.DAT (.DHT).
- Type '1' to continue.

Creation of destination attribute GAUSS data set Creation of this data set is optional and depends on whether we will use destination attributes or interactions in our model specifications.

- Type '1' to indicate that we do want to create the destination attribute GAUSS data set.
- Type 'DEST' to indicate that the raw data are called **DEST.TXT**.
- Type '6' to indicate that the data contain six destination attributes.
- You are now prompted for the names of the six attributes . You should name them (in lower case) as: 1. english, 2. french, 3. univ, 4. income, 5. temp, 6. precip.
- Type in '1' to indicate that there are non-integer elements in **DEST.TXT**.

- Type 'DESTAGG' so that the new GAUSS data set will be called DESTAGG.DAT (.DHT).
- Type '1' to continue.

Creation of distances GAUSS data set Creation of this data set is optional and depends on whether we will use distance variables in our model specifications.

- Type '1' to indicate that we wish to create the distances GAUSS data set
- Type 'DIST' to indicate that the raw data are called **DIST.TXT**.
- Type '2' to indicate that there are two distance variables in the data.
- Name the two variables (in lower case letters) as: 1. dist, 2. lndist.
- Type '1' to indicate that there are non-integers in the data.
- Type 'DISTAGG' so that the new GAUSS data will be called DISTAGG.DAT (.DHT).
- Type '1' to continue.

Creation of GAUSS data defining available alternatives

- Type '1' to indicate that we need to create this GAUSS data set.
- Type '0' to indicate that for this example all the alternatives are not available to each population segment.
- Type 'DEFINE' to indicate that the raw data are named **DEFINE.TXT**.
- Type 'DEFAGG' to indicate that the new GAUSS data will be called DEFAGG.DAT (.DHT).
- Type '1' to indicate that we wish to return to the main selection list.

9.1.2 Adding or Deleting Choice-maker variables

Having sent the choice-maker data to the file SOCIAL.DAT (.DHT), we can add and delete new variables to it if we wish. Since we will not be using a newly created variable in the final specification, here we will simply create one and then delete it for illustration purposes.

- Type '4' to indicate that we are adding or deleting choice-maker variables to some GAUSS data set.
- Type 'SOCIAL' to indicate that we will add or delete a variable to SOCIAL.DAT (.DHT). Note that the column-by-column contents of this data set are displayed by name.
- Type '1' to indicate that we wish to add a variable.
- Type 'test' as the name of the variable to be added.
- Type 'educ .* single' as your mathematical definition of the new variable. TEST will simply be the product of the binary variables EDUC and SINGLE. The variable is created as soon as you hit 'enter'.
- Type '1' to indicate that again, we want to add or delete a variable, although now we will be deleting.
- Type '0' to indicate that we are deleting a variable.
- Type 'test' to indicate the name of the variable to be erased.
- Type '0' to say that we are through with creation and deletion.
- Type '1' to return to the main selection list.

9.1.3 Activation of GAUSS data for the specification

- Type '2' from the main menu to indicate that we are activating GAUSS data sets.
- Type 'SOCIAL' to activate SOCIAL.DAT (.DHT).
- Type '1' if the files were copied successfully. Typing '0' would stop the program.
- Type '1' to indicate that we wish to activate a specific destination attribute GAUSS data set.
- Type 'DESTAGG' to activate DESTAGG.DAT (.DHT)
- Type '1' to continue.
- Type '1' to indicate that we wish to activate a specific distance GAUSS data set.

- Type 'DISTAGG' to activate DISTAGG.DAT (.DHT)
- Type '1' to continue.
- Type 'DEFAGG' to activate DEFAGG.DAT (.DHT).
- Type '1' to continue.
- Type '1' to indicate that we have no **INTER.DAT** file defined and that we would like a dummy file created to guarantee that **SPEC** will run. If this file were already in existence, we would type '0'. Actually, in this case, we are just about to create the new **INTER.DAT** file so it matters little how we answer this question.
- Type '1' to return to the main selection list.

9.1.4 Creation and checking of INTER.DAT

While we have created and activated the needed GAUSS data to do our specification and estimation, we have yet to say how the interactions of choice-maker characteristics with destination attributes that appear in OMNL.SPC and NEST1.SPC are formed. This is what we will do here.

- Type '3' from the main selection list to indicate that we are modifying or checking **INTER.DAT**.
- Type '1' to indicate that we are modifying and not just checking **INTER.DAT**.
- Type '1' to indicate that we are creating a new INTER.DAT as opposed to adding to an old one.
- Type '2' to indicate that we will form two interaction specifications.
- Type 'lang' to indicate the name of the first interaction.
- Type'5' to indicate that it has five categories.
- The names of the five categories which you are prompted for should be: 1. east, 2. que, 3. ont, 4. pray, 5. bc.
- The names of the five associated destination attributes are: 1. english, 2. french, 3. english, 4. english, 5. english. Note that we could have done the same variable in two categories have we had paired que and notque with french and english.

- Type 'postsec' to indicate the name of the second interaction.
- Type '2' to indicate that it has two categories.
- The two choice-maker names that are needed are: 1. educ, 2. noteduc.
- The two destination attribute names are: 1. univ, 2. zeros. Note that this latter destination attribute was added automatically to DESTAGG.DAT (.DHT) and is simply a column of zeros.
- Now that the interaction specifications are set, we can now check them by typing '1' in response to the question.
- Type '1' after the specification for the first interaction has been shown.
- Type '1' after the specification for the second interaction has been shown.
- Type '0' to indicate that we do not wish to return to the main selection list. The program will now terminate.

We have now done everything we need to do in **SETDATA** which is really the most involved part of the whole specification and estimation process. Now we proceed to run **SPEC**. Once **SETDATA** is complete, and before **SPEC** is run, it is necessary to create the appropriate **.SPC** file within a text editor. Of course, that file has already been created for this tutorial problem but that will not be the case in your own applications.

9.2 Running SPEC

Initially we will use the data to specify a nested logit migration model of the move/stay type at the upper level with destination choice at the lower level. This requires use of the specification file **NEST1.SPC** which has been provided.

- From the GAUSS command prompt, type 'run spec'.
- Type 'NEST1' to indicate the name of the specification file we are using.
- Type '1' to indicate that this is a nested logit model

- Type '2' to indicate that there are two aggregates; namely: moving and staying.
- Assuming now that the only variable name which has not been found in the data sets is 'stay', type '0' to indicate that no variable names have been mistyped and that the unfound variable is in fact a constant.
- Type '0' to indicate that this problem is sufficiently small that it need not be processed in blocks.

Now the program should run and inform you how many observations have been created and whether the main output SPECK.DAT (.DHT) has been successfully created. The results of the specification should be checked in the file **SUMMARY.OUT**.

9.3 Running Module

Now we actually attempt to estimate the parameters of the model.

- From the GAUSS command prompt, type 'run module'.
- Type '0' to indicate that we do not need to estimate the problem in little pieces.
- In response to the appropriate form of λ to associate with the two aggregates, type '1' for the stay option to indicate that λ is exogenously fixed to 1.0 and type '10' for the move option to indicate that this is the first endogenous λ and that it is specific to the move aggregate.

Now the estimation should commence and a solution should be obtained within 100 iterations of the BFGS algorithm. The results can be found in **RESULT.OUT**.

9.4 Specifying and Running the OMNL example

We can now go straight into the OMNL example and obtain the parameter estimates within a minute or two. Note that **OMNL.SPC** acts on exactly the same GAUSS data sets as **NEST1.SPC** since the models have the same number of choice-makers and elemental alternatives. It is only the allocation of these elemental alternatives to aggregates that differs as far as the software is concerned. If a radically different model were now to be specified with different data requirements, we would need to return to

SETDATA to activate a new set of GAUSS data sets. Here are a list of the steps that differ for this model from the one that we just went through.

- Starting with the re-running of **SPEC**, we type 'OMNL' initially to indicate that we are acting on **OMNLSPC**.
- Type '0' to indicate that we are running an OMNL model.
- Type '6' to indicate that now their are six aggregates.

Everything else in the running of **SPEC** is the same. After checking the results in **SUMMARY.OUT**, we are ready to run **MODULE** for the new model specification. This is more straightforward in this case because with the OMNL model, we need not concern ourselves with the specification of λ. The model will converge more quickly and again the final results will be stored in **RESULT.OUT**.

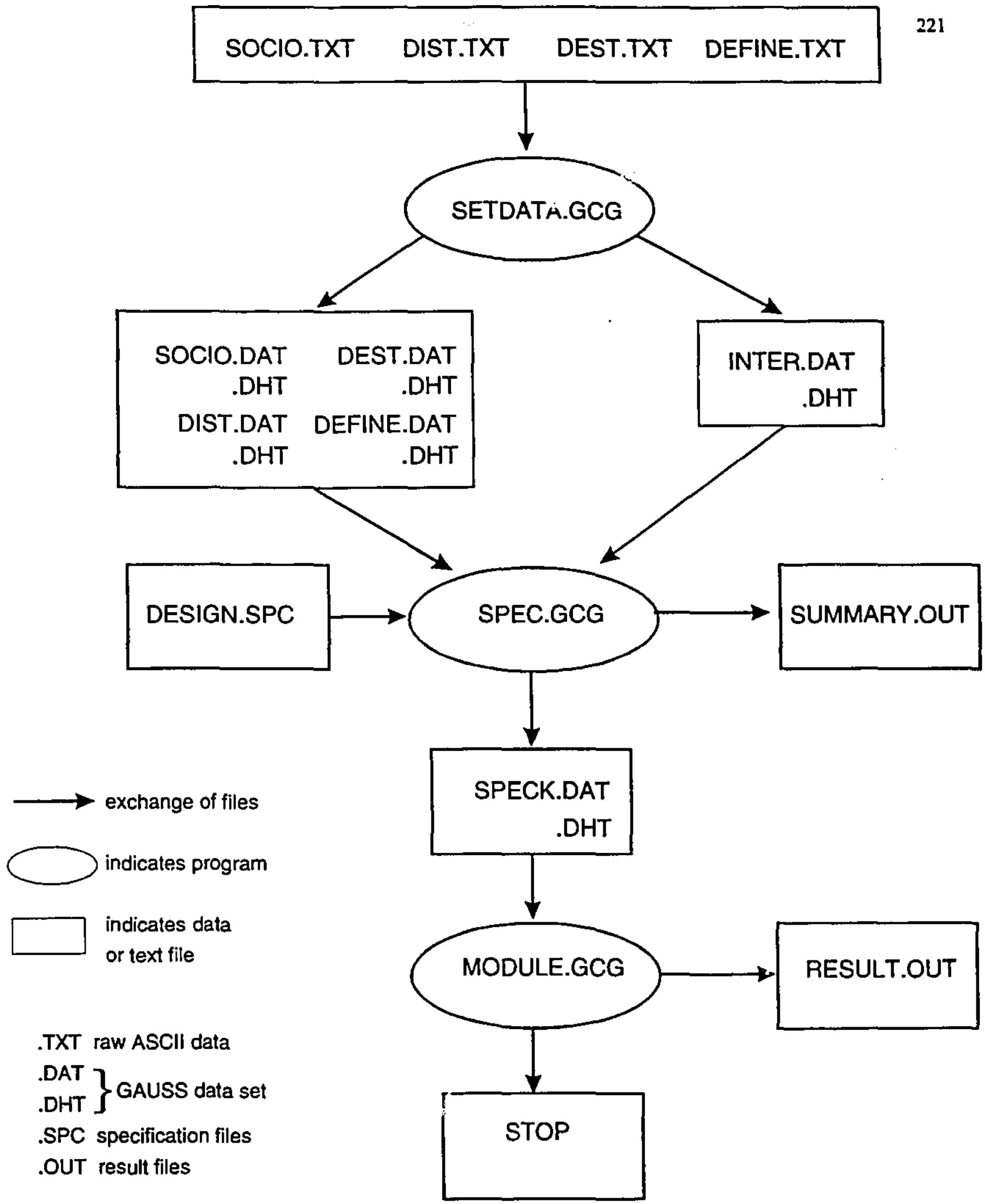


Figure A: Schematic for specification and estimation software

APPENDIX 2: CHOICEMAKER DATA AND CHOICE INFORMATION

DATA SET #1

The first table below illustrates the contents of the file SOCIO.TXT which contains the choice-maker variables and choice information for the sample models utilizing grouped data. Clearly the original ASCII text file will not contain the additional formatting and headers displayed in the table. Note the clear division between the choice information block and the variables block. The definitions of the columns are provided underneath the table.

SOCIO.TXT

STAY	EAST	QUEBEC	ONTARIO	PRAIRIES	BC	EDUC	YOUNG	MIDDLE	SINGLE	EAST	QUE	ONT	PRAY	BC	NOT QUE	NOT EDUC
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0	1	1	0	0	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0	0	1	0	0	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1	1	1	0	0	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1	0	1	0	0	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	1	1	0	0	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	0	1	0	0	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	1	1	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0	1	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	1	1	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	0	1	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1	1	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	1	0	0	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0	1	0	1	0	0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0	0	0	1	0	0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1	1	0	1	0	0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1	0	0	1	0	0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	1	0	1	0	0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	0	0	1	0	0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	1	0	1	0	0	0	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0	0	1	0	0	0	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	1	0	1	0	0	0	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	0	0	1	0	0	0	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1	0	1	0	0	0	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	0	1	0	0	0	0	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0	1	0	0	1	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	0	0	0	0	1	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1	1	0	0	1	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	1	0	0	0	1	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	1	0	0	1	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	0	0	0	0	0	1	0	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	1	0	0	1	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0	0	0	1	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	1	0	0	1	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	0	0	0	1	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1	0	0	1	0	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0	1	0	0	1	1

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	53	1	1	0	1	0	0	0	1	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	0	42	1	1	0	0	0	0	0	1	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	0	13	1	0	1	1	0	0	0	1	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	0	93	1	0	1	0	0	0	0	1	0	1	0
[illegible]	[illegible]	[illegible]	15	0	31	1	0	0	1	0	0	0	1	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	1	0	0	0	0	0	0	1	0	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	1	0	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	0	1	0	0	0	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	0	22	0	0	1	1	0	0	0	1	0	1	1
2	[illegible]	[illegible]	10	0	77	0	0	1	0	0	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	20	[illegible]	49	0	0	0	1	0	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	0	120	0	0	0	0	0	0	0	1	0	1	1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	1	0	1	0	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	1	0	0	0	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	[illegible]	22	0	1	0	1	1	0	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	18	[illegible]	0	1	0	1	0	0	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0	1	0	0	0	0	1	1	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0	0	0	0	0	0	1	1	0
[illegible]	[illegible]	2	18	82	0	0	1	0	1	0	0	0	0	1	1	1
[illegible]	[illegible]	16	31	90	0	0	1	0	0	0	0	0	0	1	1	1
[illegible]	0	[illegible]	9	16	0	0	0	1	1	0	0	0	0	1	1	1
[illegible]	7	6	32	72	0	0	0	1	0	0	0	0	0	1	1	1
[illegible]	[illegible]	6	11	35	0	0	0	0	1	0	0	0	0	1	1	1
[illegible]	8	4	31	372	0	0	0	0	0	0	0	0	0	1	1	1

Columns 1 to 6: Choice Information

Stay - # of people who remained in the same region between 1981 and 1986
East - # of people who moved to the Atlantic provinces from outside the Atlantic region from 1981 to 1[illegible]
Quebec - # of people who moved to Quebec
Ontario - # of people who moved to Ontario
Prairies - # of people who moved to the Prairie provinces
BC - # of people who moved to British Columbia

Columns 7 to 17: Population segment characteristics

Educ - people who have some post-secondary education
Young - 1 if person under the age of 30 in 1986; 0 otherwise
Middle - 1 if person aged between 30 and 40 in 1986; 0 otherwise
Single - 1 if person in a single-person household; 0 otherwise
East - 1 if person lived in the Atlantic provinces in 1981; 0 otherwise
Que - 1 if person lived in Quebec in 1981; 0 otherwise
Ont - 1 if person lived in Ontario in 1981; 0 otherwise
Pray - 1 if person lived in the Prairie provinces in 1981; 0 otherwise
BC - 1 if person lived in British Columbia in 1981; 0 otherwise
Not que - 1 if person did not live in Quebec in 1981; 0 otherwise
Not educ - 1 if person has no post-secondary education

Note that "Not Que" is the converse of "Que" and that "Not educ" is the converse of "Educ"
While these variables may seem a bit redundant, they are actually very useful for the specification of interaction variables.

DATA SET #2

Having seen what this file looks like for grouped data, we now consider a file for a problem with micro-data. These data are adapted from Ben-Akiva and Lerman (1985) and depict the choice of transit modes: either automobile or public transit. In this data we have no choice-maker characteristic such as age or income so the file contains only choice information.

BENSOC.TXT

Auto	Transit
0	1
0	1
1	0
0	1
0	1
1	0
1	0
0	1
0	1
0	1
0	1
1	0
1	0
0	1
1	0
1	0
0	1
1	0
1	0
0	1
1	0

Note that there are 21 rows which represents the fact that there are 21 choice-makers in the data. In contrast to data set #1, the choice information is binary since there is not a distribution of observed choices as is the case with grouped data. The two columns correspond to the number of alternatives with a "1" designating which alternative was chosen by the person.

APPENDIX 3: DESTINATION ATTRIBUTE DATA

DATA SET #1

These data illustrate the set of destination attributes used for regional level analysis. These data are used in sample estimations #1 and 2. It is a simple file consisting of six rows (one for each aggregate and six columns (one for each destination attribute). Note that none of the row and column labels appear in the actual data. These are included here simply to define the data. The first row of zeros associated with the stay alternative is present because the variables have no meaning for that alternative.

DEST.TXT

	english	french	univ	income	temp	precip
Stay	0.000	0.000	0.00	0.0	0.00	0.00
East	0.879	0.121	0.07	39.9	5.35	1.16
Quebec	0.056	0.944	0.06	39.0	4.57	1.01
Ontario	0.991	0.009	0.07	46.4	6.55	0.87
Prairies	0.999	0.001	0.07	43.4	1.87	0.43
BC	1.000	0.000	0.06	44.1	8.01	1.04

The first column, which lists the aggregate alternatives should be self-explanatory. The variable definitions are as follows:

english - proportion of the population that speaks english only
french - proportion of the population that speaks french or is bilingual
univ - proportion of the population with a university degree
income - average income annual income in thousands of dollars
temp - mean annual temperature in degrees celsius
precip - average annual precipitation in thousands of millimetres

DATA SET #2

Here we examine a more disaggregate destination attribute set but one which contains the same variables. These are the data which are used in the application of the aggregated logit model. Note that the destination attributes for the non-spatial stay alternative are still depicted as a row of zeros and that aggregates are shaded alternately.

DESTELEM.TXT

	english	french	univ	income	temp	precip
Stay	0.000	0.000	0.00	0.0	0.00	0.00
St. Johns	1.000	0.000	0.09	47.67	5.50	1.42
Grand Falls	1.000	0.000	0.04	39.19	4.60	0.99
Cornerbrook	1.000	0.000	0.06	41.95	5.10	1.13
Charlottetown	0.999	0.001	0.09	41.78	5.40	1.17
Summerside	0.995	0.005	0.05	37.17	5.50	1.04
Halifax	0.999	0.001	0.13	46.79	7.60	1.28
Kentville	1.000	0.000	0.08	36.93	6.80	1.18
Truro	1.000	0.000	0.07	37.29	5.30	1.14

New Glasgow	1.000	0.000	0.05	38.06	5.60	1.13
Sydney	1.000	0.000	0.05	35.54	5.70	1.40
Moncton	0.936	0.064	0.08	43.15	5.50	1.10
Saint John	0.998	0.002	0.07	42.01	5.70	1.34
Fredericton	0.992	0.008	0.15	45.63	5.40	1.11
Bathurst	0.498	0.502	0.06	37.15	4.40	1.00
Campbellton	0.565	0.435	0.05	35.36	4.10	1.05
Edmundston	0.026	0.974	0.07	36.75	3.80	1.12
other Atlantic	0.930	0.070	0.04	35.89	5.00	1.10
Rimouski	0.000	1.000	0.09	40.26	3.70	0.83
Riviere du Loup	0.000	1.000	0.06	37.92	3.00	0.88
Baie Comeau	0.001	0.999	0.05	47.40	2.00	1.07
Chicoutimi	0.002	0.998	0.06	41.26	3.40	0.95
Alma	0.001	0.999	0.06	40.52	5.30	1.39
Sept-Iles	0.017	0.983	0.05	41.90	1.10	1.13
Quebec City	0.004	0.996	0.11	42.18	4.60	1.13
Saint Georges	0.001	0.999	0.05	35.75	4.10	0.92
Thetford	0.002	0.998	0.04	35.81	3.60	1.15
Sherbrooke	0.036	0.964	0.09	36.61	5.90	1.00
Magog	0.025	0.975	0.06	35.72	5.50	1.11
Victoriaville	0.001	0.999	0.05	36.66	4.90	1.07
Trois Rivieres	0.002	0.998	0.07	37.30	4.70	1.03
Shawinigan	0.001	0.999	0.04	32.70	4.50	1.06
Drummondville	0.001	0.999	0.05	35.12	5.70	1.00
Granby	0.013	0.987	0.05	38.13	5.90	1.14
St. Hyacinthe	0.001	0.999	0.06	37.16	6.20	1.03
Sorel	0.001	0.999	0.04	38.94	5.80	0.96
Joliette	0.001	0.999	0.06	38.52	5.20	0.88
St. Jean	0.016	0.984	0.06	38.72	6.50	1.00
Montreal	0.191	0.809	0.11	43.41	6.50	1.00
Sallaberry de Valle	0.008	0.992	0.04	37.12	6.50	1.00
St. Jerome	0.001	0.999	0.04	36.46	4.60	1.03
Val d'or	0.015	0.985	0.05	41.06	4.30	0.92
Noranda	0.016	0.984	0.06	42.80	0.90	0.87
Hull	0.183	0.817	0.10	47.88	5.90	0.85
Pembroke	0.996	0.004	0.06	39.01	5.20	0.77
other quebec	0.027	0.973	0.03	35.32	5.50	1.00
Cornwall	0.957	0.043	0.05	39.41	6.60	0.93
Ottawa	0.965	0.035	0.18	56.47	5.90	0.85
Brockville	0.999	0.001	0.07	44.87	7.10	0.97
Kingston	0.998	0.002	0.12	47.10	6.70	0.87
Belleville	0.998	0.002	0.06	44.74	7.40	0.86
Cobourg	1.000	0.000	0.07	44.39	6.70	0.82
Peterborough	1.000	0.000	0.08	44.30	5.80	0.79
Lindsay	1.000	0.000	0.06	42.20	9.20	0.86
Oshawa	0.998	0.002	0.06	55.46	7.70	0.86
Toronto	0.999	0.001	0.13	59.45	7.90	0.83
Hamilton	0.999	0.001	0.09	50.41	7.40	0.82
St. Catherines-Niag	0.997	0.003	0.07	44.93	8.90	0.81
Kitchener	0.999	0.001	0.09	49.98	7.30	0.90
Brantford	1.000	0.000	0.06	43.97	7.60	0.75
Woodstock	1.000	0.000	0.05	43.13	7.30	0.86
Simcoe	0.999	0.001	0.05	43.35	7.70	0.89
Guelph	0.999	0.001	0.13	50.66	6.70	0.83

Stratford	1.000	0.000	0.07	43.86	6.30	1.05
London	1.000	0.000	0.11	47.69	7.30	0.91
Chatham	0.999	0.001	0.07	42.50	9.10	0.81
Leamington	0.999	0.001	0.05	45.32	9.30	0.82
Windsor	0.998	0.002	0.08	46.79	9.40	0.85
Sarnia	0.999	0.001	0.07	50.01	8.00	0.89
Owen Sound	1.000	0.000	0.07	40.60	6.90	1.02
Barrie	0.999	0.001	0.07	51.51	5.80	0.95
Orillia	1.000	0.000	0.05	43.63	5.60	0.91
Midland	0.994	0.006	0.04	43.85	7.00	1.03
North Bay	0.986	0.014	0.07	43.73	4.70	0.93
Sudbury	0.969	0.031	0.07	48.27	4.00	0.79
Elliott lake	0.952	0.048	0.05	44.01	4.10	0.93
Timmins	0.918	0.082	0.05	45.12	1.10	0.85
Sault Ste. Marie	0.999	0.001	0.07	41.89	4.40	0.90
Thunder Bay	0.998	0.002	0.08	48.17	2.30	0.71
Kenora	1.000	0.000	0.06	46.89	2.10	0.62
other Ontario	0.983	0.017	0.05	43.93	8.00	0.90
Winnipeg	0.998	0.002	0.10	42.65	2.40	0.53
Brandon	1.000	0.000	0.08	36.54	1.70	0.45
Thompson	1.000	0.000	0.06	53.75	3.90	0.54
Regina	0.999	0.001	0.10	45.92	2.20	0.38
Yorkton	1.000	0.000	0.05	35.22	1.10	0.43
Moose Jaw	0.999	0.001	0.05	37.92	3.50	0.38
Saskatoon	1.000	0.000	0.11	41.76	1.80	0.35
North Battleford	0.999	0.001	0.06	36.40	1.60	0.36
Prince Albert	0.999	0.001	0.06	39.36	0.10	0.40
other Prairie	0.999	0.001	0.04	35.18	1.00	0.45
Medcine Hat	1.000	0.000	0.05	40.60	5.10	0.35
Lethbridge	1.000	0.000	0.09	42.18	5.10	0.42
Calgary	1.000	0.000	0.13	52.64	3.50	0.43
Red Deer	1.000	0.000	0.06	43.70	2.60	0.46
Edmonton	0.999	0.001	0.10	47.37	2.40	0.47
Lloydminster	0.999	0.001	0.05	42.54	1.50	0.43
Grande Centre	0.991	0.009	0.05	44.70	2.10	0.47
Grande Prairie	1.000	0.000	0.07	46.98	1.20	0.45
Fort McMurray	0.999	0.001	0.06	60.48	-0.20	0.47
other Alberta	0.999	0.001	0.05	41.50	2.50	0.45
Cranbrook	1.000	0.000	0.04	40.10	5.50	0.37
Penticton	1.000	0.000	0.06	37.31	8.90	0.28
Kelowna	0.999	0.001	0.06	41.69	8.00	0.33
Vernon	1.000	0.000	0.05	39.35	7.90	0.35
Kamloops	1.000	0.000	0.06	44.44	8.70	0.24
Chilliwack	1.000	0.000	0.04	40.55	10.20	1.88
Matsqui	1.000	0.000	0.05	46.67	10.20	1.20
Vancouver	1.000	0.000	0.12	50.57	10.30	1.22
Victoria	1.000	0.000	0.12	45.56	10.00	0.87
Duncan	1.000	0.000	0.06	42.34	8.00	1.64
Nanaimo	1.000	0.000	0.06	41.59	10.00	1.13
Port Alberni	0.999	0.001	0.04	44.36	9.20	2.02
Courtenay	0.999	0.001	0.05	40.82	8.70	1.50
Campbell River	1.000	0.000	0.05	49.52	8.70	1.66
Powell River	0.999	0.001	0.05	42.17	10.50	1.09
Williams Lake	1.000	0.000	0.04	42.13	3.90	0.41

Quesnel	1.000	0.000	0.04	42.36	5.00	0.53
Prince Rupert	1.000	0.000	0.05	51.86	7.40	2.40
Terrace	0.999	0.001	0.05	50.96	5.90	1.31
Prince George	1.000	0.000	0.05	49.06	3.30	0.63
other BC	1.000	0.000	0.05	42.12	8.00	0.80

APPENDIX 4: DISTANCE DATA

The text files displayed below are the ones which would be used to make the GAUSS data set DIST.DAT. It will contain all variables representing physical distance, travel cost and time etc. In the sample estimations on grouped data to follow, two distance variables are used. For the example with the OMNL and the nested logit model, there are a total of six spatial alternatives while for the example with the aggregated model there are 122 spatial alternatives. The structure of the distance file is such that the 60 by 244 distance data cannot be shown here for the aggregated model, only for the former two models. Also, the distance file for the Ben-Akiva example is shown. Please recall that the displayed headers and lines will not actually appear in the ASCII text file provided by the user.

DATA SET #1

DIST.TXT

dist						indist					
stay	east	que	ont	pray	bc	stay	east	que	ont	pray	bc
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	6.70	[illegible]	7.86
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	7.86
0	0	0.53	0.89	2.13	2.86	0	0	[illegible]	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	[illegible]	[illegible]	[illegible]	[illegible]
0	0	0.53	0.89	2.13	2.86	0	0	6.28	[illegible]	[illegible]	[illegible]
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78
0	0.53	0	0.36	1.68	2.4	0	6.28	0	5.89	7.43	7.78

0	0.89	0.36	0	1.44	2.14	0	6.79	5.89	0	7.27	7.67
0	0.89	0.36	0	1.44	2.14	0	6.79	5.89	0	7.27	7.67
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.13	1.68	1.44	0	0.73	0	7.67	7.43	7.27	0	6.6
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0
0	2.86	2.4	2.14	0.73	0	0	7.96	7.78	7.67	6.6	0

Variable Definitions

dist - distance from associated orgin to given region in thousands of km
lndist - natural logarithm of distance in km from associated origin to given destination region

There are several things to note about the set up of this data. The first is that the number of rows is 60 which corresponds with the number of choice-makers. The total number of columns is twice the model's number of elemental alternatives because there are two distance variables. Each variable has its own block. The rows repeat themselves in blocks of 12 because for each block, the origin is the same although other choice-maker characteristics differ within each block. The identity of each origin can be gathered in this example by seeing which destination choice option has zero distance for a given population segment. Finally, note that the column of zeros which occupies the first column in each block has this value because the stay option is associated with zero distance. Nevertheless, this alternative, like all others, must be accounted for in the text file.

As mentioned, the distance text file for an application of the aggregated logit model would have the same number of rows but 244 columns. The columnar contents of one variable within this file, that is 122 columns, would be as follows: column 1 - the stay option, columns 2 to 18 - distances to elemental alternatives within the Atlantic provinces, columns 19 to 46 - distances to elemental alternatives within Quebec, columns 47 to 81 - distances to elemental alternatives within Ontario, columns 82 to 101 - distances to elemental alternatives within the prairies and columns 102 to 122 - distances to elemental alternatives within British Columbia.

As an aside, it is a property of the way GAUSS reads ASCII data that the actual text files do not need to exceed 80 columns. If the user specifies that data are to be read into a matrix with 600 columns, then the data from the 80 column file will simply be read sequentially, on a row-by-row basis so that 600 column rows are formed.

DATA SET #2

Here we consider a distance text file for a problem with individual level micro-data and we note that there is little difference from the first data set. The distance data for the Ben-Akiva problem to be illustrated later is actually travel time data in a mode choice problem. Hence, in this example, unlike the previous one there is only one block of data since there is only one variable. It contains two columns because there are two aggregate alternatives in the model (auto and transit) both of which are not sub-divided and hence are composed of one elemental alternative.

time	
auto	**transit**
52.9	4.4
4.1	28.5
4.1	86.9
56.2	31.6
51.8	20.2
0.2	91.2
27.6	79.7
89.9	2.2
41.5	24.5
95.0	43.5
99.1	8.4
18.5	84.0
82.0	38.0
8.6	1.6
22.5	74.1
51.4	83.8
81.0	19.2
51.0	85.0
62.2	90.1
95.1	22.2
41.6	91.5

It is probably for small problems like this that the idea of storing the different data types in different files becomes the most cumbersome. This software however, is fundamentally designed for spatial choice scenarios where the number of potential alternatives can be quite large. It is felt though that the data file format of this software should not be a substantial handicap for any application.

APPENDIX 5: SAMPLE ESTIMATIONS

SAMPLE PROBLEM #1

For this sample problem, we utilize the grouped choicemaker data illustrated in APPENDIX 2, the aggregate level destination attribute data in APPENDIX 3 and the aggregate level distance data shown in APPENDIX 4. To utilize these data in a specification, we would need to define an appropriate DEFINE.DAT file through running SETDATA. The correct ASCII text file for the OMNL and the nested logit model appears as below. There are 60 rows which corresponds to the # of population segments and there are six columns to correspond with the number of alternatives. Note that there are zeros where the associated alternative is unavailable to that segment.
For example, in the first block, the east destination choice alternative is unavailable to people in the Atlantic provinces because such choices are captured by the stay option.

DEFINE.TXT

1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	0	1	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	0	1	1
1	1	1	1	0	1
1	1	1	1	0	1

1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	0	1
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0

An OMNL specification using this data and all the different variable types is shown below. Note that the first two lines are concerned with the distribution of alternatives and their names, while the remaining block provides parameter-specific information. The variable types are: temp (destination attribute), young and single (choice-maker traits), lang and postsec (interactions), and dist and lndist (distance variables). The variable stay is an alternative-specific constant. Note that order of variable types does not matter. Each variable will be applied to each alternative which is associated with a 1 in the design matrix. The two interaction variables are formed from the variables in the choice-maker and destination attribute data. The binary choice-maker variables and their associated destination attributes are as follows:

lang

choice-maker trait:	east	que	ont	pray	bc
destination attribute:	english	french	english	english	english

postsec

choice-maker trait:	educ	noteduc
destination attribute:	univ	zeros

OMNL.SPC

```
1 1 1 1 1 1
stay east que ont pray bc

0 temp      0 1 1 1 1 1
0 young     1 0 0 0 0 0
0 single    1 0 0 0 0 0
0 lang      0 1 1 1 1 1
0 postsec   0 1 1 1 1 1
5 stay      1 0 0 0 0 0
0 dist      0 1 1 0 1 1
0 lndist    0 0 0 1 0 0
```

Remember also that the first column of parameter-specific information contains the starting values to be used in the estimation. Now below we see the results of this specification when it is processed by the program SPEC. These are written to SUMMARY.OUT:

```
Here are the details of the specification as you have described it..

This is an ordinary multinomial logit model

The alternatives are:
   stay   east   que   ont   pray   bc
The distribution of elemental alternatives is: 1 1 1 1 1 1
There is/are 8 beta parameter(s) in the model

*********************************************************************
There is/are 1 alternative-specific constant(s) to be estimated
The design matrix is..
 1 0 0 0 0 0

*********************************************************************
There is/are 2 socio-economic parameter(s) to be estimated
The associated choice-maker variable(s) is/are:
   young   single
The design matrix is..
 1 0 0 0 0 0
 1 0 0 0 0 0

*********************************************************************
There is/are 1 destination attribute parameter(s) to be estimated
The associated destination attribute(s) are:
    temp
The design matrix is..
 0 1 1 1 1 1

*********************************************************************
There is/are 2 distance parameters to be estimated
The associated distance variable(s) is/are
    dist   lndist

The design matrix is..
 0 1 1 0 1 1
 0 0 0 1 0 0

*********************************************************************
There is/are 2 interaction parameter(s) to be estimated
The associated interaction term(s) is/are:
    lang   postsec
The design matrix is..
 0 1 1 1 1 1
 0 1 1 1 1 1
^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
```

Having confirmed that the specification looks fine. The program MODULE is run and for this specification, it will provide the following output:

```
==============================================================================
MAXLIK Version 3.1.3                              4/26/95  9:37 pm
==============================================================================
                    Data Set:  speck
--------------------------------------------------------------


return code =   0
normal convergence

Mean log-likelihood       -529.666
Number of cases    60

Covariance matrix of the parameters computed by the following method:
Estimated Hessian from the secant update

Parameters  Estimates    Std. err.  Est./s.e.  Prob.   Gradient
------------------------------------------------------------------
stay          5.3727     0.0478  112.484   0.0000    0.0007
young        -1.0088     0.0281  -35.918   0.0000   -0.0002
single       -0.2131     0.0265   -8.051   0.0000    0.0002
temp         -0.0077     0.0069   -1.123   0.1308   -0.0024
dist         -0.6019     0.0227  -26.552   0.0000   -0.0010
lndist        0.0120     0.0060    1.997   0.0229   -0.0008
lang          1.0278     0.0321   32.017   0.0000   -0.0005
postsec       9.1500     0.3721   24.590   0.0000   -0.0000

Correlation matrix of the parameters
  1.000  -0.113  -0.186   0.645   0.302   0.091   0.315   0.280
 -0.113   1.000  -0.130   0.047   0.053   0.032  -0.074   0.028
 -0.186  -0.130   1.000  -0.038   0.011   0.043  -0.036  -0.041
  0.645   0.047  -0.038   1.000  -0.134  -0.368   0.060   0.061
  0.302   0.053   0.011  -0.134   1.000   0.746  -0.339  -0.081
  0.091   0.032   0.043  -0.368   0.746   1.000  -0.360  -0.115
  0.315  -0.074  -0.036   0.060  -0.339  -0.360   1.000   0.059
  0.280   0.028  -0.041   0.061  -0.081  -0.115   0.059   1.000

Number of iterations   44
Minutes to convergence    0.45950

The final set of lambda values across the aggregates is..
    1.000    1.000    1.000    1.000    1.000    1.000

Naive log-likelihood..-315299.497
Log-likelihood with constants..-34738.789
Log-likelihood at convergence..-31779.951
Rho-squared..    0.899
Rho-squared with constant..    0.085
Adjusted rho-squared (without constants)..    0.899
```

Likelihood ratio statistic using L(0) is..567039
and has 8 degrees of freedom

Likelihood ratio statistic using L(c) is ..5918
and has 3 degrees of freedom

Expected percent right.. 93.501
*^^^

Note that the results for the variables are presented in a different order from the manner in which they were specified. This is because the variables have been sorted by type before the estimation took place.

SAMPLE PROBLEM #2

This problem utilizes exactly the same data but applies a different model. We will utilize a two level nested logit model where the upper level is move/stay and the lower level contains the destination choice options. The definition of aggregates changes however the total number of elemental alternatives has not changed. Hence DEFINE.TXT from the previous example still applies. The specification file for this model would be as follows:

NEST1.SPC

```
1 5
stay move
stay east que ont pray bc

5 stay       1 0 0 0 0 0
0 young      1 0 0 0 0 0
0 single     1 0 0 0 0 0
0 temp       0 1 1 1 1 1
0 dist       0 1 1 0 1 1
0 lndist     0 0 0 1 0 0
0 lang       0 1 1 1 1 1
0 postsec    0 1 1 1 1 1
```

This design matrix is actually identical to that of OMNL.SPC except that the parameter order differs. As mentioned though, the order in which variables appear in an .SPC file is irrelevant. Here is how the SUMMARY.OUT file looks:

```
Here are the details of the specification as you have described it..

This is a nested logit model

The upper level alternatives are:
   stay    move

The lower level alternatives are:
   stay    east    que    ont    pray    bc

The distribution of elemental alternatives is:  1  5
There is/are  8 beta parameter(s) in the model

****************************************************************************
There is/are  1 alternative-specific constant(s) to be estimated
The design matrix is..
 1  0  0  0  0  0

****************************************************************************
There is/are  2 socio-economic parameter(s) to be estimated
The associated choice-maker variable(s) is/are:
   young   single
The design matrix is..
 1  0  0  0  0  0
 1  0  0  0  0  0
```

```
*********************************************************************
There is/are  1 destination attribute parameter(s) to be estimated
The associated destination attribute(s) are:
    temp
The design matrix is..
 0 1 1 1 1 1

*********************************************************************
There is/are  2 distance parameters to be estimated
The associated distance variable(s) is/are
    dist   lndist

The design matrix is..
 0 1 1 0 1 1
 0 0 0 1 0 0

*********************************************************************
There is/are  2 interaction parameter(s) to be estimated
The associated interaction term(s) is/are:
    lang   postsec
The design matrix is..
 0 1 1 1 1 1
 0 1 1 1 1 1
*********************************************************************
```

And finally, we observe the results of the estimation:

```
==========================================================================
 MAXLIK Version 3.1.3                          4/26/95   7:36 pm
==========================================================================
                      Data Set:  speck
--------------------------------------------------------------------------

return code =    0
normal convergence

Mean log-likelihood      -528.681
Number of cases    60

Covariance matrix of the parameters computed by the following method:
Estimated Hessian from the secant update

Parameters   Estimates    Std. err.  Est./s.e.  Prob.    Gradient
------------------------------------------------------------------
stay         6.8369       0.2013   33.963   0.0000   -0.0001
young        -0.9972      0.0257  -38.850   0.0000    0.0000
single       -0.2087      0.0248   -8.427   0.0000    0.0000
temp         -0.0312      0.0174   -1.788   0.0369    0.0006
dist         -1.3763      0.1292  -10.653   0.0000    0.0001
lndist       0.0744       0.0134    5.536   0.0000    0.0006
lang         1.3457       0.0468   28.746   0.0000    0.0001
postsec      9.5975       0.3999   23.999   0.0000    0.0000
lambda       2.4857       0.2195   11.325   0.0000    0.0001
```

Correlation matrix of the parameters

1.000	0.044	-0.079	0.041	-0.656	0.436	0.425	0.168	0.860
0.044	1.000	-0.013	-0.028	-0.082	0.044	0.122	-0.052	0.085
-0.079	-0.013	1.000	0.017	0.007	-0.018	-0.091	-0.040	-0.023
0.041	-0.028	0.017	1.000	0.313	-0.541	-0.071	0.112	-0.369
-0.656	-0.082	0.007	0.313	1.000	-0.133	-0.482	-0.064	-0.898
0.436	0.044	-0.018	-0.541	-0.133	1.000	-0.101	-0.081	0.459
0.425	0.122	-0.091	-0.071	-0.482	-0.101	1.000	0.053	0.412
0.168	-0.052	-0.040	0.112	-0.064	-0.081	0.053	1.000	0.057
0.860	0.085	-0.023	-0.369	-0.898	0.459	0.412	0.057	1.000

Number of iterations 74
Minutes to convergence 1.60733

The final set of lambda values across the aggregates is..
1.000 2.486

Naive log-likelihood..-315299.497
Log-likelihood with constants..-34738.789
Log-likelihood at convergence..-31720.842
Rho-squared.. 0.899
Rho-squared with constant.. 0.087
Adjusted rho-squared (without constants).. 0.899

Likelihood ratio statistic using L(0) is..567157
and has 9 degrees of freedom

Likelihood ratio statistic using L(c) is ..6036
and has 4 degrees of freedom

Expected percent right.. 93.511
****

Note that we can see the set of lambda values (both endogneous and exogenous) which are applying across the aggregates at the time of convergence. It is important to check this to make sure that the lambda specification took place as planned.

SAMPLE PROBLEM #3

Here a sample estimation of the aggregated logit model is illustrated. For this problem, our definition of an elemental alternative has changed and now represents cities within the regions we were modelling earlier as elemental alternatives. We are still using the same SOCIO.TXT file since we have no knowledge of city-level choices however we must utilize more spatially detailed information in regards to the other data files. The destination attribute file is DESTELEM.TXT, the distance information is from DISTELEM.TXT and the matrix of available alternatives is found in DEFELEM.TXT. Due to the large size of these files, only the first of these three is found in the Appendices, in particular APPENDIX 3.

AGG1.SPC

```
1 17 28 35 20 21
stay east que ont pray bc

5 stay      1 0 0 0 0 0
0 young     1 0 0 0 0 0
0 single    1 0 0 0 0 0
0 temp      0 1 1 1 1 1
0 dist      0 1 1 0 1 1
0 lndist    0 0 0 1 0 0
0 lang      0 1 1 1 1 1
0 postsec   0 1 1 1 1 1
```

Again, there is no difference in the design matrix. It cannot be made more spatially detailed since the spatial resolution of choice information is no higher. Each time we see a 1 in the design matrix though, we must remember that this applies to the associated aggregate and that each elemental alternative within the aggregate will be affected by the associated parameter. Note how the first row in AGG1.SPC, associated with the distribution of elemental alternatives, differs from that of OMNL.SPC. We have the same number aggregate alternatives but from the modelling perspective, what they contain is quite different. Here is what the SUMMARY.OUT file will look like with this specification. Overall, not too much different.

Here are the details of the specification as you have described it..

```
This is an aggregated logit model

The alternatives are:
   stay   east   que   ont   pray   bc
The distribution of elemental alternatives is:  1 17 28 35 20 21
There is/are  8 beta parameter(s) in the model

****************************************************************************
There is/are  1 alternative-specific constant(s) to be estimated
The design matrix is..
 1  0  0  0  0  0

****************************************************************************
There is/are  2 socio-economic parameter(s) to be estimated
The associated choice-maker variable(s) is/are:
   young   single
```

The design matrix is..
1 0 0 0 0 0
1 0 0 0 0 0

****

There is/are 1 destination attribute parameter(s) to be estimated
The associated destination attribute(s) are:
temp
The design matrix is..
0 1 1 1 1 1

****

There is/are 2 distance parameters to be estimated
The associated distance variable(s) is/are
dist lndist

The design matrix is..
0 1 1 0 1 1
0 0 0 1 0 0

****

There is/are 2 interaction parameter(s) to be estimated
The associated interaction term(s) is/are:
lang postsec
The design matrix is..
0 1 1 1 1 1
0 1 1 1 1 1
****

Now we see the results of the estimation. Note how it takes much longer to obtain a convergence since we are continually estimating 122 elemental utilities as opposed to the 6 that we did previously.

==
MAXLIK Version 3.1.3 4/27/95 1:56 am
==
Data Set: speck
----

return code = 0
normal convergence

Mean log-likelihood -527.967
Number of cases 60

Covariance matrix of the parameters computed by the following method:
Estimated Hessian from the secant update

Parameters	Estimates	Std. err.	Est./s.e.	Prob.	Gradient
stay	12.1878	0.3475	35.076	0.0000	-0.0001
young	-1.0115	0.0274	-36.926	0.0000	-0.0001
single	-0.2169	0.0271	-7.997	0.0000	-0.0001
temp	-0.0195	0.0072	-2.702	0.0034	0.0006
dist	-0.5226	0.0214	-24.468	0.0000	-0.0000
lndist	-0.1330	0.0099	-13.411	0.0000	0.0004
lang	1.3877	0.0376	36.924	0.0000	-0.0000
postsec	9.3446	0.3649	25.611	0.0000	-0.0000
lambda	2.1398	0.1074	19.929	0.0000	0.0003

Correlation matrix of the parameters

1.000	-0.018	-0.058	0.088	0.103	-0.801	0.620	0.140	0.990
-0.018	1.000	-0.115	-0.004	0.019	0.004	0.013	0.040	0.004
-0.058	-0.115	1.000	-0.007	-0.005	0.030	-0.005	-0.002	-0.036
0.088	-0.004	-0.007	1.000	-0.098	-0.201	-0.036	0.070	-0.004
0.103	0.019	-0.005	-0.098	1.000	0.359	-0.191	0.013	0.057
-0.801	0.004	0.030	-0.201	0.359	1.000	-0.625	-0.131	-0.815
0.620	0.013	-0.005	-0.036	-0.191	-0.625	1.000	0.064	0.601
0.140	0.040	-0.002	0.070	0.013	-0.131	0.064	1.000	0.098
0.990	0.004	-0.036	-0.004	0.057	-0.815	0.601	0.098	1.000

Number of iterations 73
Minutes to convergence 12.98250

The final set of lambda values across the aggregates is..
1.000 2.140 2.140 2.140 2.140 2.140

Naive log-likelihood..-315299.497
Log-likelihood with constants..-34738.789
Log-likelihood at convergence..-31678.031
Rho-squared.. 0.900
Rho-squared with constant.. 0.088
Adjusted rho-squared (without constants).. 0.900

Likelihood ratio statistic using L(0) is..567242.93
and has 9.00 degrees of freedom

Likelihood ratio statistic using L(c) is ..6121.52
and has 4.00 degrees of freedom

Expected percent right.. 93.508

SAMPLE ESTIMATION #4

Up to this point, all the estimations have been with grouped data where individual choice-makers have been grouped into homogeneous population segments. The purpose of this sample estimation is two fold: 1) to show that estimation with individual level micro-data is not really different from the perspective of the software and 2) to show that the software is equally applicable to non-spatial choice problems. To do this, we replicate the results from the small sample problem shown in Ben-Akiva and Lerman (1985). There are some points to note about the required data. First of all, we need no external ASCII file for this problem in order to create DEFINE.DAT, this will be done automatically in SETDATA since all alternatives are available to all choice-makers. The required choice-maker file to make SOCIO.DAT is found in APPENDIX 1 under the name BENSOC.TXT. In that file there are no choice-maker variables, only choice information. This is perfectly acceptable. There is no destination attribute file required for this problem while the required DIST.DAT file is made from BENDIST.DAT found in APPENDIX 3. The data in this file are not distances per se, but travel times. The alternatives in this problem are not spatial units but modes of transportation.

BEN.SPC

```
      1    1
    car   transit

.1   const     1 0
-.05 time      1 1
```

Note that the .SPC file for this problem is quite small given that there are only two parameters and two alternatives. We have a generic variable named time which encompasses transit travel time and automobile travel time. Typical specifications would have treated these variables as separate but associated with the same parameter. There is no circumstance in this software however, where more than 1 variable is associated with the same parameter.

The SUMMARY.OUT file looks like this:

Here are the details of the specification as you have described it..

```
This is an ordinary multinomial logit model

The alternatives are:
    car  transit
The distribution of elemental alternatives is:  1  1
There is/are  2 beta parameter(s) in the model

*************************************************************************
There is/are  1 alternative-specific constant(s) to be estimated
The design matrix is..
 1 0

*************************************************************************
There is/are  0 socio-economic parameter(s) to be estimated
*************************************************************************
There is/are  0 destination attribute parameter(s) to be estimated
*************************************************************************
```

There is/are 1 distance parameters to be estimated
The associated distance variable(s) is/are
time

The design matrix is.. 1 1

****
There is/are 0 interaction parameter(s) to be estimated
****

Finally, the results of estimation are as follows:

```
==============================================================================
 MAXLIK Version 3.1.3                          4/26/95  10:04 pm
==============================================================================
                  Data Set:  speck
------------------------------------------------------------------------------

return code =    0

Mean log-likelihood      -0.293621
Number of cases    21

Covariance matrix of the parameters computed by the following method:
Estimated Hessian from the secant update

Parameters   Estimates    Std. err.  Est./s.e.  Prob.    Gradient
--------------------------------------------------------------------
const        -0.2376      0.7604   -0.312   0.3773     0.0000
time         -0.0531      0.0207   -2.571   0.0051    -0.0000

Correlation matrix of the parameters
  1.000  0.156
  0.156  1.000

Number of iterations    7
Minutes to convergence    0.01367

The final set of lambda values across the aggregates is..
    1.000    1.000

Naive log-likelihood..  -14.556
Log-likelihood with constants..  -14.532
Log-likelihood at convergence..  -6.166
Rho-squared..   0.576
Rho-squared with constant..   0.576
Adjusted rho-squared (without constants)..    0.439
Likelihood ratio statistic using L(0) is..  17
and has  2  degrees of freedom

Likelihood ratio statistic using L(c) is ..  17
and has  1  degrees of freedom

Expected percent right.. 82.809
```

Bibliography

[1] Amemiya, T. (1978). 'On a two-step estimation of a multivariate logit model', *Journal of Econometrics*, 8, 13-21.

[2] Amrhein, C.G. (1995). 'Searching for the elusive aggregation effect: evidence from statistical simulations', *Environment and Planning A*, 27, 105-119.

[3] Amrhein, C.G. and R. Flowerdew (1992). 'The effect of data aggregation on a Poisson regression model of Canadian migration', *Environment and Planning A*, 24, 1381-1391.

[4] Anas, A. (1981). 'The estimation of multinomial logit models of joint location and travel mode choice from aggregated data', *Journal of Regional Science*, 21(2), 223-42.

[5] Anas, A. (1982). *Residential location, markets and urban transportation.* New York: Academic Press.

[6] Anas, A. (1983). 'Discrete choice theory, information theory and the multinomial logit and gravity models', *Transportation Research B*, 17, 13-23.

[7] Anderson, W.P. and Y.Y. Papageorgiou (1994). 'An analysis of migration streams for the Canadian regional system, 1952-1983: 1. Migration Probabilities', *Geographical Analysis*, 26, 15-36.

[8] Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer.

[9] Arbia, G. (1989). *Spatial data configuration in statistical analysis of regional economic and related problems.* Dordrecht: Kluwer.

[10] Avriel, M. (1976). *Nonlinear Programming: Analysis and Methods.* Englewood Cliffs, New Jersey: Prentice-Hall.

[11] Batty, M. and P.K. Sikdar (1982a).'Spatial aggregation in gravity models: 1. an information theoretic approach', *Environment and Planning A*, 14, 377-405.

[12] Batty, M. and P.K. Sikdar, (1982b). 'Spatial aggregation in gravity models: 2. One dimensional population density models', *Environment and Planning A*, 14, 525-553.

[13] Batty, M. and P.K. Sikdar (1982c). 'Spatial aggregation in gravity models: 3. two-dimensional trip distribution and location models', *Environment and Planning A*, 14, 629-658.

[14] Batty, M. and P.K. Sikdar (1982d), 'Spatial aggregation in gravity models: 4. generalisations and large scale applications', *Environment and Planning A*, 14, 795-822.

[15] Ben-Akiva, M. and T. Watanatada (1981). 'Application of a Continuous Spatial Choice Logit Model' in *Structural Analysis of Discrete Data*, (C. Manski and D. McFadden eds.), 321-43, Cambridge, Mass.: MIT Press.

[16] Ben-Akiva, M. N. Litinas and K. Tsunokawa (1985). 'Continuous Spatial Choice: the continuous logit model and distributions of trips and urban densities', *Transportation Research A*, 19(2), 119-54.

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

[18] Ben-Akiva, M.E. and D. Bolduc (1991). *Multinomial Probit with Autoregressive Error Structure*, Cahier de Recherche no. 91-22, Groupe de Recherche en Economie de l'Energie et des Resources Naturelles (GREEN), Universite Laval, Quebec, Canada, G1K 7P4.

[19] Berndt, E.U., B.H. Hall, R.E. Hall and J.A. Hausman (1974). 'Estimation and inference in non-linear structural models', *Annals of Economic and Social Measurement*, 3/4, 653-65.

[20] Black, W. (1983). 'A generalization of destination effects in spatial interaction modelling', *Economic Geography*, 59(1), 16-34.

[21] Bolduc, D. (1994). *A Practical Technique to Estimate Multinomial Probit Models in Transportation: Computational Details and an Application to a Dissaggregate Mode Choice Problem*, Serie de Document de Travail du CREST, No 9421, Institute National de la Statistique et des Etudes Economique.

[22] Bolduc, D., M.G.Dagenais and M.J.I. Gaudry (1989). 'Spatially Autocorrelated Errors in Origin-Destination Models: A New Specification Applied to Aggregate Mode Choice', *Transportation Research B*, 23B, 361-372.

[23] Boots, B.N. and P.S. Kanaroglou (1988). 'Incorporating the effects of spatial structure in discrete choice models of migration', *Journal of Regional Science*, 28, 495-507.

[24] Borgers, A. and H. Timmermans (1987). 'Choice model specification, substitution and spatial structure effects', *Regional Science and Urban Economics*, 17, 29-47.

[25] Borgers, A., R. van der Heijdena and H. Timmermans (1989). 'A variety seeking model of spatial choice behaviour', *Environment and Planning A*, 21, 1037-1048.

[26] Bouthelier, F. and C. Daganzo (1979). 'Aggregation with multinomial probit and estimation of disaggregate models with aggregate data: a new methodological approach', *Transportation Research B*, 13B, 133-146.

[27] Brownstone, D. and K.A. Small (1989). 'Efficient estimation of nested logit models', *Journal of Business and Economic Statistics*, 7(1), 67-74.

[28] Case A. (1992) 'Neighborhood Influence and Technological Change', *Regional Science and Urban Economics*, 22, 491-508.

[29] Caulkins P., R. Bishop and N. Bouwes (1986), 'Travel cost model for lake recreation: a comparison of two methods for incorporating site quality and substitution effects', *American Journal of Agricultural Economics*, 68, 291-297.

[30] Courchene, T.J. 1994. *Social Policy in the Millennium*. Vancouver: C.D. Howe Institute.

[31] Cramer, J.S. (1991). *The Logit model: an introduction for economists*. London: Edward Arnold, London.

[32] Curry, L. (1972). 'Spatial Analysis of gravity flows', *Regional Studies*, 6, 131-147.

[33] Curry, L., Griffith, D.A. and E.S. Sheppard (1975). 'Those gravity parameters again', *Regional Studies*, 9, 289-296.

[34] Daganzo, C. (1979). *Multinomial Probit: The theory and its application to demand forecasting*. New York: Academic Press.

[35] Daly, A. (1982). 'Estimating Choice Models Containing Attraction Variables', *Transportation Research B*, 16(1), 5-15.

[36] DaVanzo, J. (1978). 'Does unemployment affect migration? - evidence from micro-data', *Review of Economics and Statistics*, 60, 504-514.

[37] Day, K.M. (1992). 'Inter-provincial migration and local public goods', *Canadian Journal of Economics*, 25(1), 123-142.

[38] Ewing, G. (1986). 'Spatial pattern in distance-deterrence parameters and Fotheringham's theory of competing destinations', *Environment and Planning A*, 18, 547-551.

[39] Fischer M.M. and E. Aufhauser (1988). 'Housing choice in a regulated market', *Geographical Analysis*, 20(1), 47-69.

[40] Fotheringham, A.S. (1981). 'Spatial structure and distance-decay parameters', *Annals of the Association of American Geographers*, 71(3), 425-436.

[41] Fotheringham, A. S. (1983). 'A new set of spatial interaction models: the theory of competing destinations', *Environment and Planning A* 15, 15-36.

[42] Fotheringham, A. S. (1985). 'Spatial competition and agglomeration in urban modelling', *Environment and Planning A*, 17, 213-230.

[43] Fotheringham, A.S. (1986) 'Modelling Hierarchical destination Choice', *Environment and Planning A*, 18, 401-18.

[44] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. New York: John Wiley & Sons.

[45] Gaudry, M.J.I. and M.G. Dagenais (1979). 'The dogit model', *Transportation Research B*, 13B, 105-11.

[46] Goldfeldt, S.M. , R.E. Quandt and H.F. Trotter (1966). 'Maximization by quadratic hill-climbing', *Econometrica*, 34, 541- 551.

[47] Gordon, I.R. (1985). 'Economic explanations of spatial variation in distance deterrence', *Environment and Planning A*, 17, 59-72.

[48] Gould, P. (1970). 'Is statistix inferens the geographical name for a wild goose?', *Economic Geography*, 46, 439-448.

[49] Greene, W. (1990). *Econometrics*. New York: MacMillan.

[50] Griffith, D.A. and K.G. Jones (1980), 'Explorations into the relationship between spatial structure and spatial interaction', *Environment and Planning A* 12, 187-201.

[51] Haining, R. P. (1990). *Spatial Data Analysis in the social and environmental sciences*. Cambridge: Cambridge University Press.

[52] Hansen, E.R. (1987). 'Industrial location in Sao Paulo, Brazil: a nested logit model', *Regional Science and Urban Economics*, 17, 89-108.

[53] Haynes, K. E., D.H. Good and T. Dignan (1988). 'Discrete spatial choice and the axiom of independence from irrelevant alternatives', *Socio-economic Planning Sciences*, 22(6), 241-251.

[54] Haynes, K.E. and A.S. Fotheringham (1990). 'The impact of space on the application of discrete choice models', *The Review of Regional Studies*, 20(2), 39-49.

[55] Hensher, D.A. (1986). 'Sequential and full information maximum likelihood estimation of a nested logit model', *The Review of Economics and Statistics*, 56, 657-667.

[56] Horowitz, J. (1986). 'Modeling choices of residential location and mode of travel to work', in *The Geography of Urban Transportation*, (S. Hanson, ed.), New York: Guilford Press.

[57] Huang, H.Y.(1970). 'Unified approach to quadratically convergent algorithms for function minimization', *Journal of Optimization Theory and Applications*, 5, 405-23.

[58] Johnston, R.J. (1973). 'On friction-of-distance and regression coefficients', *Area*, 5,187-191.

[59] Jones, K. and N. Bullen (1994). 'Contextual models of urban house prices: a comparison of fixed- and random-coefficient models developed by expansion,' *Economic Geography*, 68(3), 252-272.

[60] Kanaroglou, P.S. (1994). 'Sampling and Discrete Choice Analysis', *The Professional Geographer*, 46(3), 359-367.

[61] Kanaroglou, P.S., K.-L. Liaw and Y.Y. Papageorgiou (1986). 'An analysis of migratory systems: 2. Operational Framework', *Environment and Planning A*, 18, 1039-1060.

[62] Kerr, C. (1950), 'Labor markets: their character and consequences', *American Economic Review*, 40, 278-295.

[63] Kitamura, R., L.P. Kostyniuk and K. Ting (1979). 'Aggregation in Spatial Choice Modeling', *Transportation Science*, 13(4), 325-342.

[64] Leonardi, G. and Y.Y. Papageorgiou (1992). 'Conceptual Foundations of Spatial Choice Models', *Environment and Planning A*, 24, 1393-1408.

[65] Lerman, S.R. (1975). *A Disaggregate Behavioral Model of Urban Mobility Decisions*. Ph.D. dissertation. Department of Civil Engineering, MIT, Cambridge, Mass.

[66] Lerman, S. R. (1979). 'Neighborhood choice and transportation services', in *The economics of neighborhood* (D. Segal, ed.), 83-118, New York: Academic Press.

[67] Liaw, K.-L and J. Ledent (1988). 'Joint effects of ecological and personal factors on elderly interprovincial migration in Canada', *Canadian Journal of Regional Science*, 11(1), 77-100.

[68] Liaw, K.-L. (1990). 'Joint effects of personal factors and ecological variables on the interprovincial migration pattern of young adults in Canada: a nested logit analysis', *Geographical Analysis*, 22(3), 189-207.

[69] Lo, L. (1990). 'A translog approach to consumer spatial behaviour', *Journal of Regional Science*, 30(3), 393-413.

[70] Lo, L. (1991a). 'Spatial structure and spatial interaction: a simulation approach', *Environment and Planning A*, 23, 1279-1300.

[71] Lo, L. (1991b). 'Substitutability, spatial structure and spatial interaction', *Geographical Analysis*, 23(2), 132-146.

[72] Lo, L. (1992). 'Destination interdependence and the competing destinations model', *Environment and Planning A*, 24, 1191-1204.

[73] Lo, L. (1993), *Unbiased distance parameter: illusion or solution?*, Paper presented at the June 1993 meetings of the Canadian Association of Geographers, Ottawa, Canada.

[74] Louviere, J.J. (1983), 'A conceptual and analytical framework for understanding spatial and travel choices', *Economic Geography*, 57(4), 304-314.

[75] Lowe, J.M. and A. Sen (1994). 'Gravity model applications in health planning', Unpublished Paper.

[76] Luce, R.D. (1959). *Individual Choice Behaviour*. New York: Wiley.

[77] Manski, C. (1977). 'The Structure of Random Utility Models', *Theory and Decision*, 8, 229-254.

[78] McFadden D., 1974, 'Conditional logit Analysis of qualitative choice behavior', in (P. Zarembka, ed.), in *Frontiers of Econometrics*, 105-142, New York: Academic Press, 105-142.

[79] McFadden, D. (1978). 'Modeling the Choice of Residential Location', *Transportation Research Record*, 673, 72-77.

[80] McFadden, D. (1981). 'Econometric models of Probabilistic Choice,' in *Structural Analysis of Discrete Data with Econometric Applications*, (C.F. Manski and D. McFadden,eds) , Cambridge, MA: MIT Press, 198-272.

[81] McFadden, D. , W. Tye and K. Train (1977). 'An application of diagnostic tests for the irrelevant alternatives property of the multinomial logit model', *Transportation Research Record*, 637, 39-46.

[82] McMillen D. P. (1992). 'Probit with Spatial Autocorrelation', *Journal of Regional Science*, 32, 335-348.

[83] Meyer, R.J. and T.C. Eagle (1982). 'Context-induced parameter instability in a disaggregate-stochastic model of store choice', *Journal of Marketing Research*, 19, 62-71.

[84] Meyer, J.W. and A. Speare (1985). 'Distinctively elderly mobility: types and determinants', *Economic Geography*, 61, 79-87.

[85] Miller, E.J. and S.R. Lerman (1981). 'Disaggregate modelling and decisions of retail firms: a case study of clothing retailers', *Environment and Planning A*, 13, 729-746.

[86] Newbold, B. (1994). *Primary, Return and Onward Migration in Canada (1976-1986)*. Ph.D. dissertation. Department of Geography, McMaster University, Hamilton, Ont.

[87] Openshaw, S. (1984), 'Ecological fallacies and the analysis of areal census data', *Environment and Planning A*, 16, 17-31.

[88] Parsons G. and M. Kealy (1992). 'Randomly drawn opportunity sets in a random utility model of lake recreation', *Land Economics*, 68, 93-106.

[89] Parsons G. and M.S. Needelman (1992). 'Site aggregation in a random utility model of recreation', *Land Economics*, 68, 418-433.

[90] Pirie, G.H. (1976). 'Thoughts on revealed preferences and spatial behaviour', *Environment and Planning A*, 8, 947-955.

[91] Powell, M.J.D. (1964). 'An efficient method for finding the minimum of a function of several variables without calculating derivatives', *Computer Journal*, 7, 152-62.

[92] Putman, S. H. and S-H. Chung (1989), 'Effects of spatial system design on spatial interaction models. 1: the spatial system definition problem', *Environment and Planning A*, 21, 27-46.

[93] Schwab, M.G. and T.R. Smith, (1985). 'Functional invariance under spatial aggregation from continuous spatial interaction models', *Geographical Analysis*, 17(3), 217-230.

[94] Shaw, R.P. (1985). *Intermetropolitan migration in Canada: changing determinants over three decades*. Toronto: New Canada Publications.

[95] Sheppard, E.S. (1978). 'Theoretical underpinnings of the gravity hypothesis', *Geographical Analysis*, 10(4), 386-402.

[96] Sheppard, E.S. (1980). 'The Ideology of Spatial Choice', *Papers of the Regional Science Association*, 45, 195-213.

[97] Smith, K. and Y. Kaoru (1986), 'Modelling recreation demand within a random utility framework', *Economic Letters*, 22, 395-399.

[98] Sobel, K. L. (1980). 'Travel demand forecasting by using the nested multinomial logit model', *Transportation Research Record*, 775, 48-55.

[99] Soot, S. and A. Sen (1991). 'A spatial employment and economic development model', *Papers of the Regional Science Association*, 70, 149-166.

[100] Tardif, T.J. (1979). 'Definition of Alternatives and Representation of Dynamic Behavior in Spatial choice Models', *Transportation Research Record*, 723, 25-30.

[101] Tay, R.S. and P.S. McCarthy (1994). 'Benefits of improved water quality: a discrete choice analysis of freshwater recreational demands', *Environment and Planning A*, 26, 1625-1638.

[102] Thill, J-C. (1992). 'Choice Set Formation for Destination Choice Models', *Progress in Human Geography*, 16, 361-82.

[103] Thill, J-C. and Y. Van de Vyvere (1989). 'Workplace and locational choice of residence: a hierarchical approach', *Sistemi Urbani*, 3, 339-365.

[104] Thill, J-C. and J.L. Horowitz (1991). 'Estimating a Destination Choice Model from a Choice-based Sample with Limited Information', *Geographical Analysis*, 23(4), 298-315.

[105] Timmermans, H. and A. Borgers (1985). 'Choice set constraints and spatial decision-making processes', *Sistemi Urbani*, 3, 211-220.

[106] Train, K. (1978). 'A validation test of a disaggregate mode choice model', *Transportation Research*, 12, 167-174.

[107] Train, K. (1986). *Qualitative Choice Analysis: Theory, Econometrics and Application to Automobile Demand.* Cambridge, Ma: The MIT Press.

[108] Webber, M. J. (1980). 'A theoretical analysis of aggregation in spatial interaction models', *Geographical Analysis*, 12(2), 129-141.

[109] Wilson, A.G. (1970). *Entropy in Urban and Regional Modelling.* London: Pion.

[110] Wismer, D.A. and R. Chattergy (1978). *Introduction to Non-linear Optimization.* New York: North Holland.

[111] Wrigley, N. (1985). *Categorical Data Analysis for geographers and environmental scientists.* London: Longman.