Report D – Revised Comprehensive Report

Development of Long-Distance Multimodal Passenger Travel Modal Choice Model

Prepared for:

Brad Gudzinas
Office of Highway Policy Information
Federal Highway Administration
1200 New Jersey Avenue, S.E.
Washington, DC 20590

July 23, 2012

<table>
<tr><td>1. Report No.</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient Catalog No.</td></tr>
<tr><td colspan="2">4 Title and Subtitle
Report D – Revised Comprehensive Report
Development of Long Distance Multimodal Passenger Travel Modal Choice Model</td><td colspan="2">5 Report Date
July 23, 2012
6 Performing Organization Code
100003810-03</td></tr>
<tr><td colspan="2">7 Author(s)
M.S. Anderson
J.M. Simkins</td><td colspan="2">8 Performing Organization Report No.</td></tr>
<tr><td colspan="2">9 Performing Organization Name and Address
Battelle
505 King Avenue
Columbus, OH 43201-2693</td><td colspan="2">10 Work Unit No. (TRAIS)
11 Contract or Grant No.
DTFH61-07-D-00010</td></tr>
<tr><td colspan="2">12 Sponsoring Agency Name and Address
United States Department of Transportation
Federal Highway Administration
Office of Highway Policy Information
1200 New Jersey Avenue SE
Washington DC 20590</td><td colspan="2">13 Type of Report and Period Covered
Comprehensive Report on Statistical Models Fit to Long-Distance Travel Data
6/2011 to 5/2012
14 Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15 Supplementary Notes</td></tr>
<tr><td colspan="4">16 Abstract

This report includes an executive summary as well as details of research conducted to develop a set of statistical models to predict long-distance passenger travel mode. A literature and practice review was conducted to gather information on the current knowledge and practices for long-distance passenger travel modal choices. Using this information, a set of multinomial logistic regression models were developed to predict mode choice given various demographic and socioeconomic attributes of the traveler, trip characteristics such as length and purpose, land-use characteristics, as well as availability of transportation infrastructure. The models use long-distance travel data from the 2001 NHTS as well as other sources. K-fold cross-validation is used to assess the predictive ability of the models.</td></tr>
<tr><td colspan="2">17 Key Words
Long-Distance Travel, Passenger Travel Mode Choice, Multinomial Logistic Regression Model</td><td colspan="2">18 Distribution Statement
No restrictions.</td></tr>
<tr><td>19 Security Classification (of this report)
Unclassified</td><td>20 Security Classification (of this page) Unclassified</td><td>21 No. of pages
84</td><td>22 Price
N/A</td></tr>
</table>

Form DOT F 1700.7 (8-72)

This page intentionally left blank.

EXECUTIVE SUMMARY

According to the 2001 National Household Travel Survey (NHTS), Americans take 2.6 billion long-distance trips (defined in the NHTS as 50 miles or more one-way) per year, or 7.2 million trips per day. About 90 percent of long-distance trips are taken by personal vehicle while ten percent use public/commercial transportation modes. Over seven percent of long-distance trips are taken by air while two percent are by bus. Train travel represents almost one percent of long-distance trips (BTS, 2006). The Office of Highway Policy Information is interested in learning more about what factors influence the choice of travel mode for long-distance trips. Thus, the objective of this research is to develop quantitative mathematical methods to analyze how long-distance passenger travelers make their modal choices. Factors including -- but not limited to -- social, economic, demographic, trip length, trip purpose, available infrastructure facilities such as rail, airport, and highways, and available modal choices (air, train, bus, and personal passenger vehicles) were evaluated. Prior to the model development, a comprehensive literature and practice review was conducted, with the goal of assessing current knowledge on long distance multimodal passenger travel modeling.

Overall, literature and practices that explore the following topics were reviewed:

- Research performed on mathematical techniques for long distance passenger travel modal choice modeling;
- Data sources used for long distance passenger travel modeling that could supplement the NHTS data; and
- Factors that were found to influence long-distance passenger travel mode choice.

Mathematical Techniques for Long Distance Passenger Travel Modal Choice Modeling

The literature review showed that analyses of long-distance multimodal passenger travel mode choice range in complexity from simple summary statistics and cross tabulations to more sophisticated mathematical modeling techniques such as nested multinomial logistic regression. There have been numerous research projects conducted on U.S.-based long-distance travel mode choice using the 2001 NHTS and the 1995 American Travel Survey (ATS) that utilize descriptive techniques such as summary statistics, cross-tabulations, and graphical representations to understand the relationship between mode choice and attributes such as socioeconomic and demographic factors as well as trip aspects (distance, duration, purpose).

The more complicated analyses involve discrete choice modeling which are statistical procedures that model choices made by people among a finite set of alternatives. In terms of long-distance travel, discrete mode choice models consider the travel mode that travelers choose for a particular long-distance trip based on certain attributes about the traveler or the trip to be taken. Although discrete choice models can take many forms, the majority of the mode choice models encountered in the literature review are based on some form of logistic regression. Logistic regression models are used to predict the probabilities of the different possible outcomes of a categorical dependent variable (mode choices), given a set of independent variables (socioeconomic characteristics, trip purpose, trip length, etc). Various forms of logit models were encountered in the literature review. Some examples include the most basic binary logit model which models a dichotomous choice in mode (e.g., airplane vs. train), a multinomial logit model which generalizes binary logistic regression by allowing more than two discrete outcomes (e.g., airplane, train, bus, automobile), and other forms of the multinomial logit such as nested or mixed logit models.

Data Sources Used for Long-Distance Passenger Traveler Modeling

The lack of available U.S.-based data on long distance travel is the main hindrance to long distance travel research. Most research identified in this literature review focused on the U.S. has made use of either the relatively recent 1995 ATS and 2001 NHTS surveys or their precursor National Travel Surveys (NTS) conducted by the U.S. Census Bureau. According to the Bureau of Transportation Statistics (BTS), the NHTS provides the only authoritative source of information at the national level on the relationships between the characteristics of personal travel and the demographics of the traveler. Even though the 1995 ATS and 2001 NHTS are the richest and most used data sources on domestic long-distance travel, there are some drawbacks. First, the surveys do not contain information on level-of-service variables such as travel time and travel cost. Second, geographical information at the origin and destination of trips is aggregated to protect the confidentiality of respondents. Because of these issues, researchers have had to look at external data sources such as published fare and schedule guides for airline, railroad, and bus to consider travel cost and time or limit their analysis by only focusing on travel to and from a Metropolitan Statistical Area (MSA) to compensate for the data shortcomings. Although there do exist other travel surveys that have some data on long-distance travel, the literature review found that they lack the richness and size of the ATS or NHTS. National surveys (e.g. versions of the NTS) prior to the 1995 ATS were not reviewed in detail given the time elapsed since they were performed.

The American Automobile Association (AAA) has developed an approach for forecasting actual domestic travel volumes based on macroeconomic drivers such as unemployment, output, household net worth, asset prices including stock indices, interest rates, and housing market indicators. The report also includes variables related to travel and tourism, including prices of gasoline, airline travel and hotel stays as well as historical travel volume estimates from travel survey databases.

Factors That Influence Mode Choice

For a lot of studies that examine long-distance travel, the focus has been primarily on the impact of socioeconomic factors at the individual and household levels. In these studies, the relationship between factors such as age, income, gender, and household location (urban vs. rural) were examined. Another area of focus for long-distance travel studies is the incorporation of land-use factors. Research has found that land-use factors have a significant impact on travel mode choice. For example, Algers (1993) found that the total number of trips over 100 kilometers was sensitive to the characteristics of the destination including population size and number of jobs. Using the 1989 Netherlands National Travel Survey, Limtanakool et al (2006) studied the effects of land use attributes such as population density, proximity to infrastructure, and land use diversity on travel mode choice and concluded that spatial configuration of land use and transport infrastructure has a significant impact even when socioeconomic characteristics and travel time are taken into account. Other studies have found that travel time and travel costs heavily influence mode choice.

One constant across all the research encountered is that the relationship between mode choice and certain factors varies by trip purpose. For example, the mode share for automobiles is higher for personal or social trips, while air travel is the preferred method for business travel. This finding as well as all others from the literature and practice review presented here was used to develop the mode choice models presented below.

Mathematical Models for Predicting Long-Distance Passenger Mode Choice

The research team decided on the 2001 NHTS as the primary data source for this modeling effort. The 2001 NHTS is a national survey of daily and long-distance travel. The survey includes demographic characteristics of households, people, vehicles, and detailed information on long-distance travel for all purposes by all modes. NHTS survey data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and a host of household attributes. According to the Bureau of Transportation Statistics (BTS), the NHTS provides the only authoritative source of information at the national level on the relationships between the characteristics of personal travel and the demographics of the traveler. The NHTS collected travel data from a national sample of the civilian, non-institutionalized population of the United States. There were approximately 66,000 households in the final 2001 NHTS dataset. The final datasets contained about 45,000 long distance trips.

Predictive factors from the NHTS that were used in the modeling included characteristics of the traveler (age, race, employment status, frequency of internet use, frequency of public/commercial transportation use), characteristics of the trip (distance, number of nights away, number of people traveling, and whether it included a weekend), household and land-use characteristics such as household income, number of vehicles, population density, and urban/rural status.

Although the NHTS gives detailed information on individual and trip level demographic information, several variables from external data sources were included in the model. These variables account for economic and environmental factors that were identified as determinants of individual travel choice mode by other studies but that are not present in the NHTS data. Two main factors governing individual choice of travel mode that these variables particularly seek to include are the economic burden of particular modes of travel as well as the availability and access to transportation infrastructure. Along with demographic information, this additional information can serve as a means to increase the resolution of predictions about travel mode choice based on observed data. Economic variables include the Research and Innovative Technology Administration (RITA) Air Travel Price Index and the Consumer Price Index (CPI) Private and Public Transportation Components. The number of different types of transportation sites within a 25 mile radius of the traveler's origin was also used. These include airports, bus depots, light and transit rail stations and standard rail stations.

Discrete choice models are statistical procedures that model choices made by people among a finite set of alternatives. Specifically, discrete choice models statistically relate the choice made by each person to the attributes of the person and the attributes of the alternatives available to the person. In terms of long-distance travel, discrete mode choice models consider the travel mode that travelers choose for a particular long-distance trip based on certain attributes about the traveler or the trip to be taken. Although discrete choice models can take many forms, the majority of the mode choice models involving transportation are logit based. For this research, logistic regression models were used to predict the probabilities of the different possible outcomes of a categorical dependent variable (mode choices of personal vehicle, air, bus, and train), given a set of independent variables (characteristics of the traveler, trip, and household, land-use factors, economic variables, and availability of transportation infrastructure).

A separate model was developed for each trip purpose: business, pleasure, and personal business. The 2001 NHTS provides an analysis weight for each long-distance trip. The weight is defined at the person trip/travel period level. These weights reflect the selection probabilities and adjustments to account for nonresponse, undercoverage, and multiple telephones in a household. Point estimates of population parameters as well as coefficients of predictors are impacted by the value of the analysis weight for each observation. To obtain estimates that are minimally biased the analysis weight was used to weight the results.

Coefficients associated with each predictive factor were estimated using the maximum likelihood estimation technique using the SAS® (version 9.3) statistical software package. The SURVEYLOGISTIC procedure was used to take into account the complex nature of the 2001 NHTS sample design. Model coefficients for the predictor variables as well as marginal probability effects were estimated from the model.

Validation of the long-distance passenger travel modal choice models was conducted by testing the models on long-distance travel survey data. The same 2001 NHTS dataset used for model calibration was used for model validation. *K*-fold cross-validation is a statistical technique for assessing how the results of the statistical model will generalize to an independent dataset. The data is first partitioned into *k* equally (or nearly equally) sized segments, or folds. Then, *k* iterations of calibrating and validation are performed such that a different fold of the data is held out for validation while the remaining *k*-1 folds are used to calibrate the model within each iteration. For this research, 10-fold cross-validation was conducted separately to validate each of the three multinomial mode choice models (one for each trip purpose). In each iteration, the fitted model was applied to the validation dataset (i.e., predicted probabilities for each mode of transportation were calculated for each trip in the validation dataset). Aggregate mode shares were calculated by summing the calculated probabilities for each trip record in the validation dataset. These were compared against the observed aggregate mode shares of the validation dataset in order to observe how well the model could replicate the observed mode shares. This process was repeated nine times, each time choosing a different segment of the data to be held out as the validation dataset. Once all iterations were complete, the comparison of predicted versus observed aggregate mode shares were combined across the ten iterations and statistics summarizing the predictive ability of the model were calculated.

Major findings from this research are as follows:

- Summary statistic and model results provide evidence that mode choice varies by trip purpose and that separate models are warranted;
- There were a much greater number of factors found to significantly influence mode choice observed across trip purpose types for personal vehicle and air travel outcomes than bus and train outcomes. This is due, in part, to the low frequency of bus and train trips in the NHTS;

- Characteristics of the survey respondents who were taking the trips tended to be more significant predictors of travel mode choice than the characteristics of the trips themselves. Specifically, familiarity with public/commercial transportation systems through frequent usage resulted in a large decrease in the likelihood of taking personal vehicles for business travel (eight percent) as well as a smaller but still significant decrease in the likelihood of taking personal vehicles for pleasure travel (three percent). Interestingly, high public/commercial transportation use was highly statistically significant for predicting increases in the use of air travel (four percent for business, 1.2 percent for pleasure). For business travel, frequent web use also increased chances of taking air travel by about 4.5 percent. Income was also a strong predictor of travel mode choice for both business and pleasure travel. Lower income travelers were more likely to take personal vehicles and less likely to take air travel. The lower likelihood of air travel as income decreases shows the stronger statistical significance trend, and this reinforces the hypothesis that fixed attributes like income are much stronger determinants of travel mode. Overall, income and behavioral variables seemed to display the highest statistical significance in model results. This indicates that people's travel mode choices may be driven largely by fixed attributes that revolve around residence and demographics rather than consideration of the dynamic costs and benefits of different modes of travel;
- Marginal effects for variables describing trip characteristics other than distance tended to have mixed effects for different travel mode outcomes. A weekend trip had a statistically significant marginal effect for personal vehicle and air travel for the two largest travel purpose types (business and pleasure). There was a two to three percent decrease in the probability of taking a personal vehicle and a two percent increase in the probability of taking air travel if the trip included a weekend for business and pleasure travel. The number of persons on the trip also significantly impacted likelihoods of different mode choices; for business travel it corresponded to a 0.5 percent decrease in the chances of taking personal vehicle per person and a 0.5 percent increase in the chances of taking air travel while for pleasure travel it increased chances of taking bus travel by 0.12 percent per person. Lastly, for pleasure travel, the number of nights away increased the probability of taking personal vehicles by 0.19 percent a night and decreased the probability of taking bus travel by 0.15 percent a night.
- The results suggest that respondents' demand for different modes of travel is relatively decoupled from cost considerations such as the price of airfares or gasoline and that the preference set may be fairly inelastic in the short run – that is, not responsive to changes in price;
- Available transportation infrastructure only appeared to be influential for business travel. The number of airports in a 25 mile radius increased the chances of taking air travel by 1.7 percent per airport. Other existing transportation infrastructure did not appear to play a significant role in travel choice, but this could also be a product of large numbers of observations in the data set that chose personal vehicle as the primary mode of transport and thus do not display any preferences towards certain types of existing networks.
- One of the most consistently significant variables in predicting mode choice was route distance of a trip from origin to destination. The probability of choosing to travel in a

personal vehicle decreases exponentially with travel distance while the probability of choosing air travel increases exponentially with travel distance; and

- The model predicts very well for the personal vehicle and air modes but loses some predictive power for the bus and train modes. The relative lack of predictive power for bus and train modes indicates that the survey data may not be sufficient to accurately assess some outcomes and that alternative sampling techniques should be explored in future national travel surveys that provide more data for bus and train trips.

ABBREVIATIONS

Abbreviation	Definition
AAA	American Automobile Association
ATPI	Air Travel Price Index
ATS	American Travel Survey
BLS	Bureau of Labor Statistics
BTS	Bureau of Transportation Statistics
CATI	Computer-Assisted Telephone Interviewing
CPI	Consumer Price Index
CSTDM	California Statewide Travel Demand Model
DKSA	D.K. Shifflet and Associates
FHWA	Federal Highway Administration
GTFS	General Transit Feed Specification
IIA	Independence of Irrelevant Alternatives
ITRD	International Transport Research Documentation
LIRR	Long Island Rail Road
LDPTM	Long Distance Personal Travel Model
MSA	Metropolitan Statistical Area
MTA	Metropolitan Transit Authority
NHTS	National Household Travel Survey
NPTS	Nationwide Personal Transportation Survey
NTAD2011	National Transportation Atlas Database 2011
NTS	National Transportation Survey
OECD	Organization for Economic Co-operation and Development
RITA	Research and Innovative Technology Administration
SUV	Sport Utility Vehicle
TSAM	Transportation Systems Analysis Model
TRIS	Transportation Research Information Service
USDOT	United States Department of Transportation

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

Page

List of Tables

TABLE OF CONTENTS (CONTINUED)

Page

List of Figures

1.0 INTRODUCTION

The Federal Highway Administration's (FHWA) Office of Highway Policy Information has been conducting research on both passenger and freight origin-destination data gathering and estimation methods and approaches. This research includes: (1) exploring new long-distance passenger travel data gathering methods; (2) a national transportation modeling framework; and (3) multimode passenger travel origin-destination. The Office of Highway Policy Information is initiating additional new research on another key component of long-distance passenger travel: passenger travel mode choice modeling.

According to the 2001 National Household Travel Survey (NHTS), Americans take 2.6 billion long-distance trips (defined in the NHTS as 50 miles or more one-way) per year. This equates to 7.2 million trips per day. As shown in Figure 1-1, about 90 percent of long-distance trips are taken by personal vehicle while ten percent use public/commercial transportation modes. Over seven percent of long-distance trips are taken by air while two percent are by bus. Train travel represents almost one percent of long-distance trips (BTS, 2006). The Office of Highway Policy Information is interested in learning more about what factors influence the choice of travel mode for long-distance trips. Thus, the objective of this new research component is to develop quantitative mathematical methods to analyze how long-distance passenger travelers make their mode choices. Factors including social, economic, demographic, trip length, trip purpose, available infrastructure facilities such as train, airport, and highways, indicators of travel costs, and available mode choices (air, train, bus, and personal passenger vehicles) were evaluated.

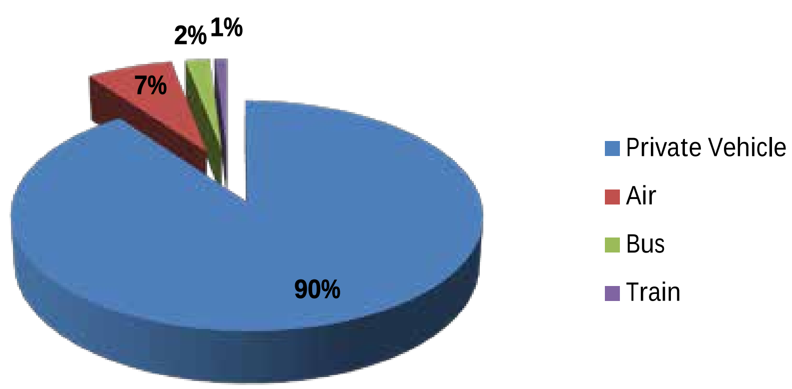

Figure 1-1. Percent of Long-Distance Passenger Trips by Mode Share Choice According to 2001 NHTS.

The research began with a comprehensive literature and practice review conducted to assess current knowledge on long distance multimode passenger travel modeling. Details of this review are provided in Section 2.0. Section 3.0 contains a detailed discussion of the mathematical models and inputs to the models used to estimate mode choice for long-distance passenger travel. Section 3.0 also contains the model coefficients and a discussion of the results. This is followed by validation of the mathematical models in Section 4.0. Finally some overall conclusions are presented in Section 5.0 followed by a bibliography in Section 6.0.

This page intentionally left blank.

2.0 LITERATURE AND PRACTICE REVIEW

Prior to the model development, a comprehensive literature and practice review was conducted to assess current knowledge on long distance multimode passenger travel modeling. This section discusses that review and is organized as follows: Section 2.1 sets the stage for the literature and practice review by providing some background material. This is followed in Section 2.2 by the review methodology, including research criteria and sources used in the search. Findings from the literature and practice reviews are included in Sections 2.3 and 2.4, respectively. Finally, a discussion of the results in terms of how they influence the model development is provided in Section 2.5.

2.1 Background

The research team believes that a crucial first step in developing and implementing a rigorous set of quantitative methods to analyze how long-distance passenger travelers make their modal choices is finding a complete, detailed, and accurate data source or sources that can be used as inputs to the mathematical models. At a minimum, the data source(s) should:

- Contain the passenger-selected mode of travel (i.e. air, rail, bus, private passenger vehicle) for long distance trips;
- Contain detailed information on the long-distance travel trips and traveler (e.g. social, economic, and demographic characteristics of traveler, availability of passenger vehicle, trip length, trip purpose, available infrastructure facilities such as rail, airport and highways);
- Be nationally representative of long distance travel in the U.S.; and
- Be available from public domain sources without the requirement of programs or purchases from private commercial vendors.

Prior to the literature search, the research team proposed that data from the 2001 NHTS be the primary data source for this modeling effort. It meets the four criteria above. The 2001 NHTS is a national survey of daily and long-distance travel. The survey includes demographic characteristics of households, people, vehicles, and detailed information on long-distance travel for all purposes by all modes. NHTS survey data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and a host of household attributes. According to the Bureau of Transportation Statistics (BTS), the NHTS provides the only authoritative source of information at the national level on the relationships between the characteristics of personal travel and the demographics of the traveler.

Overall, literature and practices that explore the following topics were reviewed:

- Research performed on mathematical techniques for long distance passenger travel modal choice modeling;
- Data sources used for long distance passenger travel modeling that could supplement the NHTS data;
- Applied practices used by organizations such as tourism/travel bureaus and economic development agencies; and
- Factors that were found to influence long-distance passenger travel mode choice.

Although the NHTS contains a large amount of data needed to model long distance passenger travel modal choices, the research team recognizes that the dataset is ten years old and may need to be updated and that other data sources could supplement the data from the NHTS because they contain additional, value added attributes related to long distance travel. Thus, the identification of such data sources was a primary objective of this review process. Efforts were made to quantify the strengths of the data sources (i.e. they have information that would provide an added value to the modeling, they can be linked to NHTS data, and they are publically available). Only data sources that meet these criteria would be considered as inputs to the modeling. One example is the Metropolitan Travel Survey Archive located at the University of Minnesota and funded by FHWA and BTS. This is an archive to store, preserve, and make publicly available, via the internet, over 80 travel surveys conducted from almost 45 metropolitan areas, states and localities. Another example is the American Travel Survey (ATS) conducted by BTS in 1995 which collected information from approximately 80,000 households about their long-distance travel through 1995.

The review of current knowledge by researchers or the applied practices of travel-related organizations was used to identify best-practices. Specifically, it helped identify the list of factors that have been found to be significantly related to long-distance travel mode choice. Although the main focus of this research is to develop mathematical models for domestic long-distance passenger travel, the review of current knowledge and best practices focuses both on domestic and international research into long-distance passenger travel. International research was included to increase the number of research studies reviewed mainly because the number of long-distance travel data sources and research domestically is limited but also because it was hypothesized that international research would provide additional insights into long-distance multimodal passenger travel modal choice modeling that could be used in developing models in this research.

2.2 Methodology

In order to identify as many publicly available documents as possible for consideration in the assessment of current knowledge on long distance multimodal passenger travel modeling, the research team sought out documents from various sources. The types of documents included international, regional, and national-level reports, workshop and conference proceedings, summaries and presentations, peer-reviewed journal articles, and other published documents. The research team used the following key methods in the document identification process:

- Searched the bibliographic database TRID. TRID is a newly integrated database that combines the records from the Transportation Research Board's Transportation Research Information Services (TRIS) Database and the Organization for Economic Co-operation and Development’s (OECD) Joint Transport Research Centre’s International Transport Research Documentation (ITRD) Database. TRID provides access to over 900,000 records of transportation research worldwide.
- Searched the websites of universities known to be active in transportation research (e.g., Virginia Tech, University of Florida, and University of California).
- Performed web-based literature searches using standard search tools and databases. Key words used to perform searches included variations on phrases such as “long distance passenger mode choice modeling“, and “passenger stated preference”.

- Referred to the references listed in the documents identified through other methods to provide potential new sources of information.

As part of the practice review, the research team reached out to several tourism/travel bureaus and economic agencies via telephone and email to learn about applied practices involving long-distance multimodal passenger travel mode choice modeling used by such organizations. Phone interviews were conducted once the appropriate contact at the organization was reached. Organizations that responded via email received a follow-up phone call and were interviewed.

2.3 Literature Review Findings

Findings from the literature review are organized around the following three main areas that correspond to the objectives of the literature review:

1. Research performed on mathematical techniques for long distance passenger travel modal choice modeling;
2. Data sources used for long distance passenger travel modeling; and
3. Factors that were found to influence long-distance passenger travel mode choice.

2.3.1 Mathematical Techniques for Long-distance Passenger Travel Modeling

The literature review showed that analyses of long-distance multimodal passenger travel mode choice range in complexity from simple summary statistics and cross tabulations to more sophisticated mathematical modeling techniques such as nested multinomial logistic regression. One common theme in most of the research is that analysis on mode choice is performed separately by trip purpose. This section discusses both domestic and international research conducted starting with the more simple analyses and continuing through more sophisticated modeling applications.

2.3.1.1 Analyses Based on Cross Tabulation and Descriptive Statistics

There has been numerous research projects conducted on U.S.-based long-distance travel using the 2001 NHTS and the 1995 ATS. Several reports published by the BTS provide interesting facts and figures as well as highlight main findings from the NHTS and ATS. Examples of these documents include:

- America on the Go … Findings from the National Household Travel Survey (BTS, 2006)
- A Picture of Long-Distance Travel Behavior of Americans Through Analysis of the 2001 National Household Travel Survey (Sharp et al., 2006)
- NHTS 2001 Highlights Report (BTS, 2003).

These publications focus mainly on frequencies based on long-distance travel such as:

- Long-distance trips by mode
- Longs distance trips by trip purpose
- Destinations of long-distance trips (e.g., within same state, out of state, international)

In addition, cross-tabulations are provided of travel mode by various factors such as:

- Trip purpose
- Trip distance
- Geography (urban vs. rural area)

- Traveler demographics
- Access and egress modes
- Gender
- Household income.

Among the key findings from these cross tabulations include:

- Long-distance trips originating in urban and metropolitan areas are more likely to use public/commercial transportation modes than trips originating in rural and non-metro areas.
- About eight percent of long-distance trips that use a public/commercial transportation mode use a different mode in each direction of travel.
- Almost 90 percent of long-distance trips are by personal vehicle.
- Mode choice varies somewhat by trip purpose and distance.
- Personal vehicle is the most frequent mode used to initially access long distance public/commercial transportation, but on the arrival end a greater mix of modes is used. (BTS, 2006)

Other research on the 2001 NHTS and 1995 ATS based on cross tabulations and other non-modeling techniques has been targeted to a specific area or hypothesis within long-distance travel. Bricka (1999) examined regional variations in long-distance travel in the U.S. by comparing geographically diverse regions (New York, Massachusetts, and Oklahoma). Using national- and state-level data, comparisons of long-distance trips were performed to identify differences in trip length, purpose, mode, and demographic characteristics of the travelers. Mallett (1999a) studied the long-distance travel behavior of low-income households in comparison with higher income households using the ATS. Also, Georggi and Pandyala (1999) provided a detailed analysis of long-distance travel behavior for two key socioeconomic groups of the population - the elderly and low income. Mallett (1999b) researched long-distance travel behavior by women, presenting data on women's long-distance travel broken out by trip purpose, trip mode, age, race/ethnicity, and household type.

2.3.1.2 Model-Based Analyses

Discrete choice models are statistical procedures that model choices made by people among a finite set of alternatives. Specifically, discrete choice models statistically relate the choice made by each person to the attributes of the person and the attributes of the alternatives available to the person. In terms of long-distance travel, discrete mode choice models consider the travel mode that travelers choose for a particular long-distance trip based on certain attributes about the traveler or the trip to be taken. Although discrete choice models can take many forms, the majority of the mode choice models encountered in the literature review are logit based. The mathematical framework of logit models in based on the theory of utility maximization which is discussed in detail in Ben-Akiva and Lerman (1985). Utility theory assumes that travelers prefer an alternative with the highest utility where utility is a representation of the attractiveness of the mode choice alternatives as derived from the traveler. Logistic regression models are used to predict the probabilities of the different possible outcomes of a categorical dependent variable (mode choices), given a set of independent variables (socioeconomic characteristics, trip purpose, trip length, etc).

2.3.1.2.1 Binary Logit Models

The most basic logistic regression models are binary logit models where the dependent or response variable is dichotomous in nature (e.g., airplane vs. rail). The literature review identified three studies where binary logit modeling was employed to analyze travel mode choice behavior.

- The Delaware Transportation Institute and the State of Delaware Department of Transportation examined factors that affect or can alter mode choice by analyzing transit trips in New Castle County, DE. (Racca and Ratledge, 2003) The study used data from a Delaware DOT Household Survey conducted annually from 1995 through 2001. Binary logit models were fitted to three trip purposes separately (transit trips, passenger trips, and walking trips) using factors such as mode travel time and costs, socioeconomic variables, access to alternative modes, vehicle availability, and time of day of transit service. Since driving trips were not studied in this case, the focus was more on short-distance trips.
- Using the 1998 Netherlands National Travel Survey, Limtanakool et al (2006) fit a series of binary logit models survey data to address the question of how socioeconomic factors, land use attributes, and travel time affect mode choice (train vs. private car) for long distance travel and how their role varies across trip purpose. Separate models were fit to commuting trips, business trips, and leisure trips. The authors state that their study was one of the first to incorporate land use attributes such as population density, proximity to infrastructure, and land use diversity. Specifically, they considered the associations between mode choice and land use attributes at the origin and destination sides while controlling for the influence of socioeconomic characteristics of persons, households, and travel times.
- In Japan, Kitagawa et al (2005) fit a series of binary logit models to estimate the mode share (air vs. high speed rail) between the cities of Keihanshin and Fukuoka. Analysis was performed based on data from a web survey in 2004. Factors of interest were based on cost and waiting time (e.g., fare cost, travel time, egress time, egress expense, access time, access expense, and time outside vehicle).

2.3.1.2.2 Multinomial Logit Models

Several researchers have employed a multinomial logit model to predict modal choice for long distance travelers. A multinomial logit model is a regression model which generalizes binary logistic regression by allowing more than two discrete outcomes. This type of model is more useful when dealing with long-distance travel because of the multiple modes available from origin to destination. Figure 2-1 presents an example of a simple multinomial logit model specification.

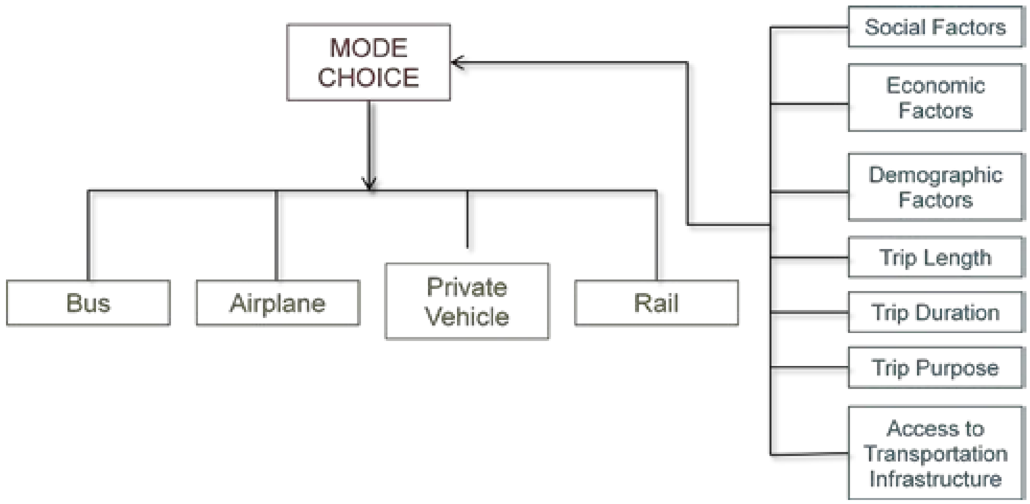

Figure 2-1. Visualization of Simple Multinomial Logit Model.

One assumption of the multinomial logit model is that the model error terms are independent and identically distributed. As a result, when the multinomial logit model is used to model choices, it relies on the assumption of independence of irrelevant alternatives (IIA) which is not always desirable. Ben-Akiva and Lerman (1985) give the definition as "the ratio of the chosen probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives." They continue on to show that IIA can produce wrong estimates when a new mode with similar characteristics is introduced into the mode choice set. As such, nested logit models are sometimes used as an extension of the multinomial logit model to capture the correlation of alternatives when alternatives are not independent. Nested logit models relax the independence assumption by grouping similar alternatives into nests.

Presented below are summaries and/or the published abstracts from national followed by international research studies conducted where multinomial or nested multinomial logit models were employed for long-distance passenger travel mode choice.

- Using data from the 2001 NHTS, Rasmidatta (2006) fits both binary and nested multinomial logit models to study the behavior of long-distance travelers in the U.S. Separate models are fit to each of three different trip purposes: business, personal business, and pleasure. Models were fit using data from the NHTS national sample and then validated with samples from Texas and Wisconsin. Data from external sources were not used. Factors used in the analysis included socioeconomic variables, availability of a car to the traveler, and various trip aspects (length, number of nights away). Rasmidatta also used Neural Networks to study the long distance travel behavior. The author found that the advantages to using Neural Networks are that there are no assumptions for the data and parameters (i.e. they can handle multicollinearity and lack of correlation). A Neural Network is an iterative non-linear model that can solve problems like people naturally solve problems. Rasmidatta found that both the nested logit models and Neural Networks generated strong results with the Neural Networks having slightly better prediction.
- Ashiabor et al (2007) developed nested and mixed logit models to study national level intercity transportation in the United States. The models are used to estimate the market

share of automobile and commercial air transportation between 3091 counties and 443 commercial service airports in the United States. Models were calibrated using the 1995 American Travel Survey and separate models were developed for business and non-business trip purposes. Factors explored in the models are travel time, travel cost, and traveler's household income. Given an input county-to-county trip demand table, the models were used to estimate county-to-county travel demand by automobile and commercial airline between all counties and commercial service airports in the United States. The model has been integrated into a computer software framework called the Transportation Systems Analysis Model (TSAM) that estimates nationwide intercity travel demand in the United States.

- Prior to their research, Ashiabor et al (2007) indicate that four major attempts were made between 1976 and 1990 to develop disaggregate national-level intercity mode choice models in the U.S. All of these attempts made use of national travel studies conducted by the Bureau of the Census and BTS. These four attempts along with the model developed by Ashiabor et al is presented in Table 2-1 (Table 2-1 is extracted from Table 1 of Ashiabor et al. report).
- The first major attempt was made by Stopher and Prashker (1976) which used the 1972 National Travel Survey (NTS). This was followed by Grayson (1982) utilizing the 1977 NTS. Morrison and Winston (1985) and Koppelman (1990) fit nested logit models using data from the 1977 NTS. Finally, Ashiabor et al fit nested logit and mixed logit models to the 1995 ATS as part of the mode choice model in their transportation system analysis model. Each of the authors created separate models for business and non-business trips and included modes of transportation that included automobile, air, train, and bus.
- Using data from a Swedish national travel study conducted in 1984-1985, Algers (1993) developed a system of models consisting of nested logit models partly estimated by the use of simultaneous estimation techniques. There are different models for business and private trip purposes. The models studied traveler socio-economic characteristics, travel cost, travel time (including access and egress time), and some elements of land use attributes at the destination.
- In Australia, Khan et al (2007) develops a mode choice module capability that could be incorporated into the Brisbane Strategic Transport Model. The module consists of a series of multinomial logit models for eight trip purpose categories (home based work – white collar, home based work - blue collar, home based education – primary and secondary, home based education – tertiary, home based shopping, home based other, work based work, and other non-home based trips). Data came from the Southeast Queensland Travel Survey. Factors incorporated into the models including demographic characteristics (number of adults and number of vehicles in household), land use characteristics (employment density), and travel characteristics (travel cost and travel time related to transit, access/egress, and waiting). Models developed for this study were mainly reflective of local travel as opposed to long-distance travel but were included here to provide another example of modeling techniques used in mode choice research.
- Rand (Burge et al., 2011) conducted a stated preference study in Great Britain to study long distance passenger travel defined as one-way trips over 50 miles. Using data from this survey, nested multinomial logit models were created to predict mode choice (rail, high speed rail, air, car) based on a series of factors that included socio-economic

characteristics, values of time, cost sensitivity, out-of-vehicle components such as frequency, interchanges, and access and egress time, rail service components such as rail reliability and crowding, and whether there existed an additional preference for high speed rail over classic rail.

Table 2-1. Major National-Level Intercity Travel Demand Models for the United States.

Author	Model Type	Data and Scope	Modes of Transportation	Variables in Utility Function	Market Segmentation
Stopher and Prashker (1976)	Multinomial logit	1972 NTS	Automobile, commercial air, bus, rail	Relative time, relative distance, relative cost, relative access-egress distance, departure frequency	Trip purpose (business, nonbusiness)
Alan Grayson (1982)	Multinomial logit	1977 NTS	Automobile, commercial air, bus, rail	Travel time, travel cost, access time, departure frequency	Trip purpose (business, nonbusiness)
Morrison and Winston (1985)	Nested logit	1977 NTS	Automobile, commercial air, bus, rail	Travel time, cost, party size, average time between departures	Trip purpose (business, nonbusiness)
Koppelman (1990)	Nested logit	1977 NTS	Automobile, commercial air, bus, rail	Travel time, cost, departure frequency, distance between city pairs, household income	Trip purpose (business, nonbusiness)
Ashiabor et al. (2007)	Nested logit and mixed logit models	1995 ATS	Automobile, commercial air, train, SATS	Travel time, travel cost, household income, region type	Trip purpose (business, nonbusiness) Household Income

Source: Logit Models for Forecasting Nationwide Intercity Travel Demand in the United States, Ashiabor et al (2007)

2.3.1.2.3 Other Model-Based Analyses

As discussed in the previous section, the assumption of independently and identically distributed error terms in the multinomial logit model leads to the IIA property. The nesting of travel modes within the multinomial logit model has been used to overcome the IIA property. In addition, researchers have developed other alternatives to the multinomial logit model. Presented below are summaries and/or the published abstracts from research studies conducted that explore alternative models when modeling long-distance travel choice behavior.

- Bhat (1995) developed a new heteroscedastic extreme value model of intercity mode choice that overcomes the “independence of irrelevant alternatives” (IIA) property of the commonly used multinomial logit model. The proposed model allows a more flexible cross-elasticity structure among alternatives than the nested logit model. Using data from a 1989 Rail Passenger Review conducted by VIA Rail to develop travel demand models to estimate shifts in mode choice, Bhat evaluates the multinomial logit, alternative nested logit structures, and the heteroscedastic model in examining the impact of improved rail service on business travel in the Toronto-Montreal corridor. Bhat finds that the heteroscedastic extreme value model is found to be superior to the multinomial logit model in terms of its ability to predict mode choice and to limit overestimation of certain modes as was observed in the multinomial logit model.

- Koppelman and Sethib (2005) observed that the relaxation of the independently and identically distributed error term assumption has been undertaken along a number of isolated dimensions leading to the development of a rich set of discrete choice models, that are more flexible than the multinomial logit model. In some cases, these more general models lose the mathematically convenient closed-form structure of the multinomial logit.
- In their research, Koppelman and Sethib combine the most flexible isolated closed-form extensions of the multinomial and nested logit models in an integrated model structure to yield a behaviorally rich, yet computationally tractable choice model. Specifically, they combine the generalized nested logit model that allows for non-independent errors, the Heteroscedastic multinomial which allows non-constant errors across observations, and the Covariance Heterogeneous nested logit model which allows for non-constant correlation structure across observations. The resulting model, called the heterogeneous generalized nested logit model extends the ability to represent the complex behavioral processes involved in choice decision-making. The value and need for the additional modeling complexity of the heterogeneous generalized nested logit model was tested in the empirical context of mode and rail service class choice behavior for long distance intercity travel. An incremental modeling approach was adopted where the researchers started from the simple multinomial logit model and sequentially relaxed some of its restrictive assumptions to estimate progressively more flexible model structures. The researchers found that the statistical fit and behavioral appeal of the estimated models improved substantially with each additional relaxation of the restrictive model assumptions, strongly supporting the concept of integrating generalizations of multinomial and nested logit models.
- Koppelman and Wen (2000) provide another alternative model to the multinomial logit model. They ran various models on rail data from 1989 between Toronto and Montreal to estimate demand for high speed rail. They also argue that the IIA property of the multinomial logit model inappropriate for many choice situations in which some pairs or sets of alternatives share the same unobserved attributes (because the model imposes a restriction of zero covariance between the utilities of pairs of alternatives). The nested logit model relaxes the zero covariance restriction of the multinomial model but imposes the restriction of equal covariance among all alternatives in a common nest and zero covariance otherwise. The paired combinatorial logit model relaxes these restrictions further by allowing different covariances for each pair of alternatives. This relaxation enables the estimation of differential competitive relationships between each pair of alternatives. The closed form of the paired combinatorial logit model retains the computational advantages of other logit models while the more flexible error correlation structure, compared to the multinomial model and nested logit models, enables better representation of many choice situations. This paper describes the derivation, structure, properties and estimation of the paired combinatorial logit model. The empirical results demonstrate that the paired combinatorial logit model is statistically superior to the multinomial and nested logit models and may lead to importantly different travel forecasts and policy decisions.

The research presented in this section demonstrates a broad range of techniques on a variety of data sources both domestically and internationally. It also highlights some of the important issues in developing a sound, technically correct modeling system. The next section focuses in greater detail on the data sources used in the research.

2.3.2 Data Sources used for Long-distance Multimodal Passenger Travel Modal Choice Modeling

As evidenced in Section 2.3.1, there were many long-distance multimodal passenger travel modal choice studies conducted outside the United States using various data sources. In the Netherlands, modeling was performed on the 1998 Netherlands National Travel Survey, a household survey asking respondents to record all journeys made in a particular day. In Sweden, the Swedish National Travel Study was used to study long-distance travel. In the Swedish survey, respondents are asked to report on their recent trips. Because both the air and rail authorities have had a major role in the survey, long-trip surveying, trips over 100 km is an important facet of the design. In Japan, a web survey of passengers who traveled between Keihanshin and Fukuoka in 2004 was conducted to examine passenger's mode choice between air and rail. In Canada, the 1989 Rail Passenger Review conducted by VIA Rail (the Canadian national rail carrier) consisting of travel surveys conducted in the corridor to collect data on intercity travel by four modes (car, air, train and bus) were used to model intercity mode choice.

These studies provided a broad understanding of the different data sources available for long-distance travel in terms of size, scope, methods, and data collected. The main focus of this section, however, is data sources used to model long-distance multimodal passenger travel modal choice within the United States. The ultimate goal of this research project is to develop quantitative mathematical methods to analyze how long-distance passenger travelers make their modal choices within the U.S. As such, one objective of this literature review was to identify those data sources used for long distance passenger travel modeling that could supplement the 2001 NHTS data or that could be used instead of the 2001 NHTS in this research effort.

The lack of available U.S.-based data on long distance travel is the main hindrance to long distance travel research. As a result, not many people have conducted research on long distance travel within the U.S. Fortunately for this effort, most of the limited research focuses on long distance travel impact on mode choice (Rasmidatta, 2006). Most research identified in this literature review focused on the U.S. has made use of either the 1995 ATS or the 2001 NHTS which are the most recent sources of long-distance travel data or the past NTS surveys conducted by the U.S. Census Bureau. Ashiabor et al (2007) indicate that four major attempts were made between 1976 and 1990 to develop disaggregate national-level intercity mode choice models in the U.S. All of these attempts plus the research of Ashiabor et al. made use of the 1995 ATS or the previous NTS surveys. These are detailed in Table 2-1 in Section 2.3.1.2.2.

Even though the 1995 ATS and 2001 NHTS are the richest and most used data sources on domestic long-distance travel, there are some drawbacks. First, the surveys do not contain information on level-of-service variables such as travel time and travel cost. Second, geographical information at the origin and destination of trips is aggregated to protect the confidentiality of respondents. Trips in the survey are only identified by state and whether they are in an MSA. As a result, some researchers have developed synthetic travel time and cost data from published fare and schedule guides, such as the official airline, railroad, and bus guides

(Ashiabor et al., 2007). However, analysis was restricted to those trips that originated and destinated in metropolitan statistical areas (MSAs) because it was very difficult to estimate travel times and costs for any trip originating or ending in non-MSA areas (Ashiabor et al., 2007). Ashiabor et al. also found that the major constraints in developing credible models are related more to the national survey databases than the modeling techniques. Specifically, the two major issues are the restriction of the minimum level of geographic detail to MSA and the absence of information related to airports and access and egress distance to airports and terminals. National surveys (e.g. versions of the NTS) prior to the 1995 ATS were not reviewed in detail given the time elapsed since they were performed.

This literature review examined the possibility that there existed data sources at smaller levels of aggregation (i.e., state and local level) that contained information on long-distance travel. A few states have conducted household travel surveys where modal choice is investigated but they mainly focus on daily travel and short distance trips. For example, the Delaware Transportation Institute and the State of Delaware Department of Transportation regularly conducted a household telephone survey that examined factors that affect or can alter mode choice. Analysis was conducted on data from 1995-2001 by analyzing transit trips in New Castle County, DE. The survey contained information on mode travel time and costs, socioeconomic variables, access to alternative modes, vehicle availability, and time of day of transit service. Since driving trips were not studied in this case, the analysis focus was more on short-distance trips and compared walking, transit, and bus modes.

Another domestic data source investigated was the Metropolitan Travel Survey Archive located at the University of Minnesota and funded by FHWA and BTS. This is an archive to store, preserve, and make publicly available, via the internet, over 80 travel surveys conducted from almost 45 metropolitan areas, states and localities. Located in the archive for each survey, if available, are the raw data, documentation, rectangularized dataset, xml dataset, and an analysis tool whereby users can summarize and analyze the survey data on the website. Generally, the reports did not address long-distance travel but short home-based travel. Although it is possible that the surveys could have some records of long-distance trips, intercity travel is not the focus of the Metropolitan Travel Survey Archive and thus the number of such trips would be extremely limited.

2.3.3 Factors that Influence Long-distance Passenger Travel Modeling

The majority of academic research is heavily concentrated on short-distance travel or trips conducted in daily urban systems (Limtanakool et al., 2006). Among the most important factors affecting mode choice for daily trips or short-distance trips are the socioeconomic characteristics of travelers. For a lot of studies that examine long-distance travel, the focus has been primarily on the impact of socioeconomic factors at the individual and household levels (Algers, 1993; Georggi and Pendyala, 1999; Mallett, 1999a,b; O'Neill and Brown, 1999). Although it is unclear whether conclusions about mode choice for short-distance trips can be translated easily to long-distance trips given that the latter involve more time and monetary expense, this limited number of studies dealing with long-distance travel suggests that these also play a significant part in modal choice for these types of trips (Limtanakool et al., 2006).

Analysis of the 1995 ATS shows that 80 percent of trips greater than 100 miles are taken by private car (Georggi and Pendyala, 1999). Mallett found that mode choice for long-distance trips does not differ much between men and women; however, women do tend to travel by bus more often and men by airplane (1999b). Age also affects mode choice. Georggi and Pendyala (1999) found that the elderly are significantly more dependent on the bus mode than the rest of the population. Also, the automobile mode share diminishes significantly for people over 75 years of age as the airplane and bus are used instead of the automobile more frequently (Georggi and Pendyala, 1999). In regards to income, Mallett (1999a) found that about two-thirds of people in low-income households did not make a single long-distance trip in 1995 with the most important limiting factors being the availability of a vehicle. Moreover, lower income groups were found to be much more likely to travel by automobile or bus when compared to other income groups (Georggi and Pendyala, 1999). Air travel was a more popular choice for long-distance travel as income levels increased.

In addition to socioeconomic factors, Rasmidatta (2006) examined trip characteristics such as trip distance, trip purpose (business, personal business, pleasure), and whether the trip was a short weekend, long weekend, or non-weekend trip using the 2001 NHTS. The analysis shows that socioeconomic factors such as occupation, income, race, whether a household is in an urban or rural area, and availability of a vehicle as well as trip characteristics such as nights away on a trip, household members on the trip, route distance, and trip characteristics all have a significant impact on mode choice.

Another area of focus for long-distance travel studies is the incorporation of land-use factors. Research has found that land-use factors have a significant impact on travel mode choice. In addition to studying traveler socioeconomic characteristics, travel time, and travel cost, Algers (1993) examined some elements of destination land use. He found that the total number of trips over 100 kilometers was sensitive to the characteristics of the destination including population size and number of jobs. However, the influence of travel mode choice on land use factors was not the main focus of his research. Bricka (1999) analyzed variations in long distance trips in New York, Massachusetts, and Oklahoma via descriptive cross-tabulation of survey results from the three locations. She finds that mode choice between states can be explained by differences between rural and urban areas such as dissimilar demographic profiles, availability of modes, and urban form. Using the 1989 Netherlands National Travel Survey, Limtanakool et al (2006) studied the effects of land use attributes such as population density, proximity to infrastructure, and land use diversity on travel mode choice. They concluded that spatial configuration of land use and transport infrastructure has a significant impact even when socioeconomic characteristics and travel time are taken into account.

One last set of factors used to model long-distance passenger travel mode choice is travel costs and travel time. These factors are not part of the 1995 ATS or 2001 NHTS or previous NTS surveys so Ashiabor et al. (2007) generated synthetic level-of-service variables from external data sources. They found that travel time and costs were a significant factor in travel mode choice. In general, as travel times and costs increased, the utility of any of the modes decreased. Furthermore, an examination of the travel costs for modes over the range of income levels in their research show that high-income travelers are less sensitive to travel cost.

One constant across all the research encountered is that mode choice varies across types of trip. Georggi and Pendyala (1999) assert that the mode choice varies considerably with trip purpose and trip length. For example, they found that personal vehicle is higher for personal or social trips, while air travel is the preferred method for business travel. As a result, researchers have consistently fit separate logit-based models to data by trip purpose. The amount of models fit is a function of the number of trip purposes in the data source.

In summary, the existing literature has shown that mode choice for long-distance travel is based on the socioeconomic characteristics of the traveler, aspects of the trip such as distance and duration, land use characteristics, as well as travel time and costs. In addition, the effect of these factors on mode choice varies by trip purpose.

2.4 Practice Review Findings

Of the 20 tourism/travel bureaus and economic development agencies that were contacted, a handful responded with valuable information for the practice review. Other organizations reported that they did not focus on long-term travel or simply did not respond to emails or voicemail messages left for them. This section presents the findings from those few organizations where useful information was obtained.

2.4.1 Metropolitan Transportation Commission in California

The Metropolitan Transportation Commission in Oakland, California is primarily concerned with intra-regional, short-distance travel and thus does not measure traveler mode choices themselves. However, when this information is needed, it is received from the California Statewide Travel Demand Model (CSTDM). This model is used by the State of California Department of Transportation to forecast personal travel by California residents on a typical weekday. The Long Distance Personal Travel Model (LDPTM) is one component of this application that applies to all trips greater than 100 miles. At the traffic zone level, the following travel information are modeled: trip frequency, party size, destination, main mode, and access/egress modes. Survey data as well as micro-simulation is used during model estimation to determine model parameters tied to decision choice probabilities. The main mode modeling uses a nested logit model to determine which of four potential modes of travel (car, air, conventional rail, and high speed rail) will be chosen for the long distance trip between the home and the destination zone. Separate models are created by type of trip (recreation/other purpose, or business/commute). Factors examined in the models are related to accessibility (e.g., cost, in-vehicle time, access/egress mode), demographics of the traveler or household (e.g., household size, income), and airport interchanges. The length of the trip is not considered. In addition, reliability indices such as the percent of flights/trains that arrive on time are assigned to each mode type (car, air, conventional rail, high speed rail), and used as explanatory parameters to explain long distance trip mode choice.

2.4.2 State of Michigan

Contacts in the State of Michigan reported that their state does not predict long distance travel by mode. They occasionally conduct surveys to capture information on long distance trips by car, rail or bus separately. Primarily, however, they use a report prepared by D.K. Shifflet and Associates (DKSA) in 2009 for the Michigan Economic Development Corporation that breaks out long-distance trips by mode when they need long-distance travel mode information. The report is titled “Michigan 2009 Visitor Profile” and focuses on a number of different travel and tourism related metrics, including U.S. and Michigan travel segments as well as travel parameters such as purpose of stay, travel composition and size, stay length, activities, daily spending, trip timing, accommodation type, and mode of transportation for leisure travel.

Data for this report comes from a DKSA-conducted, ongoing, monthly survey of U.S. consumers' travel behavior—the *PERFORMANCE/Monitor*SM. Approximately 50,000 households are surveyed per month via mail and online panels. Each survey collects the previous three months of travel behavior. DKSA uses an overlapping monthly mail sequence which reduces sample bias for maximum accuracy. More than 75,000 traveling households respond to the survey each year. This results in more than 154,000 Stays at destinations throughout the U.S. According to DKSA, the 50,000 average monthly contacts as well as returned questionnaires are balanced to the U.S. population across six demographic variables (age, gender, income, education, number of adults, and state of residence). This rebalancing ensures findings are reflective of the U.S. population and enables findings to be projected to the entire U.S. population.

The report contains graphics on the frequency of automobile vs. air transportation as the main mode as measured in leisure person days in 2009. Also shown is the travel distance by auto and air as measured in leisure person-days for Michigan, the U.S., as well as the five geographical regions within Michigan. A representative from DKSA reported that these reports have been completed for a number of states and their respective economic development or travel bureaus. Some examples include Indiana, Tennessee, New Jersey, and Illinois.

2.4.3 American Automobile Association (AAA)

Each quarter, AAA publishes a holiday forecast that combines information from several sources to provide a prospective assessment of likely travel patterns for the upcoming holiday season. The most recent report is the AAA Independence Day 2001 Travel Forecast (AAA, 2011). This report is comprised of two key components – the actual travel forecast and the holiday traveler profile. The actual travel forecast is based on economic conditions while the holiday traveler profile is developed employing survey data on travel behaviors. AAA partners with two organizations in preparing the travel forecast: IHS Global Insight and D.K. Shifflet and Associates (DKSA). IHS Global Insight developed the approach for forecasting actual domestic travel volumes based on macroeconomic drivers such as employment, output, household net worth, asset prices including stock indices, interest rates, housing market indicators and variables related to travel and tourism, including prices of gasoline, airline travel and hotel stays. The historical travel volume estimates come from the ongoing travel survey database of DKSA who interviews over 50,000 U.S. households per month tracking trip incidence, party composition, traveler behavior, and spending all after the trips have been taken.

Actual travel is forecasted by person-trips, where a person-trip is defined as a trip that involves travel of 50 miles or more away from home. In particular, AAA and IHS Global Insight forecasts total US holiday travel, travel by mode of transportation, and travel by US census region.

According to a representative from IHS Global Insight, travel mode forecasts are based off of historical travel numbers and trends received from DKSA. IHS Global Insight then ties those historical trends to the current fuel price, miles traveled, and disposable income (of the sampled population) to predict the mode of travel. The change in each of those parameters, especially as compared to previous analyses, are crucial in forecasting future holiday travel modes. The other microeconomic factors mentioned in the report, such as unemployment, output, household net worth, are used to predict travel demand and not mode forecasting. However these and other factors may be applied to the travel mode model results in a qualitative/general way, to confirm the output with AAA findings or other market findings.

The previous examples provide a small sampling of the applied practices of travel/tourism bureaus and economic development agencies in relation to long-distance multimodal passenger travel mode choice. Many agencies contacted were non-responsive or were not able to provide valuable information for this report. Time on this research effort did not permit a more detailed examination.

2.5 Discussion

This section presents the results of a comprehensive literature and practice review to assess current knowledge on long distance multimodal passenger travel modeling. This review served as a precursor to the development of quantitative mathematical methods to analyze how long-distance passenger travelers make their modal choices. Knowledge gained during this review on the mathematical techniques used, data sources that served as the inputs, as well as the factors that were found to significantly affect long distance travel mode choice was used in the development of mathematical models for this research.

While the literature review found a range of analysis techniques, the predominant methodology employed to model modal choice was logistic regression modeling. Given the number of mode choices available to long-distance passenger travels, a multinomial logit model was used in the next phase of this research. The literature identified that mode choice varies considerably across trip purpose and trip length. Thus, separate models were fit by trip purpose. The decision on which model structure (i.e. nested logit model, mixed logit model) that was utilized and how many models were created was based on a thorough analysis of the final dataset and input factors that were to be used.

Going into this review, the research team believed that a crucial first step in developing and implementing a rigorous set of quantitative methods to analyze how long distance passenger travelers make their mode choices is having a complete, detailed, and accurate data source that can be used as inputs to the mathematical models. This review strengthens that belief. At a minimum, the data source(s) should:

- Contain the passenger-selected mode of travel (i.e. air, train, bus, personal passenger vehicle) for long distance trips;
- Contain detailed information on the long-distance travel trips and traveler (e.g. social, economic, and demographic characteristics of the traveler, availability of passenger vehicle, trip length, trip purpose, and available infrastructure facilities such as train, airport and highways);
- Be nationally representative of long distance travel in the U.S.; and
- Be available from public domain sources without the requirement of programs or purchases from private commercial vendors.

Based on the literature review, the research team decided to use the 2001 NHTS as the main data source when developing models for long-distance multimode passenger travel mode choice. It meets most of the four criteria above. Data from other sources were used to supplement the NHTS data in areas where the NHTS is lacking (level-of-service variables such as available transportation infrastructure, travel cost (indirectly through financial proxies such as those used by AAA), and land use characteristics at the origin or destination were used to capture information not present in the NHTS. The 2001 NHTS is preferred over the 1995 ATS solely because the NHTS contains more recent information. Although there do exist other travel surveys that have some data on long-distance travel, the literature review found that they lack the richness and size of the ATS or NHTS.

3.0 MATHEMATICAL MODELS FOR PREDICTING MODE CHOICE

This section discusses the development of the mathematical models to predict mode choice starting from the input data sources and going through the model results. Section 3.1 discusses the main data source, the 2001 NHTS. Section 3.2 presents additional sources used to supplement the NHTS. A summary of predictive factors used in the mode choice modeling is given in Section 3.3 followed by a descriptive analysis of these factors in Section 3.4. The statistical background and methodology for the models is presented in Section 3.5 followed by the results in Section 3.6. Finally, a discussion of the results is presented in Section 3.7.

3.1 2001 National Household Travel Survey

The 2001 NHTS is a national survey of daily and long-distance travel. The survey includes demographic characteristics of households, people, vehicles, and detailed information on long-distance travel for all purposes by all modes. NHTS survey data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and a host of household attributes. According to BTS, the NHTS provides the only authoritative source of information at the national level on the relationships between the characteristics of personal travel and the demographics of the traveler. In addition to providing the first comprehensive look at travel by Americans, the 2001 NHTS also incorporated additional enhancements to previous sample designs (e.g. 1995 ATS and prior Nationwide Personal Transportation Surveys (NPTS)). For example, long distance travel was expanded to include trips as short as 50 miles and, for the first time, included trips made for the purpose of commuting to work – often overlooked segments of personal long-distance travel (BTS, 2003).

The NHTS collected travel data from a national sample of the civilian, non-institutionalized population of the United States. Sampling was done by creating a random-digit dialing list of telephone numbers. An eligible household excludes telephones in motels, hotels, group quarters, such as nursing homes, prisons, barracks, convents, or monasteries, and any living quarters with ten or more unrelated roommates (FHWA, 2004).

There were approximately 66,000 households in the final 2001 NHTS dataset. About 26,000 households were from the national sample, while the remaining 40,000 households were from nine add-on areas. The nine add-on areas were: Baltimore, Des Moines, Hawaii, Kentucky, Lancaster PA, New York State, Oahu, Texas, and Wisconsin. The final datasets contained about a quarter-million daily trips and 45,165 long distance trips.

NHTS data was obtained by using Computer-Assisted Telephone Interviewing (CATI) technology. Each household was assigned a specific twenty-four hour “Travel Day” to record daily travel by all household members. In addition, a twenty-eight day “Travel Period” was assigned to each household to collect longer-distance travel. Long-distance trips in the 2001 NHTS are defined as trips of 50 miles or more from home to the farthest destination traveled that started and ended within the four-week travel period. Data collected on long-distance trips includes:

- Purpose of the trip (pleasure, business, personal business);
- Means of transportation used (car, bus, train, air, etc.);
- Day of week when the trip took place;

- If a personal vehicle trip:
  - Number of people in the vehicle;
  - Driver characteristics (age, sex, worker status, education level, etc.);
  - Vehicle attributes (make, model, model year, amount of miles driven in a year); and
- Location of overnight stops and access/egress to an airport, train station, bus station, or boat pier.

Furthermore, the 2001 NHTS data contains data on the following:

- Household data on the relationship of household members, education level, income, housing characteristics, and other demographic information;
- Information to describe characteristics of the geographic area in which the sample household and workplace of sample persons are located;
- Public perceptions of the transportation system;
- Internet usage; and
- Information on each household vehicle, including year, make, model, and estimates of annual miles traveled.

For all of its strengths, there are some drawbacks to the 2001 NHTS. Each traveler provided data about household and trip characteristics; however, many data that may be important for long-distance travel mode choice decisions were not in the scope of the 2001 NHTS data collection. Examples of data not included in the NHTS data are travel costs and travel time as well as information that would identify the traveler's household or workplace information. Specifically, geographical information at the origin and destination of trips is aggregated to protect the confidentiality of respondents. Trips in the survey are only identified on both the origin and destination side by state and Metropolitan Statistical Area (MSA). Furthermore, about half the long-distance trips do not have origin or destination information below the state level. This is because either trips do not originate or destinate in an MSA or the MSA is too small in terms of population density to publish it in the dataset for confidentiality reasons. Fortunately, the research team obtained a separate file from FHWA that contained the 5-digit ZIP Code of each household in the survey. This research assumes that each trip originated at the household. Having this information was critical in assessing the availability of transportation infrastructure relative to the origin of the trip. To compensate for other variables that may be important for long-distance mode choice but not present in the NHTS, outside data sources were identified that would provide such variables (or suitable proxies for the variables) that could supplement data from the NHTS.

Another limitation of the 2001 NHTS dataset is that although each traveler provided information about the mode that they used, they did not provide information about other alternative modes or the traveler's reason for selecting a specific mode of travel over another mode. As will be discussed more in Section 3.5, this fact played a significant role in determining the type of multinomial logistic regression model to use to predict mode choice.

Another rich source of long-distance travel data is the 1995 ATS. This was a panel survey conducted by BTS in 1995 which collected information from approximately 80,000 households about their long-distance travel through 1995. Although the ATS has a larger number of long-distance trips compared to the 2001 NHTS, the 2001 NHTS was preferred over the 1995 ATS mainly because the NHTS contains more recent information. This was important as the data from the 2001 NHTS is already ten years old. In addition, the ATS also suffers from the same lack of reported geographic detail at the origin and destination side of trips to protect confidentiality that the NHTS does. While the research team was able to acquire five-digit ZIP Code information on the surveyed households for the NHTS to help with land-use and other variables, this information was not available from the ATS. For these reasons, the ATS was not considered in the model development.

The 9/11 terrorist attacks occurred in the midst of NHTS data collection efforts, and the potential impact of this event on data collection and travel behavior was investigated as a part of the review of this data source. Many studies note that air travel experienced a large initial usage “shock” in the immediate aftermath of the attacks followed by an almost complete recovery in consumer demand by the end of the NHTS study period. There is little evidence that NHTS data suffers from a severe deficiency of air travel observations due to any effect of 9/11, since airline travel statistics published by Research and Innovative Technology Administration (RITA) show that the total annual number of passengers enplaned by domestic carriers was only slightly lower in 2002 than in 2001. However, in order to capture any potential effect on travel behavior caused by the 9/11 attacks, a dichotomous variable was added to track any effect caused solely by travel dates that occurred after the event.

3.1.1 Trip Purpose and Mode Choice

A separate model was developed for each trip purpose: business, pleasure, and personal business. Business trips are ones where a business function is the primary purpose (i.e., to attend a conference, business meeting, or other business function other than commuting to and from work). Other non-business activities can occur as long as the trip is primarily for business. Pleasure trips include trips for vacations, visiting friends and relatives, sightseeing, and outdoor recreation. Personal business trips include trips for medical visits, trips to attend funerals, weddings, and other events. The “other” trip purpose was excluded from the analysis, leaving business, pleasure, and personal business as the three trip purposes used in the modeling.

The modes personal vehicle, air, bus, and train were used in the modeling. The modes “ship” and “other” were not included in the analysis because the number of data points was minimal. A personal vehicle can be a passenger car, sport-utility vehicle, van, or other vehicle owned by the household. Personal vehicles are attractive choices to long-distance travelers in that one can travel from origin to destination and still have a vehicle to use at the destination, travelers have more privacy, and they can have a more flexible schedule. However, personal vehicles can be a slower mode of travel. Vehicles such as taxis, limousines, and other car services were not included as they fell into the “other” mode category and represented a very small portion of the sample. The air mode is a faster transportation alternative but the cost for this alternative is relatively high. Bus and train modes are both ground modes that are usually slower modes which may need to stop at many stations before arriving at a destination. However, they are attractive options for those who do not own a personal vehicle or for those traveling in large groups. For each long-distance trip, the dataset contains information on all travel modes taken on the outbound side of the trip (origin to farthest destination) as well as the return trip. Multiple modes may be taken to get from the origin to destination and these are recorded in the NHTS. For example, a traveler could take a taxi to the airport, a plane to the destination city, and then a rental car to the final destination. For each trip, the NHTS identifies one mode (MAINMOD2) as the main mode that the traveler used most to get to the destination. In the previous example, the main mode would be "air". In this research, this variable identifies the mode of travel for the trip and is the only one considered in the modeling. Although it is possible for the main mode of transportation on the return trip to be different than that on the outbound side of the trip, this research focuses only on the one-way portion of the trip from origin to farthest destination.

3.1.2 Prediction Factors from 2001 NHTS

Variables used in this research came from trips in the national NHTS sample as well as those in the add-on samples. There were many variables present in the NHTS dataset but only a subset was used for model development. Those used for the modeling include ones that were identified from the literature and practice review as well as those that showed a significant correlation with mode choice in exploratory analysis. Some of the variables used were taken directly from the NHTS data files while others, as noted below in the variable descriptions were modified or redefined slightly to reduce the dimensionality of the variable. Variables used in the modeling from the NHTS include:

1. **Total income of all household members**: Income is a very important factor for people who travel long distances. In regards to income, Mallett (1999a) found that about two-thirds of people in low-income households did not make a single long-distance trip in 1995 with the most important limiting factors being the availability of a vehicle. Moreover, lower income groups were found to be much more likely to travel by automobile or bus when compared to other income groups (Georggi and Pendyala, 1999). Air travel was a more popular choice for long-distance travel as income levels increased. Household income was separated into four levels: households making less than or equal to $30,000 annually, households making over $30,000 and up to $60,000 annually, households making over $60,000 and up to $100,000 annually, and households with an income greater than $100,000 annually. An indicator variable was created for each of the four household income levels.
2. **Age of traveler**: Age is a factor that may impact mode choice. Georggi and Pendyala (1999) found that the elderly are significantly more dependent on the bus mode than the rest of the population. Also, the automobile mode share diminishes significantly for people over 75 years of age as the airplane and bus are used instead of the automobile more frequently (Georggi and Pendyala, 1999).
3. **Employment status of respondent**: An indicator of whether the traveler is employed. This variable is included on the hypothesis that employed travelers are likely to take more expensive modes of transportation than those who are unemployed.
4. **Population per square mile – block group for household**: This is a measure of land-use on the origin side of the trip. Travelers who live in heavily populated areas would be more likely to have access to different transportation infrastructures and also to route alternatives that could impact travel mode choice.
5. **Number of vehicles in household**: This variable shows the potential of a household to have a variety of personal vehicles. Households with large number of vehicles may have vehicles of different types which would allow selection based on trip purpose. For

example, large families with a minivan or SUV traveling a long distance might be more inclined to take a personal vehicle than other modes.

6. **Public transit use**: This provides a description of the traveler's public transit use in the last two months. This will serve as an indicator of a traveler's familiarity and comfort level with public/commercial transportation which may have behavioral implications for travel mode choice. From the NHTS variable PTUSED, a binary indicator variable (high_PTuse) was created to identify travelers who use public/commercial transportation at a rate of more than once or twice per month versus those who use it less than once or twice per month.
7. **Internet use**: Provides an indication of a traveler's internet use over the last six months. Travelers who use the internet more frequently would most likely have access to detailed travel information on alternative travel modes (e.g., airline costs and schedules) which could be a potentially important determinant of using travel modes such as airlines. From the NHTS variable WEBUSE, a binary indicator variable (high_webuse) was created to identify travelers who use the internet weekly versus those who use it less frequently.
8. **Nights away on trip**: The number of nights away on a long-distance trip impacts which mode to select. A family or group of travelers might want to spend more nights at a destination rather than many nights en route to the destination. Thus, shorter trips might be conducive to faster travel modes.
9. **Trip before or on/after 9/11**: The terrorist attacks on September 11, 2001 occurred during the data collection for the NHTS survey (March 2001 through May 2002). The terrorist attacks played a significant role on the behavior of intercity travelers in that after 9/11, people avoided traveling by air either out of fear or because of the increasing security and the uncertainty of passenger processing times at airports. The variable is an indication as to whether the trip occurred after 9/11 or before 9/11.
10. **Race of traveler**: Differences in race may affect mode choice for long-distance travel. Indicator variables were created for each of the following races: white, African-American, Asian, Hispanic, and other.
11. **Origin to destination route distance**: Route distance is a critical factor when choosing mode choice. Longer-distance travel will most likely encourage a traveler to select a faster mode. As trips become longer, the probability of taking personal vehicle or other ground forms of transportation should be reduced.
12. **Number of people on trip**: The greater the number of people on a trip, the greater the travel expense. Thus, families and groups of travelers in large numbers may be more likely to choose personal vehicle or perhaps bus as compared to more expensive options such as air.
13. **Location of household**: This is another measure of land-use on the origin side of the trip. Travelers who live in an urban area would be more likely to have access to different transportation infrastructures and also to route alternatives that could impact travel mode choice more so than in rural areas.
14. **Trip includes weekend**: Travelers who travel during the week or on short weekend trips may prefer a faster transportation mode such as air because they need to return for work.

For longer weekend trips, a slower transportation method may be preferred as travelers may have more time to spend and can do so at a lower cost.

Shortly after model development and the initial draft of the technical report, FHWA and the research team discussed the list of variables used from the NHTS to predict mode choice. FHWA expressed concern that although certain variables may play a role in determining a traveler's mode choice, it might prove difficult to obtain valid estimates for these variables when using the model to forecast mode choice within the national transportation modeling process. These variables are: 1) a measure of a traveler's public transit use; 2) a measure of a traveler's internet use; 3) the number of nights away on a trip; and 4) an indicator of whether the trip involved a weekend. The research team believes these variables are very informative and have an effect on the choice of transportation mode. However, the team also understands that the variables are useless in the model if no practical inputs can be easily obtained (i.e. from census data) without conducting another large scale travel survey. As a result, the research team presents in Section 3.6 both a full mode-choice prediction model with all the inputs identified in this section as well as a reduced prediction model that removes these variables.

In addition, FHWA expressed policy concerns with including the race variables in the model. Although the inclusion of race as a demographic variable in economic studies is quite common and has even been used in past long-distance travel mode choice studies (Georggi and Pendyala, 1999, Rasmidatta, 2006), the research team acknowledges and understands the policy concerns and implications and thus has not included the race variables in the reduced prediction model. Their effect is still explored in the full model.

Version 4.0 (July 2005) of the 2001 NHTS data was used in the analysis and model formulation. Both the 2001 NHTS long-distance trip dataset and the dataset of replicate weights were obtained from the NHTS Data Center located on the NHTS website http://nhts.ornl.gov/download.shtml. In addition, the United States Department of Transportation (USDOT) version of the 2001 NHTS household data (containing more detailed geographic information on survey households) was provided by Oak Ridge National Laboratory through FHWA.

3.2 Data Sources to Supplement the 2001 NHTS

Although the NHTS gives detailed information on individual and trip level demographic information, several variables from external data sources were included in the model. These variables, accounting for economic and environmental factors, are not present in the NHTS data but were identified in the literature review as determinants of individual travel choice mode. Two main factors governing individual choice of travel mode that these variables particularly seek to include are the economic burden of particular modes of travel as well as the availability and access to transportation infrastructure. Along with demographic information, this additional information can serve as a means to increase the resolution of predictions about travel mode choice based on observed data.

3.2.1 Economic Factors

The research team acknowledges the importance that travel cost plays on a traveler's choice of transportation mode. Unfortunately, the NHTS did not collect data that characterize the different travel costs associated with the available mode choices. To overcome this problem, previous research (Ashiabor et al, 2007 for example) has developed synthetic travel cost estimates for each mode of transportation between major origin/destination pairs using such resources as published airline fares, rail and bus fare schedules, and mileage between various geographic destinations. In addition, assumptions were made as to the extra costs incurred on the trip (access/egress transportation, overnight lodging, etc.). This provided a generalized cost estimate for each trip. The resources available to the current research project described in this report did not allow for this type of data collection and use. Furthermore, this method involves making a lot of assumptions about the costs of travel that the research team did not feel warranted making. So, as an alternative, this research focused on creating a generalized cost component for the model based on major economic indicators related to travel at the time of the long-distance trip. This section describes that process.

The first group of non-NHTS variables included in the model seeks to capture any existing economic effects that drive the actions of consumers of different modes of travel. Each potential mode of travel for a long distance trip has its own economic burden; for example, driving a personal vehicle incurs the cost of paying for gas and any risk of repairs while flying on a domestic airline incurs the cost of a ticket. These factors, in conjunction with demographic data such as income levels, serve as deterrents or incentives for individuals to choose one mode of travel over another. The economic effect of cost is not entirely captured through income level – if the price of an airline ticket becomes sufficiently low, an increasing portion of consumers will choose to substitute airline travel for personal vehicle travel even when holding income level constant. Listed below are the variables included that address these price changes and the resulting effect they have on the desirability of certain travel modes.

- **Air Travel Price Index**: The Air Travel Price Index (ATPI) is a statistical index that denotes the relative price levels of airfares faced by consumers over time. The Research and Innovative Technology Administration (RITA) of USDOT compiles this index, beginning in the first business quarter of 1995, by matching identical routings and airfare classes and the changes in their costs over time at quarterly intervals. Three different types of ATPI measurements are provided by RITA depending on the origin of the flight; this analysis uses only the U.S.-origin ATPI in an attempt to limit any foreign airfare price effects. RITA also provides the average airfare price over this time interval, but the price index is advantageous to a national average due to direct routing-price pair wise comparisons used in its calculation which may differ over domestic locations (an average masks potential local differences). The index takes value 100 in the first quarter of the first year (1995), and then changes based on the relative magnitude of increase or decrease in overall airfare prices in subsequent quarters. The change in the value of the ATPI over time is shown below in Figure 3-1:

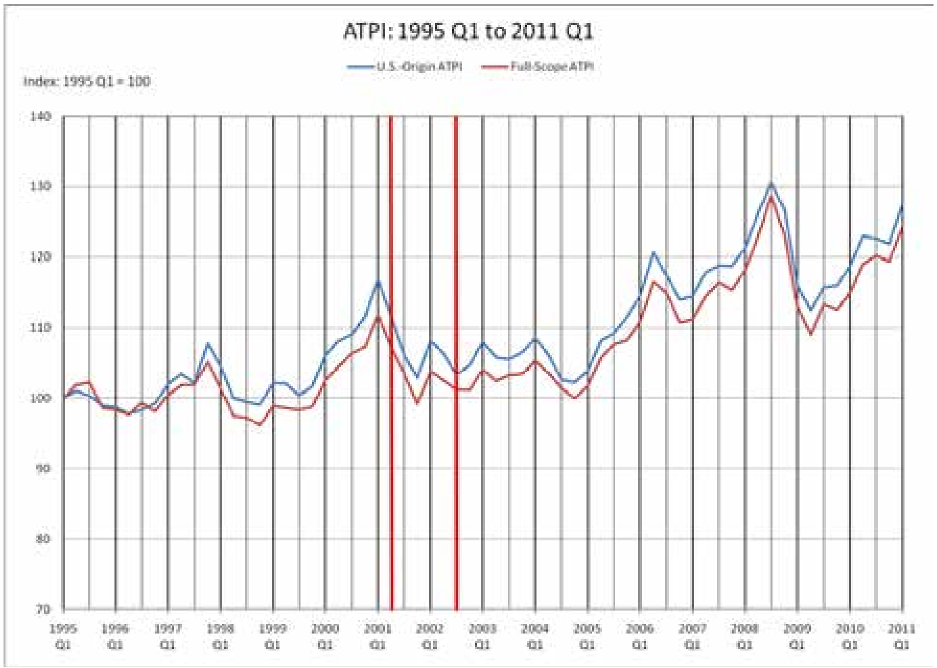

Figure 3-1. Plot of Air Travel Price Index from 1995 to 2011.

The red vertical bars denote the range of travel dates observed in the NHTS sample based on the travel periods for each respondent; the relevant airfare price levels faced by consumers over this time period vary considerably, chiefly due to the effect of the 9/11 attacks on the air travel industry and demand for air travel. This variable is intended to provide a measure of consumers' price thresholds for air travel modes, with the expected model effect being an increase in the relative price of airline travel corresponds with a decrease in the likelihood of choosing air travel as the desired mode. Also, changes in the price of airfare may be correlated with increased chances of choosing other modes as consumers substitute towards less expensive alternatives. The price of the index was recorded at the time of the travel period for each respondent. Although the purchase time of the air transportation would be preferable, it was not available in the NHTS data.

- **Consumer Price Index Private Transportation Component**: Similar to the price of airfares, economic incentives in mode choice will exist for the use of private transportation. Using the Consumer Price Index (CPI) commodity category for private transportation published the U.S. Bureau of Labor Statistics (BLS), the model can account for changes in the price of owning and operating a personal vehicle and assess any effect this has on the likelihood of choosing a transportation mode for long distance travel. The CPI for private transportation is calculated using the relative price changes in a variety of personal vehicle ownership cost categories and relative importance weights associated with each cost, with the index taking the baseline value of 100 for the years 1982-1984. The costs included in the aggregate private transportation index include:
  - the purchase and lease price of new and used motor vehicles;
  - the price of fuel;
  - the price of motor vehicle parts and equipment;

- the price of vehicle maintenance and repair; and
- the price of motor vehicle insurance and other fees.

Model sensitivity studies showed that the overall price levels for the aggregation of all of these items do not produce results that are statistically significant from models that use each of the individual price metrics; in other words, no one component cost of owning a motor vehicle seems to have a more predictive effect on mode choice than the general price level of all components. Therefore, the single index for all personal vehicle costs was included.

- **Consumer Price Index Public Transportation Component**: The BLS also publishes a monthly CPI that captures the general price levels of available public/commercial transportation options. Consumers of public/commercial transportation may be especially susceptible to changes in price in determining their mode choice for longer distance trips, as there are several disincentives to using public/commercial transportation over personal or air travel (time and privacy costs). The CPI for public transportation is calculated similarly to the methods described above for private transportation. The public/commercial transportation index also takes the baseline value of 100 for the years 1982-1984 and includes price change information on the following types of transportation:
  - Intercity bus fare;
  - Intercity train fare;
  - Ship/ferry fare; and
  - Intra-city mass transit.

  Statistical sensitivity analysis again showed that including each type of transportation's CPI individually did not yield significant improvements in model resolution, so the aggregate measure was used.

Other measures of the potential economic burden of specific travel modes were considered and excluded from the model's analysis based on analyses of their overall effect. Individual indices for both Amtrak train fares as well as cross-country bus fares were initially included in the model, but were found to be insignificant predictors. The CPI for public/commercial transportation yields virtually the same model effects, so these indices were excluded in order to avoid over-specifying the model and to avoid multicollinearity issues in the model fit. The index for airline prices was not highly correlated with either CPI measure; additionally, consumer demand for airline tickets is much more sensitive to price changes than public/commercial transportation. Also, measures of general economic conditions were initially included as potential indicators of the willingness of travelers to choose "high end" travel options during periods of prosperity; among these was the University of Michigan's Consumer Sentiment Index which tracks the general level of consumer confidence in the economy at a given point in time. These were also found to be unnecessary, as all of the price effects' impact on transportation mode choice was captured in the included indices and the effect of other economic predictors on the model's overall fit was minimal.

3.2.2 Availability of Transportation Infrastructure

A second group of variables was also included in the model analysis in order to account for factors outside those captured in the NHTS. A traveler's mode choice is likely to be affected not only by the price of a given service, but also its availability. A key component in this availability is proximity to points of access to a transportation option for both the primary and any secondary modes of trip travel. For example, a traveler's propensity to choose airline travel will not only be affected by the location of the airport itself, but also by secondary transportation options to and from the airport such as intercity rail. Some modes of transportation may be limited or may not be available in some areas, making personal vehicle the only feasible option for long distance travel. To account for any effect availability and access has on final travel mode choice, the model included variables that measure the level of transportation infrastructure and its proximity to the residence of travelers.

The locations of major hubs for various modes of transportation across the United States were assimilated to create a single set of transportation infrastructure sites. The database includes the locations of airports, standard rail stations, transit rail stations, and large bus depots. The airport locations were acquired from the National Transportation Atlas Database 2011 (NTAD2011) and represent all landing facilities in the U.S., as provided and maintained by the Federal Aviation Administration. The airports were filtered such that only those that would be used by a typical traveler were included. All private airports, heliports, ultralight ports, balloon ports and glider ports were excluded. In addition, those airports with no commercial activity and no central tower were assumed to be too small to be used by a casual traveler. Figure 3-2 shows the locations of the airports considered in this study.

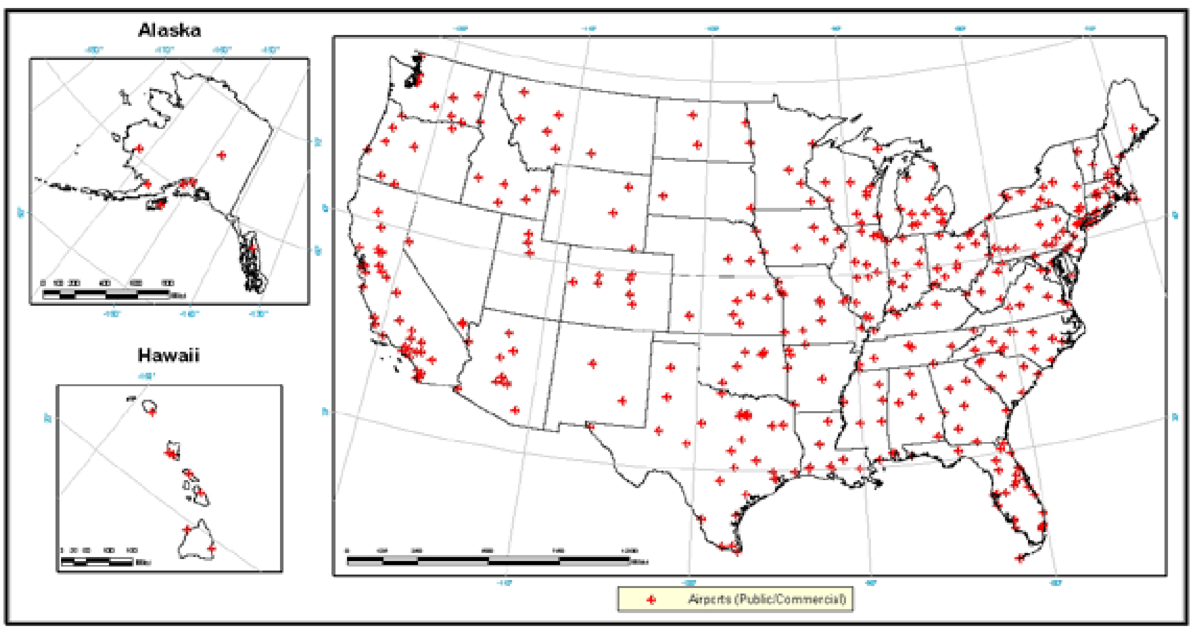

Figure 3-2. Locations of Large, Public-Use Airports from NTAD2011.

Both standard rail (i.e. Amtrak) and transit rail (e.g. light rail, subways, etc.) stations were acquired from NTAD2011 and included in the database (Figure 3-3 and 3-4). Noticeably missing from the NTAD2011 transit rail data were the New York City subway system and the Long Island Rail Road (LIRR). The locations of stations contained within these systems were acquired from the Metropolitan Transit Authority (MTA) General Transit Feed Specification (GTFS) (Figure 3-5). There are several other light rail and passenger rail systems that are not included in this dataset. Many of these were constructed or brought online after the most recent date observed in the NHTS sample data (corresponding to April, 2002) and are thus not considered to create a missing data issue given that they were not available at the time. Hence, the predictive ability of the model should not be hampered significantly when trying to predict trips during the time of the NHTS. Because of this, the count of light rail stations is included in the full prediction model in order to ascertain the general effect these stations have on mode choice. The NTAD notes that it will update its database with a significant amount of light and transit rail station data in late 2011, and there are potentially some transit rail observations missing from the infrastructure database. Given the model is going to be used to predict mode choice of future trips within the construct of a national transportation framework model and the amount of missing stations unknown at this time, this variable has been removed from the reduced prediction models which will be used in the near term for prediction.

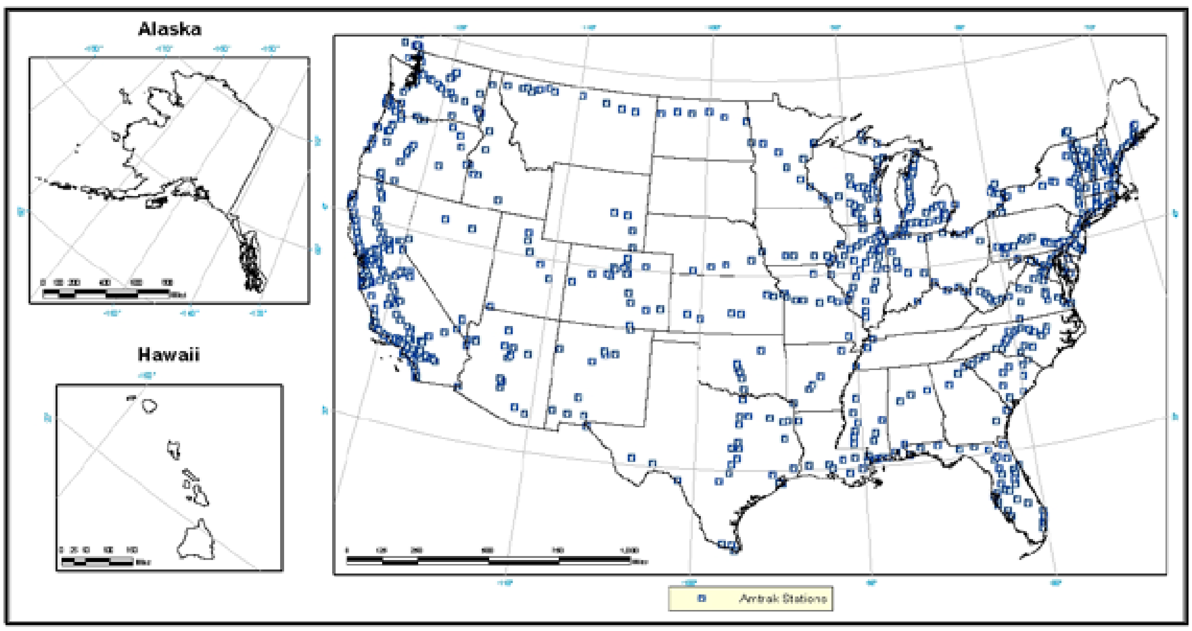

Figure 3-3. Locations of Amtrak Stations from NTAD2011.

Figure 3-4. Locations of Light Rail Stations from NTAD2011.

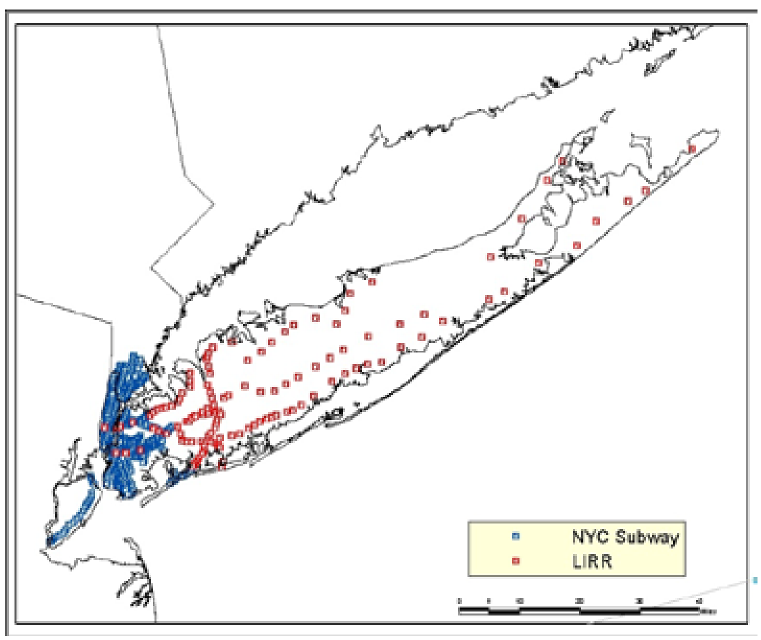

Figure 3-5. MTA NYC Subway and LIRR Stations.

While there are a vast number of single public transit bus stops throughout the country, only the major bus stations or depots were considered for this study. These stations included only major transfer or hub sites, which travelers would generally need to access for long distance travel (i.e. not intracity travel). The locations of bus stations or depots were obtained from NTAD2011 and are shown in Figure 3-6.

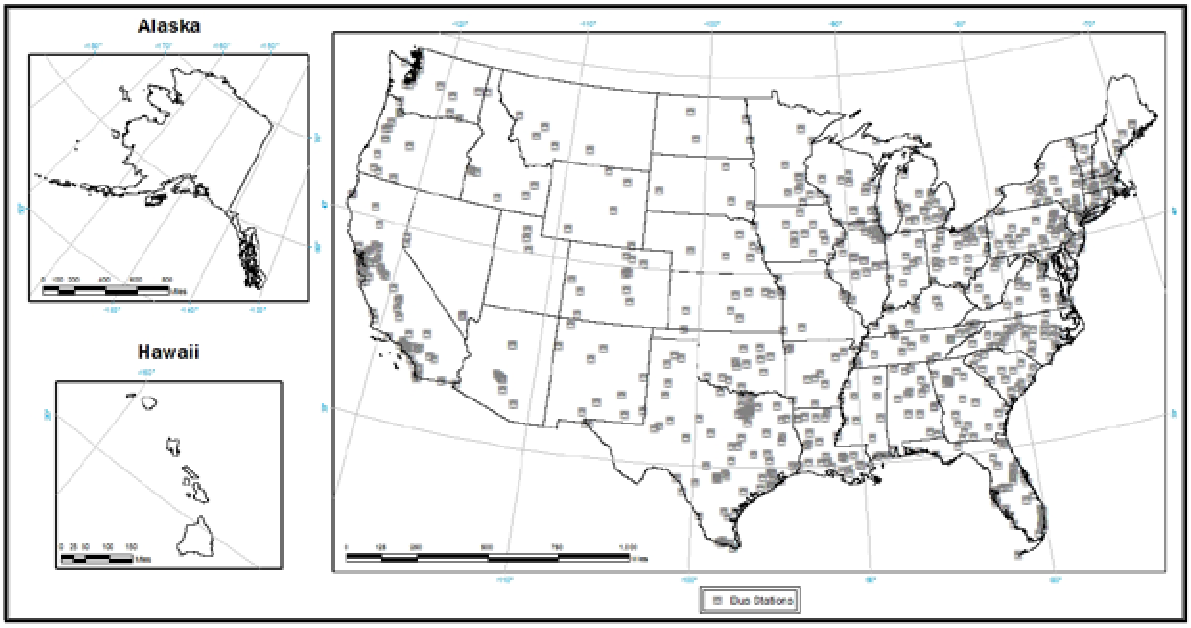

Figure 3-6. Locations of Large Bus Stations.

In order to calculate a measure of accessibility for each survey respondent, available transportation infrastructure locations were matched to each survey respondent. The highest level of geographic location information collected from the survey respondents was the 5-digit ZIP Code of residence. The 5-digit ZIP Code for each survey respondent was geocoded to the delivery-based ZIP Code centroid to represent the origin location. There were 12 ZIP Codes that could not be geocoded; two of which were not valid ZIP Codes. The remaining 10 ZIP Codes were manually assigned the data associated with the closest ZIP Code using the associated city name from the U.S. Postal Service database.

It was estimated that 25 miles was a reasonable travel distance from a respondent's residence to a transportation hub. This distance is assumed to represent a basic awareness of local travel options by each respondent as well as the ability to reach infrastructure hubs within this distance using personal vehicles or public transit as an intermediate step in the overall trip. The counts of each type of transportation mode that fell within the buffered distance were calculated to represent the respondents' access to alternative transportation (Figure 3-7).

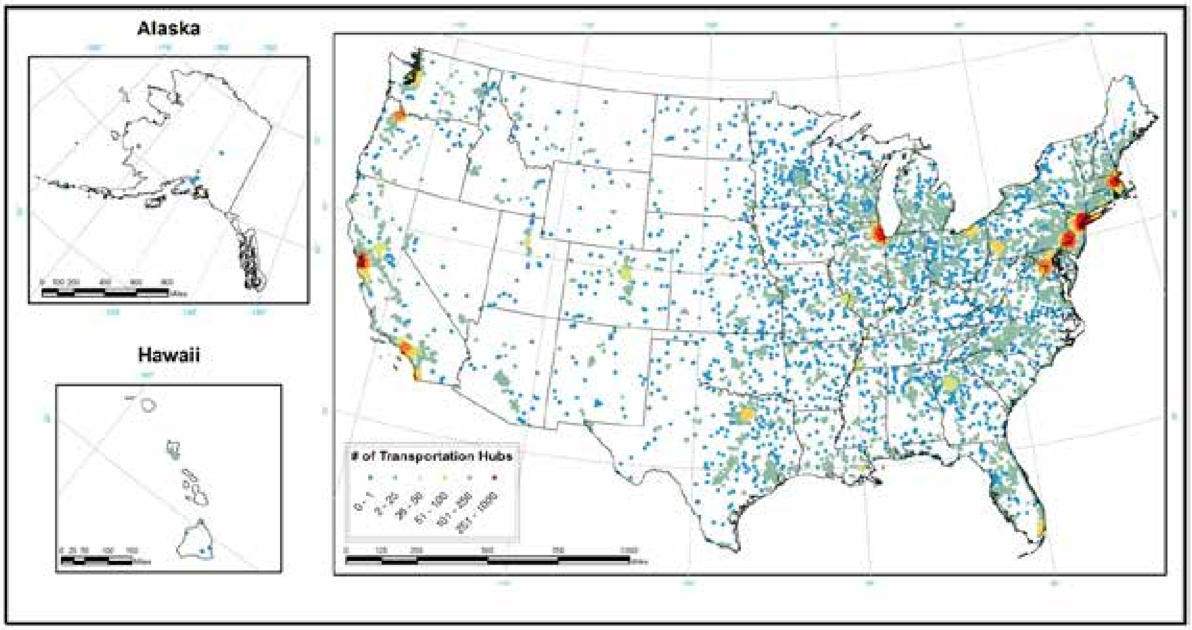

Figure 3-7. Number of Transportation Hubs Within 25 Mile Buffer for Each Survey Respondent.

Within the 25 mile buffer radius, counts of infrastructure sites were summed and included in the model as variables measuring access to different travel mode options. If travel mode choice is dependent on level of access, model results will show a significant relationship between marginal increases in the number of travel mode infrastructure sites within the buffer distance. For example, if the choice of air travel is highly dependent on access to airports, marginal increases in the number of airport sites within a traveler's access radius (say, from 0 to 1 airport sites) should yield significantly increased probabilities of taking air travel. The final infrastructure count sums within each survey respondent's assumed 25 mile access radius were compiled in four variables included in the model listed below:

- Count of all air travel sites;
- Count of all light and transit rail sites;
- Count of all standard rail sites; and
- Count of all bus travel sites.

The final set of variables included in the model was chosen based on the statistical considerations mentioned in the discussion above as well as a series of pair-wise and overall correlation analyses. This involved using statistical software to search out combinations of different variables for highly correlated variables, which if included in the model would essentially be duplicating the analysis of any effect on the travel mode outcome and create multicollinearity problems with the logistic regression model fits. Using traditional correlation matrices, a number of price index variables acquired from the St. Louis Federal Reserve were discarded due to their high correlations with one another. Also, a number of measurement indices of consumer confidence were found to be highly correlated with measures of price levels for public and private transport and were thus discarded from consideration. Several measures of population density, metropolitan statistical area classification, and household demographics in the original NHTS data set were also found to be correlated with one another; in all cases, only one metric was chosen to be included in the final set of analysis variables as determined by examination of the correlation matrices. If the factor chosen for the model had a significant effect on mode choice, then it was noted that the outcome may be linked to either the factor in the model or one of the excluded variables that were correlated with the factor in the model.

Additionally, some preliminary maximum likelihood models using the overall sample of data were used to assess preliminary model fit and further refine the set of variables used. Some variables that displayed mixed correlation results, such as the University of Michigan consumer demand index and the RITA price index for Amtrak fares, were found to have negligible effects on predicting probabilities in preliminary model runs and were not considered further. Statistical verification of improved model fits was observed after dropping these additional variables, and variables were further tested against preliminary model runs using sample data subset by each trip purpose.

3.3 Summary of Predictive Factors Used in Mode Choice Modeling

Table 3-1 provides a summary of the prediction factors discussed in the previous few sections that were used in the mode choice analysis. The factors are grouped by type and contain information on the coding of the categorical variables.

Table 3-1. Summary of Predictive Factors Used in Mode Choice Modeling.

Type of Factor	Factor	Description
Traveler Characteristics	Traveler's Age	Integer
	Household Income	Four categorical, dichotomous variables: $0<=Income<=$30,000 (1=yes, 0=no) $30,000<Income<=$60,000 (1=yes, 0=no) $60,000<Income<=$100,000 (1=yes, 0=no) $100,000<Income (1=yes, 0=no)
	Race*	Five categorical, dichotomous variables: White (1=yes, 0=no) African-American (1=yes, 0=no) Asian (1=yes, 0=no) Hispanic (1=yes, 0=no) Other (1=yes, 0=no)
	Weekly Internet Use*	Categorical (1=yes, 0=no)
	Weekly Use of Public/Commercial Transportation*	Categorical (1=yes, 0=no)
	Traveler is Employed	Categorical (1=yes, 0=no)
	Count of Vehicles in Household	Integer (counts)
Land-Use Characteristics	Household in Urban Area	Categorical (1=yes, 0=no)
	Population per Square Mile of Household	Continuous
Trip Characteristics	Trip Occurred on Weekend*	Categorical (1=yes, 0=no)
	Number of People on Trip	Integer (counts)
	Trip Distance	Continuous
	Nights Away on Trip*	Integer (count)
Availability of Transportation Infrastructure	Count of all Airports within 25 Mile Radius of Household	Integer (count)
	Count of all Bus Depots within 25 Mile Radius of Household	Integer (count)
	Count of all Amtrak Stations within 25 Mile Radius of Household	Integer (count)
	Count of all Transit/Subway/Light Commuter Train Stations within 25 Mile Radius of Household*	Integer (count)
Economic	CPI for Private Transportation – Seasonally Adjusted	Continuous
	CPI for Public Transportation – Seasonally Adjusted	Continuous
	RITA Airline Ticket Price Index	Continuous
Other	Post 9/11	Categorical (1=yes, 0=no)

* Factor included in full set of prediction models but not included in reduced set of predicted models

3.4 Descriptive Analysis

The final data file for the passenger choice modeling was compiled using variables from the NHTS and the supplemental data sources. Trips with missing values for any of the variables were excluded, which reduced the dataset to 28,402 long-distance trips. Table 3-2 shows the unweighted number of long-distance trips used in the modeling by trip purpose and travel mode. Note that personal business trips represent a smaller subset of the data set (11 percent of trips) relative to business and pleasure travel purposes. Also, the vast majority of survey respondents took personal vehicles on their trips (88 percent of trips), regardless of purpose. Air was chosen in about 9 percent of the trips while bus (1.5 percent) and train (1 percent) were chosen less frequently. This should yield several expected consequences in analysis, namely that the analysis model should have the most informed predictions of travel mode choice for personal vehicles given the discrepancy in the resolution of the available data. In addition, it is possible that the relative lack of responses for bus and train trips, even compared to air travel, could introduce small sample biases into predictive analyses of bus and train travel outcomes, especially for personal business trips.

Table 3-2. Number of Long-Distance Trips Used in Modeling by Trip Purpose and Travel Mode.

	Personal Vehicle	Air	Bus	Train	Total
Business	8,443	1,244	105	195	9,987
Pleasure	13,416	1,195	203	61	14,875
Personal Business	3,224	186	116	14	3,540
Total	25,083	2,625	424	270	28,402

Table 3-3 displays the weighted descriptive statistics for each analysis variable. For each variable the mean and standard deviation (shown in parentheses) is provided for the following: (1) all trips; (2) by trip purpose across modes of transportation; and (3) by mode of transportation across the trip purposes.

Table 3-3. Descriptive Statistics (Means and Standard Deviations) for Each Model Factor by Trip Purpose and Transportation Mode.

Factor	All Trips	Trip Purpose			Transportation Mode			
		Business	Pleasure	Personal Business	Personal Vehicle	Air	Bus	Train
$0<=Income<= $30,000	0.11 (0.00)	0.07 (0.00)	0.12 (0.00)	0.18 (0.01)	0.12 (0.00)	0.05 (0.00)	0.15 (0.02)	0.09 (0.02)
$30,000<Income<= $60,000	0.32 (0.00)	0.28 (0.00)	0.34 (0.00)	0.35 (0.01)	0.33 (0.00)	0.17 (0.01)	0.38 (0.02)	0.26 (0.03)
$60,000<Income<= $100,000	0.32 (0.00)	0.36 (0.00)	0.31 (0.00)	0.29 (0.01)	0.33 (0.00)	0.31 (0.01)	0.30 (0.02)	0.37 (0.03)
$100,000<Income	0.25 (0.00)	0.29 (0.00)	0.23 (0.00)	0.18 (0.01)	0.22 (0.00)	0.48 (0.01)	0.17 (0.02)	0.29 (0.03)
Post 9/11	0.62 (0.00)	0.64 (0.00)	0.61 (0.00)	0.60 (0.01)	0.62 (0.00)	0.60 (0.01)	0.69 (0.02)	0.61 (0.03)
African-American	0.03 (0.00)	0.02 (0.00)	0.03 (0.00)	0.05 (0.00)	0.03 (0.00)	0.03 (0.00)	0.08 (0.01)	0.03 (0.01)
Asian	0.01 (0.00)	0.01 (0.00)	0.02 (0.00)	0.02 (0.00)	0.01 (0.00)	0.02 (0.00)	0.02 (0.01)	0.01 (0.01)
Hispanic	0.01 (0.00)	0.02 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	0.00 (0.00)
Other	0.02 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.00)	0.02 (0.00)	0.02 (0.00)	0.03 (0.01)	0.04 (0.01)
White	0.92 (0.00)	0.93 (0.00)	0.92 (0.00)	0.89 (0.01)	0.92 (0.00)	0.92 (0.01)	0.86 (0.02)	0.92 (0.02)
Urban HH	0.71 (0.00)	0.69 (0.00)	0.73 (0.00)	0.66 (0.01)	0.69 (0.00)	0.88 (0.01)	0.73 (0.02)	0.75 (0.03)
Trip occurred on weekend	0.25 (0.00)	0.09 (0.00)	0.36 (0.00)	0.23 (0.01)	0.24 (0.00)	0.34 (0.01)	0.22 (0.02)	0.19 (0.02)
Respondent is employed	0.82 (0.00)	0.97 (0.00)	0.76 (0.00)	0.66 (0.01)	0.82 (0.00)	0.86 (0.01)	0.66 (0.02)	0.93 (0.02)
CPI Private Transport, seasonally adjusted	147.98 (3.09)	147.93 (3.13)	147.99 (3.05)	148.11 (3.14)	147.99 (3.09)	147.99 (3.04)	147.63 (3.02)	147.65 (3.29)
CPI Public Transport, seasonally adjusted	209.69 (1.51)	209.70 (1.49)	209.70 (1.53)	209.63 (1.51)	209.69 (1.51)	209.69 (1.55)	209.61 (1.33)	209.76 (1.54)
Airline ticket price index	107.69 (3.69)	107.78 (3.70)	107.57 (3.69)	107.98 (3.60)	107.68 (3.70)	107.79 (3.61)	107.82 (3.71)	107.80 (3.11)
Respondent's age	43.83 (13.97)	43.49 (11.15)	43.56 (15.01)	45.89 (16.21)	43.89 (13.95)	43.95 (12.90)	39.58 (19.82)	43.55 (13.06)
Population per sq mile	3176.46 (4816.23)	3027.27 (4532.98)	3365.51 (5062.07)	2802.99 (4485.77)	2992.63 (4593.32)	4660.27 (5982.20)	3879.72 (5751.54)	4724.26 (7269.35)
Count of vehicles in HH	2.60 (1.26)	2.66 (1.29)	2.56 (1.22)	2.60 (1.30)	2.63 (1.27)	2.39 (1.11)	2.53 (1.29)	2.29 (1.23)
Weekly use of public/commercial transportation	0.12 (0.00)	0.13 (0.00)	0.12 (0.00)	0.08 (0.00)	0.09 (0.00)	0.27 (0.01)	0.29 (0.02)	0.69 (0.03)

Table 3-3. Descriptive Statistics (Means and Standard Deviations) for Each Model Factor by Trip Purpose and Transportation Mode. (Continued)

Factor	All Trips	Trip Purpose			Transportation Mode			
		Business	Pleasure	Personal Business	Personal Vehicle	Air	Bus	Train
Weekly web use	0.75 (0.00)	0.76 (0.00)	0.75 (0.00)	0.74 (0.01)	0.74 (0.00)	0.87 (0.01)	0.79 (0.02)	0.87 (0.02)
Count of all airports in 25M radius	1.16 (1.19)	1.12 (1.14)	1.22 (1.23)	1.04 (1.18)	1.09 (1.16)	1.79 (1.33)	1.14 (1.24)	1.52 (1.39)
Count of all Amtrak stations in 25M radius	2.41 (3.25)	2.42 (3.43)	2.51 (3.20)	1.99 (2.90)	2.30 (3.19)	3.44 (3.67)	2.13 (2.84)	3.43 (3.21)
Count of all bus depots in 25M radius	2.15 (2.21)	2.15 (2.23)	2.23 (2.22)	1.80 (2.05)	2.05 (2.16)	3.14 (2.43)	1.96 (1.97)	2.54 (2.07)
Count of all transit/subway/light/ commuter train stations in 25M radius	33.31 (99.86)	30.09 (90.81)	37.02 (106.71)	26.82 (93.64)	29.50 (93.30)	64.16 (134.73)	38.48 (120.98)	79.49 (168.25)
Nights away on trip	1.51 (4.65)	0.91 (3.16)	1.92 (4.41)	1.51 (7.88)	1.16 (4.28)	5.04 (6.62)	1.04 (2.23)	1.06 (3.34)
Trip distance	283.12 (604.43)	270.57 (612.52)	304.28 (630.14)	229.60 (446.21)	163.05 (230.20)	1442.20 (1385.12)	236.23 (325.20)	241.52 (538.76)
Number of people on trip	2.61 (3.92)	1.68 (2.21)	3.02 (3.74)	3.51 (6.82)	2.34 (1.76)	2.64 (4.67)	18.33 (21.03)	2.30 (5.23)

The majority of survey respondents were white, employed, and lived in urban areas. While sample weights are used in the analysis to help offset this sample composition, summary statistics indicate that model results are unlikely to estimate large race effects simply due to the sample composition. Demographic variables that indicate the type of income distribution observed in this sample show that only a small portion of respondents reported a total household income of less than $30,000 per year. The remainder of income level categories was fairly evenly distributed across all respondents, although different trip purposes and travel modes did indicate some skewed income distributions; both business travel and air travel tended to be skewed towards higher income levels. About 62 percent of all long-distance trips were taken after the 9/11 terrorist attacks (this trend holds across specific travel modes and trip purposes), so if there are significant effects on travel behavior due to the ramifications of the attacks they should be observed in the analysis.

Summary statistics of individual trip purposes do display evidence for using separate models across trip purposes. For example, business trips tended to be less likely to occur on the weekends and had fewer average nights away when compared to other trip types. Based on the research on past travel mode studies discussed in Section 2.0 [Georggi and Pendyala (1999), Ashiabor et al (2007)], there is good reason to believe that trip purpose-specific attributes like those mentioned above could lead to fundamentally different behaviors in choosing a travel mode choice. Average trip distances for each of the three trip types also varied, with pleasure trips having the longest route to destination distance.

Respondent attributes that can serve as indicators for travel preferences remained fairly constant across trip purposes, but varied somewhat for different chosen travel modes. For example, high frequency (weekly) use of public/commercial transportation, high frequency (weekly) use of the internet, and type of residence area all varied considerably between different chosen modes of transportation. This could indicate some degree of underlying self-selection propensity among people who choose different modes of transportation that is driven by factors other than those that go into the behavioral choice of travel mode. Model results for some travel mode choices, therefore, will need to be examined with respect to these observed propensities when making predictive conclusions.

3.5 Statistical Modeling Methodology

Discrete choice models are statistical procedures that model choices made by people among a finite set of alternatives. Specifically, discrete choice models statistically relate the choice made by each person to the attributes of the person and the attributes of the alternatives available to the person. In terms of long-distance travel, discrete mode choice models consider the travel mode that travelers choose for a particular long-distance trip based on certain attributes about the traveler or the trip to be taken. Although discrete choice models can take many forms, the majority of the mode choice models involving transportation are logit based. The mathematical framework of logit models in based on the theory of utility maximization which is discussed in detail in Ben-Akiva and Lerman (1985). Utility theory assumes that travelers prefer an alternative with the highest utility where utility is a representation of the attractiveness of the mode choice alternatives as derived from the traveler. Logistic regression models are used to predict the probabilities of the different possible outcomes of a categorical dependent variable (mode choices), given a set of independent variables (socioeconomic characteristics, trip purpose, trip length, etc).

3.5.1 Multinomial Logit Model

A multinomial logit model is a regression model which generalizes logistic regression by allowing more than two discrete outcomes. It is a model that is used to predict the probabilities of the different possible outcomes of a categorical dependent variable, given a set of independent variables. Figure 3-8 presents an example of a simple multinomial logit model specification. This is the same graphic displayed as Figure 2-1 but is shown again in this section to assist the reader in better understanding the prediction model. Possible levels of the dependent variable (mode choice) used for this study are shown. The independent variables are those factors used to explain or predict the mode choice (e.g. trip length, trip purpose, demographics of travel).

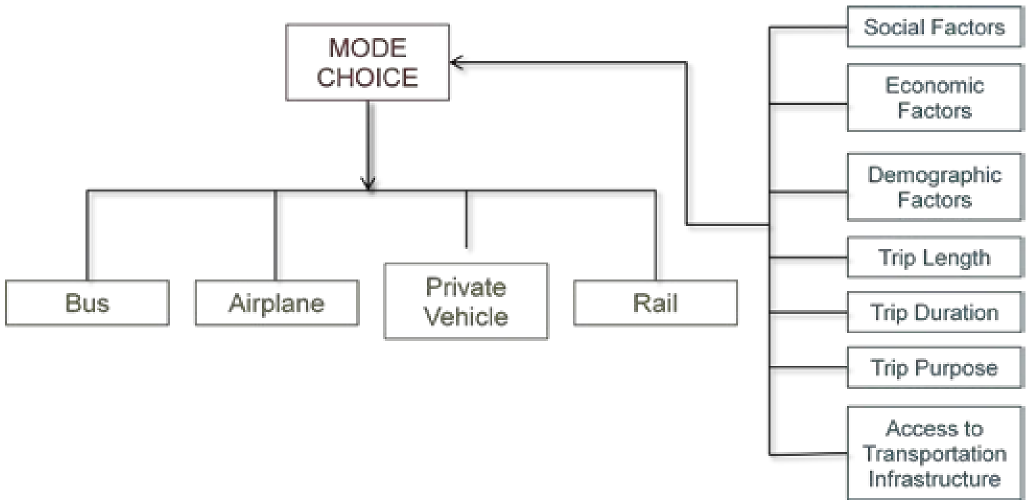

Figure 3-8. Visualization of Simple Multinomial Logit Model.

The mathematical form of the multinomial logit model is as follows. Suppose there are m total travel modes of interest ($1, 2, 3, \ldots, M$) and that there are k factors ($1, 2, 3, \ldots, K$) that are being used to predict the probability of a particular mode choice. These k factors in general may include continuous, binomial, or categorical data. To construct the logits in the multinomial case, one of the modes is considered the reference level and all other logits are constructed relative to it. Any mode can be taken as the reference level since there is no inherent ordering to the modes. Here mode M is taken as the reference level. The probability of an individual i selecting a travel mode m, out of M number of total available modes, is represented as p_{im}. The relationship between this probability and the K factors is given by the following multinomial logistic regression model

$$\log\left(\frac{p_{im}}{p_{iM}}\right) = \beta_{0m} + \beta_{1m}x_{1i} + \beta_{2m}x_{2i} + \cdots + \beta_{km}x_{ki}; \qquad m = 1, 2, \ldots, M\text{-}1, \; i = 1, 2, \ldots n. \quad (1)$$

where,

$x_{1i}, \ldots, x_{ki}$	are the k number of factors of mode m for individual i;
β_{0m}	is the mode specific constant for mode m;
$\beta_{1m}, \ldots, \beta_{km}$	are k number of coefficients of mode m which need to be estimated from the data;
M	is the set of all available travelling modes; and
n	is the number of individual/trip combinations in the dataset.

The above equation can be solved to yield the probability of an individual i selecting a travel mode m, out of M number of total available modes as

$$p_{im} = \frac{\exp(\beta_{0m}+\beta_{1m}x_{1i}+\beta_{2m}x_{2i}+\cdots+\beta_{km}x_{ki})}{1+\sum_{l=1}^{M-1}\exp(\beta_{0l}+\beta_{1l}x_{1i}+\beta_{2l}x_{2i}+\cdots+\beta_{kl}x_{ki})} \quad (2)$$

and for the reference category,

$$p_{iM} = \frac{1}{1+\sum_{l=1}^{M-1}\exp(\beta_{0l}+\beta_{1l}x_{1i}+\beta_{2l}x_{2i}+\cdots+\beta_{kl}x_{ki})} \quad (3)$$

For this research, the model in Equation (1) was fitted separately to the three different trip purposes (business, personal business, and pleasure). The number of modes (M) was equal to four (personal vehicle, air, bus, and train) where personal vehicle was considered the base level. The predictive factors included in each model are summarized in Table 3-1.

The form of the discrete choice multinomial logit model used in this research is based on the assumption that the choice of mode is a function of the characteristics of the traveler and/or the trip. This is known as a generalized multinomial logit model or unconditional multinomial logit model. The NHTS neither collected data that characterize the different available mode choices (e.g., travel time or cost under each of the mode options) nor did it provide information about other alternative modes or the traveler's reason for selecting a specific mode of travel over another mode. As such, a conditional multinomial logit model, a model where the choice of mode is a function of the characteristics of the respective modes themselves, could not be utilized without developing synthetic estimates for variables such as travel cost. This has been done in previous research (Ashiabor et al, 2007). The resources available to this research project did not allow for this type of data collection and use. Travel cost and other attributes not found in the NHTS are accounted for through economic and other proxies described in Section 3.2.

3.5.2 Model Estimation

The 2001 NHTS provides an analysis weight for each long-distance trip. The weight is defined at the person trip/travel period level. These weights reflect the selection probabilities and adjustments to account for nonresponse, undercoverage, and multiple telephones in a household. Point estimates of population parameters as well as coefficients of predictors are impacted by the value of the analysis weight for each observation. To obtain estimates that are minimally biased the analysis weight (WTPTPFIN) was used to weight the results.

Coefficients associated with each predictive factor were estimated using the maximum likelihood estimation technique in the SAS® (version 9.3) statistical software package. The SURVEYLOGISTIC procedure was used to take into account the complex nature of the 2001 NHTS sample design. This procedure was preferred over the CATMOD and PHREG procedures both of which can perform multinomial logistic regression but are based on the assumption that the sample is drawn from an infinite population by simple random sampling. If the sample is actually selected from a finite population using a complex design, these procedures generally do not calculate the estimates and their variances correctly. Namely, they fail to take into account the following characteristics of sample survey data that are present in the 2001 NHTS data and hence, generally underestimate the variance of point estimates and model coefficients:

1. Unequal selection probabilities;
2. Stratification;
3. Clustering of observations; and
4. Nonresponse and other adjustments.

The SURVEYLOGISTIC procedure fits linear logistic regression models while incorporating complex survey sample designs, including designs with stratification, clustering, and unequal weighting. In this research, the jackknife variance estimation method was used. The jackknife is a replication-based variance estimation method whereby subsamples of the original sample (replicate samples) are taken and the model coefficients are estimated for each replicate sample. The variability of the estimated model coefficients among the replicate samples is then used as a replication-based estimator of variance. Replicate weights calculated using the delete-one Jackknife method and provided on the 2001 NHTS website were used in the modeling.

Model coefficients for the predictor variables were estimated from the model. Logistic regression coefficients are difficult to interpret because they measure the effect that a change in an independent variable would have on the log odds of choosing a particular mode choice. As a result, the coefficient estimates in this analysis were transformed into marginal probability effects. Marginal probability effects are more intuitive in that they represent the effect that a change in the independent variable would have directly on the probability of choosing the mode choice. STATA (version 11) was used to calculate these marginal estimates as the SURVEYLOGISTIC procedure does not support this capability.

To assess the model fit, goodness of fit statistics such as the overall model chi-square, log-likelihood values, and pseudo- R^2 values were examined. These statistics provided evidence of a good model fit (i.e. they have values close to 1). While multinomial logistic regression does compute these measures to estimate the strength of the relationship, these correlation measures alone do not provide sound evidence for determining and estimating the accuracy or errors associated with the model. Moreover, the overall model chi-square, log likelihood values, and pseudo- R^2 values can become quite large for data with large weights and this results in the generalized R-square almost always being 1. Therefore, to assess the model's predictive ability, the model was applied to the dataset of trips to determine its predictive ability. Aggregate mode shares were calculated by summing the calculated probabilities for each trip record. These were compared against the actual mode shares of the data set of trips in order to observe how well the model could replicate the observed mode shares.

3.6 Model Estimation Results

This section contains model estimation results for both the full mode-choice prediction models (Section 3.6.1) and the reduced mode-choice prediction models (Section 3.6.2). Results from the full model are presented to collectively assess the predictive ability of all variables identified in Section 3.2 for mode choice. These results identify those variables that are highly predictive of mode choice so that a general understanding of long-distance travel mode choice can be realized without the worry that some inputs are not readily available for future mode prediction. The reduced model can be used in a more practical sense to predict future mode choice within a transportation modeling framework as it contains only readily available input variables identified as having an influence on mode choice.

3.6.1 Full Prediction Models

Coefficient estimates and their standard errors for the multinomial logit models of travel mode choice are presented in Table 3-4, with one set of coefficient results for each travel purpose type. Separate model estimates are presented for each travel mode. Note that there are no coefficient estimates for the personal vehicle mode as that mode was the reference level. Thus, the logits for all other modes are constructed relative to it. Also for those categorical variables with more than two levels (income and race) one of the levels for each variable was used as the reference category and thus no coefficients were estimated. For income, estimates for all levels were made relative to the greater than $100,000 category while for race, estimates for all levels were made relative to white travelers. Coefficient estimates significant at the 1, 5, and 10 percent level of significance are noted with a '**', '*', and '+', respectively.

Table 3-4. Coefficient Estimates (Standard Errors) for Full Set of Multinomial Logit Models of Travel Mode Choice.

	Business				Pleasure				Personal Business			
	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train
Post 9/11 (d)		-0.3461	-0.5787	-1.3608*		-0.2518	0.8113	0.1744		-0.5193	0.7281	-3.2055
		(0.2737)	(0.9486)	(0.6298)		(0.2288)	(0.5605)	(0.7392)		(0.5656)	(0.4754)	(2.8332)
Trip occurred on weekend (d)		0.4426*	-0.5726	1.7378*		0.7664**	0.1232	0.0515		0.9560**	-0.5491	2.3317
		(0.2009)	(0.8082)	(0.6680)		(0.1476)	(0.3343)	(0.3950)		(0.3362)	(0.4833)	(2.2458)
Nights away on trip		0.0510	-0.1893	-1.0593*		-0.0197	-0.2212*	0.0109		-0.0140	-0.1253+	-0.9871
		(0.0525)	(0.3612)	(0.5286)		(0.0155)	(0.1068)	(0.0222)		(0.0128)	(0.0686)	(0.6754)
Number of people on trip		0.1130**	0.2971**	-0.0467		-0.0175	0.1838**	0.1134		-0.0937	0.2711**	0.1448
		(0.0402)	(0.0677)	(0.2922)		(0.0348)	(0.0194)	(0.1570)		(0.1563)	(0.0576)	(0.1033)
Respondent's age		-0.0020	0.0683+	0.0133		-0.0067	-0.0076	-0.0089		-0.0075	-0.0350+	-0.0088
		(0.0086)	(0.0365)	(0.0216)		(0.0044)	(0.0121)	(0.0170)		(0.0118)	(0.0207)	(0.0388)
Trip distance		0.0054**	0.0036	0.0036		0.0043**	0.0014**	0.0021**		0.0049**	0.0024**	0.0031
		(0.0007)	(0.0025)	(0.0026)		(0.0003)	(0.0005)	(0.0005)		(0.0006)	(0.0009)	(0.0021)
Count of vehicles in HH		-0.4201**	0.0419	-0.0142		-0.1945*	-0.2017	-0.6012		0.0005	-0.2473	-0.3821
		(0.0975)	(0.1326)	(0.2602)		(0.0823)	(0.1423)	(0.3746)		(0.1306)	(0.2898)	(1.0886)
Urban HH (d)		0.5276*	0.3258	-0.4369		0.8068**	-0.3751	0.2356		0.4905	-0.7594	-0.7856
		(0.2560)	(1.0243)	(0.8495)		(0.2750)	(0.3639)	(0.8725)		(0.4459)	(0.7307)	(1.7244)
Population per sq mile		-0.0000	0.0000	0.0000		-0.0000	0.0000	-0.0000		0.0001	-0.0001	0.0000
		(0.0000)	(0.0001)	(0.0001)		(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0001)
Count of all bus depots in 25M radius		-0.0118	-0.1071	-0.1948		-0.0109	-0.0543	0.0950		-0.0632	-0.2657*	0.2986
		(0.0625)	(0.1715)	(0.1694)		(0.0455)	(0.0770)	(0.1533)		(0.1174)	(0.1228)	(0.4052)
Count of all airports in 25M radius		0.3199**	-0.1117	-0.1150		0.1221	-0.2026	-0.1826		-0.1406	0.9176**	0.7236
		(0.1157)	(0.5570)	(0.3497)		(0.0862)	(0.1630)	(0.3672)		(0.2679)	(0.1899)	(0.8637)
Count of all Amtrak stations in 25M radius		-0.0273	-0.0822	0.0336		0.0243	0.0087	-0.0356		-0.0405	0.0993	-0.1798
		(0.0314)	(0.1618)	(0.0717)		(0.0220)	(0.0663)	(0.0915)		(0.0744)	(0.1165)	(0.3293)
Count of all transit/subway/light/ commuter rail stations in 25M radius		-0.0019	0.0017	0.0017		0.0004	0.0002	0.0039		0.0002	0.0003	-0.0025
		(0.0012)	(0.0036)	(0.0032)		(0.0006)	(0.0013)	(0.0033)		(0.0024)	(0.0018)	(0.0077)
CPI Private Transport, seasonally adjusted		-0.0338	-0.2557	-0.2005		-0.0663+	0.0888	-0.0205		-0.0105	0.1069	0.0294
		(0.0403)	(0.1906)	(0.1462)		(0.0342)	(0.0675)	(0.1106)		(0.0815)	(0.0827)	(0.5872)
CPI Public Transport, seasonally adjusted		-0.0161	-0.1172	0.0669		0.0833	-0.1087	0.0057		0.0161	0.1254	-0.9336
		(0.0597)	(0.2254)	(0.1722)		(0.0584)	(0.0670)	(0.1574)		(0.1064)	(0.1942)	(0.8419)

Table 3-4. Coefficient Estimates (Standard Errors) for Full Set of Multinomial Logit Models of Travel Mode Choice. (Continued)

	Business				Pleasure				Personal Business			
	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train
RITA airline ticket price index		0.0109	0.0541	0.0328		0.0012	0.0123	0.0676		-0.0194	0.0589	-0.4087
		(0.0316)	(0.1609)	(0.0734)		(0.0268)	(0.0486)	(0.0817)		(0.0603)	(0.0650)	(0.3565)
Weekly use of public/commercial transportation (d)		0.7521**	2.4052**	3.4173**		0.4212*	1.4132**	1.3816**		0.2743	0.6062	1.8978+
		(0.2167)	(0.7928)	(0.8193)		(0.1727)	(0.2421)	(0.4469)		(0.4617)	(0.6275)	(0.9929)
Weekly web use (d)		1.3580**	-0.2750	0.1228		0.1543	0.3689	0.0758		0.4193	0.4555	2.7481*
		(0.3729)	(0.9399)	(0.8525)		(0.1534)	(0.2951)	(0.5304)		(0.4511)	(0.6058)	(1.2755)
$0<=Income<= $30,000 (d)		-2.1512**	-1.4185	0.5590		-1.0238**	0.7819*	0.3160		-1.2558+	0.8275	-0.3230
		(0.5319)	(2.9038)	(0.9606)		(0.2806)	(0.3874)	(0.8443)		(0.7367)	(0.9144)	(27.0395)
$30,000<Income<= $60,000 (d)		-2.3332**	0.6971	-0.2128		-0.7273**	0.4677	-0.0407		-0.9963*	0.4815	-0.4246
		(0.3171)	(1.8419)	(0.8065)		(0.2357)	(0.3919)	(0.7911)		(0.4473)	(0.8650)	(1.8705)
$60,000<Income<= $100,000 (d)		-0.6823**	0.6674	-0.5281		-0.5683**	0.4795	0.1507		-0.8241*	0.9803	0.6000
		(0.1901)	(1.7838)	(0.6104)		(0.1883)	(0.4290)	(0.8674)		(0.3880)	(0.7145)	(1.3690)
$100,000<Income (d)	Omitted – Reference Category											
African-American (d)		-0.2763	-0.8361	-0.3266		-0.1580	1.3575**	-0.7992		-1.1674	0.0936	-25.5720**
		(0.4554)	(1.5783)	(1.1267)		(0.3887)	(0.4414)	(24.1714)		(0.9106)	(0.6214)	(6.0584)
Asian (d)		1.5184+	-28.4430	-1.3924		0.0668	-0.3040	-2.0277		-28.7018**	0.8211	-28.5955**
		(0.9113)	(28.7475)	(30.7627)		(0.6076)	(0.8522)	(23.6137)		(6.4247)	(1.2124)	(10.4811)
Hispanic (d)		0.3891	-24.7044**	-29.3618**		-0.2779	-0.2395	-24.6207**		-3.0742	-4.2341	-23.8825**
		(0.5068)	(5.8215)	(10.0109)		(0.6820)	(1.3593)	(5.5158)		(24.4496)	(22.5459)	(6.9829)
Other (d)		0.5121	-1.2343	1.1816		0.2508	-0.8336	1.3186		-1.3535	-0.4788	-27.8318**
		(0.8494)	(23.0448)	(31.5731)		(0.3891)	(1.2065)	(0.8097)		(24.8385)	(1.9753)	(7.6088)
White (d)	Omitted – Reference Category											
Respondent is employed (d)		Omitted – Business trips were assumed to occur only for survey respondents who were employed				0.1672	-0.7194**	-0.2637		0.3540	-0.2628	-0.4557
						(0.1479)	(0.2397)	(0.4686)		(0.3494)	(0.4162)	(1.4084)
Constant		3.1080	47.0182	7.6112		-12.4997	3.3135	-10.8349		-4.1895	-53.9195	226.8973
		(13.7928)	(50.0787)	(32.1389)		(13.6170)	(21.9081)	(41.3653)		(27.7211)	(35.6857)	(214.9397)

Note: Multinomial logit model coefficients were estimated relative to the reference mode of personal vehicle travel

+ Indicates statistical significance at the 10% level
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level
(d) Dichotomous variable

Raw model coefficient results for maximum likelihood models can indicate statistical significance and the direction of an effect that is attributable to a certain variable, but do not give meaningful insight into the actual probabilistic changes attributable to specific variables. To show a more useful interpretation, coefficient estimates in this analysis were transformed into marginal probability effects. Marginal effects give the marginal probabilistic change in an outcome that is attributable to a given variable; for example, for a single unit of change in one variable (a marginal change), the marginal effect coefficient gives the increase or decrease in probability of observing an outcome due to that single unit change. As a more concrete example, consider the marginal effect coefficient associated with whether the trip occurred on a weekend. The marginal coefficient for personal vehicle travel (-0.0299) for business trips means that the probability of taking a personal vehicle decreases by almost three percent when the trip includes a weekend versus when it does not. Marginal effects are calculated conditional on all other model coefficients at the sample averages, which often make them more useful in predictive analyses than odds ratios, another type of transformation of maximum likelihood model results that does not condition on other model coefficients. The transformed model coefficients in their marginal effects form along with their standard errors (in parentheses) are shown below in Table 3-5.

Table 3-5. Marginal Effects (Standard Errors) Estimates for Full Set of Multinomial Logit Models of Travel Mode Choice.

	Business				Pleasure				Personal Business			
	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train
Post 9/11 (d)	0.0191	-0.0152	-0.0006	-0.0034	0.0007	-0.0066	0.0056	0.0003	0.0002	-0.0046	0.0044	
	(0.0129)	(0.0124)	(0.0011)	(0.0044)	(0.0071)	(0.0059)	(0.0040)	(0.0013)	(0.0090)	(0.0072)	(0.0057)	
Trip occurred on weekend (d)	-0.0299+	0.0222+	-0.0005	0.0082	-0.0222**	0.0214**	0.0007		-0.0080	0.0109	-0.0029	
	(0.0161)	(0.0120)	(0.0007)	(0.0100)	(0.0048)	(0.0042)	(0.0023)		(0.0135)	(0.0127)	(0.0038)	
Nights away on trip	0.0001	0.0023	-0.0002	-0.0023	0.0019*	-0.0005	-0.0015*		0.0009	-0.0001	-0.0007	
	(0.0031)	(0.0023)	(0.0004)	(0.0023)	(0.0008)	(0.0004)	(0.0007)		(0.0008)	(0.0002)	(0.0008)	
Number of people on trip	-0.0051*	0.0049*	0.0003	-0.0001	-0.0010	-0.0005	0.0013**	0.0002	-0.0008	-0.0008	0.0016	
	(0.0021)	(0.0019)	(0.0003)	(0.0006)	(0.0010)	(0.0009)	(0.0002)	(0.0004)	(0.0021)	(0.0016)	(0.0015)	
Respondent's age	-0.0000	-0.0001	0.0001		0.0002	-0.0002	-0.0001		0.0003	-0.0001	-0.0002	
	(0.0004)	(0.0004)	(0.0001)		(0.0001)	(0.0001)	(0.0001)		(0.0002)	(0.0001)	(0.0002)	
Trip distance	-0.0002**	0.0002**			-0.0001**	0.0001**	0.0000*		-0.0001			
	(0.0000)	(0.0000)			(0.0000)	(0.0000)	(0.0000)		(0.0001)			
Count of vehicles in HH	0.0181**	-0.0182**	0.0001		0.0071**	-0.0048*	-0.0013	-0.0009	0.0014		-0.0015	
	(0.0042)	(0.0041)	(0.0001)		(0.0026)	(0.0021)	(0.0010)	(0.0016)	(0.0020)		(0.0019)	
Urban HH (d)	-0.0199*	0.0207*	0.0003	-0.0011	-0.0144*	0.0170**	-0.0030	0.0003	0.0016	0.0039	-0.0054	
	(0.0097)	(0.0093)	(0.0010)	(0.0023)	(0.0064)	(0.0051)	(0.0031)	(0.0014)	(0.0095)	(0.0059)	(0.0073)	
Population per sq mile												
Count of all bus depots in 25M radius	0.0010	-0.0005	-0.0001	-0.0004	0.0005	-0.0003	-0.0004	0.0002	0.0021	-0.0005	-0.0016	
	(0.0028)	(0.0027)	(0.0002)	(0.0005)	(0.0015)	(0.0011)	(0.0005)	(0.0003)	(0.0020)	(0.0012)	(0.0018)	
Count of all airports in 25M radius	-0.0135**	0.0139**	-0.0001	-0.0003	-0.0014	0.0031	-0.0014	-0.0003	-0.0042	-0.0013	0.0054	
	(0.0050)	(0.0046)	(0.0005)	(0.0008)	(0.0030)	(0.0021)	(0.0011)	(0.0008)	(0.0058)	(0.0027)	(0.0055)	
Count of all Amtrak stations in 25M radius	0.0012	-0.0012	-0.0001	0.0001	-0.0006	0.0006	0.0001	-0.0001	-0.0002	-0.0004	0.0006	
	(0.0014)	(0.0014)	(0.0002)	(0.0002)	(0.0007)	(0.0006)	(0.0005)	(0.0002)	(0.0012)	(0.0008)	(0.0009)	
Count of all transit/subway/light/ commuter rail stations in 25M radius	0.0001	-0.0001										
	(0.0001)	(0.0001)										
CPI Private Transport, seasonally adjusted	0.0021	-0.0014	-0.0002	-0.0004	0.0011	-0.0017*	0.0006		-0.0005	-0.0001	0.0006	
	(0.0019)	(0.0017)	(0.0003)	(0.0005)	(0.0010)	(0.0008)	(0.0005)		(0.0011)	(0.0007)	(0.0009)	
CPI Public Transport, seasonally adjusted	0.0007	-0.0007	-0.0001	0.0001	-0.0014	0.0021	-0.0008		-0.0009	0.0001	0.0007	
	(0.0027)	(0.0026)	(0.0003)	(0.0004)	(0.0016)	(0.0015)	(0.0005)		(0.0016)	(0.0009)	(0.0013)	

Table 3-5. Marginal Effects (Standard Errors) Estimates for Full Set of Multinomial Logit Models of Travel Mode Choice. (Continued)

	Business				Pleasure				Personal Business			
	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train
RITA airline ticket price index	-0.0006 (0.0014)	0.0005 (0.0013)	0.0001 (0.0002)	0.0001 (0.0001)	-0.0002 (0.0008)		0.0001 (0.0003)	0.0001 (0.0002)	-0.0002 (0.0008)	-0.0002 (0.0006)	0.0003 (0.0006)	
Weekly use of public/commercial transportation (d)	-0.0795* (0.0389)	0.0386* (0.0150)	0.0061 (0.0074)	0.0348 (0.0380)	-0.0321** (0.0093)	0.0116* (0.0056)	0.0167** (0.0050)	0.0038 (0.0054)	-0.0071 (0.0111)	0.0026 (0.0057)	0.0046 (0.0082)	
Weekly web use (d)	-0.0451** (0.0100)	0.0453** (0.0098)	-0.0003 (0.0012)	0.0002 (0.0017)	-0.0061 (0.0044)	0.0037 (0.0035)	0.0023 (0.0018)	0.0001 (0.0008)	-0.0057 (0.0055)	0.0033 (0.0047)	0.0024 (0.0029)	
$0<=Income<= $30,000 (d)	0.0460** (0.0068)	-0.0469** (0.0059)	-0.0008 (0.0012)	0.0017 (0.0040)	0.0107 (0.0068)	-0.0188** (0.0038)	0.0075 (0.0048)	0.0006 (0.0021)	0.0013 (0.0149)	-0.0079 (0.0098)	0.0065 (0.0115)	
$30,000<Income<= $60,000 (d)	0.0743** (0.0104)	-0.0750** (0.0101)	0.0009 (0.0024)	-0.0003 (0.0014)	0.0133* (0.0066)	-0.0168** (0.0051)	0.0036 (0.0031)		0.0043 (0.0113)	-0.0075 (0.0095)	0.0032 (0.0072)	
$60,000<Income<= $100,000 (d)	0.0280** (0.0087)	-0.0277** (0.0077)	0.0007 (0.0021)	-0.0010 (0.0015)	0.0091 (0.0061)	-0.0131** (0.0039)	0.0037 (0.0036)	0.0003 (0.0015)	-0.0010 (0.0111)	-0.0063 (0.0076)	0.0073 (0.0089)	
$100,000<Income (d)	Omitted – Reference Category											
African-American (d)	0.0118 (0.0165)	-0.0107 (0.0155)	-0.0006 (0.0010)	-0.0006 (0.0020)	-0.0129 (0.0189)	-0.0041 (0.0086)	0.0179+ (0.0106)	-0.0009 (0.0179)	0.0062 (0.0100)	-0.0068 (0.0086)	0.0006 (0.0041)	
Asian (d)	-0.1285 (0.1340)	0.1315 (0.1329)	-0.0013 (0.0012)	-0.0017 (0.0144)	0.0014 (0.0179)	0.0018 (0.0163)	-0.0018 (0.0044)	-0.0014 (0.0057)	0.0048 (0.0240)	-0.0124 (0.0146)	0.0075 (0.0184)	
Hispanic (d)	-0.0125 (0.0304)	0.0204 (0.0303)	-0.0022 (0.0020)	-0.0057 (0.0058)	0.0102 (0.0163)	-0.0061 (0.0133)	-0.0014 (0.0074)	-0.0027 (0.0042)	0.0161 (0.0136)	-0.0092 (0.0134)	-0.0069* (0.0033)	
Other (d)	-0.0314 (0.1928)	0.0276 (0.0605)	-0.0007 (0.0064)	0.0045 (0.2053)	-0.0073 (0.0165)	0.0071 (0.0123)	-0.0040 (0.0037)	0.0042 (0.0086)	0.0089 (0.0587)	-0.0067 (0.0591)	-0.0023 (0.0079)	
White (d)	Omitted – Reference Category											
Respondent is employed (d)	Omitted – Business trips were assumed to occur only for survey respondents who were employed				0.0024 (0.0045)	0.0042 (0.0035)	-0.0061* (0.0027)	-0.0005 (0.0011)	-0.0012 (0.0057)	0.0029 (0.0044)	-0.0017 (0.0033)	

Note: All marginal effects coefficient estimates not listed in the table or otherwise denoted were estimated at values < 0.0001 and had no statistical significance

+ Indicates statistical significance at the 10% level

* Indicates statistical significance at the 5% level

** Indicates statistical significance at the 1% level

(d) Dichotomous variable

Note that some coefficient estimates that displayed statistical significance are no longer significant at any level once transformed into marginal effects. This can be due to a variety of reasons, but generally indicates that an overall relationship and its direction can be observed in the data but the exact effect may be more uncertain due to variability. Note also that personal vehicle outcome marginal effects are included in Table 3-5; although the raw coefficient estimates for multinomial logit models must be calculated relative to a base outcome, marginal effects can be estimated for each individual outcome – this is another reason for using these estimates in analyses where predictive conclusions are necessary. Some marginal effects were extremely small in magnitude (less than 0.0001, or a 0.01 percent change) and were not statistically significant. These were excluded from Table 3-5 because their predictive effect is minimal and the practical interpretations of their marginal effects are not useful for any further analysis. As in Table 3-4, one category for each of the income and race factors was used as the reference category and thus no marginal effects were estimated. Marginal effects significant at the 1, 5, and 10 percent level of significance are noted with a '**', '*', and '+', respectively.

The model results display some consistent patterns in both coefficient and marginal effects estimates. First, there are a much higher number of statistically significant relationships observed across trip purpose types for personal vehicle and air travel outcomes. This is not entirely unexpected given the earlier discussion of the much lower number of observations for bus and train travel outcomes. Second, characteristics of the survey respondents who were taking the trips tended to be more significant predictors of travel mode choice than the characteristics of the trips themselves. This indicates that people's travel mode choices may be driven largely by fixed attributes that revolve around residence and demographics rather than consideration of the dynamic costs and benefits of different modes of travel. The marginal effects also suggest that respondents' demand for different modes of travel is relatively decoupled from cost considerations such as the price of airfares or gasoline and that the preference set may be fairly inelastic in the short term – that is, not responsive to changes in price. This is difficult to state emphatically because the exact cost of each travel option for each trip is not known but the evidence leads to this conclusion based on the economic variables used in the model. Small marginal effects, which often include the value zero within the range of one standard deviation, for price indices indicate that respondents tended to be fixed in their travel mode preferences conditional on the fixed residence and demographic attributes.

Marginal effects for variables describing trip characteristics other than distance tended to have mixed effects for different travel mode outcomes. There was little evidence that the 9/11 terrorist attacks had a noticeable effect on travel mode choices, as no marginal effects were significant. This is especially noteworthy for air travel modes as it shows that perceptions of terrorism safety may not be major drivers of respondents' choices. A weekend trip had a statistically significant marginal effect for personal vehicle and air travel for the two largest travel purpose types. There was a two to three percent decrease in the probability of taking a personal vehicle and a two percent increase in the probability of taking air travel if the trip included a weekend for business and pleasure travel. The number of persons on the trip also significantly impacted likelihoods of different mode choices; for business travel it corresponded to a 0.5 percent decrease in the chances of taking personal vehicle per person and a 0.5 percent increase in the chances of taking air travel while for pleasure travel it increased chances of taking bus travel by 0.13 percent per person. This is likely due to vehicle use efficiency reasons for business travel and the appeal of bus sightseeing tours for pleasure travel. Lastly, for pleasure travel, the number of nights away increased the probability of taking personal vehicles by 0.19 percent a night and decreased the probability of taking bus travel by 0.15 percent a night. The route travel distance was highly significant for both business and pleasure travel, and will be discussed separately.

Variables describing characteristics about respondents' place of residence also displayed mixed results. Classification of a residence as a "rural" or "urban" area was a significant predictor for personal vehicle and air mode choices, and corresponded to approximately a two percent increased chance of taking air travel for business and a 1.97 percent increased chance of taking air travel for pleasure with corresponding decreases in probability for taking personal vehicles for urban areas as compared to rural areas. Conditional on urban or rural classification, population density did not appear to have any significant effect on travel mode choice. Available transportation infrastructure only appeared to be influential for business travel; the number of airports in a 25 mile radius increased the chances of taking air travel by 1.39 percent per airport. The accessibility or airports within driving or public transit distance seems to be a primary driver of choosing this mode for work travel, but does not appear to matter for other types of travel. This could again be related to time and efficiency constraints involved with business travel that are not present for other types. Other existing transportation infrastructure did not appear to play a significant role in travel choice, but this could also be a product of large numbers of observations in the data set that chose personal vehicle as the primary mode of transport and thus do not display any preferences towards certain types of existing networks.

Respondent's demographic and behavioral variables were the most consistently significant predictors of travel choice for business and pleasure travel. Familiarity with public/commercial transportation systems through frequent usage resulted in a large decrease in the likelihood of taking personal vehicles for business travel (eight percent) as well as a smaller but still significant decrease in the likelihood of taking personal vehicles for pleasure travel (three percent). Interestingly, high public/commercial transportation use was highly statistically significant for predicting increases in the use of air travel (four percent for business, 1.2 percent for pleasure). This seems to indicate that a major factor in using air travel revolves around comfort with using the public transit system as an intermediate mode to get to or from an airport. For business travel, frequent web use also increased chances of taking air travel by about 4.5 percent; this result, as well as a corresponding decrease in chances of taking personal vehicles, was statistically significant at the 1 percent level. Past studies have cited familiarity with using online travel reservations as a potential predictor of demand for air travel, and this seems to be borne out by the model results (Civil Aviation Authority, 2005; Morrison et al, 2001). Income was also a strong predictor of travel mode choice for both business and pleasure travel. Relative to the reference category of income greater than $100,000 per year, the three lower income brackets were more likely to take personal vehicles and less likely to take air travel. The lower likelihood of air travel as income decreases shows the stronger statistical significance trend, and this reinforces the hypothesis that fixed attributes like income are much stronger determinants of travel mode. The marginal effects show that a household income that is unable to support the higher cost of air travel appears to display preferences towards personal vehicles based solely on income and not the price of airline tickets. It is possible that the price threshold for air travel faced by respondents during the survey period is high enough that consumers did not display any price sensitivity, but airline prices were relatively low during this period and displayed a reasonable range of variability during the period after 9/11. Overall, income and behavioral variables seemed to display the highest statistical significance in model results.

One of the most consistently significant variables in the model was route distance of a trip from origin to destination (measured in miles). This result reflects expected respondent preferences for transportation mode choice that can be observed in the larger U.S. population of travelers – longer trips place a higher inconvenience burden on personal vehicle travel and make other modes of travel, particularly air travel, more desirable. This is due to the physical, time, and financial burdens of traveling in a personal vehicle over increasingly large distances, and there is an expected “break even” point at which the desirability of personal vehicle travel begins to be outweighed by the convenience of other modes. The marginal effects coefficients listed in Table 3-5 give the marginal changes in probability of choosing each mode per additional mile traveled. On a per mile basis this is not a practical result to use in analysis of travel behavior since there might be very small overall marginal probability changes observed for short distance trips. In order to better observe the overall relationship between route distance and travel mode choice, Figures 3-9 through 3-11 display the trend in predicted probabilities the model outputs for travel mode choice for respondent observations at different route distances. The overall trend is estimated directly from the NHTS dataset using a nonparametric polynomial smoothing function to produce probability distributions that approximate the continuous change in predicted probabilities of mode choice over the range of trip route distances. This smoothing function gives a more concise picture of the significant relationships present in the data than a standard scatter plot graph.

Note that the graphs display probabilities of taking a certain mode of travel on the vertical axis, and thus the “break even” point for this representation shows a route distance at which the predicted probability of taking a private vehicle is approximately equal to the predicted probability of taking air travel. The predicted probabilities for each mode choice at a given trip distance shown in the figures represents a smoothed average across all the trips in the NHTS file at that distance. For example, the probability of taking private vehicle or air travel is each about 50 percent at about 700 miles for business travel (a small percentage of travelers choose train or bus). For some NHTS trips around 700 miles, the probability of taking a private vehicle for the given trip is greater than that of air travel (e.g., 70 percent for private vehicle, 20 percent for air, and 10 percent for bus/train) while for other trips around that distance the probability of taking air travel for the given trip is greater than that of a private vehicle (30 percent for private vehicle, 60 percent for air, and 10 percent for bus/train). The differences in predicted probabilities for trips are a result of the values for other predictors in the model. The data values used in Figures 3-9 through 3-11 show the smoothed mean predicted probabilities at a given distance.

The coast-to-coast driving distance in the main body of the U.S. is around 3,000 miles, so this is used as the upper bound of the figures (note that there are some outlier observations with route distances higher than 3,000 miles).

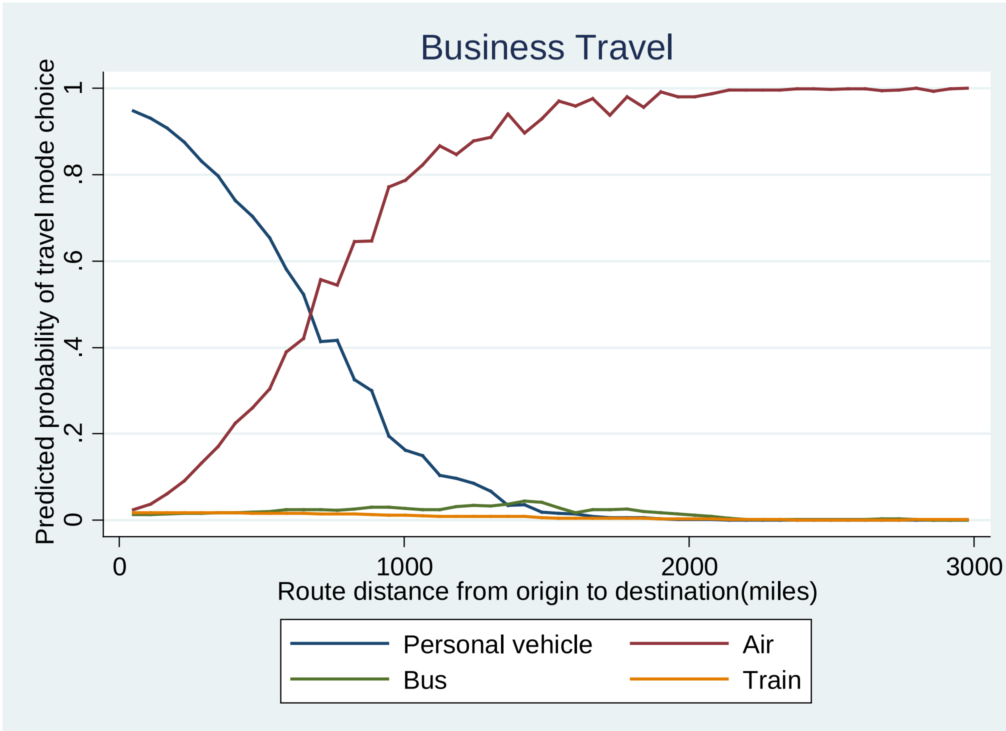

Figure 3-9. Fitted Polynomial Trend of Route Distance vs. Predicted Travel Mode Choice – Business Travel.

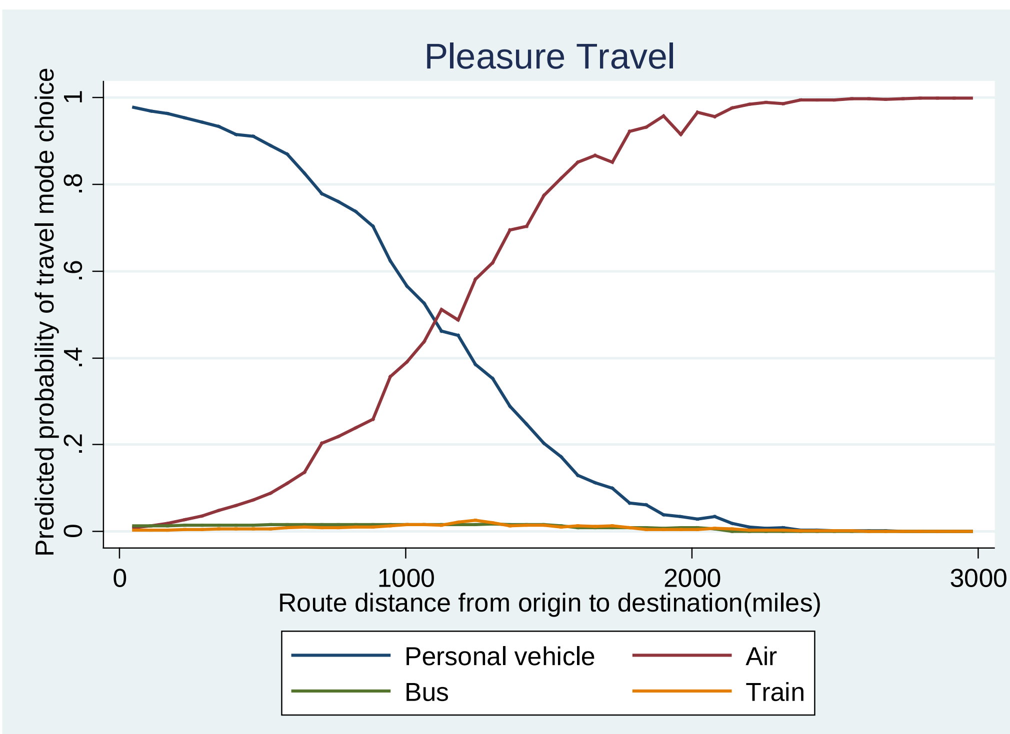

Figure 3-10. Fitted Polynomial Trend of Route Distance vs. Predicted Travel Mode Choice – Pleasure Travel.

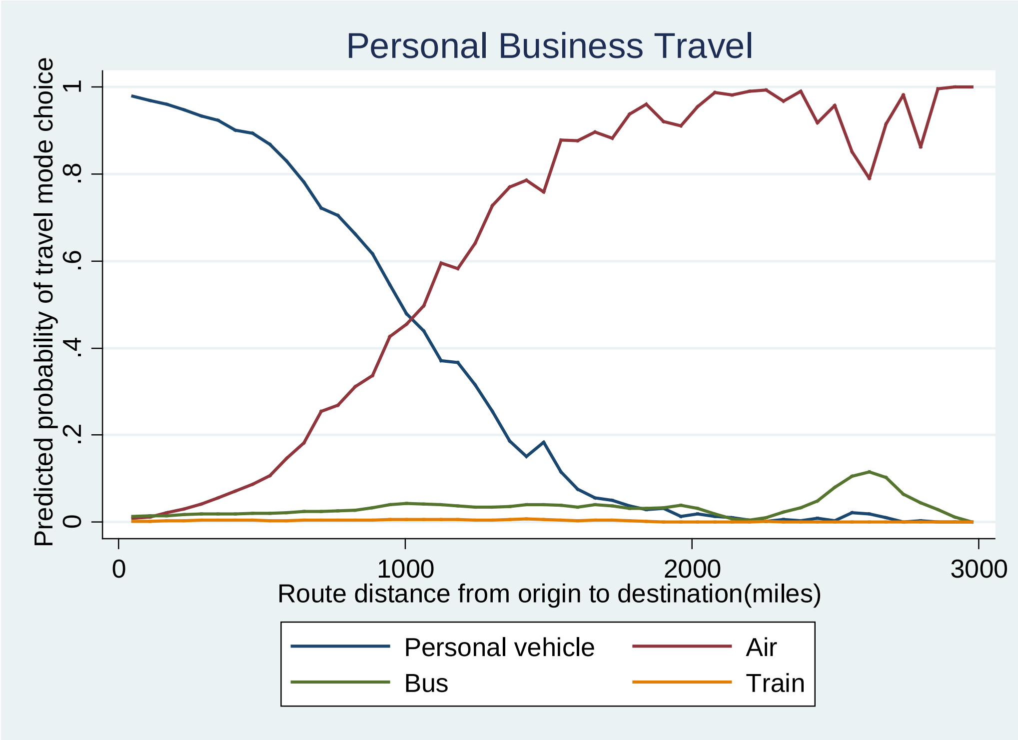

Figure 3-11. Fitted Polynomial Trend of Route Distance vs. Predicted Travel Mode Choice – Personal Business Travel.

The above figures show several clear trends. First, the probability of choosing to travel in a personal vehicle decreases exponentially with travel distance. Second, the probability of choosing air travel increases exponentially with travel distance. Last, there is a limited range of "break even" points across travel purpose types where the probability of taking air travel begins to exceed the probability of taking a personal vehicle. For business travel, this point occurs around 700 miles and is the lowest of all three travel modes. This is consistent with the need for efficient, short travel periods to conduct business with a lower upper distance tolerance for personal vehicle use that reaches time and convenience constraints more quickly. Pleasure trips have a much higher tolerance for personal vehicle use, with a "break even" point around 1,100 miles where air travel becomes more likely. This fits with more relaxed constraints surrounding pleasure travel, where the desirability of personal vehicles can remain higher over longer distances as part of sightseeing road trips. Personal business travel displays a trend that is between the other two types of travel, and also shows some signs of limited data issues due to the large swings in predicted probabilities at high route distances. Overall, bus and train travel modes do not display high predicted probabilities, and other than a few small increases at higher route distances do not display any significant trends.

The graph for personal business travel displays a noticeable increase in predicted probabilities of taking bus travel at route distances around 2,700 miles. This is due to a group of personal business travel observations in the dataset who all appear to have taken a group trip by bus at this distance, giving more weight to predicted probabilities of this mode and resulting in a corresponding decrease in the probability of taking air travel (the overwhelmingly predominant predicted mode at surrounding trip distances). While this increase is visible in the overall shape of the probability distribution for bus travel, note that the highest level of predicted probability is only about a 10 percent likelihood of taking bus travel which is still relatively small compared to the 80 percent likelihood of taking air travel at this distance. Thus this likely an artifact of the group of observations in the NHTS data at this distance as opposed to a predictive trend.

3.6.2 Reduced Prediction Models

Coefficient estimates and their standard errors for the reduced multinomial logit models of travel mode choice are presented in Table 3-6, with one set of coefficient results for each travel purpose type. Separate model estimates are presented for each travel mode. Note that there are no coefficient estimates for the personal vehicle mode as that mode was the reference level. Thus, the logits for all other modes are constructed relative to it. Also for the categorical variable income that has more than two levels, the greater than $100,000 category was used as the reference category and thus no coefficients were estimated. Estimates for all other levels were made relative to the greater than $100,000 category. Coefficient estimates significant at the 1, 5, and 10 percent level of significance are noted with a '**', '*', and '+', respectively.

The reduced model coefficients indicate that many of the relationships observed in the fully specified model are preserved for the smaller subset of variables. Income categorical variables remained statistically significant as predictors of increased use of air travel at the 1 percent and 5 percent levels. Route distance also remained a significant determinant for the choice of using air travel relative to private vehicles. In contrast to the fully specified model, the number of people on the trip was a significant predictor for taking a bus, with larger numbers of people indicating increased probabilities of bus use across trip purposes at the 1 percent level.

Marginal effects were calculated for the reduced form model in the same way as the fully specified model, and are presented in Table 3-7. Some marginal effects were extremely small in magnitude (less than 0.0001, or a 0.01 percent change) and were not statistically significant. These were excluded from Table 3-7 because their predictive effect is minimal and the practical interpretations of their marginal effects are not useful for any further analysis. As in Table 3-6, one category for the income factor was used as the reference category and thus no marginal effects were estimated. Marginal effects significant at the 1, 5, and 10 percent level of significance are noted with a '**', '*', and '+', respectively.

Many marginal effects retained similar significance and magnitude levels to the fully specified marginal effects with several exceptions. Despite the fact that the marginal effect for number of persons remained a significant factor for increased probabilities of bus usage for long distance trips, the marginal effect per additional trip person was relatively small meaning that a sufficiently large group of travelers would be needed to cause a noticeable shift in predictive probability. Income categorical variables all had increased marginal effect magnitudes in the reduced form model in addition to retaining their predictive significance. Relative to household incomes of greater than $100,000 per year, incomes of $60,000 to $100,000, $30,000 to $60,000, and less than $30,000 per year had 3.56 percent, 9.46 percent, and 5.91 percent lower chances of taking air travel for business travel, respectively. These decreased probabilities corresponded with similarly significant increases in the likelihood of taking a private vehicle relative to incomes over $100,000 per year. The same patterns for income are observed for pleasure and personal business trips although the marginal effects are less. Interestingly, for business travel only there was a statistically significant increase in the probability of taking private vehicles after 9/11.

Table 3-6. Coefficient Estimates (Standard Errors) for Reduced Set of Multinomial Logit Models of Travel Mode Choice.

	Business				Pleasure				Personal Business			
	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train
Post 9/11 (d)		-0.3531 (0.2541)	-0.5725 (0.9464)	-1.0829+ (0.5920)		-0.1737 (0.2224)	0.7769 (0.5260)	0.2369 (0.6768)		-0.5377 (0.5250)	0.5793 (0.4608)	-2.6685+ (1.3447)
Number of people on trip		0.1110** (0.0361)	0.2605** (0.0607)	-0.1512 (0.2728)		-0.0310 (0.0372)	0.1829** (0.0180)	0.1088 (0.1628)		-0.0705 (0.1386)	0.2669** (0.0492)	0.0982 (0.1006)
Respondent's age		-0.0072 (0.0095)	0.0657* (0.0330)	0.0105 (0.0164)		-0.0099* (0.0043)	-0.0148 (0.0118)	-0.0054 (0.0165)		-0.0104 (0.0104)	-0.0318+ (0.0189)	-0.0147 (0.0287)
Trip distance		0.0058** (0.0006)	0.0034+ (0.0019)	0.0023 (0.0018)		0.0040** (0.0002)	0.0004 (0.0003)	0.0021** (0.0005)		0.0045** (0.0005)	0.0018* (0.0008)	0.0016 (0.0019)
Count of vehicles in HH		-0.4384** (0.0956)	-0.0979 (0.1592)	-0.2043 (0.2367)		-0.2318** (0.0788)	-0.2373 (0.1482)	-0.7529+ (0.3842)		-0.1222 (0.1342)	-0.3283 (0.2358)	-0.3332 (0.8535)
Urban HH (d)		0.6408* (0.2691)	0.2312 (1.0199)	-0.2282 (0.5998)		0.7808** (0.2579)	-0.3105 (0.3549)	0.0353 (0.9082)		0.6458 (0.4284)	-0.8227 (0.6469)	-0.3068 (1.5105)
Population per sq mile		-0.0000 (0.0000)	0.0001 (0.0000)	0.0001 (0.0001)		-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)		0.0000 (0.0000)	-0.0000 (0.0000)	0.0001 (0.0001)
Count of all bus depots in 25M radius		0.0007 (0.0650)	0.0079 (0.2565)	-0.0728 (0.1280)		-0.0139 (0.0431)	-0.0176 (0.0723)	-0.0120 (0.1224)		-0.1146 (0.1274)	-0.2510+ (0.1422)	0.2232 (0.3528)
Count of all airports in 25M radius		0.1905+ (0.1043)	-0.2012 (0.5653)	0.0318 (0.2121)		0.1616* (0.0772)	-0.1873 (0.1572)	0.2414 (0.2138)		-0.0377 (0.2053)	0.8451** (0.1712)	0.4085 (0.4573)
Count of all Amtrak stations in 25M radius		-0.0165 (0.0315)	-0.0436 (0.1364)	0.0393 (0.0639)		0.0308 (0.0212)	0.0282 (0.0543)	0.0197 (0.0757)		-0.0331 (0.0887)	0.0797 (0.1178)	-0.1282 (0.2121)
CPI Private Transport, seasonally adjusted		-0.0359 (0.0370)	-0.2365 (0.1753)	-0.1489 (0.1321)		-0.0576+ (0.0337)	0.0707 (0.0620)	-0.0052 (0.1025)		-0.0184 (0.0739)	0.0734 (0.0963)	0.0004 (0.3137)
CPI Public Transport, seasonally adjusted		-0.0274 (0.0564)	-0.1239 (0.2246)	0.0592 (0.1408)		0.0726 (0.0566)	-0.1279+ (0.0713)	0.0021 (0.1488)		0.0269 (0.1141)	0.0997 (0.1681)	-0.7717+ (0.4309)
RITA airline ticket price index		0.0150 (0.0331)	0.0338 (0.1630)	0.0113 (0.0702)		0.0084 (0.0267)	0.0206 (0.0468)	0.0593 (0.0780)		-0.0248 (0.0594)	0.0397 (0.0636)	-0.2597 (0.1709)
$0<=Income<= $30,000 (d)		-2.2477** (0.5391)	-1.3403 (2.1476)	-0.0840 (0.7611)		-1.1929** (0.2741)	0.9018* (0.3481)	0.0421 (0.7595)		-1.5568* (0.6870)	0.8899 (0.7500)	-0.7114 (22.5901)
$30,000<Income<= $60,000 (d)		-2.4231** (0.3107)	0.1595 (1.6071)	-0.6127 (0.6879)		-0.8548** (0.2303)	0.5094 (0.3653)	-0.2554 (0.7445)		-1.3163** (0.4475)	0.6325 (0.8326)	-1.3072 (2.2160)
$60,000<Income<= $100,000 (d)		-0.7265** (0.2001)	0.5673 (1.6429)	-0.7161 (0.5725)		-0.6306** (0.1855)	0.4996 (0.4151)	0.1149 (0.8629)		-1.0345* (0.4317)	1.0823 (0.6641)	0.1292 (2.1829)
$100,000<Income (d)	Omitted – Reference Category											
Respondent is employed (d)		Omitted – Business trips were assumed to occur only for survey respondents who were employed				0.3009+ (0.1556)	-0.7529** (0.2293)	-0.2560 (0.4560)		0.5761 (0.3616)	-0.1892 (0.3994)	-0.3480 (0.8966)
Constant		7.0166 (13.1780)	48.6000 (50.6280)	5.4695 (24.6072)		-11.6122 (13.0819)	9.9690 (22.2701)	-10.9894 (39.9300)		-3.6399 (27.4661)	-41.0353 (29.0254)	184.8157+ (108.0689)

Note: Multinomial logit model coefficients were estimated relative to the base reference outcome of private travel

+ Indicates statistical significance at the 10% level * Indicates statistical significance at the 5% level ** Indicates statistical significance at the 1% level

(d) Dichotomous variable

Table 3-7. Marginal Effects (Standard Errors) Estimates for Reduced Set of Multinomial Logit Models of Travel Mode Choice.

	Business				Pleasure				Personal Business			
	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train	Private Vehicle	Air	Bus	Train
Post 9/11 (d)	0.0431* (0.0211)	-0.0176 (0.0143)	-0.0025 (0.0047)	-0.0229 (0.0177)	-0.0023 (0.0077)	-0.0051 (0.0063)	0.0066 (0.0045)	0.0008 (0.0021)	0.0070 (0.0150)	-0.0094 (0.0094)	0.0046 (0.0040)	-0.0022 (0.0106)
Number of people on trip	-0.0041 (0.0057)	0.0060** (0.0020)	0.0012 (0.0008)	-0.0031 (0.0054)		-0.0009 (0.0010)	0.0016** (0.0002)	0.0004 (0.0005)	-0.0009 (0.0022)	-0.0012 (0.0022)	0.0021** (0.0007)	0.0001 (0.0003)
Respondent's age	-0.0001 (0.0007)	-0.0004 (0.0005)	0.0003 (0.0002)	0.0002 (0.0003)	0.0004* (0.0002)	-0.0003* (0.0001)	-0.0001 (0.0001)		0.0004+ (0.0002)	-0.0002 (0.0002)	-0.0002* (0.0001)	
Trip distance	-0.0004** (0.0001)	0.0003** (0.0000)			-0.0001** (0.0000)	0.0001** (0.0000)		0.0000+ (0.0000)	-0.0001** (0.0000)	0.0001** (0.0000)		
Count of vehicles in HH	0.0269** (0.0070)	-0.0231** (0.0050)	-0.0003 (0.0008)	-0.0035 (0.0042)	0.0107** (0.0027)	-0.0063** (0.0022)	-0.0019 (0.0013)	-0.0024* (0.0011)	0.0048 (0.0031)	-0.0020 (0.0023)	-0.0025 (0.0019)	-0.0002 (0.0010)
Urban HH (d)	-0.0260 (0.0191)	0.0305** (0.0112)	0.0009 (0.0046)	-0.0054 (0.0129)	-0.0153* (0.0073)	0.0183** (0.0052)	-0.0031 (0.0035)	0.0001 (0.0029)	-0.0018 (0.0097)	0.0099 (0.0063)	-0.0079 (0.0079)	-0.0002 (0.0015)
Population per sq mile												
Count of all bus depots in 25M radius	0.0013 (0.0045)	0.0001 (0.0034)		-0.0014 (0.0024)	0.0006 (0.0016)	-0.0004 (0.0012)	-0.0001 (0.0006)	-0.0000 (0.0004)	0.0037 (0.0026)	-0.0019 (0.0021)	-0.0019 (0.0014)	0.0001 (0.0008)
Count of all airports in 25M radius	-0.0096 (0.0088)	0.0102+ (0.0051)	-0.0010 (0.0030)	0.0004 (0.0041)	-0.0037 (0.0027)	0.0045* (0.0021)	-0.0016 (0.0014)	0.0008 (0.0006)	-0.0060 (0.0041)	-0.0008 (0.0035)	0.0065** (0.0022)	0.0002 (0.0011)
Count of all Amtrak stations in 25M radius	0.0003 (0.0022)	-0.0009 (0.0017)	-0.0002 (0.0007)	0.0008 (0.0012)	-0.0011 (0.0008)	0.0009 (0.0006)	0.0002 (0.0005)	0.0001 (0.0003)		-0.0006 (0.0015)	0.0006 (0.0010)	-0.0001 (0.0004)
CPI Private Transport, seasonally adjusted	0.0056 (0.0035)	-0.0017 (0.0019)	-0.0011 (0.0009)	-0.0028 (0.0026)	0.0010 (0.0012)	-0.0016+ (0.0009)	0.0006 (0.0005)		-0.0002 (0.0015)	-0.0003 (0.0013)	0.0006 (0.0009)	
CPI Public Transport, seasonally adjusted	0.0009 (0.0042)	-0.0015 (0.0030)	-0.0006 (0.0013)	0.0012 (0.0027)	-0.0010 (0.0017)	0.0021 (0.0016)	-0.0011+ (0.0007)		-0.0007 (0.0033)	0.0005 (0.0020)	0.0008 (0.0012)	-0.0005 (0.0025)
RITA airline ticket price index	-0.0011 (0.0024)	0.0008 (0.0017)	0.0002 (0.0007)	0.0002 (0.0013)	-0.0006 (0.0009)	0.0002 (0.0007)	0.0002 (0.0004)	0.0002 (0.0002)	0.0003 (0.0014)	-0.0004 (0.0010)	0.0003 (0.0005)	-0.0002 (0.0009)
$0<=Income<= $30,000 (d)	0.0631** (0.0155)	-0.0591** (0.0067)	-0.0037 (0.0037)	-0.0003 (0.0145)	0.0117 (0.0081)	-0.0232** (0.0035)	0.0114+ (0.0059)	0.0002 (0.0025)	0.0093 (0.0144)	-0.0185** (0.0062)	0.0095 (0.0103)	-0.0004 (0.0078)
$30,000<Income<= $60,000 (d)	0.1021** (0.0175)	-0.0946** (0.0115)	0.0013 (0.0081)	-0.0088 (0.0100)	0.0175* (0.0078)	-0.0217** (0.0052)	0.0050 (0.0036)	-0.0007 (0.0025)	0.0143 (0.0119)	-0.0193** (0.0070)	0.0057 (0.0083)	-0.0007 (0.0035)
$60,000<Income<= $100,000 (d)	0.0448* (0.0171)	-0.0356** (0.0096)	0.0031 (0.0097)	-0.0123 (0.0096)	0.0107 (0.0072)	-0.0160** (0.0042)	0.0049 (0.0043)	0.0004 (0.0028)	0.0044 (0.0103)	-0.0154* (0.0062)	0.0109 (0.0079)	0.0001 (0.0014)
$100,000<Income (d)	Omitted – Reference Category											
Respondent is employed (d)	Omitted – Business trips were assumed to occur only for survey respondents who were employed				0.0010 (0.0051)	0.0080* (0.0038)	-0.0081* (0.0032)	-0.0009 (0.0019)	-0.0072 (0.0066)	0.0090 (0.0055)	-0.0016 (0.0035)	-0.0002 (0.0016)

Note: All marginal effects coefficient estimates not listed in the table or otherwise denoted were estimated at values < 0.0001 and had no statistical significance

+ Indicates statistical significance at the 10% level
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level

(d) Dichotomous variable

3.6.3 Model Limitations

One assumption of the multinomial logit model is that the model error terms are independent and identically distributed. As a result, when the multinomial logit model is used to model choices, it relies on the assumption of independence of irrelevant alternatives (IIA) which is not always desirable. Ben-Akiva and Lerman (1985) give the definition as "the ratio of the chosen probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives." They continue on to show that IIA can produce imprecise estimates when a new mode with similar characteristics is introduced into the mode choice set. As such, more complicated models such as the nested logit model or mixed logit model are sometimes used as an extension of the multinomial logit model to capture the correlation of alternatives when alternatives are not independent. Despite this shortcoming, this research utilizes the multinomial logit model. This was done primarily because of the limitations imposed by the statistical software SAS. As mentioned previously, the NHTS utilizes a complicated sampling design that involves a large amount of clustering (i.e., multiple members of a household are surveyed regarding their long-distance trips). To ensure that the effect of this clustering, as well as other survey issues such as nonresponse, unequal selection probabilities, and stratification are taken into account when calculating variances for model estimates, the SURVEYLOGISTIC procedure was used. One limitation to this procedure is that it is not designed to accommodate nested or mixed logit models. SAS can handle such models but only with other procedures that are in turn not equipped to deal with complicated survey design data. Given the amount of clustering, the research team believed it more important to account for the survey design in the analysis rather than focus on a more complicated model that might relax the IIA. FHWA requested the models be developed in SAS. There may be alternative software packages that could fit more complicated models while accounting for the complicated survey design. However, this was not explored because the resources available to this research did not allow for further investigation and FHWA preferred the use of SAS for model development in this task order. This is an area for further research.

3.7 Discussion

This report presents a detailed discussion of the mathematical models and inputs to the models used to estimate mode choice for long-distance passenger travel. The report examines the effects that the traveler (in terms of their socioeconomic, demographic, and behavioral attributes), the trip (in terms of distance, purpose, length, and traveling party size), the availability of transportation infrastructure, and land-use characteristics has on the selection of travel mode for long-distance travel as measured by a generalized multinomial logit model. Major findings from this research are as follows:

- Summary statistic and model results provide evidence that mode choice varies by trip purpose and that separate models are warranted;
- There were a much greater number of factors found to significantly influence mode choice observed across trip purpose types for personal vehicle and air travel outcomes than bus and train outcomes. This is due, in part, to the low frequency of bus and train trips in the NHTS;
- Characteristics of the survey respondents who were taking the trips tended to be more significant predictors of travel mode choice than the characteristics of the trips

themselves. This indicates that people's travel mode choices may be driven largely by fixed attributes that revolve around residence and demographics rather than consideration of the dynamic costs and benefits of different modes of travel;

- The results suggest that respondents' demand for different modes of travel may be relatively decoupled from cost considerations such as the price of airfares or gasoline and that the preference set may be fairly inelastic in the short run – that is, not responsive to changes in price;
- Available transportation infrastructure only appeared to be influential for business travel;
- Respondent's demographic and behavioral variables were the most consistently significant predictors of travel choice for business and pleasure travel;
- One of the most consistently significant variables in predicting mode choice was route distance of a trip from origin to destination. The probability of choosing to travel in a personal vehicle decreases exponentially with travel distance while the probability of choosing air travel increases exponentially with travel distance; and
- The model predicts very well for the personal vehicle and air modes but loses some predictive power for the bus and train modes. The relative lack of predictive power for bus and train modes indicate that the survey data may not be sufficient to accurately assess some outcomes and that alternative sampling techniques should be explored in future national travel surveys that provide more data for bus and train trips.

A more thorough assessment of the model's strengths and predictive power is presented in the following section. It will also describe some of the model limitations and suggestions for further research that could overcome these limitations.

4.0 VALIDATION OF MATHEMATICAL MODELS

Validation of the mode choice models presented in the previous section is provided here through the discussion of the validation methodology (Section 4.1) and the results (Section 4.2). The reduced models will be used in a more practical sense to predict future mode choice within a transportation modeling framework as it contains only readily available input variables identified as having an influence on mode choice. Thus, all validation procedures were conducted on the reduced set of models to assess their predictive ability.

4.1 Methodology

Validation of the long-distance passenger travel modal choice models was conducted by testing the models on long-distance travel survey data. The same 2001 NHTS dataset used for model calibration was also used for model validation. One common method for validating transportation models is holdout validation where the dataset of long-distance trips is divided into two non-overlapping parts; one solely used to develop and calibrate the models and another for validating the models. This approach is used to determine if over-fitting of the model is present and provides accurate estimates for the predictive performance of the models. The downside to this approach is that it does not use all the available data. Given the limited number of long-distance trips in the 2001 NHTS and the fact that the trips were segregated by trip purpose (i.e., business, personal business, and pleasure) in order to account for the differences in mode choice by trip purpose, the holdout method was not preferable. In addition, results from holdout validation are highly dependent on the choice for the calibration/validation split. This has the potential to lead to skewed results in terms of poor prediction performance if data in the validation set that may be valuable for calibration is held out in the validation set (Refaeilzadeh, 2009). To deal with these challenges and to utilize all the NHTS data, the mode choice models were validated with a technique called *k*-fold cross-validation.

Cross-validation is a statistical technique for assessing how the results of the statistical model will generalize to an independent dataset. Holdout validation, described above, is the most basic form of cross-validation. In *k*-fold cross validation, the data is first partitioned into *k* equally (or nearly equally) sized segments, or folds. Then, *k* iterations of calibration and validation are performed such that a different fold of the data is held out for validation while the remaining *k*-1 folds are used to calibrate the model within each iteration. The value of *k* is usually dependent on the size of the dataset. Small values of *k* (e.g., 2 or 3) lead to calibration datasets that are not as close to the full dataset size which is not desirable while extremely large values of *k* increase the overlap between calibration datasets across iterations and lead to small validation datasets which could result in less-precise predictions. General consensus in the data mining and model fitting community is that $k = 10$ is a common choice that balances these factors (Refaeilzadeh, 2009).

For this research, 10-fold cross-validation was conducted separately to validate each of the three multinomial mode choice models (one for each trip purpose). For example, the number of business trips (10,008) was randomly divided into ten segments of approximately 1,001 long-distance trips. In the first iteration, one segment was withheld as the validation dataset while a multinomial logit model was fit to the other nine segments. Then, the fitted model was applied to the validation dataset (i.e., predicted probabilities for each mode of transportation were calculated for each trip in the validation dataset). Using the coefficients of the predictor variables, the model predicts the probability that the given traveler for each trip would choose each of the four mode choices. For example, on a given trip where private vehicle was the actual mode of choice, the probability of taking a private vehicle, air, bus, and train from the model might be 70 percent, 20 percent, 3 percent, and 2 percent, respectively. Aggregate mode shares were calculated by summing the calculated probabilities for each trip record in the validation dataset. These were compared against the observed aggregate mode shares of the validation dataset in order to observe how well the model could replicate the observed mode shares. This process was repeated nine times, each time choosing a different segment of the data to be held out as the validation dataset. Once all iterations were complete, the comparison of predicted versus observed aggregate mode shares were combined across the ten iterations and statistics summarizing the predictive ability of the model were calculated.

4.2 Results

For each of the ten iterations in the 10-fold cross validation, the aggregate mode shares across all trips in the validation dataset were calculated for each mode by summing the calculated probabilities for each trip record. These were compared against the aggregate mode shares of the dataset in order to observe how well the model could replicate the observed mode shares. The results of this comparison across the ten iterations are shown in Table 4-1 for each trip purpose.

Table 4-1. Comparison of Actual and Model-Predicted Aggregate Mode Shares by Trip Purpose.

Trip Purpose	Actual Mode	Unweighted Number of Trips	Model Predicted Mode			
			Personal Vehicle	Air	Bus	Train
Business	Personal Vehicle	8,445	93	5	1	2
	Air	1,244	31	66	1	2
	Bus	105	70	15	14	1
	Train	214	90	6	1	2
Pleasure	Personal Vehicle	13,438	95	4	1	0
	Air	1,202	39	59	1	1
	Bus	203	68	2	29	1
	Train	62	79	14	6	1
Personal Business	Personal Vehicle	3,245	96	3	1	0
	Air	186	48	49	3	1
	Bus	116	36	2	62	0
	Train	14	80	16	3	1

Notes: Shaded Cells represent percentage of trips where actual and model-predicted modes agree. Due to rounding, some row percentages do not add exactly to 100%.

For business travel, both the personal vehicle and air modes show predicted probabilities that indicate the models are highly predictive (93 percent agreement for personal vehicles and 66 percent agreement for air), as evidenced by the relatively low of number of “wrong” predictions. This reinforces the effects observed in the marginal and raw coefficient estimates for the business travel model presented and discussed in Section 3.6, where personal vehicle and air travel display several clear trends that are highly statistically significant. However, bus and train travel do not show the same high level of predictive ability (14 percent for bus and 2 percent for train).

The results for pleasure and personal business are similar to that of business trips. Note that these two models predict the likelihood of actually using a personal vehicle very well (95 percent for pleasure trips and 96 percent for personal business trips). Air travel is correctly predicted 59 percent of the time for pleasure trips and 49 percent for personal business trips. A larger percentage of bus trips are predicted correctly (29 percent for pleasure trips and 62 percent for personal business trips). However, train travel is predicted poorly for all three trip types.

Figure 4-1 shows the distribution of the predicted aggregated mode shares relative to the observed aggregated mode shares by iteration for each mode of transportation for business trips. Figure 4-2 and Figure 4-3 show the same information for pleasure and personal business trips, respectively. These graphs are useful for assessing the variability in results across the iterations of the cross-validation. From the figures, the following observations can be made concerning the proportion of predicted mode shares relative to the observed mode shares across iterations:

- Results are consistently high for personal vehicle travel across all three trip purposes;
- Results for air travel are consistent for business and pleasure travel but are more variable for personal business travel due in most part to the smaller number of personal business trips;
- The predictive ability for bus travel varies depending on the iteration for all three trip purposes due mainly to the small number of bus trips in the NHTS; and
- Train travel is consistently poorly predicted across iterations for all trip purposes.

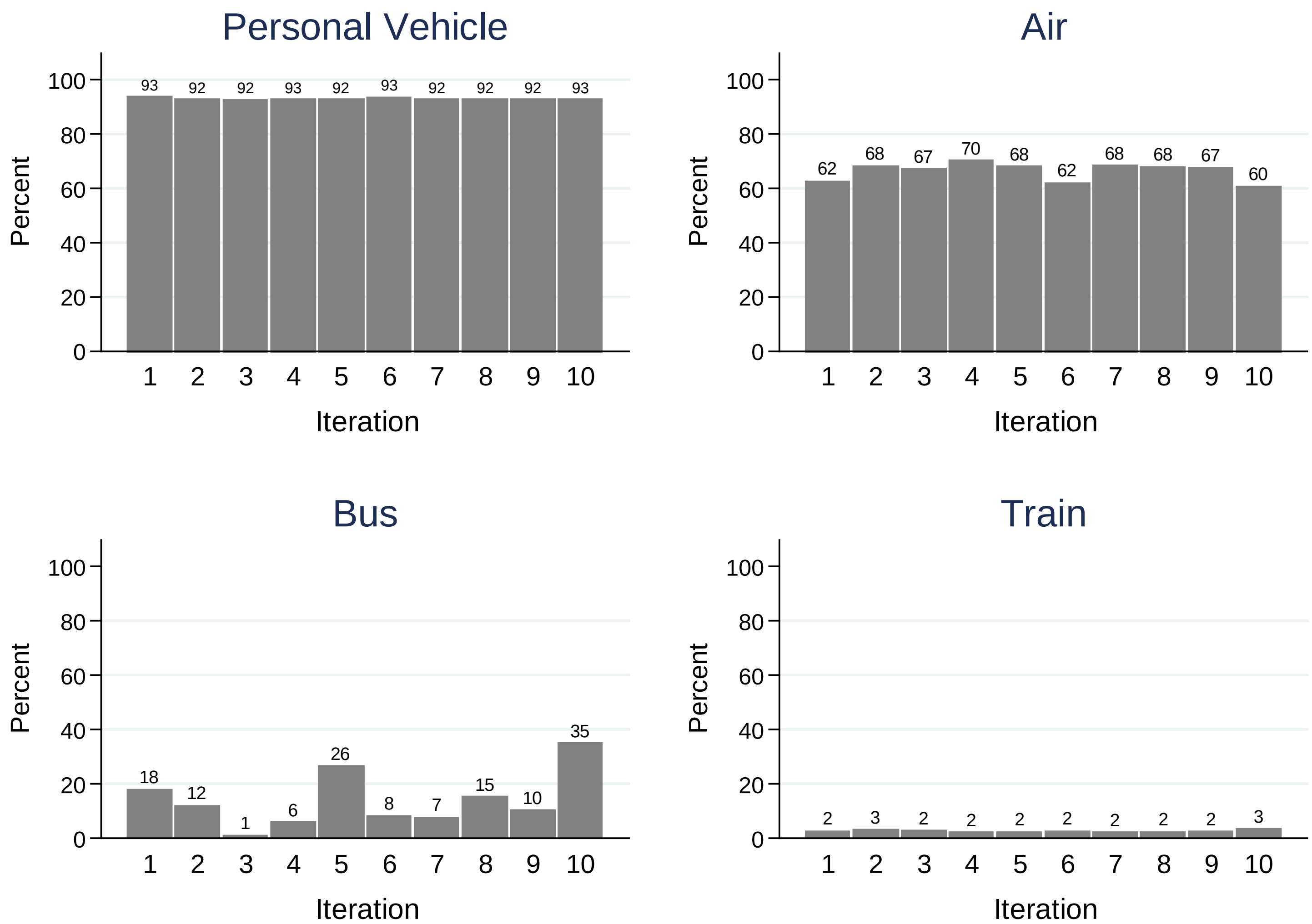

Figure 4-1. Distribution of Predicted Aggregated Mode Shares Relative to Observed Aggregated Mode Shares Across Iterations by Mode Choice (Business Trips).

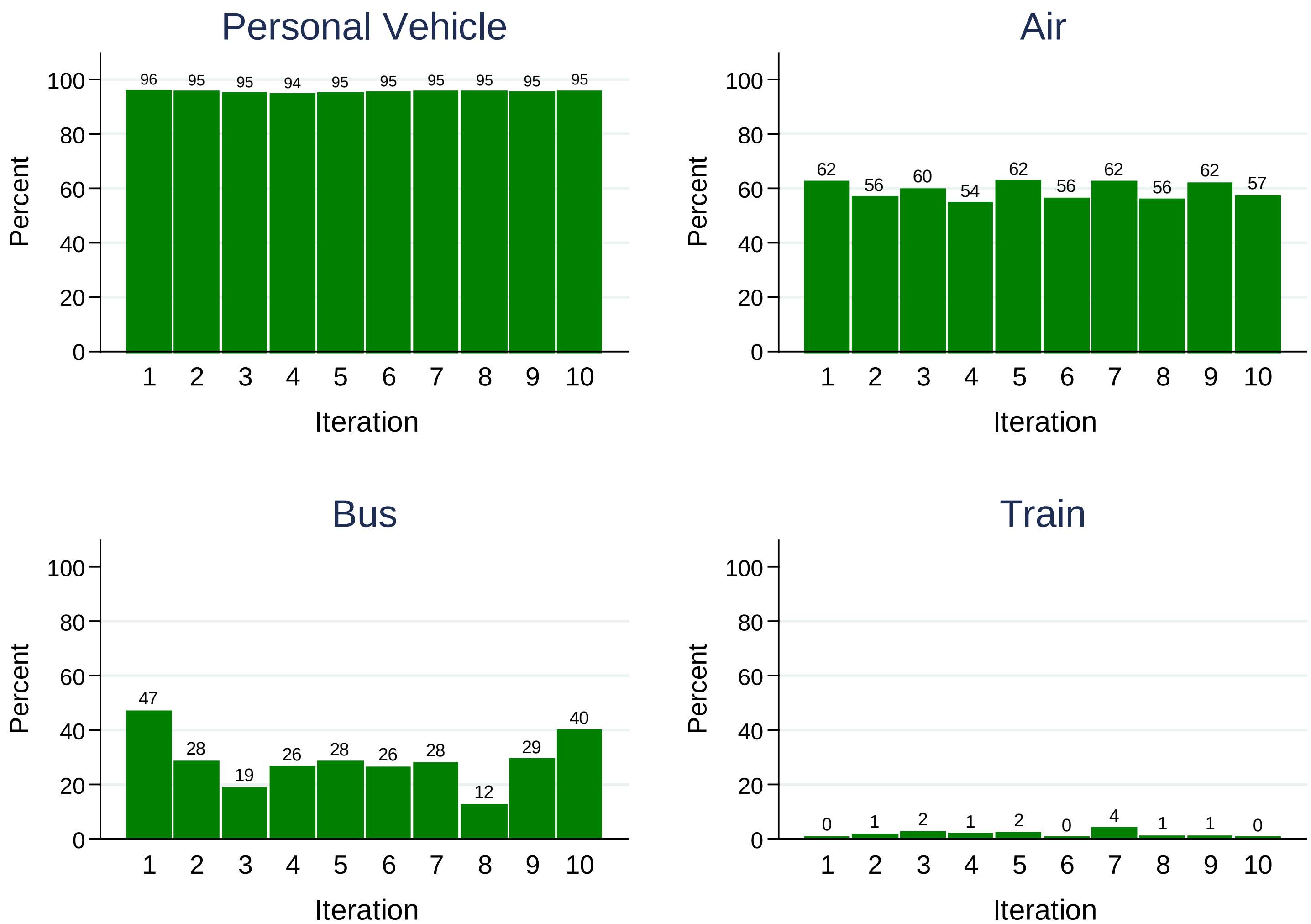

Figure 4-2. Distribution of Predicted Aggregated Mode Shares Relative to Observed Aggregated Mode Shares Across Iterations by Mode Choice (Pleasure Trips).

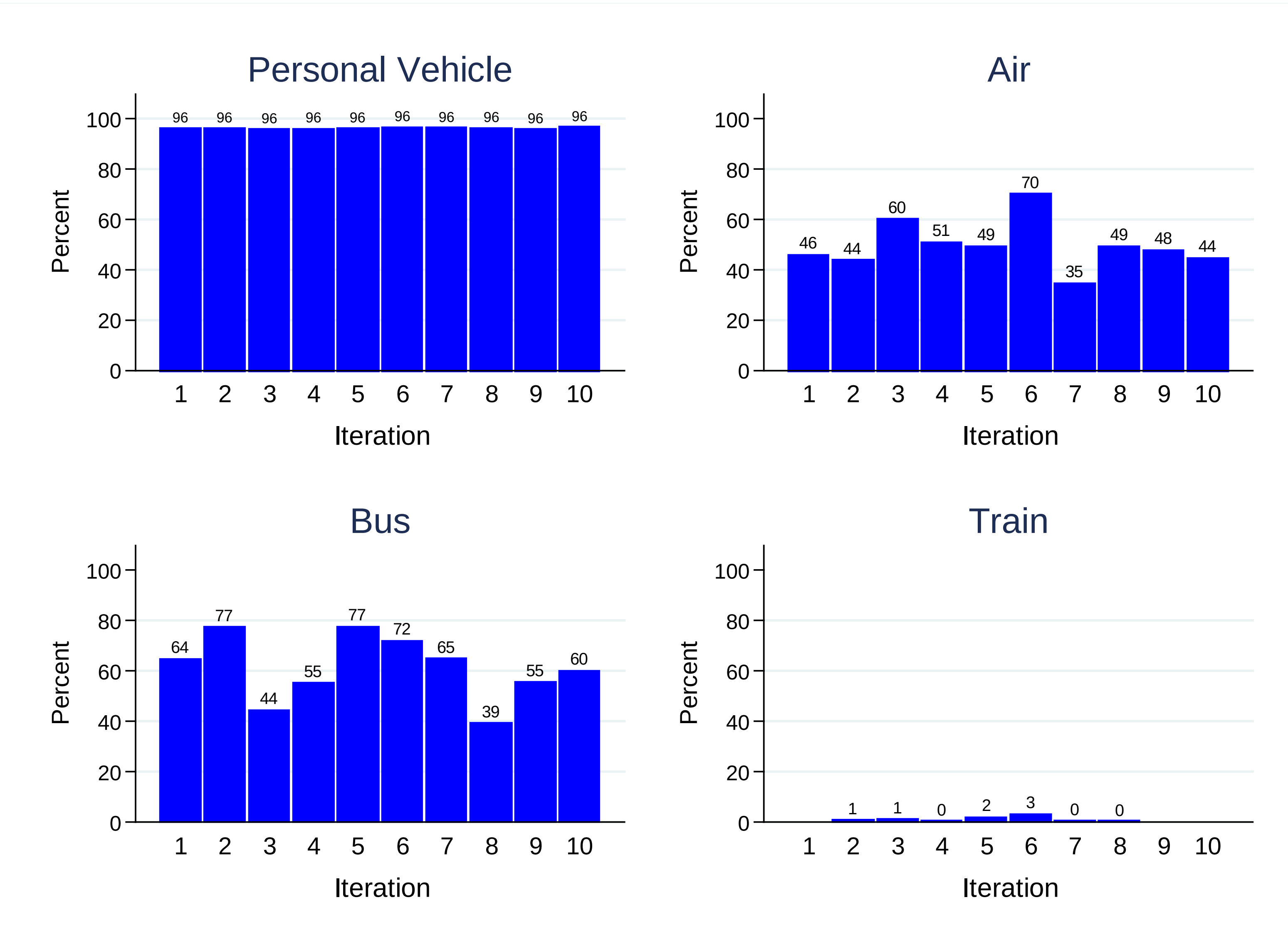

Figure 4-3. Distribution of Predicted Aggregated Mode Shares Relative to Observed Aggregated Mode Shares Across Iterations by Mode Choice (Personal Business Trips).

The results of the cross-validation show that the models predict personal vehicle and air travel well. The high predictive power coupled with the consistency of results across iterations implies that the logistic regression models are not over-fit. However, the models don't fare as well for bus and train passenger travel. When bus and train were actually used, the model most frequently predicted personal vehicle. This is most likely the product of low respondent observation counts; the models do not have the same data granularity to generate predictions as with personal vehicle and air travel. Even after accounting for survey weights, bus and train travel comprise such a small proportion of overall observations that any survey and sample bias are significant risks. The low number of observations makes it very difficult to determine those factors that influence a long-distance traveler's decision of taking personal vehicle versus taking the bus or train. For business travel in particular, the types of persons that choose to take bus or train modes are likely to be highly variable which compounds this issue.

In order to determine which factors influence long-distance passenger travels to choose bus or train travel, more data will be needed. This will be challenge given that the long-distance trip frequency in general is low for the majority of U.S. households. The 2001 NHTS long-distance sample shows that about one-half of all surveyed households did not take long-distance trips (defined by distances of 50 miles or more) during their assigned four-week travel period. The difficulty in capturing these long-distance trips is only going to become greater as the U.S. and the rest of the world are experiencing shifts in travel behavior due to the rise of the internet, economic crisis, terrorism, and other factors. The bottom line is that although it is difficult to get survey data using traditional sampling designs such as those used for the ATS and NHTS, it is even more difficult to get data on bus and train travel. In the 2001 NHTS, only three percent of all long-distance trips were taken by bus and train. The same holds true for the 1995 ATS where about two percent of all long-distance trips were taken by bus and less than one percent by train.

In order to gather more data on bus and train long-distance travel, the research team believes that the sample design and data collection techniques for future national household long-distance transportation surveys needs to be modified to address these concerns. The following are some possible improvements to the design that would result in more long-distance passenger travel data. These are presented as ideas that would warrant more research to determine their feasibility and value to long-distance travel surveys. Each of them has inherent advantages and disadvantages that would need to be explored. These potential improvements are a few of the long-distance travel survey items that are currently being investigated by the research team as part of a separate project with FHWA focusing on designing a completely new approach for a long-distance travel survey instrument.

- **Abandon household sample frames**: Instead of relying solely on household sample frames to form the sample, one idea would be to sample trips in process. Travelers could be intercepted at train or bus stations as well as airports. In addition, a sample could be made of all ticket purchases. This would be a form of area-based sampling with facilities or locations serving as the area of interest rather than households.
- **Use of Multi-frame Sampling Designs:** Dual-frame or multi-frame sampling designs for surveys primarily seek to prevent noncoverage bias. As the U.S. population becomes increasingly mobile and the emergence of positioning- and event-related technology advances, new technology-based frames are becoming more available such as those based upon Facebook, Twitter, Four Square, etc. These technologies tend to focus more on

individuals rather than households. In these situations, the use of a multi-frame design may become more appealing and would be worth additional investigation.

- **Collect data more frequently**: Event driven data collection such as pulse surveys could be used to capture more long-distance trips as they are occurring. Future surveys, especially those attempting to characterize rare events, have the novel capability of being designed around "event-driven" data collection. This concept involves detecting and collecting information on household travel events in a passive manner, by accessing data sources that are made available to the survey upon receiving explicit informed consent of the survey participants to do so. Example data sources include cell phone tracking information and social media postings (Facebook, Twitter, etc.). For example, if a survey participant's cell phone (or the cell phone of an individual within a selected household) is noted to have moved a distance that exceeds a given threshold, this finding would indicate that a long-distance travel event occurred. This concept is known as "geo-fencing." This concept could leverage technology to trigger a survey once a participant travels a certain number of miles. In addition to capturing data on long-distance trips, this method would effectively shorten the recall period for the survey participant which could help to reduce recall bias.
- **Use of social media to connect to participants**: Social media data could be mined to identify past, current, and future long-distance trips. It could also be used for self-reporting of trip events and/or as a data collection mechanism.

5.0 CONCLUSIONS

This report presents the research undertaken to develop a long-distance multimodal passenger travel modal choice model. This research started with a literature and practice review that served as a precursor to the development of quantitative mathematical methods to analyze how long-distance passenger travelers make their modal choices. Findings from this review helped identify mathematical techniques and models that have been used on mode choice modeling over the last several years. In addition the review assisted with identifying data sources used for long-distance modeling and factors that were found to influence long-distance passenger travel mode choice.

The report presents a detailed discussion of the mathematical models and inputs to the models used to estimate mode choice for long-distance passenger travel. The report examines the effects that the traveler (in terms of their socioeconomic, demographic, and behavioral attributes), the trip (in terms of distance, purpose, length, and traveling party size), the availability of transportation infrastructure, and land-use characteristics has on the selection of travel mode for long-distance travel as measured by a generalized multinomial logit model. Major findings from this research are as follows:

- Summary statistic and model results provide evidence that mode choice varies by trip purpose and that separate models are warranted;
- There were a much greater number of factors found to significantly influence mode choice observed across trip purpose types for personal vehicle and air travel outcomes than bus and train outcomes. This is due, in part, to the low frequency of bus and train trips in the NHTS;
- Characteristics of the survey respondents who were taking the trips tended to be more significant predictors of travel mode choice than the characteristics of the trips themselves. This indicates that people’s travel mode choices may be driven largely by fixed attributes that revolve around residence and demographics rather than consideration of the dynamic costs and benefits of different modes of travel;
- The results suggest that respondents’ demand for different modes of travel may be relatively decoupled from cost considerations such as the price of airfares or gasoline and that the preference set may be fairly inelastic in the short run – that is, not responsive to changes in price;
- Available transportation infrastructure only appeared to be influential for business travel;
- Respondent’s demographic and behavioral variables were the most consistently significant predictors of travel choice for business and pleasure travel;
- One of the most consistently significant variables in predicting mode choice was route distance of a trip from origin to destination. The probability of choosing to travel in a personal vehicle decreases exponentially with travel distance while the probability of choosing air travel increases exponentially with travel distance; and
- The multinomial logit models developed to predict long-distance passenger travel mode choice predict personal vehicle and air travel well. This is encouraging given that 97 percent of all long distance travel was conducted via these two methods according to

the 2001 NHTS. However, the models do not predict bus and train passenger travel very well. When bus and train were actually used, the models most frequently predicted personal vehicle. This is largely due to the fact that bus and train travel comprise such a small proportion of overall observations even after accounting for survey weights.

More data will be needed to effectively predict bus and train long-distance passenger travel. Traditional national long-distance travel surveys have not been able to capture this data. Thus, modifications to the sample design and data collection techniques for future national household long-distance transportation surveys would be warranted to address these concerns. Concepts such as transitioning from a household-based frame to frame that focuses on locations where long-distance travelers are located (e.g., airports, train and bus stations), using social media to connect to long-distance travelers, and leveraging technology such as GPS to help identify when long-distance trips occur are possible ways to increase the amount of long-distance data and to better target bus and train travel modes. As part of this data collecting effort, the research team would recommend a targeted study to identify measureable factors that could differentiate inclination for long-distance passenger travelers to use bus and train relative to private vehicle.

6.0 REFERENCES

Algers, S. "Integrated Structure of Long-Distance Travel Behavior Models in Sweden." *Transportation Research Record.* 1993. <http://pubsindex.trb.org/view.aspx?id=385102>.

Ashiabor, S., Baik, H., and Trani, A., "Logit Models for Forecasting Nationwide Travel Demand in the United States", Transportation Research Record, Journal of the Transportation Research Board, No. 2007, Transportation Research Board of the National Academies, Washington, D.C. pp 1-12.

Baik, Hojong, et al. "Forecasting Model for Air Taxi, Commercial Airline, and Automobile Demand in the United States." *Journal of the Transportation Research Board.* 7 Oct. 2008. <http://trb.metapress.com/content/v36420106814q868/>.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge, Massachusetts.

Bhat, Chandra R., "An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel." *Transportation Science.* 1997. <http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Endogenous_Segment.pdf>.

Bhat, Chandra R., "A Heteroscedastic Extreme Value Model of Intercity Mode Choice." *Transportation Research Part B: Methodological.*1995. <http://www.ce.utexas.edu/prof/bhat/ABSTRACTS/CITYNEW2.pdf>.

Bricka, Stacey. "Variations in Long-Distance Travel." *TRB Transportation Research Circular E-C026—Personal Travel: The Long and Short of It.* 1999. <http://onlinepubs.trb.org/Onlinepubs/circulars/ec026/06_bricka.pdf>.

Burge, Peter, et al. *Modelling Demand for Long-Distance Travel in Great Britain: Stated preference surveys to support the modelling of demand for high-speed rail*. Santa Monica, CA: RAND Corporation, 2011. <http://www.rand.org/pubs/technical_reports/TR899>.

Civil Aviation Authority. "Demand for Outbound Leisure Air Travel and its Key Drivers". 2005.

Cohen, Michael, et al. "NHTS 2001 Highlights Report." *U.S. Department of Transportation, Bureau of Transportation Statistics.* 2003. <http://www.bts.gov/publications/highlights_of_the_2001_national_household_travel_survey/pdf/entire.pdf>.

Costa, Bernardo. "Michigan 2009 Visitor Profile". D.K. Shifflet and Associates. Ltd. McLean, Virginia. 2010.

Georggi, Nevine Labib, and Ram M. Pendyala. "Analysis of Long-Distance Travel Behavior of the Elderly and Low Income." *Transportation Research Circular E-C026—Personal Travel: The Long and Short of It.* 1999. <http://onlinepubs.trb.org/Onlinepubs/circulars/ec026/02_georggi.pdf >.

Grayson, A. Disaggregate model of mode choice in intercity travel, In *Transportation Research Record: Journal of the Transportation Research Board, No. 385*, 1982, pp 36-42.

IHS Global Insight, American Automobile Association. “AAA Independence Day 2011 Travel Forecast”. June 2011. <http://aaanewsroom.net/Assets/Files/Independence_Day_2011_Report.pdf>

Khan, Omer A., et al. “Developing passenger mode choice models for Brisbane to reflect observed travel behavior from the South East Queensland Travel Survey.” *30th Australasian Transport Research Forum.* 2007. <http://eprints.qut.edu.au/8879/>.

Kitagawa, Toyohiro, et al. “Various Factors Affecting Modal Choice Behavior of the Inter-City Passenger Between Keihanshin and Fukuoka.” *Proceedings of the Eastern Asia Society for Transportation Studies.* 2005. <http://www.easts.info/on-line/proceedings_05/199.pdf>.

Koppelman, Frank S., and Chieh-Hua Wen. “The paired combinatorial logit model: properties, estimation, and application.” *Transportation Research Part B 34.* 2000. <http://directory.umm.ac.id/Data%20Elmu/jurnal/T/The%20Journal%20of%20Strategic%20Information%20Systems/Vol8.Issue1.Mar1999/209.pdf>.

Koppelman, Frank S., and Vaneet Sethib. “Incorporating variance and covariance heterogeneity in the Generalized Nested Logit model: an application to modeling long distance travel choice behavior.” November 2005. <http://www.sciencedirect.com/science/article/pii/S0191261504001596>.

Koppelman, F. S. Multidimensional model system for intercity travel choice behavior, In *Transportation Research Record: Journal of the Transportation Research Board, No. 1241*, 1990, pp 1-8.

LaMondia, Jeffre, et al. “Traveler Behavior and Values Analysis in the Context of Vacation Destination and Travel Mode Choices: A European Union Case Study.” *The University of Texas at Arlington.* 2010. <www.ce.utexas.edu/prof/bhat/ABSTRACTS/Vacation_destination_15March2010.doc>.

Last, Jorg, and Wilko Manz. “Unselected mode alternatives: What drives modal choice in long-distance passenger transport?” *10th International Conference on Travel Behaviour Research, Lucerne*. August 2003. <http://www.ivt.ethz.ch/news/archive/20030810_IATBR/last.pdf>.

Lave, C. and R. Crepeau. “Travel by Households without Vehicles.” *1990 NPTS Travel Mode Special Reports. FHWA (FHWA-PL-94-019)*. December 1994. <http://nhts.ornl.gov/1990/doc/mode.pdf>.

Limtanakool, Narisra, et al. “The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium- and longer-distance trips.” *Journal of Transport Geography 14.* 2006. <http://igitur-archive.library.uu.nl/socgeoplan/2008-1210-201337/Limtanakool%2c%20Dijst%2c%20Schwanen%202006%20(JoTrGeo).pdf>.

“Long Distance Transportation Patterns: Mode Choice.” *America on the Go.* Bureau of Transportation Statistics. May 2006. <http://www.bts.gov/publications/america_on_the_go/long_distance_transportation_patterns/pdf/entire.pdf>.

Lundqvist, Lars, and Lars-Göran Mattsson. *National transport models: recent developments and prospects.* Stockholm: 2002. Print.

Mallett, William J. "Long-Distance Travel by Low-Income Households." *TRB Transportation Research Circular E-C026—Personal Travel: The Long and Short of It.* 1999a. <http://onlinepubs.trb.org/Onlinepubs/circulars/ec026/04_mallet.pdf>.

Mallett, William J. "Long-Distance Travel by Women: Results from the 1995 American Travel Survey", presented at the Annual Meeting of the Transportation Research Board, Washington D.C., January 1999b.

Monzon, Andres and Alvaro Rodriguez-Dapena. "Choice of mode of transport for long-distance trips: Solving the problem of sparse data." *Transportation Research Part A: Policy and Practice.* August 2006. <http://www.sciencedirect.com/science/article/pii/S0965856405001606>.

Morrison, et al. "Predicting Usage of the Internet for Travel Bookings: An Exploratory Study". *Information Technology and Tourism*. Vol. 4, 2001.

O'Neill, W., Brown, E., 1999. Long-distance trip generation modeling using ATS. Paper presented at the conference Personal Travel: The Long and Short of IT, Transportation Research Board, June 28-July 1, Washington D.C.

"Nationwide Personal Transportation Survey: Travel Mode Special Reports." December 1994. <http://nhts.ornl.gov/1990/doc/mode.pdf>.

"Personal Travel: The Long and Short of It." *Transportation Research Circular.* 1999. <http://onlinepubs.trb.org/Onlinepubs/circulars/ec026/ec026.pdf>.

"Personal Travel The Long and the Short of It: ISSUES INVOLVED IN ANALYSIS USING THE NPTS AND THE ATS." *Office of Highway Policy Information, Federal Highway Administration.* 1998. <http://nhts.ornl.gov/1995/Doc/npts_ats.pdf>.

Picard, Richard; Cook, Dennis (1984). "Cross-Validation of Regression Models". *Journal of the American Statistical Association* 79 (387): 575–583

Racca, David P., and Edward C. Ratledge. "Factors That Affect and/or Can Alter Mode Choice." *Delaware Transportation Institute and Delaware Department of Transportation.* July 2003. <http://www.ce.udel.edu/dct/research/publications_files/Rpt.%20159%20Factors%20that%20Affect%20and%20or%20Can%20Alter%20Mode%20Choice.pdf>.

Rasmidatta, Isaradatta. "Mode choice models for long distance travel in United States of America." *The University of Texas at Arlington.* Dec. 2006. <http://gradworks.umi.com/32/21/3221209.html>.

Refaeilzadeh, Payam, et al. "Cross Validation". *Encyclopedia of Database Systems*, Springer 2009.

Sharp, Joy, et al. "A Picture of Long-distance Travel Behavior of Americans Through Analysis of the 2001 National Household Travel Survey." *National Household Travel Survey*

Conference.2004. <http://onlinepubs.trb.org/onlinepubs/archive/conferences/nhts/Sharp.pdf >.

Stopher, P., Prashker, J. Intercity Passenger Forecasting: The Use of Current Travel Forecasting Procedures, *Transportation Research Forum: Annual Meeting Proceedings*, 1976, pp 67-75.

ULTRANS Institute of Transportation Studies, UC Davis. HBA Specto Incorporated. “Model Development, Long Distance Personal Travel Model – Final System Documentation”. 2011.

U.S. Department of Transportation, Bureau of Transportation Statistics, “NHTS 2001 Highlights Report, BTS03-05, Washington D.C., 2003.

U.S. Department of Transportation, Federal Highway Administration, “2001 NHTS User’s Guide, Washington D.C., 2004.