Dissertations   Theses and Dissertations

Fall 2003

Intercity travel demand : a utility-consistent simultaneous trip generation and mode choice model

Guilin Li
New Jersey Institute of Technology

ABSTRACT

INTERCITY TRAVEL DEMAND:
A UTILITY-CONSISTENT SIMULTANEOUS
TRIP GENERATION AND MODE CHOICE MODEL

by
Guilin Li

An intercity travel decision includes a complex set of subdecisions, such as when to travel, where to travel, which mode to choose, and others. The main focus of this dissertation is to examine trip frequency and mode choice of intercity non-business travel.

The objective of this study is to understand intercity travel behavior using disaggregate models. The proposed conceptual framework for intercity travel behavior leads to a nested logit/continuous choice model that is rigorously linked to the utility maximization theory. Compared to a traditional intercity travel demand model, the proposed model is utility consistent in that trip generation and mode choice models flow from one utility function. Thus, the resultant model embodies the interrelationship of trip generation and mode choice.

Applying the model to the NorthEast Corridor, the calibrated results show that trip generation of non-business travelers is interdependent with mode choice. The factors influencing mode choice may exert an impact on trip generation directly or indirectly.

INTERCITY TRAVEL DEMAND:
A UTILITY-CONSISTENT SIMULTANEOUS
TRIP GENERATION AND MODE CHOICE MODEL

by
Guilin Li

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation

Interdisciplinary Program in Transportation

January 2004

INTERCITY TRAVEL DEMAND:
A UTILITY-CONSISTENT SIMULTANEOUS
TRIP GENERATION AND MODE CHOICE MODEL

Guilin Li

Dr. Rongfang (Rachel) Liu, Dissertation Advisor   Date
Assistant Professor of Civil and Environmental Engineering, NJIT

Dr. Athanassios K. Bladikas, Committee Member   Date
Associate Professor and Chair of Industrial and Manufacturing Engineering,
and Director of Interdisciplinary Program in Transportation, NJIT

Dr. Lazar N. Spasovic, Committee Member   Date
Professor of Management
and Director of International Intermodal Transportation Center, NJIT

Dr. Janice R. Daniel, Committee Member   Date
Assistant Professor of Civil and Environmental Engineering, NJIT

Dr. Xuehao Chu, Committee Member   Date
Economist, Senior Research Associate, Center for Urban Transportation Center, USF

BIOGRAPHICAL SKETCH

Author: Guilin Li

Degree: Doctor of Philosophy in Transportation

Date: January 2004

Undergraduate and Graduate Education:

- Doctor of Philosophy in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 2004

- Master of Science in Transportation,
  New Jersey Institute of Technology, Newark, NJ, 2001

- Master of Science in Traffic Engineering,
  Tongji University, Shanghai, P. R. China, 1999

- Bachelor of Science in Traffic Engineering,
  Tongji University, Shanghai, P. R. China, 1996

Major: Transportation Engineering

Presentations and Publications:

Liu, R., and Li, G..
"The Subjective Values of Safety and Security in the Context of Intercity Travel".
Proceedings of the 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.

Liu, R., and Li, G..
"Dynamic Travel Behavior Analysis Based on Stochastic Decision-Making Styles".
Presented at the 10th International Conference on Travel Behavior Research, Lucerne, Switzerland, August 10-15, 2003.

Liu, R., and Li, G..
"Intercity Travel Choices Affected by Safety and Security Measures".
Presented to TRB Committee A1C06, Social and Economic Factors in Transportation at the 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., January 2003.

Li, G., Wu, J., and Zhou, S..
    "OD Estimation Study of Elevated Roads".
    Presented at the 7th Cross Strait Conference on Urban Transportation, Xi-An, China, August
    23-26, 1999.

Li, G. , Lin, H. and Zhou, S..
    "Operation Analysis of the South-North Elevated Beltway in Shanghai".
    Traffic and Transportation. 1999, Vol. 1.

To

Shangwu Zhou, a professor in Tongji University, Shanghai, P.R. China

Ying Xiao and Jason Li

And my beloved family

ACKNOWLEDGMENT

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Congestion in intercity corridors has been increasing steadily, which has raised serious concerns about its adverse impacts on regional economic development, national productivity and competitiveness, and environmental quality (Bhat, 1995). To alleviate congestion, many major investment projects, such as high-speed and Maglev rail projects were proposed by different agencies. To evaluate the adequacy and efficiency of these projects, public agencies need analyses of the intercity travel demand due to the limited allocated financial resources. The quality of decisions of project selection is impacted by the accuracy of the travel demand prediction and the sensitivity of this demand to travel cost and enhancement of levels of service. Meanwhile, intercity passenger carriers welcome reliable forecasts of intercity demand so that they can be more responsive to their patronage and to remain competitive. Therefore, intercity travel behavior research is needed to estimate and evaluate expected policy impacts prior to implementation.

The following section introduces the problem statement. Section 1.2 presents the method and objective of this research. Section 1.3 defines the research scope. The dissertation organization is included in Section 1.4.

1.1 Problem Statement

Intercity travel behavior analysis can be used for demand forecasting, service pricing and improvement impact studies. Travel behavior research, however, is far from complete in the intercity travel market, especially when compared to urban travel analysis. Research

is needed to understand the travelers' preferences and willingness to choose among many existing or potential alternatives, such as intercity bus, automobile, conventional rail, high-speed rail, Maglev, and airplane.

On the basis of data used, the travel demand models can essentially be categorized into two types: aggregate models and disaggregate models. Aggregate models are based on zonal-level data. Disaggregate models make use of individual travelers' data, which reflect the causal mechanism establishing the behavioral link between the context of travel, the attributes of the transportation system, the individual's characteristics and the travel decisions. Using individual traveler behavior, this disaggregate mechanism enables the model to correctly forecast total intercity travel demand and its distribution among modes for the evaluation of large capital investments and the impact of policy changes. Therefore, many researchers consider disaggregate methods to be more promising and more reasonable than the aggregate approach (McFadden and Reid, 1975). This leads us to the investigation of behavioral models of individual travel demand.

As in classical urban transportation demand analysis, disaggregate travel behavior analysis separates the travelers' decisions into four subdecisions: trip generation, mode choice, destination choice and route choice. Since the 1970's, there has been voluminous studies of disaggregate travel behavior. The majority of these studies focus on the urban context. Some of the limited research in the intercity travel demand field (Georggi and Pendyala, 1999; O'Neill and Brown, 1999) touches on trip generation, in which the classical categorical analysis or regression models were employed. However, most, if not all, of the intercity travel demand studies concentrate on mode choices. In these models, the trip generation is presumed to be independent of mode choice. However,

mode choice and trip generation decisions are made by the same individual, which means he\she has the same causal mechanism for all his/her subdecisions. The current models are incapable of reflecting the interrelationship between trip generation and mode choice and of modeling them with the same mechanism.

1.2 Research Objective and Methods

Based on their needs and environment, travelers face a complex set of subdecisions within intercity travel. These subdecisions pertain to travel purpose, frequency, timing, destination, and mode of travel. Numerous studies presumed that the frequency, timing, and destination of trips are predetermined (Watson, 1973; Watson, 1974; Stopher and Prashker, 1976; Grayson, 1981; Stephaned, et al., 1984; Banai-Kashani, 1984; Lyles, et al., 1990; Wilson, et al., 1990; Abdelwahab et al., 1992; Forinash and Koppelman, 1993; Aljarad, 1993; Bhat, 1995; Bhat, 1997; Vosgva, 1998; etc.). These studies concentrated on understanding mode choice. From the standpoint of the individual decision-maker, however, all of these subdecisions are interrelated rather than separated. From the researchers' point of view, the model should reflect this interrelatedness and one utility maximization objective for different subdecisions.

The primary objective of this research is to apply the same causal mechanism and utility maximization theory, to trip generation and model choice, and to include the interrelationship between the two subdecisions. Identifying and evaluating the factors that affect travel decisions is also part of this dissertation.

This research documents the derivation of the trip generation model and mode choice model from one underlying utility index. A general nested logit function is constructed for the mode choice component and a continuous regression model is derived for trip generation. These models complete the proposed nested logit/continuous model. This model only employs one set of individuals' travel choice survey data to overcome the inconsistency between the trip generation dataset and mode choice data used by previous analyses.

1.3 Research Scope

This research is limited to the analyses of non-business intercity travel behavior since the decision-maker of a business trip is not necessarily the traveler. The spatial range of the study is limited to a corridor context. As a result, the route choice subdecision is not included in this study. Also, the destination choice subdecision is excluded in this research. The reason is that the most important determinant for destination choice of non-business trips is whether the individual has friends or relatives in a city. The cities cannot be interchangeable in these choice situations (Sonesson, 2001). Therefore, as indicated in the title of this study, the subdecisions are confined to trip generation and mode choice.

This study presents a conceptual framework for non-business intercity travel analysis. Accordingly, a nested logit/continuous model is derived for modeling the conceptual framework. The study area is limited to a corridor from the Washington DC Metropolitan Statistical Area (MSA) to the New York MSA. Because of the constraints in dissertation time and cost, no new survey was performed. A subset of data for the

corridor was extracted from the 1995 American Travel Survey. It is supplemented with web-based and published data. This resultant dataset is used to test and develop the proposed models' structure.

1.4 Dissertation Organization

The rest of this dissertation is organized as follows: Chapter 2 reviews the existing literature on an intercity travel demand analysis. Specifically, several methods that are similar to the model proposed here are discussed in terms of their advantages and disadvantages. In Chapter 3, an intercity travel decision paradigm is presented within which the simplified model is developed. In Chapter 4, the nested logit/continuous model is proposed. Chapter 5 describes the data requirements and data preparation. In Chapter 6, the results of the application of the nested logit/continuous model for the NorthEast Corridor are presented. Chapter 7 discusses model validation and reports the validation results. A summary of the findings and conclusions of the study and recommendations for further research are given in Chapter 8.

CHAPTER 2

LITERATURE REVIEW

The intercity travel demand analysis-related research is reviewed in this chapter. It includes the basic intercity travel demand model structure, the progression of the intercity demand models, and typical models currently used. Finally, there is a review of the development and application of discrete/continuous models in different fields.

Intercity travel is the travel between cities or other points of interest that are separated by some significant distance. The transportation literature generally refers to long-distance travel as intercity travel. The term long-distance travel is defined as trips of a certain minimum distance. However, the thresholds for long distance travel in various countries are different. All these values are derived by population surveys. The thresholds can vary from 50 miles (UK) to 100 miles (USA) (Limtanakool, Dijst and Lanzendorf, 2003). Even within the USA, there exists a plethora of definitions (US DOC, undated). In this study, 100 miles is used to define intercity travel, regardless of any overnight stay.

2.1 Intercity Travel Demand Model Structure

Intercity travel behavior is different from urban travel behavior in certain aspects, such as travel frequency. However, it still follows the general four-step model for urban travel behavior: trip generation, trip distribution, mode choice, and trip assignment. Intercity travel decision making is typically assumed to consist of trip generation, destination choice, mode choice, and route choice. As mentioned in Section 1.3, this study will focus

6

on trip generation and mode choice. Therefore, the literature review focuses on these two modules.

2.1.1 Conventional Travel Demand Models

Conventional travel demand models separate the demand functions into four steps. When used for intercity travel, the model consists of two sequential steps that predict intercity travel by mode (Koppelman and Hirsh, 1984). The first step forecasts the total intercity travel volume for city pairs. The second step distributes the volume via a logit model. Typically, the number of trips is formulated as a function of the socioeconomic characteristics of city pairs and composite measures of the level of service. Today these models are still in applications such as forecasting high-speed rail ridership (Brand, Parody, Hsu, and Tierney, 1992). These models provide some insight into intercity travel behavior. However, the model obscures much of the information in the data. Its behavioral implication that individuals decide their travel behavior in stages does not appear to be consistent with reality. Hence, it has limitations as an estimator of intercity travel demand (Peers and Bevilacqua, 1976).

2.1.2 Trip Generation/Frequency Models

Trip generation is defined as the number of individual trips generated in a given period of time. Traditionally, in travel demand modeling, trip generation is the first component that provides the possibility for the next steps, such as destination choice and mode choice.

In the context of urban travel, a trip can be home-based or non-home-based. In practice, according to Ortuzar (1994), it is also classified by purpose, such as trips to work, trips to school or college, shopping trips, social and recreational trips and other trips. Alternatively, the trips can be classified by person type based on income level, car ownership, household size and structure, which is often used as the model segmentation base. In an intercity travel context, a trip is usually categorized as a business trip or non-business trip. It also can be further classified as business, combined business/pleasure, convention, conference or seminar, visiting relative or friends, rest or relaxation (the 1995 American Travel Survey).

Trip generation analysis requires identification of the factors that affect trip generation. Often, the variables taken into account are characteristics of the traveler, and personal trip attraction (Ortuzar, 1994), as well as the attributes of alternatives. The characteristics of travelers include household income, car ownership, household structure, and household size. The personal trip attraction factors include the destination's socio-economic, industrial, or residential context. The performance of the available alternatives may influence the number of trips, particularly for intercity non-business trips. For instance, reduction in travel time or cost may induce more frequent trips. The opposite may reduce trip frequency or consolidate trips. The effects of changes in modal attributes on travel frequencies may also exert effects on choice of mode for some trip purposes as stated by Domencich and McFadden (1995).

The early trip generation models, based on aggregate data, predicted total trips between city pairs (ITE, 1992). The modeling methods generally include regression models, cross-classification analysis, or a combination of both. These methods still have

applications due to their mathematical feasibility, data availability, and ease of interpretation (USDOT, 1999).

Disaggregate trip generation models were developed to be consistent with other components of the transportation demand modeling system, such as mode choice models and destination choice models. These models are based on individual level data. Trip generation was assumed to be a process of choosing one option from the following alternatives: making no trip, one trip, two trips and so on. Therefore, these models predict the probability of an individual making a certain number of trips within a given period. The detailed progression of the disaggregate trip generation models is presented below.

As early as 1970, Stopher and Lisco suggested the logit model for trip generation. Subsequently, Talvitie (1973) proposed a disaggregate trip generation model, at only a conceptual level. Charles River Associates (CRA) (Domencich and McFadden, 1995) developed a binary logit model to determine the probability of an individual undertaking a shopping trip on a given day. In this model, the inclusive price, which is defined as a weighted average price over all possible destinations, was used as an independent variable. Another attempt to improve the trip generation model was carried out by CRA (Tye et al., 1982) which provided no significant improvement since it simply utilized different independent variables, such as the number of licensed drivers minus number of workers in the household. However, it was stated that application of the logit model to this choice set would create a problem due to the logit model's basic assumption of Independence of Irrelevant Alternative (IIA). This assumption states that the relative probability of each pair of alternatives is only related to the characteristics of this pair of

alternates and independent of the presence or characteristics of all other alternatives. Lastly, the authors concluded that the "IIA property of the MNL appears to be a fatal defect in its use for trip generation".

The use of choice theories for improving the explanatory power of trip generation models was tested by Tardiff (1977). A linear regression model and two variations of logit models were compared. One of the models was a binary logit model, in which one option is larger-than-thirteen trips per 28 days; another option is less-than thirteen trips. Another logit model examined was the linearized form. It forecasts the share of households that undertake a certain number of trips. The calibration results showed that the linear regression model was better in performance than the logit models. However, the regression model is not based on choice theory but rather on the hypothesis that those households with larger sizes, and/or more resources to travel, will make more trips. A trip generation model consistent with choice theory and similar to the mode choice model would be more plausible.

Sheffi (1979) developed an ordered logit trip generation model. The choice set of alternatives is trip frequencies of zero, one trip, two trips, and so on. These choices are known as rank ordered, or nested choices. This model performed well in predicting behavior. In this model, a particular alternative, i.e. two trips, implies that all lower ranked alternatives (of making the first trip) have to be chosen. It is assumed that decision-makers choose the options step by step. Another property of this model is that binary logit is used, which means that it does not have the IIA problem. For estimation, there are two types of approaches. One of them is step-by-step, which applies the binary logit estimation at least "n-1" times if the largest number of trips is "n". The other one is

the simultaneous maximization likelihood method. This estimation method combines all the steps into one likelihood function. The cost of estimation by a step-by-step approach could be high as a result of the need to estimate binary logit models "n-1" times. The estimated coefficients for the same variables in different binary logit steps may not be the same. This is not consistent with travel behavior. Thus, the simultaneous estimation method is favored.

Vickerman and Barmby (1985) employed this ordered logit model to estimate shopping trip frequency in England. The results indicated that the model was favorable with regard to the applicability of this model. The most recent application was Damodaran (1988) who combined this model with mode choice. This provides a combined logit model for intercity travel behavior analysis, which is included in Section 2.3.2.

2.1.3 Mode Choice Models

Mode choice analysis determines the number of personal trips between each origin-destination pair being made by each available mode of travel. The bulk of endeavors within intercity travel demand modeling has centered around the question of mode choice.

As early as 1973, Watson developed the first disaggregate model for intercity mode choice, in which he discussed two alternatives (rail versus auto) using the information regarding individual travelers on the Edinburgh-Glasgow route in Scotland. It was concluded that the use of disaggregate, behavioral, stochastic models in a predictive framework is preferable to the aggregate approach because the predictions of disaggregate models are extremely promising. Subsequently, there were a large number

of studies on disaggregate mode choice within the intercity context (Grayson, 1981; Banai-Kashani, 1984; Wilson, et al., 1990; Lyles and Mallick, 1990; Abdelwahab et al., 1992; Forinash and Koppelman, 1993; Bhat, 1997; Mehndiratta Hansen, 1997; and Vovsha, 1998). The studies contain probabilistic models which only focus on making a specific choice, once the traveler has decided to make a trip. Studies progressed from a binary logit model, to a multinomial logit model, and to the nested logit model.

The first model developed was a binary logit model which was employed within the New York City-Buffalo corridor in the U.S. In this model, (n-1) binary models should be calibrated if n modes are to be evaluated. Later, the concept of a binary logit model was extended to a multinomial logit model (Stopher and Prashker, 1976), which was used to model the choice of one among a set of mutually exclusive alternatives. Its simple and elegant mathematical structure makes it easy to estimate and interpret. Most importantly, it is consistent with utility maximization theory. However, the IIA problem of this model drove the researchers to relax the assumption.

Models with non-identical or non-independent random components commonly use different distributions to model the error terms (Bhat, 1995). The normal distribution for non-identical, non-independent error terms results in a multinomial probit model. The type I extreme value distribution to model the identical, non-independent random components leads to a nested logit model. First derived by Ben-Akiva (1973), the nested logit model, is designed to capture correlations among alternatives. Recently, the nested logit model has become widely popular. This model relaxes the IIA assumptions across alternatives. It allows interdependence within a nest.

Recently, some researchers are developing a Mixed Logit model. The model contains both probit-like disturbances and an additive Independently and Identically Distributed (i.i.d.) extreme value (or Gumbel) disturbance within the multinomial logit model. It is a highly flexible model that can approximate any random utility model. Mixed logit is considered to be the most promising state-of-the-art discrete choice model. The literature on this model includes Hensher and Greene 2001, Walker 2002, and Train 2002.

2.2 Early Intercity Travel Demand Models

2.2.1 Origin of the Intercity Travel Demand Model

As early as 1961, Lansing et al. applied simple gravity models for New York and Chicago. The initial gravity model described the relationship between the total traffic between each of these two cities and the demographic and socioeconomic characteristics of the city pairs. In this model, only population, per capita income, and distance were included as independent variables. In 1969, Quandt and Young improved the initial model. Later, it was employed in the Northeast Corridor Project to forecast the ridership on potential and existing modes of intercity travel along the Washington DC - New York - Boston corridor (U.S. DOT, 1970).

2.2.2 Direct Demand Models

Another type of early intercity travel demand model was the direct demand model. This model combined trip generation, trip distribution between cities, and modal choice in a single demand equation. The data used for these models are observational data on geographic aggregates.

Most of the direct travel demand models were developed in connection with the Northeast Corridor Project. Of most interest are the Kraft-SARC model (Kraft, 1963), the Quandt-Baumol abstract mode model (Quant and Baumol, 1966), and the Blackburn model (Blackburn, 1970). The number of observed round trips by purpose and by mode between zonal (or city) pairs is used in this class of models as the units of observation. Hence, it circumvents the trip distribution and separate modal split problems.

Meyer (1971) showed that the Kraft-SARC Model is, in its implicit form, an example of a direct and specific model. The Kraft-SARC Model is as follows:

$$T_{ijm} = \beta_{m0} \prod_{r}^{R} A_{ijr}^{\beta mr} \prod_{n}^{m} \prod_{s}^{S} C_{ijns}^{\alpha_{mns}}, (m = 1, 2, ..., M) \tag{2.1}$$

where:

T_{ijm} : Travel demand between i and j and by mode m,

A_{ijr} : Observations on the rth socioeconomics activity variable for (i, j)[th] city pair, and

C_{ijms} Observation on the sth generalized cost variable (including level of service or the impedance variable) for mode m from i to j.

Involved here are R socioeconomic activity variables for the city pair, and s generalized cost variables of mode m.

The Kraft-SARC model implies the presence of cross elasticity among modes. The model includes travel time and travel cost by all modes available in the city pair. Because of this, it is more applicable to intercity travel analysis than the Quandt-Baumol abstract mode model, which is formulated as follows:

$$\log RPM_i = a_0 + a_1 \log X + a_2 \log F + a_3 \log(F_i / F) + a_4 \log S + a_5 \log(S_i / S)$$
$$+a_6 \log D + a_7 \log(D_i / D) + E. \tag{2.2}$$

where:

RPM_i: The number of revenue passenger miles carried on mode i,

X: The number of characteristics of the city pairs that influence demand but of no concern here,

F: The mode with lowest fare,

F_i / F: The relative fare of mode i,

S: The speed of the fastest mode,

$S_i / S)$: The relative speed of mode i,

D: The departure schedule of the most frequent mode, and

D_i / D: The relative schedule of mode i.

Here, the Revenue Passenger Miles (RPM) are related to characteristics of the city pairs, the lowest fare of all modes, the speed of the fastest mode, and the departure schedule of the most frequent mode. In the Quandt-Baumol abstract mode model, the travel characteristics of the best mode, in terms of the lowest fare, the highest speed and the most frequent schedule, determine the volume of travel between city pairs. A more detailed review of the model can be found in Lave (1972).

In theory, the method used in this dissertation most closely resembles the Blackburn (1970) model, which specifies the expected value of the traffic volume after aggregation as follows:

$$E\{X_{ijk}\} = P_i E\left\{\alpha_1 P_j^{\alpha_2} Y_j^{\alpha_3} e^{w_0 - \alpha_4 C_{ijk}} \left| C_{ijk} \le C_{ijn} \right.\right\} \times \Pr\left\{C_{ijk} \le C_{ijn}\right\} \qquad n = 2, 3, \ldots m + 1 \tag{2.3}$$

$$C_{ijk} = p_{ijk} + \exp\{w_1\}t_{ijk} + w_k \qquad k = 2,...2m+1 \qquad (2.4)$$

where:

X_{ijk} : Individual's demand for trips from origin i to destination j by the kth mode,

P_i : Population of i,

C_{ijk} : Trip cost from i to j by the kth mode,

Y_j ; Per capita income at j,

p_{ijk} Cost of trip from i to j by k[th] mode,

t_{ijk} : Journey time between i and k on the kth mode, and

w_i : A random variable with joint multivariate normal probability distribution.

Blackburn proposed an individual choice model for determining alternative modes of traveling and alternative number of trips. The individual can have any number of trips within the study periods. The model was derived from utility maximization theory, which is also the foundation of this dissertation. The model aggregates demand over individual demand functions to get an aggregate model of passenger demand. It is assumed that modes, characterized in terms of inclusive prices, are perfect substitutes so that the individual always selects the cheapest mode. Most of Blackburn's analytical efforts are then devoted to aggregate demand from the demand of individuals.

An additional model is the multistage sequential model developed by McLynn and Woronka (1969). This model involves two stages: a combined trip generation-distribution stage and a mode split stage. Rice et al. (1981) have shown mathematically that there was no difference between the direct and two-stage models and that one could be derived from the other.

2.3 Simultaneous Travel Demand Models

For some time, the sequential four-step model has been employed. Mainly, this is due to the fact that it is easy to implement in practice. However, the sequential method is not consistent with the individual's decision-making process, since decisions of whether to make a trip, the destination, and the travel mode are seldom undertaken by an individual in stages. Therefore, a simultaneous model is becoming more widely used.

Early simultaneous models were either aggregate or disaggregate. Basically, the aggregate simultaneous model incorporates feedback into the four-step travel forecasting procedure. It results in a combination of trip distribution, modal split, and traffic assignment (Boyce and Zhang, 1997; Tatineni, 1992). The other aggregate simultaneous model is an equilibration procedure which provides a simultaneous solution to the trip generation, distribution, modal split, and assignment problems (Safwat and Magnanti, 1988). This modeling approach was applied to intercity transportation planning in Egypt (Moavenzaden, Markow, Brademeyer and Safwat, 1983; Safwat, 1987). In these studies, it was concluded that the approach was able to predict rational behavioral responses of users to policy specifications. Since the disaggregate approach is adopted in this study, the following review is confined to the disaggregate model.

The first disaggregate simultaneous model estimates the combined probability of destination and mode choices within a joint logit model (Ben-Akiva, 1974). This model can estimate the joint probability that an individual will take a trip to a certain destination by a particular mode. Under such a structure, the choice set contains all combinations of destinations and modes that are feasible for an individual. The multinomial logit model

(MNL) was used to estimate the choice probability. However, the joint model may violate the IIA assumption.

This violation can be avoided by the use of a nested logit model. This model was first proposed by Ben-Akiva (1974) and has been applied subsequently by others, such as Sobel (1980). The simultaneous mode and destination choice model by Ben-Akiva derived the combined probability by assuming that the distribution of error terms is type I extreme value. In this nested logit model, it is easy for the choice set to contain all feasible combinations of destinations and modes. One estimation method for this model is full information estimation. This leads to the simultaneous model between the mode choice and destination choice. The probability also can be written as a product of conditional and marginal probabilities that represent different decision sequences. However, the parameters were sensitive to the structure used. It was suggested that the simultaneous structure is preferred since there were a priori reasons to accept it.

Later in 1976, Adler and Ben-Akiva extended the nested logit model to include trip generation in addition to destination and mode choice. However, the trip frequency was restricted to zero or one. It did not address the problem of a large choice set resulting from all possible combinations of trip frequency, mode, and destination.

The following are some recent models that are similar to this study.

2.3.1 Charles River Associates (CRA) Model for Urban Travel Demand

Domencich and McFadden (1996) developed a model for urban travel demand analysis that is close to the direct demand model. This study examined the number of directed round trips between any zonal pair for a given purpose and mode as a function of the number of individuals and socioeconomic factors, the appropriate measure of level of

activity, and other relevant characteristics, as well as the travel times and costs of alternatives. The individual components of the model are "interrelated" through the attributes of the trip - time and cost variables - so that the separate components link together in an overall demand model. The distilling of information on attributes across a broad, though discrete, set of choice alternatives into a single "index of desirability" or inclusive price of travel is the success of the nested logit model. Obviously, the model is not completely interrelated because the utility is assumed to be separable. Different components are estimated separately, although they share the same variables: inclusive cost. Also, it only considers the choice between no trip and one trip due to limitations in the study case. The model's permitted shopping trip frequency allows just one or zero shopping trips per household per day, which may be limited for many applications.

2.3.2 Damodaran Model for Intercity Travel

Damodaran (1988) proposed another conceptual framework for intercity travel behavior under which a nested logit model structure was developed to combine mode choice and trip generation. The structure is shown below in Figure 2.1.
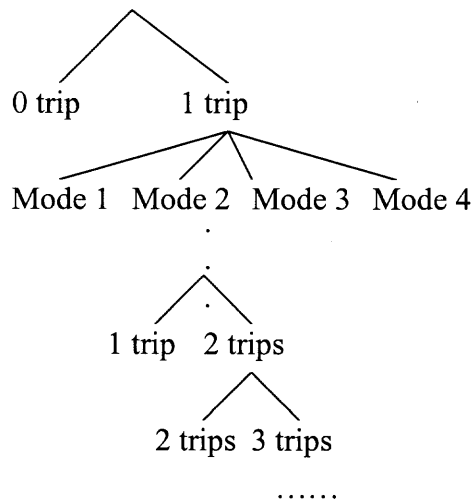
Figure 2.1 Structure of Damodaran model.

The models for trip generation and mode choice were linked through the "inclusive utility value", but it is only included in the first stage of the nested logit model for trip generation. In Damodaran's study, the trip generation model was based on a model proposed by Sheffi (1979) for nested or ordered alternatives. There were limitations when faced with many trips per study period. This model represented the linkage between conditional probabilities through a structure of decision-making. Although the decision-making is assumed to be sequential, it is possible to estimate the simultaneous structure through the simultaneous maximum likelihood method. The nests can be different subdecisions of choice or different dimensions within one subdecision, and the higher choice model incorporates a composite variable that represents the inclusive utility of the individual across the lower level, which is used by Ben-Akiva (1974), Sobel (1980), and McFadden (1978). The structure is similar to the nested trip frequency model, except that the mode and destination choices are also repeated for every trip. In this structure, each trip can be considered independent of the previous one. This simplification was made to avoid the replication of mode choice for every additional trip. So, one mode choice model is used for any number of trips. The estimation order is as follows. The mode choice is estimated first. A binary logit for a 0 or larger-than-1 trip model is estimated with an "inclusive utility" from the mode choice model. Another binary logit model for 1 trip and more-than-2 trips are estimated without any inclusive value. Many binary logit models are similarly estimated sequentially. Again, in this structure, the burdensome work results from the large amount of binary logit estimations when the trip frequency is high, although the mode choice model and trip generation are

both based on the utility maximization theory and they are combined through the "inclusive utility".

2.3.3 Multidimensional Model System for Intercity Travel Choice Behavior

Koppellman and Hirsh (1986) and Koppelman (1989) proposed a hierarchical structural system which includes trip frequency, trip destination, mode choice, and service class choice. The structure is shown in Figure 2.2.

Trip frequency choice

    0 trip   1 trip   2 trips   3 trips   ......

Trip destination choice

    City 1   City 2   City 3   ......

Mode choice

    Auto   Rail   Bus   Airplane

Service class choice

    Coach Metroliner
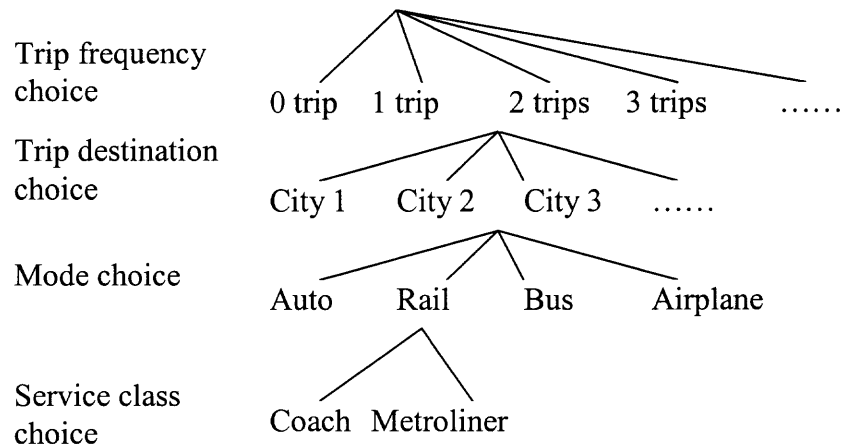
Figure 2.2 Koppellman intercity disaggregate model.

The submodels are considered to be interdependent. Travel choices in the hierarchy are interrelated. Linkages among models are used to represent relationships among travel choice. Each travel choice in the hierarchy is made conditional upon all higher-level choices. The higher-level choice is influenced by the expected choices at the lower levels. This interrelationship is embodied through inclusion of composite variables that represent the combined attributes of all alternatives in the lower-level of the hierarchy. However, in a trip generation model, a linear regression approach is used due to the somewhat cumbersome formulation of a choice model for frequency choice. The

composite variable that would represent the service characteristics of destinations is excluded. Here, the trip generation is not based on utility maximization and the interrelationship between the trip frequency and mode choice is not included.

2.3.4 Use of Roy's Identity for Comprehensive Travel Demand Modeling

Kockelman and Krishnamurthy (2002) proposed a model that adequately represents trip generation rates while simultaneously recognizing the trip attributes such as time of day, mode, and destination. They specified the indirect utility function as

$$v = \alpha_0 + \sum_i \alpha_i \ln P_i + \sum_{ij} (1/2)\beta_{ij} \ln P_i^e \ln P_j^e + \sum_i \gamma_i \ln Y^e \ln P_i^e + \gamma_Y \ln Y^e. \tag{2.5}$$

where:

P_i : Unit price of trip i, and

Y : Income constraint.

Roy's Identity, which will be introduced in Chapter 4, was used to derive trip generation for different trip purposes, which correspond to home-based non-work, home-based work, and non-home-based trips, respectively. In this aspect, it is economically rigorous. The most valuable contribution is the generalization of the constraints to time and cost (money) limitations. The effective price and income is from the nested logit estimation for the other travel subdecisions, including mode choice and destination choice. However, note that the trip generation model and nested logit model for other subdecisions are estimated separately. The time to cost ratio from the nested logit estimation was used in the indirect utility function.

2.4 Discrete/Continuous Model

Recently, a procedure was developed by econometricians to integrate discrete choices and continuous choices into one framework. This framework, termed a discrete/continuous model, is both economically and statistically sound because both choices are based on microeconomic theory and one utility function maximization. The discrete choice model is derived from one underlying utility function. The utility function is used to define an implied continuous demand function via the well-known Roy's Identity. This model has been applied widely in the economics to determine demand for electricity, demand for water, demand for housing, and labor supply. The common theme in these fields is that consumer demand for goods has a discrete choice component as well as a continuous component for the alternatives.

Dubin and McFadden (1984) proposed a discrete/continuous model for space and water heat choice and electricity demand modeling. To develop the model, an indirect utility was specified. Liao and Chang (2002) analyzed the space heating and water heating choice and energy demand of the aged in the USA. In the same area, heating equipment choice and energy consumption was modeled with a discrete/continuous model by Nesbakken (2001), where a different indirect utility function was specified. Vaage (2000) applied this type of model in the appliance choice and energy demand with data from a Norwegian energy survey. The large number of applications of this procedure is due to its appealing theoretical consistency and accounting for interdependency between discrete and continuous choices.

Another important application field is in marketing science. The discrete/continuous model was utilized to model goods brand choice and the quantity to buy. Baltas (1998) reported that the discrete/continuous model, due to its strong foundation, taking simultaneity into account, concise and easy estimation, is used to model supermarket category choice and purchase quantities. Arora, Allenby and Ginter (1998) applied it to model canned vegetable soup as a supplement to the evening meal. It was concluded that the method provided more accurate characterization of the market than existing finite mixture and quantity independent models. Chiang (1991) examined consumers' coffee purchase decisions of whether to buy, what to buy, and how much to buy simultaneously with a discrete/continuous model. Another application was for the purchase of yogurt (Chingaguanta, 1993), where the discrete/continuous model was employed to study the impact of marketing variables on the category purchase, brand choice, and purchase quantity decisions of a household. The model takes into account interdependence among the three decisions. These decisions yield the greatest possible utility in combination.

The discrete/continuous model is also applied to public affairs. It is being used to model the fishing behavior in Bering Sea and the western Gulf of Alaska (Layton, Haynie, and Huppert, 2003). The decision choices modeled in this study are where the fish were caught and how much was caught.

With an emphasis on transportation applications, Mannering and Hensher (1987) provided a general overview of discrete/continuous econometric modeling. Included are many transportation-related decisions that involve both discrete choice (e.g. vehicle

choice) and continuous choices (e.g. vehicle use). Mainly, there are applications in three fields of transportation.

The first application is time use (time allocation and activity duration) models in transportation. These models aim to predict what people do with their time, what activities they participate in and how much time is allocated. Kitamura, Yamamoto, Fuji, and Sampath (1986) analyzed two types of discretionary activity and time location in a joint modeling framework. They concluded that the discrete/continuous model is a valuable tool in future endeavors. The model was extended to discrete/continuous/continuous model by Bhat (1998). Bhat (2001) also proposed a joint model for analyzing activity type and duration during the evening commute. He modeled activity type, home-stay duration before participating in out-of-home activity, and out-of-home activity duration. Hamed and Mannering (1983) analyzed activity-type, travel time, and activity duration using discrete/continuous model.

Another application subfield is in vehicle/utilization-related topics (Mannering and Winston, 1985; Mannering, 1983; Mannering, 1985) where discrete/continuous models were used to model the number of vehicles owned, the type of each vehicle, and the extent of utilization for each vehicle. In these models, two types of decisions, discrete choices of vehicle number and/or type and continuous choice of vehicle mileage traveled, are clearly interrelated and incorporated. These applications indicate that the failure to include such interrelationships could have significant implications for the validity of any result.

The discrete/continuous model was also applied to describe the commuter's route choice and departure time choice (Mannering, Abu-Eisheh and Arnadottir, 1990). It is reported that the model results are surprisingly accurate and show promise for applications in a traditional user equilibrium framework.

The discrete/continuous model is appropriate in modeling not only passenger transportation, but also freight transportation, such as commercial vehicle choice and shipment size (Holguin-Veras, P.E. 2002; Adbelwahab, 1998). It is also used to model the different types of tickets purchased and the amount of transportation-Train kilometers traveled (Vuuren and Rietveld, 2000).

The common characteristic in the above applications is that discrete and continuous choices are interrelated and the outcome of one choice has influence on the others. As Mannering and Hensher (1987) pointed out, "there are many other transport problems that encompass interrelated discrete and continuous decisions. In this sense, researchers have only begun to scratch the surface with regard to transport applications of discrete/continuous models" and, "First and foremost there is an urgent need to apply discrete/continuous modeling techniques to the many transport problems that involve interrelated discrete and continuous decisions". Although this was stated almost 15 years ago, the development and application of this type of model is still ongoing. The concept of discrete/continuous modeling has received little (if any) attention in intercity passenger travel demand modeling. Therefore, a nested-logit/continuous model in the intercity travel context is proposed in this paper.

Most of the discrete/continuous models (Hanneman, 1984; Dubin and McFadden, 1984; Vaage, 2000; Nesbakken, 2001; Liao and Change, 2002) are multinomial

logit/continuous models, in which the multinomial logit model is proposed to describe the discrete choice. In travel choice models, however, as Forinish and Koppelman (1993), Greene (1997) argued, the nested logit model is more suitable. The nested logit model does not have the IIA problem with the MNL. Therefore, in this dissertation, the discrete/continuous model is expanded to the most general nested logit model.

CHAPTER 3

INTERCITY TRAVEL ANALYSIS

This chapter analyzes the intercity travel decision process. Based on this analysis, an intercity travel behavioral conceptual framework is presented. Section 3.1 reviews an intercity decision making process. Section 3.2 includes a discussion of whether trip frequency should be a discrete or continuous variable. Section 3.3 discusses the transferability of current frameworks to this study. Section 3.4 conceptualizes the travelers' decision-making process in the context of intercity travel. This section includes model needs, assumptions for the model, and the presented conceptual framework.

3.1 Intercity Travel Decision-Making

Trip purpose induces intercity travel. People make intercity travels only if they want to participate in some activities. The most common trip purposes are business, recreation, and personal activities. Business travel has a purpose usually associated with the travelers' work. The travel decision maker and cost payer are not necessarily the traveler. What is the primary concern in this kind of decision making may be the travel time, and/or on-time performance rather than travel cost. These types of trips are not within the scope of this dissertation. All other trips are classified as non-business trips. The purposes for non-business trips could be recreation (i.e. vacation, sports or concerts, etc.), shopping trips, personal activities, i.e. visiting friends, relatives or others.

Demand for non-business trips is different for different people. There is an enormous number of potential variables that have an impact on travel demand. The most frequently mentioned are household income, car ownership, and age (Tardiff, 1977, Damadaron, 1988, Koppelman, 1989). Income, a measure of the ability to afford travel, plays an important role in the determination of whether to travel due to the cost associated with intercity travel. Low income groups would not make many costly trips for social and recreational purposes, while higher income groups would be able to afford to do so. For many in the highest income groups, it would be expected to attend numerous social functions. Thus, household income would influence the number of intercity trips. It has been shown that car owning households make more daily trips, including car, bus, train, and air trips, than households without a car. However, in the intercity travel context, this is not necessarily true, especially in the way that the number of cars affects the number of trips. Other socioeconomic factors, such as household size, age, gender, occupation, and education, also impact the possibility of making intercity trips.

In contrast to urban trip length, which is within commuting distance, intercity travel is much longer. In 1995, the local mean trip length is 9.0 miles and the domestic long-distance mean trip length is 826 miles (US BTS 1999). This longer trip costs more and takes longer. Therefore, the cost and time spent on travel weigh more in the travel decision making process.

Often the non-business trip is non-essential. It could be canceled or postponed for different reasons. There are various constraints when planning a non-business trip, such as destination limitations, available alternatives, travel context and so on. If a trip

purpose is to visit relatives or friends, usually the destination is the city where the friend/relative lives. In that case, destination is predetermined with trip purpose. However, for vacation trips, there exists a choice among different cities. More attractive cities are more likely to be chosen.

Another constraint are the available mode alternatives and routes between the specific origin and destination. Within a corridor, for every available mode, the number of reasonable routes is limited. As a result, the route and mode constraints can be assumed interdetermined. When the destination to make the trip is decided, the available route and mode alternatives are already determined in the existing transportation system. Thus, only mode choice behavior is taken into account. The mode choice has to be based on the available alternatives. If the available alternatives change, travelers may adjust their trip plans within a period accordingly. Hence, the available alternatives exert an influence on the trip frequency.

Among the available mode alternatives, the traveler chooses the one according to his/her preferences and constraints. For example, the income level of a household controls, to a large extent, the mode choice because it determines the amount of money available to be spent on travel. For the lower income groups, the relative costs of different alternatives of transportation would be of great importance in choosing a mode, while the higher income groups can afford to satisfy other preferences. Vehicle and licensed driver's availability also affect mode choice. If the household does not possess a vehicle and a licensed driver, then all travel must be undertaken using commercial carriers.

Attitude or personal preference also has an impact on the mode choice. For instance, if someone has a fear of flying, he or she may never choose an airplane as a transportation mode.

The context of a trip is another factor influencing the mode choice. In the case of a big travel party, it is more likely to drive a car or van. The trip frequency affects the mode choice decision too.

The available alternatives are characterized in terms of a set of modal attributes. There are many variables involved in describing the alternatives, such as travel time, travel cost, comfort, convenience, safety/security, reliability, etc.. Some of the variables are more directly causative than others for travel decision making. It is noteworthy that the safety/security became more predominant after the 9/11 tragedy (Liu and Li, 2003). However safety/security, comfort, reliability and convenience are not easy to measure, so it is difficult to incorporate them into the model.

The most often used attributes are travel time and cost. Travel time is total elapsed time while traveling, which is usually counted from door to door because the travel status begins once travelers leave their starting locations and ends when the travelers reach their destination. Travel time can be segmented into line-haul time, access/egress time, waiting time, and transfer time. Travel time is one of the main factors impacting mode choice and trip generation. It is the attribute that travelers usually try to minimize or trade off when facing mode choice. Because the available time resource for traveling is limited, travel time also exerts an influence on trip frequency. The relative costs would have an effect on the choice of mode. For non-business trips, the absolute cost would affect the decision of whether or not to make the trip. For example, as the

travel cost increases, the household plans its trips more carefully and makes fewer such trips as a result. The extent of the effect of cost depends on different travelers' attitudes, income, and other travel resources.

From the above analysis, it is clear that the intercity decision-making process consists of several subdecisions, "when to travel", "where to travel" and "which mode to choose", all of which are interrelated. The different elements, such as socioeconomic factors, context of trips, and transportation system, have an impact on these subdecisions to different extents.

3.2 Intercity Trip Frequency

Often, trip generation refers to trip frequency. In this study, trip generation and trip frequency are used interchangeably. As defined in Section 2.1.2, they refer to the number of trips made during a given period of time. Thus, the study period length determines the possible trip quantity. Within urban travel analysis, often one day is set as a study period. The corresponding trip frequency can be very limited, such as 0, 1, 2 …. The probability of a large trip frequency is very low. For instance, in the study of Domincich and McFadden (1996), only 4 out of 80 households surveyed made local shopping trips more than once. Daily intercity trip frequency, although not impossible, it is rare. An individual may take a certain number of trips in a period unless the time frame is so short that a majority of the travelers either do not make a trip or take only one trip. However, with cross-section data for just a short period, one has to consider that visits to a city may depend on recent trips to that city. Thus, a reasonable study period should be chosen to study intercity trip generation. In the 1995 American Travel Survey, the study period for

intercity travel study is one year. The survey registered all journeys made by a sample of individuals during that period. Within the corridor context, the trip frequency can be as low as 0, or as high as 110. In the literature, the intercity trip frequency is considered to be either a discrete or continuous variable, and the rest of this section discusses these two choices.

A discrete choice situation refers to one in which a decision-maker faces an option among a set of alternatives meeting the following criteria:

- The number of alternatives in the choice set is finite;

- The alternatives are mutually exclusive;

- The set of alternatives is exhaustive.

Usually the choices that concern "how many" or "how much" of something (which is the choice of quantity) have alternative sets that are denoted by continuous variables. Standard regression procedures are appropriate for these continuous outcomes.

As Train (1986) argued, many a continuous variable can be represented, without loss of accuracy and sometimes with increased accuracy, by a discrete variable. Specifically when there is some conceivable maximum for the variable, the number of alternatives is finite and the choice situation could be discrete. In this case, choices of "how many" or "how much" are more fruitfully analyzed with discrete choice methods if the number of alternatives is fairly small.

When there are a large number of alternatives such that the discrete dependent variable is essentially indistinguishable from a continuous one, standard econometric methods for continuous variables can be used adequately to represent the choice.

In this dissertation, one-year-long intercity trip horizon is discussed. It theoretically qualifies as a discrete choice situation because, during the study period, the trip frequency is clearly a finite, exhaustive set of mutually exclusive alternatives. Practically, however, intercity trip frequency is more appropriate to be processed as a continuous variable due to its large choice set. This also makes it possible for this frequency to be non-integer, which is obviously acceptable for intercity travel demand analysis.

3.3 Transferability of Current Frameworks

Typically, in an intercity travel demand analysis, the following decisions are accounted for: whether to travel, where to travel, and by what mode to travel. Commonly, two methods are used to carry out this travel behavior analysis: the separable/sequential analysis approach and the simultaneous analysis approach. Within the former, it is assumed that the decision-making process is sequential, in the sense that the decision of whether to travel does not affect the choice of the travel destination, and neither the time nor the destination influences the choice of the travel mode. These assumptions do not appear to be consistent with reality. Within the simultaneous approach, it is assumed that decisions of time, destination, and travel mode are made jointly. Accordingly, each subdecision interdepends on the others. Most of the time, the mathematical feasibility in practice is the main reason for the use of the sequential decision-making process. However, there still exists an unresolved question as to the order of the subdecisions. At least, no researchers have been able to justify the order they used. As analyzed in Section

3.1, it is more reasonable to incorporate the interrelatedness in the conceptual framework and model.

The classical demand submodels are run independently, producing relatively disconnected estimates of trip generation, destination choice, and mode choice. Little attention has been given to the interdependent determinations of mode choice and trip frequency. Although the direct demand model combines all of the choices into one, it suffers from the lack of a behavioral basis.

Ben-Akiva (1974) proposed a joint logit model in which both destination choice and mode choice are incorporated into one utility objective. However, this is not feasible for cases of high frequency. The model was extended by Adler and Ben-Akiva (1975) to trip generation, destination choice, and mode choice. However, the large amount of feasible combinations of frequency and modes make the calculation cumbersome. Thus, it still was not suitable for a large choice set resulting from all possible combinations of frequency, mode, and destination.

The same problem occurs with the ordered logit model, or binary logit model for the frequency choice. Within the travel demand models of the literature (Sheffi, 1979; Vickerman and Barmby, 1985), the trip generation was considered as a binary choice problem of whether to make a trip. In the corridor context, individuals may make more than one trip within the one-year study period. The frequency distribution for intercity trips within the Northeast Corridor in Chapter 5 supports this statement. So, there are a large number of binary nests if the binary logit model is used. Therefore, this type of discrete choice model for trip frequency will cause burdensome work. In these studies, trip generation was not interdependent with mode choice.

The model proposed by Damodaran (1988) incorporated the mode choice into the ordered logit model. However, the interrelatedness between mode choice and trip generation happens only in the choice between zero trips and greater than zero trips. This does not appear to be consistent with reality. If the number of trips to be considered separately is even a small number, Damodaran's technique becomes very cumbersome.

The model constructed by Domencich and McFadden (1996) is not completely interrelated because the utility functions are different for subdecisions even though the methodology utilizes an "inclusive price" variable that consists of the travel cost and time of the destinations. Subdecisions are estimated separately. Their model only considers the choice between no trip and one trip due to the limitations of the study case. The model structure only allows "one" or "no shopping" trips per household per day, which is an unrealistic set of choice for intercity travel on an annual basis.

In Koppelman's (1989) trip generation model, a linear regression approach is used due to the somewhat cumbersome formulation of a choice model for frequency choice. The composite variable that represents service characteristics is not included. Therefore, trip generation is not based on the utility maximization theory. The interrelationship between trip frequency and mode choice is not implemented in the model.

In the model proposed by Kockelman and Krishnamurthy (2002), trip generation is considered continuous and derived from an indirect utility via Roy's Identity. The ratio between effective price and income, which is from the nested logit estimation for the travel subdecisions including mode choice and destination choice, is used for the trip generation model. The trip generation model and nested logit model for other

subdecisions are estimated separately. The utility function for the logit model was not the same as the one used for trip generation.

These reviewed models may be appropriate for the cases in their studies. However, when considering their transferability to the intercity corridor demand analysis, they are limited. The most common drawbacks are incapability of handling the high trip frequency, or inconsistency of the utility used in the trip generation and mode choice models, or the lack of interrelationships between trip generation and mode choice. So far, no models have solved all these problems within the intercity travel context.

3.4 Conceptualization of Individuals' Travel Choice Decisions

According to the previous analysis and the literature reviewed, a model should embody the following features:

- The model is a disaggregate model;

- Trip frequency is a continuous variable;

- Interrelatedness among the different subdecisions is incorporated;

- Different submodels flow from one underlying utility;

- Different submodels are based on one dataset.

An intercity travel decision-making process is much more complex because of other trip related aspects such as time of the year, party size, etc., which are omitted in the conceptual model. To keep the model manageable, particularly for the corridor study, the following assumptions or simplifications are used in this conceptual framework:

- **No Destination Choice in the Structure**   In the context of general intercity travel, the number of destinations can be very large. It may be necessary to identify and characterize the largest group of feasible destinations for a given purpose. This problem could be possibly overcome by clearly identifying the area

being studied and the possible destinations. However, in this study, the reason is that perhaps the most important determinant of a destination choice for non-business trips is whether the individual has friends or relatives in a city. The cities cannot be interchangeable in these choice situations. For a related argument, please refer to Sonesson (2001).

- **One-Mode Preference** Since the auto, bus, train, and airplane modes are essential substitutes in the function of providing transportation from one location to another, a passenger prefers to choose only one travel mode within the study time period. There may be some logical or institutional reasons that the passengers could use more than one mode within the study period, but the consumers' tastes may be such that they naturally prefer to choose only one of the discrete alternatives. The data from the 1995 ATS support this statement. Most research regarding choice behavior also considers this type of case where only one alternative is chosen. For the cases where one could expect a different behavior if the choice decisions are made repeatedly, please refer to Keren and Wagenaar (1987), Keren (1991), Kahneman and Tversky, (1984), Kerstholt and Raaijmakers (1997).

- **No Route Choice in the Structure** Route choice is generally limited within the corridor travel context. Therefore, it is not taken into account. However, that does not mean that it is less important.

- **No Time Choice in the Structure** It is assumed that the time of a trip (i.e. season) has no influence on mode choice. However, the season effect is already incorporated into this one-year period study.

Based on the model requirements and assumptions, an individual choice paradigm within the corridor context is proposed as in Figure 3.1. In this framework, only two subdecisions: trip generation and mode choice, are taken into consideration. Mode choice is a discrete choice. Trip generation has a continuous outcome. As the framework exhibits, the travel behavior, i.e. trip frequency and mode choice, is the result of the complex process that is influenced by many interacting factors, such as individual's characteristics, modal attributes and trip characteristics. Travelers' characteristics include socioeconomic characteristics (such as age, gender) and psychological characteristics (e.g. attitudes, belief, etc.). The socioeconomic characteristics determine

the individual's psychological characteristics. Both of them determine the travel needs, which are defined by trip characteristics. However, this travel need is subject to the constraints from the personal socioeconomic and psychological status (e.g. household income, vehicle availability, airplane-phobia, etc.) and travel conditions (e.g. travel cost, travel time, comfort, etc.). A high quality travel condition is likely to increase the trip frequency, while a poor one may lead some people to choose not to travel. Improvement in the quality of one mode may lead to a decrease in the propensity to use other modes. The most important trip characteristics that influence mode choice are the length of the trip and trip purpose. The trip purpose and the length of the trip also influence the trip frequency. Shorter trips are made with greater frequency than longer trips. Trip frequency also influences mode choice. For a frequently visited place, an automobile might be used to make the trips. Conversely, the mode choice affects the trip frequency. Travel by airplane is not likely to be made as frequently as travel by automobile.
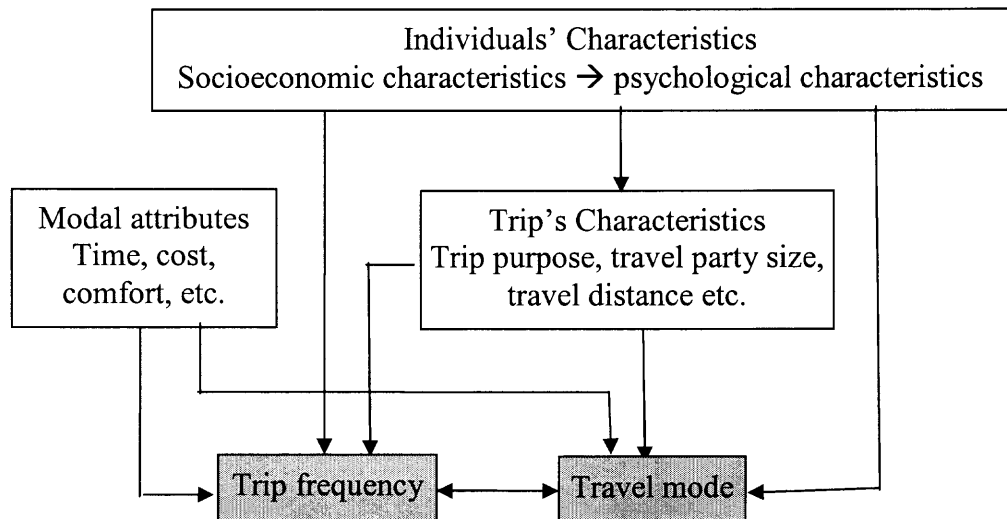
Figure 3.1 Behavioral framework of intercity travel.

To formulate intercity travel behavior, a model framework is proposed as in Figure 3.2.
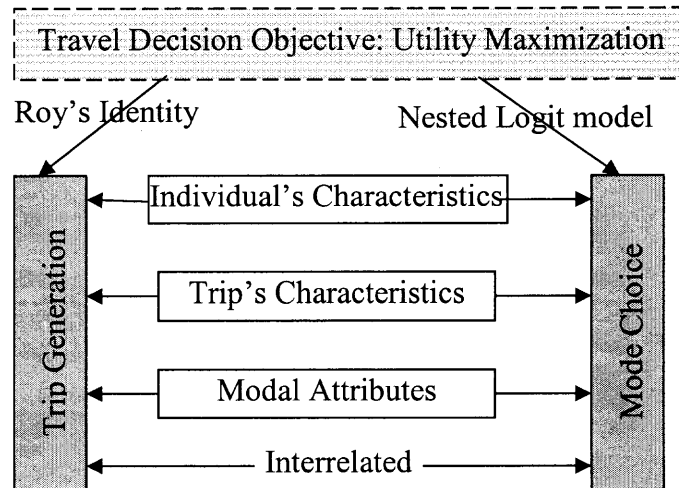
Figure 3.2 Intercity travel model framework.

The important theoretical concept applied in this model is the consumer choice theory from the econometric field, namely, utility maximization theory. The basic postulation of this theory is that every consumer chooses the option that maximizes his/her net personal utility, subject to constraints such as legal, social, physical and/or budgetary. The maximum utility theory is directed towards modeling commodity consumption. According to this theory, an individual consumes the commodity attributes according to the maximum utilities that these attributes provide.

Transportation demand is derived from activity demand. Travelers rarely travel only for the sake of traveling. They travel for some activity purpose, such as visiting friends. However, when an activity happens, a travel episode is needed to accomplish the activity because of the spatial constraints. As stated earlier, the household activity generation was modeled using the utility maximization theory. Thus, it is reasonable to

assume that one of the travel decisions, trip generation, is also based on the theory. Travelers are assumed to consume the attributes of transportation service. The available alternatives between a specific origin and destination are processed as different commodities. This implies that the demand for travel depends on the alternatives' attributes such as price, and travelers' characteristics.

Therefore, the selection of the annual set of trips can be conceptualized as an allocation problem in which the decision to pursue trips is traded-off against other time and money consuming activities. This also refers to the idea that trip generation can be considered to be a choice behavior based on the utility maximization theory.

In the model framework, an intercity travel decision has an objective in terms of utility maximization. Travelers must decide which mode to use - a discrete choice - and how many trips to make - a continuous choice. Travelers simultaneously optimize trip generation and model choice to maximize their utility according to various constraints. The optimal trip frequency can be obtained from the utility via Roy's Identity that is introduced in Chapter 4. Based on the utility, the mode choice is modeled with a nested logit model. The optimal trip frequency and chosen mode provide, in combination, the maximum utility. This is called a discrete/continuous model.

The demand for intercity travel is viewed as the result of utility-maximizing behavior. In this optimized process, trip generation and mode choice are utility consistent, which means they flow from one utility. Clearly then, the two subdecisions are interrelated since the optimal discrete choice depends partly on the outcome of continuous choice and vice versa.

CHAPTER 4

NESTED LOGIT/CONTINUOUS MODEL

In this chapter, a nested logit/continuous model is presented to implement the model framework proposed in Chapter 3. Section 4.2 introduces the random utility maximization approach and specifies an indirect utility for the travelers to maximize. For mode choice, the derivation of the nested logit model with a three-level nest using the specified indirect utility is presented in Section 4.3. In Section 4.4 the trip generation model is derived from the same specified indirect utility. The model's estimation methods are discussed in Section 4.5.

4.1 Introduction

Hanemann (1984) proposed the first unified framework for econometric models of discrete/continuous choice in which both the discrete and continuous choice stem from the same underlying random utility. Therefore, the two resulting choices are modeled in a mutually consistent manner.

Hanneman (1984), Dubin and McFadden (1984), Vaage (2000), Nesbakken (2001), and Liao and Chang (2002) proposed or applied the multinomial logit model within the discrete/continuous model. The multinomial logit model was used because it has the advantage of a simple mathematical structure and ease of estimation. However, this representation of choice behavior will result in biased estimates and incorrect predictions in cases which violate the Independence of Irrelevant Alternatives property. As many researchers (Forinash and Koppelman 1993; Greene 1997; and Hensher and

Greene 2002) suggested, the nested logit model is more appropriate for intercity travel mode choice modeling. Their results demonstrated a statistical rejection of the multinomial logit model in favor of a nested logit model. In this dissertation, the discrete choice component for a discrete/continuous model is extended to a nested logit model and this discrete/continuous model is called a nested logit/continuous model.

In this nested logit/continuous model, the travelers' discrete and continuous choices are derived from one specified utility function. Therefore, the interrelationships between mode choice decision and trip frequency are implied. In other words, when the random component of an indirect utility function is specified, the choice probabilities for different intercity travel modes can be estimated. Depending on the chosen mode, the intercity travel demand function is derived by applying Roy's Identity to the same indirect utility function.

4.2 Random Utility Maximization Model

As a representation, the utility function is used to define the level of utility or satisfaction for a certain commodity. It embodies the consumer's preference. In this study, the direct utility function is denoted as a function of the traveler's trip frequency:

$$u = u(x,b,z,s,\varepsilon).$$

$$(4.1)$$

where:

x: a continuous vector measuring the individual trip frequency;

b: a vector of attributes of alternatives;

z: a vector of all other trips which are not taken into consideration in x;

s : a vector of characteristics of the individual(age, occupation, etc.);

ε : a random component, which includes the variation in taste among individuals in the population and unobserved variations.

As Hanemann (1984) assumed, the randomness of the utility function is introduced due to the econometrician's limited ability to observe the individual's evaluation of the quality of different alternatives. For the individual traveler, ε is a set of fixed constraints or functions.

As in the aforementioned assumptions for the model in Chapter 3, travelers are restricted to choose only one mode within the study period. This results in only one $x_i>0$, and all other $x_j=0$ where i is the chosen mode and $i \neq j$. This means mutual exclusivity: $x_i*x_j=0$ all $i \neq j$. It is also assumed that $x_i=0 \rightarrow \partial U/\partial b_i = 0$, which is called "weak complementarity[1]" by Maler (1974).

Therefore, the utility u only depends on the characteristics of the chosen mode. This assumption is less restrictive than the additive separability of u.

Suppose that a traveler decides to choose only mode i. Dependent on this choice, his conditional utility can be written as a function of x_i and z:

$$\bar{u}_i = \bar{u}_i(x_i, b_i, z, s, \varepsilon) \ . \tag{4.2}$$

---

[1] In general, the nonzero demands are functions of *all* of the qualities of *all* goods. However, if weak complementarity exists between modes and their respective qualities, then the constrained demand and indirect utility functions associated with this problem can be simplified to $x_i = x_i(p_i, b_i, y, s, \varepsilon)$ and $\bar{u}_i = \bar{u}_i(x_i, b_i, z, s, \varepsilon)$. By definition, weak complementarity implies that utility does not change when the quality of an unchosen mode changes. This means that when $x_i = 0$, the constrained utility level does not depend on b_i. When researchers are specifying demand equations, it is inappropriate to include the quality of substitute modes when quality at those substitutes is thought to be weakly complementary.

The traveler maximizes this utility subject to income constraints, $\sum p_i x_i + z = y$,

and nonnegativity conditions, $x_i >= 0$, $z >= 0$, where p is the price of chosen mode i and y is

the income.

It is assumed that $\bar{u}_i$ is strictly quasiconcave[2] in x_i and z, and that none of the

indifference curves intersect the z axis i.e. x_i is essential with respect to $\bar{u}_i$. This ensures

that the conditional utility maximization has a corner solution with $x_i > 0$. If the

indifference curves do not intersect the x_i axis, there is a solution with $z > 0$. The resulting

conditional ordinary demand functions can be denoted $\bar{x}_i(p_i, b_i, y, s, \varepsilon)$ and

$\bar{z}(p_i, b_i, y, s, \varepsilon) = y - p_i \bar{x}_i(p_i, b_i, y, s, \varepsilon)$. Thus a conditional indirect utility function can

be defined as:

$$\bar{v}_i(p_i, b_i, y, s, \varepsilon) \equiv \bar{u}_i[\bar{x}_i(p_i, b_i, y, s, \varepsilon), b, z(p_i, b_i, y, s, \varepsilon), s, \varepsilon] \ . \tag{4.3}$$

So far, two utility functions have been introduced. They are $\bar{u}_i[\bar{x}_i, b, z, s, \varepsilon]$, the

"direct utility function", and $\bar{v}_i(p_i, b_i, y, s, \varepsilon)$, the "indirect utility function". Equation

4.3 means that, whether one uses a direct utility function or an indirect utility function, a

traveler's preferences are equivalent (Varian, 1978). Therefore, to represent a traveler's

preferences, an indirect utility function can be specified which implicitly represents a

direct utility function.

---

2 In economic models the assumption that a function is concave is sometimes too strong. For example, it may be unreasonable to assume that a consumer's utility function is concave. A weaker notion is quasiconcavity.

Consider a multivariate function f defined on the set S. For any real number a, the set $P_a = \{x \in S: f(x) \geq a\}$ is called an upper level set of f and the set $P_a = \{x \in S: f(x) \leq a\}$ is called a lower level set. Let f be a multivariate function defined on a convex set S. The function f is quasiconcave if every upper level set of f is convex (P_a is convex for every value of a). The function f is quasiconvex if every lower level set of f is convex (P_a is convex for every value of a).

Working with a consumer's indirect utility function to derive demand functions is much easier than using his direct utility function (Train, 1986). It circumvents the complicated constrained maximization problem of direct utility maximization. Therefore, the indirect utility function is introduced here. Well-behaved utility, $\overline{u}_i$, guarantees that all of the derived functions possess all of the standard properties. In particular, $\overline{v}_i$ is quasiconvex, decreasing in p_i and increasing in y.

A traveler chooses mode i and a trip number for the mode i, x_i, if the utility from this chosen mode i exceeds the utility from all other alternatives.

$$\overline{v}_i > \overline{v}_j, \text{ all } i \neq j \qquad (4.4)$$

Thus, given that a traveler picks mode i, the continuous choice, which is the trip frequency using mode i, can be derived via Roy's Identity[3] as:

$$\overline{x}_i(p_i, b_i, y, s, \varepsilon) = -\frac{\partial \overline{v}_i(p_i, b_i, y, s, \varepsilon)/\partial p_i}{\partial \overline{v}_i(p_i, b_i, y, s, \varepsilon)/\partial y} . \qquad (4.5)$$

The quantities $\overline{x}_i, \overline{z}$, and $\overline{v}_i$ are known numbers to the traveler. However, from the viewpoint of the econometric investigator, they are unknown because the travelers' preferences are incompletely observed.

---

[3] Roy's Identity: the demand for a goods is equal to the negative of the derivative of the indirect utility function with respect to the goods' price divided by the derivative of the indirect utility function with respect to income. In equation, the standard form of Roy' Identity is given as:

$$x_i^* = -\frac{\dfrac{dV}{dp_i}}{\dfrac{dV}{dy}}, \forall i \qquad (4.6)$$

where:

x_i^* : the optimal consumption of good per period,

V: the indirect utility,

p_i: the unit price of goods i;, and

y: the income.

Kockelman (1998) extended Roy's Identity to the time domain for use in household activity prediction. Proof of Roy's Identity can be referred to Train (1986).

To estimate the models for the probability of mode choice and individual travel demand, the indirect utility function has to be specified, including the random component ε_i and the utility function's non-random component.

The ways to introduce a random component into the utility function determine the class of random utility models. Here, it is assumed that $u(x,b,y,z,\varepsilon) = u(x,\psi(b,\varepsilon),y,z)$. As an index of the overall quality of the ith mode, the perceived quality index, ψ_i, is constructed from the b_i s, the attributes of alternative i, and the random component ε. It is denoted as a function of K (non-random) properties of alternative i, b_{ik}, and the random component ε_i.

The method in which the random elements are included in the perceived quality could be additive or multiplicative. Here, the multiplicative $\psi_i = \overline{\psi}_i(b_i)e^{\varepsilon_i}$ is adopted, where $\overline{\psi}_i(b_i)$ is non-random. The component $\overline{\psi}_i(b_i)$ is not necessarily homogenous across all travelers. Thus, it could be formed as a function of the attributes of alternatives as well as travelers' characteristics. To obtain explicit, closed-form expressions for the mode choice and the trip frequency model, the perceived quality function is assumed as:

$$\psi_i(b_i,\varepsilon_i) = \exp(\alpha_i + \beta_i y + \sum_{k=1}^{K}\gamma_k b_{ik} + \sum_{m=1}^{M}\delta_m s_{im} + \varepsilon_i) \ . \qquad (4.7)$$

Commonly, for trip generation, a higher "quality" level of an alternative implies a higher utility to the traveler. As Hanemann (1984) suggested, it is assumed that individuals maximize $\psi_i x_i$. This implies that different travel alternatives are substitutes for intercity passenger transportation service. The application of this utility model to discrete/continuous choice is suggested by Deaton and Muellobauer (1980):

$$u(x,\psi,z,s) = u^* \left(\sum_{i=1}^{N} \psi_i x_i, z, s \right).$$ (4.8)

Travelers maximize utility, subject to income constraints. It results in a corner solution meaning that only one single x is non-zero. The restriction of travel alternatives as substitutes is thereby solved. This implies that the traveler is assumed to consider one mode even though there are n available alternatives. The alternatives are deemed to be substitutes.

Suppose a traveler has decided to chose travel mode, i. The conditional utility function, $\overline{u}_i$, can be expressed as:

$$\overline{u}_i(x_i,\psi_i,z,s) = \overline{u}_i^*(\psi_i x_i, z, s)$$ (4.9)

Given that $\overline{u}_i$ is strictly quasiconcave in the variables and the variables are essential with respect to $\overline{u}_i$, the following forms for the conditional demand and indirect utility functions associated with $\overline{u}_i$, according to Muellbauer (1976), can be obtained: ·

$$\overline{x}_i(p_i,\psi_i,y,s) = \overline{x}_i^*(p_i / \psi_i, y, s) / \psi_i$$ (4.10)

And

$$\overline{v}_i(p_i,\psi_i,y,s) = \overline{v}_i^*(p_i / \psi_i, y, s)$$ (4.11)

From the above equations, the travelers adopt a `Price/Quality Ratio' to evaluate products, which is commonly within the economics field.

To utilize the above principle, the mathematical form of the indirect utility function has to be specified. Various indirect utility functions exist in the literature. Typical indirect utility functions' forms are described below:

The following function is suggested by Hanemann (1984). As Vaage (2000) suggested, however, it is extended to include individual characteristics:

$$\overline{v}_i = \left(\frac{\theta}{\rho-1}\right)\left(\frac{p_i}{\psi_i}\right)^{(1-\rho)} - \frac{e^{-\eta}\left[y+\delta_i's\right]}{\eta} \, . \tag{4.12}$$

Other utility models that are known to produce tractable demand models mentioned by Hanemann (1984) are:

$$\overline{v}_i = \left(\frac{\theta}{\rho-1}\right)\left(\frac{p_i}{\psi_i}\right)^{(1-\rho)} + \frac{1}{1-\eta}y^{1-\eta} \quad \theta > 0 \text{ and } \eta \neq 1 \tag{4.13}$$

$$\overline{v}_i = \left(y+\theta_1\left(\frac{p_i}{\psi_i}\right)+\theta_2\right)\left(\frac{p_i}{\psi_i}\right)^{-\eta} \, , \tag{4.14}$$

and

$$\overline{v}_i = \left(\log y - \theta\ln\left(\frac{p_i}{\psi_i}\right)\right)\left(\frac{p_i}{\psi_i}\right)^{-\eta} \, . \tag{4.15}$$

The criteria used to choose the indirect utility function form are:

- The utility function must have the desired properties such as a corner solution, etc.;

- The utility function must lead to a tractable demand model.

In this model, the indirect utility function (4.12) is used to derive the mode choice model and trip generation model.

4.3 Mode Choice Model

As stated previously, mode i will be chosen if, and only if,

$$\bar{v}_i \ (p_i / \psi_i, y, s) \geq \bar{v}_j \ (p_j / \psi_j, y, s) \text{ for all } j \neq i. \tag{4.16}$$

From the indirect utility specification (4.12), after some transformations, the choice probability of mode i, P_i, is:

$$P_i = \Pr\left\{\log p_i - \log \psi_i \leq \log p_j - \log \psi_j\right\} \text{ for all } i \neq j. \tag{4.17}$$

Substituting (4.7) for ψ_i in (4.17), it results in:

$$P_i = \Pr\left\{\begin{array}{l} \varepsilon_i + \alpha_i + \beta_i y + \sum_{m=1}^{M} \delta'_{im} s_m + \sum_{k=1}^{K} \gamma_k b_{ik} - \log p_i \geq \\ \varepsilon_j + \alpha_j + \beta_j y + \sum_{m=1}^{M} \delta'_{jm} s_m + \sum_{k=1}^{K} \gamma_k b_{jk} - \log p_j \text{ all } \quad j \neq i \end{array}\right\}, \tag{4.18}$$

where the β s and the δ s are the respective variable coefficients. If there is no random component in Equation 4.18, then we may have:

$$V_i = \alpha_i + \beta_i y + \sum_{m=1}^{M} \delta'_{im} s_m + \sum_{k=1}^{K} \gamma_k b_{ik} - \log p \ . \tag{4.19}$$

Hence, the probability equation is simplified as:

$$P_i = \Pr\{\varepsilon_i + V_i \geq \varepsilon_j + V_j\}. \tag{4.20}$$

where the V_i's are all of the nonstochastic components of the ψ_i's, plus the price effect.

As stated in Section 4.1, the nested logit model is more reasonable for mode choice. Therefore, in the following sections, a nested logit model is used to formulate the mode choice.

Let the set of modes be partitioned into L non-overlapping subsets denoted by B_1, B_2, ..., B_l, , ..., B_L, where the chosen mode, i, is included in B_n. The set of nests, B, is further partitioned into Q non-overlapping subsets denoted by D_1, D_2, ..., D_q, , ..., D_Q, and D_p is the nest that includes nest n.

| Top nest: D | Top nest (chosen) |
| q=1...p...Q | q=p |

| Middle nest: B | Middle nest (chosen) |
| l=1...n...L | l=n |

| Modes | Modes (chosen) |
| j=1...i...J | j=i |
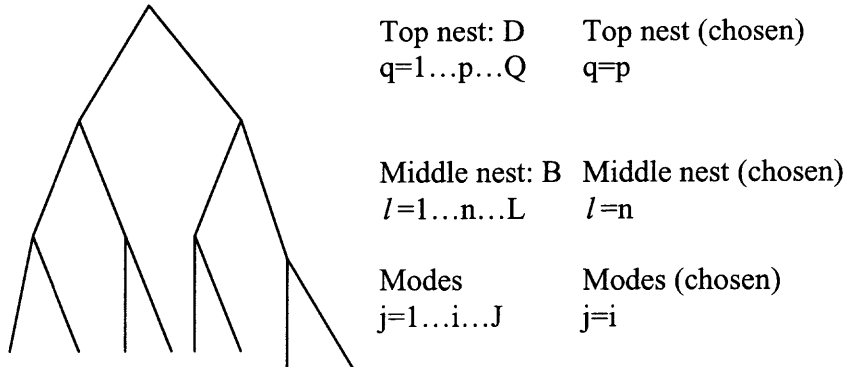
Figure 4.1 Three-level nested logit structure.

To generate a three-level nested logit model, it is specifically assumed that the cumulative density function of the vector of unobserved utility is:

$$F_\varepsilon \left(\varepsilon_{111}, ..., \varepsilon_{qlj}, ... \varepsilon_N \right) = \exp\left\{ -\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{-\varepsilon_{qlj} / \mu_l} \right]^{\mu_l / \lambda_q} \right]^{\lambda_q} \right\}. \quad (4.21)$$

This generalized extreme value distribution was first proposed by Manski and McFadden (1981). It covers the majority of nested logit model applications (McFadden 1978; Train, 2002; Morey 1997). Few studies suggest that analysts estimate models with more than three levels. This nested logit model can be nested to multiple levels to yield a broader class of functions.

Let $F_\varepsilon^{pni}(.)$ denote the derivative of $F_\varepsilon(.)$ with respect to the pni[th] argument.

$$F_\varepsilon^{pni}(\varepsilon_{pni}) = a_p * b_n * \left[\sum_{l=1}^{K} b_l \left[\sum_{j\in B_l} e^{(V_{plj}-V_{pni})/\mu_l}\right]^{\mu_l/\lambda_p}\right]^{\lambda_p-1} * \left[\sum_{j\in B_l} e^{(V_{pnj}-V_{pni})/\mu_n}\right]^{(\frac{\mu_n}{\lambda_p}-1)}$$

$$* \exp\left\{-e^{-\varepsilon_{Pnl}} * \sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j\in B_l} e^{(V_{qlj}-V_{pni})/\mu_l}\right]^{\mu_l/\lambda_q}\right]^{\lambda_q}\right\} * e^{-\varepsilon_{Pnl}}$$

(4.22)

There are four parameters here: a_q b_l μ_l λ_q. The conditions: $a_q \geq 0$, $b_l \geq 0$ and $1 \geq \mu_l \geq \lambda_q > 0$ guarantee that the function is defined over the unit interval and non-decreasing, and therefore satisfies the general extreme value distribution conditions. The condition $1 \geq \mu_l \geq \lambda_q > 0$ yields consistency with stochastic utility maximization (McFadden, 1978, Manski and McFadden, 1981). The a_q and b_l position the distribution and μ_l, λ_q determine its variance and covariance. The parameter of each mode in the lth nest will be equal for all alternatives within nest l, μ_l. It is a measure of the degree of independence in unobserved utility among the alternatives in the nest l. A higher value indicates greater independence and less correlation. λ_q is the same for all of the nests within nest q.

Thus, the choice probability can now be expressed as:

$$P_{pni} = \int_{\varepsilon_{pni}=-\infty}^{\varepsilon_{pni}=\infty} a_p * b_n * \left[\sum_{l=1}^{K} b_l \left[\sum_{j\in B_l} e^{(V_{plj}-V_{pni})/\mu_l}\right]^{\mu_l/\lambda_p}\right]^{\lambda_p-1} * \left[\sum_{j\in B_l} e^{(V_{pnj}-V_{pni})/\mu_n}\right]^{(\frac{\mu_n}{\lambda_p}-1)}$$

$$* \exp\left\{-e^{-\varepsilon_{Pnl}} * \sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j\in B_l} e^{(V_{qlj}-V_{pni})/\mu_l}\right]^{\mu_l/\lambda_q}\right]^{\lambda_q}\right\} * e^{-\varepsilon_{pnl}} * d\varepsilon_{pni}$$

. (4.23)

To simplify the derivation process, the following substitutes are defined:

$$A = \left[\sum_{j \in B_l} e^{(V_{pnj} - V_{pni})/\mu_n} \right]^{(\frac{\mu_n}{\lambda_p} - 1)} ,$$
(4.24)

$$B = \sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{qlj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q} , \text{ and}$$
(4.25)

$$C = \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{plj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_p} \right]^{\lambda_p - 1} .$$
(4.26)

So, the probability of choosing alternative pni becomes:

$$P_{pni} = \int_{\varepsilon_{pni} = -\infty}^{\varepsilon_{pni} = \infty} a_p * b_n * C * A * \exp\left\{ -e^{-\varepsilon_{Pnl}} * B \right\} * e^{-\varepsilon_{pnl}} * d\varepsilon_{pni}$$

$$= a_p * b_n * C * A / B$$
(4.27)

In common practice, this probability can be written as:

$$P_{pni} = \frac{e^{V_{pni}/\mu_n}}{\sum_{j \in B_l} e^{V_{pnj}/\mu_n}} \frac{b_n * \left[\sum_{j \in B_n} e^{V_{pnj}/\mu_n} \right]^{\mu_n/\lambda_p}}{\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{plj}/\mu_l} \right]^{\mu_l/\lambda_p}} \frac{a_p * \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{plj}/\mu_l} \right]^{\mu_l/\lambda_p} \right]^{\lambda_p}}{\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qlj}/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}} . \quad (4.28)$$

If it is assumed for the nested logit model that, $v_{qlj} = w_q + w_{ql} + \varphi_{qlj} + \varepsilon_{qlj}$, where

w_q: The utility attributable to nest q,

w_{ql}: The utility attributable to nest ql, and

φ_{qlj} : The utility attributable to alternative qlj,

then,

$$P_{pni} = \frac{e^{(W_p+W_{pn})/\mu_n}e^{\varphi_{pni}/\mu_n}}{e^{(W_p+W_{pn})/\mu_n}\sum_{j\in B_l}e^{\varphi_{pnj}/\mu_n}} \frac{e^{W_p/\lambda_p}*b_n*e^{W_{pn}/\lambda_p}\left[\sum_{j\in B_n}e^{\varphi_{pnj}/\mu_n}\right]^{\mu_n/\lambda_p}}{e^{W_p/\lambda_p}*\sum_{l=1}^{K}b_le^{W_{pl}/\lambda_p}\left[\sum_{j\in B_l}e^{V_{plj}/\mu_l}\right]^{\mu_l/\lambda_p}}*$$

$$\frac{a_p*e^{W_p/\lambda_p}\left[\sum_{l=1}^{K}b_le^{W_{pl}/\lambda_p}\left[\sum_{j\in B_l}e^{\varphi_{plj}/\mu_l}\right]^{\mu_l/\lambda_p}\right]^{\lambda_p}}{\sum_{q=1}^{Q}a_qe^{W_q/\lambda_q}\left[\sum_{l=1}^{K}b_le^{W_{ql}/\lambda_q}\left[\sum_{j\in B_l}e^{\varphi_{qlj}/\mu_l}\right]^{\mu_l/\lambda_q}\right]^{\lambda_q}} \qquad (4.29)$$

$$= P(i\,|\,n)*P(n\,|\,p)*P(p)$$

Here, most researchers have used inclusive values (IV) for nest l or upper-level nest q,

$$IV(l\,|\,q) = \ln(\sum_{j\in B_n}e^{\varphi_{pnj}/\mu_n});$$

$$IV(q) = \ln(\sum_{l=1}^{K}b_le^{W_{ql}/\lambda_q}\left[\sum_{j\in B_l}e^{\varphi_{qlj}/\mu_l}\right]^{\mu_l/\lambda_q}) = \ln(\sum_{l=1}^{K}b_l*e^{(W_{ql}/\lambda_q+\mu_l/\lambda_q[IV(l|q)])}) \qquad (4.30)$$

Thus,

$$P(i \mid n) = \frac{e^{\varphi_{pni}/\mu_n}}{\displaystyle\sum_{j \in B_l} e^{\varphi_{pnj}/\mu_n}}$$

$$P(n \mid p) = \frac{b_n * e^{W_{pn}/\lambda_p} \left[\displaystyle\sum_{j \in B_n} e^{\varphi_{pnj}/\mu_n}\right]^{\mu_n/\lambda_p}}{\displaystyle\sum_{l=1}^{K} b_l e^{W_{pl}/\lambda_p} \left[\displaystyle\sum_{j \in B_l} e^{V_{plj}/\mu_l}\right]^{\mu_l/\lambda_p}} = \frac{b_n * e^{(W_{pn}/\lambda_p + IV(n\mid q)\mu_n/\lambda_p)}}{\displaystyle\sum_{l=1}^{K} b_l e^{(W_{pl}/\lambda_p + IV(l\mid q)\mu_l/\lambda_p)}}$$

$$P(p) = \frac{a_p * e^{W_p} \left[\displaystyle\sum_{l=1}^{K} b_l e^{W_{pl}/\lambda_p} \left[\displaystyle\sum_{j \in B_l} e^{\varphi_{plj}/\mu_l}\right]^{\mu_l/\lambda_p}\right]^{\lambda_p}}{\displaystyle\sum_{q=1}^{Q} a_q e^{W_q} \left[\displaystyle\sum_{l=1}^{K} b_l e^{W_{ql}/\lambda_q} \left[\displaystyle\sum_{j \in B_l} e^{\varphi_{qlj}/\mu_l}\right]^{\mu_l/\lambda_q}\right]^{\lambda_q}} = \frac{a_p * e^{(W_p + \lambda_p IV(p))}}{\displaystyle\sum_{q=1}^{Q} a_q e^{(W_q + \lambda_q IV(q))}}$$

. (4.31)

4.4 Trip Generation Model

Using Roy's Identity and some simplifications to indirect utility (4.12), results in a model for trip frequency:

$$\overline{x}_{pni} = \theta p_{pni}^{-\rho} \eta^{*\left[y + \delta'_{pni}s\right]} e^{(\rho-1)V_{pni}} e^{(\rho-1)\varepsilon_{pni}}.$$ (4.32)

In this specification, the trip generation's income elasticity is $\eta * y$. The price elasticity is $-\rho$. This is rather different from the conventional trip generation model. Usually, the conventional trip model does not include variables such as price. Hence, it can not be used to estimate demand elasticity. In this model, inclusion of the price means that traveler's demand for intercity travel also depends on price. However, this

specification causes a problem when the trip rate is 0. The discussion for this problem can be referred to Chan (2001). The available data are suitable for this specification.

For regression convenience, function 4.33 can be transformed to logarithmic form:

$$\log \overline{x}_{pni} = \log \theta - \rho * \log p_{pni} + \eta *(y + \delta'_{pni}s) + (\rho - 1)V_{pni} + (\rho - 1)\varepsilon_{pni} \quad (4.33)$$

ε_{qni} in Equations 4.32 and 4.33 is the random component. Conditional on the choice of mode qni, $\varepsilon \mid V_{pni} > V_{qlj}$, the following can be defined:

$$A_{pni} = \left\{ \varepsilon \mid \varepsilon_{pni} + V_{pni} \geq \varepsilon_{qlj} + V_{qlj}, all \quad q, j \ and \quad l \right\} \quad (4.34)$$

Then, the conditional marginal density is

$$f_{\varepsilon_{pni} \mid \varepsilon \in A_{pni}}(\varepsilon_{pni}) = F_{\varepsilon}^{pni}(\varepsilon_{pni} + V_{pni} - V_{111}, ..., \varepsilon_{pni} + V_{pni} - V_{qlj}, ...,$$

$$\varepsilon_{pni} + V_{pni} - V_{KN}) / P_{pni}. \quad (4.35)$$

With the Extreme Value distribution, one obtains:

$$f_{\varepsilon_{pni} \mid \varepsilon \in A_{pni}}(\varepsilon_{pni}) = \frac{a_p * b_n * \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{plj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_p} \right]^{\lambda_p - 1} * \left[\sum_{j \in B_l} e^{(V_{pnj} - V_{pni})/\mu_n} \right]^{(\frac{\mu_n}{\lambda_p} - 1)} * \exp\left\{ -e^{-\varepsilon_{Pnl}} * \sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{qlj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q} \right\} * e^{-\varepsilon_{Pnl}}}{a_p * b_n * \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{plj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_p} \right]^{\lambda_p - 1} * \left[\sum_{j \in B_l} e^{(V_{pnj} - V_{pni})/\mu_n} \right]^{(\frac{\mu_n}{\lambda_p} - 1)}}{\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{qlj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}}} \quad (4.36)$$

$$= B * \exp\left\{ -e^{-\varepsilon_{Pnl}} * B \right\} * e^{-\varepsilon_{pnl}}$$

As previously used, $B = \sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{(V_{qlj} - V_{pni})/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}$, which is constant in

Equation 4.36.

Thus, the expected value of $\varepsilon \mid A$ is:

$$E[\varepsilon_{pni} \mid \varepsilon \in A_{pni}] = \int_{\varepsilon = -\infty}^{\varepsilon = \infty} B * \varepsilon * e^{-\varepsilon} * \exp\left\{-B * e^{-\varepsilon}\right\} * d\varepsilon \qquad (4.37)$$

The density function for an extreme value distribution having ln(C) as its mode, is as follows:

$$f(x) = C * e^{-x} * \exp\left\{-C * e^{-x}\right\} \qquad (4.38)$$

The expected value $\int_{x = -\infty}^{x = \infty} x * f(x) * dx$ is ln(C)+0.5772 (Morey, 1997).

Thus,

$$E[\varepsilon_{pni} \mid \varepsilon \in A_{pni}] = \int_{\varepsilon = -\infty}^{\varepsilon = \infty} \varepsilon * B * e^{-\varepsilon} * \exp\left\{-B * e^{-\varepsilon}\right\} * d\varepsilon = \ln(B) + 0.5772 \qquad (4.39)$$

The expected value of Equation 4.33 is:

$$E[\log x_{pni}] = \log \theta - \rho * \log p_{pni} + \eta * (y + \delta'_{pni} s) + (\rho - 1)[V_{pni} + \ln(B) + 0.5772]\ldots$$

$$\ldots \ldots (4.40)$$

If B from Equation 4.25 is substituted in Equation 4.40, we get:

$$E[\log x_{Pni}] = \log\theta - \rho * \log p_{pni} + \eta * (y + \delta'_{pni}s) +$$

$$(\rho - 1)[\ln(\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qlj}/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}) + 0.5772] \qquad (4.41)$$

In Equation 4.41, the term $\ln(\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qlj}/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}) + 0.5772$ is defined

as the selection term. The interaction between the mode choice and trip generation is represented by this term. Obviously, if the assumption that mode choice and trip frequency are interrelated decisions is correct, omission of this variable will cause bias in the intercity travel demand model.

In Equation 4.41, μ and λ can be identified in the nested logit model. The expression $(\rho - 1)$ shows the extent that the travel frequency is affected by the choice of mode. Hanemann (1984) noted that the necessary and sufficient condition for mode x_i to be essential with respect to the utility function is $(\rho < 1)$, i.e. the estimate of $(\rho - 1)$ should be negative. From Equation 4.41, the direct price elasticity is $-\rho$. The range of ρ can be used to test the validity of the specification, since values greater than one or less than zero are infeasible.

Consequently, the estimating conditional demand equation becomes:

$$\log x_{Pni} = a_0 - a_1 * \log p_{pni} + a_2 y + \sum_{m=1}^{M} a_{3m} s_m +$$

$$(\rho - 1)[\ln(\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qlj}/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}) + 0.5772] + \tau \qquad (4.42)$$

From the function $f_{\varepsilon_{pni}|\varepsilon \in A_{pni}}(\varepsilon_{pni}) = B * \exp\left\{-e^{-\varepsilon_{Pnl}} * B\right\} * e^{-\varepsilon_{pnl}}$, it is implied that ε

is Gumbel distributed with parameters (lnB, 1). According to the 4th basic property of the

Gumbel distribtution, (Ben-Akiva and Lerman, 1985), $\log x_i$ is Gumbel distributed with

parameters

$$((\rho-1)[\ln(\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qij}/\mu_l}\right]^{\mu_l/\lambda_q}\right]^{\lambda_q}) + a_0 - a_1 \log p_i + a_2 y + \sum_{m=1}^{M} a_{3m} s_m, \quad 1/(\rho-1)) \cdot$$

In Equation 4.42, τ is Independently and Identically Distributed Extreme-Value

distributed, i.e., $EV(1/(\rho-1), -0.5772/(\rho-1))$. Hence, it is a conventional error term

with zero mean and a variance of $\pi^2/\sqrt{6}(\rho-1)^2)$. This distribution is determined under

the assumption of a perceived quality function.

As the two choice decisions (i.e. mode choice and trip frequency) are the

derivative consequences of a single utility maximization for a traveler, the model ensures

that these decisions provide, in combination, the greatest possible utility to that traveler.

The model formulation allows the decisions to be interrelated. Specifically, if

there are factors that affect a traveler's utility that are perceived by the traveler but are

unobservable to the researcher, a change in these factors will affect all of the choice

decisions.

4.5 Model Estimation

The nested logit/continuous model proposed here can be estimated by a full information likelihood maximization technique or a two-step approach.

4.5.1 Full Information Likelihood Maximization Estimation

The full information estimation allows the most efficient use of available information. Suppose that the sample contains T individuals, which includes t^{th} individual's mode choice i and trip generation rate x. The likelihood function to be maximized can be written as

$$L = \prod_{t=1}^{T} f_{x_{i^*t}}(x_t^*)$$

$$L = \prod_{t=1}^{T} f_{x_{i^*t}|\varepsilon\in A_{i^*t}}(x_t^*) * P_{i^*t} \tag{4.43}$$

Where i*t is the index of the mode chosen by traveler t and,

x_t^* is the observed trip generation by individual t.

P_{i^*t} is the probability of individual t choosing i, which is given by P_{pni}.

$f_{x_{i^*t}|\varepsilon\in A_{i^*t}}(x_t^*)$ is given by $f_{\varepsilon_{pni}|\varepsilon\in A_{pni}}(\varepsilon_{pni})$, where ε_{pni} is replaced by x_t^*.

If it is possible to derive a closed-form solution for L, using Equation 4.43, the coefficients can be conceivably estimated by maximizing L. Similar to the argument by Amemiya (1973), this estimation is consistent and asymptotically normal and efficient.

However, the requirement of the function L being closed-form is quite severe. The normal equations may have multiple roots. Therefore, convergence to the global maximum can not be guaranteed. To simplify the problem, it is easy to exploit the two-step estimation method.

4.5.2 Two-Step Estimation Approach.

Similar to Heckeman's (1979) sample selection models, the two-step estimation procedure can be used to decompose the full information likelihood maximization estimation.

The first step is to estimate the nested logit model using the maximum log likelihood. The relationship between the observed choice i of the tth individual and the explanatory variables can be formulated with the following log likelihood function:

$$\log L = \sum_{t=1}^{T} \sum_{i=1}^{I} d_{it} \log P_{pni} \qquad (4.44)$$

Because it does not take the continuous model into account, these coefficients' estimation is not efficient but consistent.

The second step is a regression analysis of the trip generation model. Estimated coefficients from the nested logit model are used to generate the selection term. The conditional demand equation can be estimated by ordinary least squares.

If the full information is still desired, the two-step estimates can be used as an initial consistent estimator in the full-information likelihood function.

The estimation in this study is carried out using the SAS (Statistical Analysis Software) package.

CHAPTER 5

DATA PREPARATION

The models exhibited in Chapter 4 can be used by transportation modelers to predict intercity travel. In this dissertation, the model is applied to the NorthEast Corridor. Data–related topics are described in this chapter as a first step of the case study. Section 5.1 presents the data needs for the disaggregate intercity travel model. Section 5.2 discusses the different datasets used in the literature and the potential datasets for this study, specifically, the 1995 American Travel Survey (ATS). Section 5.3 presents the data preparation for this case study.

5.1 Data Need

As stated earlier, a disaggregate model attempts to disclose the individual's travel behavior pattern. According to different cases, the basic study unit could be at the individual level, or at the household level. An individual level study requires data about the individual's characteristics and the individual's revealed travel. A household level study needs data about the revealed travel and characteristics of the households.

In the conventional sequential travel decisions model, most subdecision modules are formulated at an individual level except for the trip generation. Because most trips are home based, it is logical that the household is considered as the study unit. Actually, there is no absolute reason why trip generation modeling should be carried out on a household or individual basis. In this dissertation, the analysis at the individual level has the following benefits:

- The model proposed here analyzes different subdecisions of travel choice. In this study, the mode choice model makes use of individual data. For consistency, individual data is also used for the trip generation model.

- Consistency with the intended usage for modeling and forecasting because most of the population forecasts employed are essentially individual based.

- Reflect the fact that not all members of a household have the same level of choice.

- The influence of vehicle availability on a person's trip frequency and mode choice is considered.

- Trip frequency by people of different occupations is taken into account.

- Make maximum use of survey data's information.

- Reduce heterscedasticity[1] encountered in household models.

- Be consistent in the identity of the response factor (trips) and the generator of that response (persons).

Considering the above advantages, the data used by the disaggregate model proposed in this study is suggested to be at an individual level.

As the proposed framework in Section 3.4 indicates, different individuals have different demand for non-business trips. The socio-economic characteristics of an individual are one of the influencing factors. Thus, the first section of the required dataset is the socio-economic data of the traveler. It may be gender, occupation, income, age, etc.. It is noteworthy that the socioeconomic data here also include household level data, such as household income, because the household context exerts an effect on the individual's trip making decision.

---

[1] Heteroscedasticity refers to the situation where the error terms do not have constant variance. It is thus a violation of one of the classical error assumptions (homoscedasticity), i.e. $E(\mu_i^2) \neq \sigma^2$. Often arises in cross-sectional studies.

The main objective of the proposed framework is to model individual travel behavior and there should be revealed travel data to calibrate. The actual trip data is composed of trip frequency, mode chosen, travel origin, travel destination, travel activity, party size, travel purpose, etc..

Because all travel choices are based on existing transportation supply, the available travel alternatives impact travel behavior. The level of service of the alternatives is described by travel cost, travel time, frequency, comfort, convenience, safety, reliability and so on. It is required to include the attributes of the available alternatives in the dataset.

In summary, the data requirement for the proposed framework is at an individual level. It should contain socio-economic characteristics of the travelers, revealed travel choice and related attributes, and the available travel alternatives.

5.2 Data Sources

A dataset from a travel survey at an individual level is needed for disaggregate transportation analyses. However, the availability of such data is limited. A full complete dataset embodying the above-mentioned data is not available in the US. Different data sources utilized in the literature and potential data for this study are presented in Section 5.2.1.

5.2.1 Different Data Sources

The data used by Bhat (1995) and Bhat (1997) is an intercity travel behavior dataset from the Toronto-Montreal corridor in Canada. This data is from the Rail Passenger Review conducted by VIA Rail (the Canadian national rail carrier) in 1989 to develop travel demand models to forecast future intercity travel and estimate mode shifts in response to a variety of potential rail service improvements in the Toronto-Montreal corridor. Included in this dataset are socio-demographic and general trip-making characteristics of the travelers, and detailed information on the current trip, such as travel purpose, party size, origin and destination cities, etc. This data still requires extra effort to meet the needs of a disaggregate intercity travel behavior model, such as the set of modes available to travelers and the level of service of these alternatives. Based on the geographic location of the trip, the available mode alternatives for specific origin and destination are determined. For VIA Rail, KPMG Peat Marwick collected the level of service data for each available mode. In Bhat (1995), 2,769 business travelers' behavior is formulated. In Bhat (1997), 3,593 business travelers' data are modeled.

Forinash and Koppelman (1993) used the dataset of the Ontario-Quebec corridor from Windsor in the west to Quebec City in the east, assembled by VIA Rail in 1989 to estimate the demand for high-speed rail in the Toronto-Montreal corridor and support future decisions on rail service improvement in the corridor. The data consists of travel volumes and impedance by mode and travel surveys collected on all four modes in 1988 for travel beginning and ending in 136 districts in the region. Its resultant sample in the study was 4,323 individual trips.

Similar to the American Travel Survey, there is the 1985 Canadian Travel Survey which was used by Damoradon (1988), Wilson, Damoradan, Innes (1990), and Abdelwahab and Innes and Stevens (1992). The data has socioeconomic information on the individual and the details of all trips taken by that individual. When used for disaggregate modeling, the auxiliary data items was compiled from secondary sources. For example, unpublished data on travel time, travel cost and frequency of service obtained from the strategic planning division of Transport Canada of Ottawa for some origin-destination pairs. Others were secured from published schedules or via regression equations. In Wilson, Damoradan, Innes (1990), the sample size was 1,624 trips. Abdelwahab and Innes and Stevens (1992) adopted 1,712 observations for their study.

Algers (1993) used the data of the Swedish national travel survey conducted in 1984-1985. The information in this dataset is composed of socioeconomic data of the individual and his or her household as well as trip-related information, such as access and egress modes, main mode, destination, trip purpose, party size, number of overnight stays, and type of accommodations. Because travel time is not included in the dataset, it was provided by the national transportation council, using a network analysis system (EMME/2). The construction of the mode-related cost variables is based on assumptions regarding the time of day of the trip as well as the mix of people in the travel party.

A similar dataset used in Koppelman (1989) and Grayson (1981) is the 1977 National Travel Survey (NTS) in the United States. The information in this dataset consists of actual trip characteristics, the area of residence, trip origin and destination, etc.. The actual trip characteristics include purpose, timing, duration, and the means of transportation used. The deficiency of this dataset is that it lacks transportation supply

data. This deficiency makes it necessary for it to be supplemented with information from external sources. The data were supplemented with published level of service data for the available modes and fare classes including travel time, fare, and service frequency, or estimated with a function of distance and region or of origin and destination. Different samples are used in different studies or models. For example, Grayson (1981) adopted two sample sizes, 1,658 trips along 46 routes that were the most heavily sampled, and 1,062 trips along 41 routes representing the greatest number of passenger miles.

Stephanedes, Kumar and Padmanabhan (1984) conducted a transportation survey of trips between the Twin Cities and Duluth, MN in March 1982 for their study. For the non-chosen alternatives, some external sources were used to provide the trip characteristics such as access time, distance and cost. Estimates were used, when necessary, from chosen alternative attributes and urban and rural traffic data. This study made use of 90 records resulting from this survey.

In the industrial field, some surveys were carried out for transportation projects. The typical project is a feasibility study for a large investment project for a certain corridor, such as the Maglev project. The Baltimore-Washington Maglev project obtained travel and socioeconomic data from the Metropolitan Planning Organizations: the Baltimore Metropolitan Council and Metropolitan Washington Council of Governments (Maryland Mass Transit Administration, 2000). Extensive survey data was collected at locations around the region. Other data was secured from Amtrak and local visitors' bureaus. Finally, all these different data sources were synthesized for modeling. The same type of data was collected for Boston-Montreal High-Speed Rail Feasibility Study. This dataset is composed of detailed revealed trip information, demographic

information, and stated preference data, which provides information for the development of mode choice, destination, and trip frequency models for the project (Vermont Agency of Transportation, et al. 2003). The other Maglev projects, such as Atlanta-Chattanooga Maglev Project, Pennsylvania Maglev Project, California-Nevada Maglev Project, Southern California Maglev Project, performed some survey for model development of these feasibility studies. These datasets are desired for this study. However, data acquisition is of a problem. For developing a new data base for intercity travel modeling, please refer to Koppelman and Hirsh (1986). Due to limitations of study cost and time, this study will be based on the available dataset. In the United States, the 1995 American Travel Survey (ATS) can be obtained for this study. The travel survey provides a rich source of information on intercity travel undertaken over a period of one year. Even though it is not fully complete, it can be used to test the proposed model structure when combined with data from different sources. Thus, in Section 5.2.2, the 1995 ATS is described in details. Section 5.3 presents the data collection process and preparations for the case study.

5.2.2 The 1995 ATS

Prior to the 1995 ATS, the most recent source of data on passenger flows was the 1977 national travel survey, conducted by the Bureau of the Census as a component of the census of transportation. To meet the need for passenger flow data, the BTS (Bureau of Transportation Statistics), which congress formally established, conducted the 1995 American Travel Survey.

In response to a particular interest in understanding the differences in travel patterns by state, the 1995 ATS provided detailed information on state-to-state travel and metro-to-metro travel.   It covers geographic origin and destination, housing and household characteristics of the person's household, personal characteristics, and personal trip characteristics and distance calculations. All of the factors may influence a person's travel choice over a period of time.   Housing and household characteristics include tenure, structure, vehicles, vans, utility vehicle, household type, household income, household size, etc.. The person's characteristics include age, origin and race, sex, marital status, education, and personal income.   The trip characteristics cover the number of trips away from home, reasons for trips, trip durations, trip distances, travelers in the travel party, nights away from home, types of lodging, principal transportation from origin to destination, etc..   There are no questions concerning trip related expenditures. The trip is defined here as any trip, 75 miles or longer one way, taken by any member of the household.

The samples of households were selected at the beginning of calendar year 1995 for interviewing that began in April 1995 and continued through March 1996. The sample contained about 80,000 eligible addresses. Sample households were interviewed three to four times during this period at approximately 3-month intervals. All trips made within the survey period were recorded.

5.3 Data Preparation

In the 1995 ATS dataset, 58% of all trips were less than 500 miles round trip. The dominant intercity travel is short-haul travel. Therefore, the application of the proposed model is directed towards the corridor context. The NorthEast Corridor has been studied since the beginning of intercity travel behavior research in the early 1950's. This study is limited to the "Southend" of the NorthEast Corridor: i.e., Washington to New York (225 miles). Other reasons for selecting the NorthEast Corridor are:

- The NorthEast Corridor is one of the most-densely populated and urbanized sections of the United States.

- Within the corridor, all passenger modes, such as auto, intercity bus, conventional rail, high-speed rail and air transportation are available. This provides a good arena to study intercity travel behavior.

Section 5.3.1 introduces the transportation systems and travel characteristics of the study corridor. Section 5.3.2 presents the data filtering process. Section 5.3.3 synthesizes data from different sources for modeling.

5.3.1 Corridor Background

As a typical corridor, the NorthEast Corridor contains three of the six largest consolidated metropolitan statistical areas of the United States (US DOC, U.S. Census Bureau, 2002). It has the most complete transportation systems. This corridor is critical to the national transportation system and economic development. It experiences the heaviest intercity travel density of any corridor in the nation and has the most extreme transportation problems, which are in need of immediate solutions. Hence, using this corridor as the case study has an important empirical reason.

Traditionally, included within the NorthEast Corridor is the area from Washington DC, to New York, and Boston. It is often convenient to further divide the corridor into the "South_end", i.e. Washington DC to New York (225 miles) and the "North_end", i.e. New York to Boston (231 miles). The most heavily populated areas are centered near Washington DC and New York. Only the South_end of this corridor is taken into account in this study. The study corridor covers three diverse and economically strong Consolidated Metropolitan Statistical Areas (CMSA) as defined by the U.S. Office of Management and Budget. The following are counties or cities included in the CMSAs:

1. New York-North New Jersey-Long Island, NY-NJ-CT-PA (CMSA)
Bergen-Passaic, NJ
Dutchess County, NY
Jersey City, NJ
Middlesex-Somerset-Hunterdon, NJ
Monmouth-Ocean, NJ
Nassau-Suffolk, NY
New Haven-Bridgeport-Stamford-Danbury-Waterbury, CT
New York, NY
Newark, NJ
Newburgh, NY-PA
Trenton, NJ

2. Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD (CMSA)
Atlantic-Cape May, NJ
Philadelphia, PA-NJ
Vineland-Millville-Bridgeton, NJ
Wilmington-Newark, DE-MD

3. Washington-Baltimore, DC-MD-VA-WV (CMSA)
Baltimore, MD
Hagerstown, MD
Washington, DC-MD-VA-WV

Currently, the corridor has a population of over 34 million. The average household incomes are higher than the national average. The average household size is very close to the national average. However, these households own fewer vehicles than national average. This may be due to a better mass transportation system that allows them to use public transportation.

5.3.1.1 Transportation Systems. Highways, conventional rail, high-speed rail, airline services and intercity bus transportation currently serve this densely populated corridor. Comprehensive transportation networks of the corridor is showed in Figure 5.1. The massive transportation infrastructure and services are clearly in response to the large travel demand along this corridor. Currently, the NorthEast Corridor is facing a transportation capacity crisis. This causes severe and unacceptable congestion on the transportation network.
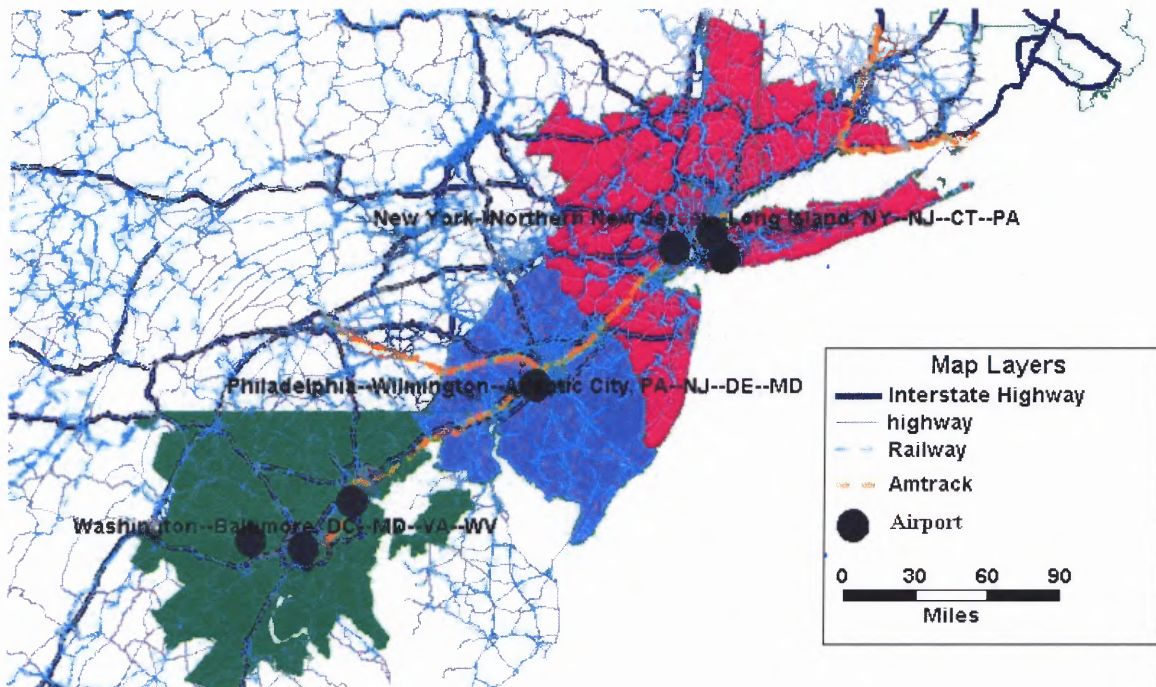
Figure 5.1 Comprehensive transportation networks of the corridor.

The highway network within this corridor is very dense. Millions of private automobiles and buses travel along these interstate highways and other major arteries. According to the annual mobility study of the Texas Transportation Institute (TTI) at the Texas A&M University issued in 1999, the areas suffer highway congestion. In the Washington DC, New York and Philadelphia areas, travel times on the roadways during peak periods were 41%, 30%, and 22% higher than under free flow conditions, respectively. They are among the nation's most heavily traveled roadways.

The most critical arterial roadways connecting Washington DC and New York is Interstate-95. Its design speed is 70-75 mph, and the engineers established a legal speed limit of 55 or 65 mph to allow for a margin of safety (Anderson, undated). The minimum numbers of freeway and tollway lanes are as follows: New York to Philadelphia, three lanes (New Jersey Turnpike); Philadelphia to Washington DC, five lanes (Interstate 95). According to the Federal Highway Administration (FHWA, 1999), the average annual daily traffic (AADT) volume on I-95 through the region exceeds 100,000 vehicles. In the Washington DC, Baltimore, Philadelphia, and Newark metropolitan nodes along I-95, traffic volume, which includes commuter traffic as well as intercity traffic, ranges from 175,000 to 200,000 AADT.

Amtrak is the only national passenger train operator. Today, trains make the trip between Washington DC and New York in as little as two hours thirty-five minutes. The Acela Express train sets operate at speeds between 110 and 150 miles per hour for much of the route. It is the busiest passenger line in the United States. Amtrak trains now carry more passengers between Washington DC and New York each day than do U.S. Airways and Delta shuttles combined. About 40,000 people ride Amtrak in the NorthEast

Corridor on a typical weekday, and more than 10,000 ride the Acela Express (King, 2002). A spike in rail travel after the 9/11/01 terrorist attacks is attributed to the appeal of Acela's shorter travel time, comfort, and safety. From city center to city center, the fastest trains are time competitive with airline services and considerably faster than automobiles or buses. This helps to alleviate highway and airport congestion.

Among the fifty busiest airports in the US, there are seven large international airports in this corridor: John F Kennedy International, LaGuardia, and Newark Liberty International Airport for the New York Metropolitan area; Reagan National, Baltimore-Washington International, and Dullas International airport for the Washington DC Metropolitan area; Philadelphia International for the Philadelphia Metropolitan area. They are the fastest growing airports on the east coast. All these airports face a critical shortage of runway capacity.

As an integral part of the nation's public transportation infrastructure for more than eighty years, intercity bus transportation also provides an affordable intercity passenger travel mode within the NorthEast Corridor, especially for those travelers who do not have access to other modes.

The major intercity bus operators include the following: Greyhound Lines, inc., the Trailways national bus system, and others. Greyhound Lines is the largest single carrier with a national network.

5.3.1.2 Travel Demand Characteristics. Due to the uniqueness of this corridor, such as different socio-economic characteristics of the population, different mass-transit systems, and available intercity travel modes, the market share of various modes in this

corridor is different from every other corridor in the nation and is presented in Figure 5.2.

The Amtrak market share in this corridor is much higher than the national average.
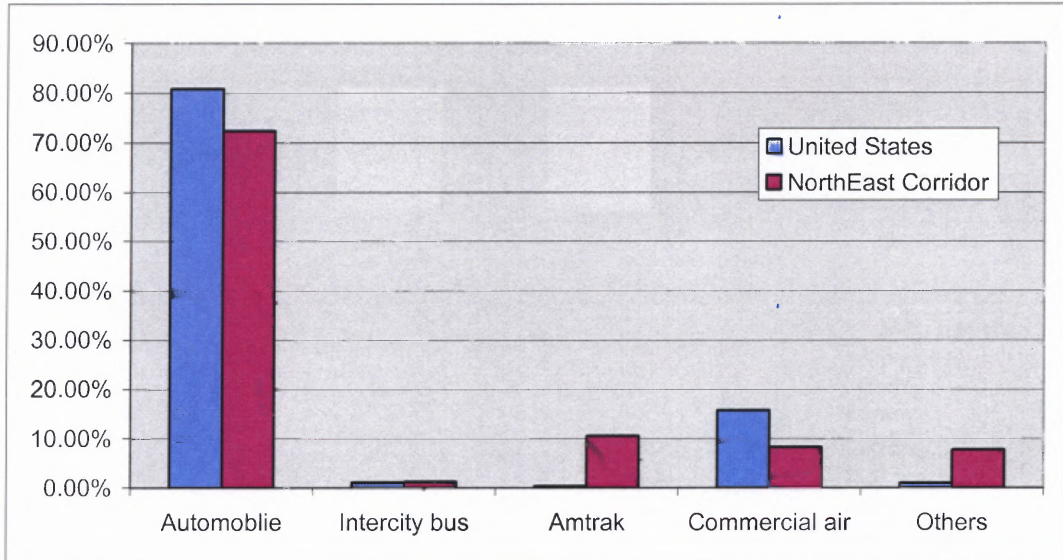
Figure 5.2 Intercity market shares of nation and corridor.

Another distinguishing point is that many travelers make more than one intercity trip within the study corridor. As shown in Figure 5.3, annually 48% of trip-makers make more than one trip. These trip-makers made 82% of the trips within this corridor. The percentage of the trip-makers that made more than five trips is 8.5%. These trip-makers made 40.5% of all trips. The 2.5% of the trip-makers who annually made more than twelve trips made 22.8% of all trips. Therefore, the high frequency trip makers' travel behavior cannot be neglected.

Trips and Individuals Distribution
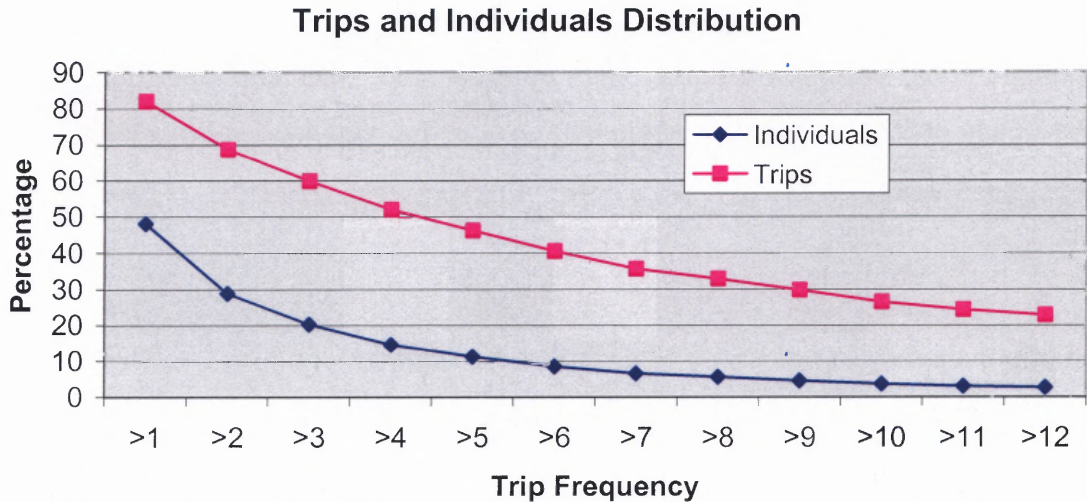
Figure 5.3 Trip frequency distribution of the study corridor.

5.3.2 Data Filtering

In this study, only the corridor from Washington DC to New York is analyzed. Hence, all of the trips, whose origin and destinations are within the New York MSA, Philadelphia MSA, or the Washington DC MSA, are chosen for this study.

Of the 256,486 personal trips in the dataset, 10,369 trips have both origin and destination within the study corridor. As defined by Kanafani (1983), the short-haul intercity travel distance is no longer than one thousand kilometers. Therefore, only the records which have a travel distance between 75-650 miles are kept which results in 10,241 trips. As mentioned earlier, the scope of this study covers only personal trips. If the reason for a trip is business, convention, conference, seminar, or school-related activity, the trip is excluded. This results in 8,155 personal trips. Excluding the minor intercity travel modes such as ship and bicycles, 3,088 respondents reported 8,125 trips. The vast majority of the respondents (82%) have only one destination for all the trips recorded in the survey calendar year. The total of these trips is 5,771. Although there are

persons who chose different modes for different trips to one destination, 97% of the travelers chose only one mode, which produced 5,390 trips. Due to the study cost and time limitations, the origin or destination representing less than 3% of all trips were excluded. This results in 4,637 trips, among which 22.8% were made by travelers who traveled more than 12 times a year, 35% were made by travelers who made more than 6 trips per year, and 60% of the trips were made by travelers who made more than 3 trips per year. Based on the resultant trip file, a person file is generated, where a new column, the number of trips made within the study corridor, is created. Five records were excluded because they had both the origin and destination in the same MSA. In addition, the records whose travel distance is less than 100 miles were excluded. At the end, there were 1,495 travelers and 3,295 trips in the file.

5.3.3 Data Preparation

Due to the constraints of time and cost, it was infeasible to collect new data. Instead, reliance has to be placed entirely on available sources. To model the intercity travel choice, the alternatives' attributes should be collected or estimated to supplement the 1995 ATS dataset. This is very common in data preparation for intercity travel demand modeling, as mentioned in Section 5.2.1. The scope of this section is to present the different sources used. The goal is to make the information contained in them compatible.

In the 1995 ATS, the available related information includes origin and destination, chosen mode, distance between origin and destination, mode and the distance to airport/station/terminal, and mode from terminal to destination. Because the detailed trip origin and destinations are at the metropolitan area level, it is impossible to collect

true access/egress data. However, average data can be generated according to the available data and some assumptions. In this study, level of service data, travel time and travel cost are generated for each available mode alternative for each trip based on the origin/destination information of the trip.

- Auto travel cost and time

Various average driving costs are assumed in the available literature. Rink (2003) reports that the average driving cost in 1995 was 41.2 cents/mile. This cost covers operating costs (gas and oil, maintenance and tires) and ownership costs (insurance costs). Trani and Teodorovic (2003) assume the average driving cost to be 31 cents per mile. Martland et al. (2002) chose 30 cents per mile for their model. Maryland Mass Transit Administration (2000) assumed 31.7 cents per mile as the driving cost. The average driving cost also can be computed as average personal auto expenditures, divided by average vehicle miles traveled according to the 1995 NPTS. The resulting average driving cost from this calculation is 19.9 cents per mile.

Considering the above average driving costs in the literature, 30 cents per mile is chosen as the average driving cost for this study. Multiplying the average driving cost with the travel distance, the travel cost for each automobile trip can be estimated. When used in a model, the cost is divided by travel party size.

With regard to travel speed, Trani and Teodorovic (2003) assume that driving speed on highways is 60 miles per hour (mph). Martland et al. (2002) used 50 mph. Considering the highway system in the corridor, 50 mph is assumed for this study.

• Public transportation

The total travel time using public transportation includes the access/egress time, waiting time, and line-haul time. Correspondingly, the travel cost includes the local cost and line-haul cost.

Alaska DOT (2001) assumed the driving speed access to a terminal to be 30 mph. The walk time to terminal is 7.5 minutes. Referring to the 1995 NPTS speed survey and the area's population and density, the average driving speed in an urban area is 34.61~34.89 mph. In this study, the driving speed access/egress to a terminal is assumed to be 30 miles per hour.

Waiting times, boarding, and alighting time are taken into account for all the access/egress modes. With the same access/egress travel time structure as in US DOC (undated), the different access/egress mode travel times are calculated in the formulas in Table 5.1. In this table, the distance is measured in mileage. The speed is 30 miles per hour.

Table 5.1 Travel Time Formulas for Different Access/Egress Modes

Access/Egress Modes	Access (Minutes)	Egress (Minutes)
Kiss'n ride/pick up	60*access distance/30	60*egress distance/30
Park'n ride/Rental	10+60*access distance/30	15+60*egress distance /30
Taxi	5+60*access distance /30	5+60*egress distance /30
Limousine/shuttle	15+60*access distance/30	20+60*egress distance /30
Transit	15+2*[60*access distance/30]	15+2*[60*egress distance /30]
Walk	10	7

With this access travel time structure in Table 5.1, the average travel time for access to different terminals can be computed for every trip origin PMSA, as in Table 5.2.

Table 5.2 Average Travel Time Between Home and Terminal (Minutes)

	Bus/Train Terminal	Air Terminal
Baltimore, MD PMSA	36.8	32.6
Bergen-Passaic, NJ PMSA	55.1	31.4
New York, NY PMSA	28.1	41
Newark, NJ PMSA	58.3	58
Philadelphia, PA-NJ PMSA	64.7	7.5*
Washington, DC-MD-VA PMSA	24.0	34.3
Wilmington, DE PMSA	35.3	24.2

*This number does not appear to be reasonable, because most of them reported the distance as 0.

No distance from terminal to destination is reported in the 1995 ATS. Therefore, all the trips are assumed to be destined to the city downtown. The estimated average distance in US DOC (undated) is used here.

Table 5.3 Average Distance from Different Terminals to Downtown (Miles)

	NY	Newark	Philadelphia	Wilmington	Baltimore	Washington
Airports	8.5	2.5	7.0	5.0	8.0	4.5
Rail/Bus Terminals	2.5	0.83	1.67	0.83	1.67	1.67

With the egress travel time structure in Table 5.1, the average travel time from terminal to destination can be calculated for each destination as showed in Table 5.4.

Table 5.4 Average Travel Time Between Downtown and Terminal (Minutes)

	Bus/train Terminal	Air Terminal
Atlantic-Cape May, NJ PMSA	24.2	21.8
Baltimore, MD PMSA	21.6	24.3
New York, NY PMSA	29.4	24.9
Philadelphia, PA-NJ PMSA	21.4	20.3
Washington, DC-MD-VA PMSA	30.5	24.8

The cost in dollars for the different access/egress modes are assumed to be as follows:

Auto, pickup or kiss 'n ride: 0.3*mileage

Auto, parking 'n ride: 20*(parking days) +0.3* mileage

Auto, rental: 45*days

Taxi: 2+1.25*mileage

Limousine: 2*mileage

Transit: 1.5 per person

Walk: free

With these assumptions, the access/egress cost is calculated for each terminal of every city and presented in Table 5.5.

Table 5.5 Access/Egress Costs for Each Terminal of Every City (Dollars)

	Bus/Train Terminal		Air Terminal	
	access	egress	access	egress
Atlantic-Cape May, NJ PMSA	2.4	4.8	*	15.4
Baltimore, MD PMSA	10.5	7.3	27.4	14.4
Bergen-Passaic, NJ PMSA	7.1	6.2	11.7	16.6
New York, NY PMSA	4.2	8.0	25.6	16.4
Newark, NJ PMSA	10.4	5.0	24.6	18
Philadelphia, PA-NJ PMSA	11.5	7.8	10	10.2
Washington, DC-MD-VA PMSA	5.6	7.7	14.8	19.4
Wilmington, DE PMSA	7.8	5.6	20.7	2.4**

* No records; **Only 1 record

For each public transportation mode, the standard one-way coach-type fare is used for line-haul time and cost estimation.

In the NorthEast Corridor, Greyhound is the most widely used city-to-city bus line. In the literature, the fare ranges from 0.39~4.96 dollars/mile (American Bus Association, 2001). However, this is not consistent with the fares from the www.greyhound.com. Therefore, the intercity bus travel time and travel cost are selected from www.greyhound.com. The ticket is reserved ten days in advance. Another benefit from using the website is that transfer information can be incorporated. A waiting time of 44 minutes for personal trips (US DOC, undated) is included.

The most popular intercity rail passenger carrier in the NorthEast Corridor is Amtrak. However, back in 1995, the Acela was not in service. The travel time and cost can be estimated using www.amtrak.com. The travel costs are comparable to the Martland et al. (2002) model: 25+0.3*Mileage. They are also close to $0.42/mile (the

Public Purpose, undated). Hence, the data from www.amtrak.com is adopted for this study.

An additional time is included for train travel which takes into account the waiting and boarding time. In Martland et al. (2002), the sum of the waiting, boarding, exiting from train, and exitiing from station, are 63 minutes. In US DOC (undated), the waiting time (including others) is 49 minutes. In this study, 60 minutes is considered the average process time from entry into the terminal to the scheduled departure time.

For airlines, various equations were used to calculate the travel time and cost (Trani and Teodorovic, 2003). However, the equations can not take transfer situations into consideration. Therefore, using www.expedia.com, the travel time and cost for air travel can be estimated between airports with most direct flights.

Waiting time, check in time, baggage claim time, exit from airplane and airports, were all considered in the literature. US DOC (undated) used 69 minutes. Skytrain (2003) assumed the whole process from waiting, boarding, and exiting from an airplane to take 75 minutes. The Martland et al. (2002) model used 90 minutes. Trani and Teodorovic, (2003) chose 90 minutes too. Given this information, a 90-minutes process time is assumed for the air transportation mode in this study.

Note that the travel costs presented so far are in 2003 dollars. With an inflation conversion factor, all costs can be converted to 1995 dollar. The inflation rate was obtained from the Federal Reserve Bank of Minneapolis. If, in 1995, goods or services were bought for one dollar, then in 2003, the same goods or services would cost 1.2 dollars. Thus the resultant costs are divided by 1.2.

As an example, the resultant travel time and travel cost of different modes for some Origin-Destinations are presented in Table 5.6.

Thus far, the prepared data include:

•   The socioeconomic characteristics of the travelers in the study area, including age, activity, vehicle ownership, household income, etc.;

•   The individual trips' characteristics, such as the chosen mode, origin and destination, party size, etc.;

•   The transportation supply's characteristics, such as the travel time and travel cost of all of the available modes for each O-D pair.

Table 5.6 Modal Attributes of Different Origin-Destination Pairs

OD pairs	Auto Time (minutes)	Auto Cost (dollars)	Bus Time (minutes)	Bus Cost (dollars)	Train Time (minutes)	Train Cost (dollars)	Air Time (minutes)	Air Cost (dollars)
Baltimore-New York	205	43	360	45	275	74	212	245
New York -Philadelphia	95	20	234	28	200	50	211	247
New York -Washington	246	51	363	39	311	70	216	138
Newark - Baltimore	194	41	344	44	284	73	252	208
Newark - Philadelphia	85	18	259	33	230	55	230	263
Newark - Washington	235	49	418	44	336	75	243	195
Philadelphia -Baltimore	110	23	264	44	213	52	169	245
Washington -Philadelphia	151	32	294	30	222	49	196	204

CHAPTER 6

CALIBRATION RESULTS AND ANALYSIS

The models presented in Chapter 4 are used to analyze the intercity travel behavior along the NorthEast Corridor using the data as prepared in Chapter 5. The scope of this chapter is to report and analyze the empirical results. The model validation is in Chapter 7.

Section 6.1 describes the determination of variables of the model. Section 6.2 presents the model evaluation methods. Section 6.3 gives the calibrated mode choice model. The trip generation model results are discussed in Section 6.4. Section 6.5 presents the resultant conceptual framework.

6.1 Determination of Variables

Travel time and travel cost are two important factors determining the mode choice. Even though distance may be one determinant of mode choice, the corridor context makes the distance variation limited. Travel cost is also directly related to distance. As a result, the distance variable is excluded.

As stated earlier, the dataset from the 1995 ATS is rich in traveler characteristics. However, most of them are categorical data, such as household income. They could be recoded as continuous variables (Aydemir, 2002; O'Neill and Brown, 1999). Referring to the literature (US DOC, undated; O'Nell, 2001), and for modeling convenience, some variables are recoded into new categorical data as listed in Table 6.1. Other variables, such as the household size and household vehicle number, are defined as continuous as in

86

the 1995 ATS dataset. These variables exert an influence on mode choice decision making.

Table 6.1 Variable Recoding for Calibration

Variables	Recoded Variables	Details
Household Income	HHINCL	Annual household income<=$25,000
	HHINCM	Annual household income<=$50,000
	HHINCH	Annual household income<=$100,000
	HHINCS	Annual household income>$100,000
Age	AgeL	Age<=20
	AgeM	Age<=45
	AgeH	Age<=65
	AgeS	Age>65
Race	Race1	White
	Race2	Black
	Race3	American Indian, Eskimo, or Aleut
	Race4	Asian or Pacific Islander
	Race5	Other race

Trip generation modeling requires the identification of factors that affect the personal level of demand for travel. For a long time, only socio-economic variables were entered in the disaggregate trip generation model because the socio-economic and demographic characteristics reflect the desire and potential for trips. All socio-economic factors available in the 1995 ATS are candidates for inclusion in the model. Generally, it is believed that household income might determine travel behavior. Other socio-economics including the household size, gender, education level, employment type, marital status and activity types also exert an effect, to different degrees.

The proposed trip generation model includes the chosen mode cost and selection term which reflects the quality of the transportation system. These characteristics of the transportation system impact the cost and ease of travel. A rapid, inexpensive, comfortable and convenient transportation system can cause travel to be economically and psychologically attractive. All other considerations being equal, more trips are generated using an attractive modal system than using an unattractive one. Improvements on one modal system may cause not only a diversion from other modes, but also an increase in demand.

Theoretically, all information associated with the individual, trip, and transportation system are potential variables. During the modeling, only the variables collected in the database are considered. These various factors impact the model, to different extents. The model identifies the most significant factors, which define the main aspects of the model. Therefore, a systematic process of eliminating variables is set for the specification. It includes the following:

- The causal relationship between the dependent variables and different independent variables is studied to determine the expected sign of the independent variable and compare it with the calibrated sign of the independent variable.

- T-test whether the single regression coefficient is significantly different from zero.

- Select variables based on a statistical index such as R-Square, adjusted R-Square.

- Meanwhile, consider the potential multicollinearity, which would otherwise cause many problems.

6.2 Model Evaluation

Different measures are utilized to assess the discrete choice model and continuous model. In the context of a logit model, the overall goodness-of-fit is measured by the McFadden Likelihood Ratio Index (LRI), which is the percentage increase in the log likelihood function above the value when the parameters are zero, and defined by

$$\rho = 1 - \frac{LL(\beta)}{LL(0)} \tag{6.1}$$

where:

$LL(\beta)$: Log of likelihood at the estimate, which is also the maximum value, and

$LL(0)$: Log of likelihood at zero, when all the parameters are set to zero, and which is also the initial value with which the iteration starts.

To exclude the effect from the inclusion of multiple independent variables, the ratio is adjusted to:

$$\rho = 1 - \frac{LL(\beta) - k/2}{LL(0)} \tag{6.2}$$

where:

k: the number of parameters.

As pointed out by experienced econometrists (Train, 2002), the likelihood ratio index is not at all similar in its interpretation to the R-Square used in regression, despite both statistics having the same range. R-Square indicates the percent of the variation in the dependent variable that is "explained" by the estimated model. The likelihood ratio has no intuitively interpretable meaning for values between the extremes of zero and one.

In comparing two model estimates on the same data and with the same set of alternatives, (e.g. LL(0) is the same for both models), it is usually valid to say that the model with the higher ρ explains the data better.

As for the log-linear regression model, overall fit and significance are used. Overall fit is used to measure how well the predictions match the observed number of trips. R-Square is used as an overall fit measure. The R-Square is computed using the following equation:

$$R^2 = 1 - \frac{\sum_i (y_i - Y_i)^2}{\sum_i (y_i - \overline{y_i})^2}$$
(6.3)

where:

y_i : observed dependent variable, here is log(trips),

Y_i : predicted dependent variable, and

$\overline{y_i}$: mean observed dependent variable.

As Equation 6.3 indicates, R-Square is used to describe the extent that the variation in dependent variables could be explained using the regression equation as opposed to simply using the mean value. Thus, there is no absolute basis for comparison of the R-Squares. Actually, in the aggregate model, the R-Squares are routinely high. However, in terms of the values one normally encounters in disaggregate models with cross-section data, the R-Squares are much lower. Sometimes it is noteworthy that R-Squares in cross sections of individual data are as high as 0.2 (Greene, 1997). As more independent variables are included in the equation, the R-Square naturally increases or

remains the same because each independent variable will explain some of the variation in the dependent variable. An adjusted R-Square tries to yield a more honest value excluding the influence of the number of the independent variables. It is computed using the formula:

$$R_a^2 = 1 - \left(\left(1 - R^2\right) * \left(\frac{N-1}{N-k-1} \right) \right) \tag{6.4}$$

where:

N: number of observations, and

k: number of independent variables.

P-value, the probability that the coefficients were not significant, is used to describe whether an independent variable is statistically significant. The null hypothesis of the t-statistic is that the coefficient is not significantly different from zero. For this test, a t-statistic is calculated from the following equation:

$$t_k = \frac{\left| \widehat{\beta}_k - 0 \right|}{SE(\widehat{\beta}_k)}$$

$$SE(\widehat{\beta}_k) = \frac{SEE}{\sqrt{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}} \tag{6.5}$$

$$SEE = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2}{n-2}}$$

where:

t_k : T-statistic for k[th] independent variable,

$\widehat{\beta_k}$: Estimate of kth independent variable,

$SE(\widehat{\beta_k})$: Standard error of the kth coefficient,

Y_i : Actual value of dependent variable at given x_i ,

$\widehat{Y}_i$: Predicted value of independent variable at x_i ,

n : Number of observation, and

SEE : Standard error of estimate, which is a statistical measure used to determine the variability of the actual dependent values from predicted values.

Based on the degrees of freedom (n-k-1), different critical t-values can be obtained from a t-distribution. Comparing the t-statistics to a series of t-values, the p-value can be determined. This p-value represents the level of significance.

Observing the predictions of the model given a particular set of values for the independent variables one may intuitively determine the model's predictive ability. Such a discussion is presented in Chapter 7.

6.3 Discrete Choice Model-Mode Choice

The intercity travel modes considered in this study are car, intercity bus, train, and air. There are many potential nest structures for the nested logit model with these four modes. Different nest structures were tested for this study.

6.3.1 Nest Structures

The nest structures that appeared in the literature are illustrated in Figure 6.1. Different nest structures imply different patterns of competitiveness among the alternatives. Of these six structures, some of them seem intuitively plausible, while others do not. For example, Nest A is composed of two nests. One nest is air transportation. The other nest is ground transportation which includes automobile, intercity bus, and train. If the ground transportation nest is further divided into public ground transportation and private transportation, it forms Nest F. If all the modes are separated into expensive nest and inexpensive nest, it leads to Nest E. The expensive nest includes airplane and train. The inexpensive nest is composed of automobile and intercity bus. If the modes are divided into ground public transportation and other groups, it forms Nest D. Even though Nest B or C was mentioned by Koppelman et al. (1998) and Forinash et al. (1993), they are not intuitively logical for this model. However, all the structures are estimated because Daly (1987) suggested that non-intuitive structures may be statistically superior. Using Statistical Analysis Software (SAS), the final estimation result of the nested logit model with car and air nested, bus and train nested, is showed in Table 6.2, which is recognized as a correctly specified. The estimation results for the other nested logit models are not shown because the scale parameters for the nest exceed one and are inconsistent with stochastic utility maximization (McFadden, 1978; Hunt, 2000; and Hensher and Greene, 2002). The estimated results of these nest structures suggest that some nest structures are incongruent with the nest structure prevailing in the structure of an individual's preference.
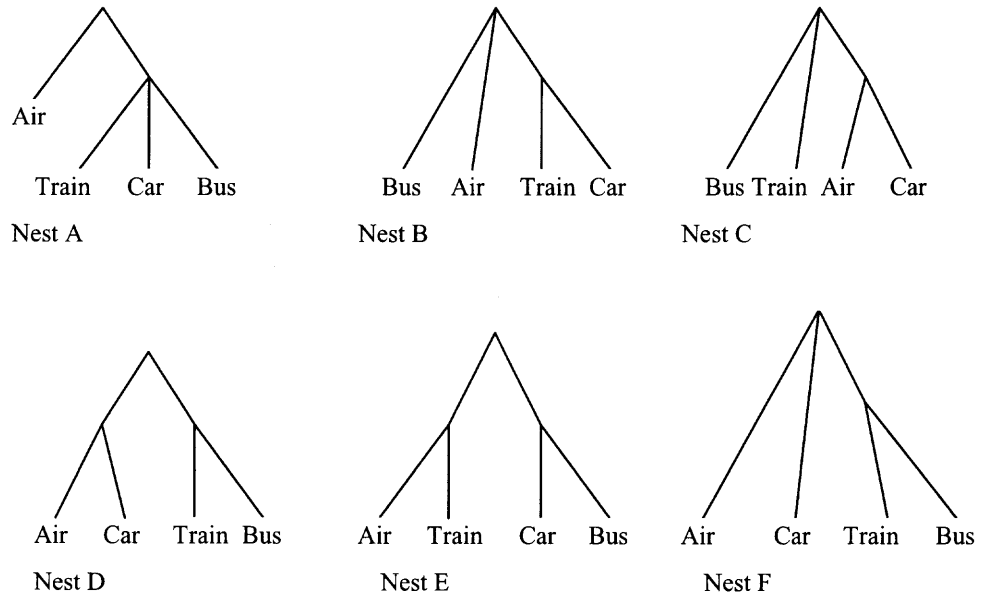
Figure 6.1 Different tested nest structures.

6.3.2 Model Results and Discussion

As proposed in Chapter 4, the utilities of different alternatives are specified as:

$$V_i = \alpha_i + \beta_i y + \sum_{m=1}^{M} \delta'_{im} s_m + \sum_{k=1}^{K} \gamma_k b_{ik} - \log p .$$
(6.6)

The automobile mode is used as the base for the alternative constants and alternative specific variables. Even though four constants are set for different modes, only the constant for bus is significantly different from zero. Therefore, the constants for train and airplane are also set to zero. During calibration, the coefficient of logp is allowed to be any negative number. Because variations exist for the travel time and cost of every mode, a common coefficient is set for all modes. In modeling the level of service variables, both travel cost and travel time, are tested. The p-value for travel time is 0.9804. It is plausible that travel time is less important for personal travelers in this dataset than travel cost. Another possible reason may be the collinearity between travel

time and travel cost. Hence, travel time is excluded in the final model. Many socioeconomic characteristics were considered for inclusion in the model. They are household types, household income, available vehicles, household size, gender, age, race, marital status and education level. The trip context characteristics tried include travel party size, and nightaway. Except for the independent variables included in Table 6.2, most of the other variables such as gender, race, marital status, education level have coefficients that are not significantly different from zero. Other variables such as age and travel party do not have the expected signs. For the alternative specific variables, such as vehicles, if the coefficient for an alterative isn't significantly different from zero, the coefficient of that variable specific to the alternative is set to zero. The variables that are not significant or have unexpected signs are excluded from the final result. Because of the limited capability of SAS 8.0, the scale parameter of the bottom nest is set to one.

Table 6.2 Intercity Mode Choice Model Results

Parameter	Estimate	p-value
Constant: Bus	0.73	0.0166
Log (price)	-0.53	<.0001
HHincM: Bus	-1.28	<.0001
HHincH: Bus	-2.60	<.0001
HHincS: Bus	-5.00	<.0001
Airplane	1.37	0.0001
Vehicles: Airplane	-0.67	<.0001
HHsize: Bus	-0.69	0.01
Train	-1.12	0.0001
Airplane	-0.43	<.0001
IV (Auto, Air)	0.52	<.0001
IV (Bus, Train)	0.46	<.0001
Number of Observations	1495	
Log Likelihood	-1094	
Likelihood Ratio (R)	1958	
Adjusted McFadden's LRI	0.469	

(1) The car mode is used as the base for the alternative constants and alternative specific variables.

Therefore, in the mode choice model, the utility function is:

$$V_i = \alpha_i + \beta_i y + \delta_i \, Vehicles + \eta_i \, Hhsize - \upsilon \log(price)$$

$$\alpha_i, \beta_i, \delta_i, \eta_i \quad are \quad alternative-specific \quad parameters \qquad (6.7)$$

$$\upsilon \quad is \quad common \quad parameter$$

Travel cost is statistically significant for explanation of the mode choice decision-making. The negative coefficient of travel cost implies a negative elasticity of demand. The model predicts lower probability that the alternative will be chosen as the price of this alternative increases.

Household income, household size, and household vehicles are the socioeconomic variables that affect the model goodness-of-fit in a significant way. They have the expected signs. However, they affect the utility of different alternatives in different ways. The income parameters show that the highest income group favors air travel relative to others, and the lowest income group favors bus travel. As compared to small households, the travelers from bigger households are more likely to drive a car for intercity travel.

The two scale parameters are within the reasonable range from zero to one, which means the model here is consistent with random utility maximization theory. These coefficients actually indicate the correlations between the alternatives within the nests (Hunt, 2000). In the nest of car and air, the correlation is $1 - 0.52^2 = 0.73$. The correlation in the nest of bus and train is $1 - 0.46^2 = 0.79$.

6.4 Continuous Model-Trip Generation

As derived in Chapter 4, the proposed trip generation is specified as follows:

$$\log x_{Pni} = a_0 - a_1 * \log p_{pni} + a_2 y + \sum_{m=1}^{M} a_{3m} s_m +$$

$$(\rho - 1)[\ln(\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qlj}/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}) + 0.5772] + \tau \qquad (6.8)$$

In this model, the price of the chosen mode is included in a logarithm form. The coefficient of log(price) is expected to be negative. The household income and other socioeconomic characteristics are the independent variables in linear form. The coefficients of household income are expected to be positive. A selection term defining the interrelationship between mode choice and trip generations is included in the proposed trip generation model. The selection term is the following:

$$\ln(\sum_{q=1}^{Q} a_q \left[\sum_{l=1}^{K} b_l \left[\sum_{j \in B_l} e^{V_{qlj}/\mu_l} \right]^{\mu_l/\lambda_q} \right]^{\lambda_q}) + 0.5772$$

$$(6.9)$$

where:

V : The utilities resulting from the nested logit model, and

μ and λ : the calibrated or specified parameters in the nested logit mode.

For estimation convenience, the position parameters a_q and b_l are set to one. Therefore, the selection term is an inclusive value plus 0.5772. With estimates from the nested logit model, the inclusive value can be calculated. If the proposed model is correctly specified, the selection term should be significant in the trip generation model and the coefficient of the selection term should be within an interval from -1 to 0.

To facilitate a comparison, the proposed trip generation model with a selection term, and a classical trip generation model without a selection term are calibrated. The two models are specified in log linear form and have the same set of independent variables except the selection term.

Table 6.3 Trip Generation Model Results

Parameter	Model with a Selection Term		Model without a Selection Term	
	Estimate	p-value	Estimate	p-value
Intercept	1.74	<.0001	1.55	<.0001
RACE1	-0.42	0.0012	-0.37	0.0021
RACE2	-0.28		-0.24	
RACE3	-0.73		-0.68	
RACE4	-0.48		-0.42	
RACE5	0.00		0.00	
HHINCL	-0.24	0.0013	-0.29	0.0024
HHINCM	-0.10		-0.12	
HHINCH	-0.05		-0.07	
HHINCS	0.00		0.00	
AgeL	-0.12	0.0012	-0.13	0.0007
AgeM	-0.08		-0.05	
AgeH	0.09		0.11	
AgeS	0.00		0.00	
HHSIZE	-0.05	<.0001	-0.06	<.0001
Selection term	-0.83	<.0001		
Log(price)	-0.18	<.0001	-0.12	<.0001
Adjusted R-Square	0.126		0.063	

As presented in Table 6.3, the trip generation model having a selection term has a better goodness-of-fit measure than the model without a selection term. The signs of different independent variables are similar to each other. All the variables have p-values suggesting that the null hypothesis that the parameter estimate is equal to zero can be rejected at a confidence level of 99 percent.

The constants represent the average effect of all the factors that influence trip generation, but are not included in the model. The positive values mean that the travelers prefer to make an intercity trip.

Intercity trip generation is heavily influenced by characteristics of the individual and the household. It is noticeable that travelers from smaller households make more intercity trips during a given period than those from bigger households. Possibly this increase is due to the fact that the trip generation is for individuals and larger households may have more constraints with respect to disposable income and time. The coefficients for ages imply that younger people are less likely to take an intercity trip in a given period. Persons between 45 and 65 years old make more intercity trips than other groups. Different races also differ in trip generation. From the coefficients for different income groups, the higher the household income, the higher the coefficients. This trend is consistent with economic theory according to which trip generation is expected to increase with income.

Level of service also affects trip making. This type of variables included in this model is price and selection term. As in the mode choice model, the log (price) is very significant in explaining trip generation. The log (price) has the expected negative sign, which implies that the higher the price, the fewer the travelers make intercity trips. This is consistent with economic theory. Price elasticities of trip generation are -0.18 and -0.12, for the models with and without a selection term, respectively. Taking the selection term into account leads to an increase of the price elasticity. The selection term is highly significant in the models. Also, as expected, the coefficient of the selection term is well inside the prescribed interval of -1 to 0. The absolute value of the coefficients shows the

extent that the trip generation is affected by the mode choice. As argued earlier, the omission of this variable in standard models implies a misspecification bias.

6.5 Resultant Model Framework

A conceptually valid model was proposed in Chapter 3 and Chapter 4. Figure 6.2 is an illustration of the resultant model framework explaining the choice of mode and trip generation.

In this structure, there is a relationship between the mode choice and trip generation. Characteristics such as race and age are allowed to influence the trip generation directly, measured by their respective parametric estimates. The model includes observable variables that may influence the mode choice, such as household income, household size, and prices of mode alternatives. These characteristics are also allowed to influence the trip generation directly, measured by their respective parametric estimates, and indirectly through their effect on the selection term. In addition, the observable characteristics such as household vehicles and unobservable characteristics related to the traveler preference of mode choice also exert an impact on trip generation via the selection term. Therefore, the selection term incorporates the information not only from quality of the transportation system, but also the travelers' characteristics. The calibrated result shows that this term is significant in explaining the trip generation. The inclusion of the selection term improves the trip generation model theoretically and statistically.
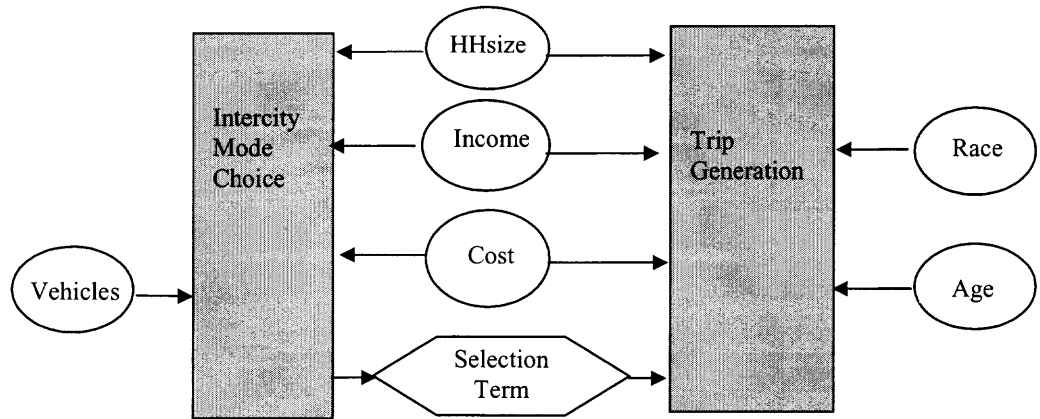
Figure 6.2 Resultant intercity travel behavior framework.

Chapter 7

VALIDATION AND POLICY IMPLICATIONS

This chapter presents the validation process and the result for the calibrated model in Chapter 6. Before being used to produce future year forecasts or transferred to other regions, a model needs to be verified to insure that it is able to replicate observed conditions within reason. The process of verifying a calibrated model in this manner is commonly termed "validation". The validation process involves checking for reasonableness and predictive ability.

7.1 Model Validation Process

Travel demand modeling consists of data preparation, model calibration, model validation and model application, as shown in Figure 7.1. There exist some feedback between application, validation and calibration. In some literature, there is model estimation before model calibration. However, here the entire process of developing travel models is termed calibration and includes statistical estimation procedures and parameter values adjustment.

Data Preparation ⟶ Calibration ⟶ Validation ⟶ Application
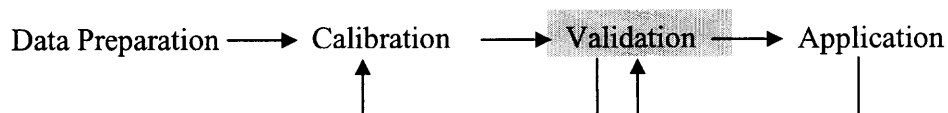
Figure 7.1 A typical travel demand modeling process.

The first method of validation is a reasonableness check. Usually it includes the estimated coefficients, theoretical and logical expectation checking, the statistical strength of the independent variable coefficients, which is defined by p-values in this study, observation of prediction ranges of the model for particular values on the independent variables, and the consistency of the calibrated results with the assumptions and theory used to formulate them. Most of the reasonableness check is done in the model calibration stage.

The validation also includes sensitivity analysis, which refers to the response of travel behavior to the transportation system, socioeconomic or policy changes. Often, the elasticity is used to express sensitivity. A common elasticity analysis for mode choice is performed using a direct or cross elasticities. They are used to estimate the percent change in demand given a percent change in supply. The elasticity analysis of trip generation usually refers to the household income or price elasticity.

To test the ability of the model to predict future behavior, validation requires comparing the model predictions with information other than that used in estimating the model. Thus, this rigorous validation process involves two independent samples. The independency of the two data sources is to ensure a sufficient validation. However, the endeavor to collect two independent datasets is seldom feasible. Under this circumstance, the second-best approach is splitting the available sample into two half samples randomly. One sample is used to for calibration. The calibrated models are used to predict the second sample's travel behavior. Different data splitting methods can be found in Watson (1973) and Meyer and Miller (2001).

Given two independent datasets, there are two ways to validate the models. First, models developed from sample one can be used to predict the behavior of sample two and the predicted behavior can then be compared with observed behavior. Second, the parameters of the models estimated on sample one and two are compared. In most cases, the first method is used. The key in the validation process is how to define how well the estimated values are reasonably similar to the observed values. For the mode choice model, different researchers utilized various measures to evaluate the discrepancy between predicted and actual values (Watson, 1973, 1974; Stephanedes, Kumar, and Paumanabhan, 1984; and MacFadden and Talvitie, 1977). In this study, the Prediction Absolute Error (PAE) and total percent correct are chosen for the validation of the mode choice model. Analogously, the correlation between predicted and actual values is used for the validation of the trip generation model. The details are discussed in the next two sections.

7.2   Mode Choice Model Validation

As mentioned in Section 7.1, the first validation for the mode choice model can be done with observation of prediction ranges for particular values on the independent variables. Table 7.1 presents the calculation results based on a representative intercity personal traveler from New York to Washington DC, whose household size is 3, the vehicles available are 2, the traveler is white and 20~45 years old.

Table 7.1 Mode Choice Prediction for a Representative Traveler

Mode	Low Income	Middle Income	High Income	Very High Income
Car	66.2%	75.8%	81.3%	79.6%
Bus	29.5%	16.7%	7.3%	0.9%
Train	2.9%	5.9%	9.6%	12.7%
Air	1.5%	1.7%	1.8%	6.9%

Table 7.1 shows that the high income group has the highest probability to travel by automobile. Because low income and middle income groups have fewer vehicles, they have lower probability to use automobiles for intercity travel. The very high income group can afford the airplane and train. Therefore, they choose the automobile less than the high income group. Generally the calibrated probabilities are consistent with the expected trend.

The second method to validate a nested logit is to estimate models with the same specification to the calibration sample and complete sample. If the estimates and specifications are both accurate, then the estimations obtained from the two samples should be similar. Comparison of estimates not only provides a test of the accuracy, but indicates the problem of the discrepancies between predicted and observed values.

Table 7.2 Mode Choice Models (Calibration Sample vs. Complete Sample)

Parameter	Complete sample		Calibration sample	
	Estimate	p-value	Estimate	p-value
Constant: Bus	0.73	0.0166	0.66	0.121
log(price)	-0.53	<.0001	-0.56	<.0001
HHincM: Bus	-1.28	<.0001	-1.06	0.0147
HHincH: Bus	-2.60	<.0001	-2.65	<.0001
HHincS: Bus	-5.00	<.0001	-5.08	<.0001
Airplane	1.37	0.0001	1.47	0.0024
Vehicles: Airplane	-0.67	<.0001	-0.52	0.0293
HHsize: Bus	-0.69	0.01	-0.43	0.1218
Train	-1.12	0.0001	-0.85	0.0053
Airplane	-0.43	<.0001	-0.50	0.0013
IV (Auto, Air)	0.52	<.0001	0.42	0.0192
IV (Bus, Train)	0.46	<.0001	0.49	0.0009
Number of Observations	1495		748	
Log Likelihood	-1094		-553	
Likelihood Ratio (R)	1958		968	
Adjusted McFadden's LRI	0.469		0.461	

The parameter estimates in Table 7.2 are close to each other in the calibration and complete samples. This closeness indicates the stability of the model specification. The difference in household size coefficients may be due to the sample size and the sample splitting method. The p-values for independent variables from the calibration sample are the same or higher than that those from the complete samples. This may also be due to the sample size difference.

The other validation method for the discrete choice model is to compare the actual and predicted riderships. The adopted indices are Prediction Absolute Error and total percent correct at the aggregate and disaggregate level, respectively.

Prediction Absolute Error (PAE), a statistic that allows a quick comparison of the predictive ability of models, is defined as the absolute difference between the predicted and actual ridership divided by the observed ridership.

$$PAE = \frac{\sum_{m}^{M} |Estimated\_ridership - observed\_ridership|_m}{\sum_{m}^{M} observed\_ridership_m} \qquad (7.1)$$

Where:

m: mode alternatives.

Table 7.3 is the prediction success table. If i denotes the row and j denotes the column in the table, the ijth element of the table is the number of person who actually chose mode i and were predicted to choose mode j. The percent correct for an alternative is the element in the diagonal for a particular column divided by the column total. The total percent correct is the sum of the elements in the diagonal divided by the total number. This is the percentage of the entire sample that is correctly predicted.

Table 7.3 presents the predicted and actual riderships for different alternatives. The total percent correct of this model is 59.0%. However, the PAE of the calibrated nested logit model, 0.99%, indicates that the predicted share is very close to the observed market share. Therefore, this model is very good at an aggregate level.

Table 7.3 Prediction Success Table

	Auto	Bus	Train	Air	Observed Total	Actual Share
Auto	419	59	58	19	556	74%
Bus	60	10	10	4	84	11%
Train	58	8	11	4	81	11%
Air	21	4	2	1	27	4%
Predicted Total	558	81	81	28	748	
Predicted Share	75%	10%	11%	4%		
Percent Correct	75%	13%	14%	3%		
Total Percent Correct	59.0%					
PAE	0.99%					

7.3 Trip Generation Model Validation

The first disaggregate validation check for trip generation is the total average trip rates. The surveyed average trip rate within the Northeast Corridor is 2.19 trips per traveler. The validated sample' trip rate is 2.18 with the model having a selection term. The trip rate without the selection term is 2.15. The low prediction may be attributed to a lot of records in the data with a trip frequency of one or two. However, this indicates that the model with a selection term performed better than the model without a selection term

Similar to the mode choice model validation, the calibrations from calibration sample and complete sample can be compared to check the specification's stability. Results in Table 7.4 show that the coefficients have the same signs. The differences may be caused by the splitting method.

Table 7.4 Trip Generation Models for Calibration vs. Complete Samples

Parameter	Complete sample		Calibration sample	
	Estimate	p-value	Estimate	p-value
Intercept	1.74	<.0001	1.78	<.0001
RACE1	-0.42	0.0012	-0.55	0.0039
RACE2	-0.28		-0.42	
RACE3	-0.73		-0.74	
RACE4	-0.48		-0.60	
RACE5	0.00		0.00	
HHINCL	-0.24	0.0013	-0.26	0.0023
HHINCM	-0.10		-0.09	
HHINCH	-0.05		-0.03	
HHINCS	0.00		0.00	
AgeL	-0.12	0.0012	-0.14	0.1303
AgeM	-0.08		-0.10	
AgeH	0.09		0.06	
AgeS	0.00		0.00	.
HHSIZE	-0.05	<.0001	-0.07	<.0001
Selection term	-0.83	<.0001	-0.75	<.0001
log(price)	-0.18	<.0001	-0.17	<.0001
Adjusted R-Square	0.126		0.131	

The last test for the trip generation model validation is to calculate the coefficient of determination (R-Square) to evaluate the relationship between the observed and predicted trips at the individual level. The R-Square in model calibration indicates the model's success in fitting the logarithm of the observed trip rate. Consequently, as Tardiff (1977) suggested, a coefficient analogous to the R-Square of the linear model will be calculated as follows:

$$R_a^2 = 1 - \frac{\sum \left(Actual\_Frquency - Predicted\_Frequency\right)^2}{\sum \left(Actual\_Frequency - Mean\_of\_Frequency\right)^2} \tag{7.2}$$

The results of Table 7.5 indicate that the proposed model with a selection term is much better than the classical model without a selection term regardless of whether the comparison is done in terms of R-Square or adjusted R-Square.

Table 7.5 Different R-Squares for Models

	Model with a selection term	Model without a selection term
R-Square	0.122	0.054
Adjusted R-Square	0.049	0.015

7.4 Elasticities and Policy Implications

The private corridor transportation carriers can employ the model developed here to estimate the likely consequence of different market strategies. Public agencies also can use this model to conduct a feasibility study of large investment projects. Different elasticity analysis is one of the methods to provide valuable information for private and public agencies. As an example, here the analysis focuses on the price elasticities.

As the resultant model indicates, changes in travel cost affect both the trip frequency and the mode choice. For this reason, the elasticities of the choice probability and trip frequency with respect to travel cost are reported.

There is a direct elasticity and cross elasticities for a nested logit model. Table 7.6 lists the choice elasticities formula. These are the elasticity of the probability of choice with respect to the change of price for train. The direct elasticity is the change of the probability of using the train with respect to the change of the train price. The cross elasticities are the changes of the probabilities of choosing automobile, intercity bus, or airplane with respect to the change of the train price.

Table 7.6 Cost Elasticities of Nested Logit Model

Direct elasticity	Cross elasticity
n not in nest $(1-P_n)\beta$	n and n' not in same nest $-P_n\beta$
n in nest m $[(1-P_n)+(\dfrac{1}{\mu_m}-1)(1-P_{n|m})]\beta$	n and n' in the nest m $-[P_n+(\dfrac{1}{\mu_m}-1)P_{n|m}]\beta$
Where; P_n is the probability of choosing train. β is the estimated coefficient of log(price). $P_{n|m}$ is the probability of choosing the train conditional on the choice of nest m. μ_m is the scale parameter of nest m.	

Note that these elasticity formulas are different from those in the literature (Wen and Koppellman, 2001) because the cost is included in the utility in a logarithmic form. The calculated results are presented in Table 7.7. The calculation is based on a representative intercity personal traveler from New York to Washington DC, whose household size is 3, the vehicles available are 2, the traveler is white and 20~45 years old.

For automobile travel, the travel party size is 3.5. The elasticity is calculated with respect to the change in the train's cost.

Table 7.7 Resultant Elasticities

	Low Income	Middle Income	High Income	Very High Income
Car	0.015	0.031	0.051	0.067
Bus	0.070	0.193	0.404	0.649
Train	-1.082	-0.960	-0.748	-0.503
Air	0.015	0.031	0.051	0.067

Table 7.7 shows that, for the low income group, one percent increase in train price leads to 1.082 percent decrease of train patronage, 0.070 percent increase of bus use, and 0.015 percent increase in car or air patronage. The elasticities in Table 7.7 signal that intercity personal travel emphasizes travel cost. Therefore, for the policy-maker, the most efficient way to increase the rail mode share is via reducing the rail travel cost.

The elasticities for different income groups indicate that the lower income households have the higher direct elasticity with respect to rail cost. As a result, from the view point of attracting riders for rail, the marketing should target the lower income households. From the cross elasticities and direct elasticity in Table 7.7, when the price of train increases, the most affected mode is the bus. Of all the shifted travelers, most of them shift from train to intercity bus.

The elasticity of trip generation with respect to a certain variable, such as the cost, is defined as the rate of change of trip generation with respect to that variable. In the model calibration for trip generation, the coefficients for log(price) are common for different income groups. Therefore, there is only one coefficient for the trip generation model with or without a selection term. For the trip generation model with a selection term, the elasticities are not exactly the coefficient of log(price) because the elasticity is

attributed not only to log(price), but also to the selection term. To avoid a complicated derivation, here only the calculated trip generation elasticities are presented. The elasticities for different household income groups are listed in Table 7.8. With a classical regression model without a selection term, different groups have the same elasticities due to the price change. This does not appear to be consistent with reality. The model with a selection term indicates that the low household income group has higher elasticity to the price change, than the other groups. This means that the price reduction exert more influence on the low income groups than other groups. The introduction of the selection term into the model increases the elasticities. Therefore, the classical regression model underestimates the trip rate change attributable to the price change. Consequently, the patronage from a rail price change policy is underestimated.

Table 7.8 Trip Generation Elasticities for Different Household Income Groups

	Classification	Model with a Selection Term	Model without a Selection Term
Elasticity to Price Change	Low Income	0.17	0.12
	Middle Income	0.17	0.12
	High Income	0.16	0.12
	Very High Income	0.15	0.12

CHAPTER 8

CONCLUSIONS AND FURTHER STUDIES

A conceptual intercity travel framework has been proposed in this dissertation. Corresponding to the conceptual framework, a nested logit/continuous model has been developed. The nested logit model component of the nested logit/continuous model formulates travelers' mode choice. The continuous component models trip generation. Both components are derived from one indirect utility function. Thus, an interrelationship between these two components has been established.

This model was tested within the southern end of the NorthEast Corridor. The data used were extracted from the 1995 ATS data. Because this dataset lacks the level of service of intercity alternatives, it is supplemented with travel cost and travel time, which is estimated from secondary sources. The mode choice and trip generation models are calibrated in two steps. Derived from the mode choice model, the selection term is included in the trip generation model to indicate the extent that the trip generation is affected by mode choice. The results indicate that the structure used in this study can be used to model intercity passenger travel demand.

8.1 Mode Choice Model

The utilities for different alternatives are slightly different from those in the literature due to the specification of the indirect utility function. The costs of the alternatives are included in logarithmic form. Different nested logit structures were tested. The

goodness-of-fit measure indicates that the proposed model is acceptable. All the independent variables have the expected signs with significant values.

8.2 Trip Generation Model

In the model proposed here, the trip generation model has, as its independent variables, not only socio-economic factors, but also factors of cost of the chosen mode and a selection term. For comparison purposes, a traditional regression trip generation model was also calibrated. The results indicate that the proposed model has much better goodness-of-fit measures than the classical one. The selection term and cost of the chosen alternatives are highly significant. Every independent variable, except for the household size, has the expected sign, similar to the sign of previous studies. The possible explanation for the unexpected sign of household size is that the trip generation in this study is at the individual level. This model incorporates the interrelatedness between trip generation and mode choice in the framework and model.

Overall, the model estimation results are consistent with the proposed conceptual framework and the corresponding model structure.

8.3 Contributions

The methodology and findings of this study can theoretically and empirically improve the intercity demand modeling in corridor, statewide, regional or nation-wide transportation planning. The proposed model fills a gap in the literature of intercity travel demand analysis. With this model, a complex set of intercity travel subdecisions can be integrated together. The optimal trip frequency and mode choice provide one maximized

utility of traveler's decisions in combination. This utility consistency makes it possible to incorporate interrelatedness and simultaneity in modeling.

The developed model can also be employed by private carriers to estimate the likely consequences of particular policies and investment planning. The travel cost and selection term, a transportation system quality index, is included in the trip generation model. Consequently, the impact of policy changes can be reflected in both mode choice and trip generation models. Therefore, the possible responses of travelers, both in mode shift and induced demand, are included in the model. Thus, this model improves the accuracy of patronage forecasting of different carriers. This accuracy can be further reflected in their revenue forecastings.

8.4 Future Studies

Although the data prepared are enough for testing the proposed model structure, the results could be further improved. Limitations may result from the access/egress data deficiency, i.e. incapable of obtain detailed trip origin and destination data.

Another limitation is the characteristics of alternatives. The factors of transportation supply affecting mode choice and trip generation should not be confined to travel time and travel cost. More level of service measures would be beneficial if they were included in the model.

Another problem with the current data is that the data do not contain samples for non-travelers. When the study scope is extended to the general population, this problem can be circumvented.

In addition, in the model proposed here, travelers are assumed to use only one mode within the study period. However, it is possible that, in different situations, several modes are used. This type of traveler behavior needs to be analyzed.

Research based on the above suggestions would improve future intercity travel modeling theoretically and empirically and could provide additional insights to transportation agencies. It would also improve the decision-making process of government and private carriers.

APPENDIX A

LIST OF VARIABLES IN THE FINAL DATA FILE

Variables	Description
PERSONTRIPS	Trips made within the corridor within one year
OMETNAME	MSA of trip origin
DMETNAME	MSA of trip destination
VEHICLES	Total number of vehicles owned
HHSIZE	The number of persons within one household
HHINC	Total household income
AGEHH	Age of householder
EDUCHH	Education attainment of householder
AGE	Age of the respondent
RACE	Race of the respondent
SEX	Sex of the respondent
MARITAL	Marital status of the respondent
EDUCATN	Education attainment of the respondent
PERSINC	Personal income
TRPARTY	Average travel party
NITAWAY	Average night away from home
TRANSOD	Travel Mode
RTEDUSOD	Travel Distance
TIME	Travel time
COST	Travel monetary cost

APPENDIX B

MEAN AND STANDARD DEVIATION FOR SELECTED VARIABLES

Variable	Mean	Standard Deviation
PERSONTRIPS	2.20	2.71
VEHICLES	1.72	1.39
HHSIZE	2.95	1.39
HHINC	2.64	0.89
AGEHH	2.59	0.70
EDUCHH	4.56	1.96
AGE	2.28	0.88
RACE	1.53	0.99
SEX	1.53	0.50
MARITAL	2.58	1.80
EDUCATN	3.86	2.17
TRPARTY	4.05	5.06
NITAWAY	2.13	2.50

REFERENCES

Abdelwahab, Walid M; Innes, J. David; and Stevens, Albert M. (1992). "Development of Disaggregate Mode Choice Models of Intercity Travel in Canada". Canadian Journal of Civil Engineering. Vol. 19, pp. 965-974.

Abdelwahab, Walid M. (1998). "Elasticities of Mode Choice Probabilities and Market Elasticities of Demand: Evidence from a Simultaneous Mode Choice/Shipment-Size Freight Transportation Model". Transportation Research E. Vol. 34, No. 4, pp. 257-266.

Abu-Eisheh, S. and F. Mannering. (1988). "Discrete/Continuous Analysis of Commuters' Route and Departure Time Choice". Transportation Research Record. No. 1138, pp. 27-34.

Adler, T.J. and M. Ben-Akiva. (1975). "A Joint Frequency, Destination, and Mode Choice Model for Shopping Trips". Transportation Research Record. No. 540.

Alaska Department of Transportation and Public Facilities. (2001). Transit Alternatives Analysis: New Seward Highway Corridor Study. Prepared by CH2MHILL and Cambridge Systematic Inc. http://projects.ch2m.com/Sewardhwy/TransitAlts.pdf.

Algers, Staffan. (1993). "Integrated Structure of Long Distance Travel Behavior Models in Sweden". Transportation Research Record. No. 1413, pp. 141-149.

Aljarad Saad N. (1993). Disaggregate Mode Choice Modeling for Intercity Non-Business Travellers in the Saudi Arabia-Bahrain Corridor. Dissertation. Indiana University

Amemiya, T. (1973). "Regression analysis when the dependent variable is truncated normal". Econometrica. Vol. 41, pp. 997-1016.

American Bus Association. (2001). 2001 Industry survey. Destinations Magazine 2001. http://www.buses.org/industry/IndustrySurvey.pdf.

Anderson, Steve. (undated). "New Jersey Turnpike: Historic Overview". http://www.nycroads.com/roads/nj-turnpike/

Arora, Neeraj; Allenby, Greg M.; and Ginter, James L.. (1998). "A Hierarchical Bayes Model of Primary and Secondary Demand". Marketing Science. No. 17, pp. 29-44.

Aydemir, Zava. (2002). Household Gasoline Demand in the U. S.: An Estimation of a Multiple-Discrete/Continuous Choice Model. http://www.eea-esem.com/eea-esem/esem2002/prog/getpdf.asp?pid=1383&pdf=/papers/eea-esem/esem2002/1383/gasdemand140202.pdf.

Baltas, George. (1998). "An Integrated Model of Category Demand and Brand Choice". Journal of the Market Research Society. Vol. 48, pp. 295-306.

Banai-Kashani, Alireza. (1984). "Travel Demand (Modal Split) Estimation by Hierarchical Measurement". Journal of Advanced Transportation. Vol. 18, No. 1, pp. 37-54.

Ben-Akiva, Moshe E. and Lerman, Steven R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge, Massachusetts.

Ben-Akiva, Moshe E. (1974). "Structure of Passenger Travel Demand Models". Transportation Research Record. No. 526, pp. 12-18.

Ben-Akiva, Moshe E. (1974). "Alternative Travel Behavior Structures". Transportation Research Record. No. 526, pp. 26-42.

Ben-Akiva, Moshe; McFadden, Daniel; Train, Kenneth, et al. (2002). "Hybrid Choice Models: Progress and Challenges". MEA Discussion Paper Series from Mannheim Research Institute for the Economics of Aging, University of Mannheim.

Bhat, C. R. (1995). "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice". Transportation Research B. Vol. 29, No. 6, pp. 471-483.

Bhat, C. R. (1997). "An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel". Transportation Science. Vol. 31, No. 1, pp. 34-47.

Bhat, C. R. (1998). "A Model of Post Home-Arrival Activity Participation Behavior". Transportation Research B. Vol. 32, No. 6, pp. 387-400.

Bhat, C. R. (2001). "Modeling the Commute Activity-Travel Pattern of Workers: Formulation and Empirical Analysis". Transportation Science. Vol. 35, No. 1, pp. 61-79.

Billeheimer, John W. (1972). "Segmented, Multimodal, Intercity Passenger Demand Model". Highway Research Record. No. 392, pp. 47-57.

Blackburn, Anthony. (1970). A Non-Linear Model of the Demand for Travel. Chapter 8 of "The Demand for Travel: Theory and Measurement". pp. 163-180.

Bolduc, Denis and Sangam, Dimitri. (1999). "Two-Step Estimation of Discrete/Continuous Econometric Models with Interdependent Multinomial Choices". Society for Computational Economics in its series Computing in Economics and Finance '99 with number 1323.

Boyce, David E., and Zhang, Yu-Fang. (1997). "Calibrating Combined Model of Trip Distribution, Modal Split, and Traffic Assignment". Transportation Research Record. No. 1607, pp 1-5.

Brand, Daniel. (1973). "Separable and Simultaneous Travel-Choice Behavior". Behavioral Demand Modeling and Valuation of Travel time, TRB Special Report 149, pp. 187-206.

Brand, Daniel, et al. (1992). "Forecasting High-Speed Rail Ridership". Transportation Research Record. No. 1341.

Buckeye, Kenneth R. (1992). "Implications of High-Speed Rail on Air Traffic". Transportation Research Record. No. 1341, pp. 19-27.

Chan, Tat Y. (2001). *Demand for Soft Drinks: Characteristics, Corners and Continuous Choice*. Dissertation. Yale University.

Chiang, J. (1991). "A Simultaneous Approach to the Whether, What and How Much to Buy Questions". Marketing Science. Vol. 10, No. 4, pp. 297-315.

Chintagunta, Pradeep K. (1993). "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households". Marketing Science. Spring 1993. pp. 184-208.

Cohen, Gerald S.; Erlbaum, Nathan S.; and Hartgen, David. (1978). "Intercity Rail Travel Models". Transportation Research Record. No. 673, pp. 21-25.

Cubukcu, K. Mert. (2001). *Factors Affecting Shopping Trip Generation Rates in Metropolitan Areas*. Student Report, The Ohio State University Knowlton School of Architecture.

Daganzo, Carlos F. and Kusnic, Michael. (1993). "Two Properties of the Nested Logit Model". Transportation Science. Vol. 27, No. 4, pp. 395-400.

Daly, Andrew. (1987). "Estimating "tree" logit models". Transportation Research B. Vol. 21, No. 4, pp. 251-267.

Daly, Andrew. (2001). "Alternative Tree Logit Models: Comments on a Paper of Koppelman and Wen". Transportation Research B. Vol. 35, No. 8, pp. 717-724.

Damodaran, Sundararajan. (1988). *"A Disaggregate Model to Predict the Intercity Travel Demand"*. Dissertation. The University of New Brunswick.

Deaton, A., and Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.

Domencich, Thomas A. and McFadden, Daniel. (1995). *Urban Travel Demand, A Behavioral Analysis*. A Charles River Associates Research Study.

Dubin, Jeffrey A. and McFadden, Daniel L. (1984). "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption". Econometrica. Vol. 52, No, 2. pp. 345-362.

Ewing, Reid H. ( ). "The Nature of Travel Decision-Making". pp. 66-72.

Federal Highway Administration. (1999). *Highway Performance Monitor System (HPMS) Data*. Washington DC.

Forinash, Christopher V. and Koppelman, Frank S. (1993). "Application and Interpretation of Nested Logit Models of Intercity Mode Choice". Transportation Research Record. No. 1413, pp. 98-106.

Georggi, Nevine Labib and Pendyala, Ram. M. (1999). "Analysis of Long-Distance Travel Behavior of the Elderly and Low Income". *Conference Proceedings of Personal Travel The Long and Short of It*. June 28–July 1, 1999, Washington, D.C.

Golob, TF; MJ Beckmann; and Y. Zahavi. (1981). "A Utility Theory Travel Demand Model Incorporating Travel Budgets". Transportation Research B. Vol. 15, pp. 375-398.

Goulias, Konstadinos G. and Kitamura, Ryuichi. (1993). "Analysis of Binary Choice Frequencies with Limit Cases: Comparison of Alternative Estimation Methods and Application to Weekly Household Mode Choice". Transportation Research B. Vol. 27, No. 1, pp. 65-78.

Grayson, Alan. (1981). "Disaggregate Model of Mode Choice in Intercity Travel". Transportation Research Record. No. 835, pp. 36-42.

Greene, William H. (1997). *Econometric Analysis*. Prentice Hall, Upper Saddle River, New Jersey 07458.

Hamed, M. and F. Mannering. (1993). "Modeling Travelers' Post-work Activity Involvement: Toward a New Methodology". Transportation Science. Vol. 17, No. 4, pp. 381-394.

Hanemann, W. Michael. (1984). "Discrete/Continuous Models of Consumer Demand". Econometrica. Vol. 52, No. 3, pp. 541-562.

Hanley N, Koop G; Wright R; and Alvarez-Farizo, B. (2001). "Go Climb a Mountain: an Application of Recreation Demand Models to Rock Climbing in Scotland" Journal of Agricultural Economics. Vol. 52, No. 1, pp. 36-52.

Hansen, Stein. (1973). "Structure of Disaggregate Behavioral Choice Model". Behavioral Demand Modeling and Valuation of Travel time, TRB Special Report 149, pp. 129-142.

Heckman, J.J.. (1979). "Sample Selection Bias as a Specification Error". Econometric. Vol.7, No. 1, pp. 153-162.

Hendel, I.. (1999). "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns". Review of Economic Studies. Vol. 66, pp. 423 – 446.

Hensher, David A and Greene, Willima H. (2002) "Specification and Estimation of Nested Logit Models". Transportation Research B. Vol. 36, No. 1, pp. 1-18.

Hensher, David A and Greene, Willima H. (2003). "A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit". Transportation Research B, 2003.

Hensher, David A and Greene, Willima H. (2003). "The Mixed Logit Model: The State of Practice and Warnings for the Unwary". Transportation Research B. 2003.

Heyen, Keith Allan. (1992). "*Semiparameteric Estimation of Joint Discrete/Continuous Choice Models*". Dissertation. University of Wisconsin-Madison.

Holguin-Veras, Jose. (2002). "Revealed Preference Analysis of Commercial Vehicle Choice Process". Journal of Transportation Engineering. July/August 2002, pp. 336-346.

Hunt, Gary. (2000). "Alternative Nested Logit Models Structures and the Special Case of Partial Degeneracy". Journal of Regional Science. Vol. 40, No. 1, pp. 89-113.

Institute of Transportation Engineers. (1992). *Transportation Planning Handbook.*

Kahneman, D. and Tversky, A. (1984). "Choices, values and power". American Psychologist. No. 39, pp. 341-350.

Kanafani, Adib. (1983). *Transportation Demand Analysis*. Mcgraw-Hill Book Company.

Kellogg, Mark H. and Diewald, Walter J. (1974). *An Investigation of the Use of Accessibility in Direct Generation Models*. Report UMTA-VA-11-0001-74-1, Virginia Polytechnic Institute and State University.

Keren, G. (1991). "Calibration and probability judgments: Conceptual and methodological issues". Acta Psychologica. No. 77, pp. 217-273.

Keren, G., and Wagenaar, W.A. (1987). "Temporal aspects of probabilistic predictions". Bulleting of the Psychonomic Society. No. 25, pp. 61-64.

Kerstholt, J.H., and Raaijmakers, G.W. (1997). "Decision making in dynamic task environments". In R. Ranyard, W.R. Crozier, & O. Svenson (Eds.), *Decision Making: Cognitive Models and Explanations*. New York: Routledge.

Khan, Ata M. (1985). "Towards the Development of Innovative Models of Intercity Travel Demand". Transportation Quarterly. Vol. 39, No. 2.

King, Leo. (2002). "Acela woes recede". Destination: Freedom, the Newsletter of the National Corridors Initiative, Inc. Vol. 3 No. 34, August 19, 2002.

Kitamura, R.; T. Yamamoto; S. Fujii, and S. Sampath. (1996). "A Discrete-Continuous Analysis of Time Allocation to Two Types of Discretionary Activities Which Accounts for Unobserved Heterogeneity". Transportation and Traffic Theory. J.B. Lesort (ed.), Elsevier, pp. 431-453.

Kockelman, Kara Maria. (1998). "*A Utility-Theory-Consistent System-of-Demand-Equations Approach to Household Travel Choice*". Dissertation. University of California, Berkeley.

Kockelman, Kara Maria. (1999). "Application of a Utility-Theory-Consistent, System-of-Demand-Equations Approach to Household Travel Choice." Transportation Research Record. No. 1676, pp. 20-29.

Kockelman, Kara. (2001). "A Model for Time- and Budget-Constrained Activity Demand Analysis". Transportation Research B. Vol. 35, No. 3, pp. 255-269.

Kockelman, Kara Maria and Krishnamurthy, Sriram. (2002). "Use of Roy's Identity for Comprehensive Travel Demand Modeling: Application to Austin, Texas". Paper Submitted for the 82nd Annual Meeting of the Transportation Research Board. Submitted: July 2002.

Koppellman, Frank S.. (1989). "Multidimensional Model System for Intercity Travel Choice Behavior". Transportation Research Record. No. 1241, pp. 1-8.

Koppellman, Frank S. (1989). *Non- Linear Utility Functions in Models of Travel Choice Behavior*. Report, The Transportation Center, Northwestern University.

Koppelman, Franz S. and Hirsh, Moshe. (1984). "Intercity Travel Choice Behavior: Theory and Empirical Analysis". Proceedings, Fifth International Conference on Travel Behavior, Institut National de Recherche sur les Transports et leur Securite, Aix-en-Provence, France, 1991. pp. 227-244.

Koppelman, Frank S. and Hirsh, Moshe. (1986). "Intercity Passenger Decision Making: Conceptual Structure and Data Implications". Transportation Research Record. No. 1085.

Koppelman, Frank S. and Wen, C.H. (1998). "Alternative Nested Logit Models: Structure, Properties and Estimation." Transportation Research B. Vol. 32, No. 5, pp. 289-298.

Koppellman, Frank S.; Sethi, Vaneet; and Wen, Chieh-hua. (2001). "Alternative Nested Logit Models: A Response to Comments by Andrew Daly on an Earlier Paper of Frank Koppelman and Chieh-hua Een". Transportation Research B. Vol. 35, No. 8, pp. 725-729.

Kraft, G., Systems Analysis and Research Corporation. (1963). "Demand for Intercity Passenger Travel in Washington-Boston Corridor". National Technical Information Service Report, No. PB166-884, Springfield, VA.

Kumar, Ajay and Levinson, David. (1993). "Specifying, Estimating and Validating a New Trip Generation Structure: A Case Study of Montgomery County, Maryland". Transportation Research Record. No. 1413, pp. 107-113.

Kuppam, Arun R.; Pendyala Ram M.; and Rahman Shela. (1999). "Analysis of the Role of Traveler Attitudes and Perceptions in Explaining Mode-Choice Behavior". Transportation Research Record. No. 1676.

Lansing, J.B., et al. (1961). "Analysis of Interurban Air Travel". Quart.J.Econ. Vol. 75, No.1, pp. 87-85.

Lave, Lester B. (1972). "The Demand for Intercity Passenger Transportation". Journal of Regional Science. Vol. 12, No. 1.

Layton, David F.; Haynie, Alan; and Hupper, Daniel. (2003). "Modeling Fishing Behavior under Spatial Regulation: A Discrete/Continuous Expected Profit Model". Preliminary Results.
http://www.feem.it/NR/Feem/resources/pdf/PRE2003-07-27-28-Haynie.pdf.

Liao, Huei-Chu and Chang, Tsai-Feng. (2002). "Space-Heating and Water-Heating Energy Demands of the Aged in the US". Energy Economics. Vol. 24, No. 3, pp. 267-84.

Limtanakool, Narisra; Dijst, Martin; Lanzendor, Martin. (2003). "International Comparison of Long-Distance Travel: the United Kingdom and the Netherlands". Transportation Research Board 2003 Annual Meeting CD-ROM.

Liu, R., and Li, G. (2003). "The Subjective Values of Safety and Security in the Context of Intercity Travel". Accepted to present at the 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.

Lyles, Richard W. and Mallick, Sayeedur R. (1990). "An Intercity Model of Modal Split Considering Four Modes." Journal of Advanced Transportation. Vol. 24, No. 3. pp. 191-212.

Maler, K. (1974). *Environmental Economics*. Baltimore: Johns Hopkins Press.

Mannering, F. (1983). "An Econometric Analysis of Vehicle Use in Multivehicle Households". Transportation Research A. Vol. 17, No. 3, pp. 183-189.

Mannering, F., (1985). "Selectivity Bias in Models of Discrete/Continuous Choice: an Empirical Analysis". Transportation Research Record. No. 1085, pp. 85-62.

Mannering, F. and Winston, Clifford. (1985). "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization". Rand Journal of Economics. Vol. 16, No. 2, pp. 215-236.

Mannering, F. (1997). "Modeling Driver Decision-Making: a Review of Methodological Alternatives". *Human Factors in Intelligent Transportation Systems*. W. Barfield and T. Dingus, editors. Lawrence Erlbaum, New Jersey, USA. pp. 187-216.

Mannering Fred and Hensher David A. (1987). "Discrete/Continuous Econometric Models and Their Application to Transport Analysis". Transport Reviews. Vol. 7, No.3, pp. 227-244.

Mannering, F.; S. Abu-Eisheh; and A. Arnadottir. (1990). "Dynamic Traffic Equilibrium with Discrete/Continuous Econometric Models". Transportation Science. Vol. 24, No. 2, pp. 105-116.

Mandel, Benedikt; Gaudry, Marc; and Rothengatter, Werner. (1994). "Linear or nonlinear utility functions in logit models? The impact on German High-speed rail demand forecasts". Transportation Research B. Vol. 28, No. 2, pp. 91-101.

Manski, Charles F. and McFadden, D. (1981). *"Structural Analysis of Discrete Data with Econometric Application"*. The MIT Press Cambridge, Massachusetts London England.

Martland, Carl; D. Lu, Lexcie; Shi, Dalong; and Sussman, Joseph M. (2002). "Performance-Based Technology Scanning for Intercity Rail Passenger Systems". Transportation Research Board, 2003 Annual meeting.

Maryland Mass Transit Administration. (2000). *Baltimore-Washington Maglev Project Description.*

McFadden, Daniel and Talvitie, Antti et al. (1977). *Demand Model Estimation and Validation.* The Urban Travel Demand Forecasting Project Phase Final Report Series. Volume V. Special Report UCB-ITS-SR-77-9.

McFadden, D. (1978). *Modeling the Choice of Residential Location*. Chapter 3 in Spatial Interaction Theory and Planning Models (Eds. A. Karlqvist, L. Lundqvist, F. Snikars, and J. Weibull) Amsterdam: North-Holland Press.

McFadden, D. (1981). *Econometric Models of Probabilistic Choice*. Chapter 5 in "Structural Analysis of Discrete Data with Econometric Application" edited by Manski, Charles F. and McFadden, D. The MIT Press Cambridge, Massachusetts London England.

McFadden, D. (2000). *Disaggregate Behavioral Travel Demand's RUM Side a 30-year Retrospective*. Department of Economics, University of California, Berkeley.

McFadden, D.; Winston, C.; and A. Boersch-Supan. (1985). *Joint Estimation of Freight Transportation Decisions under Nonrandom Sampling*. In: Daughety A.F. (Ed.), Analytical Studies in Transport Economics, Academic Press, New York, pp. 137-157.

McFadden, D., and Reid, F. (1975). "Aggregate Travel Demand Forecasting from Disaggregate Behavioral Models". Transportation Research Record: Travel Behavior and Values. No. 534, pp. 24-37.

McLynn, J. M. and Woronka, T. (1969). "Passenger Demand and Modal Split Models". Arthur B. Young and Co.

Mehndiratta, R.M. and M. Hansen. (1997). "Analysis of Discrete Choice Data with Repeated Observations: Comparison of Three Techniques in Intercity Travel Case". Transportation Research Record. No. 1607, pp. 69-73.

Meyer, John R. (1971). *Techniques of Transport Planning, Volume 1*. Washington, D.C.: The Brookings Institution, Transport Research Program.

Meyer, Michael D., and Miller, Eric J. (2001). *Urban Transportation Planning*. Second Edition.

Miller, Eric J. (1992). "Intercity Passenger Travel Demand Modeling: Present State and Future Possibilities". Canadian Transportation. pp. 378-388.

Moavenzadeh, F.; M. J. Markow; B. D. Brademeyer; and K. N. A. Safwat. (1983). "A Methodology for Intercity Transportation Planning in Egypt". Transportation Research A. Vol. 17, No. 6, pp. 481-491.

Morey Edward R. (1997). *Two RUMs uncloaked: Nested-Logit Models of Site Choice and Nested-Logit Models of Participation and Site Choice*. Forthcoming as Chapter 4 in "Valuing the Environment Using Recreation Demand Models," (C.L. Kling and J.A. Herriges, Eds), Edward Elgar Publishing Ltd.

Morikawa, Takayuki; Ben-Akiva, Moshe; and Yamada, Kikuka. (1991). "Forecasting Intercity Rail Ridership Using Revealed Preference and Stated Preference Data". Transportation Research Record. No. 1328.

Muellbauer, J. (1976). "Community Preferences and the Representative Consumer". Econometric. No. 44, pp. 979-999.

Nerhagen, Lena. (2000). *Mode Choice Behavior, Travel Mode Choice Models and Value of Time Estimation - A Literature Review.* *http://ctek.du.se/CTEKnet/documents/wp2000-6.pdf.*

Nesbakken, R. (2001)). "Energy Consumption for Space Heating: A Discrete-continuous Approach". Scandinavian Journal of Economics. Vol. 103, No. 1, pp. 165-184.

Ortuzar, J. de D. and Willumsen, L.G.. (1994). *Modeling Transport.* John Wiley and Sons, UK .

O'Neill, W.; P. Fulton; and E. Brown. (1999). "Evaluating the Role of Distance and Location in Statewide Travel Demand Forecasting Using the American travel Survey". Conference Proceedings of Personal Travel The Long and Short of It. June 28–July 1, 1999, Washington, D.C.

O'Neill, W. and E. Brown. (1999). "Long-Distance Trip Generation Modeling Using ATS". Conference Proceedings of Personal Travel The Long and Short of It. June 28–July 1, 1999, Washington, D.C.

Peers, John B. and Bevilacqua, Michael, Alan M. Voorhees and Associates, Inc. (1976). "Structural Travel Demand Models: An intercity Application". Transportation Research Record. No. 569, pp. 124-135.

Parkin, John. (1999). *Transportation Engineering - Module 3G2 Trip Generation ". Class note.* http://www.bolton.ac.uk/campus/builtenv/bengcv/tripgene.html.

Pas, Eric I. (1996). "Recent Advances in Activity-Based Travel Demand Modeling". Activity-Based Travel Forecasting Conference Proceedings.

The Public Purpose. (undated). *US Transit Costs per Passenger Mile Compared to Other Modes: 1970-1997.* The Transport Fact Book. http://www.publicpurpose.com/ut-tr$&oth-1997.htm.

Quandt, R. E. and K. H. Young. (1969). "Cross-sectional Travel Demand Models: Estimates and Tests." J. Region, Sci.. Vol. 9. pp. 201-214.

Quandt, R. E. (1970). *The Demand for Travel: Theory and Measurement.* Health Lexington Books.

Quandt, R. E. and Baumol, W. J. (1966). "The Abstract Mode Model: Theory and Measurement". Journal of Regional Science. Vol. 6, No. 2, pp. 13-26.

Rallis, Tom. (1977). *Intercity Transport, Engineering and Planning.* London and Basingstoke: MacMillan.

Recker, W.W. (2001). "A Bridge between Travel Demand Modeling and Activity-based Travel Analysis". Transportation Research B. Vol. 35, pp. 481-506.

Rice, R.G.; Miller, E.J.; Steuart, G.N.; Ridout, R.; and Brwon, M. (1981). *Review and Development of Intercity Travel Demand Models.* Research report No.77, University of Toronto/York Joint program in Transportation, Toronto, ON.

Rink, Jim. (2003). *Average Driving Cost Up 1.5 Cents Per Mile For 2003.* AAA Michigan, Retrieved June 2003 from the World Wide Web: http://www.autoclubgroup.com/michigan/about_us/press_releases.asp?articleID= 999&view=,.

Safwat, K. N. A. (1987). "Application of a Simultaneous Transportation Equilibrium Model to Intercity Passenger Travel in Egypt". Transportation Research Record. No. 1120, pp. 52-59.

Safwat K. N. and Magnanti T. L. (1988). "A Combined Trip Generation, Trip Distribution, Modal Split, and Trip Assignment Model". Transportation Science. Vol. 18, No. 1, pp. 14-30.

Sethi, Vaneet. (2000). *Modeling Complex Substitution Patterns with Variance and Covariance Heterogeneity in Long Distance Travel Choice Models.* Dissertation, Northwestern University.

Sheffi, Yosef. (1979). "Estimating Choice Probabilities among Nested Alternatives". Transportation Research B. Vol. 13, pp. 189-205.

SkyTran. (2003). *Fast Inter-City Travel by SkyTran.* http://www.skytran.net/04Technical/pod13.htm

Sobel, Kenneth.L. (1980). "Travel Demand Forecasting with the Nested Multinomial Logit Model". Transportation Research Record. No. 775, pp. 48-55.

Sonesson, Thomas. (2001). "Inter-Urban Travel Demand Elasticities with Emphasis on Trip Generation and Destination Substitution". Journal of Transportation Economics and Policy. Vol. 35 Part 2. pp. 301-326.

Stephaned, Yorgos J.; Kumar, Vijaya; and Padmanabhan, Balakrishnan. (1984). "A Fully Disaggregate Mode-Choice Model for Business Intercity Travel". Transportation Planning and Technology. Vol. 9, pp. 12-23.

Stopher, Peter R. and Prashker, Joseph N. (1976). "Intercity Passenger Forecasting: the Use of Current Travel Forecasting Procedures". Transportation Research Forum, pp. 67-75.

Stopher, P. R. and Lisco, T.E. (1970). "Modeling Travel Demand: a Disaggregate Behavioral Approach Issues and Applications". Transportation Research Forum Proceedings. Vol. 11, pp. 195-214.

Talvitie, Antti. (1973). "Aggregate Travel Demand Analysis with Disaggregate or aggregate Travel Demand Models". Transportation Research Forum Proceedings. Vol. 14. pp. 583-603.

Tardiff, Timothy.J. (1977). "Trip Generation as a Choice Process". ASCE, Transportation Engineering Journal. Vol.103, No. TE2, pp. 337-348.

Tatineni, Maya. (1992). "Using a Combined Model of Travel Choice to Analyze Effects of Cost Variations on Travel Patterns". CAT Research News, Journal of Transportation Research. Vol. 29, No. 1. pp. 7-23.

Timmermans, H.J.P. and Mazanec, J.A. (2001). *A Decision Support System for Outdoor Leisure Planning*. http://alexandria.tue.nl/extra2/200142120.pdf.

Train, Kenneth. (1986). *Qualitative Choice Analysis*. Cambridge, MA, MIT Press.

Train, Kenneth. (2002). *Discrete choice methods with simulation*. Cambridge University Press.

Trani, Drs. A. and Teodorovic, D. (2003). "Demand in Air Transportation Systems" Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University.

Tye, W.B.; Sherman, L., et. al. (1982). *Application of Disaggregate Travel Demand Models*. NCHRP report No. 253. Transportation Research Board, Washington, D.C.

U.S. Bureau of Transportation Statistics. (1999). *Transportation Statistics Annual Report 1999*. BTS 99-03.

U.S. Department of Commerce, office of the under secretary of commerce for transportation. (Undated). *Demand for Intercity Passenger Travel in the Washington-Boston Corridor*. Systems Analysis and Research Corporation Report prepared for U.S. Department of Commerce.

U.S. Department of Commerce, U.S. Census Bureau. (2002). *http://www.census.gov/*.

United States Deportment of Transportation, Office of High Speed Ground Transportation. (1970). *Northeast Corridor Transportation Project Report.* NECTP-209.

United States Deportment of Transportation, Federal Highway Administration. (1999). *Guidebook on Statewide Travel Forecasting.* Federal Highway Administration Report FHWA-HEP-99-007.

United States Deportment of Transportation, Bureau of Transportation Statistics. (1995). *1995 American Travel Survey.*

United States Deportment of Transportation. (1997). *Model Validation and Reasonableness Checking Manual.* From Travel Model Improvement Program. http://tmip.fhwa.dot.gov/clearinghouse/docs/mvrcm/.

Vaage, Kjell. (2000). "Heating Technology and Energy Use: a Discrete/Continuous Choice Approach to Norwegian Household Energy Demand". Energy Economics. Vol. 22, No. 6, pp. 649-666.

Varian, H. (1978). *Microeconomic Analysis.* New York: Norton and Co.

Vermont Agency of Transportation, et al. (2003). *Boston to Montreal High-Speed Rail Planning and Feasibility Study Phase I, Final Report.* Prepared by Parsons Brinckerhoff Quade & Douglas with others.

Vickerman, R.W. and Barmby T.A. (1985). "Household Trip Generation-Alternative Empirical Approach". Transportation Research B. Vol. 19, No. 6, pp. 471-479.

Vickerman, R. W. and Barmby, T. A. (1984). "The Structure of Shopping Travel: Some Developments of the Trip Generation Model". Journal of Transport Economics and Policy. Vol. 18, No. 2, pp. 109-121.

Volusia County Metropolitan Planning Organization, Florida. (2000). *Volusia County MPO Long-Range Transportation Planning-Refinement.* http://www.volusiacountympo.com/documents/documents_lrtp.html

Vovsha, P. (1998). "Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area". Transportation Research Record. No. 1607, pp. 6-15.

Vuuren, Daniel J. van, and Rietveld, Piet. (2000). "Modeling the Demand for Train Kilometers: A Microeconometric Approach". Draft. http://econpapers.hhs.se/scripts/redir.pl?u=http%3A%2F%2Ffmwww.bc.edu%2F RePEc%2Fes2000%2F1425.pdf&h=repec:ecm:wc2000:1425.

Walker, Joan. (2002). "The Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification". Submitted for Publication in Transportation Research Record. http://web.mit.edu/jwalker/www/JW_TRB02.pdf.

Watson, Peter L. (1973). "Predictions of Intercity Modal Choice from Disaggregate, Behavioral, Stochastic Models". Highway Research Record. No.446, pp. 28-35.

Watson, Peter L. (1974). "Comparison of the Model Structure and Predictive Power of Aggregate and Disaggregate Models of Intercity Mode Choice". Transportation Research Record. No. 524.

Wen, Chieh-Hua and Koppelman, Frank S. (2001). "The Generalized Nested Logit Model". Transportation Research-B. Vol. 35, No. 7, pp. 627-641.

Westin, R., and Gillen, D.. (1978). "Parking Location and Transit Demand; a Case Study of Endogenous Attributes in Disaggregate Model Choice Model". Journal of Econometrics. Vol. 8. pp. 75-101.

Wilson, Frank R.; Damodaran, Sundar; and Innes, J. David. (1990). "Disaggregate Mode Choice Models for Intercity Passenger Travel in Canada". Canadian Journal of Civil Engineer. Vol. 17, pp. 184-191.