

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9007777

**Aggregation and welfare analysis with mixed continuous/discrete
choice models**

Genc, Murat, Ph.D.

Washington State University, 1989

Copyright ©1989 by Genc, Murat. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

©1989

MURAT GENC

All Rights Reserved

**AGGREGATION AND WELFARE ANALYSIS WITH MIXED
CONTINUOUS/DISCRETE CHOICE MODELS**

By

MURAT GENÇ

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
Department of Economics

May 1989

©Copyright by MURAT GENÇ, 1989
All Rights Reserved

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of
MURAT GENÇ find it satisfactory and recommend that it be accepted.

Fred Goate
Chairperson

Will S. Hallagan

Paul J. Hall

Acknowledgements

This section not only expresses my gratitude to many individuals and institutions, but also reflects the way I view the world after working on this topic for such a long time! It should not be too surprising when I say that writing this section can be seen as an application of mixed continuous/discrete models. I here demonstrate how this can be done.

Writing a section like this essentially involves choices over both discrete and continuous variables as it is the case in mixed/continuous choice models. The discrete choice is to decide on whom to thank, and the continuous choice is to decide on how much to thank. However, in this specific case, the choice set of the discrete variables consists of an infinite number of alternatives in contrast to the typical setting of these models.

The list of these alternatives starts with my advisor, Fred Inaba, without whose guidance and support this dissertation would have been impossible. The next two are the other members of my committee, Ray Batina and William Hallagan. They both have been very patient and cooperative. The list continues with Nancy Wallace, who made an enormous contribution to this research even though she was miles away. The list of the alternatives also includes the other faculty members of the Economics Department who have provided me with the knowledge I needed to carry out this research.

Another alternative is John Kuhlman, who spent one summer for making the data set manageable while I was having a vacation. My other fellow students are also in the choice set, for both the encouragement and fun they have provided me with throughout the years I have been here.

Since the choice set is infinite, it is not possible to list all the elements of it. However, I should point out that all the members of my family, my past teachers, and all my friends, who all helped me to get to this point in one way or another, are, most naturally, members of this set. The following institutions that have provided various degrees of financial support are also included in this choice set: Department of Economics at W.S.U., Ministry of Education of Republic of Turkey, Washington State Department of Transportation, and Burlington Northern Foundation.

The continuous part of this decision problem relates to the amount of indebtedness I have to each element of this choice set. In contrary to the general procedure, no empirical estimation of the magnitude of this variable is required, because the model suggests that it will be infinite for all the alternatives.

If you have seen any errors or flaws either in this tiny model or in my dissertation, they all belong to me. None of individuals listed above should be held responsible for any error that might have occurred in this dissertation.

v

6.

**AGGREGATION AND WELFARE ANALYSIS WITH MIXED
CONTINUOUS/DISCRETE CHOICE MODELS**

ABSTRACT

by Murat Genç, Ph.D.
Washington State University
May 1989

Chairperson: Fred Inaba

Aggregation methods for making aggregate predictions from disaggregate pure discrete choice models have previously been developed and analyzed by numerous authors. This dissertation first introduces two more methods which are some variations of one of the existing methods. The dissertation then modifies these methods to incorporate heterogeneity of choice sets. The resulting adjusted methods are shown to produce more accurate predictions.

After this modification, the dissertation extends the aggregation methods to mixed continuous/discrete choice models where the choice sets of individuals consist of both discrete and continuous alternatives. Once aggregation methods are defined for such models, the focus of the analysis is the comparison of the aggregation methods in terms of their accuracy. Following a section on theoretical comparisons, some empirical results which are obtained from a multinomial mixed/continuous transportation demand model are presented to provide some evidence as to the relative performance of the methods.

Related to the aggregation problem is the measurement of aggregate welfare change of consumers when evaluating alternative policies based on models where agents are assumed to face discrete and/or continuous choices. The final section of the dissertation first develops a measure of welfare change for evaluating the welfare change of a single profit maximizing firm whose behavior is modeled within a mixed continuous/discrete choice framework. Then, it is shown that the aggregation methods analyzed in the previous sections can be applied to measuring aggregate welfare change as well. An empirical example, which deals with measuring the welfare change due to elimination of some alternatives, completes the dissertation.

Table of Contents

Acknowledgements	iii
Abstract	v
1. Introduction	1
1.1 Statement of the Problem	1
1.2 Relevance to Previous Work	2
1.2.1 Relevance to Previous Work on Aggregation	2
1.2.2 Relevance to Previous Work on Welfare Analysis	4
1.3 Overview of the Dissertation	5
2. Aggregation and Heterogeneity of Choice Sets in Discrete Choice Models	7
2.1 Introduction	7
2.2 General Structure of Discrete Choice Models	9
2.3 The Aggregation Problem	12
2.4 Heterogeneity of Choice Sets	16
2.5 Empirical Analysis	21
2.6 Conclusions	26
3. Aggregation in Mixed Continuous/Discrete Choice Models	34
3.1 Introduction	34
3.2 General Structure of Mixed Continuous/Discrete Choice Models	35
3.3 Aggregation in Mixed Continuous/Discrete Choice Models	40
3.4 Comparisons of the Aggregation Methods	45
3.5 Empirical Comparisons of Aggregation Methods	53

3.6 Summary and Conclusions	57
4. Aggregation and Welfare Analysis with a Mixed Continuous/Discrete Choice Model	62
4.1 Introduction	62
4.2 A Mixed Continuous/Discrete Transportation Demand Model	63
4.3 Welfare Analysis	66
4.4 Aggregation and Welfare Analysis	70
4.4.1 The Aggregation Problem	70
4.4.2 Aggregation Methods	72
4.5 An Empirical Example	75
4.6 Summary and Conclusions	81
5. Summary and Conclusions	82
References	88

List of Tables

Table	Page
2.1 Predicted aggregate demand resulting from unadjusted and adjusted Representative Agent methods vs. Actual aggregate demand (Percentage deviation from the actual value)	28
2.2 Predicted aggregate demand resulting from unadjusted and adjusted Classification methods vs. Actual aggregate demand (Percentage deviation from the actual value)	29
2.3 Predicted aggregate demand resulting from unadjusted and adjusted Statistical Differentials I methods vs. Actual aggregate demand (Percentage deviation from the actual value)	30
2.4 Predicted aggregate demand resulting from unadjusted and adjusted Statistical Differentials II methods vs. Actual aggregate demand (Percentage deviation from the actual value)	31
2.5 Predicted aggregate demand resulting from unadjusted and adjusted Statistical Differentials III methods vs. Actual aggregate demand (Percentage deviation from the actual value)	32
2.6 Aggregation errors by different criteria and aggregation methods.	33
3.1 Predictions of aggregation methods and actual aggregate demand (bushels)	59
3.2 Percentage deviations of the predictions from actual demand	60
3.3 Total aggregation errors by different criteria (Ranking of the method under the criterion)	61
4.1 Aggregate welfare change by different aggregation methods	79

Chapter 1

Introduction

1.1 Statement of the Problem

Mixed continuous/discrete choice models deal with the situations where the choice sets of the decision-makers consist of both discrete and continuous alternatives. A consumer who is to decide on which good to buy among a certain number of goods and how much of it to consume is a typical example of such a situation in economics. In general, a mixed continuous/discrete choice model provides us with the estimates of the probability that a decision-maker will choose any of the alternatives he or she faces and the quantity of that alternative he or she will consume, given the characteristics of the decision-maker and the alternatives. However, in many cases one is typically interested in predicting the behavior of aggregates of individuals rather than of particular decision-makers. This necessitates some level of aggregation in using mixed continuous/discrete choice models.

The transformation of disaggregate predictions into aggregate predictions is quite simple in principle. Once the individual quantities are obtained, they can be summed to obtain the aggregate demand for each alternative. In practice, however, this is rarely a feasible approach because it requires information on the characteristics of each decision-maker in the population for which the predictions are made and on each alternative that those decision-makers have in their choice sets. The aggregation problem in this context

is to find the best way of predicting aggregate demand without having such extreme data requirements.

Another point of interest is the measurement of the aggregate welfare change of consumers when evaluating alternative policies. The welfare change of an individual can be calculated once a model is estimated. However, as in the case of aggregate demand, predicting aggregate welfare change by summing up individual welfare changes requires extreme data requirements. This research aims to resolve this problem by making use of the results obtained for prediction of aggregate demand.

1.2 Relevance to Previous Work

This research is essentially based on two different bodies of literature. The first body of literature consists of earlier studies on the aggregation issue in discrete choice models whereas the other one involves the studies on measuring the welfare change within the same framework. The main goal of this research is to further these two sets of literature and then combine them to obtain an aggregate measure of welfare change within the mixed continuous/discrete choice framework.

1.2.1 *Relevance to Previous Work on Aggregation*

The aggregation problem has received extensive attention in the literature, particularly in travel demand literature. A number of aggregation methods have been developed and analyzed by Koppelman (1974, 1976a, 1984), Domencich and McFadden

(1975), McFadden and Reid (1975), Reid (1978), Koppelman and Ben-Akiva (1977), Bouthelier and Daganzo (1979), Talvitie (1973), Westin (1974), and Dunne (1985).¹ Following the taxonomy developed in Koppelman (1976a), these aggregation methods can be categorized in five groups: the representative agent method (known as the naive method), the classification method, the statistical differentials method, the explicit integration approach, and, the sample enumeration method.

These methods differ in terms of their computational costs, accuracy, data requirements, and the complacency in their implementation in any particular situation. Accordingly, the consensus in the literature is that none of them clearly dominates the others. However, the representative agent, classification, statistical differentials, and sample enumeration methods are the ones that are mostly used and analyzed, largely due to their relative simplicity.

An important result of the previous studies is that the differences in the choice sets of the individuals crucially affect the performance of these methods in terms of accuracy. Indeed, it has been suggested for the classification method that the population be divided into sub-groups according to the set of feasible choices. However, beyond that, very little has been done to incorporate heterogeneity of the choice sets into these aggregation methods.² In aiming to fill this gap in the literature, this research adjusts these methods in a simple way to account for heterogeneity of choice sets of the individuals, and suggests new methods of aggregation that provide more accurate predictions.

1. A good summary of these methods is available in Ben-Akiva and Lerman (1985).

2. Koppelman and Ben-Akiva (1977) do suggest an adjustment for the representative agent method.

All the studies mentioned above involve only pure discrete cases. The only study in the aggregation literature that deals with aggregation in a mixed continuous/discrete choice model is Westin and Gillen (1978). In their study, they apply the explicit integration approach to a binary probit mixed continuous/discrete choice model. There are currently no extensions of the other methods to aggregation issues arising in mixed continuous/discrete choice models. Thus, the second issue that this research takes up is the extension of aggregation methods to mixed continuous/discrete choice models in such a way that accounts for heterogeneity of choice sets.

1.2.2 Relevance to Previous Work on Welfare Analysis

There exists an extensive literature that deals with the theory and measurement of consumer's welfare change. However, there is no general theory available for measuring welfare changes when consumers' choice sets include discrete variables. A few authors, notably Small and Rosen (1981), Hanemann (1982, 1984b), and Hau (1983, 1985, 1986, 1987), have rigorously derived Hicksian measures of welfare change for particular classes of pure discrete choices.

However, these studies do not address the aggregation problem. These studies are either interested in only measuring the welfare change of an individual or they assume away the aggregation problem by implicitly applying the representative agent method. Furthermore, although these studies deal with cases where discreteness occurs, the models considered in all of them correspond to pure discrete cases rather than cases with both discrete and continuous choices.

Thus, there remains two main tasks to further this body of literature: developing a measure of welfare change of an individual within a mixed continuous/discrete choice framework, and then, finding a measure of aggregate welfare change by making use of the aggregation methodology developed in the other body of literature on aggregation. This research aims to undertake these tasks.

1.3 Overview of the Dissertation

Of the aggregation methods mentioned earlier, this research focuses on the representative agent, classification, and statistical differentials approaches. The starting point of the analysis, which starts in the following chapter, is to formalize the underlying assumptions behind these methods. This also allows us to introduce two new variations of the statistical differentials approach by making use of the similarities between the underlying assumptions of these approaches. Then, the necessity of incorporating the heterogeneity of choice sets of the agents is shown in Chapter 2 within a pure discrete choice framework. Following this, all the aggregation methods that this research focuses on are modified by making use of two restrictions that result from the heterogeneity of choice sets and the built-in structure of pure discrete choice models. An empirical analysis that focuses on the performance of these adjusted and unadjusted methods completes the analysis in Chapter 2.

The focus of analysis in Chapter 3 is the generalization of the aggregation methods to mixed continuous/discrete choice models. Following the general structure of such models, the aggregation problem within this framework is shown to be essentially

the same as in the pure discrete choice case. Hence, the same approaches used in pure discrete choice case are applied to develop aggregation methods for mixed continuous/discrete choice models. The analysis then focuses on the comparisons of the aggregation methods, both theoretically and empirically, in terms of their performances.

Chapter 4 takes up the issue of measuring aggregate welfare change within the framework of mixed continuous/discrete choice. First, a measure of welfare change is developed for evaluating the welfare change of a single profit maximizing agent whose behavior is modeled within a mixed continuous/discrete choice framework. Then, it is shown that the same approaches employed before can be applied to the aggregation problem in this context as well. An empirical example, which applies these approaches to measuring welfare change when some alternatives are eliminated, completes the analysis. Chapter 5 provides a brief summary of the analysis by combining the results of the three main chapters.

Chapter 2

Aggregation and Heterogeneity of Choice Sets in Discrete Choice Models

2.1 Introduction

The aggregation problem has received extensive attention in the literature, particularly in travel demand literature. A number of aggregation methods have been developed and analyzed by Koppelman (1974, 1976a, 1984), Domencich and McFadden (1975), McFadden and Reid (1976), Reid (1978), Koppelman and Ben-Akiva (1977), Bouthelier and Daganzo (1979), Talvitie (1973), Westin (1974), and Dunne (1985).¹ Of these methods, the "representative agent", classification, statistical differentials, and sample enumeration methods are the ones that are mostly used and analyzed, largely due to their simplicity. The representative agent method, known also as the naive method, is based on the assumption that each individual in the population behaves like the representative agent whose characteristics correspond to the mean value of the characteristics of the individuals in the population. The classification method divides the population into nearly homogeneous sub-groups and then applies the representative agent method to these groups. The aggregate prediction is then calculated as the weighted average of the estimates for these sub-groups. The statistical differentials method extends the representative agent method by including the variance of the characteristics as a measure of departures from homogeneity. The sample enumeration method uses a sample

1. A good summary of these methods is available in Ben-Akiva and Lerman (1985).

to represent the entire population and estimates the aggregate prediction as the sum of the predictions for each individual in the sample.

An important result of the previous studies is that the differences in the choice sets of the individuals crucially affect the performance of these methods in terms of accuracy. Indeed, it has been suggested for the classification method that the population be divided into sub-groups according to the set of feasible choices. However, beyond that, very little has been done to incorporate heterogeneity of the choice sets into these aggregation methods.² In aiming to fill this gap in the literature, this chapter adjusts the methods of representative agent, classification, and statistical differentials in a simple way to account for heterogeneity of choice sets.

The remainder of the chapter is organized as follows: Section two outlines the general structure of the discrete choice models. Section three formally defines the aggregation problem and representative agent, classification, and statistical differentials methods. Section four focuses on the issue of heterogeneous choice sets and modification of these methods to incorporate the differences in the choice sets. Section five presents an empirical analysis which is based on a multinomial logit model. Section six is the conclusions section.

2. Koppelman and Ben-Akiva (1977) do suggest an adjustment for the representative agent method.

2.2 General Structure of Discrete Choice Models

Discrete choice models are used to analyze situations in which a decisionmaker can be described as facing a choice among a finite and exhaustive set of mutually exclusive alternatives. These models provide us with the probability that a decisionmaker will choose a particular alternative from a set of alternatives, given the data observed by the researcher. These probabilities and thus the discrete choice models can be shown to be derived from utility theory by the use of the concept of random utility.

The random utility model is based on a precise distinction between the behavior of the decisionmaker and the analysis of the researcher.³ Consider a decisionmaker n who has a choice set C_n . Each alternative in C_n would give some utility to the decisionmaker if it were chosen. Let U_{in} be the utility from alternative i in C_n . This utility depends on various factors among which are the characteristics of the alternative, and the characteristics of the decisionmaker. If a_{in} denotes the vector of all relevant characteristics of alternative i as faced by person n , and s_n denotes the vector of all relevant characteristics of person n , then the utility U_{in} can be written as a function of these factors;

$$U_{in} = U(a_{in}, s_n) \quad \forall j \in C_n. \quad (2.2.1)$$

The decisionmaker chooses the alternative that provides the highest utility. Thus, alternative i in C_n is chosen if and only if⁴

$$U_{in} > U_{jn} \quad \forall i \in C_n, \quad j \neq i$$

3. For further explanation and discussion see Train, Ben-Akiva and Lerman, and McFadden.

4. Under the assumption that no ties occur.

which in turn implies that i is chosen if and only if

$$U(a_{in}, s_n) > U(a_{jn}, s_n) \quad \forall i, j \in C_n, \quad j \neq i. \quad (2.2.2)$$

Note that everything so far is deterministic. The decisionmaker chooses the alternative that gives the highest utility which he or she can evaluate. At this point, the probability that person n will choose alternative i is either zero or one depending upon whether or not alternative i provides the highest utility.

It is at this point where random utility model arises. The fact that the researcher does not observe all the relevant factors and does not know the utility function exactly results in randomness. Suppose that z_{in} denotes the characteristics of the alternatives that the researcher can observe. Similarly, suppose that h_n is the vector of observed characteristics of person n . Let $V'(z_{in}, h_n, \beta)$ be the part of the utility function that depends only on factors that can be observed and the form of which is known up to a vector of parameters, β , to be estimated. Then if ε_{in} denotes the other part of the utility that represents all the unknown factors and aspects of the utility, equation (2.2.1) can be rewritten as

$$U_{in} = U(a_{in}, s_n) = V'(z_{in}, h_n, \beta) + \varepsilon_{in}. \quad (2.2.3)$$

The probability that the utility of alternative i is higher than that of any other alternative given the observed components of utility for each alternative is⁵

$$P_{in} = \text{Prob} \{ U_{in} > U_{jn}, \quad \forall j \in C_n, \quad j \neq i \}. \quad (2.2.4)$$

5. For different interpretations of what this probability means, see Train p.11 and p.235.

By use of equation (2.2.3) this can be rewritten as

$$P_{in} = \text{Prob} \left\{ V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}, \quad \forall j \in C_n, \quad j \neq i \right\} \quad (2.2.5)$$

which in turn is

$$P_{in} = \text{Prob} \left\{ \varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \quad \forall j \in C_n, \quad j \neq i \right\}. \quad (2.2.6)$$

Since $V_{in} - V_{jn}$ is observed but $\varepsilon_{jn} - \varepsilon_{in}$ is a random variable, the right-hand side of (2.2.6) is simply a cumulative distribution. More precisely, it is a joint cumulative distribution; the probability that each random variable $\varepsilon_{jn} - \varepsilon_{in}$ is below $V_{in} - V_{jn}$, respectively, for all j in C_n , $j \neq i$. Different assumptions on the distributions of ε 's give rise to different functional forms for the choice probabilities, and thus, to different discrete choice models such as probit, logit, etc.

The final point is about the transition from probabilities to demand. What has been said up to this point is on predicting individual behavior, more precisely, on predicting the probability that any particular individual will choose any alternative that he or she faces. But what one is really interested in is the aggregate demand. Conceptually this transition is quite simple. Since P_{in} is the probability that alternative i will be chosen by decisionmaker n , the aggregate demand for alternative i is simply the expected total number of individuals choosing alternative i . If N_T denotes the total number of decisionmakers, the aggregate demand for alternative i can be expressed as

$$N_T(i) = \sum_{n=1}^{N_T} P_{in}. \quad (2.2.7)$$

2.3 The Aggregation Problem

Let, for convenience, x_n be the vector of all the attributes (characteristics) that affect the choice of individual n .⁶ T is the population for which the aggregate predictions are needed. It will be assumed that a discrete choice model has been estimated so that a consistent estimate of the probability that an individual n will choose an alternative i among a set of alternatives he or she faces is available. Let $P(i|x_n)$ denote this probability. Note that $P(i|x_n)$ is a function of some estimated coefficients and the elements of the vector of the attributes x_n . When these coefficients and the attributes are both known, the probability of choosing alternative i can easily be calculated for any individual. Since it is assumed that the coefficients are already estimated, they simply are some known parameters in terms of evaluating the function $P(i|x_n)$. This leaves only the attributes that one needs to know in order to evaluate $P(i|x_n)$. Consistent with this conditionality, the notation $P(i|x_n)$ means "the probability that an individual n will choose alternative i given the attributes that affect his choice".

Let n_i be a Bernoulli random variable that takes the value 1 if alternative i is chosen by individual n and zero otherwise:

$$n_i = \begin{cases} 1, & \text{if } i \text{ is chosen;} \\ 0, & \text{otherwise.} \end{cases}$$

Thus n_i represents individual n 's choice for alternative i and $P(i|x_n)$ is the probability that it will take the value 1. Then the expectation of n_i , conditional on x_n , is

$$E\{n_i|x_n\} = \text{Prob}\{i \text{ is chosen}\} = P(i|x_n).$$

6. So, $x_n = (x_n^1, \dots, x_n^R)$ where $x_n^r, r = 1, \dots, R$ is attribute r that affects the choice of individual n .

Now, let $N_T(i)$ be the sum of n_i 's, i. e., total number of the individuals who choose alternative i . Then

$$E[N_T(i)|x_1, \dots, x_{N_T}] = \sum_{n=1}^{N_T} P(i|x_n). \quad (2.3.1)$$

Thus, it is formally shown that $N_T(i)$ is the expected total number of individuals choosing alternative i , given the attributes.

Clearly, if one has the values of the attributes for all the individuals in population T , equation (2.3.1) is the expression that should be used to predict the aggregate demand for alternative i . However, this information is rarely available for the researcher. In most cases such a calculation is simply not feasible. Also, even if this extreme data requirement is met, implementation of this computation can be too complicated for large data sets. Thus, another way of predicting aggregate demand with fewer data requirements should be found. The trade-off is of course between the degree of data requirements and the ease in implementing the method and the accuracy of the method.

Given that one does not necessarily know the attributes for each and every individual in the population, x_n is, in general, a random variable and thus, the expectations above need to be expressed as unconditional expectations in order to yield a useful expression. If x_n is assumed to be continuous with joint density function $f_n(x_n)$, then

$$E[n_i] = \int_{x_n} P(i|x_n) f_n(x_n) dx_n$$

and

$$E[N_T(i)] = \sum_{n=1}^{N_T} \int_{x_n} P(i|x_n) f_n(x_n) dx_n. \quad (2.3.2)$$

The aggregation methods mentioned in the introduction differ from each other based on their representation of the distribution of the attributes.

The representative agent method (Koppelman (1976a), Domencich and McFadden (1975)) is based on the assumption that the attributes are concentrated at one point:

$$f_n(\mathbf{x}_n) \approx \begin{cases} 1, & \text{if } \mathbf{x}_n = \bar{\mathbf{x}}; \\ 0, & \text{otherwise.} \end{cases} \quad n = 1, \dots, N_T; \quad (2.3.3)$$

where $\bar{\mathbf{x}} = (\bar{x}^1, \dots, \bar{x}^R)$ and $\bar{x}^r = \frac{1}{N_T} \sum_{n=1}^{N_T} x_n^r$, $r = 1, \dots, R$.⁷ This implies that

$$E[n_i] = P(i|\bar{\mathbf{x}})$$

and

$$E[N_T(i)] = \sum_{n=1}^{N_T} P(i|\bar{\mathbf{x}}) = N_T P(i|\bar{\mathbf{x}}) \equiv N_T^R(i) \quad (2.3.4)$$

where $N_T^R(i)$ denotes the expected total number of individuals choosing alternative i predicted by the representative agent method.

The classification method (Koppelman (1976a)) is based on the assumption that the attributes are concentrated at more than one point:

$$f_n(\mathbf{x}_n) \approx \begin{cases} N_g/N_T, & \text{if } \mathbf{x}_n = \bar{\mathbf{x}}_g, \quad g = 1, \dots, G; \\ 0, & \text{otherwise.} \end{cases} \quad n = 1, \dots, N_T \quad (2.3.5)$$

where $\bar{\mathbf{x}}_g$ is the mean vector of the attributes of the individuals in group g and N_g is the number of those individuals. Note that N_g 's must satisfy $\sum_{g=1}^G N_g = N_T$. This assumption implies that

$$E[n_i] = \sum_{g=1}^G \frac{N_g}{N_T} P(i|\bar{\mathbf{x}}_g)$$

7. Thus, $\bar{\mathbf{x}}$ is the vector of the means of the attributes of all the individuals in the population.

and

$$E[N_T(i)] = \sum_{n=1}^{N_T} \sum_{g=1}^G \frac{N_g}{N_T} P(i|\bar{x}_g) = \sum_{g=1}^G N_g P(i|\bar{x}_g) \equiv N_T^C(i) \quad (2.3.6)$$

where $N_T^C(i)$ denotes the total expected number of individuals choosing alternative i predicted by the classification method.

The statistical differentials method (Talvitie 1973) starts with an approximation of the choice function $P(i|\mathbf{x}_n)$ rather than an assumption on the density function of the attributes. It is based on the second-order Taylor's series expansion of $P(i|\mathbf{x}_n)$ about $\bar{\mathbf{x}}_n$:

$$\begin{aligned} P(i|\mathbf{x}_n) &= P(i|\bar{\mathbf{x}}_n) + \sum_{r=1}^R \frac{\partial}{\partial x_n^r} P(i|\mathbf{x}_n) \Big|_{\bar{\mathbf{x}}_n} (x_n^r - \bar{x}_n^r) \\ &\quad + \frac{1}{2} \sum_{r=1}^R \sum_{s=1}^R \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|\mathbf{x}_n) \Big|_{\bar{\mathbf{x}}_n} (x_n^r - \bar{x}_n^r)(x_n^s - \bar{x}_n^s). \end{aligned}$$

Then it follows that

$$\begin{aligned} E[N_T(i)] &= \sum_{n=1}^{N_T} \int_{\mathbf{x}_n} P(i|\mathbf{x}_n) f_n(\mathbf{x}_n) d\mathbf{x}_n \\ &\approx \sum_{n=1}^{N_T} \left[P(i|\bar{\mathbf{x}}_n) \int_{\mathbf{x}_n} f_n(\mathbf{x}_n) d\mathbf{x}_n + \sum_{r=1}^R \frac{\partial}{\partial x_n^r} P(i|\mathbf{x}_n) \Big|_{\bar{\mathbf{x}}_n} \int_{\mathbf{x}_n} (x_n^r - \bar{x}_n^r) f_n(\mathbf{x}_n) d\mathbf{x}_n \right. \\ &\quad \left. + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|\mathbf{x}_n) \Big|_{\bar{\mathbf{x}}_n} \int_{\mathbf{x}_n} (x_n^r - \bar{x}_n^r)(x_n^s - \bar{x}_n^s) f_n(\mathbf{x}_n) d\mathbf{x}_n \right] \\ &= \sum_{n=1}^{N_T} \left[P(i|\bar{\mathbf{x}}_n) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|\mathbf{x}_n) \Big|_{\bar{\mathbf{x}}_n} \text{Cov}(x_n^r, x_n^s) \right]. \end{aligned}$$

By assuming a common distribution function for the attributes, one obtains

$$N_T^{SDI}(i) \equiv N_T \left[P(i|\bar{\mathbf{x}}) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|\mathbf{x}_n) \Big|_{\bar{\mathbf{x}}} \text{Cov}(x^r, x^s) \right] \quad (2.3.7)$$

where $N_T^{SDI}(i)$ denotes the total expected number of individuals choosing alternative i predicted by statistical differentials method.

Although equation (2.3.7) is what numerous authors applied in the past studies, there are two more variations of this method that one can obtain. The first one is based on recognizing the existence of some nearly homogeneous sub-groups in the population, as the classification method suggests. Then the aggregate demand can be calculated as the sum of the total demands in each sub-group for which the total demand is obtained by using the statistical differentials approach. Thus, the first variation can be expressed as

$$N_T^{SDII}(i) \equiv \sum_{g=1}^G \left\{ N_g \left[P(i|\bar{x}_g) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|x_n) \Big|_{\bar{x}_g} \text{Cov}(x_g^r, x_g^s) \right] \right\} \quad (2.3.8)$$

where $\text{Cov}(x_g^r, x_g^s)$ the covariance of the r th and s th attributes in group g .

Note that in deriving equation (2.3.8) it is implicitly assumed that there is a common distribution function for the attributes in each group g which differs from the other common distributions in the other groups. If one is willing to assume that each common distribution function in each group g shares the same variance-covariance matrix, another variation of the statistical differentials method is obtained. This second variation can be expressed as

$$N_T^{SDIII}(i) \equiv \sum_{g=1}^G \left\{ N_g \left[P(i|\bar{x}_g) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|x_n) \Big|_{\bar{x}_g} \text{Cov}(x^r, x^s) \right] \right\} \quad (2.3.9)$$

2.4 Heterogeneity of Choice Sets

In general, every individual might have a different choice set, i. e., an alternative i is not necessarily feasible for all individuals. So, C_n 's are potentially different for each

individual. If N_{Ti} denotes the total number of individuals for whom alternative i is feasible, then

$$N_{Ti} \leq N_T \quad \forall i. \quad (2.4.1)$$

Furthermore, if alternative i is not feasible for individual n , then the probability that individual n will choose alternative i is zero

$$P(i|\mathbf{x}_n) = 0 \quad \text{if } i \notin C_n. \quad (2.4.2)$$

By making use of this last expression, equations (2.3.1) and (2.3.2) can be rewritten as

$$E[N_T(i)|\mathbf{x}_1, \dots, \mathbf{x}_{N_T}] = \sum_{n=1}^{N_{Ti}} P(i|\mathbf{x}_n) \quad (2.4.3)$$

and

$$E[N_T(i)] = \sum_{n=1}^{N_{Ti}} \int_{\mathbf{x}_n} P(i|\mathbf{x}_n) f_n(\mathbf{x}_n) d\mathbf{x}_n. \quad (2.4.4)$$

Using equation (2.3.2) rather than (2.4.4) yields predictions with an upward bias, although they both theoretically add up to the same number. Suppose that the researcher has data on the attributes of each individual in the population. Then, once the choice function $P(i|\mathbf{x}_n)$ is estimated, the predictions of the individual probabilities can be summed to find the aggregate predictions. However, if the feasibility of the alternative i is not taken into consideration, the evaluation of the choice function will in general result in a positive probability even though alternative i is not feasible. Thus, leaving out the feasibility of an alternative as a criterion in evaluating the choice function causes overprediction of the aggregate demand for any particular alternative. Afterall, there is no need to evaluate the choice function for an alternative if that alternative is not feasible for the individual in consideration.

Another point that needs to be emphasized at this point is the existence of a restriction that should be imposed in deriving aggregation methods. It is known that the total number of individuals in the population is N_T . Together with the requirement that each individual must choose one and only one alternative in this framework, this implies that the overall sum of the predictions for each alternative should add up to N_T .

Formally,

$$\sum_{i=1}^I \hat{D}_i = N_T \quad (2.4.5)$$

where $\hat{D}_i$ is the estimator to be used for predicting $E[N_T(i)]$, and I is the total number of alternatives.

A closer look at the aggregation methods defined in the previous section will confirm that this equality holds for the representative agent and classification methods, although it has not been imposed explicitly in the derivation of these methods.⁸ However, the statistical differentials method does not satisfy this restriction. This might be seen as one of the reasons for statistical differentials method producing the least accurate results in past studies that made comparisons between the aggregation methods.⁹

Using equations (2.4.4) and (2.4.5) in the derivation of previously defined aggregation methods thus yields adjusted aggregate predictions. Preserving the assumption on the density of the attributes and using equation (2.4.4) in the derivation of the representative agent method results in

$$E[N_T(i)] = \sum_{n=1}^{N_{Ti}} P(i|\bar{x}) = N_{Ti} P(i|\bar{x}).$$

8. For example, consider the representative agent method: $\sum_i N_T P(i|\bar{x}) = N_T \sum_i P(i|\bar{x}) = N_T$, since $\sum P(i|\bar{x}) = 1$.

9. For example, Koppelman (1976a), Reid (1978), and Dunne (1985).

By imposing equation (2.4.5), the adjusted representative agent method is obtained as

$$N_{Ti}^R(i) \equiv \frac{N_{Ti} P(i|\bar{x})}{\sum_i N_{Ti} P(i|\bar{x})} N_T \quad (2.4.6)$$

where $N_{Ti}^R(i)$ denotes the expected total number of individuals choosing alternative i predicted by the adjusted representative agent method. By adjusted representative method it is meant that both the differences in the choice sets and the restriction on the overall demand are taken into consideration.

Similarly, the adjusted classification method is obtained by using equations (2.4.4) and (2.4.5) in the derivation of classification method without changing anything else:

$$\begin{aligned} E[N_T(i)] &= \sum_{n=1}^{N_{Ti}} \sum_{g=1}^G \frac{N_g}{N_T} P(i|\bar{x}_g) = \frac{N_{Ti}}{N_T} \sum_{g=1}^G N_g P(i|\bar{x}_g) \\ \Rightarrow N_{Ti}^C(i) &\equiv \frac{\frac{N_{Ti}}{N_T} \sum_{g=1}^G N_g P(i|\bar{x}_g)}{\sum_i \frac{N_{Ti}}{N_T} \sum_{g=1}^G N_g P(i|\bar{x}_g)} N_T. \end{aligned} \quad (2.4.7)$$

Here $N_{Ti}^C(i)$ denotes the total expected number of individuals choosing alternative i predicted by the adjusted classification method.

The use of equations (2.4.4) and (2.4.5) in the derivation of statistical differentials method changes the method as follows:

$$\begin{aligned} E[N_T(i)] &= N_{Ti} \left[P(i|\bar{x}) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|x_n) \Big|_{\bar{x}} \text{Cov}(x^r, x^s) \right] \equiv \hat{D}_i^{SDI} \\ \Rightarrow N_{Ti}^{SDI}(i) &\equiv \frac{\hat{D}_i^{SDI}}{\sum_i \hat{D}_i^{SDI}} N_T \end{aligned} \quad (2.4.8)$$

where $N_{Ti}^{SDI}(i)$ is the total expected number of individuals choosing alternative i predicted by adjusted statistical differentials I method.

The adjusted variations of the statistical differentials method are similarly obtained. The resulting expression for the first variation is

$$E[N_T(i)] = \sum_{g=1}^G \left\{ N_{gi} \left[P(i|\bar{x}_g) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|x_n) \Big|_{\bar{x}_g} \text{Cov}(x_g^r, x_g^s) \right] \right\} \equiv \hat{D}_i^{SDII}(i)$$

which implies that

$$N_{Ti}^{SDII}(i) \equiv \frac{\hat{D}_i^{SDII}}{\sum_i \hat{D}_i^{SDII}} N_T \quad (2.4.9)$$

where $N_{Ti}^{SDII}(i)$ is the total expected number of individuals choosing alternative i predicted by adjusted statistical differentials II method.

The adjusted statistical differentials III method differs in only the covariance term. Hence, the second variation of the statistical differentials approach is adjusted as follows:

$$E[N_T(i)] = \sum_{g=1}^G N_{gi} \left[P(i|\bar{x}_g) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial x_n^r \partial x_n^s} P(i|x_n) \Big|_{\bar{x}_g} \text{Cov}(x^r, x^s) \right] \equiv \hat{D}_i^{SDIII}(i)$$

$$\Rightarrow N_{Ti}^{SDIII}(i) \equiv \frac{\hat{D}_i^{SDIII}}{\sum_i \hat{D}_i^{SDIII}} \quad (2.4.10)$$

where $N_{Ti}^{SDIII}(i)$ is the total expected number of individuals choosing alternative i predicted by adjusted statistical differentials III method.

Note that when an alternative is feasible for all the individuals, the adjusted methods are equivalent to the unadjusted methods.¹⁰ Thus, the unadjusted methods are special cases of the adjusted methods. Another point is that these newly defined methods seem to require additional information as to which alternatives are feasible for

10. Given that the restriction on the overall sum of predictions over the alternatives is imposed for both adjusted and unadjusted methods.

how many individuals in the population. However, since it is assumed that the model has already been estimated this information should already be available. Thus, this really is not an additional data requirement. Even in the cases where these aggregation methods are to be used to extend the results of the estimated model to a larger population for which the model is not estimated, the feasibility of the alternatives is known in most of the cases. If this information is not available in these situations, it is more reasonable to use the ratio of the number of individuals for whom an alternative is feasible in the sample to the total number of individuals in the sample, rather than assuming an alternative is feasible for every individual in the population.

2.5 Empirical Analysis

A multinomial logit model, which was part of a transportation demand model earlier developed by Inaba and Wallace (1987), is used for the empirical test of the performance of the adjusted aggregation methods suggested in the previous section. Although the main focus of the analysis is the performance of the suggested adjusted methods against the performance of the unadjusted methods, the relative performance of the aggregation methods within each group is also investigated.

Tables 2.1–2.5 show the predicted aggregate demand for each alternative as calculated by both adjusted and unadjusted aggregation methods. The percentage deviation of each prediction from its actual value is also reported in these tables. The actual values are simply obtained by applying complete enumeration to the sample at hand, i. e., by evaluating the probability of choosing alternative i for each individual in the sample

and adding them up. There are 32 alternatives in total, each of which represents a joint mode/destination choice.¹¹ The total number of individuals in the sample is 183.

A first-hand look at the numbers in tables 2.1–2.5 indicates that the adjusted methods provide better predictions. The predictions from the representative agent method are presented in Table 2.1. For 21 of the 32 alternatives, the adjusted representative agent method is found to produce better predictions than the unadjusted representative agent method. Furthermore, the average of the percentage deviations from the actual values is found to be 98 for the unadjusted representative agent method, whereas it is only 45 for the adjusted representative agent method.

Similar results are obtained for the other methods. The adjusted classification method, for which the results are presented in Table 2.2, gives better predictions than the unadjusted classification method for 18 alternatives of the 32 alternatives. The average percentage deviation is also considerably less for the adjusted classification method compared to the unadjusted classification method: 53 vs 125.

The results presented in Tables 2.3–2.5 for the different variations of the statistical differentials approach coincide with the results for the representative agent and classification methods. Only the statistical differentials II method is found to give worse predictions for more than 16 alternatives when it is adjusted. However, the average percentage deviation for the unadjusted statistical differentials II method is 446 whereas it is only 138 for the adjusted statistical differentials II.

11. For details of the model, see chapter 4.

When the average of the percentage deviations from the actual values for each aggregation method is compared to each other within the two groups of aggregation methods, the representative agent method seems to be the best among both the adjusted and unadjusted methods. The same comparisons also indicate that the classification approach is the second best in each group.

Although these results indicate that the adjusted methods provide better predictions, a more thorough analysis is needed in order to draw firmer conclusions. To serve this purpose, Table 2.6 shows the aggregation errors resulting from each of the aggregation methods. In an attempt to have results independent of the measure of aggregation error used in the analysis, four different measures of aggregation error are used in calculating the aggregation errors. As pointed out by Koppelman (1976b), developing a suitable error measure requires two decisions. The first is how to express the error that occurs in making predictions for one alternative. The second is how to aggregate the errors from single predictions to some overall measure of the total error.

The first three measures used in this analysis use a quadratic loss function in aggregating the single predictions, resulting in different kinds of *root mean square error* based on their representation of the individual errors. The first one, $RMSE_I$, is based on the error measure described in Koppelman (1976b). It expresses the magnitude of the error from a single prediction as a proportion of the actual value:¹²

$$RMSE_I = \sqrt{\sum_i^I \left[\frac{\hat{D}_i - A_i}{A_i} \right]^2} \quad (2.5.1)$$

12. In Koppelman (1976b), it is expressed as a proportion of the predicted value.

where

$RMSE$ = root mean square error

$\hat{D}_i$ = predicted value for the demand for alternative i

A_i = actual value for the demand for alternative i

I = total number of alternatives.

The second measure, $RMSE_{II}$, is precisely what Reid (1978) has used:

$$RMSE_{II} = \sqrt{\sum_i^I A_i \left[\frac{\hat{D}_i - A_i}{A_i} \right]^2}. \quad (2.5.2)$$

The difference between these two measures is that $RMSE_{II}$ weights the individual errors by the actual values to reflect the relative importance of the individual errors.

The other two measures use an absolute measure for the individual errors rather than a relative measure. $RMSE_{III}$ takes the square of the individual differences between the predicted value and the actual value;

$$RMSE_{III} = \sqrt{\sum_i^I (\hat{D}_i - A_i)^2}, \quad (2.5.3)$$

whereas $RMSE_{IV}$ takes the absolute value of the same differences:

$$RMSE_{IV} = \sum_i^I |\hat{D}_i - A_i|. \quad (2.5.4)$$

The results reported in Table 2.6 strongly support the hypothesis that incorporation of the heterogeneity of the choice sets increases the accuracy of the aggregation methods by a large amount. For all the methods investigated in this chapter, the aggregation error is reduced by at least 22 percent when the adjusted methods are used

in place of unadjusted ones. This improvement is robust with respect to measure of aggregation error, since a reduction is observed for all measures used in the analysis. The amount of the reduction reaches to 86 percent for the statistical differentials method II when $RMSE_I$ is used as the aggregation error measure. These results suggest the use of the adjusted aggregation methods developed in this study. Given that these adjusted aggregation methods do not require additional information, they should be preferred over the unadjusted methods.

In terms of the comparisons of the methods against each other, the classification method is the most accurate method among the unadjusted methods. For all the aggregation error measures other than $RMSE_I$, it provides the least error. It also is the second best method among the unadjusted methods when $RMSE_I$ is used as the basis of comparison. Within the same class, the statistical differentials III method is the worst when $RMSE_I$ or $RMSE_{II}$ is taken to be the basis of comparison, and the statistical differentials I method is the worst when $RMSE_{IV}$ is used. Thus, if these different variations of the statistical differentials approach are taken together, this approach produces the worst predictions among the unadjusted methods.

These results are consistent with the earlier findings about the unadjusted methods that have been investigated in past studies. The results for the unadjusted methods obtained here coincide with these findings, too. The classification method is the best when $RMSE_{II}$ or $RMSE_{IV}$ is used as the measure of aggregation error for the unadjusted methods. Within the same class, the statistical differentials I method is the worst when $RMSE_I$ or $RMSE_{II}$ is used, and the statistical differentials III is the worst

when $RMSE_{III}$ or $RMSE_{IV}$ is used. However, it should be noted that the statistical differentials II method is the best one among the adjusted methods when $RMSE_{III}$ is used as the basis of comparison.

It is important to note that it is not possible to reach a definite conclusion as to which method is the best one in either class. It is true that the classification method is the best whereas the statistical differentials approach is the worst for most of the measures of aggregation error. But not for all the measures of aggregation error. The ranking of the methods changes when a different measure for aggregation error is used. For example, the representative agent method is the most accurate one among the unadjusted methods when $RMSE_I$ is used as the aggregation error measure whereas it is the least accurate method when $RMSE_{III}$ is used. Given that all the measures aggregation error used here are equally plausible, none of the methods can be said to dominate all the other methods.

2.6 Conclusions

Of the previously developed aggregation methods, the representative agent, classification, and statistical differentials methods differ from each other by their representation of the density function of the attributes. By formally laying out the assumptions on which these methods are based, it is shown that taking the heterogeneity of the choice sets into account is, in fact, quite simple. Furthermore, incorporation of heterogeneity of choice sets is not only simple, but also yields more accurate predictions in principle. These results have also been confirmed with empirical results that are obtained from

a multinomial logit model. Adjusted representative agent, classification, and statistical differential methods defined in this chapter take care of nonhomogeneity of the choice sets and do not need additional data requirements compared to unadjusted methods that have previously been developed. Therefore, preference should be given to these methods over the unadjusted aggregation methods considered here.

Table 2.1

Predicted aggregate demand resulting from unadjusted and
adjusted Representative Agent methods vs. Actual aggregate demand
(Percentage deviation from the actual value)

Alternative	Unadjusted	Adjusted	Actual
BPT*	69.53 (894)	17.66 (152)	7.00
TST	1.46 (86)	5.75 (43)	10.15
TPT	2.09 (85)	8.20 (41)	13.85
TRV	3.16 (88)	13.63 (50)	27.19
TCL	0.97 (88)	3.81 (51)	7.80
TGF	4.75 (10)	2.55 (52)	5.28
TOG	6.84 (20)	3.67 (36)	5.73
TMP	1.26 (34)	0.67 (64)	1.90
SST	1.22 (68)	3.42 (10)	3.79
SPT	1.65 (68)	4.62 (10)	5.15
SRV	2.49 (73)	6.95 (24)	9.13
SCL	0.87 (75)	2.42 (29)	3.41
SGF	3.19 (175)	0.99 (15)	1.16
SOG	6.54 (222)	2.03 (0)	2.03
SMP	0.83 (52)	0.26 (53)	0.54
MST	1.51 (51)	3.58 (17)	3.07
MPT	2.09 (52)	4.96 (15)	4.32
MRV	2.82 (56)	6.70 (5)	6.40
MCL	1.06 (60)	2.52 (5)	2.64
MGF	1.11 (226)	0.13 (63)	0.34
MOG	2.13 (200)	0.24 (66)	0.71
MMP	0.37 (51)	0.04 (83)	0.25
UST	8.49 (41)	10.30 (72)	6.00
UPT	35.70 (55)	43.33 (88)	23.00
TBPT	7.58 (60)	16.90 (11)	19.00
TMST	2.05 (2)	3.12 (55)	2.01
TMPT	3.26 (9)	4.97 (39)	3.57
TMRV	4.76 (1)	7.25 (51)	4.81
TMCL	1.33 (22)	2.02 (19)	1.70
TMGF	0.41 (95)	0.07 (67)	0.21
TMOG	1.32 (104)	0.22 (65)	0.65
TMMP	0.15 (26)	0.03 (88)	0.21
SUM	183	183	183
Av.P.D**	98	45	

* Alternative IJ should be read as 'by J to I', i. e., BPT means Barge-to-Portland. I takes the values ST (Seattle), PT (Portland), RV (River), CL (California), GF (Great Falls), OG (Ogden), and MP (Minneapolis). J takes values B=Barge, T=Truck, S=Single car rail, M=Multiple car rail, U=Unit train, TB=Truck-Barge, and TM=Truck-Multiple car.

** Average percentage deviation from the actual value.

Table 2.2

Predicted aggregate demand resulting from unadjusted and adjusted Classification methods vs. Actual aggregate demand
(Percentage deviation from the actual value)

Alternative	Unadjusted	Adjusted	Actual
BPT*	33.69 (381)	6.32 (10)	7.00
TST	2.36 (77)	6.85 (32)	10.15
TPT	3.37 (76)	9.77 (29)	13.85
TRV	5.13 (81)	16.37 (40)	27.19
TCL	1.66 (79)	4.81 (38)	7.80
TGF	2.71 (49)	1.07 (80)	5.28
TOG	4.48 (22)	1.77 (69)	5.73
TMP	0.73 (61)	0.29 (85)	1.90
SST	2.11 (44)	4.35 (15)	3.79
SPT	2.83 (45)	5.84 (13)	5.15
SRV	4.55 (50)	9.39 (3)	9.13
SCL	1.64 (52)	3.38 (1)	3.41
SGF	2.42 (108)	0.55 (52)	1.16
SOG	4.37 (115)	1.00 (51)	2.03
SMP	0.52 (4)	0.12 (78)	0.54
MST	2.33 (24)	4.08 (33)	3.07
MPT	3.36 (22)	5.88 (36)	4.32
MRV	4.50 (30)	7.88 (23)	6.40
MCL	1.87 (29)	3.27 (24)	2.64
MGF	3.55 (936)	0.30 (14)	0.34
NOG	7.31 (930)	0.61 (14)	0.71
MMP	0.40 (63)	0.03 (86)	0.25
UST	9.94 (66)	8.92 (49)	6.00
UPT	43.42 (89)	38.93 (69)	23.00
TBPT	8.72 (54)	14.36 (24)	19.00
TMST	4.29 (113)	4.83 (140)	2.01
TMPT	6.18 (73)	6.96 (95)	3.57
TMRV	10.59 (120)	11.92 (148)	4.81
TMCL	2.62 (54)	2.95 (74)	1.70
TMGF	0.30 (43)	0.04 (82)	0.21
TMOG	0.96 (49)	0.12 (81)	0.65
TMMP	0.08 (60)	0.01 (95)	0.21
SUM	183	183	183
Av.P.D.**	125	53	

* Alternative IJ should be read as 'by J to I', i. e., BPT means Barge-to-Portland, I takes the values ST (Seattle), PT (Portland), RV (River), CL (California), GF (Great Falls), OG (Ogden), and MP (Minneapolis). J takes values B=Barge, T=Truck, S=Single car rail, M=Multiple car rail, C=Unit train, TB=Truck-Barge, and TM=Truck-Multiple car.

** Average percentage deviation from the actual value.

Table 2.3

Predicted aggregate demand resulting from unadjusted and
adjusted Statistical Differentials I methods vs. Actual aggregate demand
(Percentage deviation from the actual value)

Alternative	Unadjusted	Adjusted	Actual
BPT*	4.04 (42)	1.59 (77)	7.00
TST	0.96 (91)	5.79 (43)	10.15
TPT	1.19 (91)	7.17 (48)	13.85
TRV	3.52 (87)	23.47 (14)	27.19
TRCL	0.54 (93)	3.28 (58)	7.80
TGF	12.39 (135)	10.25 (94)	5.28
TOG	25.97 (354)	21.49 (275)	5.73
TMP	1.28 (32)	1.06 (44)	1.90
SST	0.38 (90)	1.63 (57)	3.79
SPT	0.43 (92)	1.84 (64)	5.15
SRV	3.52 (61)	15.16 (66)	9.13
SCL	0.24 (93)	1.06 (69)	3.41
SGF	17.86 (1440)	8.56 (638)	1.16
SOG	38.97 (1815)	18.67 (818)	2.03
SMP	1.76 (224)	0.84 (55)	0.54
MST	0† (100)	0† (100)	3.07
MPT	0† (100)	0† (100)	4.32
MRV	0.62 (90)	2.26 (65)	6.40
MCL	0.02 (99)	0.07 (97)	2.64
MGF	11.04 (3119)	1.92 (461)	0.34
MOG	26.13 (3580)	4.55 (541)	0.71
MMP	0.94 (281)	0.16 (34)	0.25
UST	0.64 (89)	1.19 (80)	6.00
UPST	5.53 (76)	10.35 (55)	23.00
TBPT	2.20 (88)	7.56 (60)	19.00
TBMT	1.77 (12)	4.17 (107)	2.01
TMPT	2.30 (35)	5.42 (52)	3.57
TMRV	7.84 (63)	18.44 (284)	4.81
TMCL	1.03 (40)	2.42 (42)	1.70
TMGF	1.52 (626)	0.40 (90)	0.21
TMOG	8.39 (1193)	2.19 (238)	0.65
TMMP	0.13 (36)	0.03 (83)	0.21
SUM	183.13	183	183
Av.P.D**	449	153	

* Alternative IJ should be read as 'by J to I', i. e., BPT means Barge-to-Portland, I takes the values ST (Seattle), PT (Portland), RV (River), CL (California), GF (Great Falls), OG (Ogden), and MP (Minneapolis). J takes values B=Barge, T=Truck, S=Single car rail, M=Multiple car rail, U=Unit train, TB=Truck-Barge, and TM=Truck-Multiple car.

** Average percentage deviation from the actual value.

† 0 was assigned when the method produced a negative probability.

Table 2.4

Predicted aggregate demand resulting from unadjusted and
adjusted Statistical Differentials II methods vs. Actual aggregate demand.
(Percentage deviation from the actual value)

Alternative	Unadjusted	Adjusted	Actual
BPT*	29.61 (323)	18.92 (170)	7.00
TST	1.59 (84)	1.69 (83)	10.15
TPT	2.10 (85)	3.93 (72)	13.85
TRV	6.59 (76)	15.74 (42)	27.19
TCL	1.05 (86)	2.24 (71)	7.80
TGF	3.68 (30)	0† (100)	5.28
TOG	20.90 (265)	4.79 (16)	5.73
TMP	0.88 (54)	0† (100)	1.90
SST	1.05 (72)	0.51 (87)	3.79
SPT	1.32 (74)	1.53 (70)	5.15
SRV	5.03 (45)	8.26 (9)	9.13
SCL	1.00 (71)	1.68 (51)	3.41
SGF	2.90 (150)	0† (100)	1.16
SOG	9.65 (374)	1.49 (27)	2.03
SMP	0.83 (53)	0† (100)	0.54
MST	1.27 (59)	0† (100)	3.07
MPT	2.89 (33)	1.41 (67)	4.32
MRV	3.54 (45)	2.34 (63)	6.40
MCL	1.18 (55)	0.62 (76)	2.64
MGF	22.12 (6352)	3.55 (934)	0.34
MOG	36.87 (5093)	5.92 (734)	0.71
MMP	0.03 (88)	0† (100)	0.25
UST	8.97 (49)	16.13 (169)	6.00
UPT	33.93 (48)	56.35 (145)	23.00
TBPT	11.98 (37)	4.85 (74)	19.00
TMST	3.59 (79)	4.27 (113)	2.01
TMPT	4.37 (22)	6.15 (72)	3.57
TMRV	16.01 (233)	17.26 (259)	4.81
TMCL	2.41 (41)	3.38 (98)	1.70
TMGF	0.03 (84)	0† (100)	0.21
TMOG	0.68 (5)	0† (100)	0.65
TMMP	0.02 (92)	0† (100)	0.21
SUM	238.06	183	183
Av.P.D.**	-46	138	

* Alternative IJ should be read as 'by J to I', i. e., BPT means Barge-to-Portland. I takes the values ST (Seattle), PT (Portland), RV (River), CL (California), GF (Great Falls), OG (Ogden), and MP (Minneapolis). J takes values B=Barge, T=Truck, S=Single car rail, M=Multiple car rail, C=Unit train, TB=Truck-Barge, and TM=Truck-Multiple car.

** Average percentage deviation from the actual value.

† 0 was assigned when the method produced a negative probability.

Table 2.5

Predicted aggregate demand resulting from unadjusted and
adjusted Statistical Differentials III methods vs. Actual aggregate demand
(Percentage deviation from the actual value)

Alternative	Unadjusted	Adjusted	Actual
BPT*	27.52 (293)	15.27 (118)	7.00
TST	1.96 (81)	5.97 (41)	10.15
TPT	2.74 (80)	8.71 (37)	13.85
TRV	6.42 (76)	21.19 (22)	27.19
TCL	1.24 (84)	3.85 (51)	7.80
TGF	3.28 (38)	0† (100)	5.28
TOG	14.72 (157)	4.63 (19)	5.73
TMP	0.75 (60)	0† (100)	1.90
SST	1.31 (65)	2.81 (26)	3.79
SPT	1.54 (70)	3.43 (33)	5.15
SRV	7.12 (22)	14.66 (60)	9.13
SCL	1.02 (70)	2.29 (33)	3.41
SGF	3.08 (166)	1.24 (7)	1.16
SOG	9.13 (349)	3.28 (61)	2.03
SMP	0.74 (35)	0.20 (62)	0.54
MST	1.32 (57)	2.72 (11)	3.07
MPT	2.01 (53)	3.72 (14)	4.32
MRV	3.96 (38)	7.55 (18)	6.40
MCL	1.19 (55)	2.23 (16)	2.64
MGF	10.94 (3091)	1.65 (382)	0.34
MOG	34.04 (4696)	5.13 (623)	0.71
MMP	0.15 (40)	0† (100)	0.25
UST	6.57 (10)	7.67 (28)	6.00
UPT	27.66 (20)	29.62 (29)	23.00
TBPT	5.13 (73)	11.56 (39)	19.00
TMST	2.93 (46)	3.19 (59)	2.01
TMPT	4.04 (13)	4.84 (35)	3.57
TMRV	12.37 (157)	13.51 (181)	4.81
TMCL	1.73 (2)	2.13 (25)	1.70
TMGF	0.10 (53)	0† (100)	0.21
TMOG	0.50 (22)	0.00 (100)	0.65
TMMP	0.05 (75)	0† (100)	0.21
SUM	197.28	183	183
Av.P.D.**	317	82	

* Alternative IJ should be read as 'by J to I', i. e., BPT means Barge-to-Portland. I takes the values ST (Seattle), PT (Portland), RV (River), CL (California), GF (Great Falls), OG (Ogden), and MP (Minneapolis). J takes values B=Barge, T=Truck, S=Single car rail, M=Multiple car rail, U=Unit train, TB=Truck-Barge, and TM=Truck-Multiple car.

** Average percentage deviation from the actual value.

† 0 was assigned when the method produced a negative probability.

Table 2.6

Aggregation errors by different criteria
and aggregation methods.

Error Measure	Representative Agent	
	Unadjusted	Adjusted
$RMSE_I$	10.33	3.12
$RMSE_{II}$	25.57	7.64
$RMSE_{III}$	72.00	28.83
$RMSE_{IV}$	177.77	84.03
Error Measure	Classification	
	Unadjusted	Adjusted
$RMSE_I$	14.21	3.64
$RMSE_{II}$	16.76	7.15
$RMSE_{III}$	45.83	23.54
$RMSE_{IV}$	153.23	79.23
Error Measure	Statistical Differentials I	
	Unadjusted	Adjusted
$RMSE_I$	54.90	13.84
$RMSE_{II}$	49.54	18.92
$RMSE_{III}$	67.58	37.02
$RMSE_{IV}$	262.06	153.22
Error Measure	Statistical Differentials II	
	Unadjusted	Adjusted
$RMSE_I$	56.54	8.19
$RMSE_{II}$	45.47	9.28
$RMSE_{III}$	52.61	20.93
$RMSE_{IV}$	179.45	83.64
Error Measure	Statistical Differentials III	
	Unadjusted	Adjusted
$RMSE_I$	81.72	13.15
$RMSE_{II}$	58.81	15.87
$RMSE_{III}$	60.30	46.88
$RMSE_{IV}$	212.10	165.56

Chapter 3

Aggregation in Mixed Continuous/Discrete Choice Models

3.1 Introduction

This chapter deals with the aggregation problem in mixed continuous/discrete choice models. These models are similar to pure discrete choice models for which the aggregation problem is addressed in chapter 2 in the sense that they involve an optimal choice on a discrete variable. The difference is that these models also contain an optimal choice on a continuous variable in addition to the discrete choice. Thus, the underlying economic problem in mixed continuous/discrete choice models is a more generalized case where the individuals face both discrete and continuous alternatives than the pure discrete choice case where there are only discrete alternatives.

The continuous choice in a mixed continuous/discrete choice model typically refers to the decision on the magnitude of the discrete alternative that is to be chosen amongst the others.¹ These are two simultaneous choices that are jointly determined by the same underlying factors. Thus, both the probability that the decisionmaker will choose an alternative and the magnitude of the alternative are modelled in a mixed continuous/discrete choice model. However, the final outcome of the model is the magnitude of the alternative, which represents the demand for the alternative that will be chosen by the decisionmaker.

1. For example, in a labor supply model the discrete decision is to work or not to work and the continuous decision is how much to work.

Since the demand for an alternative is a direct result of a mixed continuous/discrete choice model, the aggregation problem here is somewhat different than the aggregation problem in a pure discrete choice model where the demand of a decisionmaker for an alternative is simply the probability of choosing the alternative.² For finding aggregate demand in a mixed continuous/discrete choice model, one needs to work with individual demands that are typically nonlinear functions of the underlying probabilities. The strategy of this chapter is to apply the aggregation approaches employed in chapter 2 to develop aggregation methods for these models.

The remainder of the chapter is organized as follows: Section two outlines the general structure of mixed continuous/discrete choice models. Section three defines the aggregation problem and develops different aggregation methods based on the approaches employed in chapter 2. Section four presents a theoretical discussion on the performance of the developed aggregation methods. Empirical results on the relative performance of the aggregation methods are presented in the fifth section. The final sections summarize the main results of the analysis.

3.2 General Structure of Mixed Continuous/Discrete Choice Models

Mixed continuous/discrete choice models have been developed to describe economic choice problems in which an optimal continuous choice depends to some degree on the results of an optimal discrete choice. These models have been shown to have

2. Alternatively, the demand in a pure discrete choice model refers to the expected number of times that an alternative is chosen.

a consistent theoretical base that can be derived from an underlying theoretical model of utility maximization by an individual consumer or profit maximization by individual firms (Hanemann 1984a, Duncan 1980). This section summarizes the utility maximizing consumer case.³

Let $x_1, \dots, x_n$, and z denote the commodities where z is taken as numeraire. The utility of the consumer is assumed to depend on the amount of goods in his or her consumption set, the attributes of the goods, denoted by $a_1, \dots, a_n$, and his or her own characteristics, denoted by the vector s . So the utility of the consumer can be written as $U(x, z, a, s)$. The consumer chooses the bundle (x, z) so as to maximize his or her utility subject to the budget constraint

$$\sum p_i x_i + z = y \quad (3.2.1)$$

where p_i is the price of good x_i and y is the level of income of the consumer. Other than the budget constraint, the consumer faces the nonnegativity constraints

$$x_i \geq 0 \quad i = 1, \dots, N \quad (3.2.2)$$

$$z \geq 0. \quad (3.2.3)$$

The discreteness appears when some of the relations in (3.2.2) hold as strict equalities. The discrete choice is the choice of which x_i 's are zero; the continuous choice is the decision on the magnitude of the nonzero x_i 's.⁴ This representation allows for choosing more than one good. But, a typical assumption in almost all of the literature

3. This summary is largely based on Hanemann (1984a).

4. Note that if x_i 's are assumed to take fixed values such as $\bar{x}_i$ then what we have is a pure discrete choice model.

on mixed continuous/discrete choice is that only one of the commodities (alternatives) is selected. This results in the following additional constraint.

$$x_i x_j = 0 \quad \text{all } i \neq j. \quad (3.2.4)$$

As in a pure discrete choice model, although the consumer's utility is deterministic for her, it contains some elements that are unobservable to the researcher. If the vector ϵ denotes these unobservable components, the utility of the consumer can be written as $U(x, z, a, s, \epsilon)$.⁵

The consumer can choose to consume any one of the goods. Suppose that he or she has decided to consume only good i . Then the maximum utility that the consumer can obtain, given that good i is chosen, depends on the price and attributes of good i , the person's income and characteristics. This maximum attainable utility can be written as $\bar{v}_i(p_i, a_i, y, s, \epsilon)$ which is a conditional indirect utility function. It is the indirect utility that the consumer faces given that he or she has chosen alternative i . It is called the conditional indirect utility because it is conditional on the choice of alternative i .

The conditional demand functions are $\bar{x}_i(p_i, a_i, y, s, \epsilon)$ and $\bar{z}(p_i, a_i, y, s, \epsilon) = y - p_i \bar{x}_i(p_i, a_i, y, s, \epsilon)$, both are also conditional on the choice of alternative i . Then the conditional indirect utility function can be written as

$$\bar{v}_i(p_i, a_i, y, s, \epsilon) \equiv \bar{U}_i[\bar{x}_i(p_i, a_i, y, s, \epsilon), a_i, \bar{z}(p_i, a_i, y, s, \epsilon), s, \epsilon] \quad (3.2.5)$$

5. For the individual consumer ϵ is a set of fixed constants (or functions), but for the researcher ϵ is a random variable.

and due to Roy's identity,⁶

$$\bar{x}_i(p_i, a_i, y, s, \epsilon) = - \frac{\partial \bar{v}_i(p_i, a_i, y, s, \epsilon) / \partial p_i}{\partial \bar{v}_i(p_i, a_i, y, s, \epsilon) / \partial y}. \quad (3.2.6)$$

The discrete choice of which alternative to select depends on the relative conditional indirect utilities that the alternatives will provide. If this discrete choice is represented by a set of binary valued indices $\delta_1, \dots, \delta_N$, where $\delta_i = 1$ if $x_i > 0$ and $\delta_i = 0$ otherwise, then the choice can be expressed as

$$\delta_i(p, a, y, s, \epsilon) = \begin{cases} 1, & \text{if } \bar{v}_i(p_i, a_i, y, s, \epsilon) \geq \bar{v}_j(p_j, a_j, y, s, \epsilon), \text{ all } j. \\ 0, & \text{otherwise.} \end{cases} \quad (3.2.7)$$

Thus, the probability of choosing good i is simply

$$p_i(p, a, y, s, \epsilon) = \text{Prob}\{\bar{v}_i(p_i, a_i, y, s, \epsilon) \geq \bar{v}_j(p_j, a_j, y, s, \epsilon), \forall j\} \quad (3.2.8)$$

which, of course, depends on the joint density of ϵ 's, $f_\epsilon(\epsilon_1, \dots, \epsilon_n)$.⁷

The unconditional ordinary demand functions and the unconditional indirect utility function can then be expressed as

$$x_i(p, a, y, s, \epsilon) = \delta_i(p, a, y, s, \epsilon) \bar{x}_i(p_i, a_i, y, s, \epsilon) \quad (i = 1, \dots, N) \quad (3.2.9)$$

$$V(p, a, y, s, \epsilon) = \max[\bar{v}_1(p_1, a_1, y, s, \epsilon), \dots, \bar{v}_N(p_N, a_N, y, s, \epsilon)]. \quad (3.2.10)$$

As Hanemann points out, the crucial building blocks for mixed continuous/discrete discrete choice model are the conditional indirect utility functions $\bar{v}_i(\cdot)$, $i =$

6. Note that the quantities $\bar{x}_i$, $\bar{z}$, and $\bar{v}_i$ are known numbers to the consumer, but they are random variables from the point view of the researcher. But if the random terms, ϵ 's, enter the conditional indirect utility functions appropriately, they might disappear from the ratio on the right-hand side of (3.2.6), resulting in nonrandom quantities from researcher's point of view too.

7. δ_i is known to the consumer but it is a random variable for the researcher. Thus, p_i can be thought as $p_i = E(\delta_i)$.

1, ..., N, and the joint density function of the ε 's, $f_\varepsilon(\varepsilon_1, \dots, \varepsilon_n)$. Once these have been specified, the discrete choice probabilities and the conditional and unconditional densities of the x_i 's can be formed by using equations (3.2.6) and (3.2.9), and applying a change of variable method. The way the random terms enter the conditional utility functions and different assumptions on their joint distribution, of course, result in different mixed continuous/discrete choice models.

Note that an alternative way of expressing equation (3.2.9) is

$$\bar{x}_i(p_i, a_i, y, s, \varepsilon) = \begin{cases} \bar{x}_i(p_i, a_i, y, s, \varepsilon), & \text{if } \bar{v}_i(p_i, a_i, y, s, \varepsilon) \geq \bar{v}_j(p_j, a_j, y, s, \varepsilon) \forall j \neq i; \\ 0, & \text{otherwise} \end{cases}$$

which results in the following general expression for the unconditional demand functions:

$$x(\cdot) = \begin{cases} \bar{x}_1(p_1, a_1, y, s, \varepsilon), & \text{if } \bar{v}_1(p_1, a_1, y, s, \varepsilon) \geq \bar{v}_j(p_j, a_j, y, s, \varepsilon) \forall j \neq 1 \\ \bar{x}_2(p_2, a_2, y, s, \varepsilon), & \text{if } \bar{v}_2(p_2, a_2, y, s, \varepsilon) \geq \bar{v}_j(p_j, a_j, y, s, \varepsilon) \forall j \neq 2 \\ \vdots & \\ \bar{x}_N(p_N, a_N, y, s, \varepsilon), & \text{if } \bar{v}_N(p_N, a_N, y, s, \varepsilon) \geq \bar{v}_j(p_j, a_j, y, s, \varepsilon) \forall j \neq N \end{cases} \quad (3.2.11)$$

In terms of the estimation of mixed continuous/discrete choice models, it suffices to say that equation (3.2.11) can easily be expressed as a statistical switching regression model which can be estimated by any of the econometric techniques developed for such models.⁸ The emphasis of this analysis is on the empirical application of an already estimated mixed continuous/discrete choice model, so the estimation technique used for estimating the model is not important.

8. See Hanemann (1984a), Inaba and Wallace (1987), Maddala (1983), and Duncan (1980).

3.3 Aggregation in Mixed Continuous/Discrete Choice Models

Suppose that a mixed continuous/discrete choice model has already been estimated so that the unconditional demand for any alternative i is obtained as a function of the exogenous variables in the model and some estimated coefficients:

$$x_i(\mathbf{p}, \mathbf{a}, y, \mathbf{s}, \beta, \alpha) = p_i(\mathbf{p}, \mathbf{a}, y, \mathbf{s}, \beta) \bar{x}_i(p_i, a_i, y, \mathbf{s}, \alpha) \quad i = 1, \dots, N \quad (3.3.1)$$

where α and β are vectors of estimated coefficients. Let, for convenience, ω be the vector of all the exogenous variables in the model, i.e., $\omega = (p_1, \dots, p_N, a_1, \dots, a_N, y, \mathbf{s})$. Then equation (3.3.1) can be rewritten as

$$x_i = G^i(\omega, \alpha, \beta) \quad i = 1, \dots, N \quad (3.3.2)$$

where $G^i(\cdot)$ is a function that represents $p_i(\cdot) \bar{x}_i(\cdot)$.

Equation (3.3.2) can be used to predict the unconditional demand of any decision-maker for any alternative i if the value of ω vector is known for that decisionmaker. A common use of the estimation results involves providing simulations for some chosen population or reporting aggregate predictions of unconditional demand rather than reporting the unconditional demand for a single individual. So, a typical application is to forecast the aggregate unconditional demand for a population of T individuals. This implies that equation (3.3.2) needs to be aggregated over the population of concern.

If ω_t denotes the set of exogenous variables for individual t , then the unconditional demand of this individual for alternative i is simply

$$x_{it} = G^i(\omega_t, \alpha, \beta) \quad i = 1, \dots, N; \quad t = 1, \dots, T \quad (3.3.3)$$

where T is the total number of decisionmakers in the population of concern. Then the aggregate demand for any alternative i can be expressed as

$$X_i = \sum_{t=1}^T x_{ti} = \sum_{t=1}^T G^i(\omega_t, \alpha, \beta) \quad i = 1, \dots, N. \quad (3.3.4)$$

As in the pure discrete choice case, the aggregation problem arises because of the fact that the values of the exogeneous variables in ω are not known for all the individuals in the population of concern. Thus, it is not feasible to evaluate equation (3.3.4) to predict aggregate demand. The aggregation problem in this context is really not different from the aggregation problem in pure discrete choice models. The only difference is that the function $G(\cdot)$ directly refers to quantity demand function in mixed continuous/discrete choice models whereas it is the choice function in pure discrete choice models. Other than this conceptual distinction, they are identical in the sense that both of these functions are functions of some exogeneous variables and some estimated coefficients.

This suggests that the aggregation approaches presented in chapter 2 are valid for mixed continuous/discrete choice models as well. Recall that all those approaches but the statistical differentials approach were based on different representations of the joint density function of the attributes (exogeneous variables). The statistical differentials approach used a Taylor's series expansion for which the only requirement is having a continuously differentiable function. Thus, aggregation methods for mixed continuous/discrete choice models can essentially be obtained just by replacing the choice function in pure discrete choice models by the function $G(\cdot)$. The only modification

that is needed is to eliminate the restriction on the overall sum of the demand for all the alternatives in pure discrete choice models. In contrast to pure discrete choice models, there is no information on the overall demand in mixed continuous/discrete choice models.⁹

It should also be noted that the function $G(\cdot)$ reduces to the choice function in pure discrete choice models when the conditional demand function, $\bar{x}_i(\cdot)$, is assumed to be a constant number. Hence, the aggregation problem in pure discrete choice models is a special case of the aggregation problem in mixed continuous/discrete choice models. Thus, the theoretical discussion of the different aggregation methods for mixed continuous/discrete choice models that will be presented in the next section will also be valid for pure discrete choice models.

Before developing aggregation methods for mixed continuous/ discrete choice models, equation (3.3.4) needs to be rewritten as

$$X_i = \sum_{t=1}^{T_i} G^i(\omega_t, \alpha, \beta) \quad i = 1, \dots, N \quad (3.3.5)$$

to incorporate the heterogeneity of choice sets. Here, T_i denotes the total number of decisionmakers in the population for whom alternative i is feasible.¹⁰ As chapter 2 suggests, this is necessary because if an alternative is not feasible for an individual, then the demand for that alternative is 0 for this individual.

9. Recall that in pure discrete choice models, the sum of the demands for each alternative has to add up to the total number of individuals in the population. (See equation (2.1.5).)

10. $T_i \leq T \forall i$.

The fact that the values of the variables in ω is not necessarily known for each individual in the population implies that ω is a random vector in evaluating equation (3.3.5). Thus, the aggregation problem is to estimate the total expected demand for alternative i ,

$$E[X_i] = E\left[\sum_{t=1}^{T_i} G^i(\omega_t, \alpha, \beta)\right] = \sum_{t=1}^{T_i} E[G^i(\omega_t, \alpha, \beta)]. \quad (3.3.6)$$

If ω_t is assumed to be continuous with joint density function $f_t(\omega_t)$, then equation (3.3.6) can be written as

$$E[X_i] = \sum_{t=1}^{T_i} \int_{\omega_t} G^i(\omega_t, \alpha, \beta) f_t(\omega_t) d\omega_t. \quad (3.3.7)$$

As in the pure discrete case, the representative agent method is obtained by assuming that $f_t(\omega_t)$ is approximated by a degenerate discrete probability mass function:

$$f_t(\omega_t) \approx \begin{cases} 1, & \text{if } \omega_t = \bar{\omega}; \\ 0, & \text{otherwise.} \end{cases} \quad t = 1, \dots, T; \quad (3.3.8)$$

where $\bar{\omega} = (\bar{\omega}^1, \dots, \bar{\omega}^S)$ and $\bar{\omega}^s = \frac{1}{T_i} \sum_{t=1}^{T_i} \omega_t^s$, $s = 1, \dots, S$. Then it follows that

$$X_i^R = T_i G^i(\bar{\omega}, \alpha, \beta) \quad i = 1, \dots, N \quad (3.3.9)$$

where X_i^R denotes the total expected demand for alternative i as predicted by the representative agent method.

The assumption that underlies the classification method can be stated as:

$$f_t(\omega_t) \approx \begin{cases} T_{ki}/T_i, & \text{if } \omega_t = \bar{\omega}_k, \quad k = 1, \dots, K; \\ 0, & \text{otherwise.} \end{cases} \quad t = 1, \dots, T; \quad i = 1, \dots, N \quad (3.3.10)$$

where $\bar{\omega}_k$ is the mean vector of the attributes of the individuals in group k for whom alternative i is feasible and T_{ki} is the number of those individuals.¹¹ Assumption

11. Here, $\sum_{k=1}^K T_{ki} = T_i$.

(3.3.10) results in the following expression for the classification method for mixed continuous/discrete choice models:

$$X_i^C = \sum_{k=1}^K T_{ki} G^i(\bar{\omega}_k, \alpha, \beta) \quad i = 1, \dots, N \quad (3.3.11)$$

where X_i^C denotes the total expected demand for alternative i as predicted by the classification method.

As seen in Chapter 2, another aggregation approach is the statistical differentials method which is based on a Taylor's series approximation of the function $G^i(\omega_i, \alpha, \beta)$. There are three variations of this method based on different assumptions on the distribution of the attributes. The first variation assumes that there is a single common distribution for the attributes with the variance-covariance matrix $\text{Cov}(\omega^r, \omega^s)$:

$$X_i^{SDI} = T_i \left[G^i(\bar{\omega}, \alpha, \beta) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial \omega_i^r \partial \omega_i^s} G^i(\omega_i, \alpha, \beta) \Big|_{\bar{\omega}} \text{Cov}(\omega^r, \omega^s) \right] \quad i = 1, \dots, N \quad (3.3.12)$$

where X_i^{SDI} denotes the total expected demand for alternative i as predicted by statistical differentials I method.

The second variation assumes some nearly homogeneous sub-groups as the classification method does, for each of which a different common distribution is assumed with a variance-covariance matrix $\text{Cov}(\omega_k^r, \omega_k^s)$. This results in:

$$X_i^{SDII} = \sum_{k=1}^K T_{ki} \left[G^i(\bar{\omega}_k, \alpha, \beta) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial \omega_i^r \partial \omega_i^s} G^i(\omega_i, \alpha, \beta) \Big|_{\bar{\omega}_k} \text{Cov}(\omega_k^r, \omega_k^s) \right] \quad (3.3.13)$$

where $\text{Cov}(\omega_k^r, \omega_k^s)$ the covariance of the r th and s th attributes in sub-group k .

The third variation assumes that each distribution for each sub-group has the same variance-covariance matrix:

$$X_i^{SDIII} = \sum_{k=1}^K T_{ki} \left[G^i(\bar{\omega}_k, \alpha, \beta) + \frac{1}{2} \sum_r \sum_s \frac{\partial^2}{\partial \omega_r^r \partial \omega_s^s} G^i(\omega_t, \alpha, \beta) \Big|_{\bar{\omega}_k} \text{Cov}(\omega^r, \omega^s) \right] \quad (3.3.14)$$

where X_i^{SDIII} denotes the total expected demand for alternative i as predicted by the statistical differentials III method.

3.4 Comparisons of the Aggregation Methods

Section 3 presented five different aggregation methods that can be used in mixed continuous/discrete choice models. Two of these methods, namely the representative agent and classification methods, have already been used in some applications in previous work. For example, Zorn (1988) and Winston (1981) have employed the representative agent method whereas Goodman (1988), Duncan and Leigh (1980), and King (1980) have used the classification method in their studies. Although the statistical differentials approach has not been applied yet, it certainly has the potential for future use.

Given that there are a number of aggregation methods one can use, it is important to know if any of them will perform better than the others. There is not yet a consensus as to how disaggregate choice models should be aggregated up to population estimates, including the pure discrete choice models for which this issue has been taken up in

previous work. This section provides a theoretical discussion on the relative performance of the aggregation methods presented in the previous section.¹²

Recall that the aggregate demand for any alternative i is simply the sum of the demand of each individual in the population for that alternative. Suppose that there are $k, k = 1, \dots, K$, sub-groups that partition the population. Then, the aggregate demand for alternative i can be redefined as the sum of the total demand in each sub-group k , which is equivalent to the original definition. Thus, equation (3.3.6), which states the aggregation problem, can be rewritten as

$$E[X_i] = \sum_{k=1}^K \sum_{t=1}^{T_{ki}} E[G^i(\omega_{tk})] \quad i = 1, \dots, N \quad (3.4.1)$$

where $\sum_{k=1}^K T_{ki} = T_i$, and ω_{tk} is the vector of attributes of individual t in sub-group k .¹³

The classification method is based on the assumption that these sub-groups are nearly homogeneous with respect to the attributes of the individuals. Assumption (3.3.10) on which the classification method is based implies that a common distribution whose mean value is $\bar{\omega}_k$ is assumed for the attributes in sub-group k . Thus, there is a different common distribution for each sub-group with a different mean value. Each individual in a sub-group is assumed to behave like the representative individual whose attributes take the values equal to the mean value of the distribution. The resulting aggregation procedure, as stated by equation (3.3.11), is to predict the demand of the

12. Earlier studies on pure discrete choice models were not only inconclusive but they were also based on only empirical comparisons. Some of these studies are Koppelman (1976a, 1976b), Reid (1978), and Dunne (1985). The methods that these studies investigated did not incorporate the heterogeneity of the choice sets, either.

13. α and β are dropped from the list of arguments of $G^i(\cdot)$ for simplicity.

representative individual in each group, and then take the weighted sum of these demands where the weights are the number of individuals in each group for whom the alternative is feasible:

$$X_i^C = \sum_{k=1}^K T_{ki} G^i(\bar{\omega}_k)$$

which can be rewritten as

$$X_i^C = \sum_{k=1}^K \sum_{t=1}^{T_{ki}} G^i(E(\omega_{tk})). \quad (3.4.2)$$

Thus, how well the classification method predicts the aggregate demand can be determined by comparing equation (3.4.2) with equation (3.4.1). Recall that if $G^i(\cdot)$ is a convex function, then by Jensen's inequality,

$$E G^i(\omega_{tk}) \geq G^i(E(\omega_{tk})). \quad (3.4.3)$$

Thus, if $G^i(\cdot)$ is convex, it follows that

$$X_i = \sum_{k=1}^K \sum_{t=1}^{T_{ki}} E[G^i(\omega_{tk})] \geq \sum_{k=1}^K \sum_{t=1}^{T_{ki}} G^i(E(\omega_{tk})) = X_i^C. \quad (3.4.5)$$

Therefore, if the function of interest is convex, then the classification will under-predict the aggregate demand. If the function is concave, then it over-predicts the aggregate demand.

Assumption (3.3.8), on which the representative agent method is based, implies that there is a single common distribution for the attributes of all the individuals in the population whose mean value is $\bar{\omega}$. For the representative agent method, it is assumed that all the individuals in the population behave like the representative individual whose attribute vector takes this mean value. Thus, the representative agent method is a

special case of the classification method where the existence of homogeneous sub-groups is ignored. As a special case of the classification method, it is immediate that Jensen's inequality applies when $G^i(\cdot)$ is convex (or concave). Thus, the representative agent method will also under-predict (over-predict) the aggregate demand when the function of interest is convex (concave) as the classification method does.

In terms of comparing the predictions of the classification and the representative agent methods, recall that the collection of sub-groups for $k = 1, \dots, K$ is a partition of the population. Thus, the mean value of the distribution that the representative agent method assumes can be written as

$$\begin{aligned}\bar{\omega} &= \frac{1}{T_i} \sum_{t=1}^{T_i} \omega_t = \frac{1}{T_i} \sum_{k=1}^K \sum_{t=1}^{T_{ki}} \omega_{tk} \\ &= \sum_{k=1}^K \frac{T_{ki}}{T_i} \frac{1}{T_{ki}} \sum_{t=1}^{T_{ki}} \omega_{tk} \\ &= \sum_{k=1}^K \frac{T_{ki}}{T_i} \bar{\omega}_k.\end{aligned}$$

Hence, $\bar{\omega}$ is a weighted average of the averages $\bar{\omega}_k$, for $k = 1, \dots, K$. Then, if $G^i(\cdot)$ is a convex function, it follows that

$$\sum_{k=1}^K \frac{T_{ki}}{T_i} G^i(\bar{\omega}_k) \geq G^i(\bar{\omega})$$

so that

$$\sum_{k=1}^K T_{ki} G^i(\bar{\omega}_k) \equiv X_i^C \geq T_i G^i(\bar{\omega}) \equiv X_i^R, \quad (3.4.6)$$

which reveals that when $G^i(\cdot)$ is convex, the representative agent method will provide smaller predictions than the classification method does. Since the classification method already under-predicts aggregate demand when $G^i(\cdot)$ is convex, the representative agent

method is worse than the classification method in this case. If $G^i(\cdot)$ is concave, the inequality in equation (3.4.6) will be reversed, implying that the representative agent method will over-predict aggregate demand more than the classification method will. Therefore, when the function of interest is either convex or concave, the classification method should be preferred to the representative agent method. However, when the function of interest is neither convex nor concave, it is not possible to draw a conclusion unless there is enough empirical evidence.

The three statistical differentials methods are essentially some extensions of the representative agent or the classification method in one way or another. They are all based on Taylor's series expansions but they differ from each other based on their underlying assumptions on the distribution of the attributes as well as the point of approximation. The main characteristic of each of them is that they all try to capture the departure of the behavior of the individuals from the behavior of the representative agent by including the covariances of the attributes.

The statistical differentials I method, defined by equation (3.3.12), is an extension of the representative agent method. This is most clear when equation (3.3.12) is rewritten as

$$X_i^{SDI} = X_i^R + \frac{1}{2} T_i \nabla^2 G^i(\bar{w}) @ \text{Cov}(w_i) \quad (3.4.7)$$

by making use of equation (3.3.9). Here, $\nabla^2 G^i(\bar{w})$ denotes the Hessian matrix of $G^i(\cdot)$ evaluated at the mean value $\bar{w}$, and the notation $@$ denotes the sum of the element-by-element product of two matrices of the same dimensions. $\text{Cov}(w_i)$ is the variance-

covariance matrix of the common distribution that this method assumes for the whole population.

Note that the second term in equation (3.4.7) can be written as

$$\frac{1}{2} T_i \nabla^2 G^i(\bar{\omega}) @ \text{Cov}(\omega_i) = \frac{1}{2} E[(\omega - \bar{\omega})' \nabla^2 G^i(\bar{\omega})(\omega - \bar{\omega})].$$

The term in the brackets is a quadratic form, which will always be positive (negative) for all values of ω in a neighborhood of $\bar{\omega}$ if the function $G^i(\cdot)$ is convex (concave) in that neighborhood. Since the expectation of a positive (negative) random variable is positive (negative), then it follows that the second term in equation (3.4.7) will be positive (negative) if the function of interest is convex (concave). Thus, the statistical differentials I method will provide larger (smaller) predictions than the representative agent method if the function $G^i(\cdot)$ is convex (concave).

However, since the representative agent method under-predicts the aggregate demand in the convex case, one cannot tell whether the statistical differentials I method will produce downward or upward biased predictions. It all depends on the magnitude of the second term in equation (3.4.7). If this term is large enough, the predictions from the statistical differentials I method might even exceed the actual values for aggregate demand, resulting in over-predictions. But, if the magnitude of this term is not too large, then the statistical differentials I method will be an improvement over the representative agent method, when $G^i(\cdot)$ is either convex or concave. Furthermore, depending on the magnitude of this term, it might even be an improvement over the classification method. The second term in equation (3.4.7) might be large enough to produce larger

predictions than the classification method but small enough not to exceed the actual value of aggregate demand. This, of course, is a purely empirical issue.

The statistical differentials II method, defined by equation (3.3.13), extends the classification method by including the covariances of the attributes in each group. By using equation (3.3.11), equation (3.3.13) can be rewritten as

$$X_i^{SDII} = X_i^C + \frac{1}{2} \sum_{k=1}^K T_{ki} \nabla^2 G^i(\bar{\omega}_k) @ \text{Cov}(\omega_k) \quad (3.4.8)$$

where $\nabla^2 G^i(\bar{\omega}_k)$ is the Hessian matrix of $G^i(\cdot)$ evaluated at the mean value of the attributes in sub-group k , and $@$ is defined as before. $\text{Cov}(\omega_k)$ is the variance-covariance matrix of the common distribution in sub-group k , $k = 1, \dots, K$.

As it was the case for statistical differentials I method, the second term in equation (3.4.8) is also a quadratic form, which will be positive (negative) if the function $G^i(\cdot)$ is convex (concave). Thus, the statistical differentials II method will provide larger (smaller) predictions than the classification method in the convex (concave) case. Given that the classification method already produces larger (smaller) predictions than the representative agent method in the convex (concave) case, the statistical differentials II method will be an improvement over both the representative agent and the classification methods if the second term in equation (3.4.8) is not too large in absolute value. However, it is not possible to make a comparison between the statistical differentials I and II methods without empirically identifying the relative magnitudes of the second terms in equations (3.4.7) and (3.4.8). One can only point out that the first term in equation (3.4.7) will be smaller than the corresponding term in equation (3.4.8) in the

convex case. But the difference between the second terms of these two equations can be either negative or positive.

The statistical differentials III method is also an extension of the classification method in the same way that the statistical differentials II method is. The difference is that the statistical differentials III method makes the additional assumption that the different distributions in k sub-groups share the same variance-covariance matrix, which is the same variance-covariance matrix that the statistical differentials I method uses.

This method can be written in terms of the classification method as follows:

$$X_i^{SDIII} = X_i^C + \frac{1}{2} \sum_{k=1}^K T_{ki} \nabla^2 G^i(\bar{\omega}_k) @ \text{Cov}(\omega_i). \quad (3.4.9)$$

The comparison of this method with the other methods is identical to the comparison between the statistical differentials II method and the others. However, depending upon the magnitude of the entries of $\text{Cov}(\omega)$ relative to the magnitude of the entries of $\text{Cov}(\omega_k)$, the statistical differentials III method might be an improvement over the statistical differentials II method.

Before closing this sub-section, it should be noted that the convexity of the function $G^i(\cdot)$ is quite important for all the statistical differentials methods for another reason. All the three variations of the statistical differentials approach are guaranteed to produce positive numbers as predictions of the aggregate demand only if the function of interest is convex. If the function $G^i(\cdot)$ is concave, the second terms in the equations (3.4.7), (3.4.8), and (3.4.9) will all be negative. This will imply that if the magnitude of these terms is not small enough, it is possible that these methods will produce negative

numbers as predictions of aggregate demand, which is a case one definitely needs to avoid. If the function $G^i(\cdot)$ is neither convex nor concave, one might still face this problem depending on the magnitude of the same term. Thus, use of any of the statistical differentials methods requires some care if the function of interest is not convex.

3.5 Empirical Comparisons of Aggregation Methods

Five aggregation methods for mixed continuous/discrete choice models have been described and discussed in the previous sub-sections of this chapter. This chapter applies these methods to a mixed continuous/discrete model of intercity freight demand model that was developed and estimated by Inaba and Wallace (1987) earlier. The primary purpose of this empirical analysis is to provide some empirical evidence as to how large the aggregation error can get and which aggregation method is likely to perform best.

The model involves the shipment decisions of the wheat elevators (firms) in Pacific Northwest. The decision variables of the firms are the shipment size, the continuous variable, and the mode-destination alternative, the discrete variable.¹⁴ There are at most 32 mode-destination alternatives that a firm can have as feasible. The total number of the firms in the population is 183.

Table 3.1 presents the predictions of the aggregation methods and the actual values of the aggregate demand based on the unconditional demand function that Inaba and Wallace have estimated. R. Agent is the representative agent method as defined

14. For more details of the model, see Inaba and Wallace (1987). Chapter 4 also has some more details.

by equation (3.3.8). Class is the classification method which uses 9 sub-groups within each of which the firms essentially serve the same markets and have available essentially the same modes of transportation. Stat1, Stat2, and Stat3 are the three variants of the statistical differentials approach as defined by equations (3.3.12), (3.3.13), and (3.3.14), respectively. The actual values of aggregate demand are obtained by performing equation (3.3.5) over all the firms in the population..

A convexity test based on the eigenvalues of the Hessian matrix of the unconditional demand function was applied to the demand function for each alternative. Unfortunately, all the tests were negative, implying neither convexity nor concavity. Thus, it is not possible to say anything on the relative performance of the aggregation methods without analyzing the empirical results presented in Table 3.1.¹⁵ Also, if the function were found to be convex or concave, it would be possible to identify some systematic sources for the aggregation error for the case at hand.

As pointed out in Chapter 2, there is no single, generally accepted criterion that determines the best aggregation method. Hence, the comparisons between the aggregation methods will be done in three ways. First, the predictions for each alternative from each method will be examined on a case by case basis. Second, the percentage deviation of each prediction from the corresponding actual value will be calculated, and the average of the deviations for all alternatives will be used as the basis of comparison. Third, the global measures of aggregation error that were considered in Chapter 2 will be compared.

15. Recall from the previous section that if the function is convex or concave for an alternative, the classification method definitely dominates the representative agent method.

Table 3.1 indicates that, without exception, both the representative agent and the classification methods produce substantially smaller predictions than the actual demand for each alternative. Furthermore, the predictions from the representative agent method are always smaller than the predictions from the classification method, as if the unconditional demand function is convex for each alternative. Thus, the classification method dominates the representative agent method by providing less downward biased predictions for each alternative.

Compared to the three versions of the statistical differentials approach, the classification method is found to produce larger predictions than the Stat1 method in 22 of the 32 cases, in 26 of the 32 cases than Stat2, and in 25 of 32 cases than the Stat3 method. For only one alternative do the Stat1 and Stat3 methods over-predict the aggregate demand whereas the Stat2 method results in over-predictions for three alternatives. Thus, the magnitude of the downward bias is lower for the classification method than for any of the statistical differentials methods for most of the alternatives. Therefore, in terms of this case by case comparison, the classification method can be said to dominate all the statistical differentials methods as well as the representative agent method.

Table 3.2 presents the percentage of the absolute value of the deviation of the prediction from the actual value for each alternative. The results indicate that the magnitude of the bias is quite substantial for most of the cases. For example, the total number of the cases where at least one method produces a prediction with less than a 50% (20%) bias, upward or downward, is only 10 (3) out of the 32 cases. Thus, all the aggregation methods have very high aggregation errors. The average bias is 64% even

for the classification method which produces the least bias on average and dominates all the other methods.

Note that two of the statistical differentials methods seem to be an improvement over the representative agent. The average percentage deviation is reduced to 74 and 71 percent from 77% when Stat1 or Stat3 is used instead of the representative agent method, respectively. This presents the only case so far where the Stat1 method, which is the only statistical differentials approach that have been applied in earlier studies on pure discrete choice studies, is found to produce better predictions than the representative agent method.

Table 3.3 presents the total aggregation errors of the aggregation methods based on four different measures of aggregation error that were discussed in Chapter 2. Each of these measures combines the individual errors for the alternatives in a different way.¹⁶ The classification method is found to be the best method under 3 of the 4 criteria. Under the criterion it is not the best, it is ranked second. Thus, consistent with the results of the previous comparisons, the classification method can be said to dominate the other methods when a global index for total aggregation error is used.

Similarly, Stat2 can be said to be the second best method since it is ranked first or second under 3 of the 4 aggregation error criteria. This is the method that extends the classification method by incorporating the covariances in each sub-group. Thus, use of the second variation of statistical differentials approach instead of the previously suggested Stat1 seems to provide a substantial improvement in the performance of the

16. For definitions of these errors, see Chapter 2.

statistical differentials approach. Since stat3 is also always ranked higher than stat1, it seems that both variations suggested here make the statistical differentials approach a better aggregation procedure.

It is also interesting to note that $RMSE_I$, the only criterion under which Stat2 is found to perform worse than the representative agent method, is also the most commonly used criterion in the previous work on pure discrete choice models. This suggests that if some other criteria were used in these studies, maybe the statistical differentials approach would not have been found to be the worst performing method.

Related to this is the comparisons of the rankings of the methods across different aggregation error criteria. It is only the representative agent method that is consistently ranked third under all the criteria. Stat3 is ranked fourth three times, and second once as Stat1 is ranked fifth three times and fourth once. The ranking of Stat2 ranges from first to fifth depending upon the criterion used. It is ranked last when $RMSE_I$ is used whereas it is found to be the best method when $RMSE_{III}$ is used. Such changes in the rankings of the methods across the aggregation criteria suggest that it is quite important which criterion is used in selecting an aggregation method when a method is to be selected based on some empirical comparisons.

3.6 Summary and Conclusions

Five aggregations methods for mixed continuous/discrete choice models are presented. All of these methods are based on the aggregation methods for pure discrete

choice models that are discussed in Chapter 2. They also incorporate the heterogeneity of the choice sets as suggested in Chapter 2.

It is shown that the classification method always dominates the representative agent when the unconditional demand function is convex or concave. It is also shown that the statistical differentials approach is guaranteed to produce reasonable predictions only if the function of interest is convex. However, it is found that, even then, no comparison can be made involving a statistical differentials method unless it is based on empirical results.

The results from the empirical analysis have shown that no matter which aggregation method is used, the aggregation error will be quite substantial. However, the classification method is found to perform best under various criteria, suggesting that the aggregation error might be reduced by using this method. A new variation of the statistical differentials method is found to be better than the version that have previously been suggested for applications in pure discrete cases. The same method is also found to be better than the representative agent method. Finally, it is found to be important which measure of aggregation error is used in making empirical comparisons.

Table 3.1

Predictions of aggregation methods and actual aggregate demand (bushels)

Alternative	R. Agent	Class	Stat1	Stat2	Stat3	Actual
BPT*	304307	426633	15336	387164	316058	504250
TST	37688	57821	21548	15018	42277	222258
TPT	51828	80235	25740	27877	58378	279150
TRV	81929	121837	73423	87710	121919	430709
TCL	26019	42897	13284	15793	29714	177120
TGF	14553	17578	28745	0**	0**	73995
TOG	19944	23532	53543	6897	14639	87195
TMP	4509	5152	3977	0**	0**	37261
SST	22936	37353	7359	3696	22117	98786
SPT	30124	49089	8513	10396	27344	126813
SRV	43397	73036	49294	44589	86562	191023
SCL	16788	29862	5103	10061	18428	85843
SGF	6011	8792	26070	0**	8948	26741
SOG	11218	14724	47631	8509	19544	41532
SMP	1795	2208	3283	0**	1528	11947
MST	23515	32911	1459	0**	18587	79564
MPT	31540	43314	2240	3195	24214	109256
MRV	41208	58639	11530	6603	43997	140125
MCL	17100	25388	1725	1819	16109	67079
MGF	852	3919	6918	15202	8028	8710
MOG	1523	7029	14747	15892	16468	15351
MMP	313	601	698	0**	0**	4353
UST	204866	254623	830	320184	175384	547704
UPT	1153313	1439251	42562	1536489	869263	2860633
TBPT	144288	185694	19908	26264	111679	576706
TMST	19857	33832	15818	41094	23175	53730
TMPT	29961	51379	20095	48523	35136	82998
TMRV	41470	64342	55429	141889	70348	110033
TMCL	13518	25420	9970	23867	17680	45842
TMGF	519	885	1705	0**	0**	5717
TMOG	1513	2544	7900	0**	377	15683
TMMP	209	238	172	0**	0**	4254

* Alternative IJ should be read as 'by J to I', i. e., BPT means Barge-to-Portland.

I takes the values ST (Seattle), PT (Portland), RV (River), CL (California),

GF (Great Falls), OG (Ogden), and MP (Minneapolis). J takes values B=Barge, T=Truck,

S=Single car rail, M=Multiple car rail, U=Unit train, TB=Truck-Barge, and

TM=Truck-Multiple car.

** 0 was assigned when the method produced negative predictions.

Table 3.2
Percentage deviations of the predictions from actual demand

Alternative	R. Agent	Class	Stat1	Stat2	Stat3
BPT	10	15	97	23	37
TST	83	74	90	93	81
TPT	81	71	91	90	80
TRV	81	72	83	80	72
TCL	85	76	92	91	83
TGF	80	76	61	100	100
TOG	77	73	39	92	83
TMP	88	86	89	100	100
SST	77	62	93	96	78
SPT	76	61	93	92	78
SRV	77	62	74	77	55
SCL	80	65	94	88	79
SGF	78	67	3	100	67
SOG	73	65	15	80	53
SMP	85	82	73	100	87
MST	70	59	98	100	77
MPT	70	58	98	97	77
MRV	71	58	92	95	69
MCL	75	62	97	97	76
MGF	90	55	21	75	8
NOG	90	54	4	4	7
MMP	93	86	84	100	100
UST	63	54	100	42	68
UPT	60	50	99	46	70
TBPT	75	68	97	95	81
TMST	63	37	71	21	57
TMPT	64	38	76	42	58
TMRV	62	42	50	29	36
TMCL	71	45	78	48	61
TMGF	91	85	70	100	100
TMOG	90	84	50	100	98
TMMP	95	94	96	100	100
AVP*	77	64	74	78	71

* AVP=average percentage deviation.

Table 3.3

Total aggregation errors by different criteria
(Ranking of the method under the criterion)

Aggregation Method	$RMSE_I$	$RMSE_{II}$	$RMSE_{III}$	$RMSE_{IV}$
Rep. Agent	4.39 (3)	1,796.21 (3)	1,890,196 (3)	4,718,755 (3)
Class	3.72 (1)	1,513.10 (1)	1,581,115 (2)	3,896,607 (1)
Stat1	4.48 (4)	2,479.62 (5)	3,024,749 (5)	6,533,011 (5)
Stat2	4.69 (5)	1,779.71 (2)	1,574,312 (1)	4,396,413 (2)
Stat3	4.23 (2)	1,876.06 (4)	2,150,857 (4)	4,921,702 (4)

Chapter 4

Aggregation and Welfare Analysis with a Mixed Continuous/Discrete Choice Model

4.1 Introduction

Disaggregate choice models have been widely used in many areas of economics for the past years. As the use of discrete and/or mixed continuous/discrete models gains popularity, it becomes important how to generalize the methods of applied welfare economics to these models. Although there exists an extensive literature that deals with the theory and measurement of welfare change, there is no general theory available for measuring welfare changes when agents face discreteness in their choices.

A few authors, notably Small and Rosen (1981), Hanemann (1982, 1984b), and Hui (1983, 1985, 1986, 1987), have rigorously derived Hicksian measures of welfare change for consumers faced with discrete choices. However, these studies do not address the aggregation problem. These authors are either interested in only measuring the welfare change of an individual or they assume away the aggregation problem by implicitly applying the representative agent method. As in the prediction of demand, one is typically interested in measuring the total welfare change in a population rather than estimating the welfare change of a single agent. The purpose of this chapter is to provide a methodology for measuring the aggregate welfare change when discrete and/or mixed continuous/discrete choice models are used in modelling the agent's behavior.

In order to focus on the comparisons of different methods of aggregating welfare change, a model in which the decision-making unit is a profit maximizing firm is used. The advantage of using such a model is the simplicity of the measure of welfare change for an individual. Using the change in the profits of a firm as the measure of welfare change we derive a closed-form expression for the welfare change without the stringent assumptions that are necessary when dealing with consumer welfare. This enables us to focus on the effects of applying different aggregation methods on the evaluation of the aggregate welfare change.

The organization of this chapter is as follows. The model for which a measure of welfare change is developed is described in the following section of this chapter. The third section deals with developing a measure of welfare change for a single firm. The fourth section defines the aggregation problem in measuring the aggregate welfare change and suggests five different aggregation methods to deal with this problem. A summary and conclusions section follows the fifth section where some empirical results are presented.

4.2 A Mixed Continuous/Discrete Transportation Demand Model

A multinomial mixed continuous/discrete choice model of freight transportation demand that was earlier developed by Inaba and Wallace (1987) forms the basis for the analysis in this chapter. The model is based on the individual shipment decisions of wheat elevators with respect to mode-market choice and shipment size. It is assumed that the firms (wheat elevators) are profit maximizers who take prices and transportation rates as given. As a result of a contract between the firm and a buyer at the destination,

the firm promises to deliver a given quantity, by a given date, for a given price, to a given market. Thus, a shipment is specified by the quantity of the wheat to be delivered as well as the transportation mode to be used and the market destination. Hence the firm's decision variables are the shipment size, a continuous variable, and the mode-destination alternative, a discrete variable. The model, as formulated by Inaba and Wallace, incorporates the endogeneity of these two choices.

To begin summarizing the mathematical formulation of the model, let $\mathbf{x}_{ti}$ denote the vector of exogeneous variables that enter the firm t 's decision problem, and let q_{ti} denote the quantity of wheat that firm t will ship if i is the chosen mode-market alternative. If $\bar{\pi}_{ti}(q_{ti}, \mathbf{x}_{ti})$ denotes the firm t 's profit when i is the chosen alternative and q_{ti} is the profit maximizing quantity of wheat, then the firm will choose alternative i if and only if the i th alternative provides the highest profit, i. e.,

$$\bar{\pi}_{ti}(q_{ti}, \mathbf{x}_{ti}) > \bar{\pi}_{tj}(q_{tj}, \mathbf{x}_{tj}) \quad \forall j \neq i.$$

Thus, the firm t observes the exogeneous variables, determines the profit maximizing shipment size on each feasible alternative, and then chooses the alternative that gives the highest profit.

It is assumed that the firm determines the profit maximizing shipment size for a feasible alternative in accordance with the equation

$$q_{ti} = \beta^T \mathbf{x}_{ti} + u_{ti} \quad (4.2.1)$$

where β is a vector of parameters and u_{ti} is an error term which is identically and independently distributed as normal with zero mean and variance σ_u^2 . Since this equation

gives the quantity shipped given that alternative i is chosen, it is called the conditional quantity shipment equation. The exogenous variables that enter this specification are the market boundary and the capacity of storage of the firm.

Similarly, the firm's maximum profit, given alternative i is chosen, is also assumed to be determined by a linear function of the exogenous variables. Thus, firm t 's conditional profit function is

$$\bar{\pi}_{ti} = \gamma Y_t + \alpha^T z_{ti} + v_{ti} \quad (4.2.2)$$

where v_{ti} is assumed to have an identical and independent Gumbel distribution with parameters $(0, 1)$. The vector of exogenous variables, z_{ti} , consists of the market boundary of the firm, four mode-destination dummy variables, and some characteristics of the alternatives that represent various inventory costs that are incurred when wheat is transported. The set of the cost attributes of the alternatives consists of the following characteristics: waiting time cost,¹ loading cost,² and time-in-transit cost.³ Y_t is the storage capacity of the firm t , which does not vary across the alternatives.

It is important to note that the two expressions above are generated from a common optimization problem so that the error term in equation (4.2.2) is in general correlated with the error term in equation (4.2.1). The firm computes the conditional shipment size, q_{ti} , for each feasible alternative i . However, q_{ti} is observed only when $\pi_{ti} > \pi_{tj}$ for all $j \neq i$. Thus, the model at hand is a mixed simultaneous demand system

1. Per bushel inventory holding cost.
2. Per bushel inventory holding cost.
3. Per bushel inventory holding cost of the time in transit of the mode from origin to destination.

where a continuous variable is observed based on the outcome of an unobserved latent variable.

The model has been estimated by Inaba and Wallace by making use of a procedure which is based on Lee's (1982) two-stage method that provides consistent parameter estimates for such selectivity models. In summary, the first stage of this procedure provides consistent estimates of the elements of the parameter vector α in equation (4.2.2), which are later used in the second stage to calculate the choice probabilities to be used in constructing a selectivity variable.⁴ This selectivity variable is then appended to equation (4.2.1) to correct for the selectivity problem before estimating the equation by weighted least squares in the second stage of the procedure.

4.3 Welfare Analysis

This section deals with measuring the welfare change of the firms caused by a change in the exogenous variables in the model presented in the previous section. The strategy for this analysis is to develop a measure of welfare change for a single firm and then focus on ways of measuring the aggregate welfare change of all the firms in the population.

Since the unit of the analysis is a profit maximizing firm, the change in the profit of the firm represents the change in the welfare of the firm. However, measuring the

4. Note that γ , the coefficient for the storage capacity of the firm, can not be estimated based on the formulation of the model. Since this variable does not vary across the alternatives, it does not enter the optimal discrete choice problem, and thus its coefficient is not identifiable.

change in the profit is complicated by the fact that the profit of the firm is observed for only the alternative that is chosen, and, a change in any of the exogenous variables can cause the firm to choose another alternative that it has not chosen before.⁵ Thus, it is not possible to estimate the profit of the firm after a change by simply evaluating equation (4.2.2) with the new values of the independent variables to make comparisons between the before-change profit and after-change profit of the firm.⁶

The problem arises because the observed profit of a firm is conditional on the choice of an alternative. Hence, the correct approach is to derive an expression for the unobserved unconditional profit function. Since the unconditional profit function does not depend on the alternative chosen, it allows us to calculate the change of the profit of the firm no matter what the firm decides to do when faced with a change in the exogenous variables.

Recall that the conditional profit function of firm t is specified as

$$\bar{\pi}_{ti} = \gamma Y_t + \alpha^T z_{ti} + v_{ti}$$

and it represents the profit of firm t given that alternative i is chosen. Since the error term, v_{ti} , is assumed to have an identical and independent Gumbel distribution with parameters $(0, 1)$, the conditional profit, $\bar{\pi}_{ti}$, is also a random variable whose distribution is Gumbel with parameters $(\gamma Y_t + \alpha^T z_{ti}, 1)$. Note that although $\bar{\pi}_{ti}$ is a random variable for the researcher, it is not a random variable for the firm. The firm is perfectly capable of evaluating $\bar{\pi}_{ti}$ for each alternative i and then choose the one that gives the highest

5. Except of a change in Y_t .

6. Even when the estimate of γ is available.

profit. It is the researcher who can not calculate the profit of a firm for an alternative unless that alternative is observed to be chosen by the firm. The best the researcher can do is to use the information (assumption) on the distribution of the conditional profit of the firm.

The unconditional profit function for firm i is defined as

$$\pi_i = \max[\bar{\pi}_{i1}, \dots, \bar{\pi}_{iN}] \quad (4.3.1)$$

where N is the total number of feasible alternatives for firm i . It is a random variable whose value corresponds to the profit from the alternative that the firm will choose. Since it is the maximum of random variables that are Gumbel, it has a Gumbel distribution with parameters $(\ln \sum_{i=1}^N e^{\gamma Y_i + \alpha^T z_{ii}}, 1)$. The expected value of the unconditional profit function can be expressed as

$$E[\pi_i] = \ln e^{\gamma Y_i} \sum_{i=1}^N e^{\alpha^T z_{ii}} + 0.577 \quad (4.3.2)$$

where 0.577 is the Euler constant.

The equivalent variation, the welfare change measure to be used, associated with a change in the values of the exogenous variables is defined as the amount of money that should be given to the firm before the change occurs in order to leave the firm just as well off as if the change had occurred. Thus, if π_i^F and π_i^0 denote the level of the profit of the firm after and before the change, respectively, then the equivalent variation, ev_i , is defined by

$$\pi_i^F = \pi_i^0 + ev_i$$

which results in

$$ev_t = \pi_t^F - \pi_t^0. \quad (4.3.3)$$

Note that ev will be positive (negative) when the firm is better off (worse off) after the change.

Although we have derived the unconditional profit function and the measure of welfare change, the exact amount of equivalent variation still can not be found. The reason is that, this derived welfare measure is a random variable, being the difference between two random variables.⁷ Then, a natural way of evaluating the welfare change in this framework is to focus on the mean of ev .⁸

The expected value of equivalent variation can be written as

$$E[ev_t] = E[\pi_t^F - \pi_t^0] \quad (4.3.4)$$

which simply reduces to

$$E[ev_t] = \ln \frac{e^{\gamma Y_t^F} \sum_{i=1}^N e^{\alpha^T z_{ti}^F}}{e^{\gamma Y_t^0} \sum_{i=1}^N e^{\alpha^T z_{ti}^0}} \quad (4.3.5)$$

where Y_t^F , z_{ti}^F and Y_t^0 , z_{ti}^0 denote the final and initial values of the exogenous variables, respectively.⁹

7. Note that ev has a logistic distribution.

8. It should be pointed out that there might be some problems in applying this approach to the consumer case as pointed out in Hanemann (1982).

9. It is important to note that equation (4.3.5), when divided by the marginal utility of income, is exactly what Small and Rosen (1981) have derived for a logit model that is based on utility maximizing consumers. Thus, the aggregation methods that will be suggested later on are also applicable to their measure.

Equation (4.3.5) is the expression to be used to predict the welfare change of a firm due to a change in the exogenous variables. However, it should be clear that in order to be able to make this calculation, one has to know the values that the exogenous variables take before and after the change.¹⁰ To make this dependency clear, equation (4.3.5) can be rewritten as

$$E[ev_i | Y_i^F, z_i^F, Y_i^0, z_i^0] = \ln \frac{e^{Y_i^F} \sum_{i=1}^N e^{\alpha^T z_{ii}^F}}{e^{Y_i^0} \sum_{i=1}^N e^{\alpha^T z_{ii}^0}} \quad (4.3.6)$$

where $z_i^F = (z_{i1}^F, \dots, z_{iN}^F)$ and $z_i^0 = (z_{i1}^0, \dots, z_{iN}^0)$. This is the measure of welfare change that we suggest. Given that we know what values the exogenous variables take, it gives the expected value of the amount of money that should be given to the firm before the change so that the firm will be as well off as if the change had occurred.

4.4 Aggregation and Welfare Analysis

4.4.1 The Aggregation Problem

Although a measure for evaluating the welfare change of an individual firm has been developed in the previous section, this measure is not sufficient to carry out a policy analysis. In performing a policy analysis, one is typically interested in measuring the total welfare change of all the agents in the population of concern. A common way of defining a measure of aggregate welfare change is based on the assumption that each

10. Under the assumption that all the coefficients have already been estimated.

agent in the population carries the same weight from the perspective of the policy maker. If this approach is applied to the model we have, the aggregate welfare change measure, EV , is obtained as

$$EV = \sum_{t=1}^T ev_t \quad (4.4.1.1)$$

where T denotes the total number of firms in the population.

Since each term in this summation is a random variable, the aggregate welfare change measure is also a random variable. Thus, as we did before for a single firm, we again suggest the expected value of this random variable as the measure of the aggregate welfare change. Thus, by taking the conditional expectation of both sides of equation (4.4.1.1) and using equation (4.3.6) we obtain

$$E[EV|Y_t^F, z_t^F, Y_t^0, z_t^0; t = 1, \dots, T] = \sum_{t=1}^T \ln \frac{e^{\gamma Y_t^F} \sum_{i=1}^N e^{\alpha^T z_{ti}^F}}{e^{\gamma Y_t^0} \sum_{i=1}^N e^{\alpha^T z_{ti}^0}} \quad (4.4.1.2)$$

to represent the aggregate welfare change. Evaluation of this expression gives the expected value of the sum of the amount of money that should be given to each firm in the population before the change so that each firm will be just as well off as if the change had occurred, given the values of right-hand side variables.

Equation (4.4.1.2) suggests that the aggregate welfare change is obtained by calculating the welfare change of each firm in the population and then summing them up. Although this procedure seems to be very simple, it rarely is a feasible approach in practice. First of all, calculating the welfare change of each firm in the population requires information on the values of the exogenous variables that enter the decision

problem of each firm in the population. Given that what is available in almost every case is just a sample from a larger population, this information is rarely available for the researcher. Furthermore, the implementation of this computation can be too complicated for large data sets even if this data requirement is met.

The question is then how to calculate the aggregate welfare change without the extreme data requirements that equation (4.4.1.2) imposes. Finding such a method is the aggregation problem in this context.

4.4.2 Aggregation Methods

The aggregation problem defined above essentially deals with the question of how to make equation (4.4.1.2) operational for implementing numerical computations. This equation consists of two sets of variables. The first set contains the parameters which are consistent estimates of true population parameters. Due to the underlying assumptions of the model at hand, they do not vary across the firms. Although they are random variables in their nature, for the sole purpose of evaluating equation (4.4.1.2) they can be taken as given. In other words, they can be treated as the set true values of the population parameters based on their consistency property.¹¹

The second set is the set of the exogeneous variables that enter the specification (4.2.2). These variables, which will be called attributes from hereon, are taken as given

11. The justification for this is that the prediction of aggregate welfare change as defined by equation (4.4.1.2) will be a consistent estimate of the true value of the aggregate welfare change, given the values of z vector and Y_t .

by the firms when they are solving their decision problem. They also are exogeneous for the researcher in estimating the model since estimation requires information on the values of these attributes for only the firms in the sample. But when it comes to evaluating equation (4.4.1.2), all the firms in the population are of concern. Given that it is not known what values these attributes take for all the firms in the population, these attributes are truly random variables for the researcher in calculating the aggregate welfare change.

Thus, the aggregate welfare change measure defined by equation (4.4.1.2) needs to be treated as a random variable where the randomness is due to the randomness in the attributes. If the expected value of this random variable is to be used as the best predictor, then the aggregation problem is simply to find that expectation. In other words, the aggregation problem reduces to finding the unconditional expectation of EV in equation (4.4.1.2). This is the approach that we take in developing aggregation methods for predicting the aggregate welfare change with the model we have.

Before proceeding further, recall that z_t^F and z_t^0 are the vectors of the values that the attributes of all the alternatives for firm t take after and before a change. Thus, if there are k attributes related to an alternative, then each of these vectors consists of $k \times N$ elements. Note that the value that each of these variables takes after a change can be expressed as a linear function of its initial value, i. e.,

$$z_{tik}^F = a_i^k \cdot z_{tik}^0 + b_i^k \quad i = 1, \dots, N \quad t = 1, \dots, T$$

where z_{iik}^0 is the initial value of the k th attribute of the i th alternative of firm l .¹²

Therefore, equation (4.4.1.2) can be rewritten as

$$E[EV | z_t^0, Y_t^0] = \sum_{i=1}^T \ln h(z_t^0, Y_t^0) \quad (4.4.2.1)$$

where $h(z_t^0, Y_t^0)$ is defined as¹³

$$h(z_t^0, Y_t^0) \equiv \frac{e^{\gamma(cY_t^0 + d)} \sum_{i=1}^N e^{a_i^T (z_t^0 + b_i)}}{e^{\gamma Y_t^0} \sum_{i=1}^N e^{a_i^T z_{ii}^0}} \quad (4.4.2.2)$$

Taking the expectation of both sides of equation (4.4.2.1) with respect to the random vector z_t^0 and Y_t^0 results in the unconditional expectation of EV :

$$E[EV] = E_{z_t^0, Y_t^0} \left[\sum_{i=1}^T \ln h(z_t^0, Y_t^0) \right]$$

which can be written as

$$E[EV] = \sum_{i=1}^T \int_{z_t^0, Y_t^0} \ln h(z_t^0, Y_t^0) f_i(z_t^0, Y_t^0) d(z_t^0, Y_t^0) \quad (4.4.2.3)$$

where $f_i(z_t^0, Y_t^0)$ is the joint density function of the attributes of firm l 's alternatives.

Although equation (4.4.2.3) provides us with the unconditional expectation of the aggregate welfare change, it still is not operational the way it is since it requires information on the joint density functions of each firm. If these density functions are known, then the aggregate welfare change can readily be calculated. However, this hardly is the case. Thus we need to make some assumptions on the density functions

12. So, $z_t^0 = (z_{t11}^0, \dots, z_{tN}^0) = (z_{t11}^0, \dots, z_{t1k}^0, z_{t21}^0, \dots, z_{t2k}^0, \dots, z_{tN1}^0, \dots, z_{tNk}^0)$ is a vector of $k \times N$ elements. Also, similarly, $Y_t^0 = cY_t^0 + d$.

13. Here, $a_i = (a_i^1, \dots, a_i^k)$ and $b_i = (b_i^1, \dots, b_i^k)$; $i = 1, \dots, N$, and $*$ denotes element multiplication.

in order to evaluate the aggregate welfare change. Consequently, different aggregation methods are obtained based on different representations of the joint density functions.

Now, let ω_t be the collection of all the attributes of individual t , i. e., $\omega_t \equiv (z_t^0, Y_t^0)$. Then equation (4.4.2.3) can be rewritten as

$$E[EV] = \sum_{t=1}^T \int_{\omega_t^0} \ln h(\omega_t^0) f_t(\omega_t^0) d\omega_t^0 \quad (4.4.2.3)'$$

which is identical to the right-hand side of equation (3.3.7) in Chapter 3, when the function $G^i(\cdot)$ in equation (3.3.7) is defined as the function $\ln h(\omega_t^0)$ where $h(\cdot)$ is defined by equation (4.4.2.2). What this implies is that all the aggregation methods defined in Chapter 3 apply to the case of welfare analysis as well. Measuring aggregate welfare change is a special case of the aggregation problem defined in Chapter 3, where the function of interest takes a specific form. Thus, the aggregate welfare change with respect to a change in any of the attributes can be predicted by using any of the aggregation methods defined and discussed in the previous chapter.

4.5 An Empirical Example

As seen in Chapter 3, there might be substantial differences between the performance of different aggregation methods. Given that welfare analyses are typically used to determine whether some specific policies should be undertaken or not, the bias of the predictions of the aggregate welfare change with respect to the changes the policies suggest can be very important. The main purpose of this section is to demonstrate this

by providing predictions of aggregate welfare change under a specific policy by using the model described in section 4.2.

Although the previous sections of this chapter have focused on measuring welfare change with respect to the changes in exogenous variables, another important case is the elimination of some alternatives from the choice sets of the individuals. However, the welfare change which might result from such a policy can not be calculated based on the measure of welfare change that has been developed. Furthermore, there is no previous work which considers such cases in measuring welfare changes with mixed continuous/discrete or pure discrete choice models. Therefore, a policy which suggests abandonment of a transportation mode that the firms have as feasible is analyzed in this section.

Suppose that alternative j is being considered to be abandoned. The expected value of firm i 's unconditional profit after this change can be written as

$$E[\pi_i] = \ln e^{\gamma_i} \sum_{i \neq j}^N e^{\alpha^T z_{ii}} + 0.577.$$

Then, the expected value of equivalent variation is simply

$$E[ev_i] = \ln \frac{e^{\gamma_i} \sum_{i \neq j}^N e^{\alpha^T z_{ii}}}{e^{\gamma_i} \sum_{i=1}^N e^{\alpha^T z_{ii}}} \quad (4.5.1)$$

which corresponds to equation (4.3.5).

There are three important points here. First, the welfare change for any firm will be negative since the numerator is always less than the denominator in equation

(4.5.1). Secondly, if the eliminated alternative is not feasible for firm l initially, then the welfare change for this firm is 0.¹⁴ Thirdly, note that the term e^{γ_k} cancels and thus, allows us to calculate the welfare change even though γ is not estimated in this model. Therefore, the unconditional expectation of aggregate welfare change for the population can be written as

$$E[EV] = \sum_{i=1}^{T_j} \int_{z_i} \ln \frac{\sum_{i \neq j}^N e^{\alpha^T z_{ii}}}{N} f_l(z_i) dz_i \quad (4.5.2)$$

where T_j denotes the total number of firms in the population for which alternative j is feasible.

Since equation (4.5.2) is essentially same as equation (3.3.7) in Chapter 3 on which the aggregation methods are based, all the aggregation methods defined in the previous chapter apply to equation (4.5.2). Thus, aggregate welfare change caused by this policy change can be predicted by directly applying the aggregation methods defined earlier. For example, if the representative agent method is applied, then the aggregate welfare change is obtained as

$$EV^R = T_j \frac{\sum_{i \neq j}^N e^{\alpha^T \bar{z}_i}}{N} \quad (4.5.3)$$

14. Because if an alternative is not feasible, then it does not enter the summation in the denominator either. (Recall that N has been defined as the total number of feasible alternatives.)

where $\bar{z}_i$ is the mean vector of the attributes of the i th alternative. The classification method results in

$$EV^C = \sum_{g=1}^G T_{gj} \frac{\sum_{i \neq j}^N e^{\alpha^T \bar{z}_i^g}}{\sum_{i=1}^N e^{\alpha^T \bar{z}_i^g}} \quad (4.5.4)$$

where T_{gj} is the number of firms in group g , $g = 1, \dots, G$, for whom alternative j is feasible, and $\bar{z}_i^g$ the mean vector of the attributes of i th alternative in group g .

The three variants of the statistical differentials approach result in the following expressions:

$$EV^{SDI} = T_j \left[\frac{\sum_{i \neq j}^N e^{\alpha^T \bar{z}_i}}{\sum_{i=1}^N e^{\alpha^T \bar{z}_i}} + \frac{1}{2} \sum_{r=1}^{k \times N} \sum_{s=1}^{k \times N} \frac{\partial^2}{\partial z_r \partial z_s} \frac{\sum_{i \neq j}^N e^{\alpha^T z_{ti}}}{\sum_{i=1}^N e^{\alpha^T z_{ti}}} \Big|_{\bar{z}} \text{Cov}(z_r, z_s) \right] \quad (4.5.6)$$

$$EV^{SDII} = \sum_{g=1}^G T_{gj} \left[\frac{\sum_{i \neq j}^N e^{\alpha^T \bar{z}_i^g}}{\sum_{i=1}^N e^{\alpha^T \bar{z}_i^g}} + \frac{1}{2} \sum_{r=1}^{k \times N} \sum_{s=1}^{k \times N} \frac{\partial^2}{\partial z_r \partial z_s} \frac{\sum_{i \neq j}^N e^{\alpha^T z_{ti}}}{\sum_{i=1}^N e^{\alpha^T z_{ti}}} \Big|_{\bar{z}_g} \text{Cov}(z_r^g, z_s^g) \right] \quad (4.5.7)$$

$$EV^{SDIII} = \sum_{g=1}^G T_{gj} \left[\frac{\sum_{i \neq j}^N e^{\alpha^T \bar{z}_i^g}}{\sum_{i=1}^N e^{\alpha^T \bar{z}_i^g}} + \frac{1}{2} \sum_{r=1}^{k \times N} \sum_{s=1}^{k \times N} \frac{\partial^2}{\partial z_r \partial z_s} \frac{\sum_{i \neq j}^N e^{\alpha^T z_{ti}}}{\sum_{i=1}^N e^{\alpha^T z_{ti}}} \Big|_{\bar{z}_g} \text{Cov}(z_r, z_s) \right] \quad (4.5.8)$$

where $\bar{z} = (\bar{z}_1, \dots, \bar{z}_N)$, $\bar{z}_g = (\bar{z}_{g1}, \dots, \bar{z}_{gN})$, and z_r is the r th element of z vector.

Based on these last five equations, aggregate welfare change due to the elimination of one of the transportation modes was estimated for the freight transportation demand model described in Section 4.2. The eliminated mode was unit train, which resulted in eliminating two mode-destination alternatives from the choice sets of the firms.¹⁵ This reduced the maximum total number of feasible alternatives to 30 from 32. There are

15. See Table 3.1 for a complete list of mode-destination alternatives.

183 firms in the population of interest. Of these firms, only 43 had these 2 alternatives feasible initially.

The predictions of aggregate welfare change resulting from each aggregation method are presented in Table 4.1 together with the actual aggregate welfare change. The actual aggregate welfare change is calculated by evaluating equation (4.5.1) over all the 42 firms that initially had the two eliminated alternatives feasible.

Table 4.1

Aggregate welfare change by different aggregation methods

Aggregation Method	Predicted EV (\$)	Actual EV (\$)	% Deviation	Rank
Rep. Agent	-11.88	-48.31	+75.41	5
Class	-16.60	-48.31	+65.56	3
Stat1	-12.09	-48.31	+74.97	4
Stat2	-52.78	-48.31	- 9.25	1
Stat3	-39.57	-48.31	+18.09	2

As all the predictions and the actual welfare change indicate, elimination of unit train results in a welfare loss. However, the magnitude of the change seems to be quite low in contrary to what one might expect. Even the largest loss, which is predicted by Stat2 method, is less than 53 dollars. Given that this represents the total welfare change for 43 firms, these predictions cannot be said to reflect the true welfare change. However, since the purpose of this analysis is only to compare the predictions from different aggregation methods based on an already estimated mixed continuous/discrete choice model, no effort has been made to further investigate this issue.

In terms of the comparisons between the aggregation methods, the statistical differentials method 2 is found to be the best one. It provides the closest prediction to the actual value, which is off by only 9.25 percent from the actual value. The second best method is also one of the statistical differentials methods, namely, statistical differentials II. Given that statistical differentials III method is one of the worst methods, it seems that applying statistical differentials approach to different sub-groups first results in a substantial improvement, making statistical differentials approach a better method.¹⁶ The classification method, which was found to perform quite well in the previous chapter, is ranked third here. The representative agent method is found to be the worst method, producing a bias by 75.41 percent.

It is important to note that the bias of the predictions range from 9.25 % to 75.41 % of the actual value. This indicates that which method is used in calculating aggregate welfare change can be crucial. Although the results obtained here cannot be generalized, it is irritating that the aggregation error is as much as 75.41 % and 65.56 % for the representative agent and classification methods, respectively, given that these two methods are the ones that are commonly used.¹⁷ Also, note that the statistical differentials II method is found to over-predict the welfare loss whereas the other methods are found to under-predict the welfare loss. This suggests that the same policy might be accepted or rejected depending upon which aggregation method is used.

16. Recall that the same result was obtained based on the comparisons of aggregate demand predictions in Chapter 3.

17. For example, Morrison and Winston (1986) have found a welfare gain of \$5.7 billion in their study of the economic effects of airline deregulation, where they apply the classification method to exactly the same welfare measure defined by (4.3.5). As they point out, this corresponds to a welfare gain by an amount of 35 percent of the actual 1977 airline revenues. The possibility that this figure might be wrong by a substantial amount is, of course, quite bothersome.

4.6 Summary and Conclusions

A methodology for evaluating welfare changes in a mixed continuous/discrete choice framework is presented. First, a welfare change measure for a single agent is developed. Then, based on this measure, five different ways of predicting aggregate welfare change are defined by making use of the aggregation approaches presented in Chapter 3.

These methods are compared to each other based on empirical results from a freight transportation demand model. It is found that the aggregation errors can be quite substantial. The results also confirm that the two statistical differentials methods suggested in Chapter 3 improve the statistical differentials approach. These two methods are also found to be superior to the others for this specific case where two alternatives are eliminated from the choice sets of the agents.

Chapter 5

Summary and Conclusions

This dissertation has investigated the aggregation problem in making predictions based on mixed continuous/discrete choice models. The analysis started with formalizing the assumptions underlying the aggregation methods that have already been developed for pure discrete choice models. By combining similarities between the assumptions for classification and statistical differentials methods, two variations of the statistical differentials method were introduced. Then, the issue of incorporating heterogeneity of choice sets was taken up. It was shown that when the differences between the choice sets of individuals were not taken into account, the methods produced over-predictions. Thus, the aggregation methods were modified to incorporate these differences.

The incorporation of such differences only required to use the fact that the demand for an alternative was zero if the alternative was not feasible for an individual. However, the modification of the methods involved one more restriction which was based on the fact that the individuals were required to choose one and only one alternative in a pure discrete choice framework. This, of course, was purely based on the previous applications of pure discrete choice models. In general, one does not have to impose this restriction. In this case, the suggested methods can easily be readjusted to only incorporate the differences in the choice sets.

The empirical results presented in Chapter 2, which were obtained from a multinomial logit model of transportation demand, confirmed that the adjusted methods

provided better predictions than the unadjusted methods. Since the adjustments did not require any additional information that one would not already have, the adjusted methods were concluded to be preferable to the unadjusted ones.

Chapter 3 took up the issue of aggregation with mixed continuous/ discrete models. It was shown that aggregation in mixed continuous/ discrete models was essentially the same problem as in the pure discrete choice models. Both cases were found to involve aggregating some function of some estimated coefficients and random variables. The only difference between the two cases was that the function of interest was giving the probability of choosing an alternative in pure discrete case, whereas it was generating the quantity demand for the alternative in mixed continuous/discrete case. This, of course, was based on the difference in what demand meant in these two types of models: demand for an alternative in pure discrete choice case was the expected number of times the alternative was chosen whereas it was the amount of which the alternative was chosen to be consumed in the mixed continuous/discrete case.

This implied that although the function of interest was, in general, more complicated in mixed continuous/discrete framework, the same approaches were still applicable. Thus, aggregation methods for mixed continuous/discrete choice models were developed based on the same assumptions that were made in developing aggregation methods for pure discrete models. These methods, of course, incorporated heterogeneity of choice sets as well. However, they did not involve the restriction on the overall demand that was employed for the pure discrete case, since such a restriction did not exist in mixed continuous/discrete choice models.

Chapter 4 investigated the issue of aggregation in calculating the aggregate welfare change for a population of agents whose behavior was modeled in a mixed continuous/discrete choice framework. Because of the lack of a general theory for measuring welfare change in this framework, the first task was to develop a measure of welfare change for measuring the welfare change of a single agent. In order to avoid some stringent assumptions that one would need in consumers' case, a model in which the decision unit was a profit maximizing firm was used. This resulted in a closed form expression for the welfare change measure which allowed us to focus on the aggregation issue. This closed form expression was also a function of some estimated coefficients and random variables, to which the aggregation methods developed in Chapter 3 were directly applicable. The empirical analysis presented in this chapter not only applied these methods to measuring aggregate welfare change, but also showed how welfare change could be calculated when some of the alternatives were eliminated from the choice sets of the agents.

When put together, the analyses in chapters 2 through 4 imply that the aggregation problem in a mixed continuous/discrete framework is essentially the same problem to which the same approaches can be applied, regardless of what one is interested in calculating. The point of interest can be calculating the aggregate probability of choosing an alternative, calculating aggregate quantity demand, or calculating aggregate welfare change. The only difference between these three cases is the function over which aggregation is done. Other than having a different functional form, the function of interest is

always a function of some estimated coefficients and some random variables. Thus, the same guideline should be followed in selecting an aggregation methods in all the cases.

Chapter 3 included a section on theoretical comparisons of all the five aggregation methods analyzed in the dissertation. Although this analysis showed that the classification method should be preferred over the representative agent method when the function of interest was convex or concave, this was the only conclusive result obtained from the analysis. Any other comparison of the methods was found to depend on empirical properties of the function to be aggregated. Thus, selection of the best aggregation method remained to be an empirical issue.

Given that the aggregation methods are the same in all of these three cases, the empirical analyses presented in the dissertation can be seen as a single analysis in which different functions are aggregated by using five different aggregation methods. These empirical analyses were also based on the same empirical model of freight transportation demand. The first empirical analysis involved the choice function which gave the probability of choosing an alternative. The second one aggregated the function which generated the quantity demand for an alternative. The last one aggregated the function which gave the welfare change due to eliminating two alternatives from the choice sets of the firms. Since there is no other difference between these functions other than this conceptual one and their functional forms, these results can be seen as different simulations which demonstrate the performance of the same aggregation methods in various cases.

The results from the first two simulations indicated that the classification method performed better than the other methods. However, this was not the case in the third simulation which involved the welfare change case. In this third simulation, it was the statistical differentials II method that performed the best. Although the same method was found to be second best in the simulation that involved quantity demand estimation, it was ranked fifth in the first simulation. Similarly, none of methods was ranked the same in all the simulations. Thus, no method was found to be superior to the others based on the empirical evidence obtained from these simulations. However, the results indicated that the classification and the statistical differentials II methods seemed to perform better than the others, in general. It is not clear, though, if this will remain to be case in other simulations based on different models and data.

Although no method can be suggested for use in general, there are some steps that can be taken to select an aggregation method in future applications of mixed continuous/discrete choice models. By testing for convexity of the function to be aggregated, and analyzing the magnitude of the second-order terms in the statistical differentials methods, it can be possible to select one of the methods. This, of course, will depend on the results of the convexity test. If the test gives negative results, implying neither convexity nor concavity, such a step will not be useful.

One consistent result over all the simulations was that the two variations of the statistical differentials method introduced in this dissertation performed better than the original version of the method. Thus, it can be suggested to apply some degree of classification in using statistical differentials approach.

An important result involved selection of the aggregation error criterion. The ranking of the methods was not consistent only across the simulations but also across the criteria that were used. This suggests that it is important which criterion is being used in measuring the aggregation error. An important contribution of future research on this subject will be developing a single, generally accepted criterion.

A major result from all the simulations was that the aggregation errors were quite substantial for all the methods in almost all the cases. Thus, if the aggregation issue is ignored in empirical applications of mixed continuous/discrete choice models and/or pure discrete choice models, the predictions obtained will be off by a substantial amount. However, overcoming this problem is not easy because even the methods investigated here can produce large errors. Therefore, developing a completely different new aggregation method which will produce less aggregation errors still remains to be a major area for future research.

References

- Ben-Akiva, Moshe E. and Lerman, Steven R. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, (Cambridge, MA: MIT Press).
- Bouthelier, F. and Daganzo, C. F. (1979) "Aggregation with Multinomial Probit and Estimation of Disaggregate Models with Aggregate Data: A New Methodological Approach," *Transportation Research*, 13B, 133-146.
- Domencich, T. and McFadden, D. (1975) *Urban Travel Demand - A Behavioral Analysis*, (Amsterdam: North Holland).
- Duncan, G. M. (1980), "Formulation and Statistical Analysis of the Mixed, Continuous/Discrete Dependent Variable Model in Classical Production Theory," *Econometrica*, 48, 839-852.
- and Leigh, D. E. (1980), "Wage Determination in the Union and Nonunion Sectors: A Sample Selectivity Approach," *Industrial and Labor Relations Review*, 19, 24-43.
- Dunne, J. P. (1985), "Disaggregate Choice Models and the Aggregation Issue: Some Empirical Results," *Transportation Research A*, 19, 315-324.
- Goodman, C. A. (1988), "An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing Demand," *Journal of Urban Economics*, 23, 327-353.
- Hanemann, W. M. (1982), "Applied Welfare Analysis with Qualitative Response Models," Dep. Agr. and Resour. Econ. Work. Pap. No 2-11, University of California, Berkeley.
- (1984a), "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541-561.
- (1984b), "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses," *Amer. J. Agr. Econ.*, 66, 332-341.

Hau, T. D. (1983), "Some Hicksian and Marshallian Consumer's Surplus Estimates in Discrete Choice," *Economics Letters*, 11, 203-210.

——— (1985), "A Hicksian Approach to Cost-Benefit Analysis with Discrete-choice Models," *Economica*, 52, 479-490.

——— (1986), "Distributional Cost-Benefit Analysis in Discrete Choice," *J. Transport Economics and Policy*, 20, 313-338.

——— (1987), "Using a Hicksian Approach to Cost-Benefit Analysis in Discrete Choice: An Empirical Analysis of a Transportation Corridor Simulation Model," *Transportation Research B*, 21B, 339-357.

Inaba, F.S. and Wallace, N.E. (1987), "Spatial Price Competition and the Demand for Freight Transportation," Manuscript.

King, M. A. (1980), "An Econometric Model of Tenure Choice and Demand for Housing As a Joint Decision," *Journal of Public Economics*, 14, 137-159.

Koppelman, F. S. (1974), "Prediction with Disaggregate Models: The Aggregation Issue," *Transportation Research Record*, 527, 73-80.

——— (1976a), "Guidelines for Aggregate Travel Prediction Using Disaggregate Choice Models," *Transportation Research Record*, 610, 19-24.

——— (1976b), "Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models," *Transportation Research Forum Proceedings*, 592, 17-23.

——— (1984), "The Application of Disaggregate Choice Models to Travel Demand Forecasting: Issues and Methods," *Transportation Planning Models*, 19-59.

——— and Ben-Akiva, M. E. (1977), "Aggregate Forecasting with Disaggregate Travel Demand Models Using Normally Available Data," in *Transportation Decisions in an Age of Uncertainty*, (E. J. Visser (ed), Martinus Nijhoff), (The Hague).

- 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.
- and Reid F. (1975), "Aggregate Travel Demand Forecasting from Disaggregate Behavioral Models," *Transportation Research Record*, 534, 24-37.
- Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, (Cambridge, MA: Cambridge University Press).
- Morrison, S. and Winston, C. (1986), *The Economic Effects of Airline Deregulation*, (Washington, D. C.: The Brookings Institution).
- Reid F. (1978), "Minimizing Error in Aggregate Predictions from Disaggregate Models," *Transportation Research Record*, 673, 59-65.
- Small, K. A. and Rosen, H. S. (1981), "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49, 105-130.
- Talvitie A. (1973), "Aggregate Travel Demand Analysis with Disaggregate and Aggregate Travel Demand Models," *Proceedings, Transportation Research Forum*, 14, 583-603.
- Train, K. (1986), *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, (Cambridge, MA: MIT Press).
- Westin, R. B. (1974), "Predictions from Binary Choice Models," *Journal of Econometrics*, 2, 1-16.
- Westin, R. B., and Gillen D. W. (1978), "Parking Location and Transit Demand," *Journal of Econometrics*, 8, 75-101.
- Winston, C. (1981), "A Disaggregated Model of the Demand for Intercity Freight Transportation," *Econometrica*, 49, 981-1006.

Zorn, P. M. (1988), "An Analysis of Household Mobility and Tenure Choice: An Empirical Study of Korea," *Journal of Urban Economics*, 24, 113-128.

