

ABSTRACT

Title of Dissertation: A NATIONAL TRAVEL DEMAND MODEL FOR THE U.S.: A PERSON-BASED MICROSIMULATION APPROACH

Yijing Lu, Doctor of Philosophy, 2017

Directed By: Lei Zhang, Professor
Department of Civil & Environmental Engineering

Understanding long distance travel behavior and forecasting reliable long distance travel demand are critical in evaluating intercity or regional transportation improvements and infrastructure investment projects. As the nation and various states engage in funding transportation infrastructure improvements to meet future long-distance passenger travel demand, it is imperative to develop effective and practical modeling methods for long-distance passenger travel analysis. This dissertation proposes the first integrated activity-based travel demand model system for individual's quarterly/yearly long distance or national activities and travel in the U.S at the Metropolitan Statistical Area (MSA)/Non-MSA level. The model system is developed based on a rigorous behavioral framework in long distance travel planning, and takes into account the specific attributes of the long distance travel such as low

frequency, long activity duration, different sets of mode alternatives, etc. The system includes three tiers: 1) the yearly long distance activity pattern level estimating the number of different activities a person will choose during one year; 2) the tour level which consists of tour destination choice, time of year choice, tour duration, and tour mode choice; 3) the stop level estimating the intermediate stop frequency, purpose and location. According to the different decision-making processes for different types of long distance activities (business, personal business, and pleasure), two tour-level model structures were developed, one for long distance business/personal business activities and the other for long distance pleasure activity.

Econometric model developments are conducted for the multiple model components. And estimation results are obtained based on the 1995 American Travel Survey data, transportation origin-destination (OD) skim data, and economic/demographic data. With-out-sample validation is performed for each model component and system-wide model calibration is conducted using optimization method prior to model implementation and future year policy analysis. The model system is implemented in our developed micro-simulation platform which simulates each individual's yearly long distance activities and travel in the U.S with the input of the population data, the associated transportation OD skim data and economic/demographic data. The travel demand in the year of 2040 is forecasted and two more scenarios including national-level fuel price increase and high speed rail operation are analyzed based on the calibrated long distance travel demand model.

The contributions of the dissertation lie in the following three aspects: 1). The first national travel demand model which employs a person-based microsimulation

approach is developed for the U.S. for long distance passenger travel analysis; 2). the developed person-based travel demand model enables us to conduct the travel demand analysis of high speed rail in selected inter-regional corridors in the U.S and the national-level fuel price increase; 3). a post-processing learning system which can estimate the missing information such as trip purpose for the passively collected long distance travel survey data is proposed and tested.

A NATIONAL TRAVEL DEMAND MODEL FOR THE U.S.: A PERSON-BASED MICROSIMULATION APPROACH

By

Yijing Lu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Lei Zhang, Chair
Professor William Rand
Professor Casey Dawkins
Professor Qingbin Cui
Professor Paul Schonfeld

© Copyright by
Yijing Lu
2017

Acknowledgements

It would not have been possible for me to finish this research without the help and support from so many people. It is a great pleasure to thank all the people that contributed to the preparation and the completion of this dissertation.

I am deeply indebted to my advisor, Dr. Lei Zhang, for his continuous support, inspiring advice and enthusiastic encouragement throughout the years of my Ph.D. study at the University of Maryland, College Park. I am grateful for the opportunity he gives me to work on this emerging and interesting topic. He is not only an ideal advisor in research, but also a great mentor in my pursuit of career and a good friend that cares about my personal life.

I would like to extend my thanks to the rest of my dissertation committee: Professor William Rand, Professor Casey Dawkins, Professor Qingbin Cui and Professor Paul Schonfeld, for their encouragement, insightful comments and valuable suggestions to my research.

My sincere thanks also go to Dr. Carlos Carrion, for the knowledge and insight he has shared.

I thank all my colleagues in the Transportation Systems Research Lab at the University of Maryland. The ideas and suggestions they shared with me during our group meetings have greatly contributed to the completion of this research. I would also like to thank my fellow graduate students from other groups in the transportation program. I appreciate the opportunities to take class, cooperate on projects and have discussions with them.

Finally, I would like to express my gratitude towards my family and dedicate this dissertation to them. They have been encouraging me, supporting me and loving me all the time. I am especially thankful to my husband, Bo Sun, for his understanding, love, support and help throughout the years of my Ph.D. study and my dissertation research.

Table of Contents

Acknowledgements.....	ii
Table of Contents	iv
List of Figures	vii
List of Tables	xi
List of Abbreviations	xiv
List of Mathematical Symbols.....	xvi
Chapter 1: Introduction.....	1
1.1 Background	1
1.2 Objectives and Contributions	5
Chapter 2: Literature Review.....	10
2.1 Long Distance Travel Demand Modelling.....	10
2.2 Long Distance Travel Survey Data	19
Chapter 3: Data	23
3.1 Zone System.....	23
3.2 Travel Survey Data.....	24
3.3 Transportation OD Skim and Economic/Demographic Data.....	32
3.4 Public Use Microdata Sample Data	36
Chapter 4: Model System Analysis Framework	38
4.1 Activity Pattern Level Model.....	40
4.2 Tour Level Structure	43
4.2.1 Travel Mode Choice Model	45
4.2.2 Time of Year Choice.....	49

4.2.3	Tour Duration Choice Model.....	55
4.2.4	Travel Party Size Choice Model.....	61
4.2.5	Tour Destination Choice Model.....	65
4.3	Stop Level Structure.....	69
4.3.1	Stop Frequency Choice Model.....	70
4.3.2	Stop Purpose Choice Model.....	72
4.3.3	Stop Location Choice Model	75
4.4	National Travel Demand Model Flow and Key Assumptions	81
Chapter 5:	Preliminary Base Year OD Estimations	89
Chapter 6:	Model Calibration	100
Chapter 7:	Future Year Policy Analysis	106
7.1	Future Year Policy Scenarios.....	106
7.2	Future Year Population Synthesis	112
7.3	Future Year Scenario Results Analysis	119
7.3.1	Scenario1: Base Scenario.....	120
7.3.2	Scenario 2: Fuel Price Increase.....	129
7.3.3	Scenario 3: High Speed Rail	139
Chapter 8:	Long Distance Travel Survey Instrument.....	155
8.1	Methodology for Long Distance Trip Purpose Classification	158
8.1.1	Decision Tree Learning.....	159
8.1.2	Metaleaner.....	162
8.2	Data for Long Distance Trip Purpose Classification	163
8.3	Trip Purpose Classification Results	165

Chapter 9: Conclusions and Future Research.....	174
9.1 Conclusions	174
9.2 Recommendations for Future Research	176
Bibliography	179

List of Figures

Figure 2-1: Categorization of Long Distance Travel Demand Analysis Methods	10
Figure 3-1: National Travel Demand Model Traffic Analysis Zone System	24
Figure 3-2: Percentages of Long Distance Tour Travel Modes.....	27
Figure 3-3: Percentages of Long Distance Trip Purposes	28
Figure 3-4: Travel Mode Usage by Trip Purpose	29
Figure 3-5: Trip Distribution by Time of Year.....	30
Figure 3-6: Trip Distribution by Purpose and Time of Year	30
Figure 3-7: The number of Inbound/Outbound Stop Distribution.....	31
Figure 3-8: Outbound/Inbound Stop Purpose Distribution.....	32
Figure 4-1: Long distance travel illustration.....	38
Figure 4-2: Activity-Based Long Distance Travel Demand Model System.....	40
Figure 4-3: Yearly Long Distance Activity Pattern Level.....	41
Figure 4-4: Tour Level Procedure and Model Components	45
Figure 4-5: Tour Mode Choice Validation for Business Purpose.....	48
Figure 4-6: Tour Mode Choice Validation for Pleasure Purpose	48
Figure 4-7: Tour Mode Choice Validation for Personal Business Purpose.....	49
Figure 4-8: Re-simulating Time of Year Choice Model.....	51
Figure 4-9: Time of Year Choice Validation for Business Purpose	53
Figure 4-10: Simple Time of Year Choice Validation for Pleasure Purpose	54
Figure 4-11: Full Time of Year Choice Validation for Pleasure Purpose	55
Figure 4-12: Observed Duration Distribution of Long Distance Pleasure Activities .	58

Figure 4-13: Baseline Hazard Rate for Business and Personal Business Duration Model	60
Figure 4-14: Validation results for Business Duration Model.....	61
Figure 4-15: Validation results for Personal Business Duration Model	61
Figure 4-16: Travel Party Size Choice Validation for Business Purpose	63
Figure 4-17: Travel Party Size Choice Validation for Personal Business	64
Figure 4-18: Travel Party Size Choice Validation for Pleasure Purpose.....	65
Figure 4-19: Destination Choice Validation for Business Purpose	68
Figure 4-20: Destination Choice Validation for Pleasure Purpose	68
Figure 4-21: Destination Choice Validation for Personal Business Purpose	68
Figure 4-22: Stop Level Procedure and Model Components.....	69
Figure 4-23: Inbound Stop Frequency Model Validation.....	71
Figure 4-24: Outbound Stop Frequency Model Validation	72
Figure 4-25: Outbound Stop Purpose Model Validation	74
Figure 4-26: Inbound Stop Purpose Model Validation.....	75
Figure 4-27: LOS estimation for the first stop during outbound tour leg	76
Figure 4-28: LOS estimation for the jth stop during outbound tour leg	77
Figure 4-29: Outbound Stop Location Choice Validation	79
Figure 4-30: Inbound Stop location Choice Validation	80
Figure 5-1: MSA/Non-MSA Population.....	90
Figure 5-2: Trip Distribution by Travel Mode.....	92
Figure 5-3: Trip Distribution by Travel Mode and Time of Year	93
Figure 5-4: Trip Distribution by purpose.....	93

Figure 5-5: Trips Originate/Destinate at MSA/Non-MSA level.....	95
Figure 5-6: Yearly Car Trip Distribution by TAZ originating from Washington D.C.....	96
Figure 5-7: Year Train Trip Distribution by TAZ Originating from Washington D.C	
.....	96
Figure 5-8: Yearly Air Trip Distribution by TAZ Originating from Washington D.C	
.....	97
Figure 5-9: Number of trips of different categories for different model run	98
Figure 7-1: Retail Gasoline Price Changes	107
Figure 7-2: Crude Oil Price Projections to 2040	109
Figure 7-3: Interface of PopGen	114
Figure 7-4: Comparison between Synthetic Population and Control Margins by Control Variable.....	116
Figure 7-5: MSA/Non-MSA Population in 2040.....	117
Figure 7-6: Gender distribution of synthetic 2040 population.....	118
Figure 7-7: Income group distribution of synthetic 2040 population	118
Figure 7-8: Age group distribution of synthetic 2040 population	119
Figure 7-9: Trip distribution by travel mode	121
Figure 7-10: Trip Distribution by Time of Year	122
Figure 7-11: Trip distribution by Travel Mode and Time of Year	122
Figure 7-12: Trip distribution by trip purpose	123
Figure 7-13: Trip Distribution by Income level and Travel Mode	125
Figure 7-14: Average number of trips/person during one year by income group and travel mode.....	125

Figure 7-15: Trip distribution by Gender and Travel Mode	126
Figure 7-16: Average number of trips/person during one year by gender and travel mode.....	126
Figure 7-17: Trip distribution by travel mode and age group.....	127
Figure 7-18: Average number of trips/person during one year by age group and travel mode.....	127
Figure 7-19: Trips Originate/Destinate at MSA/Non-MSA level.....	128
Figure 7-20: Comparison of trip distribution by travel mode.....	131
Figure 7-21: Comparison of trip distribution by travel mode.....	132
Figure 7-22: Comparison of trip distribution by purpose	132
Figure 7-23: Comparison of trip distribution by purpose and travel mode	133
Figure 7-24: Comparison of trip distribution by income group and travel mode	136
Figure 7-25: Comparison of trip distribution by gender and travel mode	136
Figure 7-26: Comparison of trip distribution by age group and travel mode	137
Figure 7-27: Comparison of miles by car per person during one year by income group	137
Figure 7-28: Comparison of miles/car trip by income group	138
Figure 7-29: Comparison of train trip distribution by time of year	142
Figure 7-30: Comparison of Train Trips by trip purpose	142
Figure 8-1: Trip Purpose Learning System.....	158

List of Tables

Table 3-1: Encoding Reported Trip Purposes.....	27
Table 3-2: Part of PUMA and MSA/Non-MSA Correspondence Table	37
Table 4-1: Comparison between long distance and short distance travel.....	39
Table 4-2: Trip Rate for Long Distance Business Travel.....	42
Table 4-3: Trip Rates for Long Distance Pleasure Travel	42
Table 4-4: Trip Rates for Long Distance Personal Business Travel	43
Table 4-5: Tour Mode Choice Model Estimation Results	47
Table 4-6: Full Time of Year Choice for Business Trip.....	52
Table 4-7: Simple Time of Year Choice for Pleasure Trip	53
Table 4-8: Full Time of Year Choice for Pleasure Trip.....	54
Table 4-9: Tour duration choice model estimation results	59
Table 4-10: Travel Party Size Choice Model Estimation for Business Tour	63
Table 4-11: Travel Party Size Choice Model Estimation for Personal Business Tour	64
Table 4-12: Travel Party Size Choice Model Estimation for Pleasure Tour	65
Table 4-13: Primary Destination Choice Model Estimation at Tour Level.....	67
Table 4-14: Stop frequency model estimation for tour inbound leg.....	71
Table 4-15: Stop frequency model estimation for tour outbound leg.....	71
Table 4-16: Purpose estimations for outbound stops.....	74
Table 4-17: Purpose estimations for inbound stops.....	74
Table 4-18: Stop location model estimation for tour outbound leg	79
Table 4-19: Stop location model estimation for tour inbound leg	79

Table 4-20: Long distance travel demand model input data.....	85
Table 4-21: Output of each model component	85
Table 4-22: Output of long distance travel demand model system for each person...	86
Table 6-1: 60-Iteration Calibration Hyper-Parameter.....	105
Table 6-2: 60-Iteration Calibration Hyper-Parameter.....	105
Table 6-3: 60-Iteration Calibration Results	105
Table 7-1: Comparison of trips between Base Scenario and Fuel Price Increase Scenario.....	134
Table 7-2: Comparison of trips between High Speed Rail Scenario and Base Scenario	141
Table 7-3: Percentage of trip changes by travel mode between HSR scenario and Base scenario	144
Table 7-4: Trip Changes between TAZs by Trip purpose	145
Table 7-5: Changes of number of train trips by trip purpose between TAZs	148
Table 7-6: Changes of number of car trips by trip purpose between TAZs	149
Table 7-7: Changes of number of air trips by trip purpose between TAZs	149
Table 7-8: Comparison of trip shares by travel mode between HSR and Base scenario	151
Table 7-9: Comparison of percentage of trips and trip changes between HSR and Base	154
Table 8-1: Six Sets of Long Distance Trip Purpose Categorization Schemes.....	159
Table 8-2: Model Variables Used for Long Distance Trip Purpose Estimation.....	165
Table 8-3: Model 1 Results.....	166

Table 8-4: Model 2 Results.....	167
Table 8-5: Model 3 Results.....	167
Table 8-6: Model 4 Results.....	168
Table 8-7: Model 5 Results.....	169
Table 8-8: Model 6 Results.....	169
Table 8-9: Compared Model Developments.....	170
Table 8-10: Full Model Estimation Results	170
Table 8-11: Reduced Model Results.....	171
Table 8-12: Minimized Model Results	171

List of Abbreviations

ACS	American Community Survey
ATS	American Travel Survey
BTS	Bureau of Transportation Statistics
CEDDS	Complete Economic and Demographic Data Source
CPI	Consumer Price Index
CSTDM	California Statewide Travel Demand Model
DB1B	Airline Origin and Destination Survey
EIA	U.S Energy Information Administration
FHWA	Federal Highway Administration
FLSWM	Florida statewide travel demand model
FP	Fuel price scenario
GDP	Gross Domestic Product
GSP	Gross State Product
GTC	Generalized Travel Cost
GUI	graphical user interface
HI	Household Interview
HSR	high speed rail
IIA	Independency of Irrelevant Alternatives
IPF	iterative proportional fitting
IPU	iterative proportional updating
ISTEA	Intermodal Surface Transportation Efficiency Act
ISTEA	Intermodal Surface Transportation Efficiency Act of 1991
KYSTM	Kentucky Statewide Travel Demand Model
LDM	Long Distance Model in UK
LDPT	Long Distance Personal Travel Program
LOS	Level of Service
MCA	Multiple Classification Analysis
MDCEV	Multiple Discrete-Continuous Extreme Value
MPO	Metropolitan Planning Organization
MSA	Metropolitan Statistical Area
MSTM	Maryland Statewide Transportation Model
MTAS	Multimodal Travel Analysis System
NHTS	National Household Travel Survey
NHTS	National Household Travel Survey

NMS	National Model System in Netherlands
NRTF	national road traffic forecast
NRTS	National Rail Travel Survey
NTS	U.S National Transportation Studies
NUMA	National Use Model Area
OD	Origin-Destination
PETRA	Danish activity-based travel demand model
PUMA	Public Use Microdata Areas
PUMS	Public Use Microdata Sample data
RITA	Research and Innovation Technology Administration
RV	Recreational Vehicle
SAMPERS	Sweden national travel demand model
SP	Stated Preference
SPSA	Simultaneous Perturbation Stochastic Approximation
SRS	Simple Random Sampling
SWIM2	Oregon Statewide Integrated Model
TAZ	Traffic Analysis Zone
TOY	Time of Year
TRANS-TOOLS	The most recent pan-European travel demand model
TSAM	Transportation Systems Analysis Model
TTC	Total Travel Cost
TTT	Total Travel Time

List of Mathematical Symbols

d_{mn}	the distance between county or city m in zone i and county or city n in zone j
D_{ij}	the distance between zone i and zone j
U_{ij}	the utility value of person choosing travel mode i for long distance activity j between a specific OD
$tc_{ij.rn}$	total travel cost of mode i for j th long distance activity when travel cost falls into the range of rn
tt_{ij}	total travel time using mode i for j th long distance activity
α_n	the coefficients of total travel costs
β	the total travel time coefficient
ε_{ij}	error term capturing the factors that affect utility
U_i	the utility value of person choosing to travel during the time period i ,
$logsum_i$	mode choice logsum during time period i
α	mode choice logsum coefficient
X	vector of person's characteristics
B_i	vector of person's characteristics coefficients for time alternative i
V_{ik}	representative mode utility for the tour by mode k during time i
$f(t)$	the probability that a person will survive beyond the time period t
$S(t)$	discrete time survival function
$F(t)$	the failure function giving the probability of the event has occurred by duration t
$h(t)$	hazard rate, the probability of an event occurs given that one has survived to that time t
h_{it}	the probability of an event occurs given that one (i) has survived to the that time (t)
$i (1, 2, \dots, n)$	individual
t	the discrete time
α_t	the baseline hazard function
X_{it}	the covariates or explanatory variables of individual i at time t
$\gamma, \theta, \text{and } \beta$	coefficients which need to be estimated in tour duration model
Cost (S_i, D)	cost from stop i to tour destination
Cost (O, D)	cost from tour origin to tour destination

$\text{Cost}(S, D)$	cost from stop to tour destination
$\text{Cost}(S_i, S_j)$	cost from stop i to stop j
$\text{Cost}(S_j, D)$	cost from stop j to tour destination
$\text{Cost}(O, S)$	cost from tour origin to stop
w	is a vector of weights indicating the trust of the modeler on different observed outputs in calibration
O_m	a vector representing the observed outputs
O_s	a vector representing the simulated outputs
θ	the vector of parameters from the model to be calibrated
l, u	vectors of lower bounds and upper bounds for the parameters
$F(Z; \theta)$	the link between the simulated outputs and the simulation-based model
Z	the inputs required to run the simulation-based model
$\ .\ $	the Euclidian norm
a, c, α, γ , and A	hyper-parameters in SPSA algorithm
$\text{Gain}(S, a_i)$	the information gain of the attribute a_i in the data set S
$\text{Info}(S)$	the information value of the data set S
(L_j, V_i)	the leaf node L_j in subdivision V_i
$\text{Info}(L_j, V_i)$	the information value of leaf node L_j in subdivision V_i resulting from the data split on attribute a_i

Chapter 1: Introduction

1.1 Background

The increasing interest of national transportation policies from strategic infrastructure investment to infrastructure operation and management with regard to efficiency, sustainability, and safety has attracted researchers and decision makers to call for advanced and policy-sensitive tools for analysis (Lundqvist & Mattsson, 2001). The increase of national travel also requires the analysis tools beyond the urban and regional level. The highway infrastructure investment, the high-speed rail and the airport development all depend on national travel markets. To ensure that the infrastructure meets the demand growth it is imperative to model and analyze the passenger travel behavior at the national level (FHWA, 2013).

Americans travel a lot including inter-city, interregional and international travel. According to the 1995 American Travel Survey on the long-distance travel of persons in the U.S. (Bureau of Transportation Statistics, BTS), the U.S households made over one billion national-level long-distance trips and 41 million international trips (Zhang, et.al, 2012). National long-distance trips in the U.S can be of various purposes including business, leisure, personal business, family or friend visit and so on. All of the long-distance activities could constitute the economic and recreational opportunities that would benefit both the person and the area where the long distance activities occur. Thus, it is essential for the U.S from the economic and social perspectives to have the capability to support high level personal long-distance travel,

which requires that we have sufficient data and accurate analysis tools to be able to understand the long-distance travel behavior and forecast the travel patterns in the future. Without the analysis tools we could risk making inefficient and costly investments in our transportation infrastructure and management.

The needs for analyzing transportation capital expenditure decisions at the national level in the 1970s led to two U.S. National Transportation Studies (NTS) in 1972 and 1974 respectively (Weiner, 1976). These early national travel studies inventoried existing and planned U.S. transportation systems; and estimated future travel demand, system costs, performance, and broader impacts under alternative funding scenarios. With the completion of major investments on the Interstate Highway System, the development of national-level long-distance passenger travel analysis tools in the U.S. has been stagnant since the 1970s, though there have been continual academic interests in improving the theory and methods for multimodal intercity passenger travel demand analysis with a focus on mode choice (Lundqvist & Mattsson, 2001; Koppelman & Sethi, 2005; Bhat, 1995; Winston, 1985; Mannering, 1983,).

The lack of a capable long-distance passenger travel analysis tool in the U.S. is in sharp contrast with important emerging needs for analyzing various national transportation policies related to long-distance passenger travel. With regard to the national and regional long-distance travel, air, train, bus and auto modes compete each other. So any infrastructure investments or operational and management improvements should be evaluated through a capable national travel analysis tool instead of region-level, corridor-level, or state-level models which are mainly used in

these type of analysis. The Obama administration allocated \$8 billion in the 2009 stimulus funds for high-speed passenger rail, hoping that the supertrains would operate throughout the American landscape as they do in Europe and Asia (Billitteri, 2013). The U.S federal and state planners are prompted to provide the high speed rail services through selected major corridors. However, lacking a capable national long-distance passenger travel analysis tool in the U.S. has hindered the decision makers' and politicians' ability to systematically design and quantitatively evaluate the high speed rail in a broader view. Under this circumstance, it will be desirable to quantitatively forecast the high speed rail demand, systematically design and evaluate the operational effectiveness of the investment. Besides the high speed rail, there are also other national transportation investment strategies in need of the long-distance travel analysis tool to conduct quantitative analysis and evaluation, such as reconstructing and expanding the capacity of the Interstate Highway System, and building the next-generation air transportation system. In addition to these multimodal capacity investment needs for long-distance passenger travel, there are also urgent needs to assess a variety of operational and management strategies at the national level, which could significantly improve transportation efficiency and productivity, support and stimulate economic growth, and produce positive social and environmental impacts. Examples include: (1) Pricing, which could include congestion pricing on the Interstate and National Highway System, toll roads, air and rail fare increase, as well as fuel price increase. The fuel price change can affect the overall transportation system, The air fare and driving cost of auto are more related to the fuel price than the rail fare. (2) Congestion management at airports; (3) Separation

of passenger vehicles and heavy trucks on highway facilities; (4) National transportation financing options such as fuel tax increase and mileage fees; and (5) Substitution between long-distance travel and teleconference/telecommuting.

In addition to enabling national-level infrastructure investment and operational analysis, a long-distance passenger travel demand model for the U.S. also has the following important benefits: (1) Analyze the impact of socio-demographic, economic, and transportation infrastructure changes on long-distance travel demand. We know that people in the U.S mainly rely on travel for their everyday lives. How much people travel, when and where they go for different purposes, and which mode they take to get the destination is dependent on various factors including household socio-demographic characteristics, land use, transportation infrastructure, and so on (Contrino & McGuckin, 2009). One factor changes would cause people's travel behavior change, and different factors would weigh differently in determining their travel choice. How these factors influence the travel demand in different areas at the urban level or the metropolitan level has already been studied and can be answered based on the analysis tool - travel demand model at a certain geographic level. The same question arises on people's long distance travel, which is how the impact factors such as socio-demographic characteristics, economy, and transportation infrastructure changes would influence the long-distance travel demand. Such a question cannot be addressed accurately without a long-distance travel demand model.

(2) Anticipate the influence of energy (e.g. fuel price) and environmental factors (e.g. climate change and related regulations) on long-distance passenger travel. The impacts of fuel price changes on the long-distance travel behavior of the

population are critical in developing region-level or nation-level transportation policies and operations that can abate negative effects and increase benefits.

(3) Improve the long-distance passenger travel module in statewide and even some metropolitan travel demand models and provide an authoritative tool for multi-state transportation corridor analysis. After the Intermodal Surface Transportation Efficiency Act (ISTEA) in 1991 was enacted, many state departments of transportation started to develop statewide travel demand models and use them as critical analysis tools in addressing legislative requirements in statewide planning. However, the statewide models are weak in external trips which are usually generated with information from federal and neighboring states instead of available socioeconomic data (National Travel Demand Forecasting Model Phase I Final Scope). A national long-distance travel demand model therefore can provide external trips for statewide models in base-year and future-year. Meanwhile, it can also generate the travel demand for multistate corridors based on available datasets as well as standard and rigorous procedures which can minimize the duplication and efforts of the statewide models. (4) Support large-scale evacuation planning and operations due to natural disasters or targeted attacks; and (5) Enable micro-level analysis of the spread of pandemic diseases resulting from long-distance travel.

1.2 Objectives and Contributions

National long-distance passenger travel demand analysis has been an understudied area in transportation planning. The lack of multimodal long-distance origin-destination data has seriously limited planners' ability to conduct quantitative

analysis for operational effectiveness and infrastructure investment. As the nation and various states engage in funding transportation infrastructure improvements (interstate highway tolling/expansion, high speed rail, next-generation passenger air transportation system relying more on smaller airports and aircrafts) to meet future long-distance passenger travel demand, developing a national Multimodal Travel Analysis System (MTAS) and an American Long Distance Personal Travel Program (LDPT) become the priorities and the fundamental work when planners conduct the national travel analysis.

Therefore, this dissertation aims to develop a person-based activity national travel demand model for national travel analysis. All major behavioral dimensions of long-distance travel will be considered. Compare to the traditional four-step approach, activity-based techniques offer several advantages: (1) It is easier to consider tours, multi-day and multi-stop trips, and intermodal access/egress transfers that are important for long-distance travel modeling; (2) Households and persons are the basic units of analysis, which enables detailed behavioral representations and interactions; and (3) It provides a rich framework in which travel is analyzed as a multi-day, monthly, quarterly, or yearly pattern of behavior, derived from activity participation. There are also significant differences between long-distance trips considered in the proposed activity-based model and daily/weekly trips in metropolitan/state-level tour/activity-based models developed in previous research. For instance, the long distance trips usually take days or weeks and may involve car, airplane, train, bus, or multiple modes of the four. It is often the case that households firstly choose travel time for long-distance vacation trips based on time and money budget before

selecting destinations and mode. Categorization of trip purposes is also different for long-distance trips. Cost of travel for long-distance trips is not just travel disutility, but also includes lodging, food, and etc. The same applies to the total travel time for long-distance which usually covers not only in-vehicle travel time but also the ingress/egress time, transfer time, and lodge time. The much lower frequency of long-distance travel may also imply a different decision-making process.

The dissertation has contributed to both methodology developments from the academic perspective and real world application from the industry perspective.

First, this dissertation represents the first attempt to develop an integrated activity-based travel demand model system for individual's quarterly/yearly long distance or national activities and travel in the U.S at the Metropolitan Statistical Area (MSA)/Non-MSA level which is the highest geographic resolution in the long distance travel survey data. The model system is developed considering the specific attributes of the long distance travel such as low frequency, long activity duration, long activity duration at intermediate stops on the tour legs, different sets of mode alternatives, etc. Therefore, the model system not only takes into account the people's long distance travel at the tour level, but also at the stop level. Three levels of choice are modeled. The first level is the activity pattern level which determines the number of different types of activities a person will choose during one year; the second level is the tour level which contains choices of tour destination, time of year, tour duration, and tour mode; and the lowest level is the stop level model system including the number, the purpose, and the location of each intermediate stop made during the inbound and outbound legs of the tour. National-level travel survey data are used to

estimate the model components and provide the parameters for simulation. The model is implemented in our developed micro-simulation platform which simulates each individual's yearly long distance activities and travel in the U.S. over the course of one year.

Second, high speed rail (HSR) is expected to help alleviate the heavy load of the traffic in road and air corridors and improve the inter-regional accessibility (Börjesson, 2012). However, the construction of HSR requires large quantities of investments and efforts. In order to help decision makers and politicians evaluate and systematically design the HSR, we selected northeast corridors to conduct quantitative analysis and forecast the HSR demand based on our proposed and developed person-based national travel demand model. Meanwhile, the impact of fuel price increase at the national level on the national travel demand is also analyzed to evaluate the sensitivity of the national travel demand model.

Third, as the FHWA has planned the next iteration of long distance travel survey in the U.S, advanced technologies such as GPS, smartphone, social media, etc., are explored in the travel survey methods to provide the more accurate temporal-spatial information on travel than traditional surveys. However, the passively collected travel data from these new survey methods does not provide the important travel components such as trip purpose. Therefore, in the dissertation, a trip purpose learning system based on machine learning methods is proposed and tested with valid data.

This dissertation represents a more advanced academic research endeavor for national passenger travel analysis. Findings from this dissertation are expected to

provide important insight and help guide federal and state to make decisions on corridor-level, region-level, and nation-level infrastructure investment, design, and management, as well as to research on long-distance passenger travel demand.

Chapter 2: Literature Review

2.1 Long Distance Travel Demand Modelling

Long distance travel analysis methods can be classified into four groups, see Figure 2-1 (Zhang et al., 2012). All the methods in the figure can be used to estimate the multimodal OD matrices. However, they are different in terms of data requirements, and whether travel behavior responses to policy scenarios (Zhang et al., 2012).

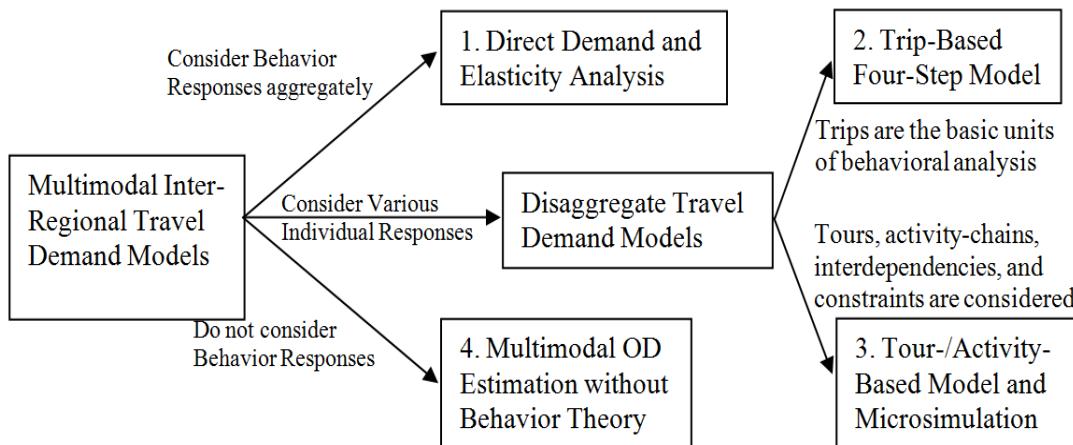


Figure 2-1: Categorization of Long Distance Travel Demand Analysis Methods¹

During the middle of 20th century, the U.S started its travel demand modeling at urban and metropolitan level. Since the passage of the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, an increasing number of states began to develop their statewide travel demand modelling in order to meet policy and legislative needs (Zhang, et al., 2013). In the US the current long distance travel

¹ Source: Zhang, L., Southworth, F., Xiong, C., & Sonnenberg, A. (2012). Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review. *Transport Reviews*, Vol. 32, No. 4, pp. 399-433.

demand models have been developed mainly serving as a component of statewide models, and the majority of the travel demand model developments are traditional or modified 4-step models. As the 4-step planning approach is easy to implement, and requires less data than tour-based or activity-based travel demand models. The United States long distance travel demand models effort reviewed in our research include those in California, Ohio, Oregon, Maryland, Florida, Kentucky, Tennessee, New Hampshire, and etc.

The California Statewide Travel Demand Model (CSTD) is a tour-based travel demand model. It can forecast all types of travel as well as long distance trips made by California residents, plus commercial vehicle travel. The travel modes in the model vary in different sub-models. The CSTD (ULTRANS & HBA, 2011) adopts statewide networks for roads and transit. Multiple data including the 2012 California Household Travel Survey, 2010 United States Census data, zonal land use data, employment, and population data were employed for model calibration and base year assignment (Bureau Transportation Statistic, 2015). The CSTD was developed expressly in order to evaluate the proposed high-speed rail (HSR) system connecting Southern and Northern California (Cambridge Systematics et al., 2006). According to the forecast analysis in the year of 2040 when the high speed rail will be operated in the CSTD study region, there will be 7.1 million high speed rail ridership from San Francisco region to Los Angeles region (Cambridge Systematics, 2016) where the population is forecasted as 47.95 million in 2040 and the distance is about 385 miles. Compared to the predicted high-speed rail ridership from San Francisco to Los Angeles, our scenario analysis (high speed rail) in the northeast corridor based on our

developed national passenger travel demand model forecast a total number of 2.6 million trips between Washington D.C and New York, And in the region there will be a total number of 25 million population in 2040 and the distance between D.C and New York is about 226 miles. Ohio statewide travel demand model uses state-of-the-art tour-based modeling approach. It includes person travel for both short distance (less than 50 miles) and long distance (larger than 50 miles). The long distance travel models are developed based on the Ohio DOT's sponsored long distance travel survey data, and can estimate the frequency and characteristic of long distance travel for assignment on transportation network (Erhardt et al., 2015). Oregon Statewide Integrated Model (SWIM2) is a second generation model and developed based on the First Generation based Statewide Model (SWIM1) and the Eugene-Springfield UrbanSim Model. SWIM2 is an integrated land use transportation model covering the whole state of Oregon. The transportation part of the SWIM2 includes the tour-based models of personal travel and commercial travel, as well as simple model of external truck travel (Parsons Brinckerhoff, 2010). Florida statewide travel demand model (FLSWM) employs the traditional 4-step model methods for the passenger and freight travel in the state of Florida (Giaimo & Schiffer, 2005). Maryland Statewide Transportation Model (MSTM) is a four-step travel demand model, and it is designed to generate link-level assignment for personal and freight travel. In terms of passenger travel, the statewide level models the short-distance or urban personal trips for the residents in the study area. The Regional level model includes a long-distance travel model for both residents and visitors with trips longer than 50 miles one-way. The freight travel module takes into account both short-distance truck travel and long

distance truck travel (University of Maryland & Parsons Brinckerhoff, 2011). Kentucky Statewide Travel Demand Model (KYSTM) models the long distance trips (over 100 miles) in Kentucky and part of the neighboring states. The model adopts a modified 4-step travel demand model, removing the mode choice module. The model can be used to evaluate traffic volume and economic impacts of new corridors (Bostrom, 2006). Tennessee Statewide Model was developed only for long distance trips over 75 miles, and it also employed a 3 step travel demand model (no mode choice model), Future year travel demand forecasting is conducted based on MPO forecasts of population and employment growth (Stammer, 2002), New Hampshire Statewide Travel Demand Model is developed to address the needs of New Hampshire Department of Transportation in the areas of statewide planning, congestion management, intermodal management, public transportation, air quality, financing policies, etc. It is a tour-based micro-simulation framework of sequential multinomial logit models (Sharma et al., 1999).

According to Giaimo and Schiffer's review on statewide travel demand modeling developments, most of the statewide travel demand models in the US do not consider the long distance travel (Giaimo & Schiffer, 2005). Among the statewide models that consider long distance travel, many employ Fratar approach to estimate future year travel demand from base year OD table. Under this method, long distance travel demand has little sensitivity to policy changes. A lot of researchers have extended their efforts on long distance travel demand modeling. Koppelman (1989) developed a behavior framework and model system for intercity travel, and the 1977 National Travel Survey data was used to estimate the model. Li (2004) examined the

frequency and mode choice of intercity non-business trip through the developed intercity travel behavior system which uses a nested logit/continuous model. Forinash and Koppelman (1993), Bhat (1995), and Lee et al. (2004) studied intercity mode choice through exploring a set of discrete choice models.

Baik et al. (2008) developed a travel demand model at the national level which can predict the annual county-to-county personal travel for commercial airline, air taxi, and automobile in the U.S at 1-year interval through 2030. The transportation systems analysis model (TSAM) adopted the trip-based four-step travel demand model process which includes trip generation, trip distribution, mode choice and network assignment. An intercity trip in the system is defined as the one with one-way route distance larger than 100 miles, excluding commute travel. Different from the traditional network assignment in four-step travel demand model, the network assignment in TSAM is developed for the commercial airline and air taxi. The TSAM outputs annual county-to-county person round-trips by travel mode, trip purpose (business and non-business) and household income group. The network assignment then outputs the annual flights between all the commercial and air taxi airports. Cambridge Systematics (2008) conducted a study in order to provide specifications for national travel demand forecasting model development. They proposed a trip-based four-step demand model structure (not an operational model) which focuses on providing travel information for statewide models. Data sources including network and zone system, demographic and employment data, freight data and travel behavior data are assessed prior to preparing the input data. Nostrand *et al.* (2013) presented a research on national long distance travel demand modeling only for leisure purpose.

An annual vacation destination and time choice model is developed, using the Multiple Discrete-Continuous Extreme Value (MDCEV) structure (Bhat, 2005), to estimate the destinations that a household would visit during a year and the time allocated for each of the vacation destinations. The model mainly relied on the 1995 American Travel Survey data for analysis, and a total of 210 zones were divided for the whole nation. The output of the model can be used to construct a national-level OD table for leisure travel. The funded FHWA project *Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework* (Outwater et.al, 2015) which was conducted at the same time as our work, developed a long-distance passenger travel demand model for the whole U.S, using a disaggregate, tour-based approach. The model component estimations are based on different travel survey including 1995 American Travel Survey, 2012 California Statewide Household Travel Survey long-distance component, 2003 Ohio Household Travel Survey, 2010 Colorado Front Range Travel Survey, and 2001 National Household Travel Survey (NHTS) which span multiple years from 1995 to 2012. They adopted a 4570-zone structure of own developed National Use Model Area (NUMA). The long distance passenger travel demand model mainly used the MDCEV approach. In the model framework, only travel information at the tour level is captured and modeled including tour generation, scheduling, tour duration, travel party size, tour destination choice and mode choice. Travel information at the stop level is not captured.

Meanwhile, in Europe, a lot of attention has been paid to national travel demand modeling during the last two decades. In 1997, the UK employed the direct demand and elasticity analysis method in the national road traffic forecast (NRTF)

model (Worsley & Harris, 2001) to predict the national level travel demand. Multiple direct demand models were developed and estimated for vehicle ownership/use, level of service (LOS) of transportation network, truck traffic, and traffic flows. With the models elasticity, a hierarchical set of switching rules are defined to evaluate the full impact of policy scenarios on the road network. After then in 2010, the UK Department of Transportation developed a Long Distance Model (LDM) to predict long distance passenger travel demand (trips over 50 miles) (Rohr et al., 2013; Burge et al., 2011). It can be used to examine policies including demand for high-speed rail (HSR) (Burge et al., 2010), and policies that can influence long-distance car, classic rail and air demand. In order to develop the model, multiple data sources such as 2002-06 National Travel Survey (NTS) data, 2004-06 National Rail Travel Survey (NRTS), and 2009 Household Interview (HI) data were employed. The nested logit choice model was used for demand estimation, and network models were developed for highway, rail and air. An Italian national travel demand model was developed and applied to different macroeconomic, transport supply, and HSR study and marketing scenarios (Ben-Akiva et al., 2010). The Italian national travel demand model system is consisted of three sub-models, demand model that can predict the future year OD matrices from the base year, the nested logit mode choice model that estimates the market share of inter-urban travel modes, and induced demand model that predicts the additional HSR demand as a result of improved HSR LOS.

The National Model System (NMS) in Netherlands was developed in 1986 and is being updated since then (HCG & TOI, 1990; Gunn, 2001). It has served as a “prototype” disaggregate model in Europe and was built based on behaviorally

oriented tour-based method (Gunn, 2001). The model system consists of a series of connected choice models including license holding and car ownership models, tour frequency, tour mode and destination choice, and time-of-day. Total 345 zones were divided in the NMS system and 1302 sub-zones were used in the sub-module of mode and destination choice models. The NMS is sensitive to a variety of socio-economic, land use, transportation systems, and policy factors. Applications were observed in the rail demand prediction for railway options, impact of raising fuel prices, effect of the introduction of motorway signaling, high speed trains demand analysis, and etc (Hofman, 2001). The Norwegian national model system followed the structure of the Dutch NMS, but with the objective of emissions (CO_2 and NO_x) prediction at the national level (HCG & TOI, 1990). Therefore, no detailed link-loadings on a national network were needed (Gunn, 2001). Sweden started its national model development in the beginning of 1980s, and has improved it to the current new version of SAMPERS which belongs to the mainstream trip-based four-step model (Zhang et al., 2012). The new model covers the trips in Sweden and to neighboring countries in detail and trips to and from other parts of Europe in a coarser way (Widlert, 2001), which results in three different model systems: regional models, domestic long distance models and international models. All models in the three systems, such as car ownership, trip frequency, destination choice, mode choice, and departure time choice, adopted discrete choice logit model, except for an ordinary least squares trip frequency model for foreigners' travel to Sweden (Beser & Algers, 2001). The Danish PETRA model was developed as an activity-based method to travel demand analysis (Fosgerau, 2002). In the model, a person's daily travel is represented in terms

of chains of tours instead of separate trips or tours. In order to reduce the complexity, the observed chains were transformed to a simplified version of chain types. The model system first deals with the cohort effects on car ownership and license holding (Lundqvist & Mattson, 2001). Then a mode and destination choice model (nested logit model) was estimated for each tour in a chain. Finally, the choice of chain type is modeled, and the accessibility measured by the logsum from the mode and destination model is incorporated in the chain type choice model. Since there is no network assignment module in PETRA, no congestion analysis is considered. A variety of applications have been observed with the model in analyzing the effects of different policy measures on people's travel behavior. From the perspective of the geography and the travelling population size, many of the European national travel modeling efforts are closely equal to statewide model studies in the U.S (Zhang et al., 2012). The pan-European models include the countries forming the European Union, which to some extent are close to the national model study in the U.S. The pan-European travel demand model developments started with the estimation of multimodal OD matrices (method 4 in Figure 2-1) without behavior framework in the MYSTIC (Peter Davidson Consultancy, 2000) project in the early 1990s and then the DATELINE project of 2004 (Brog et al., 2004; Davidson & Clarke, 2004). The MYSTIC project developed a heuristic harmonization procedure to merge various data sources from seven European countries to obtain pan-European OD matrices. The more recent DATELINE project refined the methodology in the MYSTIC project, and also used the data from a new pan-European long-distance travel survey for 16 countries. Then aggregate methods and joint aggregate-disaggregate methods in pan-

European models were employed in the 201-zone NUTS2 STREAM model (Williams, 2001) and the 1275-zone NUTS3 STEMM model (Gaudry, 2001) respectively. Finally, the most recent pan-European travel demand model is the disaggregate TRANS-TOOLS model integrating European transportation and economic models (Zhang et al., 2012; Burgess et al., 2006).

In addition to the applications of the national travel demand model in the transportation field, we also found that it is used in other areas. Epstein, Parker, and et.al (*Epstein et al., 2008*) developed an agent-based microsimulation model for intercity travel in a research on spatial-temporal epidemic dynamics. The model simulates individual's travel decisions on trip frequency and destination choice based on a zip-code-level OD system. Since travel demand analysis is not a focus of the study, no mode choice or assignment models were employed. Even so, this study demonstrated the benefit of long distance model in other fields in addition to the transportation area.

2.2 Long Distance Travel Survey Data

The most recent survey of long distance passenger travel, the American Travel Survey, was conducted in 1995. The 2001 National Household Travel Survey has a long-distance travel component but with a relatively small sample. An up-to-date long distance travel survey is necessary and provides the latest travel data needed in statewide or national travel demand modelling which the state and federal governments need to meet their policy requirements and legislative development needs and to predict future travel demand (Cohen et al., 2008; Giaimo & Schiffer,

2005; Horowitz, 2006; Horowitz, 2008; Souleyrette et al., 1996). With the rapid development of technology, GPS, smartphone, social media, etc. become researchers' and governments' new tool to supplement or replace traditional survey methods in long distance travel data collection. However, the advanced technology-based method cannot provide all the necessary long distance trip information such as travel mode, and trip purpose. Thus, the practical post-processing methods that can generate data on these missing travel characteristics are essential in a GPS/smартphone/social media-based survey.

A number of travel researchers are exploring GPS-based travel survey methods, and as a result, have developed different methods to derive trip purpose from collected trip data, mainly in the area of regular intra-regional travel. Wolf et al. (2001a, 2001b) pioneered the procedures of trip purpose derivation that are based on a set of deterministic rules. By collecting data from a sample of 19 respondents in Atlanta GA who utilized a GPS data logger and returned a completed paper trip diary, and by using data from a detailed GIS database of land use, they found that mixed-use land use parcels such as shopping center, office building and strip mall challenges the accurate trip purposes derivation. Besides the GIS land use data, respondent's socio-economic characteristics such as household composition, possession of travel modes, and home and work addresses can contribute to trip purpose derivation as well. The procedure was further developed by Schönfelder et al. (Axhausen et al., 2003; Schönfelder et al., 2002) in Europe who used a multi-stage hierarchical matching procedure that involved calculating a cluster center of stop ends. This was done by combining trip ends, identifying trips with obvious purposes and establishing

relationships between trip purposes and activity temporal information as well as the socio-demographics of the respondents. Stopher et al. (Stopher et al., 2005; Stopher et al., 2008a; Stopher et al 2008b) established a set of heuristic rules to derive trip purpose for 43 trips collected in Sydney with the parcel-level land use data and the geo-coded addresses of respondents' workplace/school and the two most frequently used grocery stores. Bohte et al. (2009) developed a GPS-based travel data collection method utilizing GPS devices, GIS technology and a web-based validation procedure. Then, based on the collected GPS data, they derived the trip purposes with the developed heuristic rules. Chen et al. (2010) employed the same approach as Schönfelder's to cluster trip ends into activity locations. Grounded on the GIS data and respondent's socio-demographic characteristic, deterministic rules were used to classify trip purposes for trips occurred in low-density area. For trips in high-density areas, trip purposes could not be deterministically identified, and a multinomial logit model was employed to calculate the probability that a trip served a particular purpose, where only four trip purpose categories were considered.

The method of deriving trip purpose based on GPS-based data was further explored with the help of artificial intelligence and machine learning. Griffin and Huang (2005) employed the decision tree method to derive trip purposes, in which the procedure was implemented in the C4.5 environment with 50 randomly generated trips which were simulated based on a series of assumptions. Alternatively, Deng and Ji(2010) used a set of attributes (attributes from GPS data, such as time stamp, spatial-temporal indices of trips and attributes from GIS data, and the social-demographic and socio-economic characteristics of respondents) to construct a decision tree in

travel models and trip purposes derivation. The decision tree was implemented in the C5.0 environment with a homogenous set of 226 GPS trips collected from 36 respondents in Shanghai. Lu et al. (2013) explored the feasibility of automating trip purpose imputation using multiple machine learning methods (decision tree, Support Vector Machine, and Metalearner) with the help of geospatial location data, land use data, and the GPS-based survey conducted by University of Minnesota. A heterogeneous sample of 2238 trip records with a 7-trip purpose categorization scheme is utilized.

Chapter 3: Data

The data employed in the dissertation contains two parts: one for model estimation and the other for the base year model simulation. The main component data source for model estimation is the travel survey data, besides the transportation OD skim data and economic/demographic information of traffic analysis zones (TAZs) which are also the input in the model simulation part. The other input data for model simulation is the population data which contains individual and household characteristics such as age, gender, employment status, household type, household income, etc.

3.1 Zone System

The traffic analysis zone in our national travel demand model system is Metropolitan Statistical Area (MSA) and non-MSA that is the remaining area of a state not belonging to an MSA. Therefore, the number of the non-MSA is equal to the number of states in the U.S. Although the non-MSA usually has larger area than MSA and it will be desirable to divide the non-MSA into smaller zones or to use county or city as the zone in the national travel demand model, the finest geographic resolution in the ATS data which we mainly relied on for model estimation is MSA and non-MSA. The United States include 3202 counties and the MSA/Non-MSA is consisted of multiple counties or cities. Aggregating the counties or cities can give us the MSA and Non-MSA.

Figure 3-1 shows the traffic analysis zone (TAZ) system in the national travel demand model. There are a total of 380 zones (excluding Puerto Rico) covering the mainland of the United States, Alaska, and Hawaii. In the 1995 ATS sample data, there are a total of 208 zones (161 MSAs and 47 non-MSAs) used for model estimation.

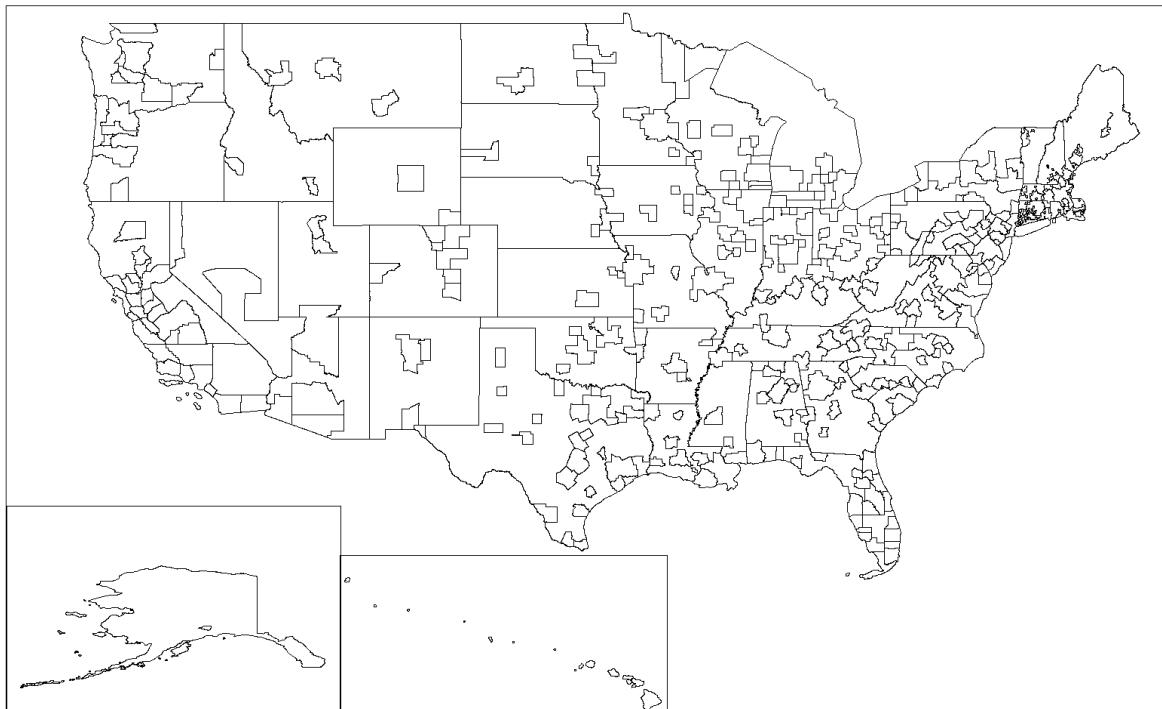


Figure 3-1: National Travel Demand Model Traffic Analysis Zone System

3.2 *Travel Survey Data*

The 1995 American Travel Survey is the primary source of the national travel survey data used in the research for model estimation. It is a long distance nationwide travel survey in the United States, and was conducted by the Bureau of Transportation Statistics (BTS) between April 1995 and March 1996. The 1995 ATS data collected

detailed long distance travel (>100 miles) information and demographic information from more than 80,000 random selected households from the 50 states and the District of Columbia, and each household was interviewed every three months during the survey period. "One of the ATS's main objectives is to comply with two requirements of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA): (1) to provide information on the number of people carried in intermodal transportation by relevant classification, and (2) to provide information on patterns of movement of people carried in intermodal transportation by relevant classification in terms of origin and destination." (Hwang & Rollow, 2000). Undeniably, the 1995 ATS is a little old, but it is the most recent dataset in long distance travel over the course of one year in the U.S, and it incorporates the information on stops during the tour legs. Although the household travel survey data, 2001 National Household Travel Survey (NHTS), has a long-distance travel component, it is a relatively small sample and has less long distance information than the 1995 ATS.

The 1995 ATS provides comprehensive information. To be specific, the gathered demographic information in the survey includes the characteristics of both the household (family type, household income, household size and etc.) and the household members (gender, age, employment status, race and etc.). The detailed long distance travel information contains the origin and the destination of the trip, stops along the way from/ to the destination, side trips originating at the destination, the means of transportation, the reason of the trip, the lodging type, the number of nights spent away from home, travel party size, and etc. All the location information was recorded at regional, state, and metropolitan level. Such detailed information

cannot be found in other U.S. national travel surveys and makes the 1995 ATS a useful source of long distance travel survey data.

The 1995 ATS database includes four data sets: household trips, household characteristics, personal trips, and personal characteristics, of which the latter three datasets were utilized in this research. The personal trip data included 556,026 trip records for both domestic and abroad long distance trips, and a total of 45374 long distance trips were made by people older than 18 years old and dozens of travel modes were recorded for the entire travel (inbound trip and outbound trip), and 438022 of the 45374 trips were made domestic. In the dissertation, only adults' domestic long distance trip will be covered. Based on the characteristics of the detailed travel modes, 18 classes of the travel modes were aggregated into 6 classes including car, air, bus, train, recreational vehicle (RV) and others. The percentage of the long distance tour travel modes in the sample can be plotted (Figure 3-2), where Others refers to the combinations of multiple travel modes and other modes such as motorcycle and bicycle. It can be observed that above 80% of the long distance activities were made by car for the entire tour, which has the highest share of the travel mode usage. The second most popular travel mode is the air which accounts for 15.6%. From the 11 activity types in 1995 ATS, three main activity types (business, personal business, pleasure) were aggregated to reduce the model computation complexity (Table 3-1) based on the activity similarities.

Figure 3-3 illustrates the percentage of the aggregated three trip purposes. Almost 60% of the long distance trips are made for pleasure, and 25.8% and 15.6% of

the long distance activities are made for business and personal business related respectively.

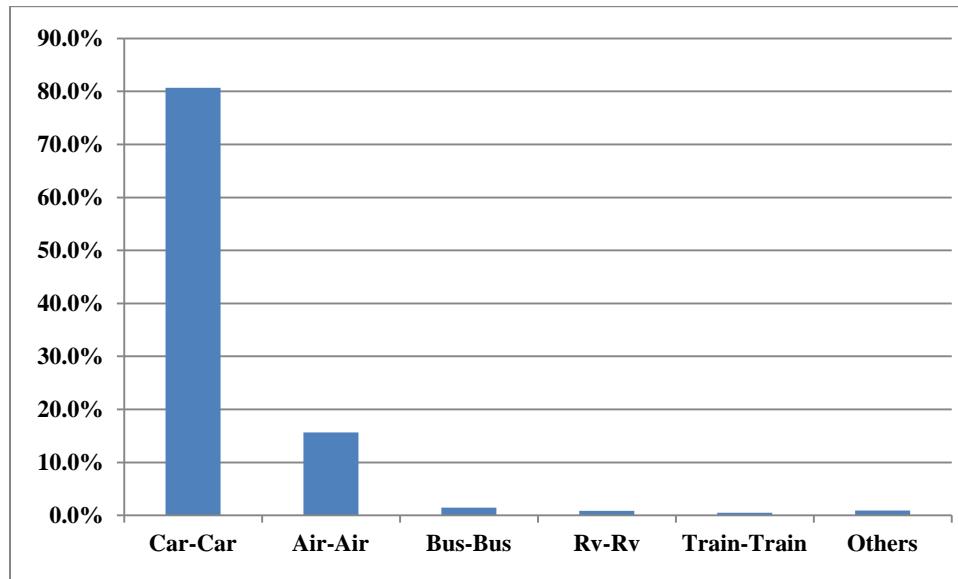


Figure 3-2: Percentages of Long Distance Tour Travel Modes

Reported Trip Purpose	Encoded Trip Purpose
Business	Business
Combined Business/Pleasure (B/P)	Business
Convention, Conference, or Seminar	Business
School-related activity	Personal Business
Visit relatives or friends	Pleasure
Rest or relaxation	Pleasure
Sightseeing, or to visit a historic/scenic attraction	Pleasure
Outdoor recreation	Pleasure
Entertainment	Pleasure
Shopping	Pleasure
Personal, family or medical	Personal Business
Others	Deleted

Table 3-1: Encoding Reported Trip Purposes

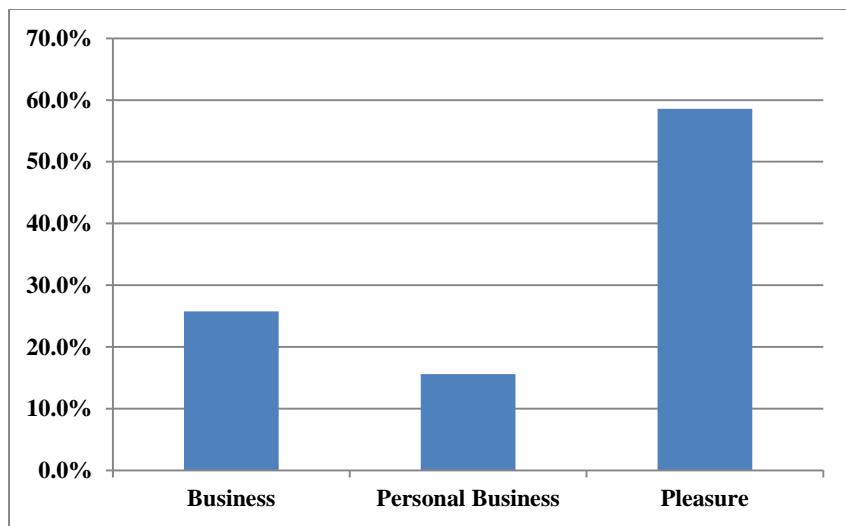


Figure 3-3: Percentages of Long Distance Trip Purposes

When people make long distance travel for different purposes, they usually have different tendency for transportation means. For example, people travelling for business purpose are usually more likely to take air mode than traveling for personal business and pleasure purposes, while people tend to make personal business and pleasure trips by car. As expected, compared to pleasure and personal business trips, people preferred choosing air for their business and personal business travel (Figure 3-4). Up to 47% and 43% of air trips were for business and personal business, and only 10% were pleasure trips. Among people travelling by car, 61% of them made pleasure trips, and only 22% and 17% of people traveled for business and personal business. With regard to people travelling by bus, 73% of them traveled for pleasure purpose. Among all the long distance trips by train, pleasure trips had the largest share (53%) and personal business trips have the least share (only 8%).

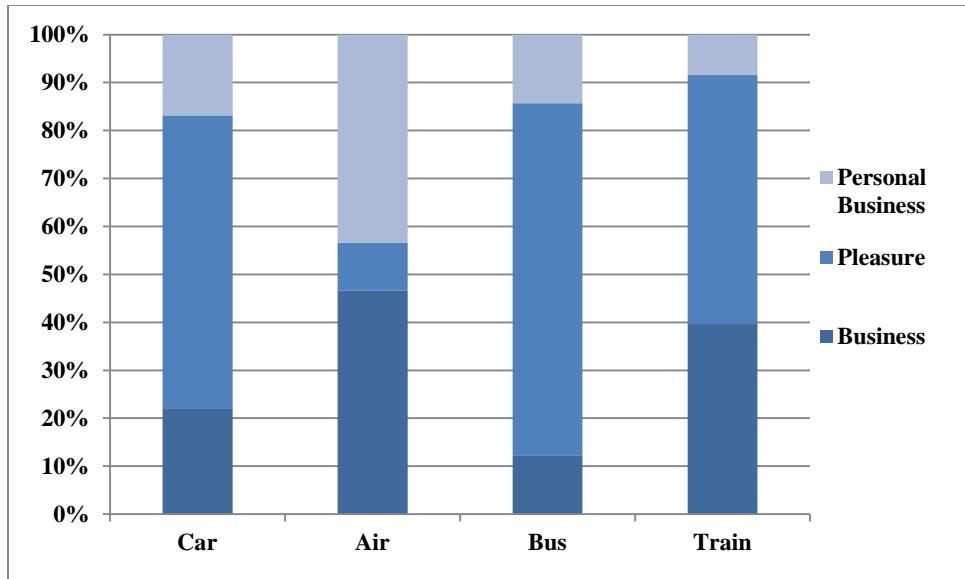


Figure 3-4: Travel Mode Usage by Trip Purpose

It is observed from 1995 ATS data (Figure 3-5) that the largest share (28%) of the long distance trips occurred during the third quarter (from July to September), whereas the fewest people (20.6%) chose to make long distance travel during the fourth quarter (from October to December). The third quarter was also the peak time period to make pleasure travel (up to 30.1%). Meanwhile, the largest proportion of the business trips (29.9%) occurred during the first quarter (from January to March), and from the first quarter to the fourth quarter the share of the business trips decreased. The second quarter (from April to June) is the most popular time period for people to make long distance travel for personal business (29.7%), and then followed by the third quarter.

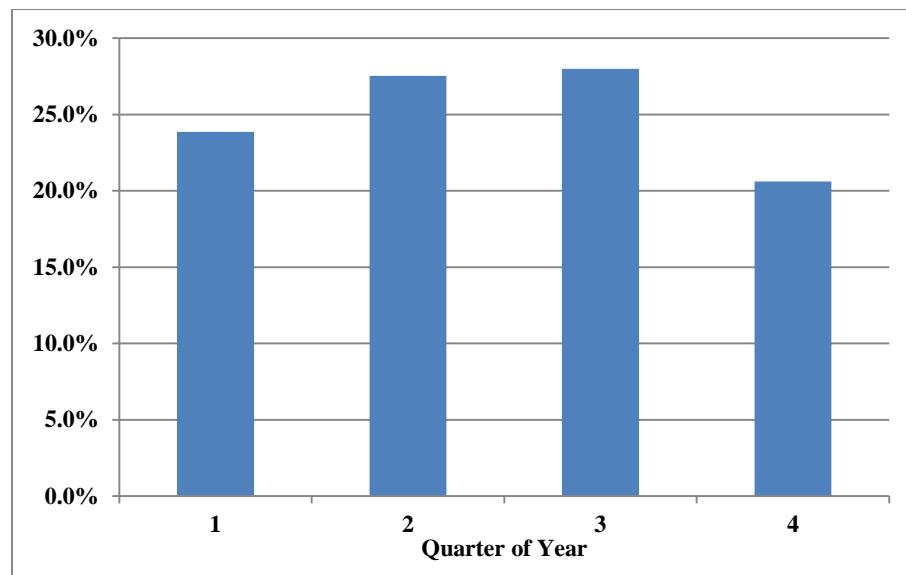


Figure 3-5: Trip Distribution by Time of Year

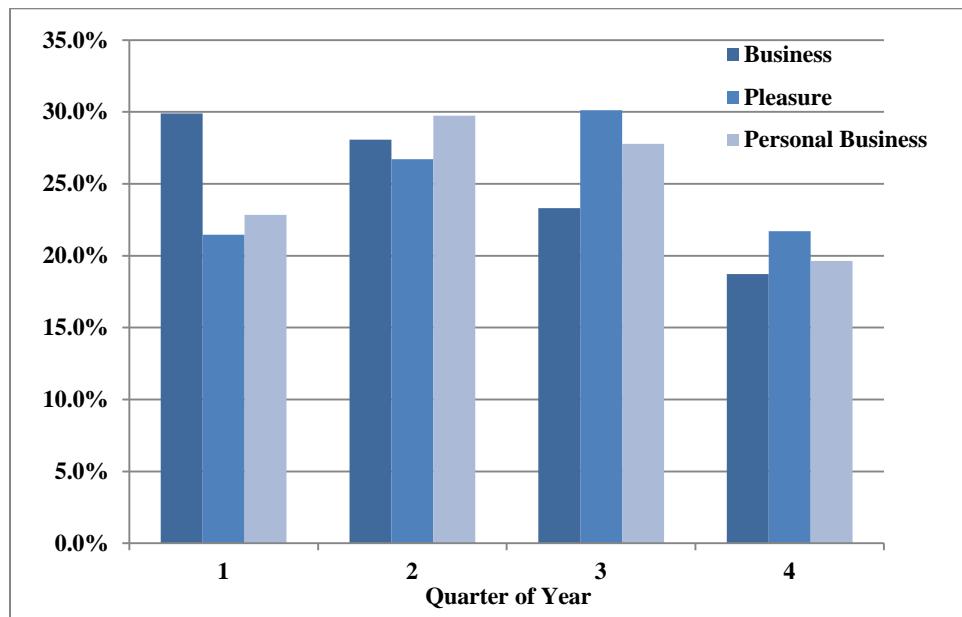


Figure 3-6: Trip Distribution by Purpose and Time of Year

When people make long distance travel, they usually would make stops during the half legs of the tour. The 1995 ATS data records the stop information in terms of ‘Number of Stops’ and ‘Reason for Each Stop’ for each long-distance trip. In this research, the stop is redefined as the one that people make during the tour legs for a

certain purpose (e.g. business, personal business, and pleasure). The stops for rest and transfer in the same travel mode or across the modes will be eliminated from the study. Figure 3-7 illustrates the distribution of the number of stops in both outbound and inbound legs of the long distance tour. It can be found out that most people (above 95%) did not stop during their long distance travel, and only few people would make one stop (0.2% or 0.1%) or four stops (0.3% or 0.3%) during either inbound or outbound trip of the long distance tour. The distribution of the stop purposes shown in Figure 3-8 indicates that most of people make stops for pleasure purpose either during the inbound or outbound leg of the long distance tour. And if they decide to stop for pleasure, more people would choose to stop on the way back home rather than on the way to the primary tour destination. The smallest share of the stops made during either inbound or outbound leg was for personal business.

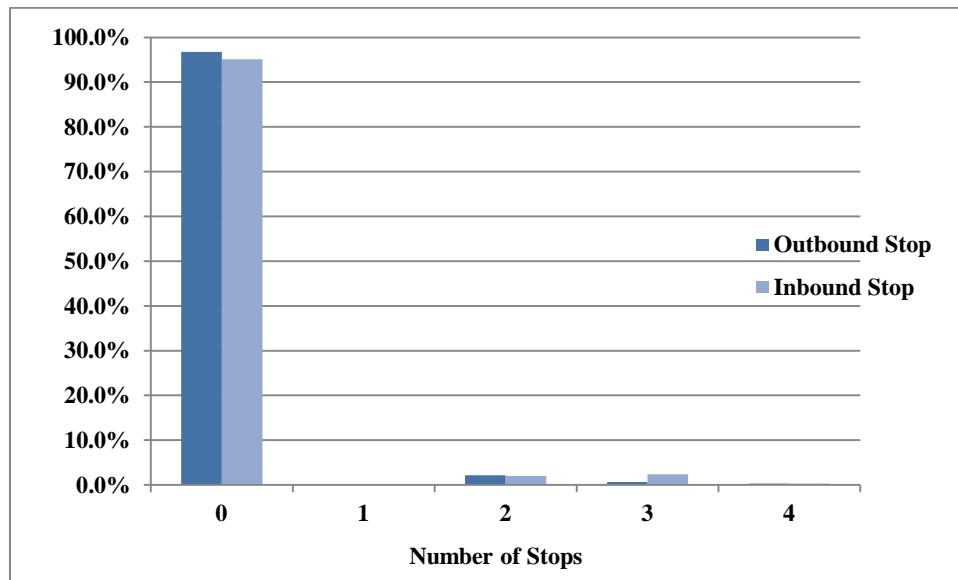


Figure 3-7: The number of Inbound/Outbound Stop Distribution

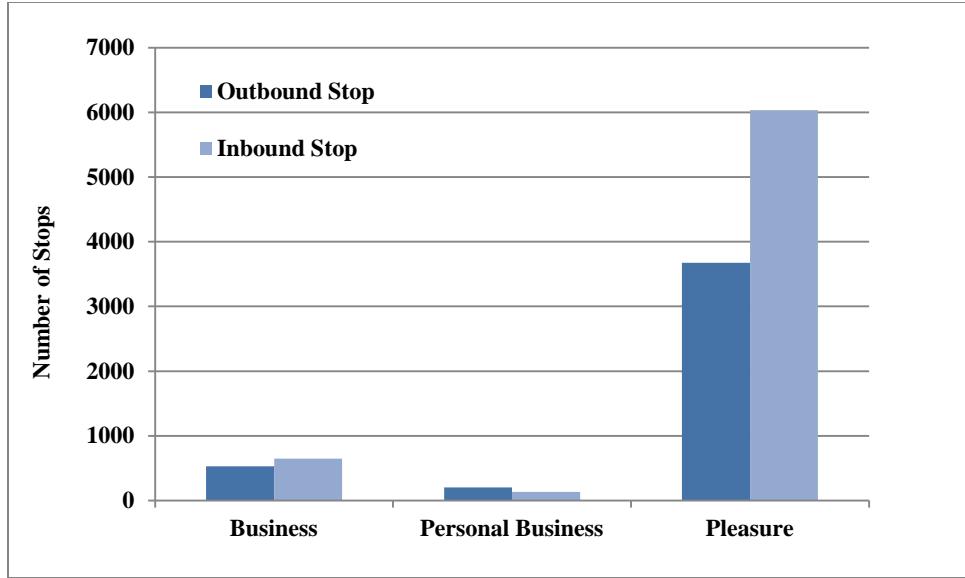


Figure 3-8: Outbound/Inbound Stop Purpose Distribution

3.3 Transportation OD Skim and Economic/Demographic Data

According to the percentages of the travel modes and the difficulty in accessing the travel data about the bus and the RV, only three travel modes (car-car, air-air, and train-train) were considered for the entire tour of the long distance activities. The OD skim data or level of service variables primarily refer to the travel time and travel cost between each origin and destination pair via different travel modes, and it can be observed as the functions of distance between TAZs. Usually an MSA or a non-MSA is made up of more than one county or city, so the distance between two TAZs can be estimated by averaging the distance between all the county-county (county-city and city-city) pairs in each zone (Equation 3.1).

$$D_{ij} = \frac{\sum_{m,n} d_{mn}}{m \times n} \quad (3.1)$$

where i, j refer to the zone of the MSA or non-MSA; m, n indicate the number of the county in zone i and zone j , respectively; d_{mn} is the distance between county or

city m in zone i and county or city n in zone j . The Census Bureau provides the geographic information of each county/city in the U.S., which assists us to estimate the great circle distance² between each zone pair.

Auto travel time and cost were derived as a series of functions with the information of the great circle distance of a zone pair, the average driving speed, the vehicle's characteristics, and etc. The cost of the vehicle driving usually contains the fuel cost, the insurance, the maintenance and the tire costs. And among them only the fuel cost is out-of-pocket expense for the trip, while the other costs are paid separately after the trip. Therefore, only fuel cost is considered for the vehicle driving cost during the long distance travel. Several assumptions are made in order to estimate the auto travel time and cost: 1) the average auto speed is 65 miles/hour; 2) the auto travel time/cost consists of two parts spent on driving and lodging; 3) people on business travel will stop for an overnight stay every 9 hours, while people taking personal business and pleasure travel will stop every 13 hours; 4) the auto average fuel efficiency is 19.7 mpg (Grush, 1998) and the average retail fuel price is \$1.48/gallon in the U.S. in 1995 (EIA, 2016); 5) the average lodge cost per person night for business travel from the low-income to the high-income are \$70, \$90, and \$110 respectively, while the lodge cost for personal business and pleasure are \$30, \$50, and \$70 respectively; 6) the travel party size is 1 person for business travel, and 2 persons for personal business and pleasure trip, which helps to estimate the vehicle driving cost for each person.

² The great circle distance: the shortest distance between two points on the surface of a sphere. Distance= $6371 * \text{acos}(\sin(\text{latitue1}) * \sin(\text{latitue2}) + \cos(\text{latitue1}) * \cos(\text{latitue2}) * \cos(\text{longitude2} - \text{longitude1}))$, where 1 and 2 refer to the point 1 and point 2 on the surface of a sphere.

Air fare and the number of layover were collected from the Airline Origin and Destination Survey (DB1B) data provided by the Bureau of Transportation Statistics, Research and Innovation Technology Administration (RITA). DB1B is a 10% sample of airline tickets from reporting carriers; therefore, in order to obtain a sample size as large as possible the DB1B data from 1994 to 1996 was employed. The air travel time is made up of access/egress time (time spent in traveling to the airport and from the airport to the final destination), air fly time, and transfer wait time between flights. Air fly time is estimated with the obtained great circle distance and the average flight speed which is assumed as 500 mph (Boeing, 2011). The average total access and egress time is assumed as 2 hours for all the air travel, and the average wait time per transfer is set as 1.5 hours. The number of layover is obtained based on the airport groups in the DB1B data which lists the airport codes of all the airports in the flight itinerary. Since the airport code is of three characters and the airports in the airport group variable are separated by colons, the number of layovers can be obtained according to the number of the characters in the variable. By multiplying the average wait time per transfer and the number of layovers, we can acquire the total transfer wait time in the itinerary between TAZs. Air fare was taken from the DB1B data, and we eliminated the first/business class fare to reduce the travel cost variance and to be in line with the fact that the majority of the travelers choose the economic class for their air travel. A MSA or non-MSA may have more than one airport from the geographic perspective, therefore, the air fare and time between TAZs should be the average value between all the airport pairs in the corresponding zones.

It is hard to get the train fare and time data in 1995 or neighboring years. Therefore, we collected the train fare and travel time in August, 2013 from Amtrak, a national railroad passenger corporation, as a proxy. The Amtrak website provides an access to look up station-to-station timetable and ticket information. The Amtrak train ticket has several classes including saver, value, flexible, and premium. Generally, saver class ticket charges the least, flexible or premium class ticket charges the most, and the price of the value class ticket is at the middle level of all the classes' fares. While collecting the train fare for our study, we chose the value class price and then converted the fare to 1995 dollar according to the Consumer Price Index (CPI). The travel time from Amtrak contains both the train travel time and the transfer waiting time from origin station to destination station. In economically developed regions, a TAZ may have multiple rail stations. Under this circumstance, the TAZ-to-TAZ train fare and time is achieved by aggregating the fare and time between all station pairs from the two zones.

Data for the zones' attractiveness indexes in this research mainly include the total population, the number of employment by industry sector, and the number of households. These economic and demographic data for each MSA and non-MSA was obtained from the Complete Economic and Demographic Data Source (CEDDS) by Woods & Poole Economics. This database offered historical, current and projected socioeconomic indicators (e.g. population, employment, households, etc.) for all the regions, states, statistical areas and counties in the U.S.

3.4 Public Use Microdata Sample Data

The input data for the base year model is the 2010 1-year ACS Public Use Microdata Sample data (PUMS) that represents about 1 percent of the total U.S. population or approximately 1.3 million housing unit records and about 3 million person records (Census Bureau, 2008). Detailed person and household information were stored in person record and housing unit record in one PUMS file for each state. Each record had a unique identifier linking the people to the proper housing unit record. The specific geographic unit was defined as Public Use Microdata Areas (PUMAs) in 2000 Census PUMS, and each PUMA contained a minimum population threshold of 100,000. The housing unit record in PUMS file contains detailed household information such as home ownership, real estate taxes, number of vehicles, number of persons in the household, household type, household unit weight, presence and age of own children, PUMA code, state code, household income, MSAPMSA code, and etc. The person record includes person's information like age, gender, race, marital status, education attainment, school enrollment, employment status, means of transportation to work, travel time to work, class of worker, personal income, person weight and etc. The weight in PUMS file for each person and housing unit can be used to expand the sample to the relevant total. Since each state is consisted of one or more PUMAs and some large metropolitan areas may be divided into several PUMAs, a PUMA could contain parts of multiple TAZs. Geographic equivalency between PUMA and MSA/non-MSA needs to be identified prior to the data being employed. All the persons or housing units located in a PUMA containing mixed TAZs should be allocated to each TAZ according to the population percentage of the TAZs in the

specific PUMA. Table 3-2 shows part of the PUMA and the MSA/Non-MSA correspondence. POP in the table indicates the total population in the MSA/Non-MSA. It can be told that the PUMA 800 in state 1 contains two TAZs (MSA/Non-MSAs) which are 1000 and 0199. According to each TAZ's (MSA/Non-MSA) population percentage in PUMA 800, we can assign 62% of the PUMA 800 population to TAZ 1000 and 38% population to TAZ 0199.

STFIPS	PUMA	CountyFIPS	MSA/NonMSA	POP	pumapop	percentage
1	100	33	2650	285900	285900	1
1	200	89	3440	315904	315904	1
1	800	9	1000	231532	372958	62%
1	800	127	0199	141426	372958	38%

Table 3-2: Part of PUMA and MSA/Non-MSA Correspondence Table

Chapter 4: Model System Analysis Framework

The long distance trip in our model system is defined as the ones greater than or equal to 50 miles one way. For each long distance activity, there is only one tour destination or primary destination, and during each leg of the tour, there could be multiple intermediate stops. People can also travel or make side trips (stops) based on their primary destination (Figure 4-1). Due to the data limitation of the side trips, our national travel demand model system will not cover the part of people's side trips or stops.

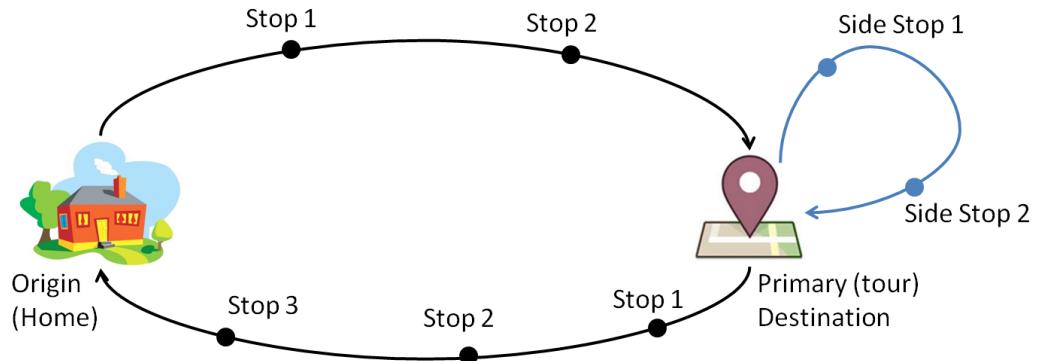


Figure 4-1: Long distance travel illustration

Compared to short distance travel or urban travel (less than 50 miles), long distance travel has its own characteristics of longer travel distance, low frequency and longer duration usually in days (see Figure 4-1). Usually, people do not take any or take several long distance trips during one year, and it is unlikely that people take long distance travel every day. The commute trip longer than 50 miles one way is not taken into account in this research. Also, the trip purposes for long distance travel are different from the purposes of short distance travel which mainly include work/school,

social/recreation, shopping, pickup/drop-off, meal, and errands/personal/family business. The long distance trip purposes are mainly business, pleasure and personal business. Compared to short distance travel, long distance travel is unlikely to be made by non-motor travel modes including walk and bike. And it is also impossible for people to take air for the short distance travel, which is a main travel mode for long distance travel.

	Long Distance Travel	Short Distance Travel
Distance	≥ 50 miles	<50 miles
Frequency	Low frequency, multiple trips per year	Daily trips, multiple trips per day
Activity Duration	Days	Hours, Minutes
Trips Connection	Usually large time gap between multiple long distance trips or activities during one year	Tight connection between trips and activities
Travel Modes	Mainly car, air, train, bus	Mainly car, transit, non-motor mode (bicycle, walk)
Travel Purposes	Business, pleasure, personal/family business	Work/school, social/recreation, shopping, pickup/drop-off, meal, errands/personal/family business

Table 4-1: Comparison between long distance and short distance travel

The activity-based national travel demand model we developed can generate the long distance passenger trips made by auto, air, and train in the U.S. in one year period. It can serve as a forecasting tool of long distance travel in the U.S. The model system has root in econometric model developments including discrete choice model and duration model. These models are employed to guarantee the maximum behavior realism and model sensitivity to regional and national projects and policies. The model is implemented in a micro-simulation framework which simulates the long

distance travel for each adult in the U.S. Since the finest spatial resolution in 1995 ATS data is metropolitan statistical area and non-metropolitan statistical area, we adopted MSA and Non-MSA as our traffic analysis zone system. The model system consists of three tiers, see Figure 4-2, 1) the yearly long distance activity pattern level which estimates the number of different types of activities a person will choose during one year; 2) the tour level model system which contains choices of tour destination, time of year, tour duration, and tour mode; 3) the stop level model system which estimates the intermediate stop frequency, the purpose and the location of each stop made during the inbound and outbound legs of the tour.

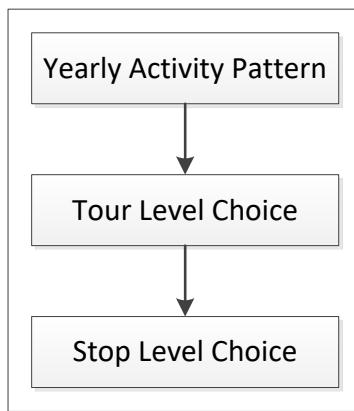


Figure 4-2: Activity-Based Long Distance Travel Demand Model System

4.1 Activity Pattern Level Model

The demand for long distance activities and travel can be considered as a choice among all the possible bundles of activities and travel annually. The model system adopts a timeframe of one year because of less frequent long distance travel , and days, weeks and even months of activity duration. Dissimilar to regular urban-level activities schedule, people choose their long distance activities with few

interactions within one year due to much less frequent long distance travel. As shown in Figure 4-3, the yearly long distance activity schedule can be presented as a set of different long distance activities per year, and all the long distance activities at this level are the primary activities. So the yearly long distance activity pattern can be presented as {B-x, PB-y, P-z}, where B, PB, and P stand for the activity type of business, personal business and pleasure respectively, and x, y, z are integers ($x, y, z \geq 0$) referring to the number of the corresponding activities during one year. In the model, the Multiple Classification Analysis (MCA) method which is a mostly-used trip generation method in traditional four-step travel demand model is employed to estimate the long distance trip rates by activity type. The long distance trip rates for each purpose are shown in Table 4-2, Table 4-3, and Table 4-4.

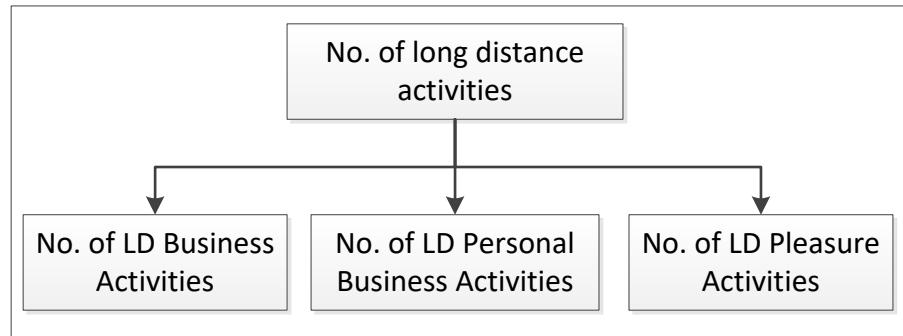


Figure 4-3: Yearly Long Distance Activity Pattern Level

		Low Income		Medium Income		High Income	
		Male	Female	Male	Female	Male	Female
Employed	MSA	1.478	0.407	2.051	0.980	2.860	1.789
	Non-MSA	1.731	0.660	2.304	1.233	3.113	2.042
Un-Employed	MSA	0.433	0.000	1.006	0.000	1.814	0.743
	Non-MSA	0.685	0.000	1.258	0.187	2.067	0.996

School	MSA	0.293	0.000	0.866	0.000	1.675	0.604
	Non-MSA	0.546	0.000	1.119	0.048	1.928	0.857

Table 4-2: Trip Rate for Long Distance Business Travel

		Low Income			Medium Income			High Income		
		Age1	Age2	Age3	Age1	Age2	Age3	Age1	Age2	Age3
Couple w/o Children	MSA	2.223	2.567	2.564	2.766	3.110	3.108	3.267	3.610	3.608
	Non-MSA	2.521	2.864	2.862	3.064	3.408	3.405	3.564	3.908	3.905
Couple w Children	MSA	1.970	2.314	2.312	2.514	2.857	2.855	3.014	3.357	3.355
	Non-MSA	2.268	2.611	2.609	2.811	3.155	3.152	3.311	3.655	3.652
Single	MSA	2.070	2.414	2.412	2.614	2.957	2.955	3.114	3.457	3.455
	Non-MSA	2.368	2.711	2.709	2.911	3.255	3.252	3.411	3.755	3.752
Non-Family	MSA	2.106	2.450	2.447	2.649	2.993	2.990	3.149	3.493	3.491
	Non-MSA	2.403	2.747	2.745	2.947	3.290	3.288	3.447	3.790	3.788

* Age1: age between 19 and 35; Age2: age between 36 and 55; Age 3: age larger than 55

Table 4-3: Trip Rates for Long Distance Pleasure Travel

		Age 1			Age 2			Age 3		
		Employed	Un-employed	School	Employed	Un-employed	School	Employed	Un-employed	School
Couple w/o Children	MSA	0.385	0.643	0.755	0.627	0.884	0.997	0.725	0.982	1.095
	Non-MSA	0.866	1.124	1.236	1.108	1.365	1.478	1.206	1.463	1.576
Couple w Children	MSA	0.243	0.500	0.613	0.485	0.742	0.855	0.582	0.840	0.952
	Non-MSA	0.724	0.981	1.094	0.966	1.223	1.336	1.063	1.321	1.433
Single	MSA	0.146	0.403	0.516	0.387	0.645	0.757	0.485	0.743	0.855
	Non-MSA	0.627	0.884	0.997	0.868	1.126	1.238	0.966	1.224	1.336

Non-Family	MSA	0.000	0.219	0.331	0.203	0.460	0.573	0.301	0.558	0.671
	Non-MSA	0.442	0.700	0.812	0.684	0.941	1.054	0.782	1.039	1.152

* Age1: age between 19 and 35; Age2: age between 36 and 55; Age 3: age larger than 55

Table 4-4: Trip Rates for Long Distance Personal Business Travel

4.2 Tour Level Structure

Each long distance activity schedule has a primary tour, and may have zero or more intermediate stops or side stops during the legs of the tour and at the destination. In our model system, the secondary tours or the side stops that occur based on the long distance primary destination are ignored due to the data limitation and its coverage of urban- or metropolitan-level travel. The tour level model system defines the characteristics of the primary tour of each long distance activity such as the tour destination, time of year, tour duration, travel party size and tour travel mode. When we develop and estimate each model component at the tour level, it is assumed that the outcomes of the upper-level model, the household, person characteristics and mobility attributes are already known. So the solid arrow in Figure 4-4 indicates that the output of the upper level can be used as an explanatory variable at the lower level, while the dash arrow means that the expected utility of the lower-level models can affect the choices at the upper level. Meanwhile, the upper level decisions of tour duration and destination will constrain the travel mode choice at the lower level; for instance, if people from Washington, D.C., only have one day to travel to California and get back, it is unlikely that he/she will drive. In reality, the proportion of the long distance tour travel time in total duration varies by person. As the 1995 ATS data has limitation on the activity duration information at the primary destination, we

simplified the temporal constraint and made assumption that for each person the total tour travel time should not exceed half of the total tour duration. When people decide to make a long distance activity, they usually have different priority considerations and decision procedures for different activity types. In the research, we made a set of assumptions about people's decision making process of long distance travel at the tour level. For example, the long distance pleasure activity (a discretionary activity) requires people to consider their time availability prior to other decisions. When they have a period of time (days, weeks or months) for pleasure, they will decide when to spend it, where to go, whom to go with and how to go sequentially. In contrast, people taking long distance business and personal business activities usually give priority to the decisions of the activity location and the time (including time of year and duration), followed by travel party size and tour mode choice. Therefore, two different tour level structures are proposed for business/personal business and pleasure (Figure 4-4). According to the direction of the dash lines in the figure, both time of year models and destination models should include the expected utility variable from the mode choice model (mode choice logsum).

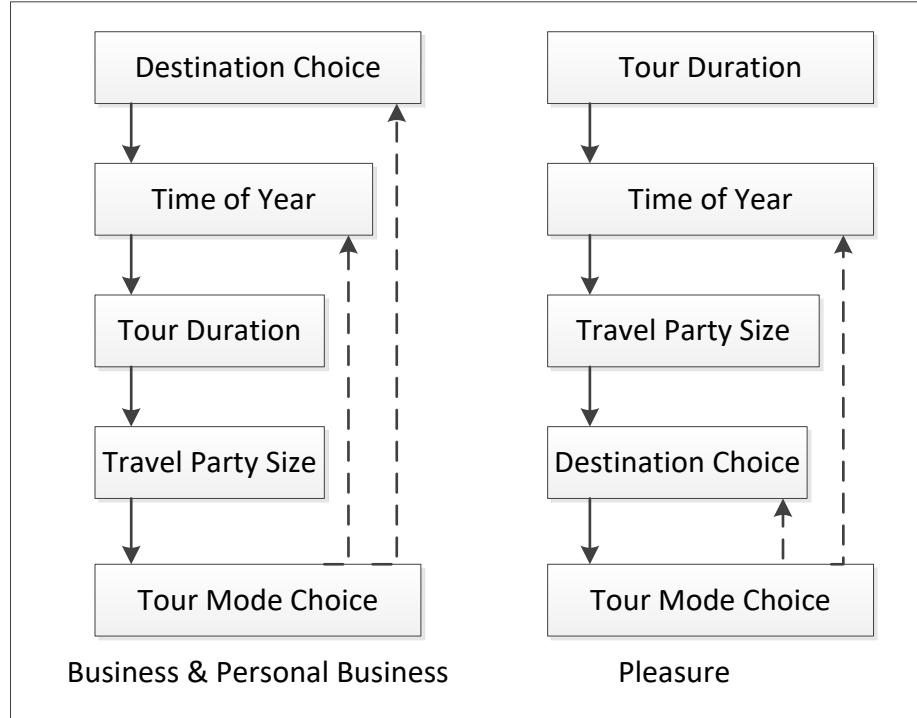


Figure 4-4: Tour Level Procedure and Model Components

All the model components are estimated mainly based on the 1995 ATS data, and most of the models are estimated using discrete choice model (multinomial logit model) except the tour duration model. In our research, we use 80 percent of the data sample to conduct the model estimation, while the rest of 20 percent of the data sample to validate the estimated model.

4.2.1 *Travel Mode Choice Model*

Three travel modes are modeled at the tour level, i.e. {(car, car), (air, air), (train, train)}, and no combination of different travel modes is considered due to the small sample size in the ATS data (See Figure 3-2). Multinomial logit model is employed to develop the travel mode choice model, and a piecewise linear utility function (Ben-Akiva & Lerman, 1985) shown in Equation (4.1) is adopted.

$$U_{ij} = \alpha_1 \cdot tc_{ij,r1} + \alpha_2 \cdot tc_{ij,r2} + \dots + \alpha_n \cdot tc_{ij,rn} + \beta \cdot tt_{ij} + \varepsilon_{ij} \quad (4.1)$$

U_{ij} : the utility value of person choosing travel mode i for long distance activity j

between a specific OD;

i: refers to the three tour level travel modes, {(car, car), (air, air), (train, train)};

j: one of the three long distance activities (business, personal business, and pleasure);

$tc_{ij,r1}, tc_{ij,r2}, \dots, tc_{ij,rn}$: total travel cost of mode i for jth long distance activity

when travel cost falls into the range of r1, r2, ..., rn;

tt_{ij} : total travel time using mode i for jth long distance activity.

$\alpha_1, \alpha_2, \dots, \alpha_n$: the coefficients of total travel cost for different travel cost ranges;

β : the total travel time coefficient;

ε_{ij} : error term capturing the factors that affect utility, but are not observable by the researcher

In order to study the high-speed rail, air travel behavior (Hess et al. 2007), and some other travel behavior with choice situations not yet revealed in the market, stated preference data is suggested in the mode choice. However, for the forecasting purpose, as there exists gap between the interviewees' stated preference/response and their actual preference/response (Wardman, 1988), the error in the model makes it unsupported for the analysis (Daly and Rohr, 1998).

Table 4-5 illustrates the estimation results for the tour mode choice. And all the variables have expected signs and are significant at the 95% confidence level. The coefficient estimation implies that as the cost of the travel mode people choose increases, the value of time will increase correspondingly. And in general, people

taking long distance pleasure trip or personal business trip have larger VOT than taking business trip. This can be explained by the fact that people going on a business trip can get reimbursement for both their travel and accommodation expense, while they need to pay for their pleasure and personal business travel out of the pocket.

Variables	Business	Pleasure	Personal Business
TTC(<=\$188)	-0.0325	-0.0095	-0.0127
TTC (>\$188 & <=\$332)	-0.0093		
TTC (>\$332 & <=\$476)	-0.0066		
TTC(>\$476 & <= \$620)	-0.0037		
TTC (>\$620)	-0.0028		
TTC (>\$188 & <=\$312)		-0.0043	-0.0057
TTC (>\$312 & <=\$436)		-0.0009	-0.0040
TTC (>\$436)		-0.0003	
TTC (>\$436 & <=\$560)			-0.0028
TTC (>\$560)			-0.0011
TTT	-0.0356	-0.0590	-0.0328
Constant-Air	-0.4400	-2.9500	-1.4900
Constant-Train	-2.9300	-3.5600	-3.7500
Rho-Square	0.536	0.754	0.682

* TTC: Total Travel Cost

TTT: Total Travel Time

Bold and Italic: variables are significant at 95% confidence interval

Table 4-5: Tour Mode Choice Model Estimation Results

Figure 4-5, Figure 4-6, and Figure 4-7 present the aggregate share of the observed tour travel mode choice and the estimated one for business, pleasure, and personal business purposes. Results show that the mode choice models can estimate the mode choice market share in the right trend but with errors. Compared to business and personal business mode choice models, pleasure mode choice model has a

weaker performance. The air mode for pleasure purposes is under estimated by almost 40%, and the large percentage is also caused by the small sample size of the air pleasure travel in the validation data set.

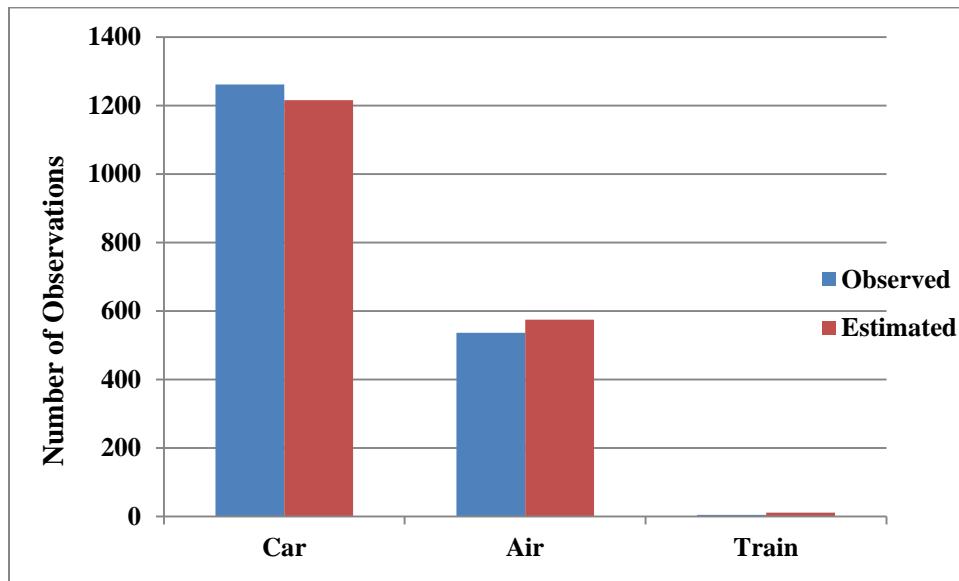


Figure 4-5: Tour Mode Choice Validation for Business Purpose

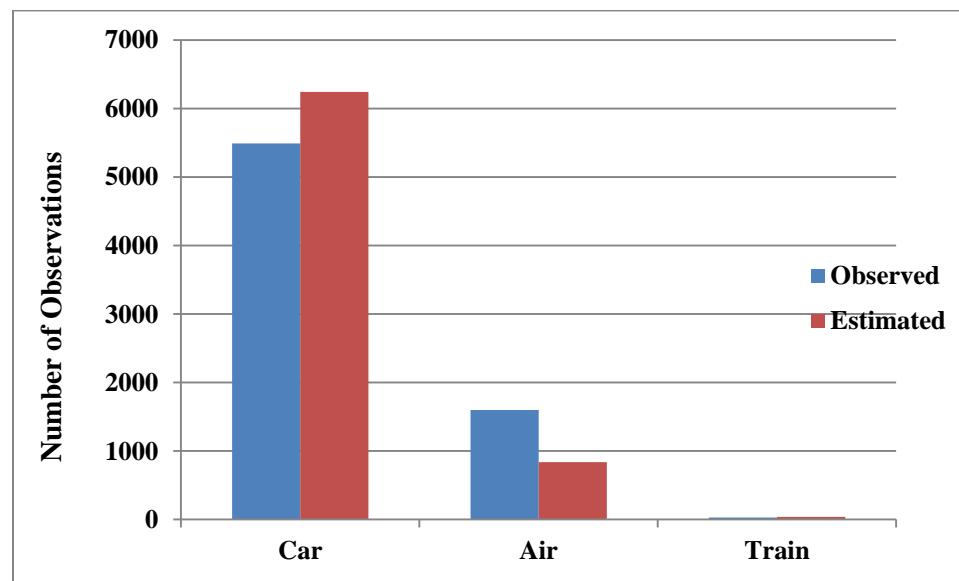


Figure 4-6: Tour Mode Choice Validation for Pleasure Purpose

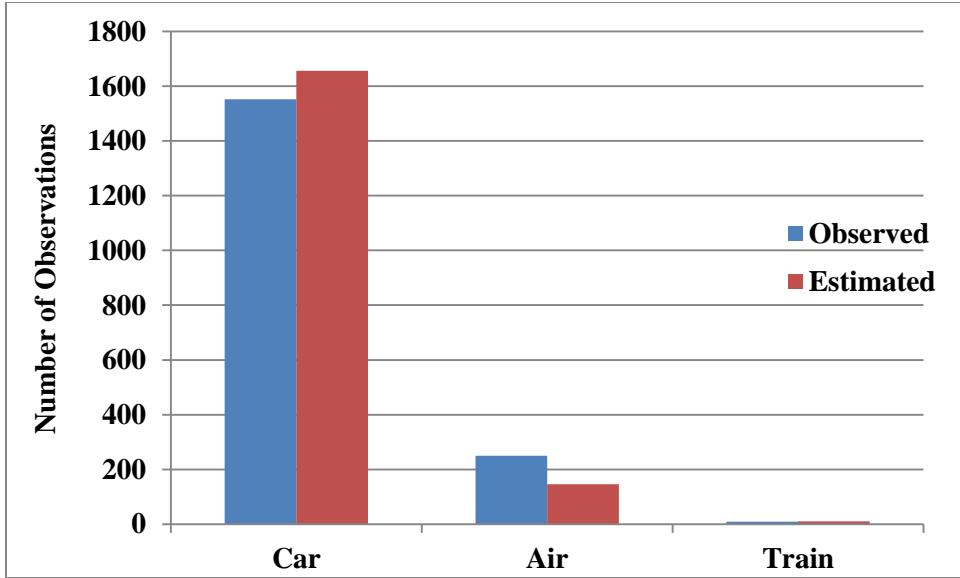


Figure 4-7: Tour Mode Choice Validation for Personal Business Purpose

4.2.2 Time of Year Choice

Due to the fact that the finest temporal resolution in ATS is quarter, our proposed model system will function at a time resolution of three-month or one-quarter. The three-month increments begin in January and end in December, thus four quarters in total. In the ATS sample data, few records are observed that depart from and arrive at home across quarters. Consequently, we adopted a choice set of only four alternatives $\{(Q1, Q1), (Q2, Q2), (Q3, Q3), (Q4, Q4)\}$ for each person when he/she decides what time of the year to travel and what time of the year to get back, where $Q1, \dots, Q4$ refer to Quarter 1 to Quarter 4. Multinomial logit model is adopted for time of year choice model, and the model employs the person and the zonal characteristics most of which are generic across the four time alternatives. Since transportation network LOS attributes vary by time periods especially the air fare with large price fluctuations in different seasons, these variables are specified as

alternative-specific based on the four time alternatives. The general form of the TOY model utility can be represented in Equation (4.2)

$$U_i = \alpha \cdot \text{logsum}_i + B_i \cdot X + \varepsilon_i \quad (4.2)$$

$$\text{logsum}_i = \ln \sum_k e^{V_{ik}}$$

where

U_i : the utility value of person choosing to travel during the time period i, where

$i=1, 2, 3, 4$, referring to (Q1, Q1), (Q2, Q2), (Q3, Q3), (Q4, Q4);

logsum_i : mode choice logsum during time period i, and it represents the total ease of travel between two TAZs across all available travel modes during time period i;

α : mode choice logsum coefficient;

X : vector of person's characteristics;

B_i : vector of person's characteristics coefficients for time alternative i;

V_{ik} : representative mode utility for the tour by mode k during time i;

At the tour level for pleasure long distance activity, a simplified time of year choice model which only takes into account the person's characteristics is firstly developed and applied when the pleasure destination is not known. The time period assigned from this simplified model will serve as an input or known attributes for the destination choice model. Once the destination is chosen, a full time of year model considering the mode choice logsum is re-run to choose the final time period (Figure 4-8).

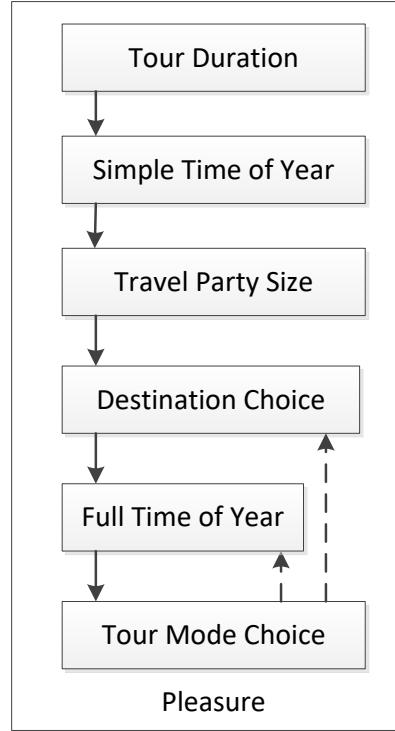


Figure 4-8: Re-simulating Time of Year Choice Model

The following three tables (from Table 4-6 to Table 4-8) present the estimation results for time of year choice under the long distance business and pleasure trip. The fourth quarter (Q4, Q4) is set as the base alternative for all the TOY models. Coefficients in bold and italic format are significant at the 95% confidence level. The mode choice logsum variables in the full TOY models infer that people tend to take their long distance trip in the quarter with larger accessibility in terms of mode choice logsum. Moreover, the mode choice logsum coefficient in pleasure full time of year model has a larger value than that in business full TOY model, which means that people are more sensitive to the mode accessibility when they take long distance pleasure trips. It can be explained by the fact that people care little about their travel expense when taking business trips due to the reimbursement of business travel. Due to the unexpected sign of the mode choice logsum coefficient and its

insignificance in personal business full time of year model, the model will not be used. Instead, the time of year distribution {(Q1,Q1): (Q2, Q2): (Q3,Q3): (Q4,Q4) = (0.228: 0.297: 0.278: 0.197)} for the personal business trip in the survey data (1995 ATS) will be used to assign each person's personal business trip a quarter based on Monte Carlo simulation. As noted, the personal business trips are usually School-related activities and Personal, family or medical activities (Table 3-1). So it is reasonable that when people travel for these activities, they are not sensitive to the time of year they choose to travel. Good estimation can be observed from the without sample validation results (Figure 4-9, Figure 4-10, and Figure 4-11), having a very small difference between the observed time of year distributions and the estimated ones.

	(Q1,Q1)	(Q2,Q2)	(Q3,Q3)	(Q4,Q4)
Mode choice logsum	0.05			
Household Income	8.02e-06	-1.39e-06	3.61e-06	0.000
Employed	0.586	0.245	-0.385	0.000
Couple with Child	0.376	0.290	0.002	0.000
Age	0.03	0.037	0.0003	0.000
Constant	-2.330	-2.678	-1.527	0.000
Pseudo R-Square	0.05			

Table 4-6: Full Time of Year Choice for Business Trip

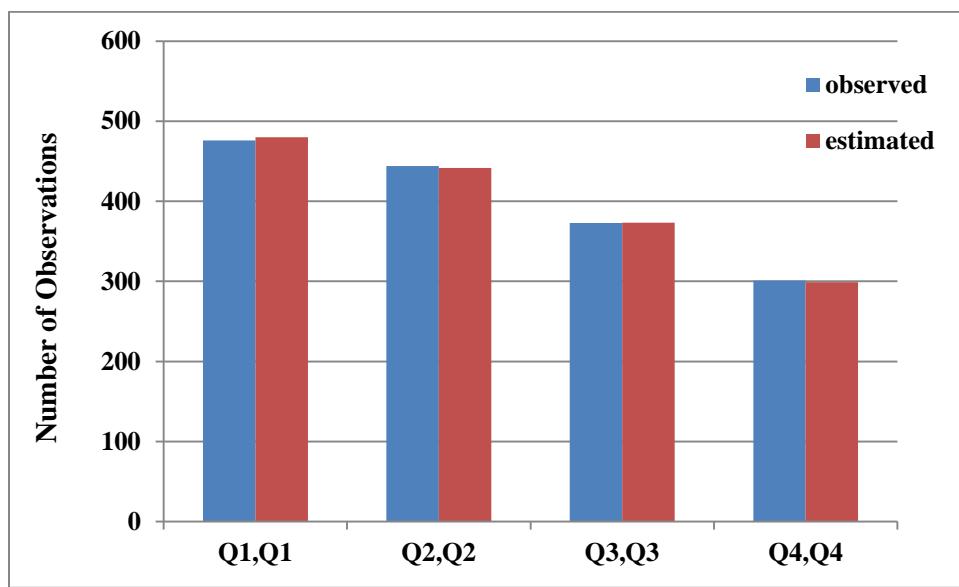


Figure 4-9: Time of Year Choice Validation for Business Purpose

	(Q1,Q1)	(Q2,Q2)	(Q3,Q3)
Household Income	<i>8.67e-06</i>	<i>4.02e-06</i>	8.09e-07
Age	<i>0.010</i>	<i>0.010</i>	<i>0.006</i>
Couple w/o Child	<i>0.312</i>	<i>0.468</i>	<i>0.42</i>
Couple w Child	<i>0.539</i>	<i>0.842</i>	<i>0.825</i>
Single	0.137	<i>0.28</i>	0.124
Employed	<i>0.308</i>	<i>0.458</i>	0.236
Unemployed	0.062	<i>0.374</i>	0.019
Constant	0.119	<i>-0.39</i>	-0.005
Pseudo R-Square	0.01		

Table 4-7: Simple Time of Year Choice for Pleasure Trip

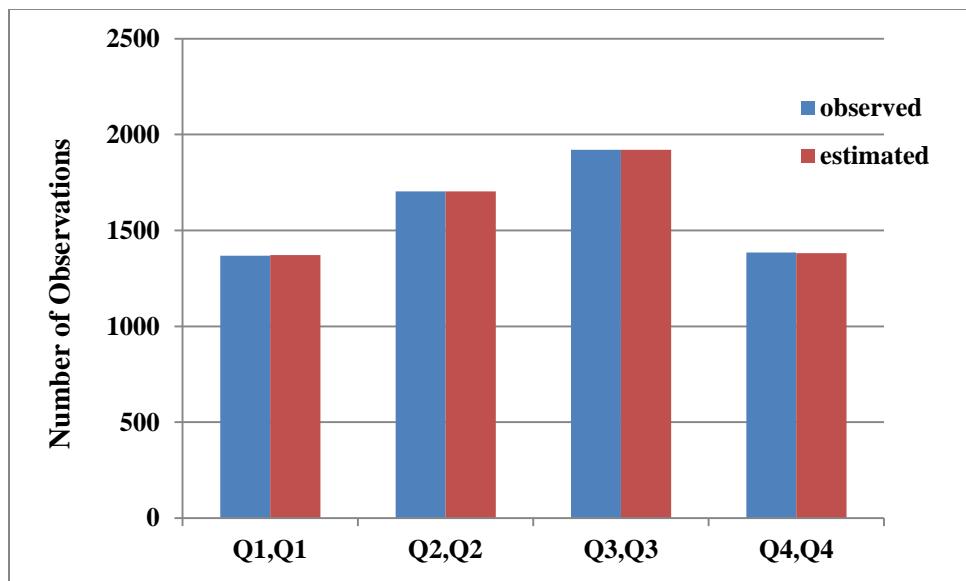


Figure 4-10: Simple Time of Year Choice Validation for Pleasure Purpose

	(Q1,Q1)	(Q2,Q2)	(Q3,Q3)	(Q4,Q4)
Mode choice logsum	0.147			
Household Income	<i>8.76e-06</i>	<i>4.38e-06</i>	7.20e-08	0.000
Employed	<i>0.376</i>	0.155	<i>0.154</i>	0.000
Male	<i>-0.256</i>	-0.066	-0.357	0.000
Couple w Child	0.223	<i>0.544</i>	<i>0.538</i>	0.000
Age	<i>0.016</i>	<i>0.01</i>	<i>0.006</i>	0.000
Constant	<i>-1.33</i>	<i>-0.8945</i>	-0.43	0.000
Pseudo R-Square	0.01			

* Coefficients in bold and italic are significant at 95% confidence level;
coefficients in italic are significant at 90% confidence level

Table 4-8: Full Time of Year Choice for Pleasure Trip

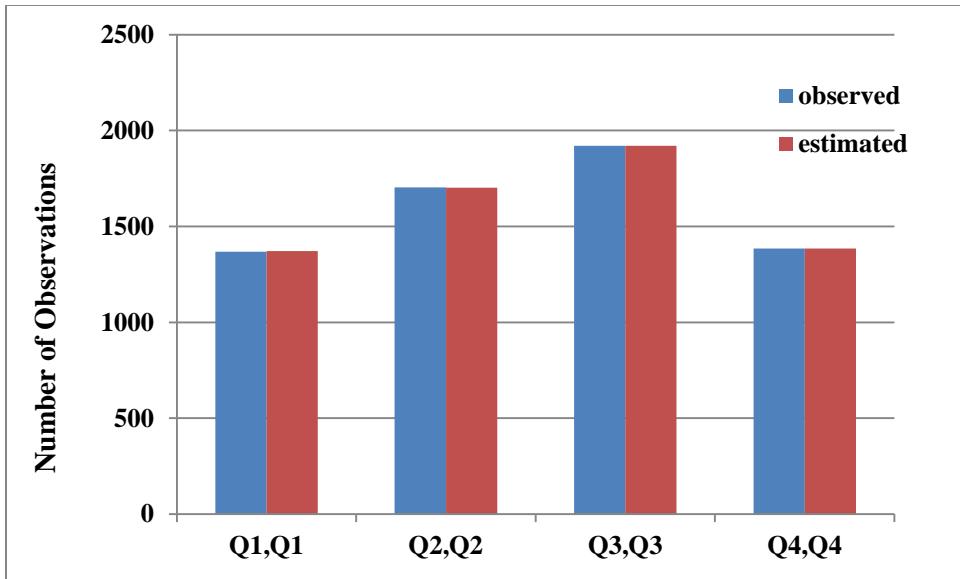


Figure 4-11: Full Time of Year Choice Validation for Pleasure Purpose

4.2.3 Tour Duration Choice Model

Different from the urban- or metropolitan-level travel demand model systems, the duration in long distance trip is measured in days away from the origin and covers the whole time period starting from the origin and ending at the origin. The tour duration is modelled as it could affect the travel distance and travel mode that people will choose when they plan the long distance travel. Hazard duration model (survival analysis) analyzing the time to the occurrence of event is employed for the tour duration model. According to the feature that the long distance travel duration is recorded as days, the discrete time survival analysis method is utilized for tour duration choice (Gokovali, 2007). In the discrete time survival analysis for tour duration, we consider each long distance tour as a subject, and all the subjects are uncensored in a one-year calendar time. The longest duration for the long distance

tour is set as 31 days. The time unit is measured as day. The survival time T is a discrete random variable with probabilities:

$$f(t) = \Pr(T = t) \quad (4.3)$$

where t represents the time interval. The discrete time survival function which describes the chance that a person will survive beyond the time period t in question without experiencing the event is given by Equation (4.4), while the failure function giving the probability of the event has occurred by duration t is given by Equation (4.5)

$$S(t) = \Pr(T \geq t) = \sum_{n=t}^{\infty} f_n \quad (4.4)$$

$$F(t) = 1 - S(t) \quad (4.5)$$

And the hazard rate, which represents the probability of an event occurs given that one has survived to that time t, is:

$$h(t) = \Pr(T = t | T \geq t) = \frac{f(t)}{S(t-1)} \quad (4.6)$$

Given the hazard rate, the discrete time survival function can also be written in Equation (4.7):

$$S(t) = (1 - h_1)(1 - h_2) \dots \dots (1 - h_{t-1})(1 - h_t) \quad (4.7)$$

The probability of the event being occurred during the time interval t is:

$$\Pr(t-1 < T \leq t) = F(t) - F(t-1) = S(t-1) - S(t) \quad (4.8)$$

Two functions including logistic regression function and complementary log-log function can be used to fit the discrete-time hazard models (Allison, 1982; Jenkins, 2000), and we adopted the logistic regression function for the hazard rate in our analysis.

$$\log\left(\frac{h_{it}}{1-h_{it}}\right) = \alpha_t + \beta \cdot X_{it} \quad (4.9)$$

where h_{it} is the hazard rate, and it is the probability of an event occurs given that one (i) has survived to the that time (t); i (1, 2,...n) refers to individual; t, taking on positive integer value, refers to the discrete time; α_t is the baseline hazard function; β is the coefficient vector of the covariates; X_{it} is the covariates or explanatory variables of individual i at time t. In the duration model, the explanatory variables or covariates are known features of the long distance tour, person and household characteristics. It is less likely that these attributes will change over time during the period of the long distance tour. Therefore, the covariates are assumed time independent. Meanwhile, multiple baseline hazard functions including log(time) and polynomial in time are tested. Little difference is shown in the model estimated coefficients (for the same covariates) and the model with-out sample validation results. In our model, the polynomial function of time was employed as the baseline hazard function. Therefore, the hazard rate function can be represented as:

$$\log\left(\frac{h_{it}}{1-h_{it}}\right) = \gamma t + \theta t^2 + \beta \cdot X_{it} \quad (4.10)$$

$$h_{it} = \frac{1}{1+\exp(-\gamma t - \theta t^2 - \beta \cdot X_{it})} \quad (4.11)$$

where γ , θ , and β are coefficients which need to be estimated.

Table 4-9 presents the tour duration model estimation results for long distance business and personal business activities. The duration of long distance pleasure activities is estimated based on the observed distribution of the pleasure duration using Monte Carlo simulation, as the duration model for pleasure trips has a very small Pseudo R2 and the validation results shows a much different pattern between the estimated distribution of the duration and the observed one. The observed duration distribution of the long distance pleasure trip is shown in Figure 4-12, and it

shows that most people usually have 2 days for their entire long distance pleasure activities.

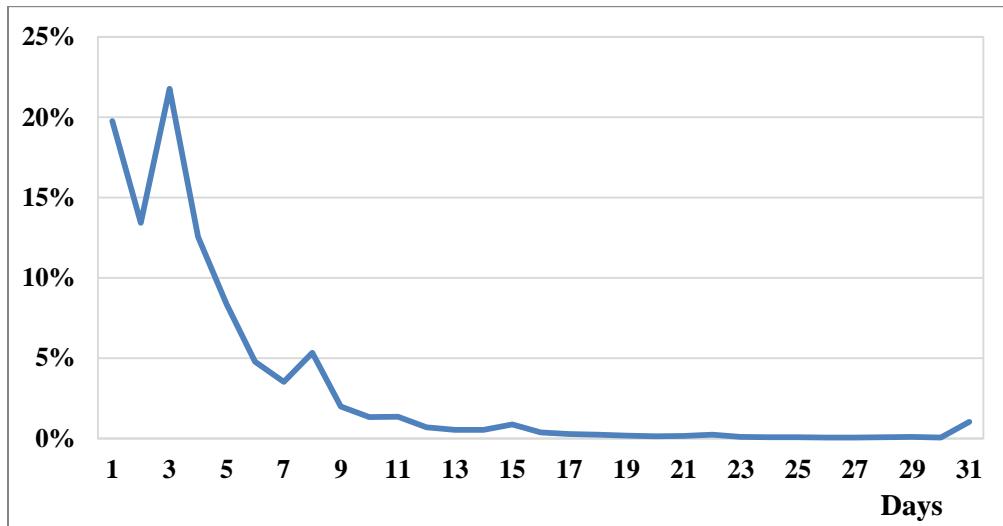


Figure 4-12: Observed Duration Distribution of Long Distance Pleasure Activities

According to the duration model estimation results, people travelling with tour primary destination in MSA zone have lower hazard rate for both business and personal business purposes. People in couple-family household with children have higher hazard rate than people in single family household and non-family household if they make long distance business trips. Meanwhile, people in couple-family household with children and non-family household have higher hazard rate than people in single family household when they travel long distance for personal business purpose. Compared to low income group, the hazard rates of medium and high income groups are lower, with the hazard rate of high income even lower than that of medium income group. Also, the hazard rate increases with person age, and decreases if the person is unemployed. People travelling in the second, the third, and the fourth quarter have lower hazard rate than people travelling in the first quarter,

and the largest extent of decreasing rate is the third quarter. The baseline hazard rates for business and personal business are shown in Figure 4-13.

	Business	Personal Business
Destination in MSA	-0.182	-0.096
Couple w Children	0.086	0.149
Single Family	-0.075	-0.148
Non-Family Household	-0.169	0.204
Household Size	0.028	0.030
Medium Income	-0.045	-0.082
High Income	-0.203	-0.279
Unemployed	-0.132	-0.194
Student	-0.052	-0.669
Tour Departure in Quarter 2	-0.196	-0.062
Tour Departure in Quarter 3	-0.326	-0.128
Tour Departure in Quarter 4	-0.141	-0.084
Age	0.006	0.003
Time interval (<i>t</i>)	-0.084	-0.188
Squared time interval (<i>t</i> ²)	0.002	0.005
constant	-0.854	-0.096
Pseudo R2	0.02	0.05

* Coefficients in bold and italic are significant at 95% confidence level;
coefficients in italic are significant at 90% confidence level

Table 4-9: Tour duration choice model estimation results

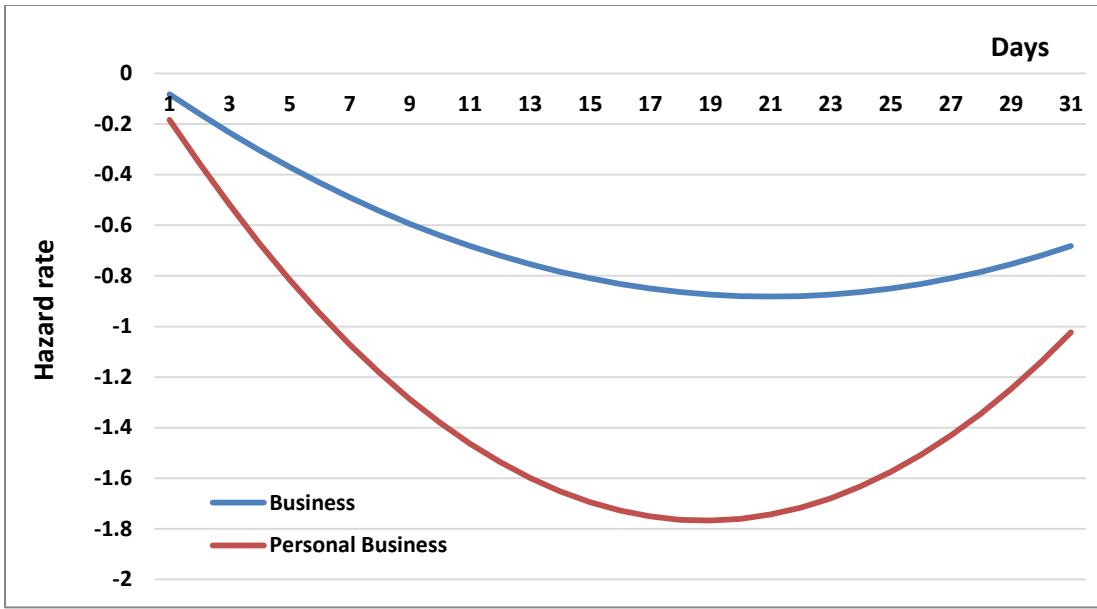


Figure 4-13: Baseline Hazard Rate for Business and Personal Business Duration Model

Based on the model estimation results, we estimated the duration of each long distance tour based on Equation (4.7), (4.8) and (4.11) using Monte Carlo simulation method. The validation results between the observed duration distribution and estimated duration distribution are shown in Figure 4-14 and Figure 4-15 for business and personal business purposes separately. It is observed that except the duration of 2 days the estimated duration distributions follow the same pattern as the observed distribution for both long distance purposes.

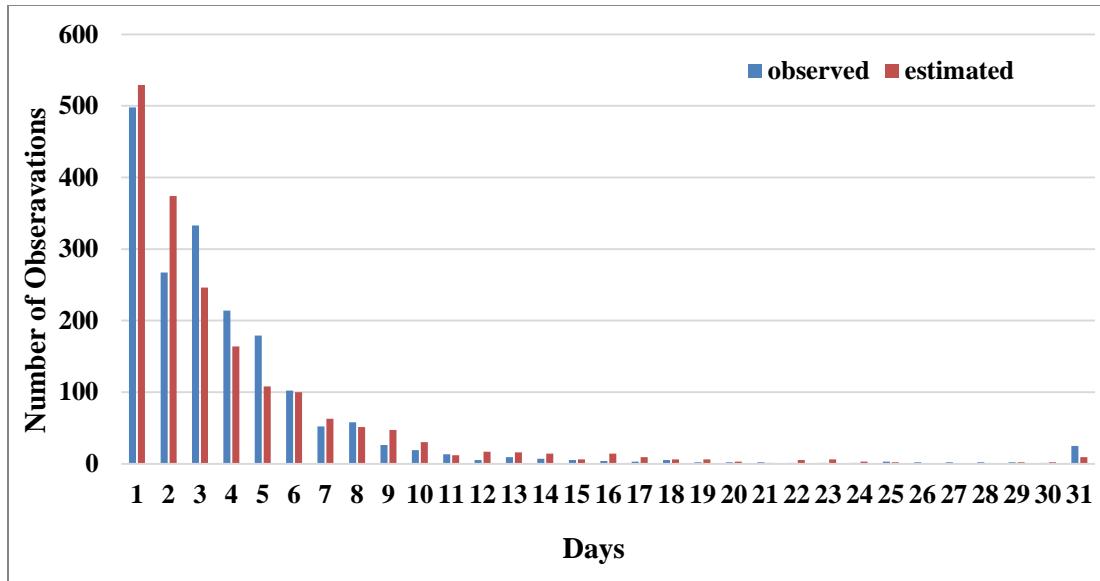


Figure 4-14: Validation results for Business Duration Model

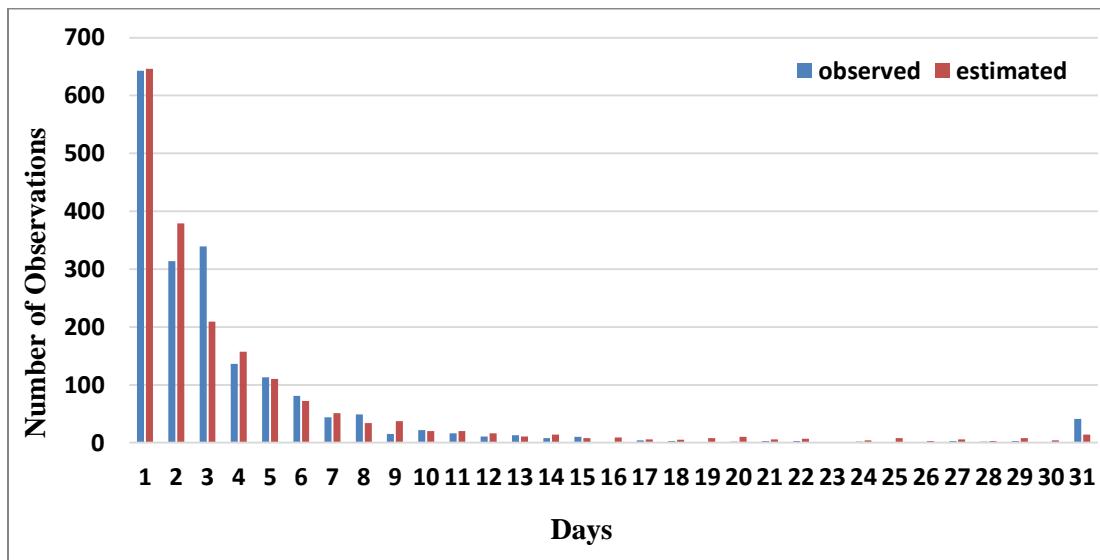


Figure 4-15: Validation results for Personal Business Duration Model

4.2.4 Travel Party Size Choice Model

The travel party size choice is modeled for each long distance tour, and it determines how many persons participating in the tour. And it is assumed that no one will get on or get off the tour during the long distance travel. The model is

multinomial logit model, and each people will have a choice set of four alternatives (travelling alone, travelling in 2 persons, in 3 persons, and in 4 and more persons) for all the three long distance activities (business, personal business and pleasure). People travelling alone is set as the base alternative for all the three models. The explanatory variables mainly include the person and the household characteristics. According to the tour-level model structure, the tour destination is determined and known before people decide the travel party size for business and personal business travel. Therefore, the zonal attributes can be utilized in the travel party size choice model for long distance business and personal business tour. The estimation results are shown in tables from Table 4-10 to Table 4-12. The coefficients in both bold and italic font are significant at 95% confidence level, while those only in italic font are significant at 90% confidence level. The results imply that when people take long distance business travel, they prefer travelling alone instead of travelling in a party if the destination is a metropolitan statistical area. And compared to people with high income, people with low and medium income tend to travel in a party and low-income level people prefer more persons on the trip. If the traveler is female, she tends to travel with companies during her long distance business travel. Aged people are more likely to travel in a two- and three-person party on their business trips. When people travel for long distance personal business and pleasure activities, they are more likely to travel alone if they live in a single family. Contrariwise, they tend to travel in a party with more persons if they live in a family with spouse and children. The same pattern as in the business tour can be observed for female, low-income and medium-income people when they take the personal business and pleasure trips. The with-out sample

validation results (Figure 4-16, Figure 4-17, and Figure 4-18) illustrate that the travel party size choice models of all the three purposes can estimate the number of persons during the entire tour with small errors.

	2 Persons	3 Persons	4 and 4+ Persons
Destination in MSA	-0.187	-0.327	-0.248
Single Family	-1.207	-0.847	-0.187
Couple with Children	-0.474	0.050	0.027
Household Size	-0.003	-0.051	0.261
Low Income level	0.660	1.063	1.161
Medium Income level	0.347	0.632	0.606
Age	0.020	0.007	-0.007
Female	0.908	0.984	1.046
Constant	-1.221	-2.052	-2.288
Pseudo R2		0.055	

* Coefficients in bold and italic are significant at 95% confidence level;
coefficients in italic are significant at 90% confidence level

Table 4-10: Travel Party Size Choice Model Estimation for Business Tour

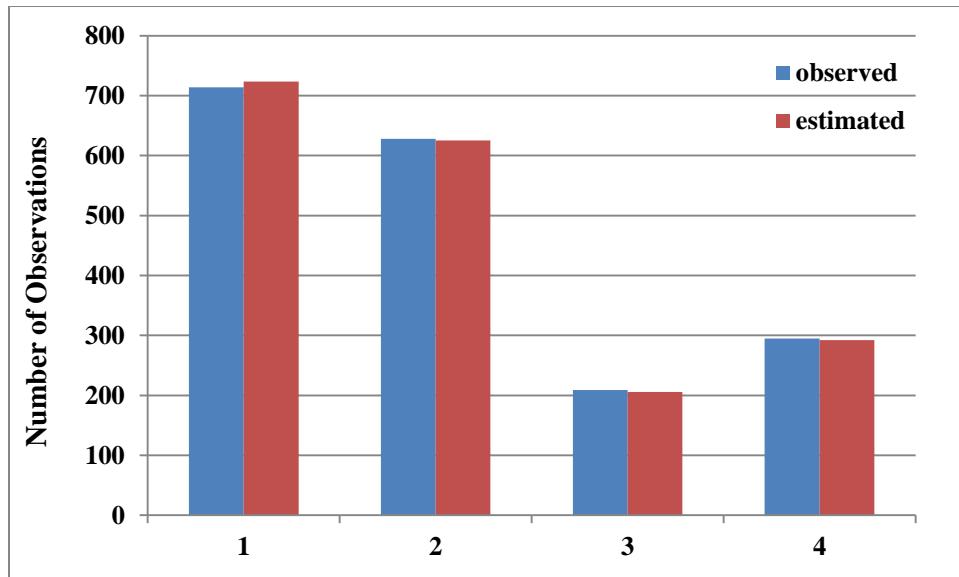


Figure 4-16: Travel Party Size Choice Validation for Business Purpose

	2 Persons	3 Persons	4 and 4+ Persons
Destination in MSA	-0.086	0.0004	-0.238
Single Family	-2.095	-1.385	-0.911

HH with Children	0.010	0.619	0.842
Household Size	-0.174	-0.017	0.270
Low Income level	0.382	0.699	0.771
Medium Income level	0.268	0.335	0.394
Age	0.022	0.012	0.008
Female	0.283	0.384	0.402
Constant	0.315	-0.988	-1.519
Pseudo R2		0.075	

* Coefficients in bold and italic are significant at 95% confidence level;
coefficients in italic are significant at 90% confidence level

Table 4-11: Travel Party Size Choice Model Estimation for Personal Business Tour

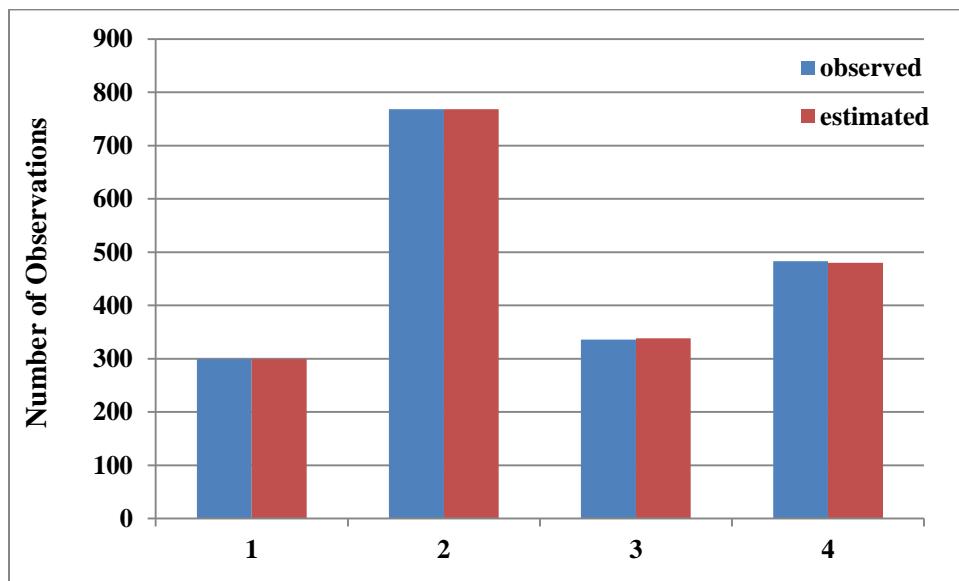


Figure 4-17: Travel Party Size Choice Validation for Personal Business

	2 Persons	3 Persons	4 and 4+ Persons
Single Family	-2.505	-1.754	-0.921
HH with Children	0.088	1.588	1.370
Household Size	-0.265	-0.184	0.395
Low Income level	0.119	0.199	0.467
Medium Income level	0.168	0.256	0.394
Age	0.023	0.013	0.016
Female	0.023	0.098	0.118
Constant	1.058	-0.196	-1.682
Pseudo R2		0.133	

* Coefficients in bold and italic are significant at 95% confidence level;
coefficients in italic are significant at 90% confidence level

Table 4-12: Travel Party Size Choice Model Estimation for Pleasure Tour

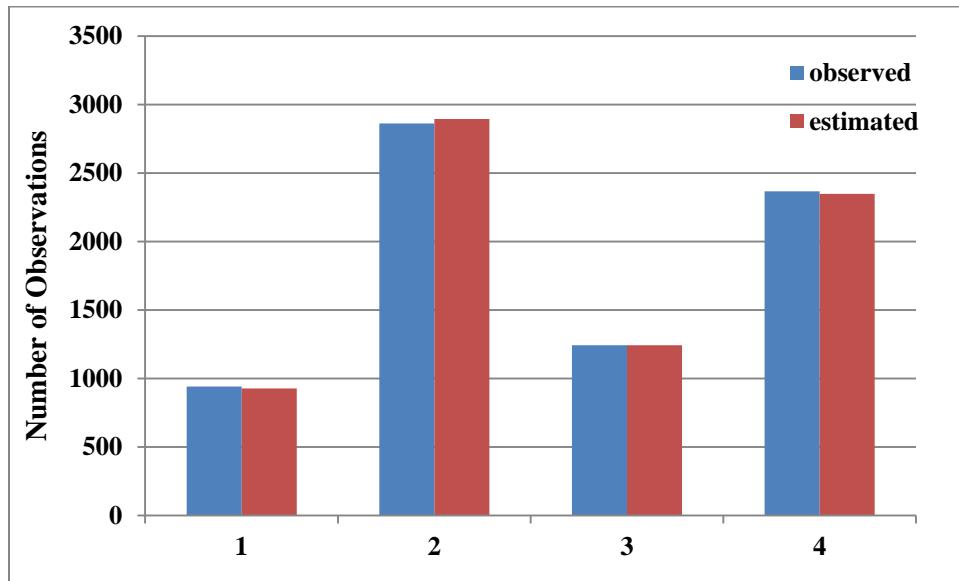


Figure 4-18: Travel Party Size Choice Validation for Pleasure Purpose

4.2.5 Tour Destination Choice Model

The destination choice determines the location of the long distance tour's primary destination. It works at the zonal level and each person will be assigned a TAZ as his/her primary destination according to the multinomial logit destination choice model. In the 1995 ATS sample data, there are a total of 208 TAZs which means that each person faces a universe choice set of 208 alternatives. In order to reduce the estimation time and complexity, the method of simple random sampling (SRS) is implemented due to the independency of irrelevant alternatives (IIA) property of multinomial logit model (Nerella & Bhat, 2004; Lemp & Kockelman, 2012). Consequently, each person will have a destination choice sub-set of 10 alternatives among which one is the person's chosen zone and nine will be randomly

selected from the rest of 207 zones. Due to the fact that people tend to take their long distance pleasure and business trips in places that are large cities, tourist attractions or vacation locations, we added several dummy variables in the long distance business and pleasure destination choice models to control people's preference for certain places. The zonal attractiveness variables in the destination choice models are mainly the zonal total employment, the number of households, and the dummy variable indicating a metropolitan statistical area. As we can see in Figure 4-4, destination is firstly determined for long distance business and personal business travel and the travel time period is unknown. Therefore, in the destination choice model for business and personal business, the mode choice logsum is calculated with the average travel time and travel cost of each travel mode across four seasons.

Table 4-13 presents the estimation results for the destination choice under all the three long distance activities. All the variables in the table are significant at 95% confidence level. The coefficients of the mode choice logsum in all the three models are positive, implying that people tend to make a long distance travel to a place with large accessibility. And the larger value in business model indicates that people are more sensitive to the accessibility when it is a long distance business travel. The variable of distance to destination shows that people are not likely to take a long distance travel to a place farther from home. The variables of distance squared and cubed are incorporated in the model to allow for a nonlinear effect of the distance. Also no matter what kind of purpose they travel for, they prefer travelling to a location with a lot of job positions which implies plenty of resources. If a zone is a metropolitan statistical area and has a large number of households, it is not likely

people would choose the zone for their long distance travel destination. As expected, people prefer choosing the large cities (Las Vegas) or tourist attractions (Florida) for their long distance business and pleasure travel.

	Business	Personal Business	Pleasure
Mode choice logsum	0.999	0.609	0.424
Distance to destination	-0.003	-0.004	-0.003
Squared Distance to destination (1000mi)	0.001	0.001	0.001
Cubed Distance to destination (10^5 mi)	-0.00001		
Destination in MSA	-0.695	-0.965	-1.256
No. of Employment	0.002	0.001	0.002
No. of Households	-0.003	-0.001	-0.002
Destination in Las Vegas	-2.179	--	-1.822
Destination in Florida	--	--	1.974

* Coefficients in bold and italic are significant at 95% confidence level;

Table 4-13: Primary Destination Choice Model Estimation at Tour Level

Figure 4-19, Figure 4-20, and Figure 4-21 show the validation results for only 206 zones that exist in the dataset for destination model estimations for the three long distance travel primary purposes. The figures depict that most of the destinations can be estimated with small difference from the observed except for a few ones with large errors due to the relatively small sample size. Among all the three purposes, the destination model for pleasure presents the overall largest error.

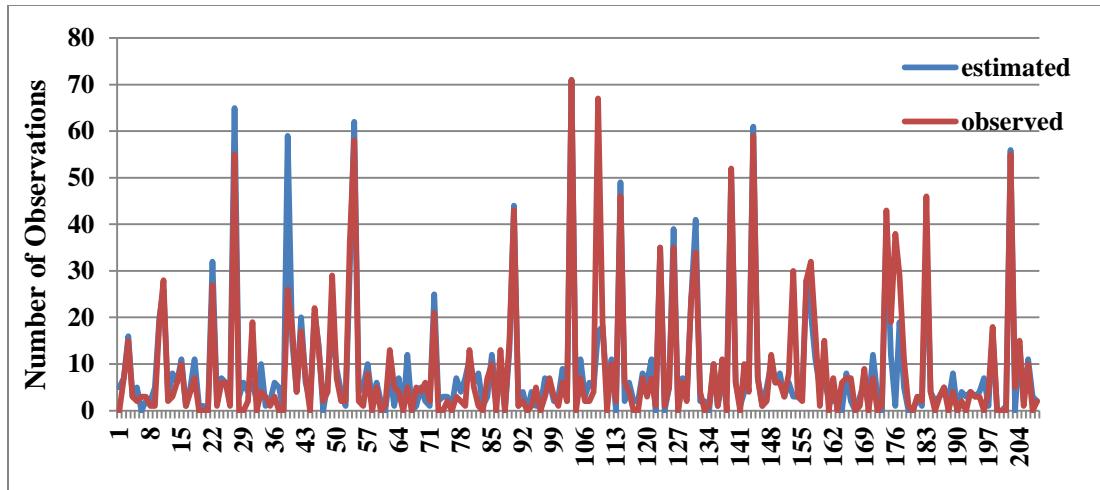


Figure 4-19: Destination Choice Validation for Business Purpose

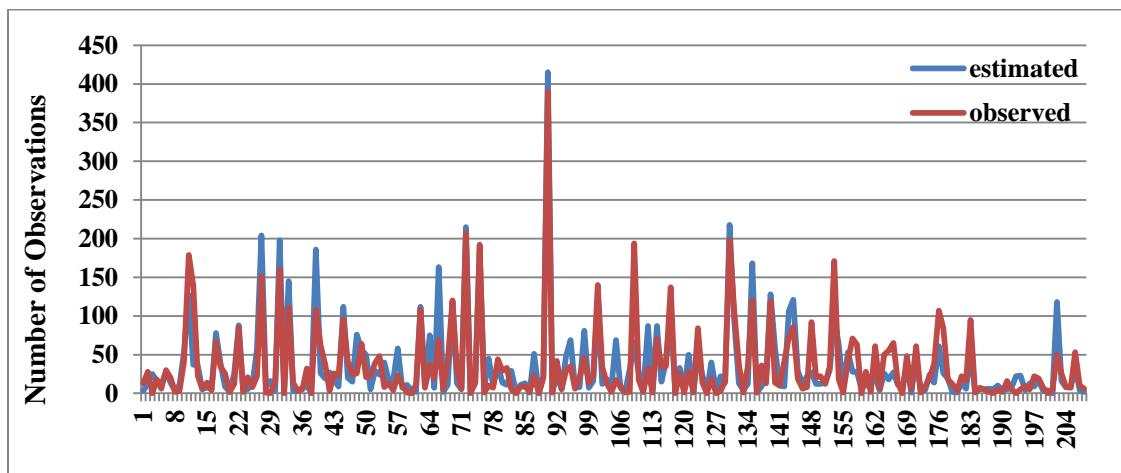


Figure 4-20: Destination Choice Validation for Pleasure Purpose

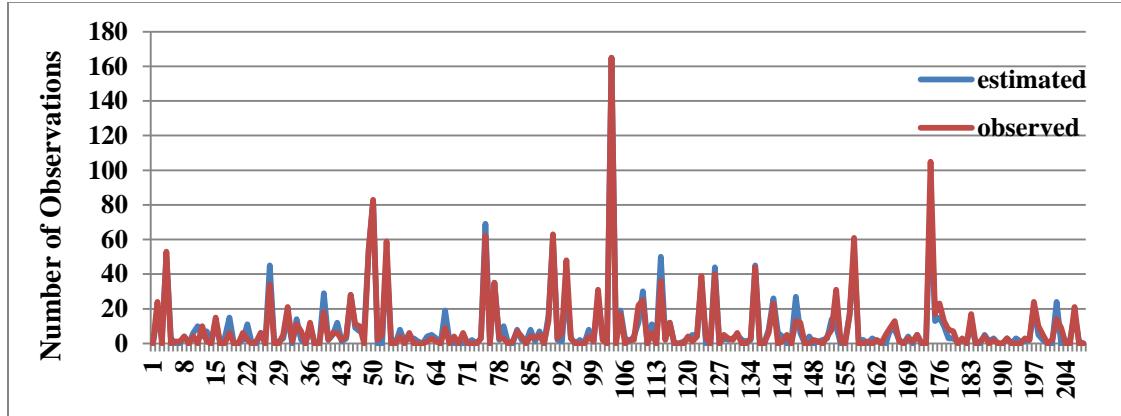


Figure 4-21: Destination Choice Validation for Personal Business Purpose

4.3 Stop Level Structure

After people have made decisions about their travel to the main destination, they will make plans for their trips on the way to and from the destination based on the remained time. It is assumed that people have the same logic to determine their stops or trips during the tour legs regardless of the main activity types. Consequently, the same model structure at the stop level will be applied to all the three tour-level activity types (business, pleasure, personal business) (Figure 4-22). The stop level structure generates the information of the intermediate stops people would make during their inbound/outbound legs of the long distance tour. A stop during the tour is defined as one people make for certain purpose like business, personal business or pleasure. The stops for rest or transfer in the same travel mode or across multiple travel modes are not the ones we are going to analyze, and are excluded from the data set. At the stop level, the information about each long distance tour such as the tour duration, travel mode, travel party size, and tour origin and destination are already known.

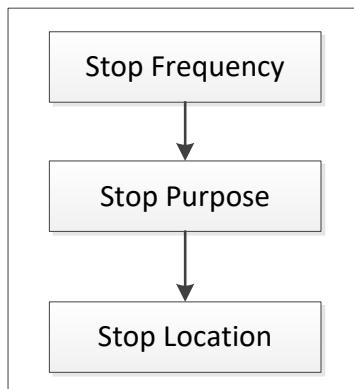


Figure 4-22: Stop Level Procedure and Model Components

4.3.1 Stop Frequency Choice Model

The stop frequency model at the higher level determines the number of intermediate stops people will have on the way from/to the tour destination. In each direction, a maximum number of 4 stops can be made which results in a maximum of 5 trips on each tour leg. The stop frequency choice model for each half tour leg is developed using multinomial logit model, and each person faces a choice set of 5 alternatives (0, 1, 2, 3, 4) for each tour leg. That people make zero stop is set as the base alternative for both models. The models mainly utilize the long distance tour characteristics as the explanatory variables such as tour duration, tour mode, the activity type of the long distance tour, time of year, and distance between tour origin and destination. As expected, people would like to make stops on inbound and outbound trips if the travel distance between origin and primary destination is large or people have plenty of time for their long distance travel. When the long distance travel is for pleasure, people tend to make stops on both trips (inbound and outbound trips), more likely one stop. People using car for their long distance travel mode prefer stopping one or two times when they head home (inbound trip). While they are on the way to the tour primary destination they are more likely to make stops especially 2 stops for certain purposes. The with-out sample validation results in

Figure 4-23 and

Figure 4-24 show that the stop frequency models perform well in estimating the number of stops during either inbound or outbound leg of the tour.

	1	2	3	4
OD Distance	0.001	0.0005	0.0002	0.001

Tour Duration	0.025	0.017	0.015	0.031
Travel Party	0.034	0.0002	-0.043	-0.025
Car mode	1.044	2.309	-3.090	-1.483
Business Tour	0.640	0.201	0.365	0.599
Pleasure Tour	1.014	0.177	0.185	0.570
Travel in Quarter 2	-0.659	0.060	1.420	0.393
Travel in Quarter 3	-0.237	0.110	1.298	0.463
Travel in Quarter 4	-0.011	-0.398	1.582	0.268
Constant	-9.078	-6.444	-3.344	-6.100
Pseudo R-Square	0.16			

Table 4-14: Stop frequency model estimation for tour inbound leg

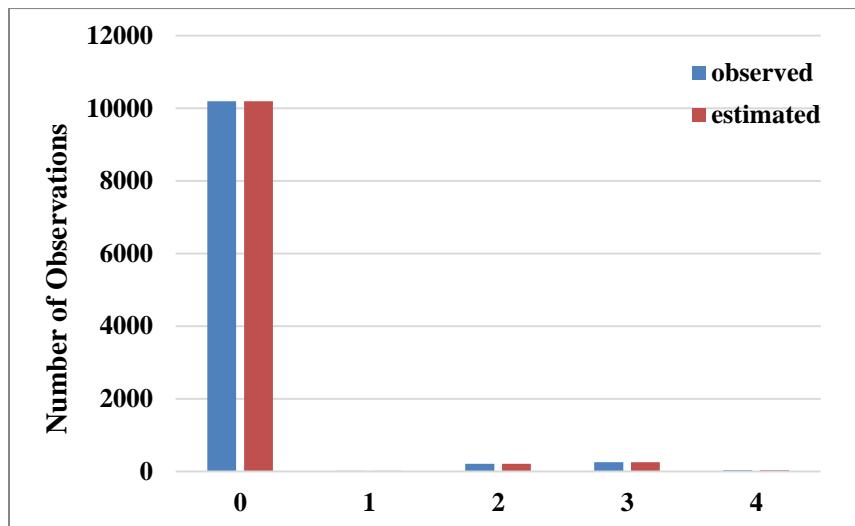


Figure 4-23: Inbound Stop Frequency Model Validation

	1	2	3	4
OD Distance	0.001	0.001	0.001	0.001
Tour Duration	0.023	0.017	0.020	0.022
Travel Party	-0.003	-0.004	0.013	0.005
Car mode	0.975	3.283	0.185	0.983
Business Tour	-0.527	0.208	0.704	0.298
Pleasure Tour	0.727	0.434	0.611	0.628
Travel in Quarter 2	-0.368	0.133	0.004	0.176
Travel in Quarter 3	0.045	-0.001	0.094	0.563
Travel in Quarter 4	-1.162	-0.096	-0.405	-0.691
Constant	-8.481	-7.512	-6.326	-8.052
Pseudo R-Square	0.05			

Table 4-15: Stop frequency model estimation for tour outbound leg

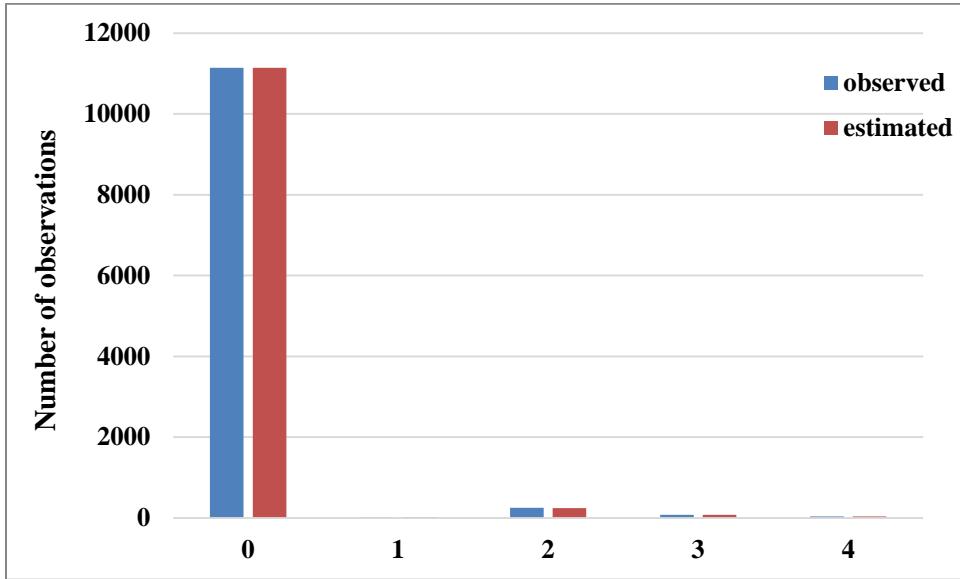


Figure 4-24: Outbound Stop Frequency Model Validation

4.3.2 Stop Purpose Choice Model

Once the number of stops a person will make during the long distance tour is obtained, the purpose of each stop will be determined (the middle-level model component in Figure 4-22) through the stop purpose choice model. The stop purpose category follows the same tour-level activity types that include business, personal business and pleasure. The model is developed for each half tour using multinomial logit model, and the pleasure purpose will be set as the base alternative. Both obtained stop-level and tour-level characteristics can be used as the explanatory variables in the stop purpose choice model for each tour leg, such as the sequence of the stop, the long distance primary activity type, travel party size, and the tour travel mode.

The estimation results for the stop purpose choice models (Table 4-16 and Table 4-17) indicate that it is less likely that people will make a business stop for their

second, third, or fourth stop if they are on the way to the primary destination. Meanwhile, people don't like to arrange a stop for business purpose either on inbound trip or on outbound trip when their primary activity of the long distance travel is pleasure or personal business. However, a stop is more likely to be a personal business stop on either half tour if people's primary activity is personal business. People travelling by car or air tend to make a business stop when they are on the way back home.

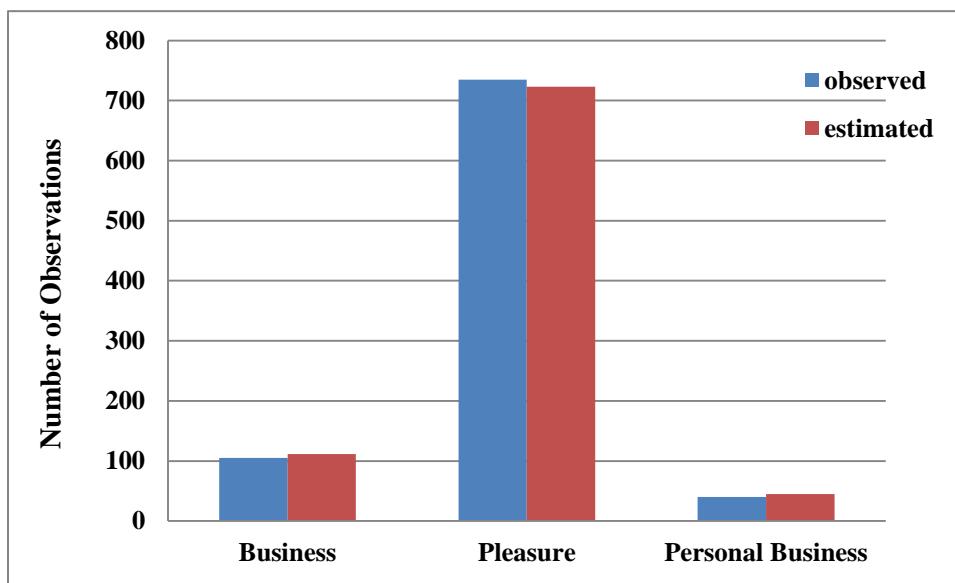


Figure 4-25 and Figure 4-26 illustrate the good performance of the stop purpose models in estimating the purposes of each stop during each half leg of the long distance tour.

	Business	PB
Second Stop	-0.462	-0.061
Third Stop	-0.395	-0.352
Fourth Stop	-0.589	0.048
Pleasure Tour	-3.904	1.259
Personal Business Tour	-2.697	3.662
Travel Party	-0.015	-0.123
Car mode	-0.067	-0.547
Air mode	1.227	0.007
Constant	0.666	-3.939

Pseudo R-Square	0.32
-----------------	------

Table 4-16: Purpose estimations for outbound stops

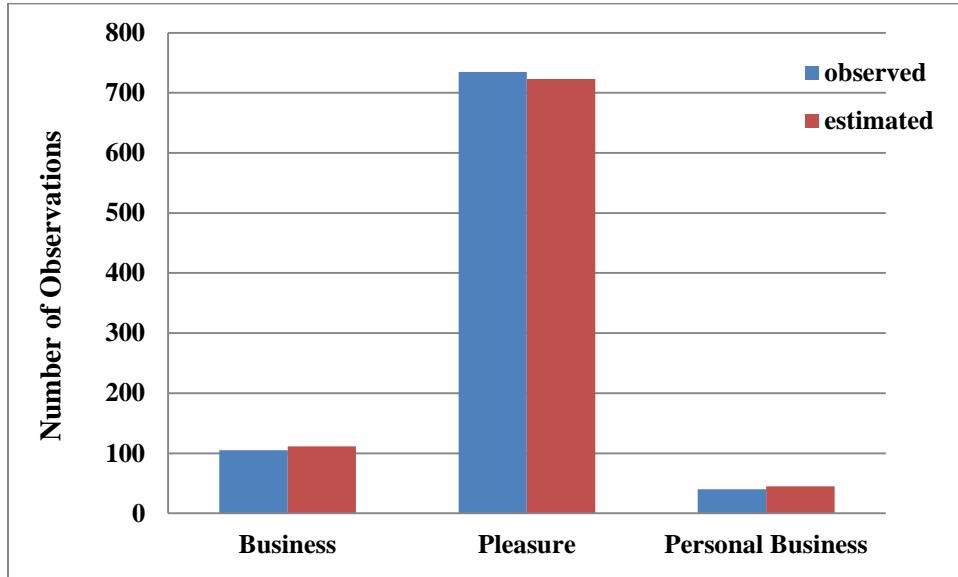


Figure 4-25: Outbound Stop Purpose Model Validation

	Business	PB
Second Stop	-0.061	-0.083
Third Stop	-0.019	-0.111
Fourth Stop	0.561	1.719
Pleasure Tour	-1.735	1.193
Personal Business Tour	-1.319	2.972
Travel Party	-0.01	0.067
Car mode	0.824	0.645
Air mode	0.872	-0.273
Constant	-1.920	-5.953
Pseudo R-Square	0.1	

Table 4-17: Purpose estimations for inbound stops

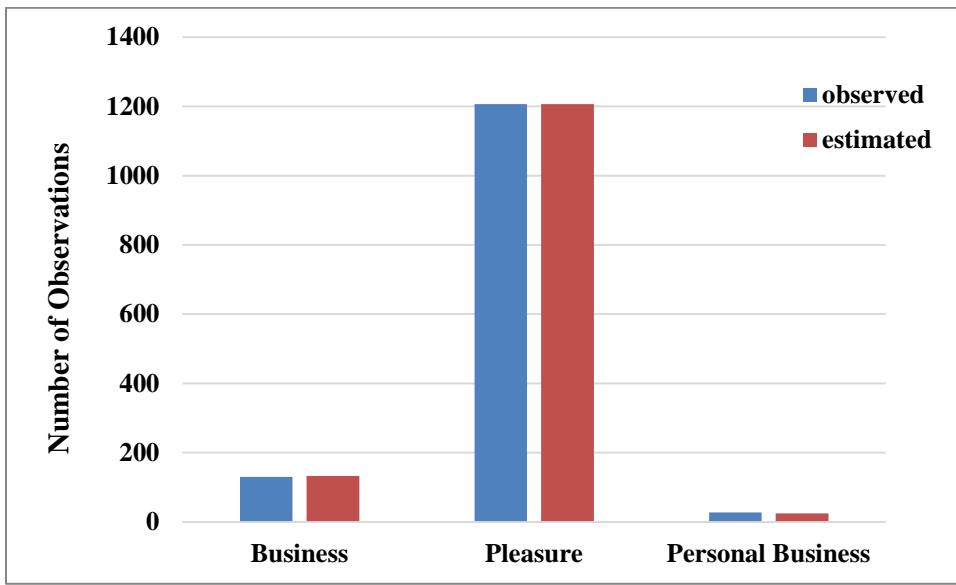


Figure 4-26: Inbound Stop Purpose Model Validation

4.3.3 Stop Location Choice Model

At the low tier of the stop-level structure, the location for each stop will be estimated with the similar method employed in the primary destination choice at the tour level. And before the stop location, we know the number of stops and the sequence of the stops during each half tour, the stop purpose and tour origin and destination. Since we assume that people only take one of the three modes (air, car, train) and no transfer among different modes for the entire tour, the travel mode for each trip on each half leg will remain the same as the one estimated at the tour level. In the stop location choice, the distance between stop origin and stop location should be larger than 50 miles. The short distance travel based on the stop origin will not be incorporated and modelled. The same methodology as the one used in tour destination choice will be employed. However, different from the tour-level primary destination

choice, the impedance of the travel to an intermediated stop in the stop location choice model should measure the additional impedance between the tour origin or stop origin and the tour primary destination if it is an outbound trip (Bowman & Bradley, 2006). And the main variables in stop location choice model are the out-of-direction or detour generalized travel cost and detour travel distance. For example, the level of service (LOS) variables for the first stop on the way to the tour primary destination are based on the additional impedance between the tour origin and the tour destination (Figure 4-27), and the LOS for the following stop is based on the additional impedance between the previous stop and the tour destination (Figure 4-28). At the meantime, the tour origin becomes the stop origin for the stop in the situation of Figure 4-27, and stop i will be called the stop origin of stop j in the Figure 4-28 situation. The same method works with the stops of the inbound direction but in an opposite way, as the anchor point is the tour origin instead of tour primary destination.

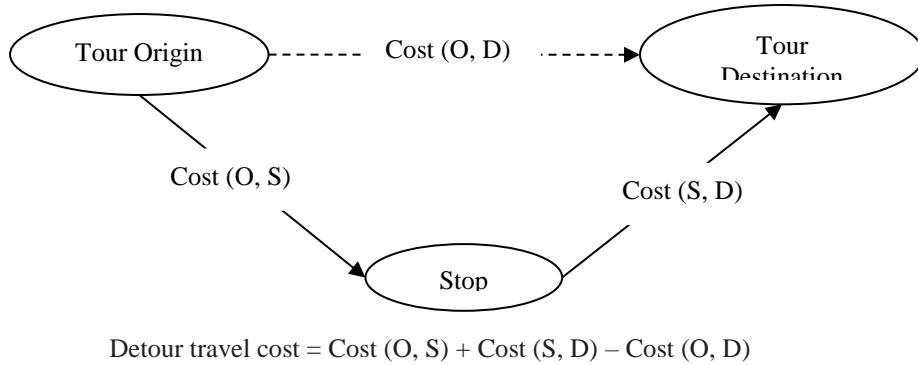
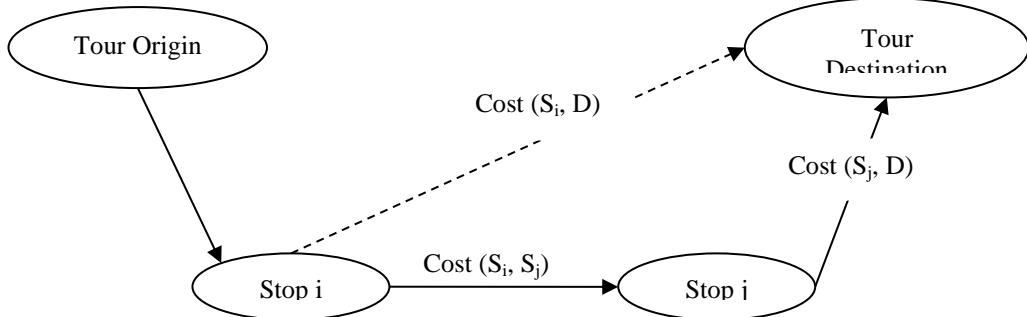


Figure 4-27: LOS estimation for the first stop during outbound tour leg



$$\text{Detour Travel Cost for Stop } j = \text{Cost}(S_i, S_j) + \text{Cost}(S_j, D) - \text{Cost}(S_i, D)$$

Figure 4-28: LOS estimation for the jth stop during outbound tour leg

The detour generalized travel cost combines the detour travel cost and travel time components according to the time of value obtained from the tour level mode choice model. Table 4-18 and Table 4-19 provide the estimation results for all the coefficients in the stop location choice models for both inbound and outbound trips. The results imply that it is less likely that people will make a stop at a location with large detour travel distance on the way to or from the primary destination. Furthermore, people would like to stop at a non-MSA location with a large number of employments, no matter whether they are on the inbound or outbound trip.

When the distance between the stop origin and tour destination is less than 150 miles and people will not be likely to take a stop at a place with higher detour generalized travel cost when they head to the tour destination. If the distance between the stop origin and tour destination is larger than 150 miles and less than 550 miles, people are still not likely to stop at a location with high detour generalized travel cost, but less sensitive compared to the situation of the distance between the stop origin and destination less than 150 miles. When the distance between the stop origin and destination is larger than 550 miles, the detour generalized travel cost has an

insignificant impact on travelers' stop location choice. The positive coefficients could be caused by the small sample size of the outbound stops for location choice.

On the way back from the tour destination, people are less likely to stop at a location where the detour generalized travel cost is high when the distance between the stop origin and the tour origin is shorter than 550 miles. However, it seems that people by car would not care about the detour generalized travel cost when the distance between the stop origin and tour origin is larger than 550 miles, and this impact is insignificant.

Figure 4-29 and Figure 4-30 shows the with-out sample validation results for the two models. From the figures, we can observe that the size of the validation sample is very small for each zone, with the largest number of trips less than 30. Based on the very few validation data records, we can say that both models can estimate the stop location well for a number of zones. Furthermore, the outbound location choice model shows a better performance than the inbound location model, with a smaller number of trips under/over-estimated.

Explanatory Variables	Coefficients
Detour Travel Distance	-0.00811
Detour GTC ($D_{So-D} < 150$ miles)	-0.10292
Detour GTC ($150 \leq D_{So-D} < 550$)	-0.00016
Detour GTC ($D_{So-D} \geq 550$)	0.00158
Detour GTC ($D_{So-D} \geq 550$, Car mode)	-0.00152
No. of Employment	0.00170
No. of Households	-0.00158
The zone is not MSA	1.82495
R-Square	0.09

* GTC: Generalized Travel Cost;

D_{So-D} : Distance from stop origin- tour destination

Coefficients in bold and italic are significant
at 95% confidence level;

Coefficients in italic are significant at 90% confidence level

Table 4-18: Stop location model estimation for tour outbound leg

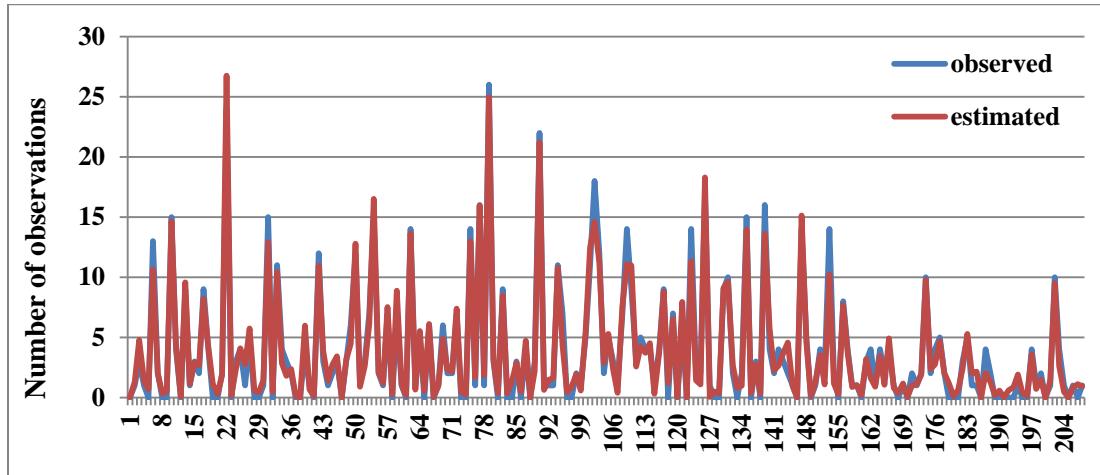


Figure 4-29: Outbound Stop Location Choice Validation

Explanatory Variables	Coefficients
Detour Travel Distance	-0.0081
Detour GTC ($D_{So-D} < 150$ miles)	-0.0361
Detour GTC ($150 \leq D_{So-D} < 550$)	-0.0007
Detour GTC ($D_{So-D} \geq 550$, Car mode)	0.0010
Detour GTC ($D_{So-D} \geq 550$)	-0.0001
No. of Employment	0.0006
No. of Households	-0.0002
The zone is not MSA	1.7805
R-Square	0.04

* GTC: Generalized Travel Cost;

D_{So-o} : Distance from stop origin- tour origin

Coefficients in bold and italic are significant at 95% confidence level;

Coefficients in italic are significant at 90% confidence level

Table 4-19: Stop location model estimation for tour inbound leg

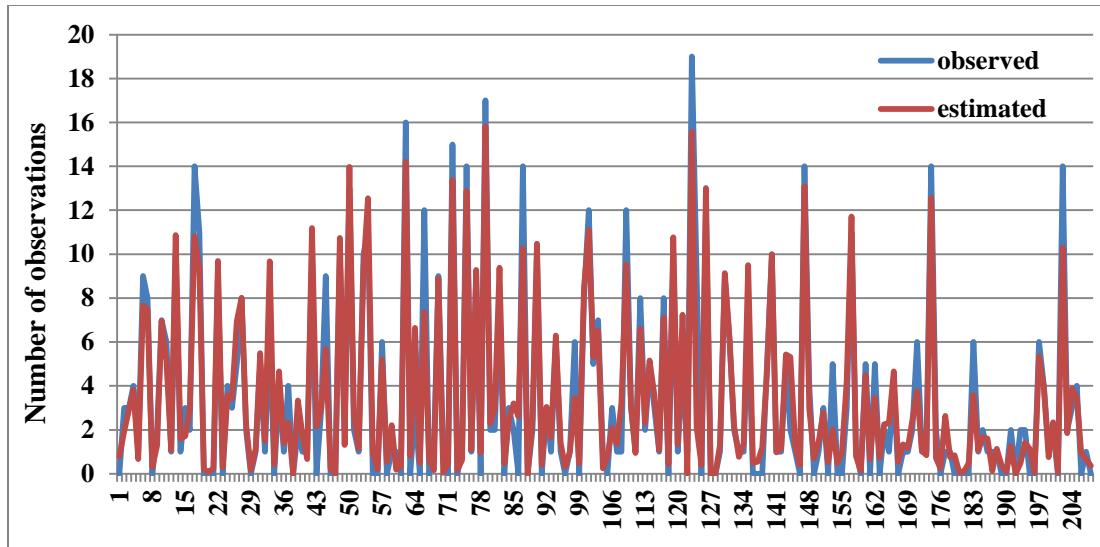


Figure 4-30: Inbound Stop location Choice Validation

The developed national passenger travel demand model has a series of analytic tools to ensure travel behavior realism and model sensitivity. All the model components in the model system employed the discrete choice modelling, specifically, multinomial logit model, except the yearly long distance activity model and tour duration model component. Most of the model components adopted discrete choice modelling methodology, because these models have limited discrete dependent variables which represent the person's decisions or choices when planning for his/her long distance travel. Another reason we chose the multinomial logit model over the other methodology is because using the multinomial logit model can give us better estimation results when we conducted the with-out sample validation. For example, we tried the negative binomial model and the Poisson model to estimate the stop frequency during each leg of the tour (stop frequency model). However, the with-out sample validation results of these models showed poor performance in matching the observed data. On the contrary, the with-out sample validation results of the stop

frequency model using multinomial logit modelling methodology shows good performance in matching the observed data (Figure 4-23 and Figure 4-24). We have the same reason for choosing multinomial logit model over ordered logit model in estimating the travel party size.

The methodology we used to estimate the number of long distance activities during the course of one year is Multiple Classification Analysis which is usually mostly used in trip generation in traditional four-step travel demand model. We chose this method over other methodologies (e.g. multinomial logit model and count models) because some of the explanatory variables (e.g. income and travel cost) that we took into account in the multinomial logit model or count models didn't give us the right sign and the model performance was poor in terms of validation. Therefore, we chose a less advanced methodology, MCA method, to generate the long distance trip rate. The discrete time survival analysis or discrete hazard duration model was chosen because we would like to model the time that people spend for their entire long distance tour, and the time is measured as days which is discrete. Most importantly, the discrete time survival analysis method shows good performance in terms of matching the observed data in with-out sample validation.

4.4 National Travel Demand Model Flow and Key Assumptions

Based on the model system, we can tell that the model assumes that people usually make plans for their long distance activities for one year months or even one year in advance. People will first make a plan of how many long distance activities they will have during one year at the end of the last year. For example, a person who

lives in Washington D.C area decides to have three long distance activities during the year, one for business, one for pleasure, and one for personal business. Then he/she has to make plans for the three long distance activities. With regard to the business travel, according to his/her company's schedule and his/her own work schedule, he/she decides to go to California for one week in the first quarter, and go there on his/her own by air. As one week he/she schedules is tight for him/her and he/she has to go back to work, he/she doesn't want to make any stops on the way from/to California. Meanwhile, he/she makes plans for his/her long distance pleasure activity. After he/she checks his/her work schedule and accrued leave from work, he/she finds out that he/she could take two weeks in July or August (the third quarter) to have a long distance pleasure activity. He/she decides to make the long distance pleasure travel with his/her spouse. As they have plenty of time (two weeks), they decide to go to Miami in Florida by car. And on the way from Washington D.C to Miami, he/she plans to make a stop in Charlotte, North Carolina to pay a visit to his/her sister for two days, and back from Miami to Washington D.C, he/she decides to stop in Orlando for pleasure purpose. Then, he/she plans his/her long distance personal business activity during the year. As noticed one year ago, he/she knows that he/she needs to go to Chicago to attend his/her best friend's wedding in May this year. The wedding is scheduled on Saturday, and he/she plans to take three days to make this long distance personal business tour with his/her spouse. As the time is limited, he/she decides to go to Chicago by air. And on the way to/from Chicago, he/she plans no stops.

Based on the structure of the model system and how the model works, we could summarize the information and data that needed for each person to simulate his/her long distance activities and travel (Table 4-20), the output of each model component (Table 4-21), as well as the model system generated or simulated long distance travel information for each person (Table 4-22). In Table 4-22, the trip is part of the long distance tour. If there are no stops (`Stops_In` = 0 and `Stops_Out` = 0) during the tour legs, there are only two trips for the long distance tour, and one trip for each direction. If there are, for instance, 2 stops during outbound leg of the tour and 2 stops during inbound leg of the tour (`Stops_In` = 2 and `Stops_Out` = 2), there will be a total of 6 trips during the entire tour, with 3 trips in each direction. If the trip is the first trip during the outbound leg of the tour, the `Trip-Origin` would be the `Tour-Origin`. If the trip is the first trip during the inbound leg of the tour, the `Trip-Origin` would be the `Tour_Dest`. If the trip is the last trip during the outbound leg of the tour, the `Trip_Dest` would be the `Tour_Dest`. If the trip is the last trip during the inbound leg of the tour, the `Trip_Dest` would be the `Tour-Origin`. If the trip is not the last trip during the outbound leg or the inbound leg of the tour, the `Trip_Dest` is the `Stop_Location`.

Input categories	Input variables	Description
Person Attributes	PersonID	Person ID
	State	State FIPs
	MSAPMSA	MSAPMSA where person lives, 4-digit code
	PUMA	PUMA where person lives, from PUMS data
	TAZ	TAZ where people lives, integer value from 1-380
	NMSA	whether TAZ is MSA or Non-MSA, 0 or 1
	PAGE	Person Age, integer value
	INCLVL	Household income level: low, middle, and high income level
	Gender	Person's gender
	EmpStatus	Person's employment status: Employed, Unemployed, School
	HHtype	Household type: Couple w/o Children, Couple w Children, Single, Non-Family
	HHsize	Household size
Transportation OD Skim	TT_car	Car travel time
	TC_car_Business/PB	Car travel cost for business and personal business purpose
	TC_car_Pleasure	Car travel cost for pleasure purpose
	TT_air	Air travel time
	TC_air	Air travel cost
	TC_air_1	Air travel cost in quarter 1
	TC_air_2	Air travel cost in quarter 2
	TC_air_3	Air travel cost in quarter 3
	TC_air_4	Air travel cost in quarter 4
	TT_train	Train travel time

	TC_train	Train travel cost
TAZ Economic and Demographic	# of Households	Number of household in TAZ
	# of Population	Number of population in TAZ
	# of Employment	Number of employment in TAZ

Table 4-20: Long distance travel demand model input data

Model Component	Model component output	Output Description
Yearly activity pattern model	Business	Person's total number of long distance business tours per year
	PB	Person's total number of long distance Personal business tours per year
	Pleasure	Person's total number of long distance pleasure tours per year
Tour destination choice model	Tour_Dest	The primary tour destination TAZ of each tour
Tour duration model	Tour_Duration	The tour duration in days
Time of year choice model	Tour_TOY	Time of year the tour is made: Q1, Q2, Q3, Q4
Travel party size choice model	Traparty	Travel party size during the tour
Travel mode choice model	TravelMode	Tour travel mode
Stop frequency Model	Stops_In, Stops_Out	Number of stops during the inbound and outbound legs of the tour
Stop purpose model	Stop_Purpose	Purpose of each stop
Stop location model	Stop_Location	Destination TAZ of each stop

Table 4-21: Output of each model component

Model Output	Description
Person ID	Person ID
N_Business	Person's yearly total number of long distance business tours per year
N_PB	Person's yearly total number of long distance Personal business tours
N_Pleasure	Person's yearly total number of long distance pleasure tours per year
Tour ID	Tour ID
Tour_Purpose	Purpose of the tour: business, personal business, pleasure
Tour_Origin	Tour origin: home TAZ
Tour_Dest	Tour primary destination TAZ
Tour_TOY	Time of Year the tour is made: Q1, Q2, Q3, Q4
TravelMode	Travel mode: car, air, train
TraParty	Travel party size during the tour
Tour_Duration	Tour duration in days
Stops_In	Number of stops made during inbound leg of tour
Stops_Out	Number of stops made during outbound leg of tour
Trip ID	Trip ID during the tour
Trip_Purpose	Trip purpose: business, personal business, pleasure
Trip_Origin	Trip origin TAZ
Trip_Dest	Trip destination TAZ

Table 4-22: Output of long distance travel demand model system for each person

From the modelling perspective, a model is usually considered to be able to reflect the maximum travel behavior reality if it takes into account all the necessary data and factors. However, such model could be too complex and have too many parameters that cannot be accurately measured using the available data. And it is likely that as the model becomes more complicated, it may not be able to generate stable and reasonable results. Our national travel demand model is developed and estimated with all the available information in 1995 ATS data, and to some extent it is simplified with a series of assumptions.

- People's long distance activities during one don't have interactions, which means that one long distance activity and travel will not affect the person's other long distance activities and travel during one year.
- People's long distance activity or the entire long distance tour occurs in the same quarter.
- The duration of the entire long distance tour will be modeled, but the duration of the activities occurring at the primary destination and the stops are not modelled due to the data limitation.
- The travel mode stays the same through the entire long distance tour, no travel mode transfer is considered. Only car, air and train are taken into account as the travel mode alternatives for the long distance travel.
- The stop is defined as the one that people make for a certain purpose (business, pleasure, and personal business) and stay for one day or more than one day. The stop made for a rest or mode transfer between the same travel mode or across different travel modes will not be included in the model.
- The maximum number of stops people could make during the tour leg is assumed to be four.
- As in the model estimation there is no data allowing us to model the time constraints, we assume that when simulating travel mode choice at the tour level, the one-way travel time should not be larger than the half time of the tour duration.
- At the stop level, the distance between the chosen stop location and the tour origin should be less than the distance between tour origin and tour

destination. The distance between the stop origin and stop destination should also be larger or equal than 50 miles.

- During the entire long distance tour, the travel party size will stay the same. It means that no one will get on or get off the tour during the long distance travel.
- The side stops occur based on the primary tour destination is not considered in the model system, and will not be estimated by the model system.
- At the tour level, people have different decision procedures and logics to determine their travel to primary destination for different purposes. At the stop level, people have the same logic and decision process to determine their stops from/to the primary destination.

Chapter 5: Preliminary Base Year OD Estimations

The long distance activity-based model system has a series of analytic tools to ensure maximum travel behavior realism and model sensitivity. All the model components at the tour level and the stop level adopt the discrete choice forms (multinomial Logit model) except for the tour duration component which employs the hazard duration methodology. A microsimulation-based framework, which simulates each person's long distance travel decision features, is developed based on the model system. The Monte Carlo simulation method is used in the framework to achieve an unbiased selection of alternatives in the case of estimating each decision.

The year of 2010 was chosen as the base year. The 2010 PUMS data provides the population information. The 1 percent 2010 PUMS data was expanded to the whole nation's population according to the person weight in PUMS file. Each individual will be assigned a TAZ (MSA/Non-MSA) based on the correspondence file of MSA/Non-MSA and PUMA. In this way, the generated population by expanding the sample according to the weight can be observed statistically representative of the true distribution in MSA/Non-MSA. The population in each MSA/Non-MSA is shown in Figure 5-1. The base year transportation OD skim data and the economic/demographic data are also collected and processed using the same methods discussed in Chapter 3. In auto skim data estimations, the average auto fuel efficiency is 21 mpg (Bureau Transportation Statistics, 2015) and the average retail fuel price is \$2.56/gallon in the U.S. in 2010 (EIA). The average unit lodge costs for different income groups are correspondingly converted to 2010 dollar based on CPI.

The other assumptions (e.g. travel speed and rest hours) in auto OD skim data estimation remain the same. The 2010 air skim data is obtained from the 2010 DB1B data. The train cost for each OD is converted to 2010 dollar based on the collected 2013 Amtrak train data, and the travel time by train for each OD is assumed unchanged from 2010 to 2013. The 2010 county-level economic/demographic data is obtained from CEDDS and further aggregated to MSA/Non-MSA level.

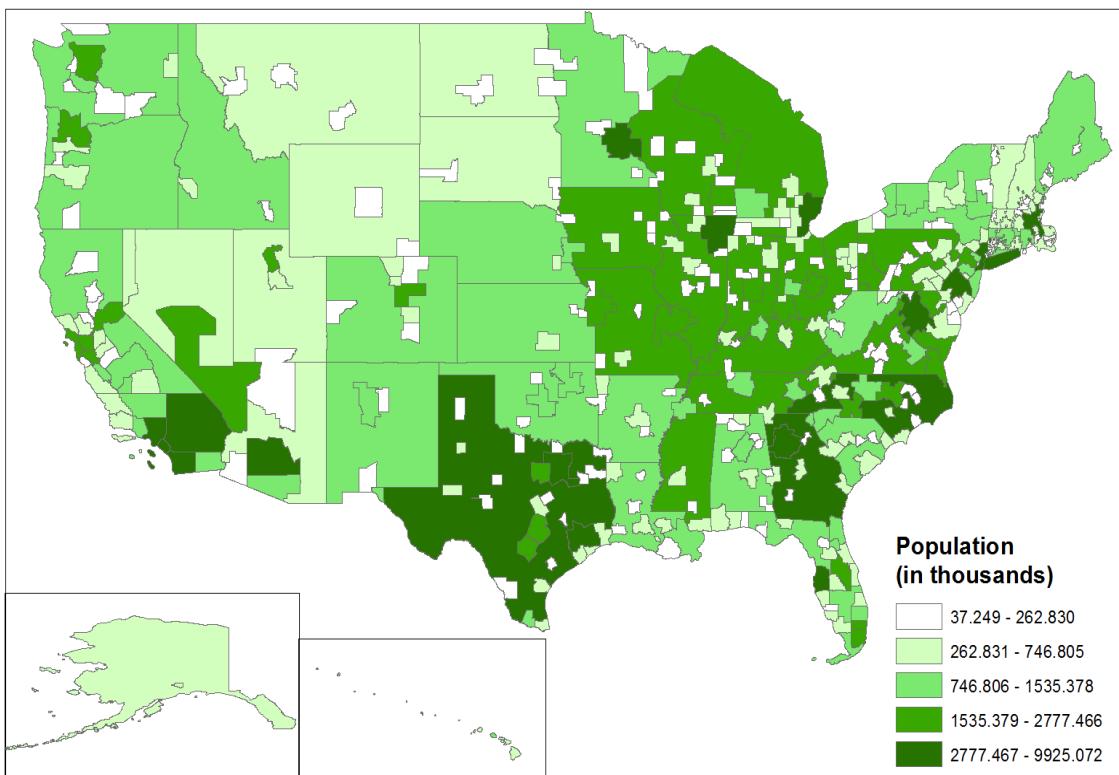


Figure 5-1: MSA/Non-MSA Population

In the simulation, we made several assumptions to add the time constraints as many as we can, as in the model estimation there is no data allowing us to model the time constraints. When the travel mode choice at the tour level is simulated, the one-way travel time should not be larger than the half time of the tour duration. Such assumption constrains the travel mode choice for certain OD pairs given the tour duration. Meanwhile, at the stop level, the distance between the chosen stop location

and the tour origin should be less than the distance between tour origin and tour destination. As in the long distance travel in our model, the tour destination is defined as the one with the furthest distance. It will conflict with the definition if the distance between the stop and the tour origin is larger than the tour OD distance.

The microsimulation tool is developed using Java. Given all the input data, our developed micro-simulator could output the long distance activity patterns of each person in the U.S. Given all the input data including 2010 PUMS population data, transportation OD skim data, and economic/demographic data, our developed micro-simulator could output the long distance travel information of each person in the U.S. Based on the simulation results (integrating the generated tour-level and the stop-level travel information), we can obtain a total of 48 national-level trip OD tables (4 quarters * 3 travel modes * 4 Purposes, 4 purposes includes business, personal business, pleasure, and back to origin). Aggregating the cell trips in each of the 48 OD tables can give us the national trip distribution by travel mode, trip purpose and time of year separately.

Aggregating the cell trips in each of the 48 OD tables can give us the national trip distribution by travel mode in each quarter (Figure 5-2). The aggregate results imply that car travel accounts for above 80% among the three travel modes, and a very small portion of people (less than 1%) would choose train for their long distance travel. The very small sample size of train travel in model estimation data (1995 ATS) could cause the bias of the model estimation coefficients of train travel. There are almost 20% of trips made by air. Figure 5-3 shows the trip distribution by travel mode and time of year. It indicates that a large number of people prefer taking their long

distance trips during the first three months (from January to March) regardless of the travel mode, and a few people would choose the fourth quarter (from October to December) to travel. The trip distribution pattern by time of year has a discrepancy compared to what is expected and the survey distribution. In the survey, as what we expected, trips made in the third quarter (from July to September) account for the largest group. One reason of this could be that the model coefficients used to simulate the travel in 2010 are estimated using 1995 ATS data which is 15 years old.. It could be desirable to use a survey data close to the year of 2010 to estimate the model. However, the most current long distance travel survey with detailed travel information is 1995 ATS. In model calibration procedure, we will try to resolve this discrepancy. Summing all the trips including back to home/origin trips yields a total of over 3.59 billion trips during a year which infers that a person would take an average of around 11 long distance trips during one year.

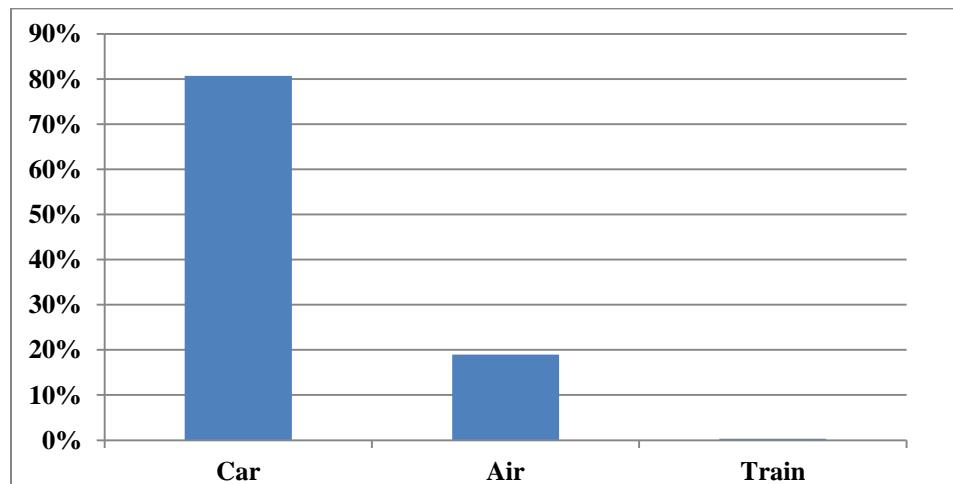


Figure 5-2: Trip Distribution by Travel Mode

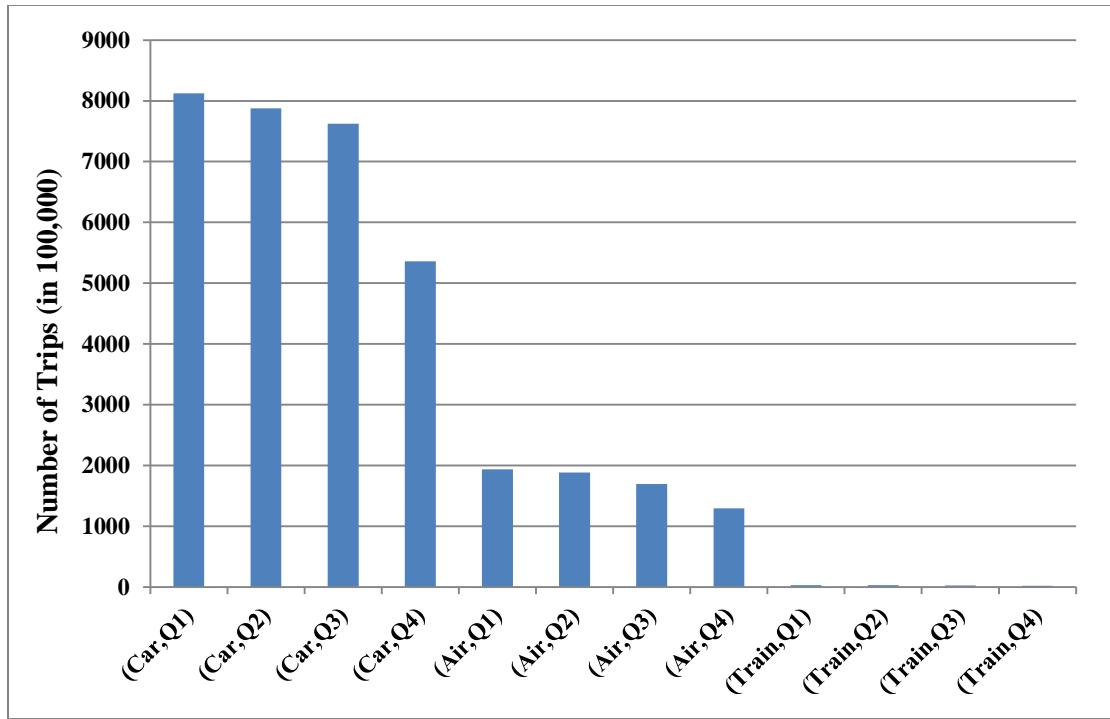


Figure 5-3: Trip Distribution by Travel Mode and Time of Year

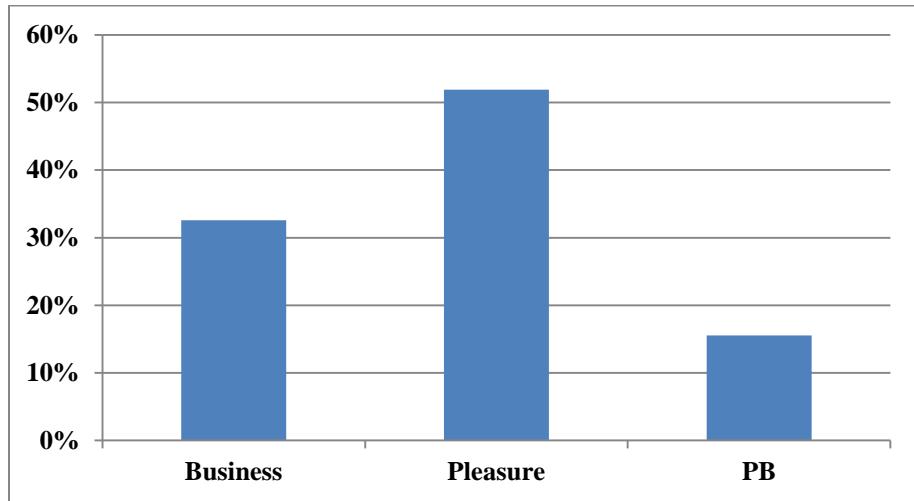


Figure 5-4: Trip Distribution by purpose

Based on the OD tables, we can obtain long distance trips from/to each MSA/Non-MSA zone (see Figure 5-5) for the nation. As noted here, trips from/to each zone plotted in the map are different from the trip generation or attraction terms in the traditional 4-step model. The trips are parts of the long distance tour. If a

person has one tour and one stop during each half leg, then this person will have 4 trips. Therefore, the number of trips generated from each zone should equal to the number of trips destined to the zone for the model year.

It is observed from Figure 5-5 that the trip distribution has a similar pattern with the population distribution. Generally speaking, the MSA zones (except several large cities such as New York, Los Angeles, Chicago, Seattle, Washington D.C) have a smaller number of trips due to the smaller size in terms of geography population. The zones with larger population (in northeast, west pacific, Texas, and some large cities in west coast) usually have a larger number of trips during one year, while the zones in the states of Montana, North Dakota, South Dakota, and Wyoming produce a relatively small number of trips which can be explained by the small number of the population and low GDP of these states.

As an illustrative example, Figure 5-6, Figure 5-7, and Figure 5-8 show the trip distribution by TAZ and travel mode for trips from the Washington D.C metropolitan area. Figure 5-6 infers that the car trips from D.C usually are centralized around D.C, and the farther the zone is away from DC, the fewer people will travel by car. Therefore, most of the car trips from D.C occur in the middle-east and east parts of the U.S around the D.C area. Since the train network is not as wide as car or air network, the trips taken by train should be distributed in the zones with train stations. The distribution of train stations reflects the distribution of Amtrak rail stations, as the train data are from Amtrak. As expected, the train trips from D.C are mainly located in east coast and centered around D.C. In general, those trips have shorter distance than the air trips and the distribution range is much smaller than that of the air trips

(see Figure 5-7). The most significant advantage of air travel is that it is unquestionably the fastest mode among all the transportation means, especially when people travel in longer distance. Moreover, in the U.S people can travel by air to most places in the nation due to the high density of airlines and airports. Consequently, trips by air departing from D.C are observed all over the U.S. (Figure 5-8). A large number of air trips to non-MSA zone in Virginia and Richmond are observed due to the fact that the DB1B data contains the air fare between D.C and the two zones. Also because the Non-MSA zone in Virginia contains multiple counties in Virginia. If we could divide the Non-MSA zone into smaller parts, we should observe that the areas near D.C don't have air trips, and the areas far from DC have a lot of air trips. If we compare the number of trips from D.C to Non-MSA in Virginia and Richmond by travel mode, we could see that a lot more people would choose car than air.

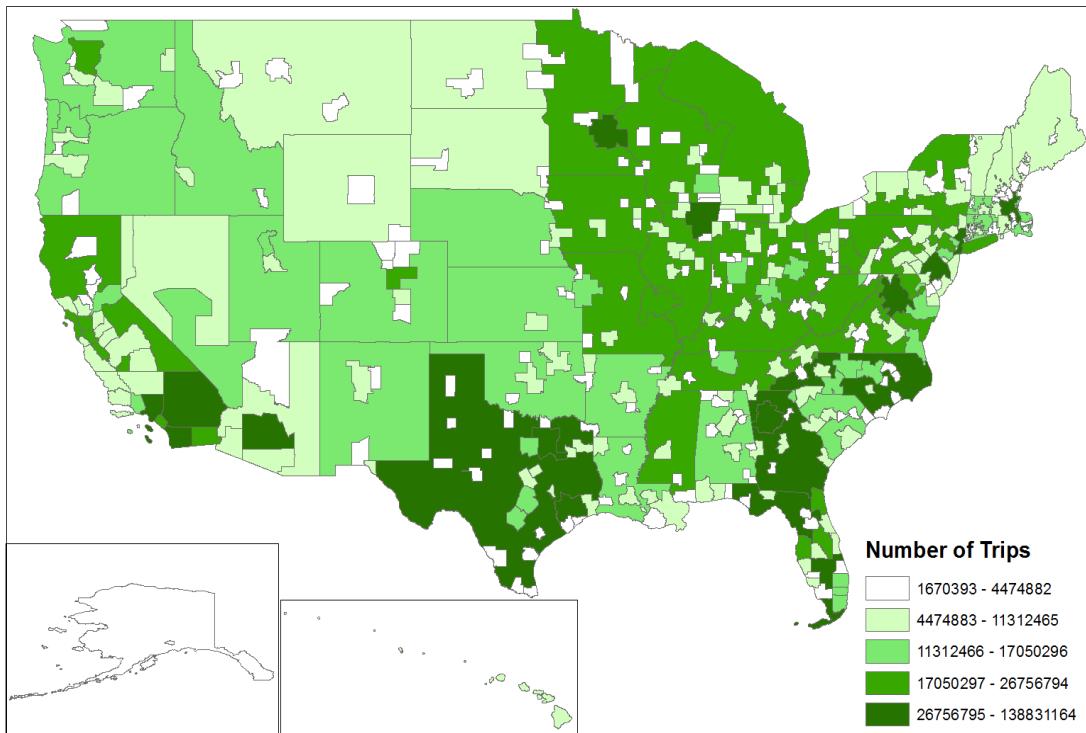


Figure 5-5: Trips Originate/Destinate at MSA/Non-MSA level

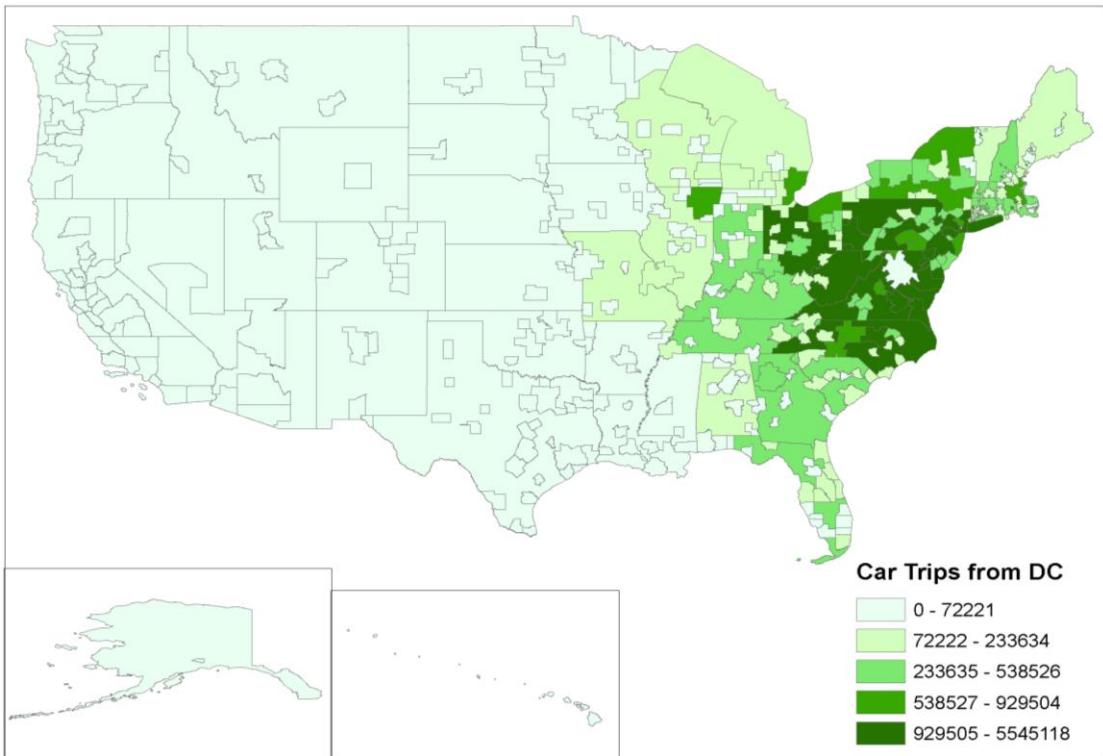


Figure 5-6: Yearly Car Trip Distribution by TAZ originating from Washington D.C

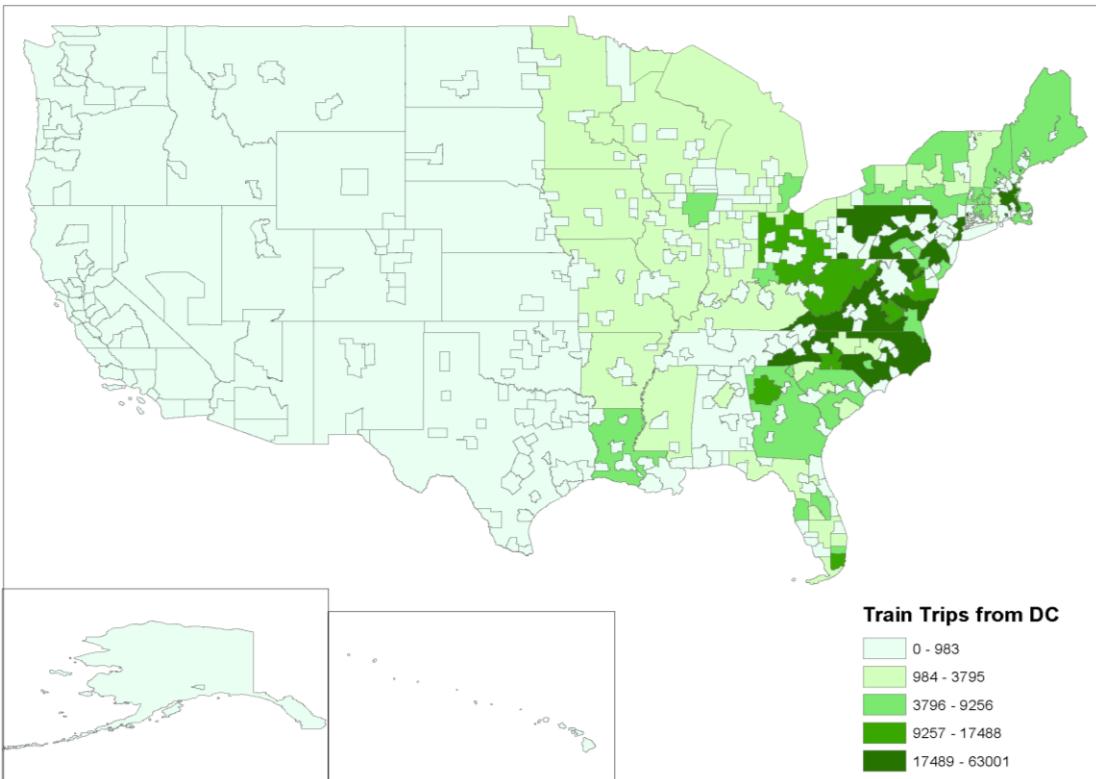


Figure 5-7: Year Train Trip Distribution by TAZ Originating from Washington D.C

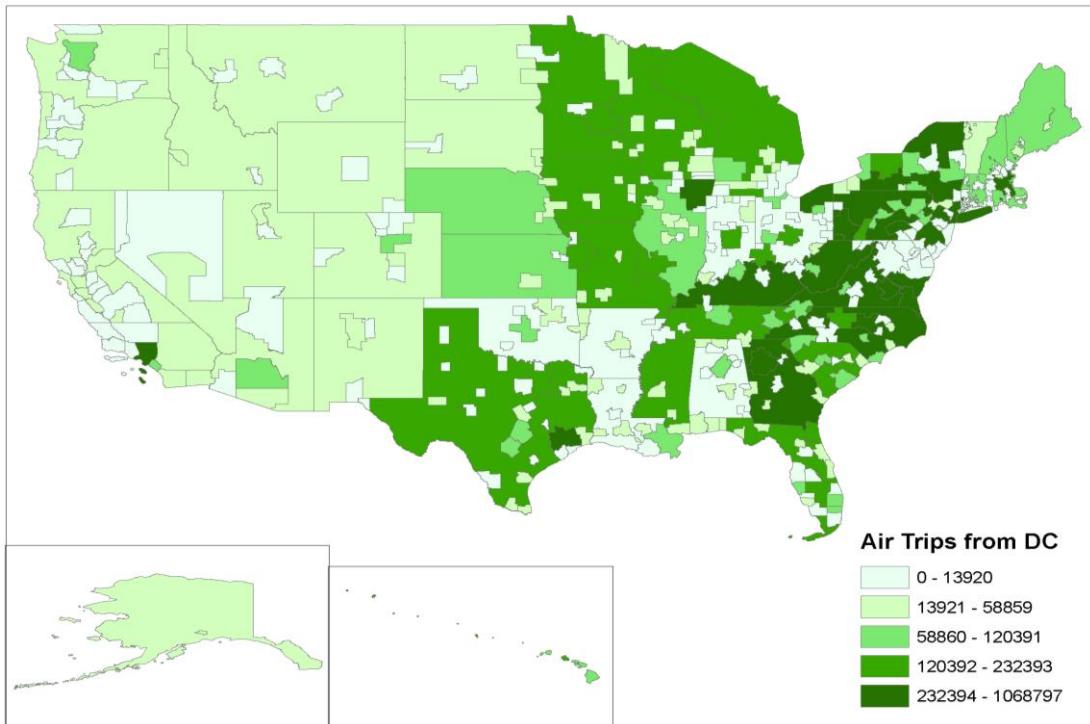


Figure 5-8: Yearly Air Trip Distribution by TAZ Originating from Washington D.C

The national travel demand model is a person-based microsimulation model, and it employs the Monte Carlo simulation method to obtain an unbiased selection of alternatives in the case of estimating each decision. Therefore, there exists random noise in the model results. To test the noise level or stability of the microsimulation-based national travel demand model, we need to run the model multiple times and analyze the results of the model runs. As it usually takes four days to complete one full run of the national travel demand model, we only conducted 5 more full runs to evaluate the model's stability. And with the base year model simulation results, we could have a total of 6 runs of model results to compare.

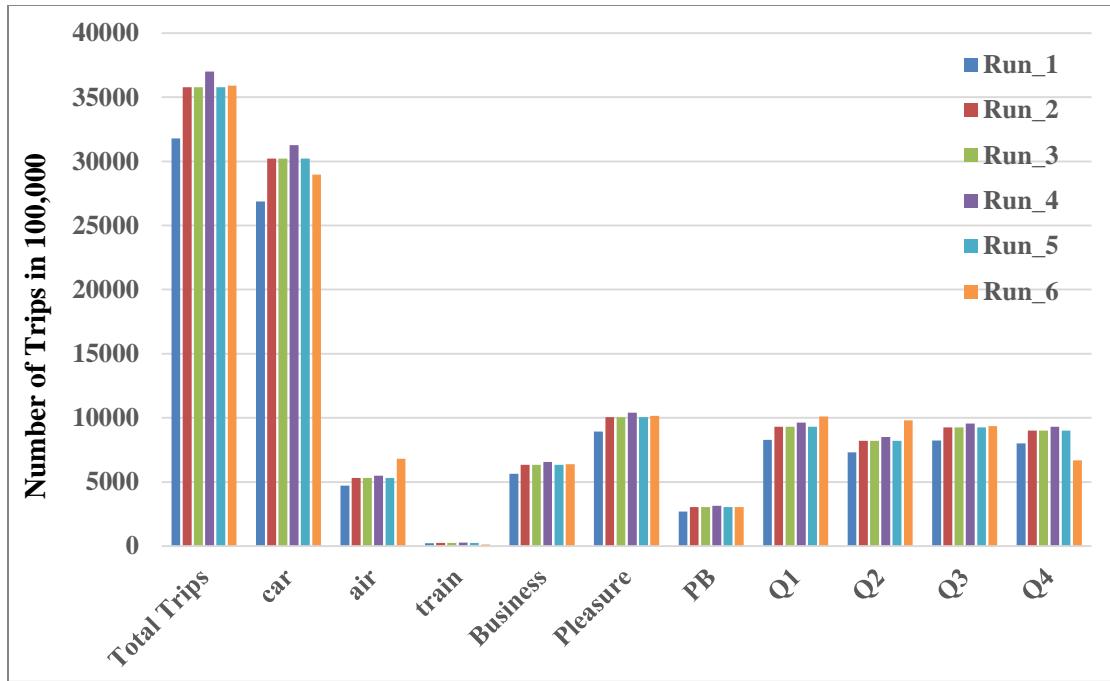


Figure 5-9: Number of trips of different categories for different model run

Figure 5-9 shows the comparison of the total number of trips, number of trips by travel model, trip purpose and time of year for the 6 model runs. As we can see that the total number of trips ranges from 3.18 billion to 3.7 billion, and four model runs generate the trips at the level of 3.58 billion trips. The number of generated car trips ranges from 2.7 billion to 3.13 billion, and 3 model runs generated a total of around 3.02 billion car trips. The number of air trips ranges from 0.47 billion to 0.68 billion, and four model runs generated the air trips at a similar level, which is around 0.5 billion trips. The number of train trips ranges from 12 million to 25 million, and four model runs generated the train trips at the level of around 25 million. From the trip comparison for different purposes, we can observe that the model is able to generate quite stable results in terms of number of trips by purpose. And with regard to the number of trips by time of year, 4 model runs (Run_2, Run_3, Run_4 and Run_5) generated a similar level of number of trips in different quarters.

Therefore, with more runs of the national travel demand model, we could expect that most of the model runs will generate a similar level of total number of trips which is around 3.58 billion, and a total number of around 3.02 billion car trips, around 0.5 billion air trips and around 25 million train trips. The microsimulation model could generate quite stable results in terms of number of trips by trip purpose. Most of the model runs will generate a similar level of number of trips in different quarters.

Chapter 6: Model Calibration

Travel demand model calibration is essential to accurately model people's travel. Model calibration is the process of adjusting the model parameter values until the simulated or estimated travel results closely match the observed travel for the base year.

In general, the process of system-wide calibration of parameters of a simulation-based model is understood to find the value of the parameters that minimize the error between observed outputs and simulated outputs. The observed outputs, usually, refer to trusted external data sets containing aggregate measures matching the base conditions (i.e. OD tables, traffic counts for base year). Thus, we want the simulation-based model to replicate these base conditions. The degree of trust in each measurement may be represented as a weight in the calibration process. The simulated output refers to the aggregation of the simulation results into measures that are equivalent to the observed output's measures.

The calibration process is formulated as a constrained minimization optimization problem as follows,

$$\text{minimize}_{\theta} w \left\| O_m - O_S(\theta) \right\|^2$$

s.t.

$$O_S = F(Z; \theta)$$

$$l \leq \theta \leq u$$

Where w is a vector of weights indicating the trust of the modeler on different observed outputs, and in our calibration, the w is set as 1; O_m is a vector representing

the observed outputs and O_s is a vector representing the simulated outputs; θ is the vector of parameters from the model to be calibrated, and l and u are vectors of lower bounds and upper bounds for the parameters; $F(Z; \theta)$ represents the link between the simulated outputs and the simulation-based model, and Z are the inputs required to run the simulation-based model. Lastly, the $\|.\|$ represent the Euclidian norm.

The algorithm selected to solve this constrained minimization problem is *Simultaneous Perturbation Stochastic Approximation* (SPSA) (Spall, 2003). This algorithm has the following advantages: 1). it accounts for the simulation error in the simulation-based model output; 2). it can be applied in stochastic gradient setting or a gradient-free setting; 3). it only requires two function evaluations per iteration regardless of the length of the vector of parameters.

In essence, the SPSA algorithm works by perturbing the components of the vector of parameters and computing difference of these perturbations with respect to the objective function of the constrained minimization problem. More detailed information can be found in Spall's work (Spall, 2003). One example for calibrating models using a variation of SPSA is the work from Antoniou et al. (2015). The implementation step of the algorithm is briefly described as follows,

Step 0 Initialization and coefficient selection

In this step, the SPSA algorithm is set up with the initial values for the vector of parameters and also the values for the vector of hyper-parameters. These hyper-parameters belong exclusively to the SPSA algorithm.

Step 1 Generation of simultaneous perturbation vector

A perturbation vector is generated using Monte Carlo simulation. This perturbation vector uses a Bernoulli distribution centered at zero.

Step 2 Objective function evaluations

Evaluate the objective function twice using the perturbation vector.

Step 3 Gradient approximation

Compute the gradient approximation using the perturbation vector and two evaluations of the objective function computed from step 2.

Step 4 Update vector of parameters

Update the values of the vector of parameters based on the gradient descent using the approximated gradient computed from step 3.

Step 5 Iteration or termination

Return to step 1 to continue iterating or terminate if there is negligible change between iterations in the objective function and/or values of the vector of parameters.

For the passenger long distance model, we initially calibrated the alternative specific constants of the time of year choice model, and the travel mode choice model using the Airline Origin and Destination Survey (DB1B) data. This survey is a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics. Data includes origin, destination and other itinerary details of passengers transported. This database is used to determine air traffic patterns, air carrier market shares and passenger flows. The model is calibrated on the base year model.

For the purpose of this calibration effort, airline OD data is summarized to the following 4 values for each quarter:

- Number of flights departing from Maryland
- Number of flights landing in Maryland
- Number of flights departing from all other states
- Number of flights landing in all other states

Consequently, airline OD data was summarized into 16 variables. The model outputs were also summarized to capture these 16 variables.

Sum of the squared differences between model simulation and airline OD values for these 16 variables was used as the objective function of the calibration. SPSA seeks to minimize the objective function by changing model parameters. For that, some of the model parameters should be selected as the inputs of SPSA to be changed. Mode choice and Time of year choice models were selected to be calibrated against airline OD data. Alternative specific constants for these models were selected as the variables to be calibrated. The followings are the variables selected for calibration:

- Business trips time of year model alternative specific constant 1
- Business trips time of year model alternative specific constant 2
- Business trips time of year model alternative specific constant 3
- Pleasure trips simple time of year model alternative specific constant 1
- Pleasure trips simple time of year model alternative specific constant 2
- Pleasure trips simple time of year model alternative specific constant 3
- Pleasure trips full time of year model alternative specific constant 1

-Pleasure trips full time of year model alternative specific constant 2

-Pleasure trips full time of year model alternative specific constant 3

-Business trips mode choice model alternative specific constant 1

-Business trips mode choice model alternative specific constant 2

-Pleasure trips mode choice model alternative specific constant 1

-Pleasure trips mode choice model alternative specific constant 2

-Personal Business trips mode choice model alternative specific constant 1

-Personal Business trips mode choice model alternative specific constant 2

Another important aspect of the calibration is hyper-parameter selection for SPSA algorithm. SPSA includes 5 hyper-parameters: a , c , α , γ , and A . More information about these parameters can be found in the literature of Spall's. These parameters were selected based on the suggestions in the optimization literature using variance of objective function, and average of gradients.

We conducted several rounds of calibration, and finally chose the calibrated parameters under 60 iterations of calibration as it shows better performance in trip distribution by time of year and other trip distributions. SPSA algorithm was coded in Java to find the calibrated variables. The initial values for calibration variables were set to their estimated values from the estimation process.

The 60-iteration calibration results are shown as below:

Hyper-Parameter	Value
A	100
a	1.3724531551126038E-18

c	0.1
α	0.6
γ	0.1

Table 6-1: 60-Iteration Calibration Hyper-Parameter

Number of iterations	60
Initial objective function value	5.3036402075560817E18
Final objective function value	4.692863010673278E18

Table 6-2: 60-Iteration Calibration Hyper-Parameter

Variable	Calibrated Value	Initial Value
toyBusiness_ASC1	-2.199	-1.469
toyBusiness_ASC2	-1.695	-0.744
toyBusiness_ASC3	-0.427	-0.370
toySimplePleasure_ASC1	-2.017	-1.228
toySimplePleasure_ASC2	-1.239	-1.285
toySimplePleasure_ASC3	-0.384	-0.541
toyFullPleasure_ASC1	-1.415	-1.120
toyFullPleasure_ASC2	-1.049	-0.653
toyFullPleasure_ASC3	-1.011	-0.403
mc_BUSINESS_Air	-1.622	-0.44
mc_BUSINESS_Train	-2.350	-2.93
mc_PLEASURE_Air	-3.960	-2.95
mc_PLEASURE_Train	-2.676	-3.56
mc_PERSONAL_BUSINESS_Air	-1.626	-1.49
mc_PERSONAL_BUSINESS_Train	-4.110	-3.75

Table 6-3: 60-Iteration Calibration Results

Chapter 7: Future Year Policy Analysis

7.1 Future Year Policy Scenarios

The year of 2040 is selected as the horizon year analysis. Two categories of transportation policies are analyzed for the year of 2040 through the calibrated model. One is the fuel price changes, and the other one is the High Speed Rail.

In recent years, the crude oil price has driven large fluctuations of the market fuel price (Figure 7-1) (EIA, 2016) which is an important component of transportation cost. Since January of 2003, the retail motor gasoline price has risen dramatically from the average \$1.5 to more than \$3. Travelers have responded to the 100 percent increase of the gasoline price in different ways. They have adjusted their travel behavior, driving habit and even changed their vehicle type to more fuel-efficient ones (CBO, 2008). According to the CBO study on evaluating the effects of the fuel price increase in a metropolitan area where the transit is available, it shows that for every 50 cents increase in the retail fuel price, there is a 0.7 percent decrease in the number of freeway trips. And the number of transit trips is increased accordingly. As the travel distance increases, the fuel cost would be a dominating cost that people would take into account while driving.

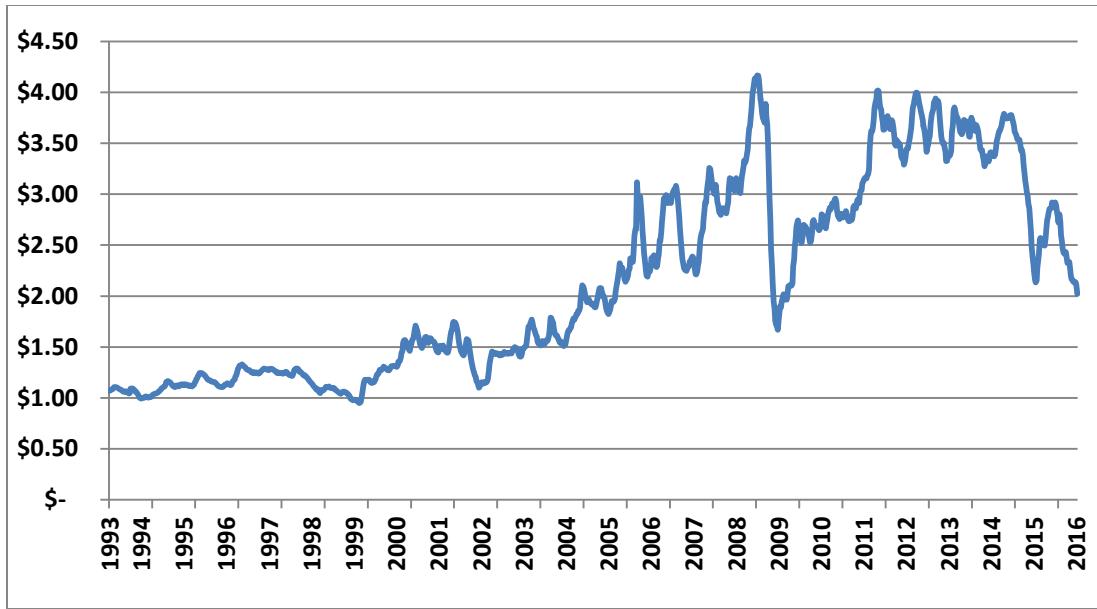


Figure 7-1: Retail Gasoline Price Changes

Meanwhile, between 2002 and 2013, jet gasoline price has increased more than four times from \$0.72 to \$2.98 per gallon and general aviation gasoline price has increased more than three times from \$1.29 to \$3.93 per gallon in nominal terms (GAO, 2014). The increase of fuel price does affect the aviation activity including both scheduled and non-scheduled air service. In order to mitigating the financial effect of the fuel price increase, commercial airlines may improve the flight fees including flight tickets, checked bag fee, and other facilities, and some commercial airlines have taken a number of other steps such as restraining the domestic seat capacity growth, reconfiguring the fleets, conducting efficient flight and ground operation, etc (GAO, 2014). All of these could affect the comfort of individual's travel experience and airline's level of service, which will definitely affect travelers' travel mode choice when making plans for long distance travel.

Compared to airline and driving, the fuel cost has accounted for less percentage of total operating cost in railway (Tipping et.al, 2015). As the fuel price

decreases, the rail, as a more energy-efficient travel mode, will lose some of the cost advantages (USDOT, 2008). Therefore, in our study, we will not reflect the fuel price changes through the railway level of service variables (e.g. rail fare) in the scenario of fuel price changes.

In our study, the impact of the fuel price increase on people's long distance travel is analyzed. According to the crude oil projections by EIA (Figure 7-2), the crude oil price is observed to increase around 1.75 times of the price in 2012. Therefore, for simplicity, we assume that the retail fuel price is also increased by the same extent which is 1.75 times of the 2012 retail fuel price. It is complicated to identify the quantitative relationship between the fuel price changes and the air fare changes. But we do have the knowledge that the longer the flight distance is, the more fuel the airplane would consume given the same number of passengers. Based on this knowledge, in this research we simplified the relationship by assuming that the air fare would be increased by \$5 if the flight distance is less than 500 miles, increased by \$10 if the flight distance is between 500 miles and 1000 miles, and increased by \$30 if the distance is larger than 1000 miles.

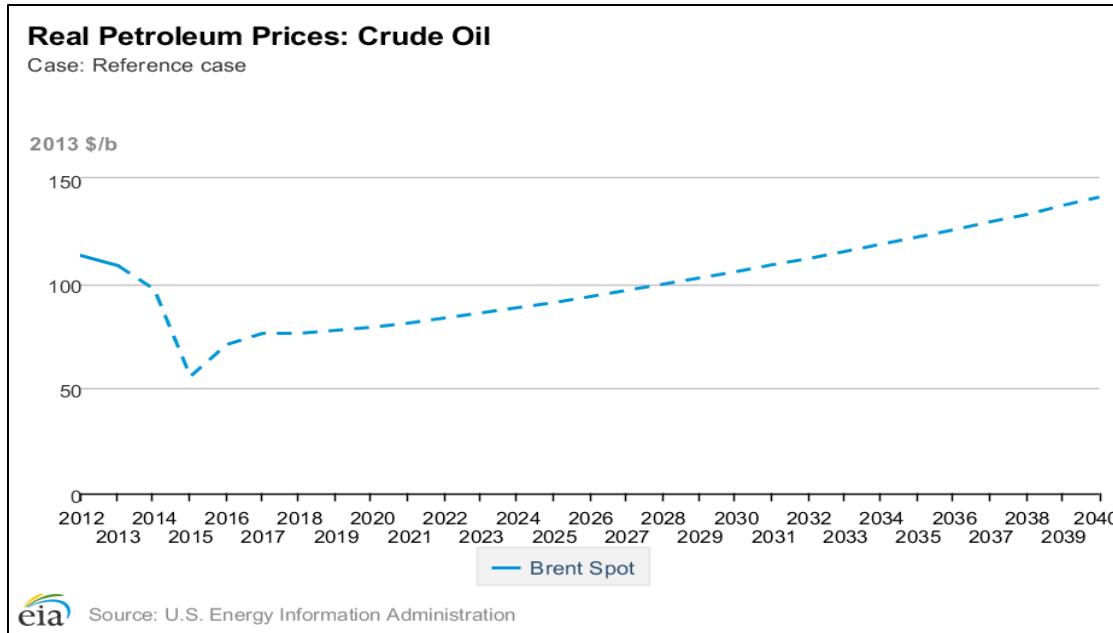


Figure 7-2: Crude Oil Price Projections to 2040

As mentioned before, High speed rail (HSR) is expected to help alleviate the heavy load of the traffic in road and air corridors and improve the inter-regional accessibility. The U.S federal and state planners are prompted to provide the high speed rail services through selected major corridors. The developed personal-based national travel model gives us ability to quantitatively forecast the high speed rail demand and evaluate the operational effectiveness of the investment to some extent. In this research, part of the northeast corridor is selected to forecast the high speed rail demand and evaluate its impact on the long distance travel market. It is desirable and more accurate that the analysis is conducted based on the stated preference (SP) data, as high speed rail along the northeast corridor does not exist and no one has experience of taking high speed rail. For simplicity, we take the method of improving the speed of the current rail and the corresponding fare as an equivalent of high speed rail.

To be specific, the travel time/cost changes in the scenarios for future year analysis are described as follows:

Base Scenario: Future Year Base Scenario

No change for the air and the train OD skim data in this scenario in the year of 2040. With regard to the car OD skim data, the vehicle fuel economy is assumed to improve to 30 MPG while the fuel price and other information maintain the same as in the year of 2010 for car travel. The increased MPG will definitely cause the cost decrease of car travel. This scenario is a basis of the other two scenarios, which means the transportation OD skim changes in the following scenarios are based on the future year base scenario.

Scenario 1: Fuel Price Increase

As we discussed before, fuel cost takes up fewer proportion of the total operating cost for railway compared to car and air. Therefore, in this scenario, the train OD skim data stays the same as the base year while the car and the air travel cost will be changed. To be specific, the retail fuel price is increased to 1.75 times of the base year fuel price for car travel, which is finally \$4.48/gallon. As the vehicle fuel economy is set as 30 MPG which is 9 MPG higher than the base year. As both the fuel price and the vehicle economy are increased, it is hard to tell the travel cost change without calculation of specific OD. Regarding the air travel, the air fare is increased by \$5 if the OD flight distance is less than 500 miles, increased by \$10 if the distance is between 500 miles and 1000 miles, and increased by \$30 if the distance is larger than 1000 miles.

Scenario 2: High Speed Rail

In this scenario, we improve the travel speed of the existing rail line as a proxy of high speed rail scenario. The Washington D.C – New York section of the northeast corridor is chosen to analyze the travel demand changes along the corridor and its effect on the effectiveness of the investment. According to Amtrak's projected planning and construction of high speed rail in northeast corridor, the travel time between New York and Washington DC including a stop in Philadelphia will be reduced to 96 minutes by 2040 (Amtrak, 2010). Therefore, in this scenario, we adopt the travel time projected by Amtrak, which is 96 minutes. Based on this, we assume one hour of travel time between New York and Philadelphia, and 0.6 hours of travel time between Washington D.C and Philadelphia. As the high speed rail provides a higher level of service in terms of the travel time which decreases a lot compared to the regular train travel, the travel cost of the high speed rail should be increased to reflect the improved service. A 30% increase of travel cost for the high speed rail is proposed and used.

In all these three scenarios, we use the base year dollar in monetary value as we have no information about the CPI between the 2040 and 2010. Besides the transportation OD skim data, the economic and demographic data for each MSA and non-MSA in the year of 2040 is used to reflect the future year changes in economy and population. Such data is obtained from the Complete Economic and Demographic Data Source (CEDDS) by Woods & Poole Economics. As mentioned before, this database also offers projected socioeconomic indicators (e.g. population, employment, households, etc.) for all the regions, states, statistical areas and counties in the U.S.

7.2 Future Year Population Synthesis

The developed person-based microsimulation national travel demand model recognizes individual as the decision maker, and the travel demand is derived from each individual's desire in spatially located activities. When implementing such model, the initial step is generating the population in the U.S, particularly each individual's socio-demographic characteristics except children under 18 years old, as the children under 18 years old is not considered as the decision maker in our study.

Population synthesis is a procedure that expands the sample drawn from a population to the full size of the population such that the synthesized population can be representative of the actual population at various aggregate levels (Ryan et al, 2009; Lim & Cargett, 2013). The main idea of the population synthesis is to combine the census sample data (both household and person) with available up-to-date aggregate distribution or margins data (Beckman et.al, 1996). There are many population synthesizers, either standalone software packages or component of microsimulation activity-based travel demand models, most of which function based on iterative proportional fitting procedure (IPF) (Bowman, 2004). IPF estimates a distribution of control variables so that the number of individuals in given categories matches the corresponding margins and meanwhile the correlation structure of the seed is maintained (Axhausen and Müller, 2010). For more details about the IPF procedure, please refer to Deming and Stephan (1940) who first introduced the IPF method.

Among the various kinds of population synthesizers such as PopSynWin, PopGen, ILUTE (Salvini & Miller, 2005), FSUMTS (Srinivasan & Ma, 2009; Srinivasan et.al, 2008), CEMDAP (Guo & Bhat, 2007), etc., we chose PopGen (Ye et.al, 2009), a standalone open source software package, developed by Arizona State University to generate the future year population for the whole U.S by using distributions of household and person variables of interest and a sample of household data.

PopGen also uses the standard IPF procedure draw households from the provided sample data to match the marginal distributions of control variables. It adopts the iterative proportional updating algorithm to estimate the household weights. Based on the household composition and the provided marginal, PopGen develops the weights for households in the sample data by using the IPU algorithm. It incorporates a heuristic approach to generate synthetic populations while matching both household-level and person-level characteristics of interest.

PopGen is a Python-based software with an friendly graphical user interface (GUI) (Figure 7-3). Through the wizard-based project setup procedure, the users can choose the region at different levels in the U.S and provide their own input data or use the default inputs for population synthesis. Once the inputs are provided, PopGen will import the data into MySQL database and works from there.

A total of 5 input files are required before conducting the population synthesis in PopGen. They are

1. Household sample file providing regional distribution of household characteristics,

2. Person sample file providing regional distribution of person attributes,
3. Household marginal file providing the marginal distribution of household attribute at specific geographic level,
4. Person marginal file providing the marginal distribution of person attributes at specific geographic level
5. Correspondence file between different geographies used in sample file and marginal file.

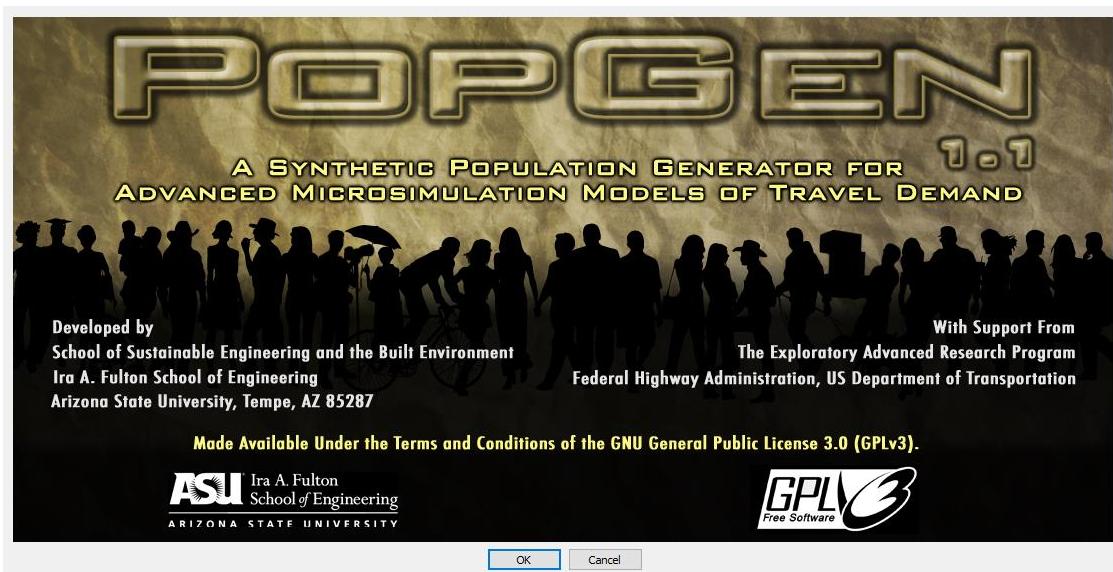


Figure 7-3: Interface of PopGen

The TAZ (MSA/Non-MSA) is consisted of multiple counties or cities, once we know the population of each county, we can have the population in each TAZ. With regard to the county in which different cities belong to different MSA, we allocate the number of population in the county to the corresponding TAZs according to the population ratios in that county. Therefore, the geographic level we choose to synthesize the population is county. The household sample file and person sample file are prepared from the 2013 ACS PUMS one-year data which is the latest ACS PUMS

data having the same PUMA system as the one in PopGen. The household marginal file and the person marginal file were prepared based on the Woods & Poole Economics' projected data on the number of households by income, the number of persons by age, and the number of person by gender in 2040. Specifically, in household marginal file, the control variable household income is divided into 4 categories which are income group 1 with income less than \$30000, income group 2 with income greater than or equal to \$30k and less than \$75k, income group 3 with income greater than or equal to \$75k and less than \$150k, and income group 4 with income greater than or equal to \$150k. In the person file, we use person age and gender as the control variables, and the person age is divided into 5 categories which are [0, 20), [20, 35), [35, 55), [55,75), and [75, the up limit in the sample]. Regarding the geographic correspondence file, we used the default one in PopGen.

We conducted the population synthesis for each state. And we randomly select one state to evaluate how the synthetic population and households match the provided control margins. The comparisons between the synthetic population and provided totals by control variable (Figure 7-4) shows that PopGen can generate the future population that can match the control totals for different variables very well. The synthetic population in 2040 for each MSA and Non-MSA can be shown in Figure 7-5.

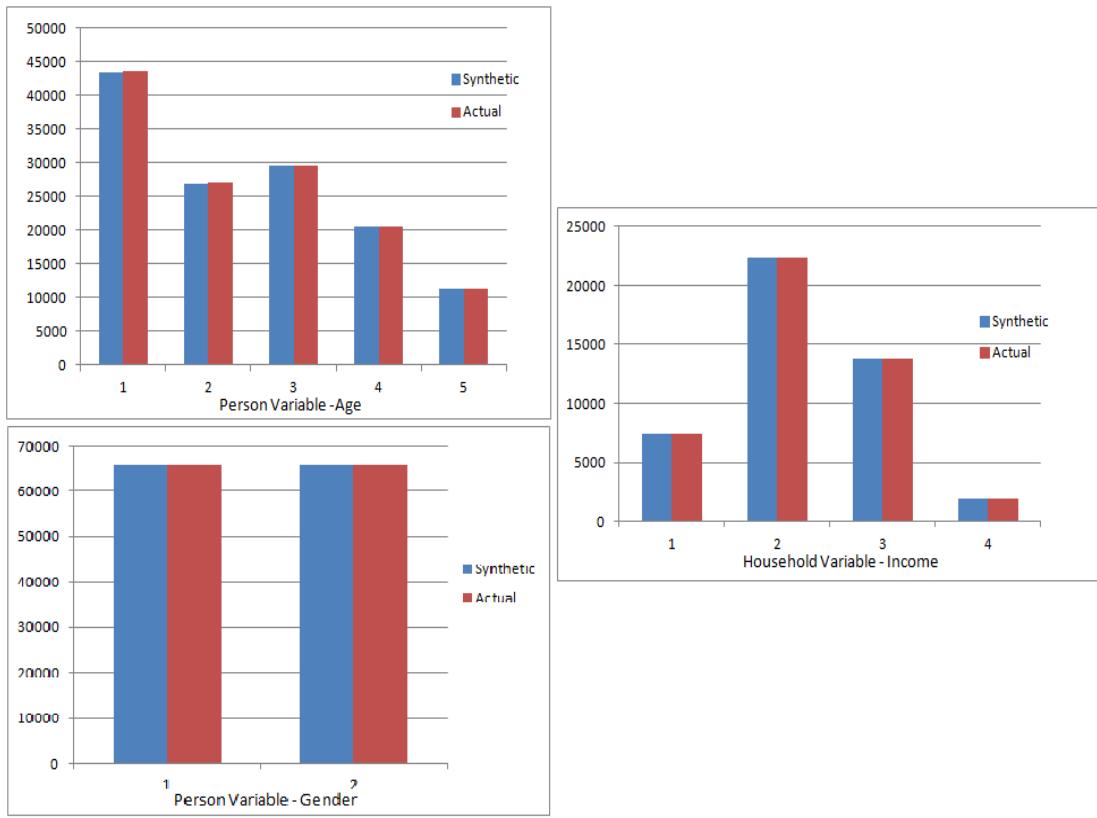


Figure 7-4: Comparison between Synthetic Population and Control Margins by Control Variable

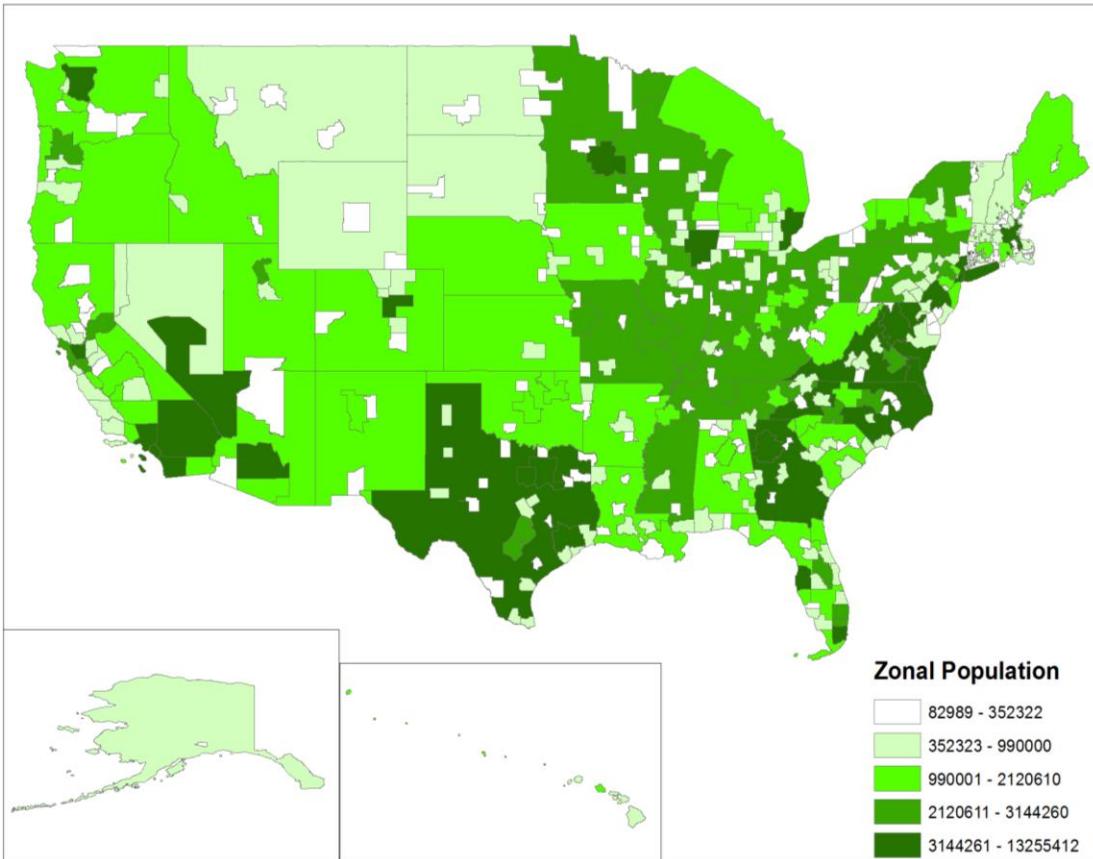


Figure 7-5: MSA/Non-MSA Population in 2040

Figure 7-6, Figure 7-7, and Figure 7-8 shows the synthetic population distribution by gender, income group and age group. It shows that female population occupies a larger share in our 2040 synthetic population, with above 50.5%, while male population has a share of less than 49.5%. Income group distribution figure indicates that high-income group (household income above \$75,000) has the largest share of the whole 2040 synthetic population, while the low-income group (household income less than \$30,000) has the smallest number of population. The number of middle-income group population (household income between \$30,000 and \$75,000) is smaller than high-income group but larger than low-income group. Also, as expected, the middle-age group (from 18 years old to 35 years old) has the largest

number of population, while the old-age group (above 60 years old) has the smallest number of population which is close to but slightly smaller than the young-age group (above 35 years old and younger than 61 years old) population.

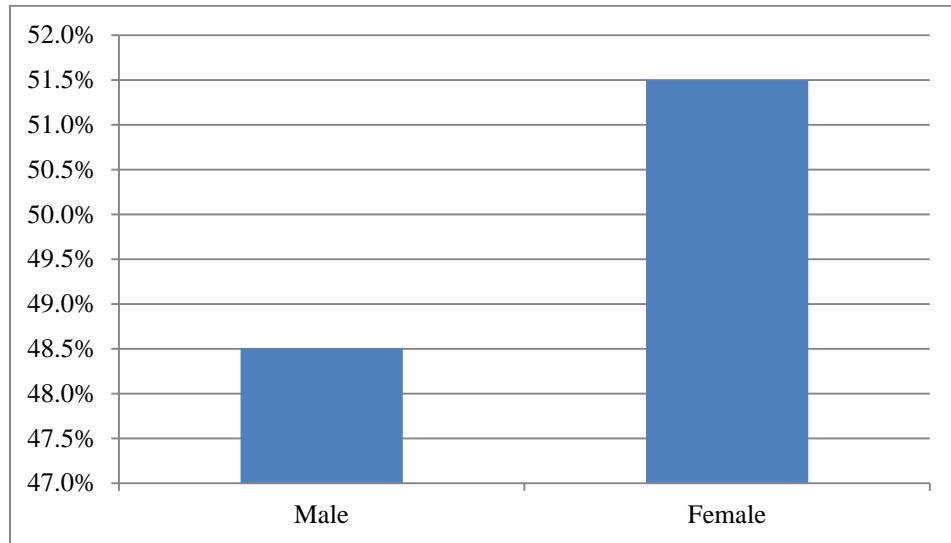


Figure 7-6: Gender distribution of synthetic 2040 population

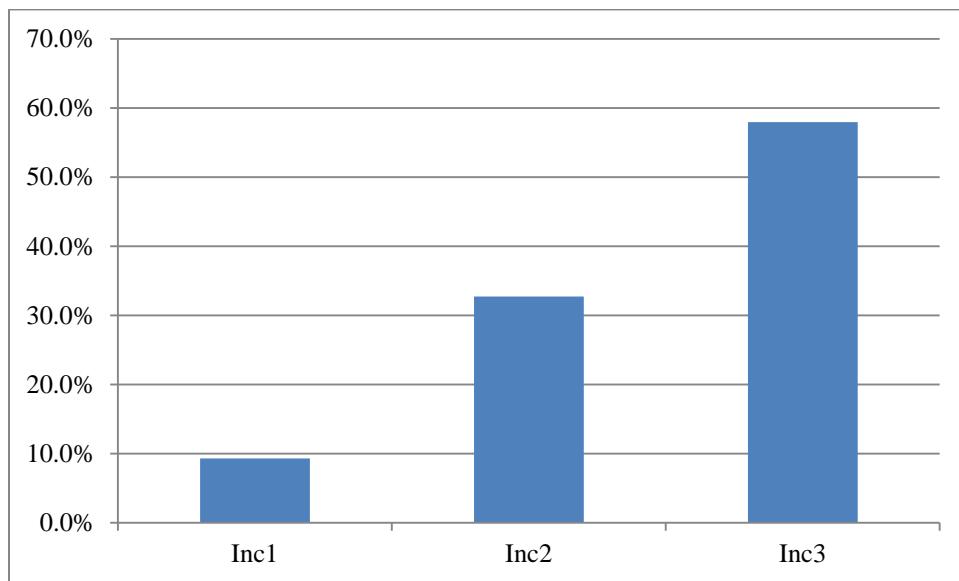


Figure 7-7: Income group distribution of synthetic 2040 population

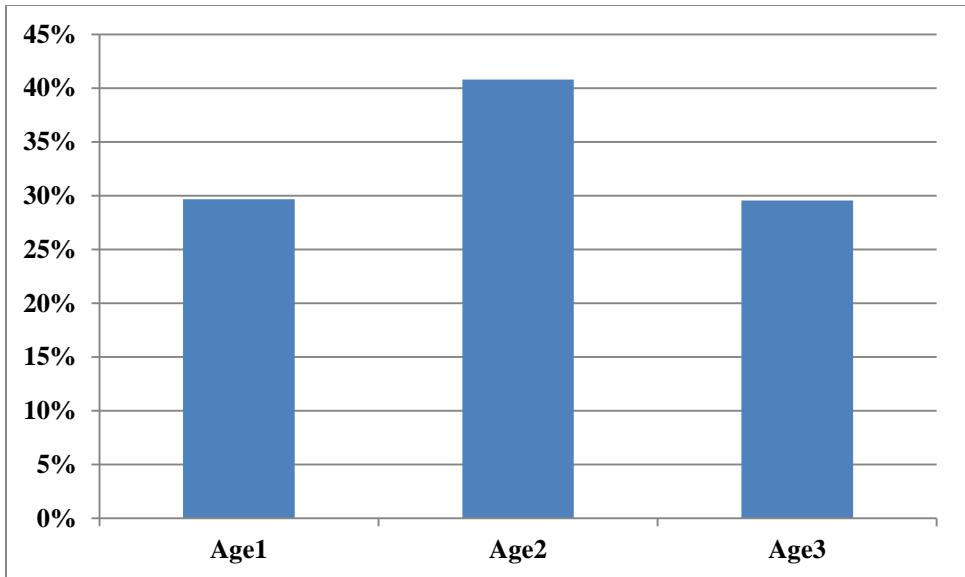


Figure 7-8: Age group distribution of synthetic 2040 population

7.3 Future Year Scenario Results Analysis

Compared to the base year model, the only difference in the future year model is the calibrated model coefficients shown in Table 6-3. All the assumptions in the model simulation are kept the same as the base year model simulation such as one-way travel time should not be larger than the half time of the tour duration, at the stop level the distance between the chosen stop location and the tour origin should be less than the distance between tour origin and tour destination, and etc.

Using the calibrated model parameters, the synthetic population in 2040, and the proposed OD skim data, we can have the long distance travel patterns and trip information in future year for different scenarios. The total number of long distance tours by purpose, travel mode, and time of year could be obtained. As in the national travel demand model the number of long distance activities (tours) is determined by the long distance trip rate (see Table 4-2, Table 4-3, and Table 4-4) and has no

sensitivity to policies and scenarios, we will analyze and compare the results of different scenarios at the level of trips for future year.

7.3.1 Scenario1: Base Scenario

Given the trip OD tables by purpose, time of year, and travel mode, we can have the total number of trips by time of year, travel mode, and purpose. Aggregating all the trips from the trip tables, we can have a total of 5.12 billion trips, which means a person would make an average of 12.5 long distance trips with one-way more than 50 miles during one year.

Figure 7-11 shows the trip distribution by travel mode at the national level, and as expected that the car is the most used transportation mode when people make long distance travel due to the large coverage of highway network and the flexibility of car travel, while the train is the least used mode due to its limited distributions of rail stations and rail tracks. Figure 7-11 presents the trip distribution by time of year. Using the 60-iteration calibrated parameters, the model can predict the trip distribution by time of year ($\{Q1, Q1\}$ -28%, $\{Q2, Q2\}$ -23%, $\{Q3, Q3\}$ -26%, $\{Q4, Q4\}$ -24%) in the right track to the distribution of 1995 ATS data ($\{Q1, Q1\}$ -23.9%, $\{Q2, Q2\}$ -27.5%, $\{Q3, Q3\}$ -28%, $\{Q4, Q4\}$ -20.6%), even though the result shows that most people choose the first quarter to make long distance travel instead of the third quarter. And for different travel modes, the trip distributions by time of year show slight difference (Figure 7-11). With regard to the car mode, it presents a similar pattern with the distribution across all the three modes, with most people choosing the first quarter to travel and then the third quarter. The second quarter is the

least popular quarter that people will choose for long distance travel. Most of the people travelling by air would like to make their long distance travel in the third quarter, and then in the first quarter. The second quarter is also the least popular quarter that people by air will choose. The train trip distribution by time of year shows a same pattern as the distribution for trips by air.

Given more iterations of calibration, the simulated time of year distribution should be close to the 1995 observed distribution. As it takes more than one week to do 60 iterations of calibration, we will explore more in the future work in the group. In this dissertation, we will continue to use the 60-iteration calibrated parameters.

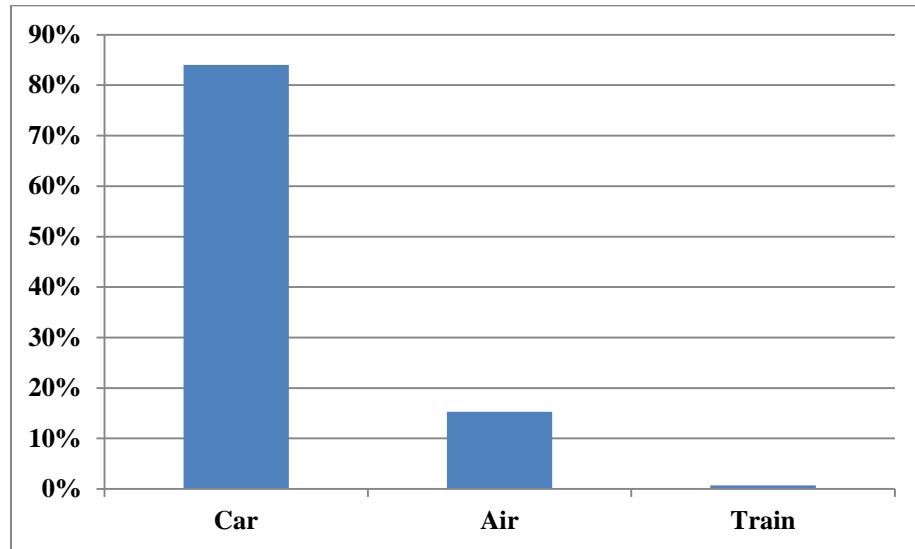


Figure 7-9: Trip distribution by travel mode

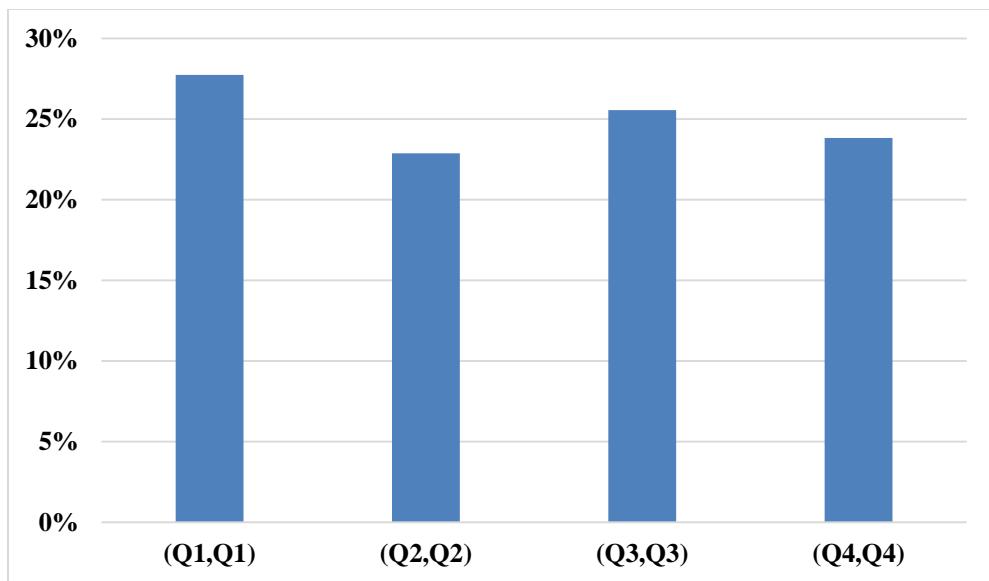


Figure 7-10: Trip Distribution by Time of Year

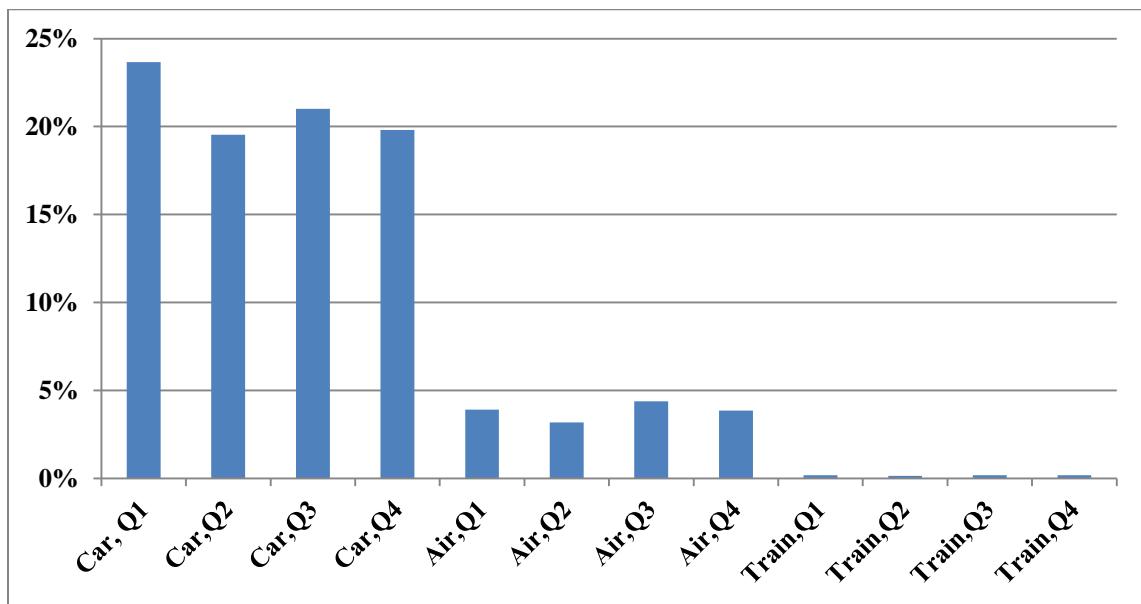


Figure 7-11: Trip distribution by Travel Mode and Time of Year

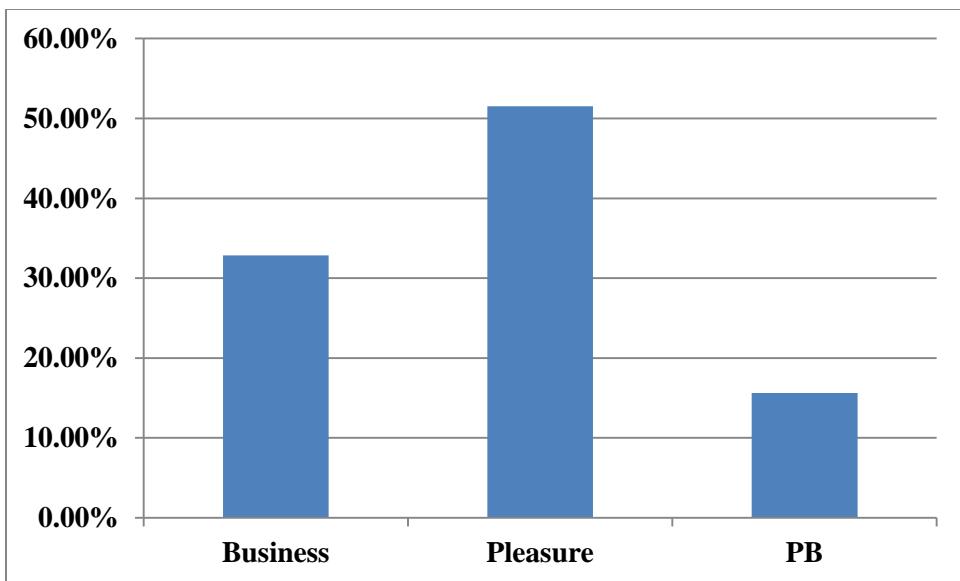


Figure 7-12: Trip distribution by trip purpose

Figure 7-13 and Figure 7-14 shows the trip distribution and average number of trips per person by income group and travel mode at the national level. As expected, high-income group (household income above \$75,000) generate the largest share of trips regardless of the travel mode, and higher income people usually travel more during one year than lower income people. Among the three travel modes, car is always the most popular transportation means for people to make long distance travel for different income groups. During one year, the high-income group people generate a total of almost 2.75 billion car trips, 0.5 billion air trips, and 23 million train trips. In average, a person in the category of high-income group could make 15 long distance trips by car, 2.5 long distance trips by air and 0.1 trips by train. The lower income a person has, the fewer long distance trips by car, air and train he/she will make during one year.

Figure 7-15 and Figure 7-16 shows the trip distribution and average number of trips per person by gender and travel mode. It shows that although female takes up

the larger share of the whole population (above 50.5%), they don't make as many long distance trips as the male. The male population generate a larger number of long distance trips than female regardless of the travel mode. A male person will make an average of above 18 long distance trips per year (almost 15 car trips, 3 air trips, and 0.13 train trips), while a female person would travel less than a male with an average of only 12.5 long distance trips per year (12.5 car trips, 1.9 air trips and 0.1 train trips).

Figure 7-17 and Figure 7-18 represents the trip distribution and average number of trips generated per year by age group and travel mode. The middle-age group people (36 years old – 60 years old) has the largest contribution to the long distance trip generation of all the three travel modes. It is reasonable that the middle-age group people make a lot of long distance trips as most of them have jobs and steady income (which would lead to business trips) and have families (which would lead to more pleasure trips). The young-age group population (18 years old – 35 years old) and old-age group population (above 60 years old) have fewer long distance trips per year for all the three travel modes, compared to middle-age group population, while the old-age group population generates the fewest long distance trips. As expected, a middle-age person travel more frequently per year, with an average of 17.3 trips (14.5 car trips, 2.7 air trips, and 0.13 train trips). A young-age person could make an average of 16.3 long distance trips (13.7 car trips, 2.5 air trips, and 0.12 train trips), and an old-age person make long distance travel the least frequently, with an average of 14.7 trips (12.5 car trips, 2.1 air trips, and 0.1 train trips).

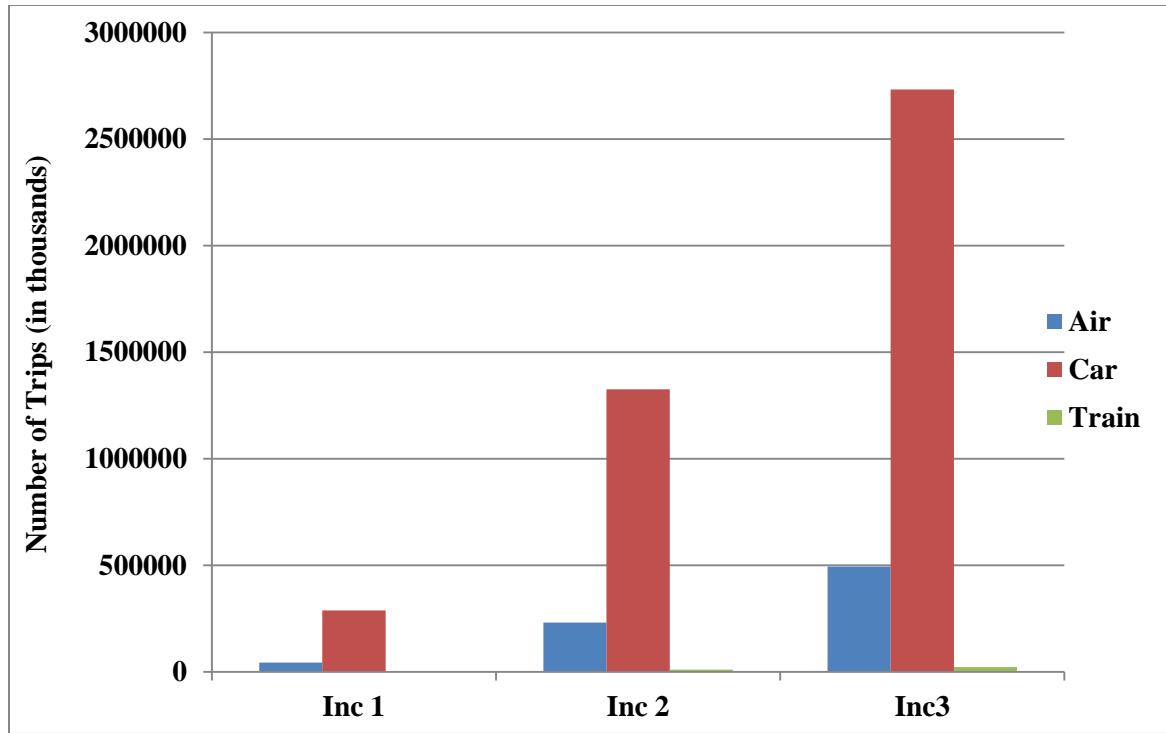


Figure 7-13: Trip Distribution by Income level and Travel Mode

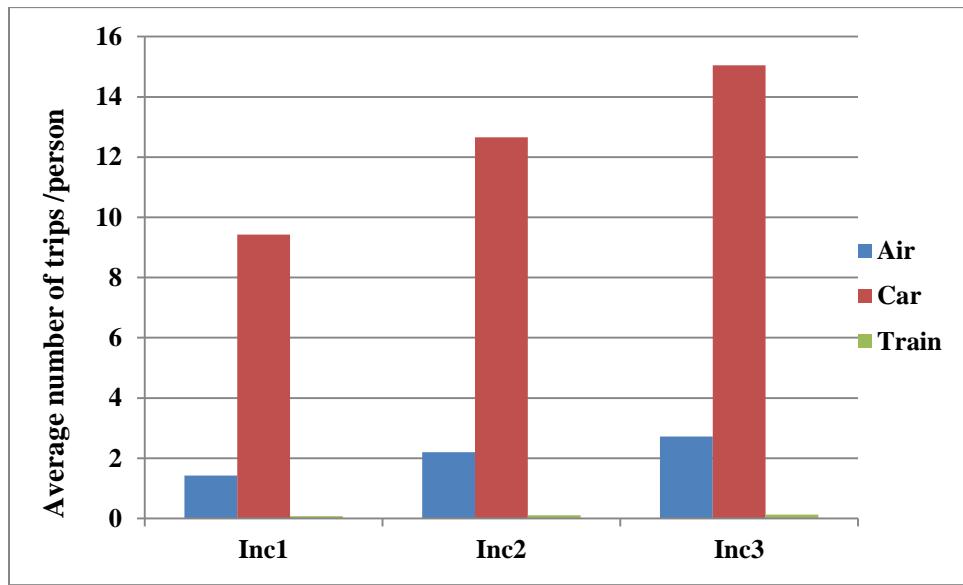


Figure 7-14: Average number of trips/person during one year by income group and travel mode

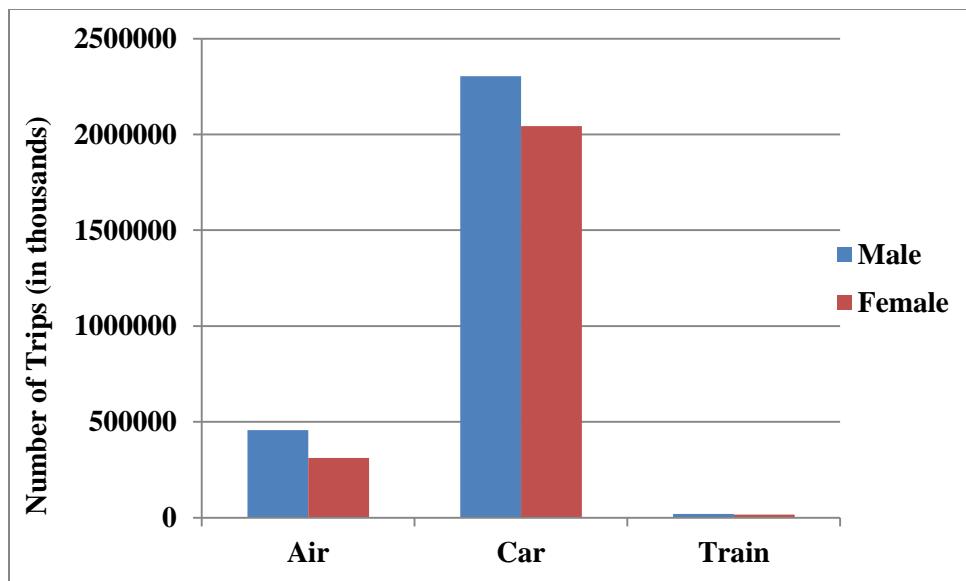


Figure 7-15: Trip distribution by Gender and Travel Mode

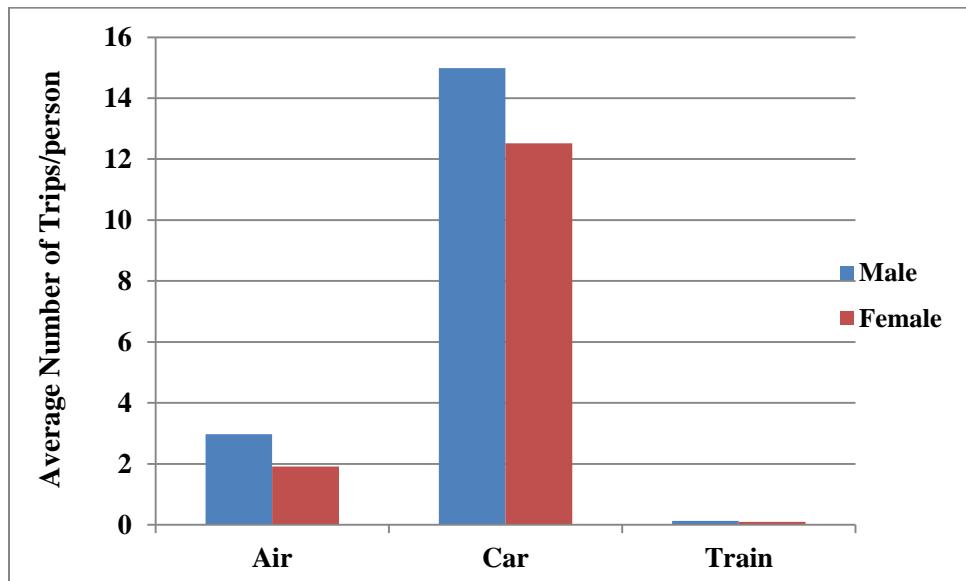


Figure 7-16: Average number of trips/person during one year by gender and travel mode

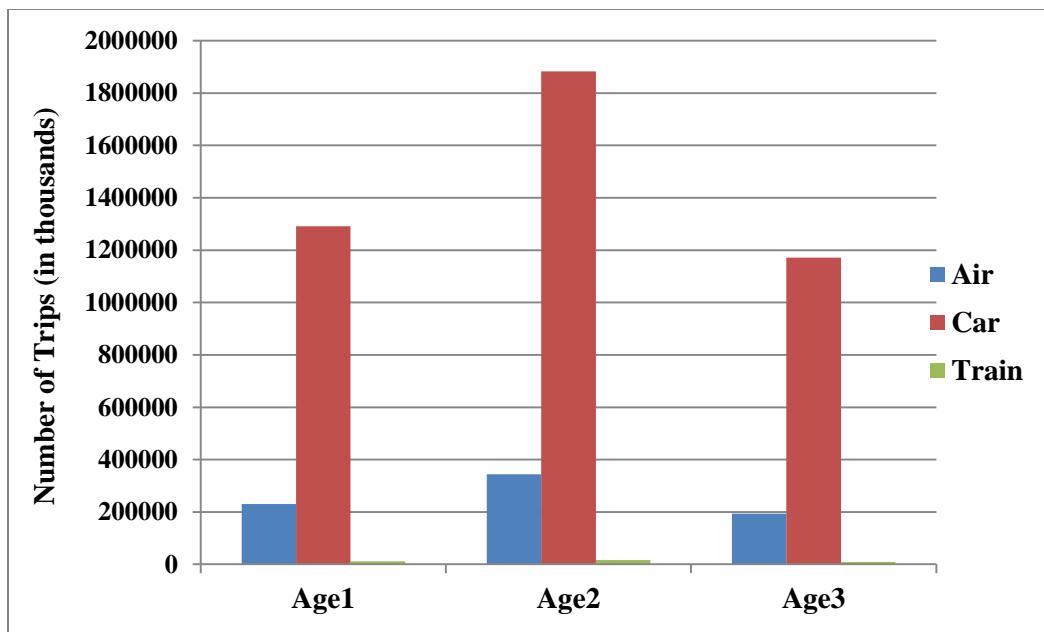


Figure 7-17: Trip distribution by travel mode and age group

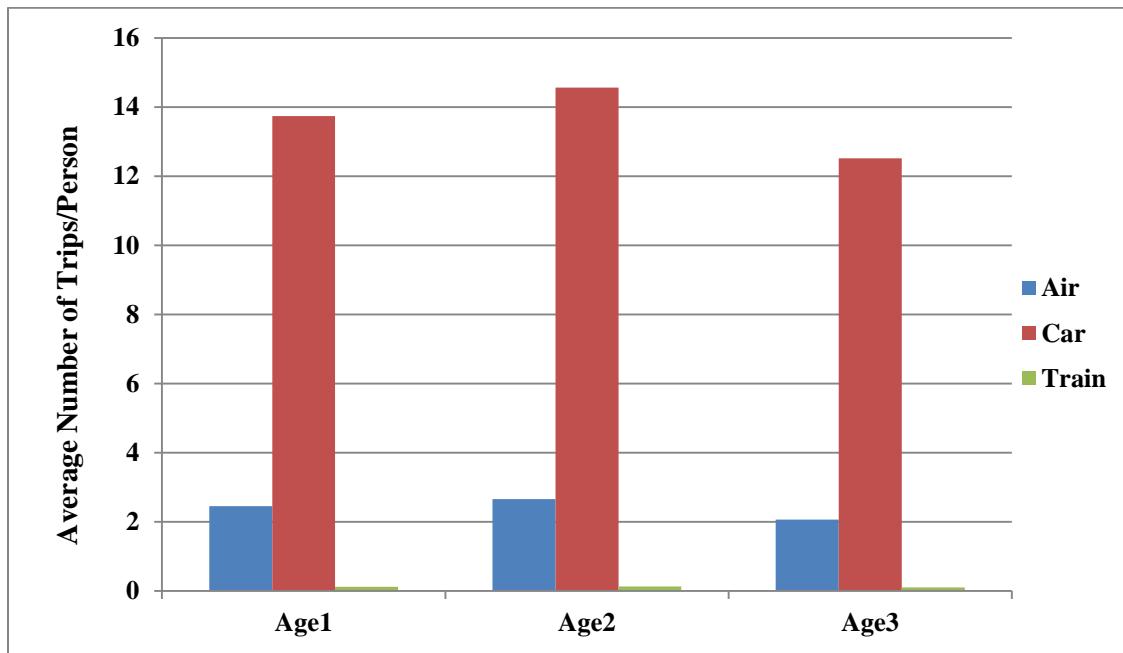


Figure 7-18: Average number of trips/person during one year by age group and travel mode

Disaggregating the national trips into each TAZ, we can have the number of trips (all three travel modes) at the TAZ (MSA/Non-MSA) level, see Figure 7-19. As the figure shows that except some large metropolitan areas like New York,

Washington D.C, Seattle, Los Angeles, Chicago, San Diego, and etc., most MSAs have fewer trips due to their smaller size in geography and population compared to Non-MSAs. The pattern that TAZ with smaller number of population usually have fewer trips than TAZ with large number of population is similar with the 2010 base year results. The zones with large population and high GDP (in east coast cities, Illinois, Ohio, Texas, and some large cities in west coast) usually have a larger number of trips during one year, while the zones in the states of Montana, North Dakota, South Dakota, and Wyoming produce a relatively small number of trips which can be explained by the small size of the population and low GDP of these states.

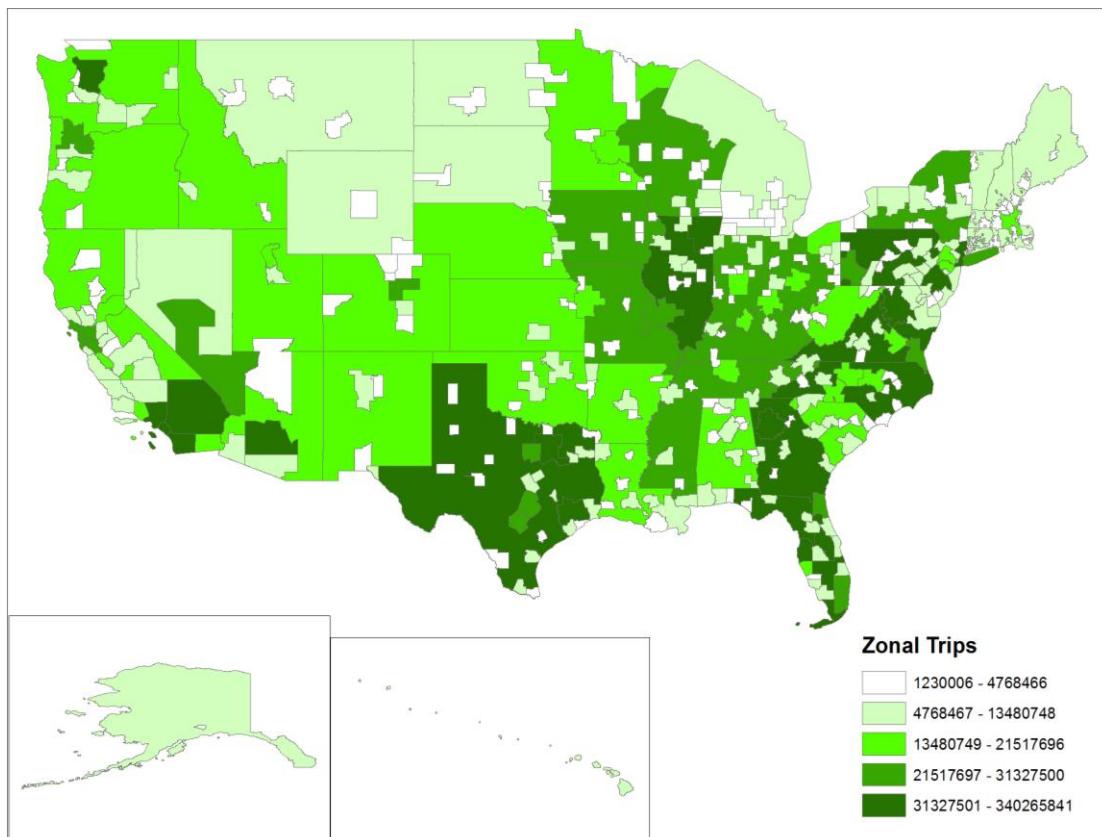


Figure 7-19: Trips Originate/Destinate at MSA/Non-MSA level

7.3.2 Scenario 2: Fuel Price Increase

This scenario assumes that the fuel price will increase which will affect the car driving cost and airfare. Details of this scenario are described in section 7.1. The driving cost increases \$0.064/mile which means that making a long distance trip by car will incur \$3.2 more at least (50 miles). And the longer distance they travel, the more people will pay for their driving cost. The air fare increases \$0.03/mile, and compared to the relatively high base air fare the increase of the air fare is very insignificant. As the number of long distance activities is decided by the pre-calculated trip rate for each purpose, the number of long distance tours by purpose for each person will not be affected by the fuel price increase and it will be the same as that in 2040 base scenario. People's choice at the tour level and stop level will be affected by the fuel price change. The model component that affects the number of long distance trips is the stop frequency model. It decides how many stops a person will make during the inbound/outbound long distance tour legs. The fuel price changes have impact on the number of stops indirectly through the variables (e.g. Car Mode, Time of Year, and Tour Duration) that are output of the high level model components.

Summing all the trips from the trip tables by time of year, purpose, and travel mode, we can have a total of around 5.17 billion long distance trips per year, which is a 0.4% increase from base scenario. As the number of long distance tours for each person during one year will not change from base scenario to the scenario of fuel price increase, the number of trips will be affected by the simulation results of the stop frequency model. In the stop frequency model (Table 4-14 and Table 4-15),

people taking long distance travel by car has a larger impact on making one and two stops than making three and four stops. Therefore, the probability of people choosing three and four stops will increase under the circumstances of not taking car mode, which would directly increase the number of trips by air and by train. And it could make the total number of trips (all of the three travel modes) increase.

Figure 7-20 shows the comparison of trip distribution by travel mode and time of year between base scenario (Base) and fuel price scenario (FP). As we can see that the trip distribution by time of year has the similar pattern as the base scenario for the three travel modes separately. And for car travel in fuel price scenario, the share of trips decreases proportionately for all the time periods, and the percentage of air trips and train trips increase proportionately for the four time periods under fuel price scenario. Consequently, the share of car trips among the total trips at the national level decreases by almost 5% due to the increased driving cost, while the percentage of air and train trips among the total trips increase by 4% and 0.3% (Figure 7-21). The fuel price increase can affect people's choice of the travel modes, while it has little impact on people's trip purpose choice. Therefore, the percentage of trips for different trip purposes doesn't change much (see Figure 7-22). Combining the trip purpose and travel mode, we can generate the trip distribution by purpose and travel mode (Figure 7-23). The total number of car trips decreases under the fuel price scenario, and most of the decreased car trips are for business purpose, and then for personal business purpose. The number of pleasure trips by car increased a little by only 0.59%. In the stop purpose choice model, estimation results indicate that travel mode have a significant impact on people's stop purpose. People do not prefer car driving for

business and personal business trips, and it will reduce the probability of people's stopping for business and personal business during their long distance outbound tour leg. Therefore, as the driving cost increases, people would limit their stops if they travel by car and most of the stops are for business and personal business purposes. During the inbound tour leg, the stop purpose model estimation results show that if people travel by car, it is not likely that they would stop for business and personal business purposes. Therefore, as the number of people travel by car decreases due to fuel price increase, the number of people making stops for business and personal business will decrease as well during the inbound tour leg. It is possible that the number of pleasure trips have a slightly not significantly increase (0.59%) according to the model estimation results.

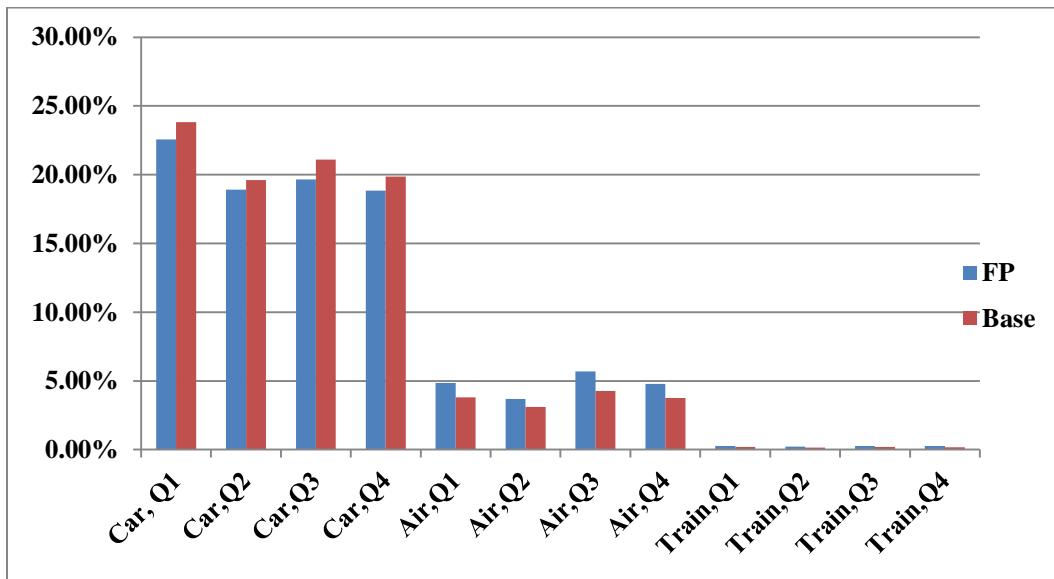


Figure 7-20: Comparison of trip distribution by travel mode and time of year

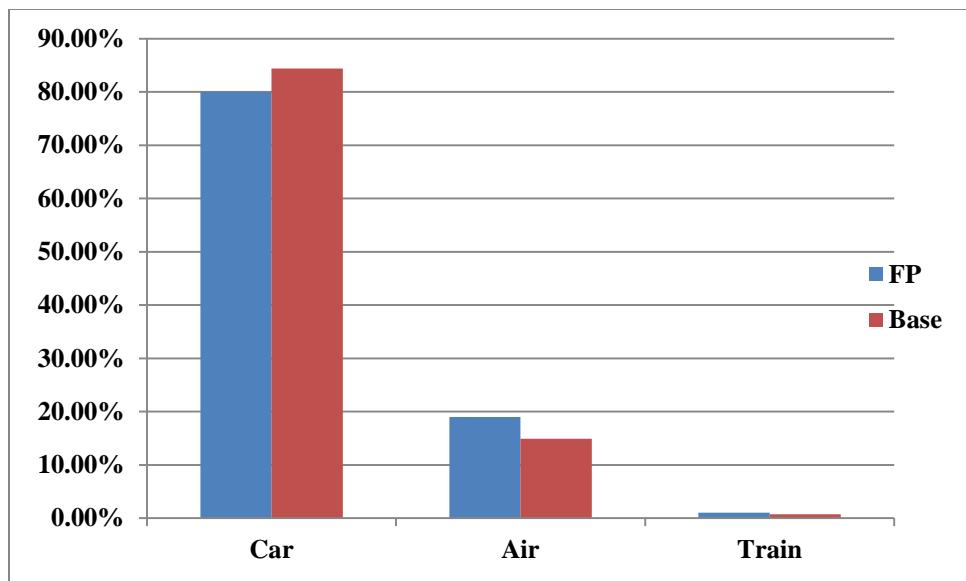


Figure 7-21: Comparison of trip distribution by travel mode

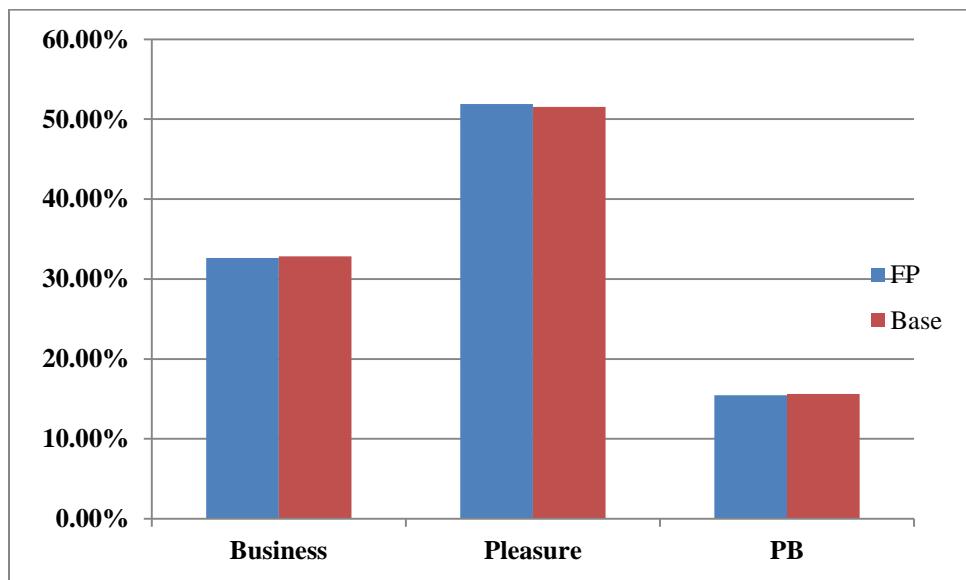


Figure 7-22: Comparison of trip distribution by purpose

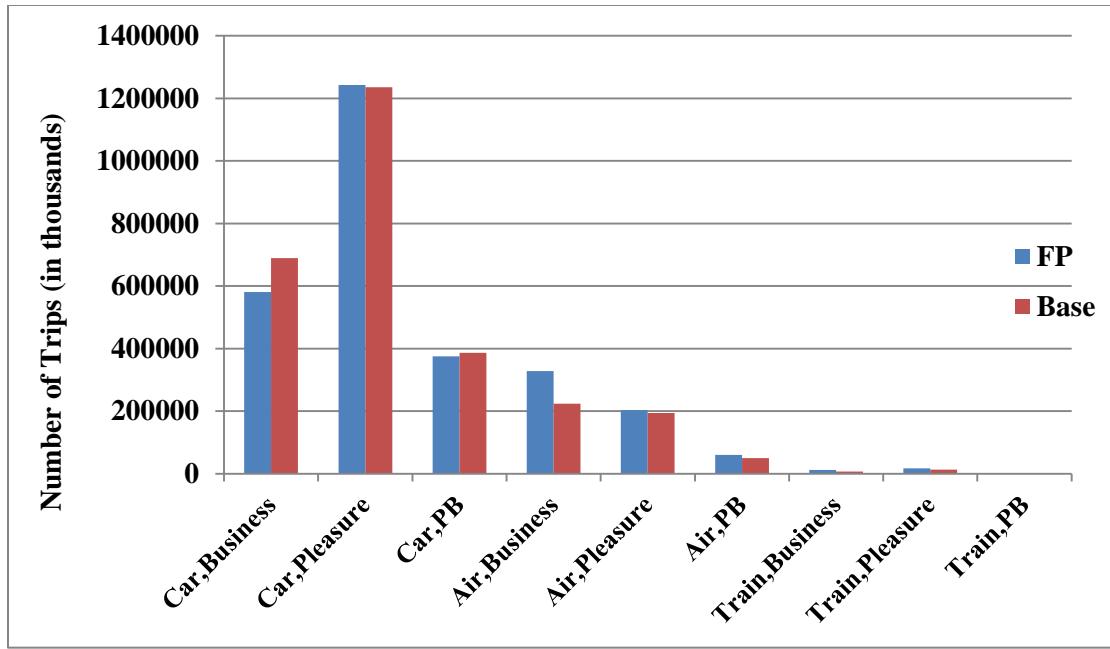


Figure 7-23: Comparison of trip distribution by purpose and travel mode

Table 7-1 presents the numeric comparison of trip changes for time of year, travel mode, and trip purpose between base scenario and fuel price increase scenario at the national level. It is observed that the total number of trips increased a little by 0.4% under the fuel price increase scenario, compared to the base scenario. The number of trips by time of year and trip purpose has little change. The number of trips by time of year changes by less than 1%, and the number of business trips and personal business trips changes by only around 0.1%, while the number of pleasure trips increases by 1.4%. The significant changes occurred for number of trips by travel mode, as the fuel price change directly influence the travel mode choice of people making long distance travel. The number of trips by car decreased by 5% under fuel price increase scenario compared to base scenario, while the number of trips by air and train increased by 28% and 42% respectively. Fuel cost is an important component of the driving cost. As the travel distance is longer, the fuel cost

will become more significant. Fuel price increase will definitely increase the driving cost of long distance travel which will make fewer people choose to travel by car. Compared to fuel cost change for driving and the relatively high air fare, the proposed amount of air fare increase due to the fuel price increase is insignificant. Therefore, it is possible that people will turn to air instead car for long distance travel. Meanwhile, the cost of travel by train doesn't change at all which could lead to a larger percentage of increase in number of trips by train.

	Base		Fuel Price Increase		% difference (FP- Base)/Base
	# of trips	% of trips	# of trips	% of trips	
(Q1,Q1)	1.43E+09	27.81%	1.43E+09	27.68%	-0.10%
(Q2,Q2)	1.18E+09	22.85%	1.18E+09	22.82%	0.26%
(Q3,Q3)	1.32E+09	25.54%	1.32E+09	25.62%	0.72%
(Q4,Q4)	1.23E+09	23.79%	1.23E+09	23.88%	0.76%
Car	4.35E+09	84.38%	4.14E+09	79.99%	-5%
Air	7.68E+08	14.91%	9.83E+08	19.00%	28%
Train	36786395	0.71%	52312913	1.01%	42%
Business	9.21E+08	32.65%	9.2E+08	32.85%	0.1%
Pleasure	1.46E+09	51.88%	1.44E+09	51.54%	1.4%
PB	4.37E+08	15.47%	4.37E+08	15.61%	-0.1%
Total trips	5151015931	100%	5171431217	100%	0.4%

Table 7-1: Comparison of trips between Base Scenario and Fuel Price Increase Scenario

The following figures show the distributional impacts of the fuel price increase on the number of trips by income group, age group, and gender. Figure 7-24 shows that high-income group (Inc3) people altogether reduced more car trips and generated more air trips and train trips than the other two groups when fuel price increases. It is also observed that on an average, a person with high income will cut slightly more car trips and have slightly more air trips and train during one year,

compared to person with low income and medium income. Meanwhile, male population is more sensitive to the fuel price change than female population. Under the fuel price increase, male population will limit their trips by car more than female population (Figure 7-25), and turn to air mode for more trips than female population. Male population will also generate more train trips than female population. The distributional impact of the fuel price increase on age group (Figure 7-26) also shows that the middle-age group people is more sensitive to the fuel price increase than young-age group people and old-age group people. The middle-age group people altogether reduced car trips and increased air and train trips to a larger extent than the young-age group and old-age group people. And generally the young-age group people are less sensitive than middle-age group people but more sensitive than old-age group people. From Figure 7-26, we can see that if fuel price increases the old-age group would also reduce their car trips and increase their air and train trips, but at a minimum degree compared to the other two age groups. According to the base scenario analysis, we can see that the male, the high-income group, and the middle age group are the main population who make a lot of long distance trips per year. Therefore, it is reasonable that when fuel price increases, the total number of trips made by the three groups respectively would reduce/increase more than the peer groups.

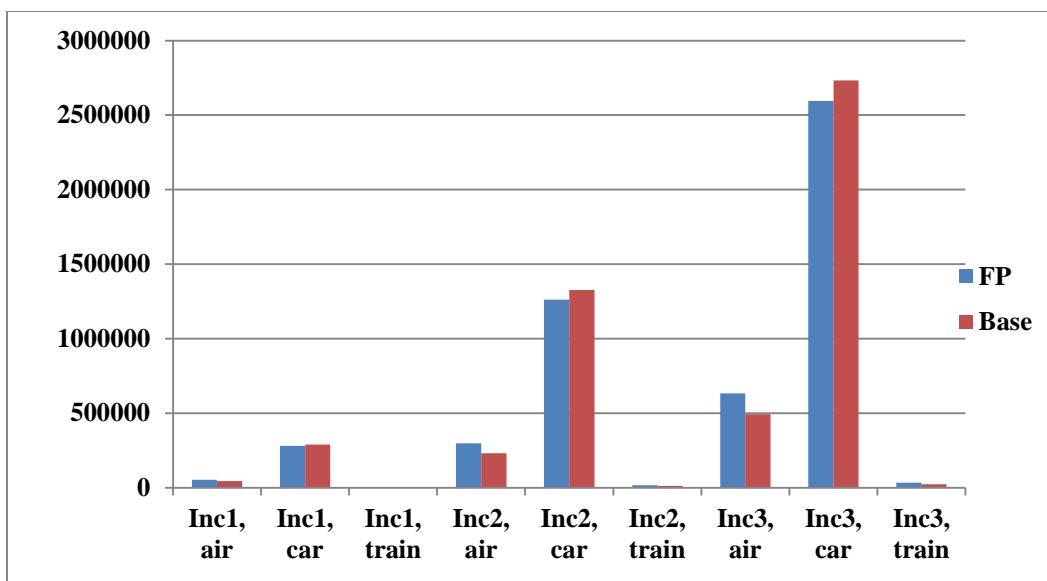


Figure 7-24: Comparison of trip distribution by income group and travel mode

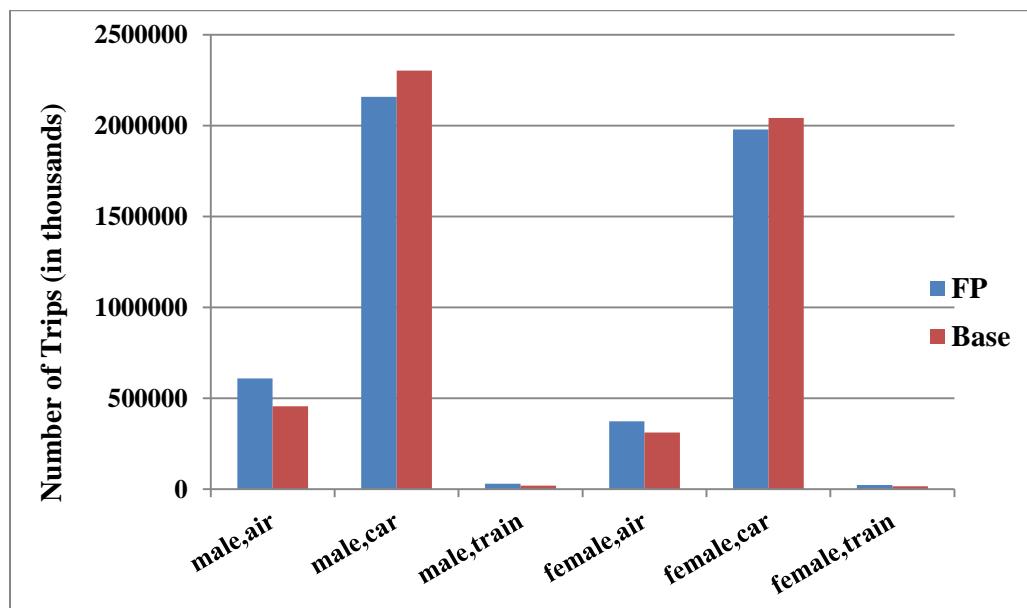


Figure 7-25: Comparison of trip distribution by gender and travel mode

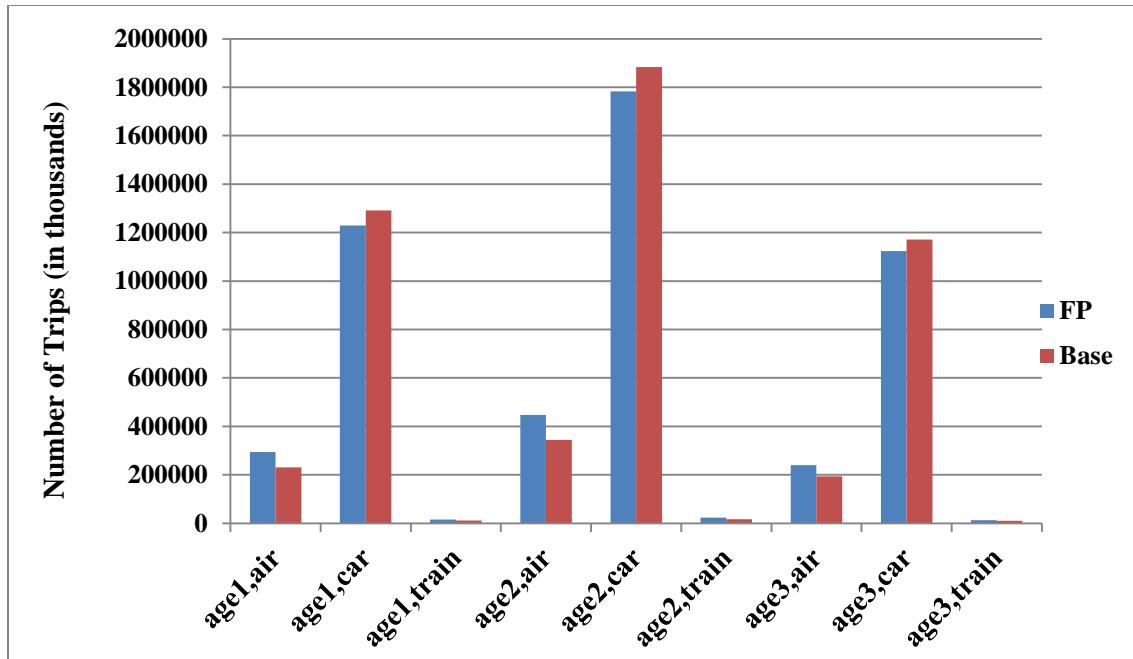


Figure 7-26: Comparison of trip distribution by age group and travel mode

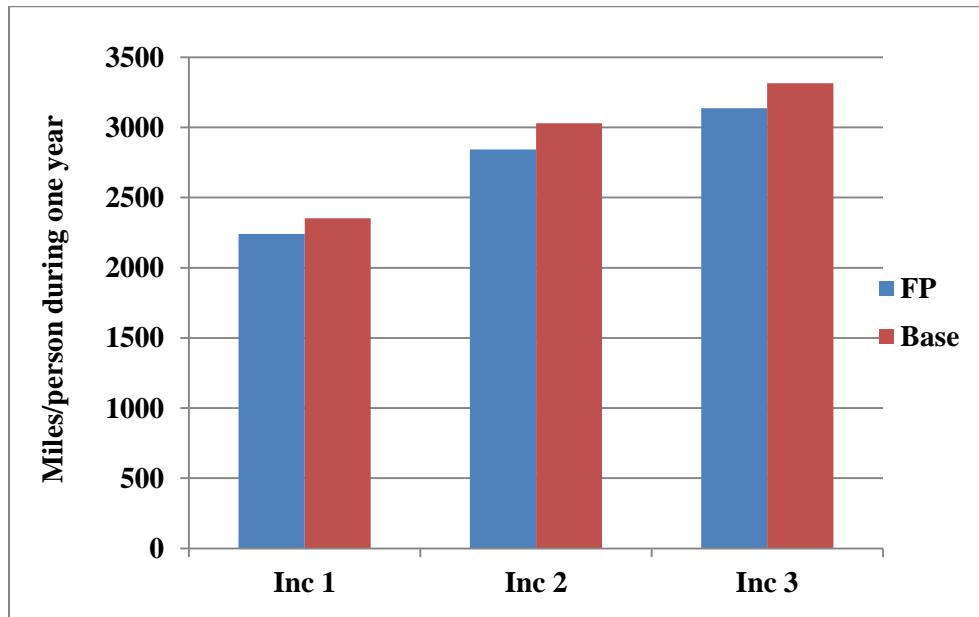


Figure 7-27: Comparison of miles by car per person during one year by income group

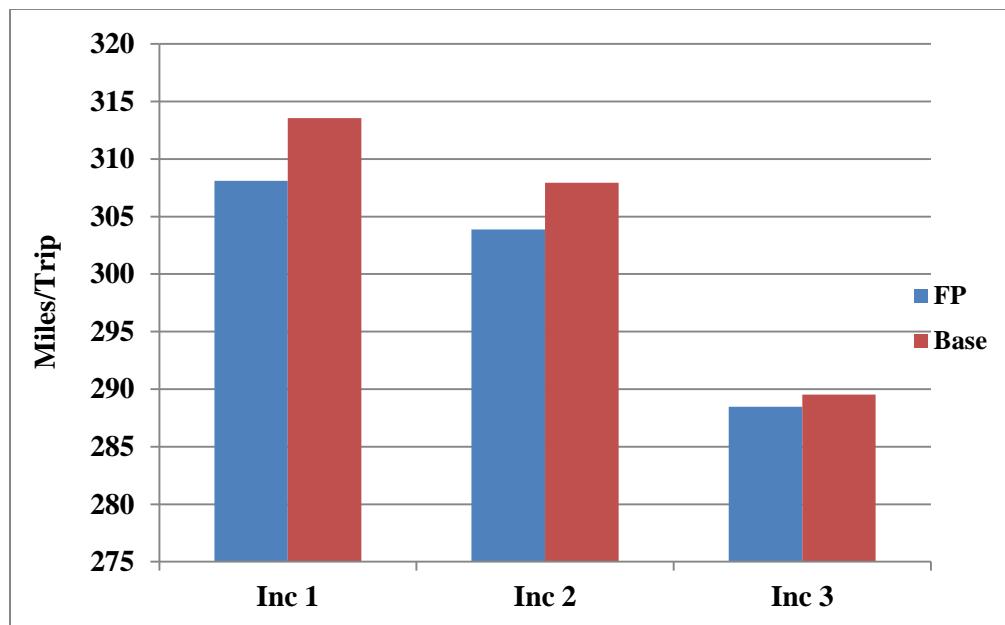


Figure 7-28: Comparison of miles/car trip by income group

Figure 7-27 and Figure 7-28 shows the impact of the fuel price increase on car driving by income group. It is seen that when the fuel price increases, not only the total number of car trips is decreased for all the three income groups but also the average driving miles of each people during one year is shrunk for all groups. Specifically, a person in middle income group would reduce an average of 185.78 miles of car travel per year, while a person in high income group would reduce an average of 176.71 miles and a person in low income group with an average of only 111.46 miles. And a high income person still has the largest average driving miles per year under the fuel price increase scenario. Although high income person generate more car trips than low income and middle income people, their travel distance per trip is shorter (Figure 7-28). The low-income person has the longest travel distance per trip. With the travel distance longer, high income will choose air instead of car for travel, while the low income is less sensitive to the driving distance due to their low income level and high air fare. It will make them more tolerant of the car driving

distance than the high income person, and it also means that low income person will be more sensitive to the driving cost per trip. Consequently, as the fuel price increases and driving cost increases, the low income person would travel less than before, reducing more miles per trip (average of 5.48 miles/trip) than middle income (4.06 miles/trip) and high income (1.05 mile/trip).

7.3.3 Scenario 3: High Speed Rail

In this scenario, the travel time between New York and Washington DC including a stop in Philadelphia is reduced to 96 minutes, travel time between New York and Philadelphia will be 1 hour, and the travel time between Washington D.C and Philadelphia will be 0.6 hours. As the high speed rail provides a higher level of service in terms of the travel time which decreases a lot compared to the regular train travel, a 30% of travel cost increase of the high speed rail (HSR) is also assumed to reflect the improved service. Therefore, all the input data (OD skim data, TAZ economic/demographic data) are the same as the 2040 base scenario except the travel time and cost between New York and Washington DC, New York and Philadelphia, as well as Washington D.C and Philadelphia are altered. As the number of long distance activities each person made is not sensitive to the travel time and cost, the number of long distance activities by purpose is the same as the 2040 base scenario for each person. The tour level and stop level choice could be affected by the changes.

It is expected that with the travel time decrease of HSR, more people would turn to train for their travel instead of car and air. However, as the train network only covers a limited area and people's preference for car and air, train is still not the popular travel mode in the U.S. Since the train travel time decrease only occurs in one

corridor, we don't expect significant changes in the number of car trips and air trips at the national level.

Table 7-2 summarizes and compares the number of trips for different categories between High Speed Rail (HSR) scenario and Base scenario at the national level. As we can see that, the total number of trips has little change under HSR, with trips increasing by only 0.004%. Comparing the number of trips and trip distributions by time of year and purpose for both scenarios, we can find out that the travel time increase of the northeast corridor train service has little impact on the number of trips (trip distribution) by time of year and purpose at the national level. The significant change is observed for the number of train trips. Compared to base scenario, the travel time increase of the northeast corridor train service makes the total number of train trips increased by 5.79%, though the HSR is only open for three lines of the northeast corridor. The HSR we assumed operates between Washington D.C, New York, and Philadelphia, and these three MSA areas are main tourism places with a lot of population and are transportation hubs in the train network which collect to other train lines. Therefore, people would consider taking train as their long distance travel mode, if part of their long distance trips fall in the lines between Washington D.C, New York, and Philadelphia. It will definitely increase the number of trips at the national level.

	High Speed Rail (HSR)		Base		% difference (HSR-Base)/Base
	# of Trips	% of Trips	# of Trips	% of Trips	
(Q1,Q1)	1.43E+09	27.81%	1.43E+09	27.81%	0.00%
(Q2,Q2)	1.18E+09	22.86%	1.18E+09	22.85%	0.00%

(Q3,Q3)	1.32E+09	25.54%	1.32E+09	25.54%	0.00%
(Q4,Q4)	1.23E+09	23.79%	1.23E+09	23.79%	0.01%
Car	4.34E+09	84.34%	4.35E+09	84.38%	-0.04%
Air	7.68E+08	14.91%	7.68E+08	14.91%	-0.01%
Train	38918060	0.76%	36786395	0.71%	5.79%
Business	9.2E+08	32.85%	9.2E+08	32.85%	0.00%
Pleasure	1.44E+09	51.54%	1.44E+09	51.54%	0.01%
PB	4.37E+08	15.61%	4.37E+08	15.61%	0.00%
Total Trips	5151212120	100%	5151015931	100%	0.004%

Table 7-2: Comparison of trips between High Speed Rail Scenario and Base Scenario

As little changes are observed between HSR scenario and Base scenario in trip distribution by time of year and trip purpose for car mode and air mode, only the train trip distribution between HSR scenario and Base scenario are compared for time of year and trip purpose. The number of train trips is increased for all the time periods (four quarters) but by different percentages (Figure 7-29). The train trips in Quarter 1 is increased the most by 6.58% under the HSR scenario, which makes the Quarter 1 becomes the most popular time period for train travel during one year while in Base scenario Quarter 3 is the most popular time period for train travel. The number of trips in Quarter 2 is increased by 6.14%, but it is still the least popular time for travel under HSR scenario. Meanwhile, the number of trips in Quarter 3 and Quarter 4 are increased by 4.95% and 5.56% respectively. Figure 7-30 shows the comparison of train trip by trip purpose between Base scenario and HSR scenario. The operation of HSR in the northeast corridor boosts the train travel and increases the total number of train trips. Among all the three trip purposes (business, person business, and pleasure), pleasure travel is affected by HSR the most and the largest number of trips is

increased for pleasure purpose. And then the business trip follows the pleasure trip as the second affected purpose. The HSR has little impact on personal business travel.

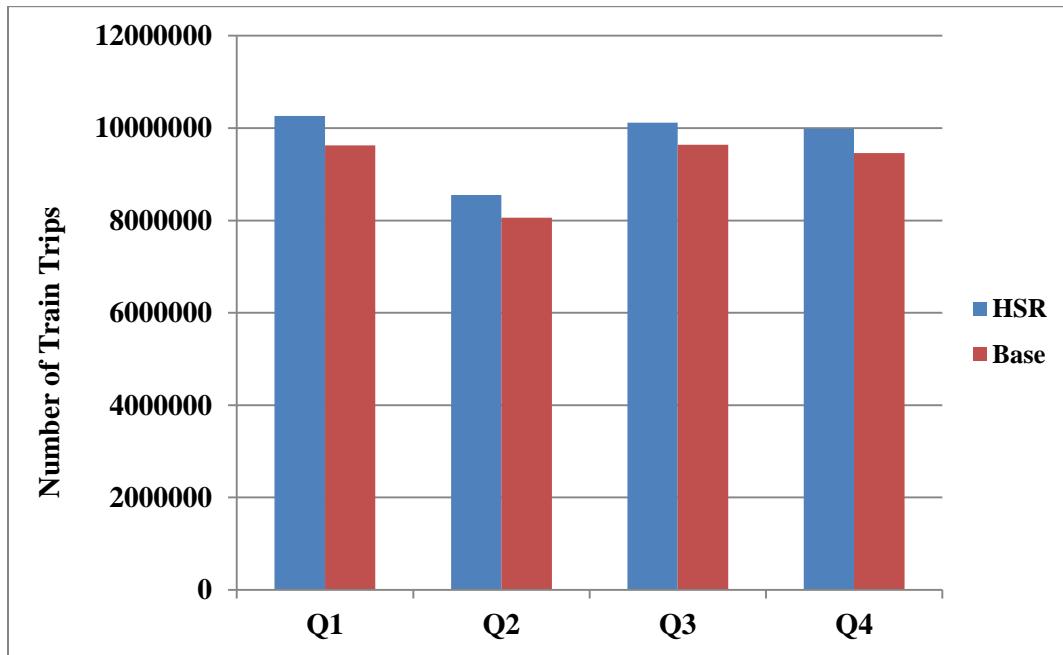


Figure 7-29: Comparison of train trip distribution by time of year

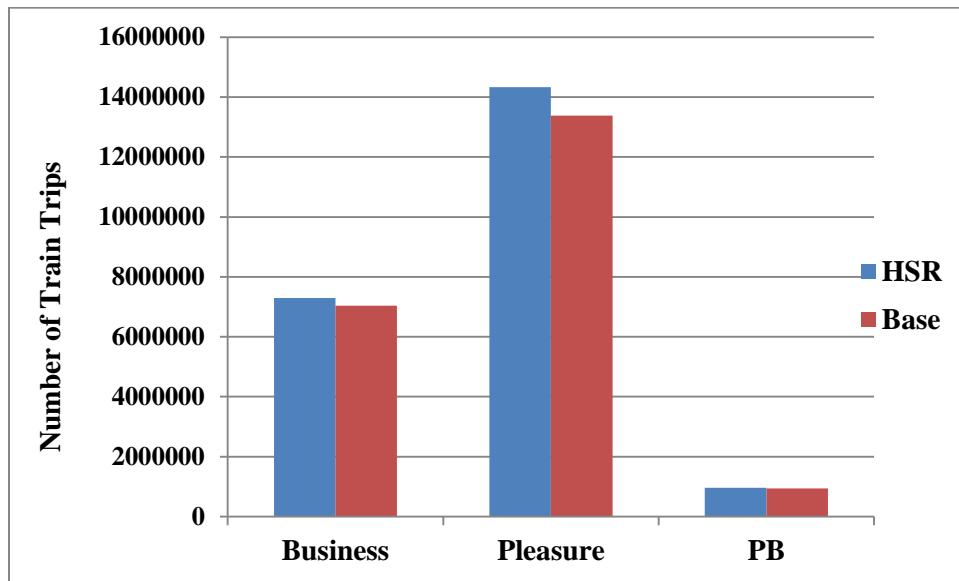


Figure 7-30: Comparison of Train Trips by trip purpose

Analyzing the trip changes by travel mode at TAZ level under HSR scenario, we present Table 7-3 showing the percentage of trip changes by travel mode for TAZs of which the train trips is increased by above 0%. Taking a view of the table, we can find out that the increased train trips mainly is concentrated along the northeast corridor and the TAZs that are connected with or close to Washington D.C, Philadelphia, and New York under HSR scenario. The largest percentage of increase in number of train trips mainly occurred at stations between Washington D.C, MSA and New York, MSA (Philadelphia, MSA, Trenton, MSA), and including D.C, MSA and New York, MSA. Philadelphia has the largest percentage of increase in number of train trips (43.76%), due to its location between the Washington D.C and New York and it connectivity to multiple rail lines. The train trips from/to New York, MSA and Washington D.C are increased by 38.64% and 22.86% respectively. To the contrary, the number of car trips and air trips of the TAZs has little changes, decreasing by less than 0.5%. And generally, the number of trips by car is decreased more than the number of trips by air, which can tell that car travel is the main competitor of train travel.

TAZ Name	Car	Air	Train
	(HSR-Base)/Base		
Philadelphia, PA-NJ MSA	-0.46%	-0.08%	43.76%
New York, NY MSA	-0.47%	-0.11%	38.64%
Trenton, NJ MSA	-0.04%	0.00%	36.17%
Washington, DC-MD-VA-WV MSA	-0.28%	-0.07%	22.86%
Lancaster, PA MSA	-0.03%	-0.19%	20.88%
Wilmington-Newark, DE-MD MSA	-0.05%	0.00%	15.31%
MD Non-MSA	-0.12%	-0.15%	14.15%
Baltimore, MD MSA	-0.01%	-0.03%	13.25%

Bridgeport, CT MSA	-0.05%	-0.01%	11.45%
Harrisburg-Lebanon-Carlisle, PA MSA	-0.05%	-0.12%	9.62%
Atlantic-Cape May, NJ MSA	-0.08%	0.00%	8.79%
PA Non-MSA	-0.01%	-0.038%	2.91%

Table 7-3: Percentage of trip changes by travel mode between HSR scenario and Base scenario

Table 7-4, Table 7-5, Table 7-6 and Table 7-7 show the trips changes among Washington D.C, New York, and Philadelphia by trip purpose and travel mode. From Table 7-4, we can know that the total number of trips (including business, personal business, pleasure, and return to home trips) between the TAZs has more or less decreased, while the number of business, personal business and pleasure trips between the TAZs has increased except the ones from Philadelphia, MSA to Washington D.C MSA. The trips from T361 to T258, for example, means that although the business, personal business and pleasure trips has increased from Washington D.C (T361) to New York (T258), fewer people with tour originating from New York will choose Washington D.C as their last stop back home under the HSR scenario.

TAZ→TAZ	Total	Business	PB	Pleasure
T361-T258	-52522	8109	1889	33156
T361-T276	-14865	24881	1068	5598
T258-T361	-57107	9551	1311	42756
T258-T276	-18437	21808	1888	10314
T276-T258	-15197	13921	2085	7967
T276-T361	-25096	-10658	-319	-2087

Note: T361- TAZ number of Washington, DC-MD-VA-WV MSA; T258- TAZ number of New York, NY MSA; T276- TAZ number of Philadelphia, PA-NJ MSA;
Changes of trips=number of Trips under HSR scenario – number of Trips under Base scenario

Table 7-4: Trip Changes between TAZs by Trip purpose

Table 7-5, Table 7-6 and Table 7-7 show the number of trips changes among

the three TAZs by trip purpose and travel mode. When the HSR is operated along the northeast corridor between Washington D.C and New York in terms of travel time increase in existing rail lines, more people will choose HSR to travel between the three TAZs while fewer people will choose car and air to travel. And the changes in number of trips between TAZs are consistent with the changes of trip from/to the TAZs. Although the total number of trips (Table 7-5) is decreased between the TAZs (Washington D.C, New York, and Philadelphia), the total number of trips by train is increased (Table 7-5). The number of business, personal business and pleasure trips by train is also increased to different degrees. It is observed that the increased train trips between the TAZs are mainly for business and pleasure purposes, and only the smallest share of the increase train trips is for personal business. And a larger number of train trips are increased between Washington D.C and New York, due to the fact that the Washington D.C and New York are two large cities attracting a large number of trips going in and out. Between the pair of Washington D.C and New York and the pair of Washington D.C and Philadelphia, more people choose to travel by train for pleasure purpose, as it is shown that the largest number of train trips is increased for pleasure purpose (179712 trips and 248059 trips). Between New York and Philadelphia, more increased train trips are for business purpose than for pleasure purpose. The percentage of change in number of train trips is also high. The high percentage of changes in train trips between New York and Philadelphia is due to the small value of the base year train trips.

The total number of car trips between the TAZs is observed to be decreased a lot more than the number of air trips between the TAZs (Table 7-6 and Table 7-7), which again indicates that between these TAZs train mode and car mode are mainly compete with each other while air doesn't have many advantages over train and car between these TAZs. The number of car trips (Table 7-6) between Washington D.C and New York is decreased the most, and the largest share of the decreased car trips are for pleasure purpose. The change of car trips between New York and Philadelphia has a similar pattern with the change between New York and Washington D.C but at different degrees. The largest share of the decreased car trips between New York and Philadelphia are for pleasure purpose, followed by business purpose. Although the number of pleasure trips are decreased more than business trips for both pairs (Washington D.C and New York, and New York and Philadelphia), the difference between increased pleasure trips and increased business trips for the pair of Washington D.C and New York is larger than the difference for the pair of New York and Philadelphia. The number of car trips between Washington D.C and Philadelphia is increased a little for business purpose and personal business (only on the side from Washington D.C to Philadelphia). The increase of the car trips is small in terms of absolute value and percentage change. Although the number of car trips is decreased a lot in terms of absolute value (above 40,000 trips) between TAZs, the percentage of change is only no more than 2% due to the originally large number of car trips between the TAZs in base scenario. Very few changed car trips are seen for personal business purpose.

Compared to trips made by car and train, the trips made by air between the TAZs are the least affected (Table 7-7) in terms of the absolute value changes. One of the reasons for this is that the air mode is not the popular travel mode between these TAZs even before the HSR is operated. Although the travel time is increased a lot due to HSR, HSR still cannot attract many trips from air by which a small number of people travel between these TAZs. Among the pairs of the TAZs, the pair between Washington D.C and New York has the largest number of air trips, and the pair between New York and Philadelphia has the smallest number of air trips. It is seen that few people will choose to travel by air between New York and Philadelphia for any purpose even at the base scenario. Therefore even though only few trips are changed (less than 5 trips), the percentage of trip changes is not small (up to 7.1%) due to the small base value. The air trips between New York and Washington D.C are decreased by around 1% (above 4000 trips at the most) for almost all the purposes except for personal business purpose on the rail line from New York to Washington D.C. which has an increase by 0.2% (203 trips). As the travel time of the train is increased, the air line from Philadelphia to Washington D.C loses the largest number of trips (22972 air trips) for business purpose which is about 13% of decrease from base scenario. But in the opposite direction (from Washington D.C to Philadelphia), there is an increase of air trips for business (1578 trips) and personal business (77 trips) purposes by less than 1%.

Looking at the national travel demand model structure and the simulation flow, we found out that the number of long distance activities per year for each person is not sensitive to any of the two scenarios in our research which means that the number

of long distance activities by purpose keeps the same for all the three scenarios. The travel mode is determined before the stop frequency and stop location choice, and will be through the entire tour. Therefore, we can see that as the HSR is operated, more people would choose to travel by train when they travel around the northeast corridor. And once people choose to travel by train for their long distance tours, they are more likely to make stops in the TAZs where the HSR is connected for all purposes. As the main competitor for train in the northeast corridor is the car travel, more people would choose to travel by train instead of car which decrease the number of car trips. And people are generally less likely to stop in the three TAZs if they travel by car.

TAZ→TAZ	Total	Business	PB	Pleasure
	Changes of Train Trips			
T361→T258	367731	37357	2495	179712
T361→T276	151171	10446	801	74760
T258→T361	358812	45753	2693	248059
T258→T276	139599	56758	3767	50108
T276→T258	140480	51489	4371	51308
T276→T361	140548	11985	354	108534
	Percentage of Change in Train Trips			
T361→T258	109.2%	49.5%	34.9%	199.7%
T361→T276	58.2%	35.0%	13.8%	81.5%
T258→T361	107.1%	50.4%	36.2%	226.4%
T258→T276	146.3%	784.5%	108.8%	89.7%
T276→T258	158.1%	1030.6%	131.5%	106.8%
T276→T361	55.3%	31.8%	5.0%	89.1%

Note: Changes of trips=number of Trips under HSR scenario – number of Trips under Base scenario
 Percentage of Change=Change of trips/number of trips under Base scenario

Table 7-5: Changes of number of train trips by trip purpose between TAZs

TAZ→TAZ	Total	Business	PB	Pleasure
	Changes of Car Trips			

T361→T258	-409744	-26987	-92	-144338
T361→T276	-164172	12849	190	-67164
T258→T361	-405003	-33939	-1585	-200790
T258→T276	-157976	-34949	-1879	-39791
T276→T258	-155677	-37570	-2285	-43340
T276→T361	-163108	329	-612	-108782
Percentage of Change in Car Trips				
T361→T258	-2.0%	-0.7%	0.0%	-3.9%
T361→T276	-1.2%	0.4%	0.0%	-2.9%
T258→T361	-2.0%	-0.7%	-0.1%	-4.1%
T258→T276	-1.2%	-1.1%	-0.2%	-1.4%
T276→T258	-1.1%	-1.2%	-0.2%	-1.3%
T276→T361	-1.2%	0.0%	-0.1%	-3.2%

Table 7-6: Changes of number of car trips by trip purpose between TAZs

TAZ→TAZ	Total	Business	PB	Pleasure
Changes of Air Trips				
T361→T258	-10509	-2261	-514	-2218
T361→T276	-1864	1586	77	-1998
T258→T361	-10916	-2263	203	-4513
T258→T276	-60	-1	0	-3
T276→T258	0	2	-1	-1
T276→T361	-2536	-22972	-61	-1839
Percentage of Change in Car Trips				
T361→T258	-0.8%	-0.6%	-0.3%	-0.8%
T361→T276	-0.2%	0.8%	0.1%	-0.8%
T258→T361	-0.9%	-0.6%	0.2%	-1.5%
T258→T276	-6.4%	-1.8%	0.0%	-7.1%
T276→T258	0.0%	5.4%	-4.5%	-4.2%
T276→T361	-0.4%	-13.0%	-0.1%	-1.0%

Table 7-7: Changes of number of air trips by trip purpose between TAZs

Table 7-8 presents the share of the trips by travel mode between TAZs under two scenarios (HSR and Base). Summing the percentage values with the same purpose and same TAZs pair of the three travel modes will give us the total (100%)

trips of the purpose between the TAZ pair. For example, from Washington D.C to New York (T361→T258), the sum of business train trip share (2.6%), business air trip share (9.41%), and business car trip share (87.98%) will be the total business trips percentage from Washington D.C to New York.

Taking an overall view of the table, we can see that although the HSR is operated in terms of travel time increase between the three TAZs, the most popular travel mode is still the car mode. The share of the train trips between the TAZs has more or less increased for all the purposes under the HSR scenario. The share of air trips between the TAZs is decreased a little by no more than 0.5% for all the purposes except that the share of the air trips from New York to Washington D.C for personal business is increased by 0.01%. The share of car trips is also observed to have decreased for almost all the TAZ pairs except the pair from Philadelphia to Washington D.C for business purpose which is increased by 0.24%. After the HSR is operated, the share of the train trips is increased for all purposes between the TAZ pairs. The share of the train trips between some pairs has increased and exceeded the share of the air trips, while most of the pairs keep the original pattern under base scenario. For example, under base scenario, from Philadelphia to Washington D.C, 3.3% of pleasure trips are made by train and 4.79% of pleasure trips are made by air. When the travel time of train travel is increased between the two TAZs, there are more people choosing to travel by train. And there are more pleasure trips made by train (6.24%) than by air (4.74%). For the other pairs, the comparison between the share of the train trips and air trips under HSR scenario keeps the same pattern as under Base scenario. It means although the share of the train trips is increased

compared to base scenario, it is still smaller (larger) than air trips under HSR scenario if it is smaller (larger) under the base scenario.

TAZ→TAZ	The Share of Train Trips							
	HSR Scenario				Base Scenario			
	Total	Business	PB	Pleasure	Total	Business	PB	Pleasure
T361→T258	3.26%	2.60%	0.53%	6.61%	1.55%	1.74%	0.39%	2.23%
T361→T276	2.72%	1.22%	0.81%	6.17%	1.72%	0.91%	0.71%	3.41%
T258→T361	3.24%	2.66%	0.53%	6.70%	1.56%	1.77%	0.39%	2.07%
T258→T276	1.71%	1.94%	0.74%	3.71%	0.69%	0.22%	0.36%	1.96%
T276→T258	1.66%	1.84%	0.54%	2.96%	0.64%	0.16%	0.23%	1.44%
T276→T361	2.70%	1.15%	0.63%	6.24%	1.73%	0.87%	0.60%	3.30%
TAZ→TAZ	The Share of Air Trips							
	HSR Scenario				Base Scenario			
	Total	Business	PB	Pleasure	Total	Business	PB	Pleasure
T361→T258	6.12%	9.41%	8.94%	6.68%	6.15%	9.48%	8.98%	6.79%
T361→T276	4.92%	6.16%	6.36%	9.12%	4.93%	6.16%	6.36%	9.21%
T258→T361	5.72%	6.98%	6.48%	5.69%	5.76%	7.04%	6.47%	5.82%
T258→T276	0.01%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
T276→T258	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
T276→T361	3.87%	3.56%	5.73%	4.74%	3.88%	4.08%	5.74%	4.79%
TAZ→TAZ	The Share of Car Trips							
	HSR Scenario				Base Scenario			
	Total	Business	PB	Pleasure	Total	Business	PB	Pleasure
T361→T258	90.62%	87.98%	90.53%	86.71%	92.29%	88.77%	90.63%	90.99%
T361→T276	92.36%	92.62%	92.83%	84.71%	93.36%	92.93%	92.93%	87.38%
T258→T361	91.04%	90.36%	92.99%	87.61%	92.68%	91.19%	93.14%	92.11%
T258→T276	98.28%	98.06%	99.25%	96.29%	99.30%	99.78%	99.64%	98.04%
T276→T258	98.34%	98.16%	99.46%	97.04%	99.36%	99.84%	99.76%	98.56%
T276→T361	93.43%	95.29%	93.64%	89.02%	94.39%	95.04%	93.67%	91.92%

Table 7-8: Comparison of trip shares by travel mode between HSR and Base scenario

Table 7-9 shows the distributional impact of HSR on number of trips at the national level by income group, gender, age group and travel mode. As we can see that although the number of car trips and air trips decreased to different degrees, the share of the air and car trips for different categories (income, gender, and age) has no

change (by air and car) or slightly decreased (by car). And the share of car trips is decreased by less than 0.05% for different categories. For example, the share of air trips by all the income groups and the share of car trips by low-income and middle-income groups don't have any changes, while the share of car trips by high-income is decreased by 0.03%.

The number of train trips made by high-income group is increased the most, and the number of train trips made by low-income group is increased the least. However, the share of the train trips by low-income and middle-income group has the same change by 0.01% from base scenario to HSR scenario, while the share of high-income group has increased by 0.03%. Among the train trips, the high-income group still has the largest share under the HSR scenario (0.48%). Although the female group has more train trips increased than the male group, the share of the train trips made by female and male increased by the same percentage (0.02%) under HSR scenario. The male group still takes up the larger share of train trips under both base scenario and HSR scenario. The middle-age group (Age2) has the largest share of the trips for all the three travel modes, and the group also has the largest increase in number of train trips among the three age groups. The number of train trips made by old-age group is increased the least. Although the number of train trips made by middle-age group increased more than the young-age group, the share of the train trips by the two groups has increased by the same percentage (0.02%).

As the HSR is operated in terms of travel time increase in the existing northeast corridor rail lines, it has a very small impact on number of trip by income, gender and age group at the national level. The car is still the most popular travel

among all the groups. Although the number and share of the train trips is increased, it still has the smallest share among the three travel modes at the national level.

Category	Base	HSR	Changes in # of Trips
	% of trips	% of trips	(HSR trips – Base trips)
Inc1, air	0.84%	0.84%	-5795
Inc1, car	5.59%	5.59%	-87273
Inc1, train	0.04%	0.05%	103475
Inc2, air	4.48%	4.48%	-21296
Inc2, car	25.74%	25.74%	-393541
Inc2, train	0.22%	0.23%	465063
Inc3, air	9.59%	9.59%	-72507
Inc3, car	53.04%	53.01%	-1355064
Inc3, train	0.45%	0.48%	1563127
Total	100%	100%	-
Male,air	8.86%	8.86%	-61495
Male,car	44.71%	44.69%	-902927
Male,train	0.39%	0.41%	1058576
Female,air	6.05%	6.05%	-38103
Female,car	39.66%	39.64%	-932951
Female,train	0.32%	0.34%	1073089
Total	100%	100%	-
Age1,air	4.47%	4.47%	-39611
Age1,car	25.08%	25.07%	-593362
Age1,train	0.21%	0.23%	689245
Age2,air	6.68%	6.68%	-37891
Age2,car	36.55%	36.54%	-784857
Age2,train	0.31%	0.33%	917319
Age3,air	3.76%	3.76%	-22096
Age3,car	22.75%	22.74%	-457659
Age3,train	0.19%	0.20%	525101
Total	100%	100%	-

Table 7-9: Comparison of percentage of trips and trip changes between HSR and Base

Chapter 8: Long Distance Travel Survey Instrument

The ability to collect and analyze trip data plays a critical role in the success of travel demand modeling at both the statewide and national levels. The most recent sources of long distance passenger trip data in the U.S are the 1995 American Travel Survey (ATS) conducted by the Bureau of Transportation Statistics (BTS) and 2001/2009 National Household Travel Survey (NHTS). However, as known to all, the ATS data are more than 20 years old and have a limited sample size, which will limit its applicability for future long distance passenger travel analysis in the U.S. The Federal Highway Administration (FHWA) has planned the development of a new set of travel surveys – the next round of the National Household Travel Survey (NHTS) and the next iteration of a long distance travel survey of the U.S. household population. It pushes the current survey research by identifying the novel, innovative and cost effective methods to capture the data and improve estimates in future FHWA long distance household travel studies. With the new long distance survey data, we can expect more applications for future long distance passenger travel analysis and conduct the proposed national travel demand modeling research.

The techniques and methods that are utilized in the travel survey have evolved over the past decades. Prior to 1970, most national surveys of the U.S. population were conducted through in-person, interviewer administered methods. Self-administered surveys followed, with widespread adoption of telephone-based survey methods occurring in the 1980s to 1990s. These traditional long distance travel surveys at the household level are capable of collecting most of the information

required for travel analysis and modeling. However, collecting certain travel-related data can place a large burden on respondents at a relatively high cost, and the resulting reporting and measurement errors associated with these data can decrease the overall data reliability. In addition, the low frequency of long distance trips for most households makes it difficult and costly to acquire a sufficiently large sample of long distance trips. Consequently, researchers are turning to advanced travel survey methods using GPS technology, smartphone, social media, etc., to provide the temporal-spatial information on travel more accurately than traditional surveys. The funded FHWA project “Design of a completely new approach for a national household travel survey instrument” designed, developed, and tested new technology and applications to collect survey data that could improve data quality, response rates, respondent burden, and bias reductions. In the project, the long distance travel study continued to feature a probability sample design as its core component. However, the affordability of the probabilistic-based survey will come into question. And the sample size available under these high-priced designs will be unable to achieve the target precision objectives required for the long distance travel study. Therefore, a non-probabilistic-based survey component as an inexpensive supplement to a probabilistic core sample is adopted for data collection in the project (Battelle Memorial Institute et al. 2013). The non-probability sample could provide a large sample size and improve the survey data quality. In the non-probabilistic-based survey, respondent’s burden can be reduced by implementing passive data collection using technologies such as GPS, smartphone, social-media and etc. for triggering the presence of a long distance trip and collecting information in real time on these trips.

However, the non-probabilistic survey methods based on new technologies cannot provide all the necessary long distance trip information such as travel mode, trip purpose, and travel time that are important components in national travel demand analysis. Therefore, the practical post-processing methods that can generate data on these missing travel characteristics are needed to supplement the data that are passively collected from the GPS/smartphone/social media-based survey. Compared to the travel mode and the travel time that can be easily estimated based on the collected spatial-temporal information, the long distance trip purpose needs to be estimated with efforts.

In this chapter, we propose post-processing methods (based on machine learning techniques) that can estimate trip purpose for the non-probabilistic-based long distance travel survey with the new technologies. Available datasets, including travel survey data and other supplementary data, are employed to test and validate the methods. In addition, the research aims to provide the support tool for long distance travel data collection and the sound methodology needed for post-processing the GPS-, smartphone-, and social media-based travel survey data. We also consider alternative trip purpose categorization schemes and the effects of different attributes on trip purpose imputation for long distance travel. Model performance under these alternative schemes and with different attributes is tested in order to provide comprehensive information for the design of future long distance travel surveys.

8.1 Methodology for Long Distance Trip Purpose Classification

The trip purpose detection system consists of four parts: model inputs, learning process, model output, and validation (Figure 8-1). Model inputs include travelers' trip spatial-temporal data, land use data, and individuals' social-demographic and economic attributes provided by travel recall survey if there is recall in the travel survey. The learning process employs machine learning methods and implements trip purpose detection algorithms. Once trip purposes are derived and output from the model that implements these machine learning methods, the validation component evaluates the classifier performance and the reliability of the results.

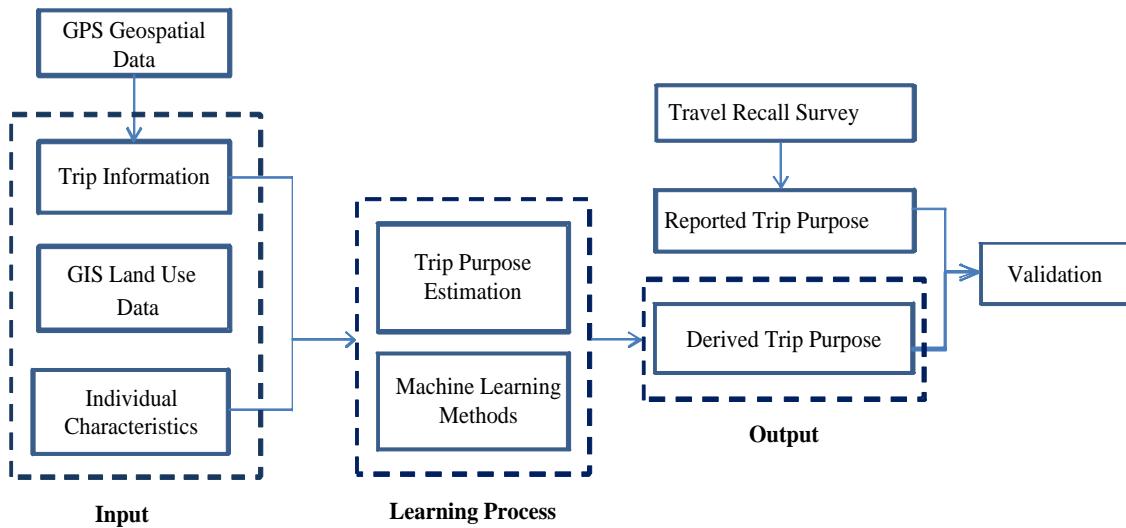


Figure 8-1: Trip Purpose Learning System

In the learning process part of the trip purpose detection system, multiple machine learning methods (e.g. decision tree learning and meta-learning) are employed and tested for trip purpose imputation. The purpose of testing is to identify the classifier that achieves the best performance. Furthermore, alternative trip purpose

categorization schemes for long distance travel have been developed and tested step by step for six different sets of trip purpose classifiers, or “models,” from binary-class to multi-class (Table 8-1). The following sections detail these machine learning methods.

Reported Trip Purpose	Decoded Trip Purpose (# Categories)					
	Model 1 (2)	Model 2 (3)	Model 3 (4)	Model 4 (3)	Model 5 (2)	Model 6 (3)
Business	Business	Business	Business	Business	Business	Business
Combined Business/Pleasure (B/P)	Business	Business	Business	Combined B/P	Non-Business	Pleasure
Convention, Conference, or Seminar	Business	Business	Business	Business	Business	Business
School-related activity	Non-Business	Personal Business	Personal Business	Non-Business	Non-Business	Personal Business
Visit relatives or friends	Non-Business	Pleasure	Social Visit	Non-Business	Non-Business	Pleasure
Rest or relaxation	Non-Business	Pleasure	Leisure	Non-Business	Non-Business	Pleasure
Sightseeing, or to visit a historic/scenic attraction	Non-Business	Pleasure	Leisure	Non-Business	Non-Business	Pleasure
Outdoor recreation	Non-Business	Pleasure	Leisure	Non-Business	Non-Business	Pleasure
Entertainment	Non-Business	Pleasure	Leisure	Non-Business	Non-Business	Pleasure
Shopping	Non-Business	Pleasure	Leisure	Non-Business	Non-Business	Pleasure
Personal, family or medical	Non-Business	Personal Business	Personal Business	Non-Business	Non-Business	Personal Business
Other	Non-Business	Deleted	Deleted	Non-Business	Non-Business	Deleted

Table 8-1: Six Sets of Long Distance Trip Purpose Categorization Schemes

8.1.1 Decision Tree Learning

Decision tree learning involves using a series of input attributes to construct a decision tree for classifying trip purposes (Witten & Frank, 2005). These attributes

include trip characteristics derived from the add-on trip location data, GIS-based land use type, and the individual's social-demographic attributes. The widely used decision tree algorithm in practice is C4.5. It is introduced by J. Ross Quinlan in 1993. C4.5 algorithm utilizes the information gain to split each node. It chooses the attribute at each node that produces the purest daughter node to split on. The information is a measurement of purity. The daughter nodes in the sub-tree will be split based on the same procedure, until all the instances at a node reach the same classification. Given a training data set S and an attribute set A (a_1, a_2, \dots, a_n), the n attributes create branches and partition the training data set S of trip information into n different subdivisions sets (V_1, V_2, \dots, V_n). The number of leaf nodes (L) denoted as v in subdivision V_i varies by the split attribute. The information gain of each attribute in the attribute set A will be calculated and the attribute with the largest information gain will be chosen to split on. The information gain is represented in Formula (8.1).

$$\text{Gain}(S, a_i) = \text{Info}(S) - \text{Average} [\text{Info}(L_1, V_i), \text{Info}(L_2, V_i), \dots, \text{Info}(L_v, V_i)] \quad (8.1)$$

Where $\text{Gain}(S, a_i)$ represents the information gain of the attribute a_i in the data set S . $\text{Info}(S)$ refers to the information value of the data set S . (L_j, V_i) represents the leaf node L_j in subdivision V_i ; $\text{Info}(L_j, V_i)$ is the information value of leaf node L_j in subdivision V_i resulting from the data split on attribute a_i . The term Average $[\text{Info}(L_1, V_i), \text{Info}(L_2, V_i), \dots, \text{Info}(L_v, V_i)]$ on the right hand side in the formula is a weighted average linked to the number of instances at each leaf node. It represents the amount of information expected to be necessary to determine the class of a new

instance, given the tree structure. Therefore, the information gain of each attribute in attribute set A based on data set S can be generated, and the attribute with the largest information gain will be selected to be split on.

Using this basic framework, each attribute in set A would be split recursively so that the information gain can reach the maximum value at each node of the tree, until all the instances at each leaf node will have only one classification.

Pruning a decision tree is a technique that reduces the size of the tree by cutting off some nodes from the tree which have little power in instance classification. Employing pruning in decision tree model can improve the computational efficiency and accuracy, and can reduce the complexity of the tree to avoid the problem of the data set over-fitting as well. The pruning methods applied to the trip purpose decision tree in the research are post-pruning and on-line pruning. Post-pruning, a bottom up pruning strategy, is executed based on a built decision tree. The relative frequencies of leaf nodes are calculated and compared, and any leaf node with dominant classification will result in a pruning of the parent node. Afterwards, error estimates of the replacement node and the old parent node would be compared to evaluate whether the pruning is advantageous. On-line pruning is different from the post-pruning in the time of pruning, and the former one implements pruning while the decision tree is being built. When a split is made on a certain node which we discussed in the Trip Purpose Estimation part, several children leaf nodes will be generated. Once a child leaf node owns less than a minimum number of instances, the parent node and its children leaf nodes will be compressed into a single node. The pruning process continues until the completion of the entire tree.

The method used to estimate the error rate associated with the decision tree learning technique is the 10-fold cross-validation approach. The full sample size is randomly divided into 10 parts, where each part has the same proportion of classes. Every part is held alternately, and the remaining nine parts are trained by the learning algorithm, so that the error rate of the held part can be calculated. The learning procedure is repeated 10 times with different training sets. The overall error rate is equal to the average of the 10 error rates.

8.1.2 *Metaleaner*

Meta-learning implies learning from the learned knowledge. Here, learning occurs from the classifiers produced by the inducers and from the classifications of these classifiers on training data. The main objective of meta-learning is to implement a number of base learning processes on a number of data subsets, and to integrate the knowledge obtained from the separately learned classifiers by an extra level of learning to boost the overall estimation accuracy.

Ensemble methods, which are typically employed for classification and combine multiple base classifiers results, are one type of meta-learning methods. Bagging or Bootstrap Aggregating is the specific ensemble method used in the trip purpose detection system. It builds data subsets by bootstrap sampling, trains the multiple classifiers grounded on these data subsets, and predicts (tests) by majority vote for classification. When larger variance exists in the training data set and the base classifier is over-fitted, the bagging method works well by decreasing the

variance without changing the bias. However, if the base classifier is under-fitting, bagging will not improve the predictive accuracy much.

8.2 Data for Long Distance Trip Purpose Classification

Trip data from the 1995 ATS were used to derive trip purpose imputation. More information about the data can be found in Section 3.1. Information on all trips is employed to help derive the trip purpose. In addition to the primary long distance trip characteristics, additional information including stops to the destination, stops from the destination, and the side trips at the destination are provided. These include the stop location at metropolitan area level and state level, travel mode used to get to the stop, reason for the stop, the number of nights at the stop, and the lodging type at the stop.

Because long distance travel represents a wide coverage area, land use data at the national level is required. Sources of land use data included land use type and intensity at the state, zone, parcel, block and building levels from local, metropolitan and state planning agencies, and graphic/digital land use information and other geospatial information. Because no geo-coded address information is available for trips in the 1995 ATS, land use data at a more aggregate level are adequate and suitable under the premise of providing the destination state or metropolitan area. This research used the NOAA Coastal Assessment and Data Synthesis System as a source of land use data at the national level. It provides the area and the corresponding percentage of different land use types by state. A total of 39 land use types can be combined into 10 land use classes. According to the particular objective

of the research, if the 10 land use classes would yield over-utilization, they are further aggregated into three land use covers including urban, agriculture, and nature.

In order to better derive long distance trip purpose, supplementary data such as travel and tourism statistics data as well as Gross State Product (GSP) data were collected and employed. It is hypothesized that people who travel to states having a higher travel and tourism population are more likely to travel for pleasure and visiting. Similarly, states having a higher GSP tend to have more enterprises and easier accessibility, leading to a higher possibility of attracting business trips. The travel and tourism data were obtained from the U.S. Census Bureau. They include yearly recreation visits in national parks and state parks by state. Meanwhile, the Gross State Product data in 1995 were obtained for each state from the Bureau of Economic Analysis.

To derive trip purposes for long distance travel, various model input variables were proposed in four categories: trip-related variables, respondent characteristics, land use attributes, and other supplementary data. A list of the model variables is found in Table 8-2.

Variable Name	Description
HHIncome	Household Income
Age	Respondent's Age
Race	Respondent's Race
EducAttainment	Respondent's education level
Activity	Activity of Respondent
TrParty	Travel party size
TrPrHousePercent	Percentage of Adult Household Members in Travel Party
TrPrTyCh	Children Under 18 Years in the Travel Party
Weekend	Whether it's a weekend trip
NiteDest	Number of nights at destination
LodgDest	Lodge type at destination

TransportOriginDest	Principal Transportation from Origin to Destination
InternationalDestFlag	U.S. or International Destination Flag
StopsTo	Number of Stops to Destination
SideTrps	Number of Side trips
Sex	Respondent's gender
Side1state	Side trip 1 destination locates in the same state as the main trip or not
SidetripDest1Lodgn	Lodge type at side trip 1 destination
SidetripDest1Reasn	Trip purpose of side trip 1
SidetripDest1Transportation	Transportation mode to side trip 1 destination
DestRegion	The region where the destination state falls in
Tourism	National Park recreation visits by state
GSP	Gross State Product
Urban	Percentage of urban land use cover by state
Agriculture	Percentage of agriculture land use cover by state
Nature	Percentage of natural land use cover by state

Table 8-2: Model Variables Used for Long Distance Trip Purpose Estimation

8.3 *Trip Purpose Classification Results*

Using the 1995 ATS data, models from binary class to multi-class were developed to estimate trip purpose and provide methodological sound support to assist the design of a GPS-, social media-, and smartphone-based long distance travel survey. The six different classifiers, or “models” defined in Table 8-1 were tested, and the classifier yielding the highest level of classification accuracy was noted.

Model 1. First, a binary classification model was developed in which long distance trip purpose was classified as either business or non-business. The results of the classifier with the highest classification accuracy are shown in Table 8-3. The classification results are encouraging: non-business trips are correctly identified with 96.1% accuracy, while for business trips, the accuracy level is 70.1%. Meanwhile, the results imply that the non-business trips are over-predicted with over twice the

number of business trips wrongly classified into non-business trips compared to vice versa. Overall, Model 1 successfully estimates trip purposes for 90.31% for all long distance trips in the 1995 ATS.

Non-business	Business	Actual Purpose	TP Rate
415,473	16,950	non-business	96.1%
36,932	86,671	Business	70.1%
Overall Accuracy: 90.31%			
Number of Leaves: 27,473; Size of the tree: 35,643			

Table 8-3: Model 1 Results

Model 2. Trip purpose imputation models that considered more than two categories were also tested. The majority of long distance travel models involve only three trip purpose categories, usually along the lines of business, pleasure (leisure/vacation), and other personal purposes. In Model 2, trips involving a combination of business and pleasure were regarded as business trips, while non-business trips were split into pleasure and personal business trips. The best classifier results for these three trip purposes are presented in **Error! Reference source not found.** An overall predictive accuracy of 81.87% was achieved, with pleasure trips acquiring the highest performance of 91.5% and personal business trips obtaining the lowest accuracy of 51.7%. Almost half of the personal business trips are wrongly classified as pleasure trips, leading to the under-prediction of personal business trips. Moreover, as the number of trip purpose categorization increases from binary to three-class (i.e., from Model 1 to Model 2), the decision tree grows much larger.

Pleasure	Business	Personal Business	Actual Purpose	TP Rate
315,520	17,032	12,328	Pleasure	91.5%

25,656	94,419	3,528	Business	76.4%
35,955	6,286	45,276	Personal Business	51.7%
Overall Accuracy: 81.87 %				
Number of Leaves: 91,241; Size of the tree: 108,145				

Table 8-4: Model 2 Results

Model 3. A model with four trip purpose categories was tested to examine the impact of having more than three trip purposes in the trip purpose classification. In Model 3, the decoding structure for business and personal business trips remains the same as that in Model 2, while the pleasure trips are further split into leisure and social visit trips (Table 8-1). Table 8-5 shows the results of the four-trip-purpose imputation. The overall accuracy decreased from 81.87% (Model 2) to 76.98%. Compared to having only one pleasure trip category, the breakout of pleasure trips into leisure and social visit categories appears to deteriorate the predictive accuracy of pleasure trips. Meanwhile, the lowest classification performance is still seen with personal business trips.

Leisure	Social Visit	Business	Personal Business	Actual Purpose	TP Rate
136,656	14,359	13,395	8,665	Leisure	79%
13,871	144,403	6,519	7,012	Social Visit	84.1%
14,544	7,430	97,597	4,032	Business	79%
14,894	16,231	7,051	49,341	Personal Business	56.4%
Overall Accuracy: 76.98 %					
Number of Leaves: 137,660; Size of the tree: 163,883					

Table 8-5: Model 3 Results

Model 4. Another 3-trip-purpose scheme was tested which involved business, non-business and combined business and pleasure (B/P) trip categories. The classification results are shown in **Error! Reference source not found.** Overall

curacy was at 90.22%. Among all trips, the combined B/P trips have the weakest predictive power with only 30.50% of accuracy, due primarily to nearly 60% of these trips being classified as non-business trips resulting in the under-predication of combined B/P trips.

Non-business	Business	Combined B/P	Actual Purpose	TP Rate
417,072	14,288	1,063	Non-business	96.50%
28,740	80,249	397	Business	73.40%
8,557	1,330	4,330	Combined B/P	30.50%
Overall Accuracy: 90.22%				
Number of Leaves: 30,342; Size of the tree: 38,672				

Table 8-6: Model 4 Results

Models 5 and 6. Two additional models were developed and evaluated which were slight alterations to Models 1 and 2, to assess the extent to which changing the decoded classification of combined B/P trips led to improved classification performance. In both Models 1 and 2, combined B/P trips were to be classified as business trips. In Model 5, combined B/P trips were to be classified as non-business trips, while in Model 6, combined B/P trips were to be classified as pleasure trips. Due to the uncertainty of the pleasure part in the combined B/P trip, it was deemed risky to define the combined B/P trip as either a leisure trip or a social visit trip for the four-trip-purpose scheme. Therefore, a similar alteration of Model 3 was not considered.

Model 5 represented an alternative binary classification of business versus non-business trips. The results of this model are represented in Table 8-7. An overall accuracy of 91.86% is achieved, which is 1.5% higher than the classification accuracy observed for Model 1.

Non-business	Business	Actual Purpose	TP Rate
432,079	14,561	non-business	96.7%
30,693	78,693	business	71.9%
Overall Accuracy:91.86%			
Number of Leaves: 24,109; Size of the tree: 30,120			

Table 8-7: Model 5 Results

A reconstructed business, pleasure, and personal business trip scheme was utilized to learn the three trip purpose classification (Model 6). The results (Table 8-8) indicate that the predicative accuracy is increased from Model 2 to a small extent (from 81.87% to 82.82%), when the combined B/P trips are coded as pleasure trips.

Pleasure	Business	Personal Business	Actual Purpose	TP Rate
332,435	15,029	11,633	Pleasure	92.6%
22,158	84,496	2,732	Business	77.2%
38,696	5,249	43,572	Personal Business	49.8%
Overall Accuracy: 82.82 %				
Number of Leaves: 86,176; Size of the tree: 100,786				

Table 8-8: Model 6 Results

Travel surveys using advanced passive tracking device such as GPS and smartphone cannot record information that requires the respondents' interaction. With such tracking devices, the data will just be collected; there is no further reference to the user. Under this circumstance, the trip purpose can be derived only based on the passively collected spatial-temporal data or the passively collected data combined with the respondent's social-economic information. In order to provide comprehensive information for future travel survey designs, and to evaluate the effect of different categories of travel information on trip purpose derivation, multiple binary classifiers of business versus non-business are developed and re-estimated

(Table 8-9). To overcome the bias towards one class caused by the imbalanced data set, we augmented the sample size of the underrepresented class (business trip) by duplicating the training examples.

Full Model. The binary classifier was learned with all kinds of information fed into the model. The results are represented in Table 8-10. The overall classification performance of the model can reach 94.57%, with 2.71% increment in accuracy compared to the model based on the original sample size (91.86%).

Variable Sets	Full Model	Reduced Model	Minimized Model
Passively Collected Spatial-temporal Data	√	√	√
Respondents' Characteristics and Other Supplementary Variables	√	√	
Information based on Respondents' answers in travel survey	√		

Table 8-9: Compared Model Developments

Business	Non-business	Actual Purpose	TP Rate
427302	10242	Business	97.7%
37790	408850	Non-business	91.5%
Overall Accuracy:94.57%			
Number of Leaves: 79949; Size of the tree: 99695			

Table 8-10: Full Model Estimation Results

Reduced Model. Without the respondents' interaction in the travel survey, the trip purpose can be derived based on the passively collected spatial-temporal data as well as the respondent's social-economic characteristics and other supplementary information. The results of the reduced model are shown in Table 8-11. The overall

classification accuracy is decreased by above 3% from full model to reduced model, and the size of the decision tree grows.

Minimized Model. The final model is developed based on passively collected spatial-temporal data only. The attributes used in the model consist of two parts. One part includes the attributes which can be directly used in the model after some simple post-processing of the passive collected data, such as number of stops, number of side trips, travel start time, etc., while the other part includes the attributes, such as travel mode, lodge type at destination, etc. which need to be derived or estimated based on certain algorithms or established models. Table 8-12 shows the results of the minimized model. The overall classification performance (71.72%) is reduced a lot by more than 20%, compared to the full model. Meanwhile, the tree size is dramatically decreased.

Business	Non-business	Actual Purpose	TP Rate
415048	22496	Business	94.9%
55855	390785	Non-business	87.5%
Overall Accuracy:91.14%			
Number of Leaves:92855 ; Size of the tree:116914			

Table 8-11: Reduced Model Results

Business	Non-business	Actual Purpose	TP Rate
356522	81022	Business	81.5%
169014	277626	Non-business	62.2%
Overall Accuracy:71.72%			
Number of Leaves:1766; Size of the tree:2466			

Table 8-12: Minimized Model Results

Estimation results show that in general, as the number of categories increases, the accuracy of a trip purpose imputation scheme tends to deteriorate, and the decision tree is inclined to become more complex. We found that non-business trips or pleasure trips can achieve satisfactory results with higher classification accuracy than business trips. Moreover, based on the comparison of the results of different trip purpose categorizations, it is more appropriate to decode the reported combined business and pleasure trips to non-business trips for binary classification and to pleasure trips for three-class classification. Unsatisfactory results can be seen for business trips and personal business trips, which can be explained by the reported errors that tend to be inevitable in the traditional travel survey, as well as the sharing of similar characteristics between personal business trips and pleasure trips (e.g., travel party, travel mode, lodging type of destination, duration). More information about respondents' travel at the destination at urban level and detailed land use data would be helpful to better distinguish business trips and personal business trips from other types of trips based on high-quality travel survey data.

Three additional models were developed and re-estimated to examine the role of different kinds of information in long distance trip purpose imputation, and to assist future long distance travel survey designs. The full model considers all available information including passively collected spatial-temporal data, respondents' socio-economic and demographic characteristic, and other supplementary data such as land use attributes. The reduced model does not require information that can only be collected through travel surveys (e.g., travel party size), but still uses respondents' socio-economic and demographic characteristics and GPS or other location data. The

minimized model is only based on passively collected spatial-temporal data. As expected, the full model reaches the highest performance accuracy (94.57%) based on the augmented sample size. The classification performance decreases gradually with more variable sets excluded. More than 3% decrease can be achieved by the reduced model excluding the information that requires respondents' interaction in the travel survey. While only the passive collected data is employed to estimate the trip purpose, the overall classification performance is reduced by above 20% compared to full model. Consequently, it can be concluded that the travel information which collected from respondents' interaction can affect the long distance trip purpose estimation to a small degree. The respondents' characteristics and other supplementary information can help improve the trip purpose classification accuracy significantly.

Chapter 9: Conclusions and Future Research

9.1 Conclusions

The needs to support high level personal long-distance national travel requires that we have accurate analysis tools to be able to understand the long-distance travel behavior and forecast the travel patterns in the future. However, national long-distance passenger travel demand analysis has been an understudied area in transportation planning.

This research demonstrates a more advanced academic research endeavor for national passenger travel analysis. It aims to provide important insight and help guide federal and state to make decisions on corridor-level, region-level, and nation-level infrastructure investment, design, and management, as well as to research on long-distance passenger travel demand. The developed national travel demand model exhibits the system logic and concept, statistically supports its basic structure, and provides OD estimations based on the model simulation.

The research represents the first attempt to develop an integrated activity-based travel demand model system for individual's quarterly/yearly long distance or national activities and travel in the U.S at the Metropolitan Statistical Area (MSA)/Non-MSA level which is the finest geographic resolution in the long distance travel survey data. The model system is developed considering the specific attributes of the long distance travel such as low frequency, long activity duration, long activity duration at intermediate stops on the tour legs, different sets of mode alternatives, etc. Therefore, the model system not only takes into account the people's long distance travel at the tour level, but also at the stop level. Three levels of choice are modeled.

The first level is the activity pattern level which generates the number of different types of activities a person will choose during one year; the second level is the tour level which contains choices of tour destination, time of year, tour duration, and tour mode; and the lowest level is the stop level model system including the number, the purpose, and the location of each intermediate stop made during the inbound and outbound legs of the tour.

Nationwide long distance travel data over the course of one year is adopted to estimate the parameters of the national travel demand model system. Data from multiple sources (e.g. Census Bureau, Airline Origin and Destination Survey, Amtrak, EIA, Boeing, CEDDS, and etc.) are collected and used to obtain the TAZ economic and demographic data and the transportation OD skim data for car, air and train travel. Each model component of the model system is validated through the with-out sample validation method.

The model is implemented in the developed microsimulation platform which simulates each individual's yearly long distance activities and travel in the U.S. The 2010 PUMS data is expanded according to the person weight in order to generate the total population data in base year. Based on the base year passenger long-distance model, we calibrated the alternative specific constants of the time of year choice model, and the travel mode choice model using the Airline Origin and Destination Survey (DB1B) which is the only observed data we have for calibration.

The calibrated model is employed to predict the future year long distance travel demand based on the synthetic future year population which we call it base scenario. The future year population is generated using PopGen at the county level for

the future year of 2040. Multiple scenarios (including fuel price increase and high speed rail) in the future year are then analyzed and compared with the base scenario. The comparison results are as what we expected. People are sensitive to the fuel price change while driving for long distance travel. As the fuel price increases, people would decrease their number of car travel or reduce the travel distance of each car trip. Although the high speed rail (HSR) is operated along the selected regional corridor in terms of train travel time increase and more people would turn to train for their long distance travel, the car is still the most popular travel mode among the three travel modes with the largest share of the trips along the regional corridor. And after the HSR is operated along regional corridor, the main change of the trip (trip distributions) occurs along the corridor. The trip distributions at the national level are seldom affected.

9.2 Recommendations for Future Research

In our national travel demand model, the high-level module generates the number of long distance activities by purpose based on the parameters called long distance trip rates. The long distance trip rates for each purpose were estimated using Multiple Classification Analysis (MCA) method based on 1995 ATS data. The value of the parameters is determined only by the person's socioeconomic and demographic characteristics. Such model feature could result in the insensitivity of the number of long distance activities for each purpose to any transportation-related policies. For instance, in our future year scenario analysis, each person's long distance activities of the three purposes don't change with the scenario, as the socioeconomic and

demographic characteristics of the person or household don't change. In reality, people's number of long distance activities for each purpose could be affected by not only their socioeconomic and demographic attributes but also the level of service of the transportation network. Due to the data limitation, we simplified the theory of people's choice on the number of long distance activities by purpose during one year. But in future, with the sufficient data, the method of predicting the long distance activity pattern during one year could be improved by using more advanced methodology (e.g. multinomial logit model) and incorporating transportation LOS attributes.

The national travel demand model is developed based on the 1995 American Travel Survey (ATS) data which provides detailed information about people's national long distance travel during one year. However, as we know that the main problem about the data is that it is more than 20 years old and has a limited sample size, and people's travel behavior and the travel pattern could change during the 20 years. With the next iteration of the long distance travel survey of the U.S. household population in future, we can utilize the new survey data set with the proposed methodology to impute the missing information to improve the national travel demand model which can better reflect people's decisions in long distance travel.

For the passenger long distance model, we initially conducted 60 iterations of calibration due to time limitation, and it results in that the trip distribution by time of year hasn't reached up to the expected pattern. With more iterations, the trip distribution by time of year should be what we expect. In the current calibration, we calibrate the alternative specific constants of the time of year choice model, and the

travel mode choice model using the Airline Origin and Destination Survey (DB1B) which is the only available data set to use for calibration. The airline OD data is aggregated at the state level for Maryland and the rest of the states as a whole for each quarter (which yields 16 variables) in order to reduce the calibration time. With more powerful computer, we could use the airline OD data at the TAZ level to conduct the calibration. Also, in our calibration, due to the data limitation (lack of OD data for car and train travel), only airline OD data is used. In future, with available OD data for car and train travel, the passenger national travel model could be calibrated by taking into account all the OD data for all the travel modes.

The national travel demand model can simulate the travel pattern and travel behavior of each person in the U.S over the course of one year. It means that the simulation could use a large amount of RAM and the results would need a lot of storage to be saved. Currently, writing each person's detailed trip and tour information of long distance activities during one year into database (e.g. MySQL) would take more time and require a lot more computer storage and resources. Therefore, we output more aggregated and summarized travel data, which will result in no access to the output of each model component and the detailed trip/tour information of each person. With distributed computation technique or other big data technique and more computer storage, the simulation program of the national travel demand model could be improved and optimized to reduce the running time and store the detailed trip/tour information of each person in the U.S.

Bibliography

- Allison, Paul D, 1982. Discrete-time methods for the analysis of event histories."Sociological methodology 13.1: 61-98.
- Amtrak, 2010. A-Vision-for-High-Speed-Rail-in-the-Northeast-Corridor.
<https://www.amtrak.com/ccurl/214/393/A-Vision-for-High-Speed-Rail-in-the-Northeast-Corridor.pdf>
- Antoniou, C., Azevedo, C. L., Lu, L., Pereira, F., and Ben-Akiva, M. (2015), W-SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. *Transportation Research Part C: Emerging Technologies*, 59, 129-146.
- Axhausen, K. and Müller, K., 2010. Population synthesis for microsimulation: State of the art. Technical Report August. Swiss Federal Institute of Technology Zurich.
- Axhausen, KW., Schonfelder, S., Wolf, J., Oliveira, M., Samaga, U., 2003. 80 weeks of GPS-traces: approaches to enriching the trip information. *Transportation Research Record*, 1870, 46 -54
- Baik, H., Trani, A. A., Hinze, N., Swingle, H., Ashiabor, S., & Seshadri, A., 2008. Forecasting Model for Air Taxi, Commercial Airline, And Automobile Demand in the United States. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2052, pp9-20.
- Battelle Memorial Institute, The Urban Institute, and University of Maryland, 2013. Design of a completely new approach for a national household travel survey instrument. Project report for Federal Highway Administration. Contract No. DTFH61-11-C-00039
- Beckman, R. J., K. A. Baggerly and M. D. McKay (1996) Creating synthetic baseline populations, *Transportation Research Part A: Policy and Practice*, 30 (6) 415–429.
- Ben-Akiva, M., Cascetta, E., Coppola, P., Papola, A., & Velardi, V., 2010. High speed rail demand forecasting: Italian case study. In European Transport Conference, 2010.
- Ben-Akiva, M.E. and Lerman, S.R., 1985. Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.
- Besser, M. and Algers, S., 2001. SAMPERS – The New Swedish National Travel Demand Forecasting Tool. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

- Bhat, C., 2005. A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research Part B*, Vol. 39, No. 8, pp. 679-707.
- Bhat, C.R., 1995. A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice. *Transportation Research Part B*, Vol. 29, No. 6, pp. 471-483.
- Billitteri, Thomas J. High-Speed Trains: Does the United States Need Supertrains?. [Online] library.cqpress.com. Retrieved on December 26, 2013.
- Boeing (Producer)., 2011. 757 Commercial Transport History. Retrieved from <http://www.boeing.com/history/boeing/757.html>
- Bohte, W., Maat, K., 2009. Deriving and validating trip destinations and modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. *Transportation Research Part C* 17, 285–297
- Börjesson, M., 2012. Forecasting demand for high speed rail. <http://vti.diva-portal.org/smash/get/diva2:669361/FULLTEXT01.pdf>
- Bostrom, R., 2006. Kentucky Statewide Travel Model (KYSTM). Presentation. Combined Kentucky-Tennessee Model Users Group Meeting, Bowling Green KY. October 26, 2006.
- Bowman, J.L., 2004. A comparison of population synthesizers used in microsimulation models of activity and travel demand. Unpublished working paper. http://jbowman.net/papers/2004_Bowman_Comparison_of_PopSyns.pdf.
- Bowman, J. L., and M. A. Bradley., 2006. Activity-based travel forecasting model for SACOG: Technical Memo Number 5: Intermediate Stop Location Models. Available at <http://jbowman.net>
- Brog, W., Erl, E., & Schulze, B., 2004. DATELINE: Concept and Methodology. Paper presented at the 2004 European Transport Conference.
- Bureau of Transportation Statistics, 2015. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transporation_statistics/html/table_04_23.html
- Burge, P., Kim, C. W., & Rohr, C., 2011. Modelling Demand for Long-Distance Travel in Great Britain: Stated preference surveys to support the modelling of demand for high-speed rail.
- Burge, P., Rohr, C. and C. Kim., 2010. Who might travel by high-speed rail? Modeling Choices for long-distance travelers in the UK, European Transport Conference, Glasgow.

Burgess, A., Snelder, M., Martino, A., Fiorello, D., Bröcker, J., Schneekloth, N. & Rudzi-kaite, L., 2006. TRANS-TOOLS (TOOLS for Transport forecasting AND Scenario testing) Deliverable 1. Funded by 6th Framework RTD Programme. TNO Inro, Delft, Netherlands.

Cambridge Systematics, 2016, California High-Speed Rail Draft 2016 Business Plan-Ridership and Revenue Forecasting. Prepared for Parsons Brinckerhoff and California High-Speed Rail Authority
https://www.hsr.ca.gov/docs/about/business_plans/DRAFT_2016_Business_Plan_Ridersihp_Revenue_Forecast.pdf

Cambridge Systematics and Mark Bradley Research and Consulting, 2006. Bay Area/California High-Speed Rail Ridership and Revenue Forecasting Study: Interregional Model System Development. Final Report. Prepared for Metropolitan Transportation Commission and the California High-Speed Rail Authority.

Cambridge Systematics., 2008. National travel demand forecasting model phase I final scope. NCHRP Project (2008): 08-36. Retrieved from <http://www.camsys.com/pubs/NCHRP08-36-70.pdf>

Census Bureau., 2008. Public Use Microdata Sample, 2000 Census of Population and Housing-Technical Documentation.
<https://www.census.gov/prod/cen2000/doc/pums.pdf>

Chen, C., Gong, H., b, Lawson, C., Bialostozky, E., 2010. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A* 44, 830–840

Cohen, H., Horowitz, A., & Pendyala, R., 2008. Forecasting statewide freight toolkit. Washington, DC: Transportation Research Board.

Congression Budget Office (CBO), 2008. Effects of Gasoline Prices on Driving Behavior and Vehicle Markets. <https://www.cbo.gov/sites/default/files/110th-congress-2007-2008/reports/01-14-gasolineprices.pdf>

Contrino, H., & McGuckin, N., 2009. Demographics Matter Travel Demand, Options and Characteristics Among Minority Populations. *Public Works Management & Policy*, 13(4), 361-368

Daly, A.J., Rohr, C. (1998) Forecasting Demand for New Travel Alternatives. In: T Gärling, T Laitila, K Westin (ed.) *Theoretical Foundation for Travel Choice Modeling*, Pergamon.

- Davidson, P., & Clarke, P., 2004. Preparation of OD Matrices for DATELINE. Paper presented at the 2004 European Transport Conference.
- Deming, W. E. and F. F. Stephan (1940) On the least squares adjustment of a sampled frequency table when the expected marginal totals are known, Annals of Mathematical Statistics, 11 (4) 427–444.
- Deng, Z., Ji, M., 2010. Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach. Traffic and Transportation Studies. p.768-777
- Energy Information Administration (EIA)., <http://www.eia.gov/>. Accessed in 2016.
- Epstein, J. M., Parker, J., Cummings, D., & Hammond, R. A., 2008. Coupled Contagion Dynamics of Fear and Disease: Mathematical and Computational Explorations. PLoS One, Vol. 3, No. 12, p. e3955.
- Erhardt, G., Freedman, J., Stryker, A., Fujioka, H., & Anderson, R., 2015. Ohio long-distance travel model. Transportation Research Record: Journal of the Transportation Research Board.
- Federal Highway Administration., 2013. Understanding Long-Distance Traveler Behavior - Supporting a Long-Distance Passenger Travel Demand Model. FHWA-HRT-13-095.
- Forinash, C. V., and F. S. Koppelman., 1993. Application and Interpretation of Nested Logit Models of Intercity Mode Choice. In Transportation Research Record 1413, TRB, National Research Council, Washington, D.C., pp. 98–106.
- Fosgerau, M., 2002. PETRA—An Activity-based Approach to Travel Demand Analysis. In National Transport Models (pp. 134-145). Springer Berlin Heidelberg.
- GAO (United States Government Accountability Office), 2014. Impact of Fuel Price Increases on the Aviation Industry.
- Gaudry, M., 2001. Test of Nonlinearity, Modal Captivity And Spatial Competition Within the STEMM Multicountry Applications for Passengers. In Lundqvist, L., & Mattsson, L.G. (Eds.), National Transport Models: Recent Developments and Prospects, Springer, New York.
- Giaimo, G. T., & Schiffer, R., 2005. Statewide Travel Demand Modeling: A Peer Exchange, Longboat Key, Florida, September 23-24, 2004. Transportation Research E-Circular, (E-C075).

- Giaimo, G.T.,& Schiffer, R. (Eds.), 2005. August. Statewide travel demand modeling: A peer exchange. *Transportation Research Circular*, #E-C075.
- Gokovali, U., Bahar, O. and Kozak, M., 2007. Determinants of length of stay: A practical use of survival analysis. *Tourism Management*, 28(3), pp.736-746.
- Griffin, T., Huang, Y., 2005. A Decision Tree Based Classification Model to Automate Trip Purpose Derivation. In the Proceedings of the 18th International Conference on Computer Applications in Industry and Engineering, Honolulu, Hawaii
- Grush, W., 1998. Usage and Vehicle Miles of Travel (VMT) per Capita. *Highway Information Quarterly*, 5(4).
- Gunn, H.F., 2001a. An Overview of European National Models. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.
- Guo, J. Y. and C. R. Bhat , 2007. Population synthesis for microsimulating travel behavior, *Transportation Research Record*, 2014 (12) 92–101.
- HCG and TOI., 1990. A Model System to Predict Fuel Use And Emissions from Private Travel in Norway from 1985 to 2025. Report to the Norwegian Ministry of Transport. Hague Consulting Group.
- Hess, S., Adler, T. & Polak, J.W., 2007. Modelling airport and airline choice behavior with the use of stated preference survey data, *Transportation Research Part E*, 43, pp. 221-233
- Hofman, F., 2001. Application Areas for the Dutch National Model. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.
- Horowitz, A.J., 2006. Statewide travel forecasting models (NCHRP Synthesis #358). Transportation Research Board.
- Horowitz, A.J., 2008. White paper: Statewide travel demand forecasting. Requested by AASHTO and presented at the conference on meeting federal surface transportation requirements in statewide and metropolitan transportation planning.
- Hwang, H.L. & Rollow, J., 2000. Data Processing Procedures and Methodology for Estimating Trip Distances for the 1995 American Travel Survey (ATS). http://cta.ornl.gov/cta/Publications/Reports/ORNL_TM_2000_141.pdf
- Koppelman, F. S., 1989. Multidimensional model system for intercity travel choice behavior. *Transportation Research Record*, (1241).

- Koppelman, F.S., & Sethi, V., 2005. Incorporating Variance and Covariance Heterogeneity in the Generalized Nested Logit model: An Application to Modeling Long Distance Travel Choice Behavior. *Transportation Research Part B*, Vol. 39, No. 9, pp. 825-853.
- Lee, J. H., K.-S. Chon, and C. Park., 2004. Accommodating Heterogeneity and Heteroscedasticity in Intercity Travel Mode Choice Model: Formulation and Application to Honam, South Korea, High-Speed Rail Demand Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1898, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 69–78.
- Lemp, J.D. and Kockelman, K.M., 2012. Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A: Policy and Practice*, 46(3), pp.602-613.
- Li, G., 2004. Intercity Travel Demand: A Utility Consistent Simultaneous Trip Generation and Mode Choice Model. Doctoral dissertation. Interdisciplinary Program in Transportation, New Jersey Institute of Technology, Newark
- Lim, P.P. and Gargett, D., 2013, October. Population Synthesis for Travel Demand Forecasting. In Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia.
- Lu, Y., Zhu, S., Zhang, L., 2013. Imputing Trip Purpose Based on GPS Travel Survey Data and Machine Learning Methods, *Transportation Research Board 92nd Annual Meeting*. Washington D.C, Paper Number: 13-3177
- Lu, Y., and Zhang, L. (2015). Imputing trip purposes for long-distance travel. *Transportation*, 42(4), 581-595.
- Lundqvist, L., & Mattsson, L. G. (Eds.), 2001. *National Transport Models: Recent Developments and Prospects*. Springer.
- Mannering, F.L., 1983. An econometric analysis of vehicle use in multivehicle households. *Transportation Research Part A*, Vol. 17, No. 3, pp. 183-189.
- Nerella, S. and Bhat, C., 2004. Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record: Journal of the Transportation Research Board*, (1894), pp.11-19.
- Outwater, M., Bradley, M., Ferdous, N., Trevino, S., & Lin, H. (2015). Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework: Implementation Report.

Parsons Brinckerhoff, HBA Specto Incorporated, and EcoNorthwest., 2010. Oregon Statewide Integrated Model (SWIM2) Model Description.
<http://www.oregon.gov/ODOT/TD/TP/docs/statewide/swim2.pdf>

Peter Davidson Consultancy., 2000. MYSTIC Toward Origin-Destination Matrices for Europe. London.

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S. and F. Tsang, 2013. Modelling long-distance travel in the UK, Transport Research Record 2344, pp 145-152

Ryan, J, Maoh, H and Kanaroglou, P., 2009. Population synthesis: Comparing the major techniques using a small, complete population of firms', Geographical Analysis, 41 (2) 181–203.

Salvini, P. A. and E. J. Miller, 2005. ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems, Networks and Spatial Economics, 5 (2) 217–234.

Schönfelder, S., K. Axhausen, N. Antille, M. Bierlaire, and E. Lausanne., 2002. Exploring the potentials of automatically collected GPS data for travel behavior analysis - a Swedish data source. GI-Technologien für Verkehr und Logistik 13, 155-179.

Sharma, S., Lyford, R., & Rossi, T., 1999. The New Hampshire Statewide Travel Model System (No. E-C011).

<http://onlinepubs.trb.org/onlinepubs/circulars/ec011/sharma.pdf>

Srinivasan, S. and L. Ma, 2009. Synthetic population generation: A heuristic data-fitting approach and validations, paper presented at the the 12th International Conference on Travel Behaviour Research (IATBR), Jaipur, December 2009.

Srinivasan, S., L. Ma and K. Yathindra, 2008. Procedure for forecasting household characteristics for input to travel-demand models, Final Report, TRC-FDOT-64011-2008, Transportation Research Center, University of Florida.
http://www.fsutmsonline.net/images/uploads/reports/FDOT_BD545_79_rpt.pdf.

Souleyrette, R.R., Hans, Z.N., & Pathak, S., 1996. Statewide transportation planning model and methodology development program. Ames: Iowa State University.

Spall, J.C. (2003), Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control, Wiley.

Stammer, Jr., R.E., 2002. Statewide Modeling Practices and Prototype Statewide Model Development. Final Report TNSPR-RES-1147. Prepared for Tennessee Department of Transportation.

- Stopher, P., Clifford, E., Zhang, J., FitzGerald, C., 2008a. Deducing Mode and Purpose from GPS data. Working Paper of the Austrian Key Centre in Transport and Logistics. University of Sydney, Sydney, Australia.
- Stopher, P., FitzGerald, C., Zhang, J., 2008b. Search for a Global Positioning System device to measure personal travel. *Transportation Research Part C* 16(3), 350–369.
- Stopher, P.R., Greaves, S., FitzGerald,C., 2005. Developing and deploying a new wearable GPS device for transport applications. Paper presented to the 28th Australasian Transport Research Forum, Sydney, 28–30 September.
- Tipping, A., Schmahl, A., Duiven, F. 2015. The Impact of Reduced Oil Prices on the Transportation Sector. <http://www.strategy-business.com/article/00312?gko=ae404>
- The National Center for Smart Growth Research and Education, University of Maryland and Parsons Brinckerhoff., 2011. ‘MSTM User Guide: Maryland Statewide Transportation Model’
- Urban Land Use and Transportation Center, HBA Specto Incorporated., 2011. CSTD莫 - California Statewide Travel Demand Model. http://ultrans.its.ucdavis.edu/files/pecas/CSTD莫_ModelOverview_Final_0.pdf
- U.S Department of Transportation (USDOT), 2008. Impact Of High Oil Prices On Freight Transportation: Modal Shift Potential In Five Corridors Technical Report. http://www.marad.dot.gov/wp-content/uploads/pdf/Modal_Shift_Study_-Technical_Report.pdf
- Van Nostrand, C., Sivaraman, V., & Pinjari, A. R., 2013. Analysis of Long-Distance Vacation Travel Demand in the United States: A Multiple Discrete-Continuous Choice Framework. *Transportation*, Vol. 40, No. 1, pp. 151-171.
- Wardman, M., 1988. A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy*, pp.71-91.
- Weiner, E., 1976. Assessing National Urban Transportation Policy Alternatives. *Transportation Research*, Vol. 10, No. 3, pp. 159-178.
- Widlert, S., 2001. National Models: How to Make It Happen. The Case of the Swedish National Model System: SAMPERS. In Lundqvist, L., & Mattsson, L.G. (Eds.), *National Transport Models: Recent Developments and Prospects*, Springer, New York.

Williams, I., 2001. Designing the STREAMS model of Europe. In Lundqvist, L., & Mattsson, L.G. (Eds.), National Transport Models: Recent Developments and Prospects, Springer, New York.

Winston, C., 1985. Research on Intercity Freight and Passenger Transportation: An Economist's Perspective. *Transportation Research Part A*, Vol. 19, No. 5-6, pp. 491-494.

Witten, I. H., Frank, E., 2005, Data Mining: Practical Machine Learning Tools and Techniques, Second Edition

Wolf, J., Guensler, R., Bachman, W., 2001a. Elimination of the travel diary: an experiment to derive trip purpose from GPS travel data. In 80th Annual Meeting of the Transportation Research Board, Washington DC., p.24.

Wolf, J., R. Guensler, and W. Bachman, 2001b. Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record: Journal of the Transportation Research Board* 1768 (1), 125-134.

Worsley, T.E., & Harris, R.C.E., 2001. GB traffic forecasts—status and development. In L. Lundqvist & L.-G. Mattsson (Eds.), National transport models: Recent developments and prospects. Stockholm: Swedish Transport and Communications Research Board

Ye, X., K. Konduri, R. M. Pendyala, B. Sana and P. Waddell, 2009. A methodology to 14 match distributions of both household and person attributes in the generation of synthetic 15 populations, paper presented at the the 88th Annual Meeting of the Transportation Re- 16 search Board, Washington, D.C., January 2009.

Zhang, L., Southworth, F., Xiong, C., & Sonnenberg, A., 2012. Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review. *Transport Reviews*, Vol. 32, No. 4, pp. 399-433.

Zhang, L., Southworth, F., Xiong, C., & Sonnenberg, A., 2012. Methodological Options and Data Sources for the Development of Long-Distance Passenger Travel Demand Models: A Comprehensive Review.

Zhang, L., Xiong, C., & Berger, K., 2010. Multimodal Inter-Regional Origin-Destination Demand Estimation: A Review of Methodologies and Their Applicability to National-Level Travel Analysis in the US. In the World Conference on Transport Research, Lisbon, Portugal.^{ss}