University of Massachusetts
Transportation Center

DISAGGREGATE ATTRACTION-END CHOICE MODELING FOR HOME-BASED WORK AND HOME-BASED SHOPPING TRIPS

JULY 1997

UMTC-96-2

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Report Date July 1997
3. Title and Subtitle Disaggregate Attraction-End Choice Modeling for Home-Based Work and Home-Based Shopping Trips	4. Author(s) Chandra Bhat Vamsi Pulugurta Ajay Govindarajan
5. Performing Organization Name and Address University of Massachusetts Transportation Center 224 Marston Hall Amherst, MA 01003	6. Performing Organization Report No. UMTC-96-2
7. Sponsoring Agency Name and Address Massachusetts Highway Department Ten Park Plaza Boston, MA 02116	8. Type of Report and Period Covered Final Report 1 October 1995 – 1 September 1997

9. Supplementary Notes

Prepared in cooperation with the Massachusetts Highway Department and the United States Department of Transportation, Federal Highway Administration

10. Abstract

The ability of travel demand models to provide good forecasts requires that they be casual; that is, the models should represent the travel decisions made by individuals (and households) and should incorporate important demographic and policy sensitive explanatory variables. This recognition has led to a shift from the aggregate modeling paradigm to the disaggregate modeling paradigm, as evident in the widespread use of disaggregate trip production and mode choice models in practice. However, this shift toward disaggregate procedures has not yet influenced the fundamental specification of trip attraction and distribution models employed in practice. This research develops (and estimates) disaggregate attraction-end choice models that will facilitate the replacement of the aggregate trip attraction and distribution models currently in use. The research also compares the proposed disaggregate attraction-end choice model with the disaggregate equivalent of the gravity model.

11. Key Words Attraction-end, Disaggregate choice modeling, Spatial structure, Gravity model	12. Distribution Statement Unrestricted. This document is available through the University of Massachusetts Transportation Center, 224 Marston Hall, Amherst, MA 01003

13. Security Classif. (of this report) Unclassified	14. Security Classif. (of this page) Unclassified	15. No. of Pages 65

Disaggregate Attraction-End Choice Modeling for Home-Based Work and Home-Based Shopping Trips

Final Report

Principal Investigator:

Chandra Bhat
Assistant Professor
University of Massachusetts

Authors:

Chandra Bhat
Vamsi Pulugurta
Ajay Govindarajan

Report of Research Conducted for:

Massachusetts Highway Department
Ten Park Plaza
Boston, MA 02116-3973
ISA #7323

University of Massachusetts Transportation Center
224 Marston Hall
Amherst, Massachusetts 01003-5223
Telephone: (413) 545-3728
FAX: (413) 545-9569

July 29, 1997

Acknowledgments

This research was funded by the Massachusetts Highway Department under the guidance of the Central Transportation Planning Staff (CTPS) of the Boston Metropolitan Organization. The authors would like to thank Karl Quackenbush, Ian Harrington and other members of the CTPS for providing valuable input and feedback during the course of the research.

EXECUTIVE SUMMARY

The objective of the research presented in this report is to develop (and estimate) disaggregate attraction-end choice models that will facilitate the replacement of the aggregate trip attraction-end and trip distribution models currently in use by the Central Transportation Planning Staff (CTPS) of the Boston Metropolitan Organization (MPO). The disaggregate models developed here use individual/household characteristics, attributes of transportation service, and the characteristics of destinations as determinants of attraction-end choice. The research advances the state of the art in disaggregate attraction-end choice modeling while, at the same time, ensuring that the proposed model formulation can be integrated within the current CTPS travel demand modeling structure.

The structure for the attraction-end choice model takes the form of a non-linear-in-parameters multinomial logit model. The non-linear-in parameters structure is necessitated because of the accommodation of multiple measures to represent the size of an attraction-end zone. A parallel conductance formula is used to develop a composite impedance measure for travel between each zone pair from the elemental impedance measures of in-vehicle time, out-of-vehicle time, and cost for each available model serving the zone pair.

The attraction-end choice model associates the individual's attraction-end traffic zone choice for each trip production to relevant exogenous variables. Since the choice of attraction-end zone is characterized by a large number of alternatives, a random subset of the feasible attraction-end zones is chosen during the estimation of the choice model.

The research estimates attraction-end models for two trip purposes: home-based work and home-based shopping/personal business. Six sets of explanatory variables are considered for inclusion in the models. These are: a) impedance variables, b) zonal size measures, c) zonal attractiveness measures, d) zonal location indicators, e) a zonal spatial structure measure that accommodates the impact of the location pattern of zones, and f) interaction of soico-demographic variables with impedance and zone-associated variables. The sample used in the analysis was drawn from the 1991 Boston Household Travel Survey and from supplemental data sources.

The empirical results for the home-based work trip purpose indicate that the variables which significantly impact attraction-end choice include the travel impedance variable, total employment in the zone, a dummy variable for zones far away from Boston Central Business District, and the soico-demographic interactions with the impedance variable. The effect of all the variables are quite intuitive. A larger impedance between the production zone and a candidate attraction-end zone makes it less likely that the candidate attraction-end zone will be chosen. Zones with high total employment are more likely to be chosen as the attraction-end (relative to zones with low employment). Zones which are distant from the Boston Central Business District draw fewer work trips than zones close to the Central Business District. The sensitivity to impedance is higher for individuals over 55 years relative to individuals younger than 55 years. The high sensitivity to impedance of individuals over 55 years might be a reflection of mobility and physical dexterity challenges that such individuals experience. The higher sensitivity of women and lower sensitivity of high income earners to impedance found in our results are consistent with the findings from earlier studies.

The variables found to affect attraction-end choice for the home-based shopping purpose include the impedance variable, two zonal size measures (retail plus service employment and zonal area), a dummy variable for zones distant from the Boston Central Business District, a zonal spatial structure measure, and socio-demographic interactions with the composite impedance measure. A comparison of the impedance coefficient between the shopping and work purposes indicates that there is much greater sensitivity to travel impedance for shopping relative to work. Zones with larger retail plus service employment and zones with larger spatial area are more likely to attract home-based shopping trips than zones with smaller retail plus service employment and smaller area. As with the work purpose, zones which are distant from the Boston Central Business District appear to draw fewer shopping trips compared to zones close to the Boston Central Business District. The spatial structure measure indicates that zones in close proximity to other shopping opportunities have a lower likelihood of being selected as the attraction-end as compared to zones in spatial isolation. The socio-demographic interactions with composite impedance are statistically significant. There was no difference in sensitivity among individuals below 65 years of age, but individuals above 65 years are more sensitive to impedance than their younger counterparts. Women appeared to be more sensitive to impedance than men and individuals in the higher income brackets are less sensitive to impedance than individuals in lower income brackets.

The performance of the choice models estimated in this research was compared with the performance of the disaggregate equivalent of the conventional gravity model.

The disaggregate equivalent of the conventional gravity model includes only the composite travel impedance measure and a single size measure (with the coefficient on the size measure constrained to one). The evaluation results clearly indicate the superior performance of the choice models compared to the gravity model. Thus, including multiple size measure, zonal attractiveness and location measures, the location pattern of zones, and socio-demographic interactions with travel impedance appear to be important in attraction-end choice modeling. However, the conventional gravity model ignores the impact of these factors.

The research develops a simple procedure to estimate the aggregate trip interchanges between each zonal pair and total trip attractions to each zone from the estimation results of the choice models. The only additional requirement that this procedure places on the analyst (compared to the requirements placed by the conventional gravity model) is that the analyst must provide trip attractions by a number of socio-economic categories for each zone rather than just total trip productions. Even if the analyst is unable to breakdown trip production by socio-economic grouping, the results of the choice models can be selectively implemented. In particular, the composite impedance formulation and the composite size measure formulation of the choice models can be directly used with the conventional gravity model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Background

Reliable travel demand estimation and forecasting is a critical prerequisite to examining transportation related issues such as evaluating the effects of alternative traffic congestion alleviating strategies, estimating travel demand on a potential new service and forecasting travel demand into the future. The 1990 Clean Air Act Amendments (CAAA) also place a number of requirements for improvement in travel demand modeling (Guensler, 1993).

The ability of travel demand models to provide a good representation of the traffic flow changes in response to potential changes in the transportation infrastructure and/or population socio-demographics requires that they be *causal*; that is, the travel demand models should represent travel decisions made by individuals (and households) and should incorporate important demographic and policy-sensitive explanatory variables (Oi and Shuldiner, 1962; Domencich and McFadden, 1975). This recognition has led to a shift in the past couple of decades from the aggregate (zonal level) paradigm toward the disaggregate (individual/household level) modeling paradigm, as evident in the widespread use of disaggregate trip production and mode choice models in practice (COMSIS Corporation 1988; Metropolitan Service District, 1989; Chang and Plager, 1994; McClennen, Quackenbush and Gallagher, 1994). However, the shift toward disaggregate procedures has yet to influence the fundamental specification of trip attraction-end and trip distribution models; in particular aggregate models continue to be used in most urban areas for these two components of the four-step transportation planning process. The aggregate models fail to take account of the characteristics of households and trip makers and the way in which these characteristics interact with the attributes of alternative attraction-end zone and of the transportation system. The aggregate models also assume (rather unrealistically) that the total trips attracted to a zone is solely a function of zonal land-use type and intensity, irrespective of its location and accessibility from other zones (since the total zonal attraction modeling does not include travel impedance attributes; see Jones 1978).

1.2 Objective of the Research

The objective of this research is to formulate and estimate disaggregate attraction-end choice models that will facilitate the replacement of the aggregate trip attraction models and the aggregate trip distribution models currently used by most metropolitan planning organizations. The traditional planning process first estimates trip productions from each zone and trip attractions to each zone and subsequently "links" trip productions to trip attractions in the trip distribution stage. This process will be revised with the introduction of attraction-end choice models. In the revised procedure, trip productions from each zone are first estimated using disaggregate methods. Subsequently, the attraction-end of each trip production is determined by the attraction-end choice model and this provides the trip interchanges between each zonal pair. The total trip attraction to each zone is implicitly determined since it is simply the summation of all paired trip interchanges involving the relevant zone as the attraction-end. As part of the objective to estimate disaggregate attraction-end choice models, the research also evaluates the performance of the disaggregate choice models with the disaggregate equivalent of the conventional gravity model.

1.3 Structure of the Report

The rest of this report is organized as follows. Chapter 2 presents an overview of the attraction-end choice models that have been proposed in earlier research. It includes a comprehensive review of the various disaggregate trip distribution models that have been developed. The exogenous variables used in these models and the methodology of inclusion of these variables is also addressed.

Chapter 3 develops the formulation for the disaggregate attraction-end choice model and presents a procedure to estimate the model. The chapter also discusses the procedure employed to define attraction-end choice alternatives and the explanatory variables considered for inclusion in the attraction-end model.

Chapter 4 focuses on the preparation and assembly of the data for analysis. The chapter includes a documentation of the various steps involved in data gathering, data cleaning, and data assembly.

Chapter 5 identifies the explanatory variables considered in the analysis and presents/discusses the empirical results of the disaggregate attraction-end choice models for home-based work and home-based shopping purposes. An approach to obtain aggregated zonal trip interchanges from the disaggregate attraction-end choice model is also presented.

Chapter 6 concludes the paper by discussing the important findings from the research.

CHAPTER 2

Literature Review

Disaggregate trip distribution models (*i.e.* attraction-end choice models) have attracted relatively less attention in travel demand literature compared to trip generation and mode choice models. Previous research studies on disaggregate models of attraction-end choice include Choukroun (1975), Koppelman and Hauser (1978), Recker and Kostyniuk (1978), Timmermans *et al.* (1982), Daly (1982) and Kitamura (1984).

One of the earliest studies in disaggregate attraction-end choice models was conducted by Choukroun (1975). Choukroun adopts a probabilistic approach for developing the trip distribution pattern of a homogenous class of travelers and subsequently generates aggregate models of trip distribution from the probabilistic model. However, Choukroun uses only travel cost as the measure of travel impedance and his model does not accommodate the effect of differences in socio-demographics on attraction-end choice.

Koppelman and Hauser (1978) estimate attraction-end choice models for non-grocery shopping trips using several model formulations and compare the performance of these alternative formulations. Based on their research, they identify factor analysis as a plausible approach to describing consumers' spatial cognitive process while choosing among alternative attraction-end choices. They also find that statistical logit-based preference models perform reasonably well in predicting the attraction-end choice of individuals. The empirical results from the study indicate that attractiveness and accessibility of the attraction-end are useful determinants of attraction-end choice.

Recker and Kostyniuk (1978) formulate a multinomial logit model for grocery shopping trips which is similar in structure to the model employed by Koppelman and Hauser. They hypothesized that urban grocery shopping trips are determined by three factors: the individual's perception of the attraction-end, the individual's accessibility to the attraction-end and the number of attraction-end alternatives available to the individual. Recker and Kostyniuk found accessibility of the attraction-end to be the primary determinant of attraction-end choice.

While the studies by Koppelman and Hauser and Recker and Kostyniuk have provided important insights into the elements of attraction-end choice, their primary focus is on estimating attraction-end choice using attitudinal consumer data. Such attitudinal data are not easily obtained and, consequently, the models of Koppelman and Hauser and Recker and Kostyniuk are not readily suitable for use in transportation planning practice.

Timmermans *et al.* (1982) apply Kelly's (1955) repertory grid methodology to identify the factors influencing the consumer choice of shopping centers. Kelly's repertory grid methodology is based on the fundamental postulate that an individual uses strictly personal constructs to guide his actions. This method is a useful technique for eliciting the criteria people use to differentiate perceptually between shopping centers. However the interviews which are required to elicit the individual personal constructs are time consuming and demanding, and not practical.

Daly (1982) focuses on disaggregate attraction-end choice models based on the multinomial logit form. Daly's model includes a composite travel cost variable and variables describing the quantity of the elementary choices within each aggregated zone. The study, however, is limited in its focus on empirical application since it ignores the socio-demographic characteristics of the individual making the trip.

Kitamura (1984) develops a model of attraction-end choice within the context of trip-chaining. He uses the proximity of a candidate attraction zone to other activity opportunities as a measure of the utility of that zone. Kitamura's study, however, develops an attraction-end choice model which takes the form of a complex simultaneous equation system and which is difficult to implement in practice.

The focus of the proposed research is to develop a disaggregate attraction-end choice model from revealed preference data (*i.e.* from observed attraction-end choices made by individuals) that will be easy to implement and that will, at the same time, be more behavioral in its representation of attraction-end choice models compared to the traditional aggregate gravity model.

CHAPTER 3

Disaggregate Attraction-End Choice Model Formulation

3.1 Definition of Attraction-End Choice Alternatives

The proposed disaggregate attraction-end choice model predicts the individual choice of travel to aggregated zones (or spatial clusters) and not to specific attraction-ends within the zones (or elemental alternatives). This approach is adopted for two reasons. First, there will be a number of choices in each individual's choice set if elemental alternatives were used. This makes the modeling process difficult, imposes infeasible data processing requirements and poses problems in defining alternatives. Second, the desired end-result is to predict trip interchanges between zone-pairs and not elemental alternatives. Therefore, from an aggregation standpoint, developing attraction-end choice models with zonal choice alternatives is preferable.

The choice of trip attraction-end is characterized by a large number of alternatives even after defining alternatives at the zonal level. However, by adopting a Identically and Independently Distributed (IID) structure for error terms across the zonal attraction alternatives, the attraction-end choice model can be consistently estimated with only a subset of the alternatives from the feasible choice set (McFadden, 1978). The subset of alternatives can be drawn in many different ways from the feasible choice set (Ben-Akiva *et al.*, 1984). The simplest method is to use a random sampling approach where the subset includes the chosen attraction-end alternative and a random sample of non-chosen alternatives from the feasible choice set. This is the method adopted in the current research. The total number of attraction-end alternatives sampled for each trip case is pre-specified (six randomly selected alternatives and the chosen alternative).

3.2 Model Structure

The alternatives in the attraction-end choice model are aggregated zones. Each zone j may include several possible elemental attraction alternatives. Let the number of elemental attractions in j be D_j . Assume the following: a) D_j is large for each zone j, b) utilities of elemental alternatives within each zone are IID (conditional on unobserved zonal attributes) and c) the systematic utility of the elemental attraction alternatives are relatively homogenous within each aggregate zone or the within-zone variance of the systematic utilities of the elemental alternatives in each aggregate zone are almost equal across all zones. With the foregoing assumptions, the utility (benefit) U_{ij} presented by attraction-end j for a trip production made by an individual q in zone i may be written as (see Daly, 1982):

$$U_{ijq} = V_{ijq} + \in_{ijq} = \mu' z_{ijq} + \eta \, log(D_j) + \in_{ijq} \tag{1}$$

where, z_{ijq} is a vector comprising : a) travel impedance measures for travel between zones i and j, b) zonal attractiveness and location attributes of candidate attraction zone j and c) interactions of the socio-demographic characteristics of the individual q with travel impedance and zonal attributes. D_j is the number of elemental alternatives in the zone j, and $\in_{ijq}$ is a random term distributed IID gumbel across zonal alternatives and individuals, and μ and η are parameter vectors to be estimated. In equation (1), D_j (the number of elemental work or shop attractions in zone j) is not easily quantifiable. However, sets of observable size variables such as employment in zone j or land area of zone j can be used to 'proxy' D_j. Let d_j represent a vector of 'proxy' size variables for zone j and let δ be a corresponding vector reflecting the contribution of the proxy size variables to the actual zone size D_j. Equation (1) can then be re-written as:

$$U_{ijq} = V_{ijq} + \in_{ijq} = \mu' z_{ijq} + \eta \, log(\delta' d_j) + \in_{ijq} \qquad (2)$$

The magnitude of the parameter η in the above equation, characterizes the presence of common unobserved zonal attributes (such as congestion or parking availability) affecting the attractiveness of elemental alternatives in a zone (Daly, 1982). If there is no correlation in the unobserved factors affecting the attractiveness of elemental alternatives within a zone, η equals one. On the other hand, if all the unobserved factors affecting the attractiveness of elemental attraction-ends within a zone are zonal attributes, η equals zero. In practice, the value of η is likely to lie between 0 and 1. η is essentially an inclusive parameter in a nested logit structure with aggregate zones at the upper level of the nest and the elemental alternatives within the zone at the lower level. If the lower level nest is not modeled, as in the current research, it can be shown that the appropriate utility expression for choice at the higher-level nest (*i.e.*, aggregate zones) is given by equation

(2) under the three assumptions listed at the beginning of the section (McFadden, 1978; Ben-Akiva and Lerman, Chapter 7, 1985).

The probability that an individual q would choose attraction-end j from the set C_i of attraction-end zones accessible from zone i may be formulated as:

$$P_{ijq} = \frac{e^{V_{ijq}}}{\sum_{j' \in C_i} e^{V_{ij'q}}} \tag{3}$$

The model specified in equation (3) may appear to be the familiar multinomial logit model. However, this is not the case as the expression in equation (3) includes the component $\eta log(\delta' d_j)$. Thus, the model takes the form of a non-linear-in-parameters multinomial logit model.

If there is only one zonal size variable in the size vector d_j, the coefficient δ on this variable should be constrained to one. In this situation the model represented by equations (2) and (3) collapses to the standard multinomial logit model. In the more general case of M zonal size measures in the vector d_j the coefficient on one of the size measures in the vector δ has to be normalized to one for identification.

3.3 A Comparison between the Gravity Model and the Disaggregate Choice Model

The disaggregate choice model specified by equations (2) and (3) can easily be shown to be more general compared to the conventional production-constrained gravity model. Consider a production zone i and a candidate attraction zone j. Let the only variable in the vector z_{ijq} in equation (2) be the travel impedance between zones i and j

(say C_{ij}). Also assume a single size variable in the vector d_j (say employment) and let η be constrained to 1. As indicated earlier, since there is only one size variable, the corresponding coefficient in the vector δ cannot be identified and is set to 1. Based on these constraints, equation (2) reduces to :

$$U_{ij} = V_{ij} + \epsilon_{ij} = \mu C_{ij} + log(d_j) + \epsilon_{ij} \tag{4}$$

From equation (4) the probability of choice of attraction-end j from the set of available alternatives is:

$$P_{ij} = \frac{e^{\mu C_{ij}} \times d_j}{\sum_{j' \in C_i} e^{\mu C_{ij'}} \times d_{j'}} \tag{5}$$

Let O_i be the total number of trip productions originating at i. The trip-interchange T_{ij} between the zones i and j can be obtained by aggregating the probability in equation (5) across all trip productions O_i. Since the probability in equation (5) does not vary among trip productions originating in zone i, T_{ij} can be written as:

$$T_{ij} = O_i P_{ij} = O_i \times \left(\frac{1}{\sum_{j' \in C_q} e^{\mu C_{ij'}} \times d_{j'}} \right) \times e^{\mu C_{ij}} \times d_j = O_i A_i d_j e^{\mu C_{ij}} \tag{6}$$

which is exactly the conventional production-constrained aggregate gravity model. Thus, the model represented by equations (4) and (5) is the disaggregate equivalent of the gravity model. As such, using the disaggregate specification in equation (2) which accommodates socio-demographics, multiple size variables and variables representing the attractiveness measures of attraction-end zones can only provide better aggregate forecasts than the gravity model (It should be noted that the disaggregate model of equation (2) and the disaggregate equivalent of the gravity model of equation (4) can be compared using a standard nested likelihood ratio test).

3.4 Model Estimation

The estimation of the model represented by equations (2) and (3) cannot be accomplished by a standard multinomial logit software or a nested logit software. Special purpose code needs to be developed to estimate the model. The estimation procedure developed in this research uses the maximum likelihood method. The log-likelihood function to be maximized takes the form:

$$L = \sum_{q=1}^{Q} \sum_{j \in C_i} y_{qj} \left(\mu' z_{ijq} + \eta \log(\delta' d_j) - \log\left[\sum_{j' \in C_i} e^{\mu' z_{ij'q} + \eta \log(\delta' d_{j'})} \right] \right) \tag{7}$$

where, y_{qi} is a set of dummy variables such that:

$$y_{qj} = 1 \qquad \text{if individual } q \text{ (in production zone } i\text{) chooses zone } j$$

$$= 0 \qquad \text{otherwise} \tag{8}$$

The maximization of the log-likelihood function is achieved by a program written and coded in the GAUSS matrix programming language using an iterative David-Fletcher-Powell (DFP) algorithm.

CHAPTER 4

Data Source and Sample

The sample used for estimation is a household activity survey conducted by the Central Transportation Planning Staff (CTPS) of the Boston Metropolitan Organization in April of 1991. The survey collected data on the socio-demographic characteristics of the trip-maker's household and each individual in the household. The survey also included a one-day (mid-week working day) activity diary to be filled out by all members of the household above five years of age. The socio-demographic information and the activity information were assembled in a "person/household socio-demographic" file and a "trip file", respectively, by CTPS.

The Boston Metropolitan area is divided into 986 traffic zones. CTPS provided a "zonal demographic and land-use" file containing zonal attribute data for each of the 986 traffic zones. CTPS has also developed a "zone-to-zone travel impedance" file for travel between each zone pair in the region.

The sample used in the current analysis was generated from the four files listed above: that is, from the a) survey trip file, b) person/household socio-demographic file, c) zonal demographic and land-use file and d) zone-to-zone impedance file. The four files listed above were imported into SPSS, a statistical package, for data manipulation. Additional zonal data of potential use in explaining attraction-end choice were generated from supplementary reference sources and coded in SPSS format (referred to as data generation step). The data generation step was followed by a cleaning process to remove missing values on relevant variables and to impute values for the missing income observations (referred to as data cleaning). Finally, the four 'clean' data files were brought together and assembled in a format suitable for attraction-end choice modeling of home-based work and home-based shopping trips (referred to as data assembly).

Sections 4.1, 4.2 and 4.3 document and detail the data generation, data cleaning and data assembly steps, respectively.

4.1 Data Generation

Three zone-based variables that might be associated with the attractiveness of a zone as an attraction-end for home based work and shopping trips were generated. These were: a) zonal unemployment rate, b) zonal crime rate and c) a zonal spatial structure variable which captures the spatial proximity of a zone to possible shopping opportunities in other zones. The first two variables were developed from supplementary reference sources, while the third variable was developed from zonal retail employment data and zone-to-zone impedance data. The procedures adopted to generate the first two variables are discussed in the next two subsections. The formulation developed to compute the third variable; that is the zonal spatial structure variable; involves empirical issues. Hence it is discussed along with the presentation of the empirical results in chapter 5.

4.1.1 Zonal Unemployment Rate

Unemployment rate information was obtained on a town basis from the Department of Employment and Training (1992). The unemployment rate represents the percentage of individuals in the labor force residing in the town who are not employed. The zones that lie within a particular town were identified from the zonal land-use file. The unemployment rate of each zone was then obtained by assigning the unemployment rate of each town to all its constituent zones. The zonal unemployment rate variable was used in home-based work trip attraction-end choice model estimation.

4.1.2 Zonal Crime Rate

Crime rate information was obtained on a town basis from a report prepared by the Massachusetts State Police Crime Reporting Unit (1992). Crime rate was measured as the number of police reported crimes for the year 1992. The number of crimes reported for each town was evenly distributed to all its constituent zones. This variable was used in both home-based work and home-based-shopping attraction-end choice model estimations.

4.2 Data Cleaning

The four files (listed in the beginning of this chapter) were subjected to a cleaning process to remove observations with missing values and/or to impute missing values on relevant variables.

Fig. 1 presents the screening procedures for the survey trip file (the trip file includes only home-based work and home-based shopping trips, selected from the larger and complete trip file). The screening criteria were applied so that the travel impedance data for the trip may be appended later based on travel mode, origin/destination of trip and trip departure time. Fig. 2 shows the cleaning procedure applied to the person/household socio-demographic file. Individuals for whom information on sex or age is not available were removed. The missing values for income were imputed based on regressing income values on relevant socio-demographic information for observed income records.

4.3 Data Assembly

The data assembly procedure involved two broad steps: a) Preparing each of the four files; the survey trip file, the person/household socio-demographic file, the zonal demographics/land-use file and the zone-to-zone impedance file; for assembly (individual file preparation step) and b) the actual assembly of the four files to obtain the master estimation files for home-based work and home based shopping attraction-end choice modeling (master estimation file development step).

4.3.1 Individual File Preparation

The steps taken to prepare each of the individual files are outlined below.

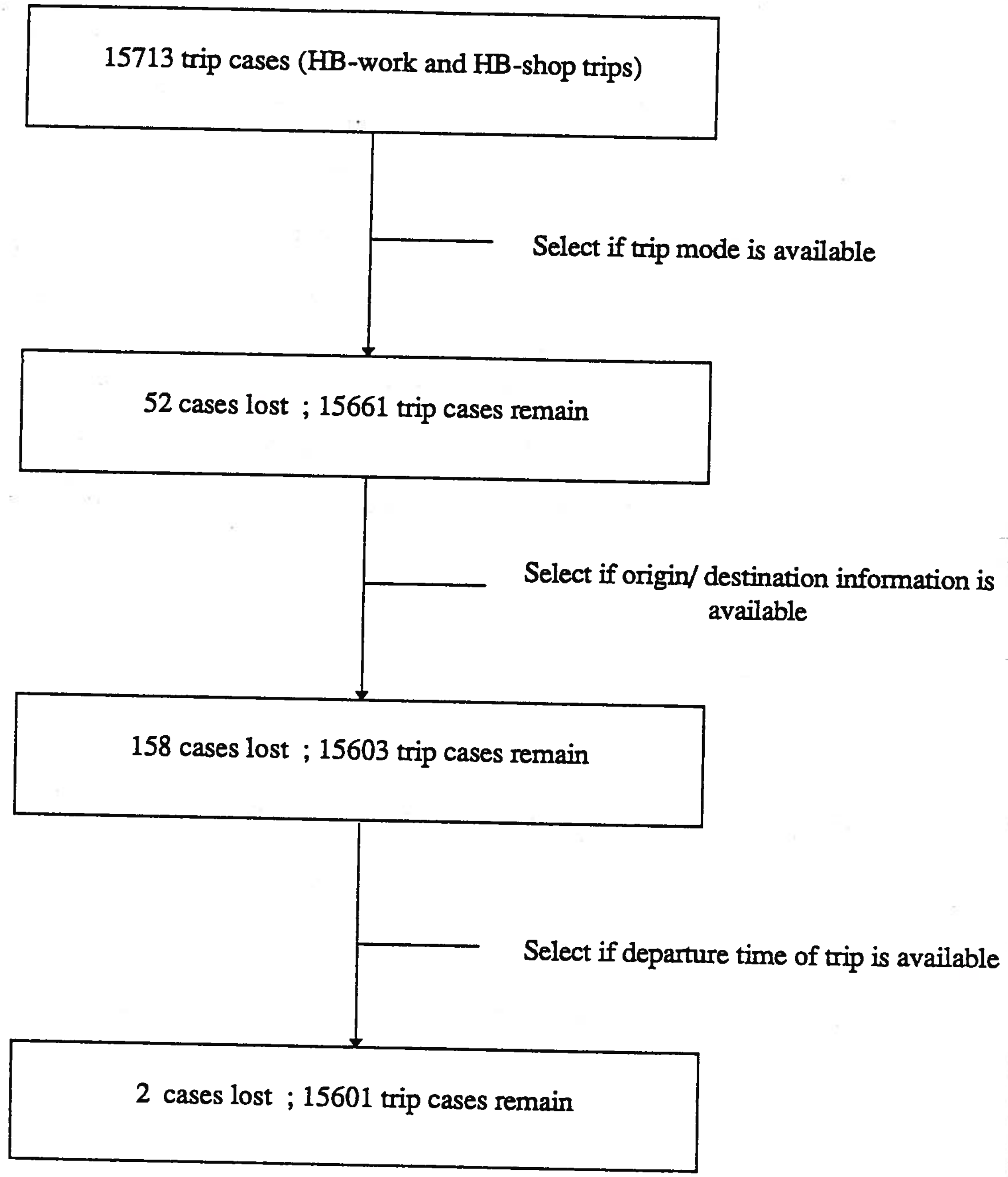


Fig 1. - Data Cleaning of Survey Trip File

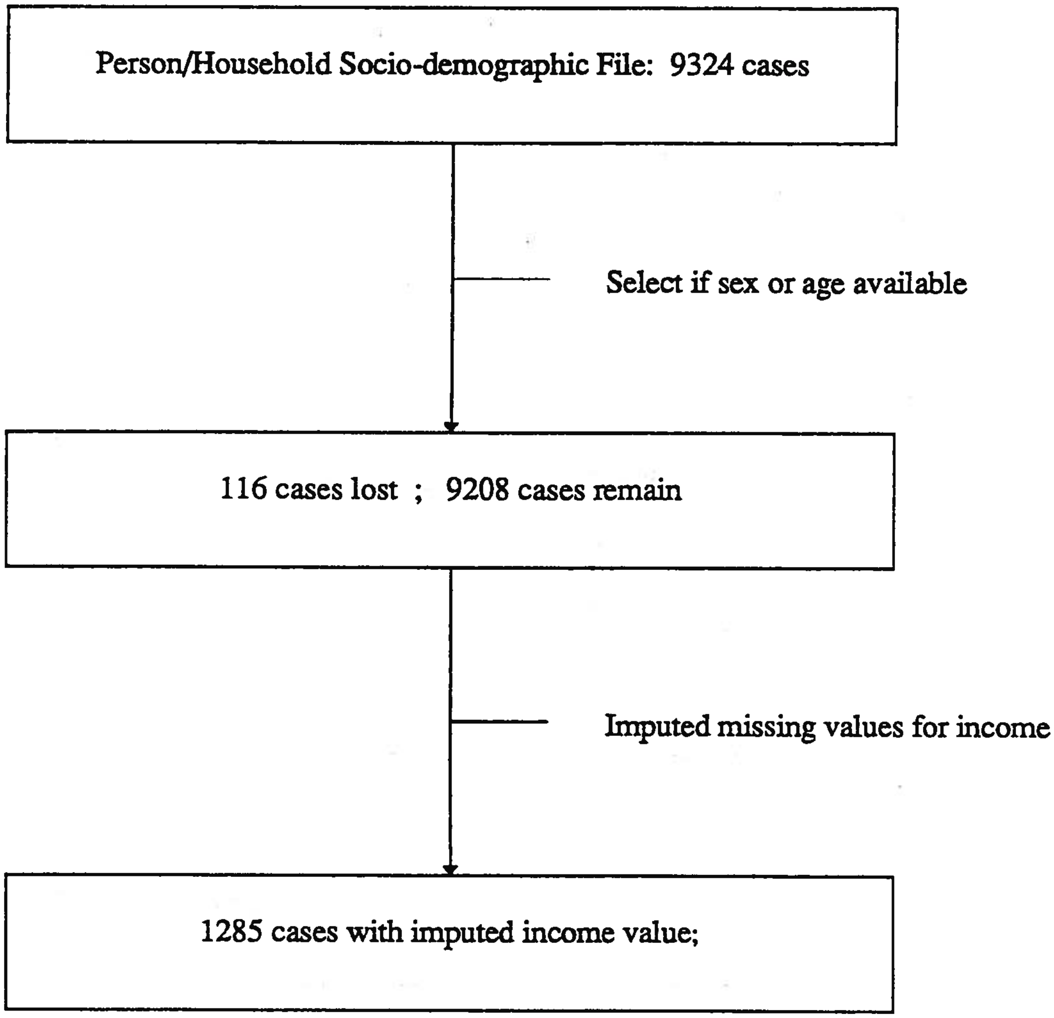


Fig 2. - Data Cleaning of Person/Household Socio-demographic File

4.3.1.1 Survey Trip File

A dummy variable was created indicating the time period of the day (AM peak, midday, PM peak, evening) in which the trip was made using the mid-point of the departure and arrival time of the trip as the basis for classification. This was done so that the impedance values of the appropriate time-period may be appended to each trip case. The origin and destination ends of each trip were converted to production and attraction ends.

4.3.1.2 Person/Household Socio-demographic File

No changes were made in these files.

4.3.1.3 Zonal Demographics and Land-use File

The three zonal variables generated in the data generation step (*i.e.*, zonal unemployment rate, zonal crime rate and zonal shopping accessibility variable) were appended to the zonal demographic and land use file.

4.3.1.4 Zone-to-zone Impedance File

The in-vehicle time, out-of-vehicle time, and travel cost variables for each zonal pair obtained from the impedance files provided by CTPS were as follows:

4.3.1.4.1 Highway Impedance Files

The values for in-vehicle travel time, and out-of-vehicle travel time were taken directly from the CTPS highway impedance files. Three variables associated with cost were created.

Cost1: Sum of the toll costs and the automobile maintenance costs.

Cost2: Sum of Cost1 and the parking cost. The average duration of parking for a shopping trip is assumed to be 2 hours, hence, the per hour parking cost was taken as the effective parking cost for each shopping trip.

Cost3: Sum of Cost1 and four times the parking costs. The average duration of parking for a work trip is to be 8 hours, hence, the per hour parking cost was taken as the effective parking cost for each work trip.

4.3.1.4.2 Transit Impedance Files

The values for in-vehicle travel time (IVTT) and travel cost (transit fare) were taken directly from the transit impedance file. The values of out-of-vehicle travel time (OVTT) was computed as the sum of transit walk time, auto access time, initial wait time and transfer wait time. The impedance values (IVTT, OVTT, transit fare) were available for two possible access modes (drive alone and walk). If only one access mode was available for a zone pair, the impedances of that access mode were taken as the relevant values. If both access modes were available, the impedance values corresponding to the best access mode was chosen (*i.e.*, the access mode with lower travel time). Absence of both access modes for a zone pair indicated lack of transit availability and a zero value was assigned for all the transit impedance variables.

4.3.1.4.3 Walk Impedance Files

The inter-zonal walk distances were obtained from the walk impedance files. The walk time was computed assuming a walk speed of 3 miles/hr. Zone pairs separated by a distance greater than 6 miles were assumed not to have walk access and hence assigned zero values.

4.3.2 Master Estimation File Development

The following are the sequence of steps involved in the assembly of the master estimation file.

a) Home-based work and home-based shopping trips were selected from the trip file and the origin-destination ends of each trip were converted to production-attraction ends.

b) The travel survey trip file was matched with the person/household socio-demographic file to append the socio-demographic characteristics of the individual to each of his/her trip productions.

c) A descriptive analysis was conducted on trip length to obtain the longest trip for each of the home-based-work and home-based -shop trip purposes. The longest trip length was about 72 miles for both trip purposes and this distance was selected as the basis for determining the feasible attraction choice set of zones for each production zone.

d) The feasible choice set was identified for each production zone; six candidate attraction zones (but not the actual chosen attraction zone) were selected randomly from the feasible attraction choice set for each trip production and the actual chosen attraction-end was added to these six randomly chosen alternatives (for a total of seven attraction-end alternatives).

e) Each trip production in the trip file was replicated seven times, and each attraction-end alternative generated in the previous step was appended to one of the seven replicated records of each trip production.

f) The impedance values by all available modes for the production-attraction zone pair of each record in the replicated trip file were appended from the zone-to zone impedance file; the impedance values were available by four time-periods in the day and the values corresponding to the time-of-day of the trip production were selected for appending.

g) The zonal attributes of each candidate attraction-end alternative were appended from the zonal demographics and land-use file.

The final sample files included 7337 trip cases (7337*7=51359 records) for the home-bases work purpose and 7963 trip cases (7963*7=55741 records) for the home-based shopping purpose.

CHAPTER 5

Empirical Analysis

The home-based work and the home-based shopping sample files (obtained after the data preparation steps outlined in the previous chapter) were imported into GAUSS for model estimation. This chapter presents the results of the empirical analysis. Section 5.1 discusses the explanatory variables considered for analysis and the functional form used for them. Section 5.2 focuses on the empirical results for home-based work and home-based shopping trips. The final section outlines an approach to obtain aggregate zonal trip interchanges from the disaggregate attraction end choice models.

5.1 Variable Specification

Six broad categories of explanatory variables were considered for inclusion in the attraction-end choice models. These are: a) impedance variables, b) zonal size measures, c) zonal attractiveness measures, d) zonal location indicators, e) a zonal spatial structure measure, and f) interaction of socio-demographic variables with impedance and zone-associated variables. Each of these sets of variables are discussed in the subsequent sections.

5.1.1 Impedance Variables

The impedance variables include in-vehicle travel time, out-of-vehicle travel time, and travel cost. The appropriate impedance measures were appended to each production zone-candidate attraction zone pair based on the time-period of the individual trip. The impedance variables are defined for each available mode between the production zone and the candidate attraction zone, and are undefined for unavailable modes (see sections 4.3.1 and 4.3.2).

Three issues had to be addressed in the introduction of the impedance variables in the attraction-end choice models. Firstly, there was substantial multicollinearity among different impedance measures (in-vehicle travel time, out-of-vehicle travel time, and travel cost) for each travel mode. Secondly, transit and walk mode were not available from all zone pairs. The differential modal availability for different zone pairs had to be accounted for. Thirdly, a functional form for the introduction of the impedance effect in the choice modeling had to be developed. Each of these issues is discussed below.

The issue of collinearity in in-vehicle time (IVTT), out-of-vehicle time (OVTT), and travel cost (COST) was resolved by converting OVTT and COST into equivalent in-vehicle time units for each available mode. The universal set of modes included were highway, transit, and walk (the bike mode accounted for very few trips for both work and shopping purposes and so is not considered here). The equivalent in-vehicle time units for each available mode (for each zone pair) was computed as follows:

$$C \text{ (Equivalent highway in-vehicle time units)} = \text{Highway IVTT} + \phi^{*}\text{Highway OVTT} + \lambda^{*}\text{Highway Cost}$$

$$T \text{ (Equivalent transit in-vehicle time units)} = \text{Transit IVTT} + \phi^{*}\text{Transit OVTT} + \lambda^{*}\text{Transit Cost}$$

$$W \text{ (Walk impedance)} = \Delta^{*}\text{Walk Time} \quad (9)$$

In the expression above, values for ϕ and λ were obtained from work mode choice modeling among motorized modes. The ratio of the parameters for out-of-vehicle time and in-vehicle time from work mode choice estimation was 1.75 and this value was used for ϕ. The money value of in-vehicle travel time from the work mode choice estimation was found to be $ 4.00 per hour. This translates to an equivalent of 0.15 minutes of in-vehicle time for each cent of travel cost which is the value employed for λ (travel times are measured in minutes and cost are measured in cents in the data). The value of Δ should be between 1 and 1.75. A value of 1 for $\Delta = 1$, would imply that the disutility of one minute of walk time is the same as a minute of in-vehicle time on motorized modes, while if $\Delta = 1.75$, then the disutility of walk time is considered to be the same as out-of-vehicle time on motorized modes. The appropriate value for Δ was determined through empirical estimation, as discussed later in this section.

The second issue in the introduction of the impedance variables related to the differential mode availability among zone pairs. The highway mode is available for all zone pairs in the Boston Metropolitan area. However, the transit and walk modes are not universally available. Consider a production zone i and two candidate attraction-end zones j and k. Let the highway impedance C_{ij} and C_{ik} (in equivalent in-vehicle time units as computed from equation 1) be the same and let zone k be served by only a subset of the modes that serve zone j. Then the overall impedance assigned for travel from zone i to zone j should be lower than that assigned for travel from zone i to zone k. This fundamental principle may be accommodated by specifying a composite impedance for travel using a parallel conductance formula. To develop this formulation, define a transit availability dummy variable y_t (for each zone pair) that takes a value of 1 if transit is available and zero otherwise, and a walk availability dummy variable y_w that takes a value of 1 if walk is available and zero otherwise. The composite impedance H is then written as:

$$H = (1-y_t)(1-y_w)C + y_t(1-y_w)\left(\frac{C}{1+\frac{C}{T^{\beta}}}\right) + y_w(1-y_t)\left(\frac{C}{1+\frac{C}{W^{\gamma}}}\right) + y_t y_w\left(\frac{C}{1+\frac{C}{T^{\beta}}+\frac{C}{W^{\gamma}}}\right) \quad (10)$$

If both transit and walk are unavailable, the first term applies and the composite impedance is just the highway impedance C. If transit is available, but walk is not, the second term applies. If walk is available, but transit is not, the third term applies. If all modes are available, the fourth term applies. β and γ are positive parameters which indicate the relative weights placed on transit and walk modes, respectively, relative to highway as the mode of travel. If $\beta>1$ ($\beta<1$) then the highway mode dictates attraction-end choice more (less) than the transit mode (when both modes are available). Similarly, if $\gamma>1$ ($\gamma<1$) then the highway mode determines attraction-end choice more (less) than the walk mode (when both modes are available). The relative importance between the transit and walk modes can be determined by comparing the values of β and γ. The transit mode determines choice more or less than the walk mode depending on whether $\beta<\gamma$ or $\beta>\gamma$.

The third issue in the introduction of travel impedance is the functional form for the effect of composite impedance on utility. The two forms considered were the linear form and the log-linear form, both of which have been used in earlier studies (Fotheringham, 1986). The linear form implies that the marginal deterrence due to travel impedance is independent of the existing impedance level, while the log-linear form implies that the marginal deterrence decreases as the existing travel impedance level increases (*i.e.*, a constant increase in the composite impedance has a higher deterrence when the initial impedance level is low than when the initial impedance level is high).

The determination of the value for Δ (related to the first issue) and the values for β and γ (associated with the second issue), and the functional form for the composite impedance term (associated with the third issue), was based on empirical estimation. The probability that a trip production from zone i will be attracted to zone j is given by the following "non linear-in-parameters" multinomial logit model (ignoring the effect of other non-impedance measures)

$$P_{ij} = \frac{e^{\alpha f(H_{ij})}}{\sum_k e^{\alpha f(H_{ik})}} \tag{11}$$

where $f(H_{ij}) = H_{ij}$ for the linear functional form and $f(H_{ij}) = ln(H_{ij})$ for the log-linear functional form. H_{ij} is the composite impedance as given by equation (10) for travel from zone i to zone j and α is the coefficient on the impedance term. The model in (11) was estimated with a specialized maximum likelihood code written in the GAUSS matrix programming language using the individual choice sample assembled for the attraction-end modeling. The model was estimated for three pre-defined values of Δ : 1, 1.35 and 1.75. For both the linear and log-linear forms, there was little sensitivity in the log-likelihood function at convergence or in the estimated values for β and γ to the value of Δ used. The best results were, however, obtained when Δ=1.00 for both the linear and the log-linear form. Between the linear and log-linear functional forms, the latter form performed substantially better. The log-likelihood at convergence for the home-based work purpose for the linear form was -11864 and for the log-linear form was -10805 (a difference of 1059 log-likelihood points). The corresponding values for the home-based shopping purpose was -8902 for the linear form and -7166 for the log-linear form (a difference of 1736 log-likelihood points). Thus, the log-linear functional form was chosen for the composite impedance term. The final results of the estimation in this log-linear specification were as follows: β =1.0752 and γ =0.8779 for the work purpose and β =1.9375 and γ =1.0712 for the shopping purpose. All these estimates were highly significant in their difference from the value of 1. These results indicate that, for the work purpose, the highway mode determines attraction-end choice more than the transit mode, but the walk mode determines choice more than the highway and transit modes. However, for the shopping purpose, the highway mode impedance is the primary determinant of attraction-end choice. Using the estimates of β, γ and Δ a composite impedance value (in equivalent highway in-vehicle time units) was computed using equation (10) and introduced in a log-linear form along with other exogenous variables discussed below.

5.1.2 Zonal Size Measures

The zonal size measures considered in the analysis included total zonal employment for the home-based work purpose, and retail plus service employment and zonal land area for the home-based shopping purpose. The method of inclusion of the size variables is not discussed here as it has been presented in detail in section 3.2. The size variables represent proxy measures of the number of elemental attraction-ends within a zone and so the expectation is that "large" zones will be more likely to be chosen as the attraction-end than "small" zones.

5.1.3 Zonal Attractiveness Measures

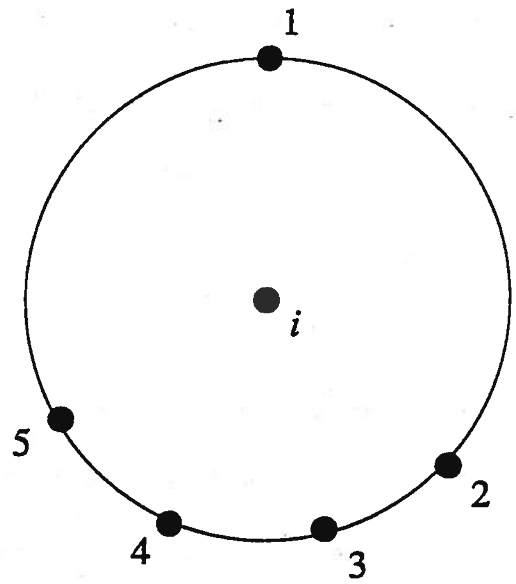
The zonal attractiveness measures included percentage unemployment rate and crime rate for the home-based work purpose, and crime rate for the home-based shopping purpose. As discussed in section 4.1.1, the unemployment rate represents the percentage of individuals in the labor force residing in the zone who are not employed. This variable was used as a "proxy" for the availability of jobs in a zone with the expectation that a higher zonal unemployment rate would result in a lower attractiveness of the zone. Crime rate was measured as the total number of police-reported crimes for 1992 (refer section 4.1.2 for details). The hypothesis was that zones with high crime rates would be less preferred as an attraction-end choice compared to zones with low crime rates, both for work and shopping trip purposes.

5.1.4 Zonal Location Indicators

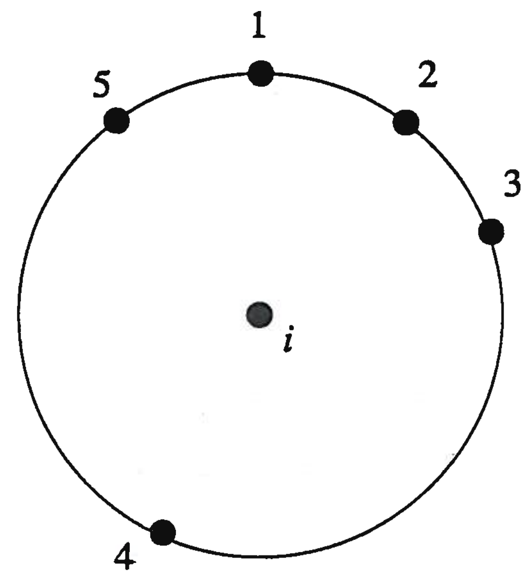
The zonal location indicators represent the geographic location of a zone relative to the location of the Boston Central Business District (CBD). The Central Transportation Planning Staff has developed a "ring" system of geographic location. Ring 0 includes the zones in, and immediately around, the Boston CBD. Ring 4 encompasses the "rural" zones in the periphery of the Boston Metropolitan Region. Rings 1,2, and 3 lie in between. Five dummy variables (ring0, ring1, ring2, ring3 and ring4) were created for each zone. For each zone, a value of '1' was assigned for the ring variable to which the zone belonged and a value of '0' was assigned for the other ring variables. These ring variables capture the effect of miscellaneous attractiveness (or unattractiveness) attributes associated with geographic location. For example, most of the large shopping malls may be located in non-rural areas (*i.e.*, in rings other than Ring 4) and there may be a general preference to shop in such malls than in disjoint shops in the "rural" areas (*i.e.*, in Ring 4). If so, and if the presence of large shopping malls is not controlled for, the effect might manifest itself in the form of larger shopping attractiveness for "non-rural" zones compared to "rural" zones. Since the effect of the ring variables actually represents the impact of a complex milieu of unknown attributes, they are similar to alternative specific constants except that the ring variables are associated with a group of zones rather than each alternative zone.

5.1.5 Zonal Spatial Structure Measure

The zonal spatial structure variable is used for the home-based shopping purpose. This variable is used to accommodate the impact of the location pattern of shopping attraction-end zones. To motivate the inclusion of such a variable, consider the example provided in Fotheringham's (1983) paper. Consider the attraction-end choice of an individual at zone *i* in the two spatial arrangements in Figure 3. All the possible attraction-ends (zones 1 through 5) are equally distant from zone *i* and are identical in all other respects (*i.e.*, they are all of the same size and attractiveness). The traditional gravity model would then estimate the same trip interchange volumes from zone i to each attraction zone. At a disaggregate choice level, the



Spatial Arrangement 1



Spatial Arrangement 2

Figure 3 : Illustration of Spatial Structure Effects

implication is that the probability of attraction-end choice for a trip produced at zone i is the same for all the zones 1 through 5. However, the positioning of the attraction zones relative to one another may have an impact on the choice probabilities and hence on aggregate trip interchanges. One possibility is that the choice probability of zone 1 may be higher in the first spatial configuration than in the second due to "competition" effects. Zone 1 may occupy a unique location in the cognitive perception of individuals in zone i in the first spatial pattern. Equivalently, zone 1 competes less with other zones in the first pattern, while there is more competition among other potential zones (2 through 5). Also, zone 1 may be more attractive in the first configuration because individuals may want to avoid congestion costs in and around a group of zones with several complementary shopping locations in close proximity to each other (Lo, 1991). An alternative possibility is that the choice probability of zone 1 is higher in the second spatial pattern than in the first due to "agglomeration" effects. The presence of several closely clustered opportunities for shopping may provide individuals at zone i with a perception of greater variety (even if this were not actually the case) or more opportunity for comparative shopping. In practice, either competition effects or agglomeration effects may be present and the appropriate effect can be inferred from estimation. To do so, the proximity of a candidate attraction-end zone j to other shopping opportunities is specified using a Hansen-type accessibility measure (see Fotheringham, 1983):

$$M_j = \left[\frac{1}{L} \sum_{l=1}^{L} \left(\frac{\log R_l}{\log H_{lj}} \right) \right] \tag{12}$$

where R_l is the sum of retail and service employment in zone l (a proxy for shopping opportunities in zone l), H_{lj} is the composite travel impedance between zones l and j, and L is the total number of zones in the Boston Metropolitan area. Large values of the proximity variable indicate more opportunities to shop in close proximity of that zone, while small values indicate zones which are spatially isolated from other shopping opportunities. As in the competing attraction-end formulation of Fotheringham (1986) and Borgers and Timmermans (1987), the utility of an attraction-end zone j is specified as a linear function of the proximity variable as θM_j, where θ is a spatial structure parameter. If $\theta < 0$, zones in close proximity to other shopping opportunities have a lower utility than zones in spatial isolation and competition forces exist. On the other hand, if $\theta > 0$, zones in close proximity to other shopping opportunities have a higher utility than zones in spatial isolation and agglomeration forces exist. If $\theta = 0$, it indicates absence of spatial structure effects in attraction-end choice.

5.1.6 Interaction of Socio-Demographic Variables with Other Variables

Previous studies have suggested that zonal attributes and impedance measures might interact with socio-demographic characteristics of the individual in determining attraction-end choice. A consistent finding in geographical literature is that women work closer to home than men because of, among other things, household and child care responsibilities. For example, Dubin (1991) and Hodge (1990) show evidence that household and child care responsibilities leave the women with less time for work-related activities. Consequently, women consider a narrower spatial range in their search for jobs. Similarly, Madden (1981) indicates that age seems to have a negative effect on commuting time, possibly due to physiological considerations that increase the sensitivity of older people to traveling long distances. There is also considerable evidence (Blumen, 1994 and Hanson and Pratt, 1990) that higher income earners travel longer commuting distances. An explanation may be that commuting is a household cost and, therefore, it makes economic sense for individuals with low income earning potential to find jobs close to their residence. The above discussion suggests that interactions of the composite impedance measure with sex, age, and income of the individual be explored for the home-based work purpose. While there is less research on the interaction effects of socio-economic attributes with travel impedance for home-based shopping, these interactions were also explored. Finally, the interactions of the socio-economic attributes with zonal attractiveness/location indicators were also considered to accommodate any differential sensitivities of different population groups to such attributes.

The interaction effects involving age/income were introduced by creating dummy variables characterizing different ranges of age/income and interacting these dummy variables with the composite impedance measure and zonal attractiveness/location indicators. Such a specification does not constrain the age/income effect to be linear or monotonic. Further, the use of dummy variables facilitates the application of the estimated model in forecasting since it is easier to predict trip productions by age/income categories than it is to predict trip productions by the continuous age/income value of individuals.

5.2 Empirical Results

The empirical results are presented in four sections. The following two section presents the estimation results for home-based work attraction-end choice and home-based shopping attraction-end choice respectively. These results represent the final specification after an extensive and systematic exploration of alternative variable specifications. Section 5.2.3

examines the sensitivity of parameter estimates and their standard errors to the number of attraction-end zonal alternatives selected in the choice set. Specifically, the section compares the estimation results for 7, 11, and 16 alternatives in the choice set. Section 5.2.4 compares the performance of the disaggregate choice model developed with the disaggregate equivalent of the conventional gravity model.

5.2.1 Home-Based Work Attraction-End Choice

The column titled "Home-Based Work Purpose" in Table 1 provides the parameter estimates and associated t-statistics for the work purpose. The variables that significantly impact work attraction-end choice include the impedance variable, the composite size variable, the dummy variable for Ring 4, and the socio-demographic interactions with the impedance variable. There were no substantial differences in the utility of zones among rings 0, 1, 2 and 3, but there were substantial differences based on whether a zone was in ring 4 or not. Also, there was considerable multicollinearity between unemployment rate and the ring 4 dummy variable. Including both variables resulted in the incorrect sign on the unemployment rate variable and only a marginal increase in the log-likelihood value over the specification with the ring4 variable alone. So, unemployment rate was dropped from the specification.

The parameter signs on all variables are in the expected direction. A larger impedance between the production zone and a candidate attraction-end zone makes it less likely that the candidate attraction-end zone will be chosen. The composite size measure is represented by the total zonal employment; the parameter on total zonal employment is normalized to one since it is the only size measure employed in home-based work modeling. The sign on the size variable

Table 1: Empirical Estimation Results

Variable	Home-Based Work Purpose		Home-Based Shopping Purpose	
	Parameter	t-stat.	Parameter	t-stat.
Log of composite impedance[1]	-1.3136	-51.04	-1.9366	-86.62
Log of Composite zonal size measure	0.0485	6.80	0.1360	6.67
Total zonal employment[2]	1.0000	-	-	-
Zonal retail plus service employment[3]	-	-	1.0000	-
Zonal land area (in square miles)	-	-	5.18	9.35
Ring 4 dummy variable	-0.7967	-14.49	-0.9395	-14.08
Spatial structure variable	-	-	-4.3618	-48.00
Socio-demographic interactions with impedance				
Female	-0.2410	-30.81	-0.0299	-3.09
55 yrs < age ≤ 65 yrs	-0.2623	-17.96	-	-
age > 65 yrs	-0.3881	-13.29	-0.0723	-5.08
20K < income ≤ 60K	0.1469	6.93	-	-
income > 60K	0.3443	16.10	-	-
60K < income ≤80K	-	-	0.0782	5.69
income > 80K	-	-	0.1478	11.19
Number of Observations	7337		7963	
Log-likelihood at zero	-14277		-15495	
Log-likelihood at convergence	-10613		-6962	

[1]The unit of the composite impedance variable is equivalent highway in-vehicle travel time (in minutes).

[2]The coefficient on this variable is constrained to one in the home-based work model.

[3]The coefficient on this variable is constrained to one in the home-based shopping model.

parameter indicates that zones with high total employment are more likely to be chosen as the attraction-end (relative to zones with low total employment). The coefficient on the size variable represents an inclusive value characterizing the presence of common unobserved zonal attributes affecting the utility of elemental alternatives within a zone (see section 3.2). The parameter is close to zero, indicating that there are several unobserved zonal attributes which have a common effect on elemental attraction-ends within the zone. The parameter on the Ring 4 dummy variable shows a lower attraction-end choice utility for zones in Ring 4 relative to zones in other rings.

The empirical results indicate significant socio-demographic interactions with the composite impedance variable (the socio-demographic interactions with Ring 4 and unemployment rate were not statistically significant). The sensitivity to impedance is higher for individuals in the "56-65 years" age group and the "greater than 65 years" age group relative to individuals younger than 55 years. Between the "56-65 years" age group and the "greater than 65 years" age group, the sensitivity to impedance is higher for the latter group. There was no substantial difference in sensitivity among individuals below 55 years. The high sensitivity to impedance of individuals over 55 years might be a reflection of mobility and physical dexterity challenges that such individuals experience. The higher sensitivity of women and lower sensitivity of high income earners to impedance is consistent with the findings from earlier studies.

5.2.2 Home-Based Shopping Attraction-End Choice

The column titled "Home-Based Shopping Purpose" in Table 1 presents the estimation results for the home-based shopping purpose. The variables in the final specification include the impedance variable, two zonal size measures (retail plus service employment and zonal area), the Ring 4 dummy variable, the zonal spatial structure measure, and socio-demographic interactions with the composite impedance measure. As for the home-based work purpose, there were no substantial differences in the utility of zones among rings 0, 1, 2 and 3, but there were significant differences based on whether a zone was in ring 4 or not. There was collinearity between the crime rate variable and the ring 4 dummy variable. Including both variables resulted in a non-intuitive sign on the crime rate variable and only a marginal increase in the log-likelihood value over the specification including the ring 4 variable alone. So, crime rate was dropped from the specification.

The sign on the composite impedance measure in Table 1 is as expected. A comparison of the impedance coefficient between the shopping and work purposes indicates that there is much greater sensitivity to travel impedance for shopping relative to work. The coefficient on the composite size variable for the shopping purpose is significantly different from one, indicating that there are unobserved zonal attributes affecting the utility of elemental attraction-ends within the zone. Among the size variables characterizing the composite size measure, the parameter on zonal area is larger than that on the sum of retail and service employment (the coefficient on retail plus service employment is constrained to one for identification). This result indicates that 1 square mile of zonal area is equivalent to about 5.35 units of retail plus service employment in terms of zonal size representation. The mean value of retail plus service employment in the sample is 1076 and that of zonal area is 2.62 square miles. Thus, effectively zonal area (in square miles) contributes substantially less to the composite size measure than does the retail plus service employment measure. The sign on the Ring 4 dummy variable indicates that the utility of a zone not in Ring 4 is greater than a zone in Ring 4. The spatial structure measure is highly significant in its effect on attraction-end choice. The negative parameter on this measure reflects the presence of "competition" forces; *i.e.*, zones in close proximity to other shopping opportunities have a lower utility than zones in spatial isolation.

The socio-demographic interactions with composite impedance are statistically significant (the interactions of crime rate and Ring 4 with socio-demographics turned out to be statistically insignificant). Differences in sensitivity among different age groups below 65 years of age were absent, but individuals above 65 years are more sensitive to impedance than their younger counterparts. Women are more sensitive to impedance than men and individuals in the higher income brackets are less sensitive to impedance than individuals in lower income brackets.

5.2.3 Sensitivity of Estimation Results to the Number of Randomly Selected Alternatives

The results presented in the previous two sections are based on the random selection of six attraction-end zonal alternatives for each production zone in addition to the actual chosen attraction-end alternative (a total of seven alternatives in the choice set). To examine the sensitivity of the estimation results to the number of alternatives in the choice set, these models were estimated with a total of 11 and 16 alternatives in the choice set. The results with 7, 11, and 16 alternatives are presented in Table 2 for the home-based work purpose and in Table 3

Table 2: Sensitivity of Home-Based Work Estimation Results to the Number of Attraction-End Alternatives

Variable	7 alternatives		11 alternatives		16 alternatives	
	Parameter	Std. error	Parameter	Std. error	Parameter	Std. error
Log of composite impedance	-1.3136	0.026	-1.3647	0.024	-1.3765	0.023
Log of Composite zonal size measure	0.0485	0.007	0.0665	0.006	0.0647	0.006
Total zonal employment	1.0000	-	1.0000	-	1.0000	-
Ring 4 dummy variable	-0.7967	0.055	-0.7856	0.052	-0.7346	0.051
Socio-demographic interactions with impedance						
Female	-0.2410	0.008	-0.2507	0.008	-0.2533	0.007
55 yrs < age < 65 yrs	-0.2623	0.015	-0.2983	0.014	-0.3152	0.013
age > 65 yrs	-0.3881	0.029	-0.1790	0.024	-0.2471	0.024
20K < income < 60K	0.1469	0.021	0.1829	0.020	0.1550	0.020
income > 60K	0.3443	0.021	0.3574	0.020	0.3370	0.020

Table 3: Sensitivity of Home-Based Shop Estimation Results to the Number of Attraction-End Alternatives

Variable	7 alternatives		11 alternatives	
	Parameter	Std. error	Parameter	Std. error
Log of composite impedance	-1.9366	-86.62	-1.9692	0.0195
Log of Composite zonal size measure	0.1360	6.67	0.1307	0.0185
Zonal land area (in square miles)	5.18	9.35	5.0074	0.5465
Ring 4 dummy variable	-0.9395	-14.08	-0.9365	0.0622
Spatial structure variable	-4.3618	-48.00	-4.0685	0.0823
Socio-demographic interactions with impedance				
Female	-0.0299	-3.09	0.0383	4.168
Age > 65 yrs	-0.0723	-5.08	-0.0921	0.0135
60K < income < 80K	0.0782	5.69	0.0150	0.0130
income > 80K	0.1478	11.19	0.1262	0.0125

for the home-based shopping purpose. The parameter estimates are relatively stable in the three cases in both Tables 2 and 3, except for the coefficient on the "age > 65 years" variable in Table 2. As expected, the standard errors tend to decrease as the number of alternatives increase; surprisingly, however, the decrease in standard errors is very marginal. Overall, the results suggest little sensitivity in both parameter estimates and associated standard errors to the number of randomly selected attraction-end alternatives.

5.2.4 Evaluation of Fit

In this section, the fit of the disaggregate choice specifications in sections 5.2.1 and 5.2.2 are evaluated with the disaggregate equivalent of the conventional gravity model. The disaggregate equivalent of the conventional gravity model includes only the composite travel impedance measure and a single size measure (with the coefficient on the size measure constrained to one, see section 3.4). Thus, the choice specifications in sections 5.2.1 and 5.2.2 are more general than the disaggregate equivalent of the gravity model (in the rest of this section, the disaggregate equivalent of the gravity model is referred to as the "gravity model" for brevity). The fit of these models are examined on both an estimation sample (used in estimation) and a holdout sample (that is not used in estimation). The overall trip samples used in sections 5.2.1 and 5.2.2 were split into an estimation sample (about two-thirds of the trip sample) and a validation sample (one-third of the trip sample). As earlier, six additional candidate zonal alternatives were generated for each trip for a total of seven alternatives in the choice set.

A measure of fit of a model in the estimation sample is the $\bar{\rho}^2$ value (referred to as the adjusted likelihood ratio index or McFadden's adjusted R^2; see Windmeijer, 1995) defined as follows:

$$\bar{\rho}^2 = 1 - \frac{L(\hat{\beta}) - Q}{L(0)}, \tag{13}$$

where $L(\hat{\beta})$and $L(0)$ are the log-likelihood function values at convergence and at equal shares, respectively, and Q is the number of parameters estimated in the model. From a formal statistical fit standpoint, the choice specifications in sections 5.2.1 and 5.2.2 are compared to the gravity model using a nested likelihood ratio test.

In the validation sample, the choice specifications in sections 5.2.1 and 5.2.2 are compared to the gravity model using the predictive adjusted likelihood ratio index. This measure is computed by calculating the predictive log-likelihood function value at the parameter estimates obtained by maximizing the estimation likelihood function and then computing the corresponding predictive adjusted likelihood ratio index.

The measures of fit in the estimation sample are provided in Table 4. The disaggregate choice specification has a higher adjusted likelihood ratio index compared to the gravity model for both the work and shopping purposes in the estimation sample (retail plus service employment is used as the single size measure for the shopping purpose in the gravity model). The nested likelihood ratio index statistic for testing the choice specifications with the gravity model are 2247.58 and 2084.48 for the work and shop purposes, respectively. A comparison of these statistics with the chi-squared value with 7 degrees of freedom for the work purpose and 8 degrees of freedom for the shop purpose indicates that one can reject the gravity model at any reasonable level of significance.

Table 4: Measures of Fit in Estimation Sample

Summary Statistic	Home-Based Work Purpose		Home-Based Shop Purpose	
	Dissaggregate choice specification	Disaggregate equivalent of gravity model	Dissaggregate choice specification	Disaggregate equivalent of gravity model
Log-likelihood at zero	-9517.54	-9517.54	-10330.94	-10330.94
Log-likelihood at convergence	-7064.33	-8188.12	-4618.32	-5691.22
Number of parameters	8	1	9	1
Number of Observations	4891	4891	5309	5309
Adjusted likelihood ratio index	0.2569	0.1396	0.5521	0.4490
Nested likelihood ratio test statistic	2247.58		2145.80	

Table 5: Measures of Fit in Validation Sample

Summary Statistic	Home-Based Work Purpose		Home-Based Shop Purpose	
	Disaggregate choice specification	Disaggregate equivalent of gravity model	Disaggregate choice specification	Disaggregate equivalent of gravity model
Log-likelihood at zero	-4759.75	-4759.75	-5164.50	-5164.50
Log-likelihood at convergence	-2536.40	-2382.00	-2344.92	-2889.44
Number of parameters	8	1	9	1
Number of observations	2446	2446	2654	2654
Predictive adjusted likelihood ratio index	0.2526	0.1382	0.5442	0.4403

Table 5 provides the results of the validation exercise. The predictive adjusted likelihood ratio index is presented in the final row of the table and indicates again that the disaggregate choice model clearly outperforms the disaggregate equivalent of the gravity model. This confirms that the results obtained from the estimation sample are not an artifact of overfitting and are in fact stable. Another interesting point that emerges from Tables 4 and 5 is that the home-based shopping models perform substantially better than their counterparts for the home-based work purpose.

5.3 Procedure to Apply Choice Model to Obtain Trip Interchanges

The disaggregate choice model results presented in Table 1 can be applied in a straightforward way to obtain aggregate trip interchanges. To illustrate this procedure, define V_{ijs} as the utility presented by attraction-end zone j for a trip production from zone i made by an individual in socio-demographic group s. Let H_{ij} be the composite travel impedance from zone i to zone j, D_j be the composite size measure for zone j, and z_j be a vector comprising the Ring 4 dummy variable for the work purpose and the Ring 4 as well as the spatial structure measure for the shopping purpose. V_{ij} can then be written as:

$$V_{ijs} = \alpha_s \ln H_{ij} + \eta \ln D_j + \mu' z_j . \tag{14}$$

The coefficient on the impedance variable is subscripted by s since the impedance coefficient is a function of socio-demographics. However, the coefficients on the size measure and the z_j vector are independent of socio-demographics, as obtained in our estimation results. The probability that zone j is selected as the attraction-end for the trip from production zone i made by an individual in a socio-demographic group s can be written as:

$$P_{ijs} = \frac{e^{V_{ijs}}}{\sum_k e^{V_{iks}}} = \frac{H_{ij}^{\alpha_s} D_j^{\eta} e^{\mu' z_j}}{\sum_k H_{ik}^{\alpha_s} D_k^{\eta} e^{\mu' z_k}}, \tag{15}$$

where k is an index of all feasible attraction-end zones for production zone i. The above probability is constant across all productions from zone i made by individuals in demographic group s. Thus, if O_{is} is the total number of trip productions from zone i made by individuals in socio-demographic group s, the number of these trips attracted to zone j is given by:

$$T_{ijs} = O_{is} P_{ijs} = O_{is} A_{is} H_{ij}^{\alpha_s} D_j^{\eta} e^{\mu' z_j}, \text{ where } A_{is} = \left(\frac{1}{\sum_k H_{ik}^{\alpha_s} D_k^{\eta} e^{\mu' z_k}} \right). \tag{16}$$

The total interchange from zone i to zone j can then be computed by summing the above expression over all socio-demographic groups s as:

$$T_{ij} = \sum_s O_{is} A_{is} H_{ij}^{\alpha_s} D_j^{\eta} e^{\mu' z_j}. \tag{17}$$

The expression in equation (17) requires trip productions by socio-demographic group. There are 18 distinct socio-demographic groups for the work purpose and 12 distinct socio-demographic groups for the shopping purpose. These groups and the corresponding impedance parameters (α_s estimates) are presented in Table 6.

Table 6: Socio-demographic Groups and Corresponding Impedance Parameters

Home-Based Work Purpose		Home-Based Shopping Purpose	
Socio-demographic group	Imped. parm.	Socio-demographic group	Imped. parm.
Female		Female	
Age ≤55, income ≤20K	-1.5546	Age ≤65, income ≤60K	-1.9665
55 <Age ≤65, income ≤20K	-1.8169	Age >65, income ≤ 60K	-2.0388
Age >65, income ≤20K	-1.9427	Age ≤65, 60K <income ≤80K	-1.8883
Age ≤55, 20K ≤income ≤60K	-1.4077	Age >65, 60K ≤income ≤80K	-1.9606
55 <Age ≤65, 20K ≤income ≤60K	-1.6700	Age ≤65, income >80K	-1.8187
Age >65, 20K ≤income ≤60K	-1.7958	Age >65, income >80K	-1.8910
Age ≤55, income >60K	-1.2103	Male	
55 <Age ≤65, income >60K	-1.4726	Age ≤65, income ≤60K	-1.9366
Age >65, income >60K	-1.5984	Age >65, income ≤60K	-2.0089
Male		Age ≤65, 60K <income ≤80K	-1.8585
Age ≤55, income ≤20K	-1.3136	Age >65, 60K ≤income ≤80K	-1.9307
55 <Age ≤65, income ≤20K	-1.5546	Age ≤65, income >80K	-1.7888
Age >65, income ≤20K	-1.7017	Age >65, income >80K	-1.8611
Age ≤55, 20K ≤income ≤60K	-1.1667		
55 <Age ≤65, 20K ≤income ≤60K	-1.4077		
Age >65, 20K ≤income ≤60K	-1.5548		
Age ≤55, income >60K	-0.9693		
55 <Age ≤65, income >60K	-1.2103		
Age >65, income >60K	-1.3574		

CHAPTER 6

Summary and Conclusions

The past few decades has seen a shift from the aggregate (or zonal) level paradigm toward the disaggregate (or individual/household level) modeling paradigm in the context of the four-step transportation planning process. Much of the advances in disaggregate modeling has been confined to trip production and mode choice modeling. In contrast, the shift toward disaggregate procedures has yet to influence the fundamental specification of trip attraction-end and trip distribution models; in particular aggregate models continue to be used in most urban areas for these two components of the four-step transportation planning process. This research formulates and estimates (disaggregate) attraction-end choice models that will facilitate the replacement of the aggregate trip attraction models and the aggregate trip distribution models currently used by most metropolitan planning organizations.

The structure for the attraction-end choice models takes the form of a non-linear-in-parameters multinomial logit model. The non-linear-in parameters structure is necessitated because of the accommodation of multiple measures to represent the size of an attraction-end zone. A parallel conductance formula is used to develop a composite impedance measure for travel between each zone pair from the elemental impedance measures of in-vehicle time, out-of-vehicle time, and cost for each available mode serving the zone pair.

The attraction-end choice model associates the individual's attraction-end traffic zone choice for each trip production to relevant exogenous variables. Since the choice of attraction-end zone is characterized by a large number of alternatives, a random subset of the feasible attraction-end zones is chosen during the estimation of the choice model.

The research estimates attraction-end models for two trip purposes: home-based work and home-based shopping/personal business. Six sets of explanatory variables are considered for inclusion in the models. These are: a) impedance variables, b) zonal size measures, c) zonal attractiveness measures, d) zonal location indicators, e) a zonal spatial structure measure that accommodates the impact of the location pattern of zones, and f) interaction of socio-demographic variables with impedance and zone-associated variables. The sample used in the analysis was drawn from the 1991 Boston Household Travel Survey and from supplemental data sources.

The empirical results for the home-based work trip purpose indicate that the variables which significantly impact attraction-end choice include the travel impedance variable, total employment in the zone, a dummy variable for zones far away from Boston Central Business District, and the socio-demographic interactions with the impedance variable. The variables found to affect attraction-end choice for the home-based shopping purpose include the impedance variable, two zonal size measures (retail plus service employment and zonal area), a dummy variable for zones distant from the Boston Central Business District, a zonal spatial structure measure, and socio-demographic interactions with the composite impedance measure. A comparison of the impedance coefficient between the shopping and work purposes indicates that there is much greater sensitivity to travel impedance for shopping relative to work.

The relative performance of the disaggregate attraction-end choice model and the disaggregate equivalent of the gravity model were assessed using an estimation sample and a validation sample. The adjusted likelihood ratio index and the nested likelihood ratio tests were used as measures of fit for the estimation sample while in the validation sample the predictive adjusted likelihood ratio test was used. A consistent result that emerged from the different measures of fit was that the disaggregate attraction-end choice model outperforms the gravity model.

The disaggregate choice models estimated in this report require trip productions by several socio-demographic groups if they are to be used for predicting trip-interchanges among zone pairs. The home-based work attraction-end model requires trip productions by sex, age (in three categories: < 55 years; between 55 and 65 years; > 65 years) and income (in three categories: < 20K; between 20K and 60K; > 60K). The home-based shop attraction-end model requires trip productions by sex, age (in two categories: < 65 years; > 65 years) and income (in three categories: < 60K; between 60K and 80K; > 80K).

It is important to emphasize that two components of the choice models estimated in this report may be readily used within the context of a traditional aggregate gravity model if trip productions by socio-economic category are not available. First, the impedance formulation in this research can be applied to develop a composite impedance measure for use in the gravity model. Second, several zonal size measures can be used at the same time in the conventional gravity model through the generation of a composite size measure.

References

1. Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge.
2. Ben-Akiva, M., G.H. Gunn, and L. Silman (1984). "Disaggregate Trip Distribution Models", *Proceedings of the Japanese Society of Civil Engineers* (JSCE).
3. Blumen, O. (1994). "Gender Differences in The Journey To Work", *Urban Geography*, 15, 223-245.
4. Borgers, A. and H. Timmermans (1987). "A Context-Sensitive Model of Spatial Choice Behavior", *Behavioral Modeling in Geography and Planning*, Eds: Golledge, R. and H. Timmermans, Croom Helm, London, 159-178.
5. Chang, Y.B. and M. Plager (1994). "Interim Regional Travel Demand Forecasting Model: Development of the 1993 Version", Memorandum, June 17.
6. Choukroun, J.M (1975). "A General framework for the Development of Gravity-Type Distribution Models", *Regional Science and Urban Economics, 5*, 177-202.
7. COMSIS Corporation (1988). "Models of Mode and Occupancy Choice in the Shirley Highway Corridor", July 7.
8. Daly, A. (1982). "Estimating Choice Models Containing Attraction Variables", *Transportation Research*, 16B, 5-15.
9. Department of Employment and Training (1992) Employment review.
10. Domencich, T.A and D. McFadden (1975). *Urban Travel Demand: A Behavioral Analysis*, North Holland Publishing Company, Amsterdam

11. Dubin, R. (1991). "Commuting Patterns and Firm Decentralization", *Land Economics*, 67, 15-29.

12. Fotheringham, A.S. (1983). "Some Theoretical Aspects of Destination Choice and their Relevance To Production-Constrained Gravity Models", *Environment and Planning*, 15A, 1121-1132.

13. Fotheringham, A.S. (1986). "Modeling Hierarchical Destination Choice", *Environment and Planning*, 18A, 401-418.

14. Guensler, R. (1993). "Critical Research Needs in Land Use, Transportation and Air Quality", *Transportation Research Circular*, 389, No. 1.

15. Hansen, S. and G. Pratt (1990). "Geographic Perspective on the Occupational Segregation Of Women", *National Geographic Research*, 6, 376-378.

16. Hodge, D.C. (1990). "Geography and the Political Economy of Urban Transportation", *Urban Geography*, 11, 87-100.

17. Jones P.M. (1978). "Destination Choice and Travel Attributes", *Determinants of Travel Choice*, Eds: D.A. Hensher and Q. Dalvi, Saxon House, England.

18. Kelly, G. A. (1955). *The Psychology of Personal Constructs*. New York: W.W. Norton

19. Kitamura, R. (1984). "Incorporating Trip Chaining into Analysis of Destination Choice", *Transportation Research*, 18B, 67-81.

20. Koppelman, F.S and J.R. Hauser (1978). "Destination Choice Behavior for Non-Grocery Shopping Trips", *Transportation Research Record*, 673, 157-165.

21. Lo, L. (1991). "Substitutability, spatial structure, and spatial interaction", *Geographical Analysis*, 21, 132-146.

22. Madden, F.J. (1981). "Why women work closer to home", *Urban Studies*, 18, 181-194.

23. Massachusetts State Police Crime reporting Unit (1992) Crime in Massachusetts.

24. McClennen, J., K. Quackenbush and J. Gallagher (1994). "Chronology of FYA Transfer Penalties Research", Memorandum, October 20.

25. McFadden, D. (1978). "Modeling the Choice of residential Location", *Spatial Interaction Theory and Planning Models*, Eds: Karlquist, A., L. Lundquist, F. Snickbars and J.W. Weibull, North Holland: Amsterdam, 75-96.

26. Metropolitan Service District (1989). "Travel Forecasting Methodology Report", Westside Light Rail Project, September 29.

27. Oi, W. and P. Shuldiner (1962). *An Analysis of Urban Travel Demand,* Northwestern University Press, Evanston, Illinois.

28. Recker , W.W. and L.P. Kostyniuk (1978). " Factors Influencing Destination Choice for the Urban Grocery Shopping Trip", *Transportation,* 7, 19-33.

29. Timmermans *et al* (1982). "The Identification of Factors Influencing Destination Choice: An Application of the Repertory Grid Methodology", *Transportation,* 11, 189-203.

30. Windmeijer, F.A.G. (1995). "Goodness-of-Fit Measures in Binary Choice Models", *Econometric Reviews*, 14, 101-116.