Federal Highway Administration

Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework

Part A: Final Report

Exploratory Advanced Research Program

DTFH61-10-R-00036

March 2015

U.S.Department of Transportation
Federal Highway Administration

Notice

Quality Assurance Statement

Form DOT F1700.7 (8-72)

1. Report No.
FHWA-***********

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework

5. Report Date
March 2015

6. Performing Organization Code:

7. Author(s):
Maren Outwater, Mark Bradley, Nazneen Ferdous, (RSG)
Ram Pendyala, Venu Garikapati (Arizona State University)
Chandra Bhat, Subodh Dubey (University of Texas at Austin)
Jeff LaMondia (Auburn University)
Stephane Hess, Andrew Daly (University of Leeds)

8. Performing Organization Report No.

9. Performing Organization Name and Address

Resource Systems Group, Inc.
55 Railroad Row, White River Junction, VT 05001

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

DTFH61-10-R-00036

12. Sponsoring Agency Name and Address

U.S. Department of Transportation
Federal Highway Administration
Washington, DC 20590

13. Type of Report and Period Covered

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract
Intercity travel is a topic of increasing importance in the United States, with many states and the federal government faced with the challenge of improving mobility and reducing impacts for these travelers. The Federal Highway Administration (FHWA) has invested in several studies to better understand intercity travel; this study is an extension of that interest, focused on exploratory research to develop a long-distance passenger travel demand model framework. The modeling framework is a tour-based microsimulation model of annual long-distance passenger travel for all households in the United States. The models schedule travel across one full year to capture business travel (e.g., conferences, meetings, and combined business/leisure) and leisure travel (e.g., visiting friends and family, personal business and shopping, relaxation, sight-seeing, outdoor recreation, and entertainment). The models are multimodal (i.e., auto, rail, bus, and air) and based on national networks for each mode to provide opportunities for evaluation of intercity transportation investments or testing national economic, environmental and pricing policies. Advanced modeling methods were tested for the scheduling, time-use, activity-participation, and joint-mode and destination models. Methods included multiple discrete-continuous extreme value (MDCEV) for the scheduling models and cross nested logit (NL) choice for the joint mode and destination models. The modeling framework was demonstrated, with application software that simulates long-distance travel for all households in the United States using an initial set of advanced modeling methods. This final report is an overview of the exploratory research over three years.

17. Key Words
Long-Distance; National Passenger Models; Intercity Travel; Scheduling Models; Joint Mode and Destination; MDCEV; Cross-Nested Logit Choice

18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	254	

TABLE OF CONTENTS

List of Figures

List of Tables

List of Equations

List of Acronyms

AADT	Average annual daily traffic
AB	Activity-based
ACS	American Community Survey
API	Application programming interface
ATS	American Travel Survey
BLS	Bureau of Labor Statistics
BTS	Bureau of Transportation Statistics
CAPI	Computer-assisted personal interviewing
CHTS	California Household Travel Survey
CNL	Cross-nested logit
CPI	Consumer Price Index
DOT	Department of transportation
DRCOG	Denver Regional Council of Governments
EAR	Exploratory Advanced Research
EPA	Environmental Protection Agency
FHWA	Federal Highway Administration
FIPS	Federal Information Processing Standard
GDP	Gross domestic product
GTFS	General Transit Feed Specification
HH	Household
HPMS	Highway Performance Management System
HSR	High-speed rail
IPF	Iterative proportional fitting
IPU	Iterative proportional updating

LD	Long distance
LDT	Long-distance transport
LEHD	Longitudinal Employer Household Dynamics
LL	Log likelihood
LODES	LEHD Origin-Destination Employment Statistics
LOS	Level-of-service
LSOT	Longitudinal Study of Overnight Travel
MACML	Maximum Approximate Composite Marginal Likelihood
MDCEV	Multiple discrete-continuous extreme value
MLE	Maximum likelihood estimation
MMNL	Mixed Multinomial Logit
MNL	Multinomial logit
MPO	Metropolitan Planning Organization
NAICS	North American Industry Classification System
NCHRP	National Cooperative Highway Research Program
NHB	Non-home based
NHPN	National Highway Planning Network
NHTS	National Household Travel Survey
NPMRDS	National Performance Management Research Data Set
NUMA	National Use Model Area
PUMA	Public Use Microdata Areas
PUMS	Public Use Microdata Sample
QCEW	Quarterly Census of Employment and Wages
O-D	Origin-destination
QWI	Quarterly Workforce Indicators

RUM	Random utility maximization
SLL	Simulated log-likelihood
TAZ	Traffic analysis zones
US DOT	United States Department of Transportation
VMT	Vehicle miles traveled

ACKNOWLEDGEMENTS

This study was a three-year collaborative effort—between the project team members, the Federal Highway Administration (FHWA), and the peer review panel—to explore advanced modeling methods for long-distance passenger travel demand. Maren Outwater, of RSG, was the Principal Investigator for the project and oversaw all aspects of the research in close partnership with Mark Bradley, also of RSG, who led the development of the demonstration model and the initial joint destination and mode choice models. Nazneen Ferdous, of RSG, led the development of data sources for the estimation, calibration, and validation of the national passenger models with support from Åsa Bergman; Nazneen also conducted model estimation support for the destination and mode choice models. Colin Smith, of RSG, provided critical assistance during the design phase and led the development of the air passenger data. Tom Adler, of RSG, provided senior technical advice throughout the project.

Ram Pendyala and Venu Garikapati, of Arizona State University, developed and implemented the population synthesis for the United States and contributed significantly to the development of the modal networks and national zone system. Chandra Bhat and Subodh Dubey, of the University of Texas at Austin, estimated the nonbusiness scheduling models and supported the development of the business scheduling models; this work was performed in partnership with Jeff LaMondia, of Auburn University, who led the development of the business scheduling models. Stephane Hess and Andrew Daly, of the University of Leeds, led the development and testing of the advanced joint destination and mode choice models.

The quality of the products was enhanced by the direction and feedback provided by the Peer Review Panel (Chad Baker, Doug Hunt, John Conlow, Rich Brancato, Richard LaFerriere, Robert Cochrane, Bill Sheridan, Kay Axhausen, Guy Rousseau, Frank Southworth, and Erik Sabina) and FHWA leaders (Brad Gudzinas and Tianjia Tang).

CHAPTER 1. INTRODUCTION

This technical report presents the findings of Task 11: Develop Framework of the DTFH61-10-R-00036 Exploratory Advanced Research (EAR) program to develop Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework. Task 11, which is the last task under Phase II (Research Tasks) of the project, unifies all the model components into a modeling system, demonstrates the integration of the modeling components with software, and produces performance metrics of interest to planning applications.

The purpose of this report is to document the development of the modeling system and the demonstration of this system to develop long-distance passenger travel for 2010. This effort involved significant data development to build road, air, rail, and bus networks, as well as a zone system for the United States in order to best serve the goal of accurately estimating long-distance travel. The models developed and presented in the Model Estimation Memorandum (Tasks 8 and 9, October 2013) were updated to ensure consistency among the variables and to expand the underlying survey coverage from a single source to a combined dataset of four states. Additional exploratory work on the joint-destination and mode-choice models was undertaken to address the challenges of estimation on such a large dataset.

The demonstration of the methods and data developed is provided with software to implement the model components for long-distance travel. This demonstration is based on the least complex of the recommended model component to evaluate the overall challenges of the modeling system and methods to implement the long-distance passenger travel models. The demonstration focuses on predicting a single day of long-distance travel in a specific month and evaluates several ways to expand this demonstration to predict annual long-distance passenger travel. As a demonstration, this does not include calibration, validation, or forecasting with the modeling system. Performance metrics have been selected to visualize the possible outputs from the modeling system; examples of these outputs have been provided.

1.1 Objectives of the Program

The US Department of Transportation (US DOT) recognizes that long-distance travel represents a significant portion of all travel and that this type of travel is economically important. In 1995, the date of the last long-distance passenger travel survey in the United States, 25% of all person-miles traveled in the United States were classified long-distance travel. Long-distance travel includes both Business and Leisure travel, both of which are major contributors to the United States economy. As a result, the US DOT desires to support continued high levels of personal mobility for both economic and social purposes.

Providing personal mobility in a cost-effective manner requires an understanding of travel behavior and an evaluation of investments in transportation infrastructure and services; it also requires anticipating the effects of any public- and private-sector decisions that impact the transportation system and its uses. The development of the Long-Distance Passenger Travel Demand Modeling Framework will be focused on addressing this overall goal.

The primary uses of the long-distance passenger travel demand model are to: a) evaluate transportation investments in transportation infrastructure and services designed to address deficiencies in the national transportation system; and (b) evaluate the impacts of transportation planning policies being considered at the national level or significant private-sector decisions that may affect the national transportation system. Intercity travel has not received the same attention in the travel demand modeling field as urban and local travel have; as a result, it has not evolved to reflect the unique nature of these types of trips.

A secondary use of the long-distance passenger travel demand model is to support statewide models around the country. There are at least 30 statewide travel demand models in the United States and each one incorporates long-distance travel as external movements into, out of, and through the state, and internal movements within the state. There are currently many methods for incorporating long-distance travel in statewide models, but there is no agreement on the best method. There is also a collective desire to improve upon these methods and the data that are used to support these methods. Insufficient or inaccurate data is a common theme in limiting the understanding of long-distance travel.

The US DOT is leading the effort to develop long-distance passenger travel demand models that will support both national modeling needs and statewide modeling needs through this Exploratory Advanced Research (EAR) program. A multimodal national transportation demand model could be used to:

- Analyze national, multistate, or megaregion travel and congestion;
- Test the effectiveness of national policies;
- Provide a framework for system performance measurement;
- Evaluate the impacts of private-sector decisions on the national transportation system;
- Provide state agencies with interstate passenger travel for statewide planning; and
- Provide metropolitan agencies with quality external trip data for urban area planning.

National Travel Modeling

There are three primary reasons to build national travel models:

- To inform infrastructure investment decisions;
- To evaluate the impacts of transportation policies on mobility and the economy; and
- To understand the impacts of private-sector transportation decisions.

It is important that the Long-Distance Passenger Travel Demand Modeling Framework recognize these uses so that it is specifically designed to address these needs. In some cases, emerging developments in various fields should be considered so that the modeling framework can integrate these developments and evaluate their impacts. Examples of emerging developments are vehicle technologies (e.g., gas, hybrid, diesel, electric, etc.) with different fuel efficiencies, communication advances, high-speed rail, and others.

Infrastructure Investment Decisions

Transportation agencies across the United States use travel demand forecasting models to inform their infrastructure investment decisions by evaluating possible alternatives against a base case to select the alternative with the highest return on the investment. Benefits are determined as a result of the travel demand and compared to the cost of the investment. It is important to understand the characteristics of the travel demand so that the benefits for each alternative can accurately reflect tradeoffs that travelers make and the impacts that will be affected by these choices.

Policy Analysis

The use of a long-distance passenger travel demand model for policy analysis in the United States was one of the primary objectives in constructing this model. Proposed is a broad spectrum of potential policy analyses that a national passenger travel model could support:

- **Modal alternatives**. There is direct competition between air, rail, bus, and auto modes for intercity and long-distance travel, and any infrastructure investments being considered should be evaluated in the context of this competition. Currently, these types of infrastructure investments are evaluated using a corridor, state, or megaregion model, often developed specifically to evaluate a single project. A distinct advantage of a national model to evaluate these investments would be that a national would be consistent from one state/region to another; be less likely to be biased in favor of a particular mode; and represent improved forecasts, since the forecasts will represent a deeper understanding of long-distance travel. In addition, some modal alternatives have the ability to induce travel.
- **Pricing**. There are many aspects of pricing that can and should affect national model forecasts. Pricing can be a strategy to manage demand or raise revenues (e.g., toll roads, air and rail fares, etc.). Pricing is also represented as assumptions on gas prices, and the effects of rising or falling gas prices can be used to evaluate overall transportation impacts. Air fares can change based on when tickets are purchased (advanced prices are lower) and are set by the private sector based on market factors. Rail fares tend to be more stable and are correlated with distance traveled, but could change to reflect advanced ticket prices (as is currently done in the United Kingdom) or demand, as is the case with congestion pricing (prices are higher during peak periods of demand). Also, traveler decisions are affected by reimbursement for business travel.
- **Economics**. Policies that aim to improve the economic conditions will, in turn, affect long-distance travel. Tourism is an important sector of the economy that can change according to economic conditions, which determines the type of traveler who will visit. In tougher economic times, travelers will visit destinations closer to home. Business travel is also affected by the economy and has increasingly adapted to new technologies that offer a substitute for business travel. Fare structures can be set differently for Business and Leisure travelers (either by time of day or day or week or season).
- **Environmental**. Environmental policies that aim to reduce emissions can have effects on long-distance travel. An increase in the gas tax will influence gas consumption and

potentially reduce vehicle miles traveled (VMT), and possibly change vehicle purchase decisions for lower-emission vehicles. Carbon taxes will affect the cost of air travel. The Environmental Protection Agency (EPA) may change fuel standards, which would result in changes in vehicle purchases. Some lower-emission vehicles (e.g., electric cars) have limitations on long-distance travel.

- **Livability**. Transportation policies pertaining to livability, like Smart Growth or Complete Streets, can change the social character of a destination and increase the attractiveness of an area for tourism or business travel. This, in turn, will affect the long-distance travel to that destination over time.
- **Safety**. Policies to reduce accidents, such as the security clearance at airports or reduced speeds on highways, can affect travel behavior by causing mode shifts or reduced travel to avoid these regulations; these policies can also induce travel from safety conscious travelers.
- **Regional**. Policies that affect different regions of the country (e.g., taxes, tolls, gas prices, etc.) may result in different travel behavior across regions. These differences will need to be reflected, to the extent possible, in the data underlying the long-distance models.
- **Airport or Rail Planning**. Policies made by airports or long-distance rail operators regarding new capacity, rail airport connections, or environmental impacts could use a national model to inform these decisions.
- **Labor Force**. Prosperous regions can attract labor forces from less prosperous regions and can create long-distance travel between regions. The labor force is adaptive to economic conditions and housing affordability, which can drive changes in long-distance travel.

The goal of the Long-Distance Passenger Travel Demand Modeling Framework is to account for changes in these types of policies so that changes in long-distance travel can be forecast. While private-sector decisions are outside the control of the US DOT, it is useful to have long-distance passenger travel demand models that are sensitive to the factors that the private sector may control. This is because sensitivity tests can then be used to evaluate decision impacts.

Statewide Travel Modeling

Statewide models can benefit from national long-distance passenger travel demand models that provide input on travel into, out of, and through a state and long-distance travel occurring between urban areas within a state. The usefulness of these inputs and outputs will depend directly on the definition of long-distance travel chosen for the framework. Long-distance trips can be defined by distance, mode, destination, and/or purpose.

Interstate Travel

Long-distance travel is usually implied with interstate travel, but there are some states where metropolitan areas cross state boundaries and interstate travel within these regions is not long distance. There are also states with smaller geographies where interstate travel (or even through travel) are also not long distance. However, statewide models depend on a definition of long-distance travel that includes interstate travel.

Intrastate Travel

Larger states are recognizing that long-distance travel within their states is behaviorally different from other travel and that there are benefits to forecasting this travel with different parameters than metropolitan area models or traditional statewide models. Smaller states do not face this same challenge. National long-distance passenger travel demand models could provide useful information for larger states on long-distance passenger travel demand.

Megaregion Travel Modeling

The objectives for a long-distance national travel model to support analysis for statewide modeling also apply to megaregion modeling objectives, which are often developed to support infrastructure investments that cross several regions or states. There are fewer megaregion models in the United States than statewide models, but these models should be considered in defining the long-distance trip and determining aspects of integration that are important (e.g., spatial scale, temporal scale, trip purpose, mode, and trip chaining).

1.2 Contents of the Report

This report comprises work completed during the course of the research study. There are three main topics covered in this final report:

- **Background research** is presented in Chapter 2 as a summary of experience in the United States and internationally and identifies the majority of the literature provided in the references (Chapter 9).
- **Model exploration** includes a discussion of estimation and application data sources, such as socioeconomic, land-use data, and household surveys, as well as zones developed for this study and the networks developed for each mode in Chapter 3. Chapter 4 presents the integrated modeling system framework and the estimated model components within this framework. There is additional research on the exploration into model form and structure that is provided in a technical memorandum.
- **A demonstration** of the long distance modeling framework, using a simplified application structure, is presented in Chapter 5. This includes presentation of software developed to apply the microsimulation of all United States households' long-distance passenger travel models. Chapter 6 is a discussion of potential performance metrics that can be produced with this new set of long-distance models.

The background research for this project explored many possible datasets to support model estimation, but all datasets had some limitations. Once the integrated modeling system framework for long-distance passenger travel was developed, the research team also developed a set of recommendations to provide the data for building and applying the model in Chapter 7. Chapter 8 presents a high-level summary and a scope of work for the implementation phase of the project. This phase is underway and includes calibration and validation of the models with various observed data and sensitivity testing for different policies.

CHAPTER 2. SUMMARY OF EXPERIENCE IN THE UNITED STATES AND INTERNATIONALLY

Experience with long-distance travel demand modeling in both the United States and internationally served as a foundation for the development of the Long-Distance Passenger Travel Demand Modeling Framework. Current long-distance travel demand modeling in the United States has been developed primarily as a component of statewide models, either directly or indirectly (as part of overall trip-making). These approaches vary from conventional 4-step planning models to integrate economic- and activity-based approaches. The treatment of long-distance travel depends on the size of the state and the decision to segment long- and short-distance travel.

The United States models reviewed include those in California, Florida, Indiana, Kentucky, Michigan, New Hampshire, Ohio, Oregon, Tennessee, and Wisconsin. There are several national or multinational models outside the United States—primarily in Europe, but also in South America and Australia—where long-distance trips are modeled, with a similar range of methods as for statewide modeling in the United States. The international models reviewed in Task 2 include those in the United Kingdom, France, Belgium, Netherlands, Germany, Denmark, Norway, Sweden, Portugal, Switzerland, Czech Republic, Italy, Australia, and Chile.

The definition of a long-distance trip varied in the literature, which reinforced the need to determine a definition for a long-distance trip that met the needs of this study. In other studies, a long-distance trip has often been defined according to the size and scope of the region of interest; regions may vary from small states to small countries, to large states or countries. The United States is the largest country examined in the literature review, with the exception of the European-level models.

Each of the models reviewed were summarized along seven dimensions: 1) context; 2) definition of a long-distance trip; 3) model structure and form; 4) segmentation; 5) approach to forecasting; 6) types of applications and procedures; and 7) data sources.

2.1 Definitions and Context

Two contexts in which long-distance travel has been modeled can be distinguished:

- Some models have been developed for specific corridors, generally to help with decision-making about infrastructure investment and pricing. An example is the Danish Great Belt Study.
- Other models have been developed for complete areas, with a view to making them available for general policy and investment analysis for several modes and over a wide area. An example is the UK Long-Distance Model.

The UK Planet HS2 models show how a general model can be adapted to a specific project, in that case giving a substantial time saving, which was essential for such a large project (over US $50 billion) attracting huge political attention.

In the United States, of those states that forecast long-distance travel, many consider daily travel demand, long-distance travel, and freight demand simultaneously. Typically, states develop these models in response to a specific need, whether it is tourism, alternative modes, or freight. Models of air traffic may not give detailed attention to other modes.

In the case of models applicable to a specific project, the definition of the trips that are included is obviously those that would or might use the project. The more general models typically have a rigorous specification of trip length, often 100 km (62 miles) or 50 miles, with some instances of thresholds greater than 50 miles. The international examples often use the 100-km threshold while the domestic examples often use the 50-mile threshold, highlighting the somewhat arbitrary nature of this threshold setting. In some cases, the models consider any travel between urban areas, without a specific distance threshold. The Long-Distance Passenger Travel Demand Modeling Framework assumes a long-distance trip is greater than or equal to 50 miles.

2.2 Model Structures and Segmentation

The majority of long-distance trip models in the United States rely on modifications to the traditional 4-step planning process. While there are many assumptions inherent in this process, it makes it: a) easier to implement long-distance models across the state; and b) more efficient for results of the long-distance modeling efforts to be compared with those from local urban models. This is especially relevant, as many long-distance travel models in the United States serve as a supplement (and are estimated simultaneously) to daily travel models.

However, there has been a recent trend toward more long-distance models utilizing the tour-based modeling approach. While this is more insightful and offers more-detailed results and opportunities for analysis, it requires extensive surveys of travelers.

The major components of the international models were found to be as follows:

- The majority of the models described have at their core a logit choice submodel describing mode choice and usually other choices, such as submode, major routes, and timing choices.
- Some of the models, chiefly those that are not specific to corridors, represent destination choice. This is usually more sensitive to network effects than mode choice (i.e., it should be placed lower in a nested logit hierarchy).
- A number of models have an elastic trip generation component, in which change in accessibility is represented as changing the total number of trips made.
- Overall growth in trips is based on population and employment growth, with income, car ownership, and purchasing power possibly taken into account.

The MATISSE model, which uses an assignment concept, and Dargay's model, which is based on elasticities, are exceptions to the general trend of these models. Estimation generally uses maximum likelihood, although in many cases this is not a full-information procedure, as sequential estimations are made. Some models use trips (origin–destination) as the basic unit, while others use return tours or production–attraction relationships.

Among the models in the United States, the most common long-distance trip purposes are business, leisure, and personal business. However, a significant number of models do not define trip purpose. Few states consider segments of long-distance travel beyond the main trip purpose. Analysis methods consider household and zone characteristics, which could be extrapolated, but states have not demonstrated this.

It was found that all of the international models are segmented by travel purpose, at least separating Business and Leisure trips; however, commuting is occasionally grouped with business. Further purpose segmentations often concern the identification of commute and education, holiday, and social ("visit friends or relatives") trips. Associated with the purpose segmentation is length of stay, perhaps isolating day trips, or distinguishing short stays from long stays with a split at 3–5 days. Further trips are sometimes split and modeled separately for medium lengths and long trips, with a split at 150–300 miles.

A key further segmentation, which for data reasons is not included in many models, is by income group. Other segmentations used in some models concern residence location (e.g., country), party size, age, sex, employment, and car ownership (sometimes considered to be car availability). Specific segmentations that are not widely used are by area type (in the UK National Travel Model) and the detailed segmentation used in MATISSE, and also INVERMO.

2.3 Approach to Forecasting

Among the international models, the key distinctions were found to be between sample enumeration and zonal approaches, though these form a spectrum rather than a clear distinction between approaches. In either case, a conventional factoring to allow for growth in population, employment, and perhaps income is often applied, using government data sources. In some cases, forecasting is made difficult by the extent of macroeconomic developments (e.g., the huge advance in income in Poland in the last 20 years or the major setback in many countries in the last three years).

Pivoting is an important part of some modeling processes. When the model has a full set of alternative-specific constants (e.g., in a mode-choice model), then it will automatically reproduce the base situation. However, when destination choice is included, pivoting can help to substantially improve accuracy, though the quality of the base matrix data that are available are sometimes poor. Some models also include steps for assignment and iteration between the assignment and the demand model to allow highway capacity to be taken into account. This is more often a highway assignment, but in some cases rail capacity is also an issue.

In the United States, many states keep the long-range models separate from their forecasting methods, treating them as a sequential process rather than integrating them. As mentioned, many states developed long-distance models to address specific concerns. As a result, these models are designed to evaluate corridors, alternative routes, and alternative modes. This is also true internationally, where many of the models are designed for particular schemes and are restricted to application for those schemes, perhaps covering variations in the infrastructure, but also in operation and pricing.

However, there are also a set of more general models, for which high-speed rail is often an important application, though it is recognized that general models are not always capable of modeling this alternative without being extended (e.g., by stated choice information). Also, the prediction of CO_2 output is important in some cases and can be handled by special modules in conjunction with, for example, user benefit. In other cases, models are packaged for use by the client (e.g., in spreadsheet systems or special-purpose software). Occasionally, the packaged application is a simplified “sketch planning” variant of the full model.

2.4 Data Sources

All of the United States models rely on extensive datasets, and utilize Census and NHTS data to some extent. The most detailed models either include an add-on to the NHTS or include data from a statewide long-distance travel survey. This is also true internationally, where national travel surveys are often a key source for modeling traveler behavior. These are often supplemented by scheme-specific surveys, whether revealed preference, stated choice, or both. Stated intentions data are also used, for example, to try to obtain information about likely increases in the frequency of travel, which is not amenable to stated choice investigation. Ticket sales data, where available, are a useful source because of the volume. Other aggregate data, such as simple counts, can also be important to improving the accuracy of the model. The importance of good network data for all the modes considered cannot be overstated.

2.5 Other Issues

Some documentation is available only in the languages of the countries concerned. In some cases, documentation is limited or not available because the study is confidential. In other cases, interest focused on the first years of operation of infrastructure and then the pattern of build-up, involving learning and perhaps “novelty” trips. This is also associated with validation of the models, which can be done immediately on opening, provided that the unstable nature of demand is taken into account at that point. In addition, the important application to high-speed rail raises the issue of the position of new alternatives relative to existing ones. Whether high-speed rail is best handled as a choice (e.g., using a logit model) or as part of a rail assignment has raised issues in several countries. Furthermore, the question of induced demand can be an important one for significant infrastructure or pricing projects. This was addressed in the Great Belt Study and validated once the project opened, but the models under-predicted induced demand in the long term, after correctly predicting short-term induced demand. The use of simple models to check complex ones is a useful procedure. Finally, it will be useful to focus on a small number of the studies reviewed to get detailed information on particular modeling issues.

2.6 Summary

Table 1 provides a list of all of the models reviewed and a summary along the seven dimensions.

Table 1: Summary Table Comparing Long-Distance Models

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
Eurotunnel Car Market Share Model	UK to France and beyond	Market for Le Shuttle vs. ferry	No specific definition (London to Paris > 200 miles)	Nested logit route choice model, linear regression growth model	Business, Motor inclusive tour, Day trip, Short stay, Long holiday, Caravans	Time-series models predict total trips, pivot from flows on existing alternatives	Determine a price elasticities/cross-elasticities across days, time periods	UK Government International Passenger Survey, stated preference surveys, road networks, economic data
Union Railways	UK to France, Belgium and beyond	Market for Eurostar trains from London to Brussels and Paris	No specific definition (London to Paris > 200 miles)	Choice model of mode, station, access mode. exponential model with accessibility LogSums for trip generation	7 trip purposes, UK residents vs. nonresidents, group size, trip duration, and season	Sample enumeration, pivot from flows on existing alternatives. Growth factors from regression model	Viability of intermediate stations, elasticity of demand to price and travel time	Combined revealed preference- stated preference data, plus Stated Intentions data
Very Fast Train	Australia	A high-speed rail project from Sydney to Melbourne via Canberra	More than 200 miles	Mode-choice model	Business vs. leisure, SE variables are party size, duration of stay, age and income	Conventional model to project total demand, then sample enumeration	Specific to the Sydney-Melbourne project	Combined revealed preference-stated preference-stated Intentions survey
High-Speed Line – South	Netherlands to Belgium and France	HSL-S from Amsterdam to Paris via Brussels	No specific definition	Predictions for business-as-usual and diversions from that	Business, commuter and leisure, domestic and international flows	Spreadsheet forecasting model allows user adjustment, no explicit forecasts incorporated	Specific to evaluation of the HSL project	Revealed preference and stated choice data, network data and traffic counts
Scilly Corridor Model	SW United Kingdom	Mode choice for travelers to the Isles of Scilly	Islands are 28 miles from the English mainland	Choice between ferry, helicopter and (fixed-wing) plane	Visitors (staying on islands, day trippers), Residents (leisure, business), high and low income	UK government forecasts for population and income, sample enumeration in a spreadsheet	Specific to the ferry and its replacement	Revealed preference and stated preference surveys

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
Norwegian National Model	Norway	Developed to support national environmental policy, extended to consider infrastructure projects	Over 100 km (62 miles)	Predicts travel frequency, mode (car, air, rail, bus and boat), destination choice	Business, work/education, social, recreation and 'service' (personal business); SE variables in the trip frequency model	Documentation not available	Environmental policy, infrastructure issues, considered unsuitable for HSR modeling	Revealed preference data from the national transport survey
Norwegian High-Speed Rail Modeling	Norway	Under development to investigate potential for HSR	No specific definition	Mode choice (car, air, bus, existing train, proposed HSR)	Travel purpose	Conventional growth factors are applied to the matrices	Specific to the high-speed rail application	Stated choice survey, base trip volumes from Norwegian National Model
Swedish National Model	Sweden	Similar to the Norwegian model: distinguishes between long-distance and short-distance traffic	Over 100 km (62 miles), although some longer commute trips modeled as short distance	Nested logit models predicts travel frequency, mode, destination choice, applied to tours	Business and Leisure travel (which is segmented by duration of stay)	Documentation not available	Range of policy analyses, particularly HSR studies	National travel survey, detailed network data
Great Belt Study	Denmark	A model of travel between east and west Denmark	No specific definition	Nested logit model for mode, route choice, travel frequency, submodel forecasts new trip generation	Business and Leisure (which is segmented into single passengers and groups	Sample enumeration, pivot from existing flows. Trips forecast using population, economic growth, fuel prices	User-friendly system delivered to client testing demand sensitivity to pricing	Stated preference surveys and national travel survey
Fehmarn/Femern Belt Study	Germany/Denmark	Model passenger and freight traffic across the Femern Belt	No specific definition (Copenhagen to Hamburg is 180 miles)	Discrete choice model of mode and route choice	Business and Leisure (which is segmented into single passengers and groups	Full details not available, included forecasting economic/trade development	Specific to the evaluation of the project; simple sketch version used for fast scenario testing	Revealed preference and stated preference surveys
UK Long-Distance Model	Great Britain (does not include Northern Ireland)	Intended to study a wide range of policy relating to long-distance travel	Trips over 50 miles entirely within Great Britain plus airport access trips over 50 miles	Nested logit choice model for demand, network models for highway, air and rail	Commute/ education, business and other, plus SE variables income, auto ownership, etc.	Population, auto-ownership forecasts are inputs, pivots from observed data, preloaded networks	Initial use is for HSR tests	Revealed preference data from the National Travel Survey, additional trip diaries, stated preference survey

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
PLANET Long Distance	England	Forecasting for proposed new HSR line (“HS2 ”) in England	All air and rail trips, car trips over 30 km (see text for other complexities)	Emphasis on railway assignment, predicts growth for car, air and rail independently	Business, commute and other travel; auto ownership	External mode-specific forecasts, then incremental modeling of mode choice and changed travel frequency	Specific to the evaluation of the project	Rail ticket sales, rail on board surveys, National Transport Model, stated preference survey
Dargay’s Model	Great Britain (does not include Northern Ireland)	Model for long-distance travel in GB	Over 50 miles	Mode, purpose and trip length components, each with elasticities with respect to a range of variables	Business, commuting, leisure day trips, visiting friends and relatives and holidays; trip length split at 150 miles	2030 forecasts, uses government forecasts of GDP and fuel price	Fast model to test a range of policy and other scenarios	National travel survey, national aggregate time-series data, survey of long-distance travel
UK National Transport Model	Great Britain (does not include Northern Ireland)	Predicts traffic volumes for surface modes	Long-distance traffic is included in the model but is known to be poorly represented	Logit model for mode and area type. Highway and rail trips iterate with a network model	8 travel purposes, plus SE variables	Trip generation input from National Trip End Model	Wide range of policy appraisal by the Department for Transport	UK National Travel Survey, national population and employment data, network data
MATISSE	France and Europe	Used for a number of French and international studies in the 1990s	Unclear from documentation	All-or-nothing assignment to mode and route for each segment and for O-D pair	15,000 segments: value of time, preferred journey time, frequency of travel, length of journey and party size	MATISSE model embedded in a system of models that give forecasts of population and transport supply	HSR applications in France and neighboring countries	Unclear from documentation
European-level Models	Europe	TransTools2 is a continental scale model to provide input to European infrastructure planning	All trips included except intrazonal, therefore minimum around 30 miles; separate models for trips over 100km (62 miles)	Nested logit model of travel frequency, mode and destination, linked to assignment models. Tour based	Business/ commuting, holiday, social and recreation	Forecasts are assigned to networks for highway, rail and air, pivots from base matrices	Model is the standard for applications involving the Trans-European Networks	DATELINE survey
Netherlands National Model	Netherlands	Extensively used for range of transport policy analysis and updated for 25 years	Model does not separate out long trips	Tree-nested logit models of travel frequency, mode, destination, time period, iterated with network models. Tour based	Detailed (several hundred) based on SE variables, separate models for each trip purpose	Choice models used to pivot from observed highway traffic matrices. Population growth modeled	Infrastructure, pricing and management policies. Hundreds of applications have been made	National travel survey (OVG), network data, stated preference surveys, and many other sources

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
Airport Catchment Competition Model	Netherlands	Assessment of policy relating to the national airport, Schiphol	Only trips beyond the catchment area, which includes the major competing airports; effectively over 300 miles	Nested logit model of access/egress mode choice within the catchment area, route choice for air journeys, main mode choice (car, train and air)	Business and Leisure	Forecasts of total traffic using growth factors	Used to formulate national policy on the development of Schiphol	Passenger counts for existing services, base case supply data
UK Air Travel Demand Models	UK	National air passenger demand model and allocation model	All domestic and international scheduled and charter air trips	Demand model is a time-series model, logit model of airport choice	Country of residence (UK vs. foreign), purpose (business vs. leisure), and destination (domestic vs. international by region)	Forecasts are produced up to the year 2050. Allow for uncertainty relating to GDP, oil prices and fuel efficiency	Predicts demand by airport, which gives air traffic movements and CO2 emissions	Historic GDP data, forecasts for future GDP growth, UK airport survey data
Portuguese National Model	Portugal	Model to evaluate transport measures that have a regional or national impact	Over 50 km (31 miles), but excludes air travel	Models choice of mode (multinomial logit model), destination and route	Business, commuting and leisure; auto ownership	Currently base-year only	None yet, but intended for testing HSR proposals	Household long-distance survey, network data
Generation, Distribution And Mode Choice Model	Northern Chile	Model of interurban trips in the north-central macrozone of Chile	<150km short trips, 150-500km medium, >500km long trips	Multinomial/Nested logit models: trip generation, trip length, joint choice of distribution and mode	Short, medium, and long trips; home-based (BH) and non-home-based (NHB) trips; 8 segments by trips purpose/ season/party size	Pivots from existing flows, no long-term forecasts were produced	Produces matrices of daily journey numbers by distance and modes for the forecast period	Intercept O-D survey in the region
KITE Long-distance Study	Switzerland, Portugal and Czech Republic	Long-distance survey for Switzerland, Portugal and the Czech Republic	Over 100km (62 miles)	Hazard models of time between long-distance trips, mixed logit models of mode choice	Distance, mode of travel, income, age, education and gender	Not relevant	Not relevant	KITE survey has both revealed preference and mode choice stated preference

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
Invermo Study	Germany	Person-level simulation model of intermodal trips	Over 100 km (62 miles)	Trip generation (synth. population, simulation), destination choice (gravity model), mode choice (logit model)	Work related, visiting friends or relatives, other personal journeys of under three days, and holiday travel over four days	Microsimulation model	Case study of air-rail interchange in Cologne/ Frankfurt/ Stuttgart	Large-scale revealed preference study
Italian High-Speed Rail Study	Italy	Demand for HSR in Italy	Routes vary from 124 km (75 miles) to 720 km (450 miles)	linear regression model of future O-D volumes, nested logit mode-choice model, induced demand model	Business and nonbusiness	Documentation not available	Predict impacts of HSR services	Revealed preference-stated preference survey for mode choice, before and after data for induced demand
California Statewide High-Speed Rail Ridership	California	HSR ridership and revenue for proposed line	Interregional trips segmented at 100 miles into short and long trips	Interregional model: frequency, destination choice and mode choice; then assignment of all trips including urban and external	Business, Commute, Recreation and Other; other variables such as household size, income, number of workers	Standalone program that loops on TAZ residence/ segment combinations, CUBE voyager for assignment	EIRs and revenue forecasting of various HSR alignments	Revealed preference an d stated preference surveys, Household Travel Surveys by Caltrans, SCAG, MTC, SACOG, network data and Census data
Integrated Florida Statewide Model	Florida	Model system including passenger and freight components	Travel greater than 40 miles from home	4-step model ("mode" is just auto occupancy, assignment is joint with freight)	Separate long-distance generation and distribution models; trip purpose are business and four types of visitor	Land-use/ population growth factors supplied by the local MPOs	Range of statewide and regional transportation planning studies	NHTS, ATS, and Florida Visitors Survey
Indiana Statewide Travel Demand Model	Indiana	Full planning tool, includes long-distance trips	Defined subjectively by the decision-maker	4 step model, with separate procedures for long-distance trips (except combined assignments)	Within the state vs. one trip end outside the state	Separate population and employment forecasts feed the model	Corridor studies, state Lrevealed preference, developing inputs to HERS	Indiana travel survey, NHTS, revealed preference data, Census, DOT road inventory

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
Kentucky Statewide Travel Model	Kentucky	Model of long-distance trips within KY and roughly halfway into neighboring states	Over 100 miles	US Macro model (3-step, no mode choice), combined with micro model of Kentucky (denser network)	Business, tourism, or other	Fratar method to forecast population growth related to each trip purpose	Studies to determine traffic volumes, economic impacts of new corridors	ATS, National Highway Planning Network, HPMS
Michigan Statewide Travel Demand Model	Michigan	4-step person-trip model for all motorized ground transportation	Any trip between urban areas	4-step model	Home-based trips (work, vacation and other), and non-home-based trips	Regional economic model forecasts SE inputs	Used for making development decisions, including bypass, freeway and/ or local road alternatives	Census, state employment agencies, roadway inventory, traffic counts
New Hampshire Statewide Travel Model	New Hampshire	Tour-based model of all travel in the state	Long-distance trips are included as tours, but not defined explicitly	Tour-based microsimulation framework of sequential multinomial logit models	Work, school, other, shopping, recreation, and chauffeuring	Uses a sample enumeration technique, SE data for 1990, 2005, and 2020	Capable of analyzing major projects and pricing policies, producing VMT estimates for air quality, external flows for MPOs	New Hampshire Activities and Travel Survey, transit surveys, stated preference survey, Census, network data
Ohio Long-Distance Travel Model	Ohio	Tour-based long-distance model	All non-work tours to destinations more than 50 miles from home	A series of (mainly logit) models of travel or not, tour pattern scheduling, internal/external, destination, mode	Household travel, work-related travel and other travel; household characteristic	Not available	It can be used to evaluate HSR	Ohio Statewide Household Travel Survey
Oregon Statewide Integrated Model	Oregon	The Long-Distance Transport (LDT) module is part of SWIM2	Noncommuting trips longer than 50 miles, but still within the state	A series of (mainly logit) models of travel or not, tour pattern scheduling, internal/external, destination, mode	Household trips (entire household), work-related trips (individual), individual non-work trips	Synthetic population is the main SE input	Sensitivity tests (e.g. increasing capacity, increasing travel costs, changing density) for 2006 to 2024	NHTS data
Tennessee Statewide Model	Tennessee	3-step statewide model, only long-distance trips	Over 75 miles	Trip generation, trip distribution and trip assignment (no mode choice)	Home-based work, home-based others and non-home-based, but combined during distribution	Uses MPO forecasts of population and employment growth	Lanes, ADT on new highways, pavement management plans for interstates	1990 NPTS and 1990 CTPP

Model	Region	Context	Definition	Structure	Segments	Forecasting	Applications	Data
Wisconsin Multimodal Intercity Passenger Demand Model	Wisconsin	Interurban model of all roads, including HSR	Over 50-miles between states, counties, and major urban areas	Cross classification trip generation mode, destination and mode-choice models that are run simultaneously	Business, personal business, and pleasure-related travel	Based on population, economic activity forecasts, induced growth	Planning analyses at a statewide level, ridership and revenue estimates	2001 NHTS add-on

CHAPTER 3. DATA SOURCES

Before the long-distance passenger travel demand model could be applied to predict long-distance travel behavior of all the households in the United States, a number of datasets had to be prepared. A brief description of the key steps involved in the preparation of the application datasets is provided in the sections that follow.

3.1 Zone Systems

This section provides an overview of the development of a new zonal system for forecasting long-distance travel at the national scale.

Zone System Creation

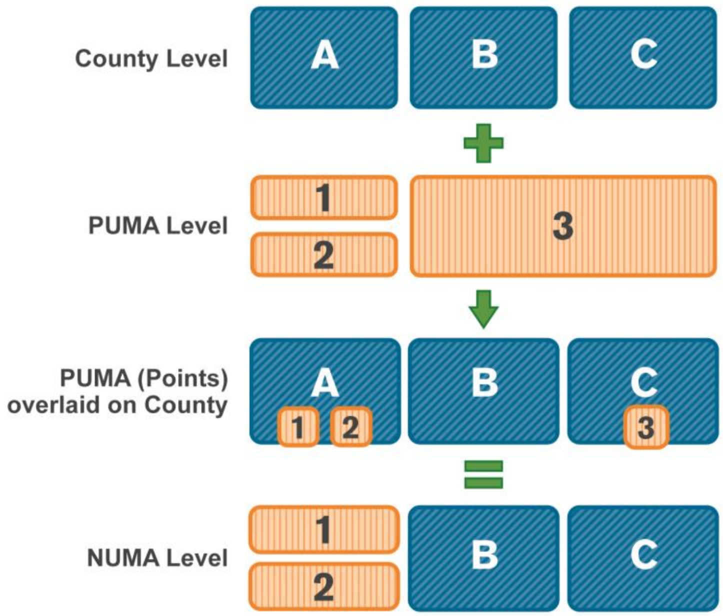
A new geographical construct termed the National Use Model Area (NUMA) was created and adopted for this effort. NUMA-level geography is a composite representation of counties and Census Bureau Public Use Microdata Areas (PUMAs) across the United States. The United States includes 3,143 counties and county equivalents (in 2013) and 2,378 PUMAs (as of the 2012 American Community Survey). Using counties or PUMAs as zones for a national-level travel model is appropriate; both offer a geographic resolution that may be considered reasonable from a long-distance travel perspective, and the number of geographical units is consistent with the number of zones typically seen in large-area travel models.

Census Tracts were considered as a means to enhance the level of detail in the zone system, but with approximately 75,000 Census Tracts, it was found to be computationally prohibitive to adopt the Census Tract as the geographic basis for defining national travel model zones. Census Tracts were found to add detail for access and egress to air and rail stations, by building travel paths that connect a Census Tract at the origin to an origin station, then connecting the origin station to the destination station, then connecting the destination station to the destination Census Tract. This method of multilevel geographies for evaluating travel paths has been implemented in urban activity-based models and was selected as the preferred method for the integrated modeling system framework (Chapter 3), but was not included in the demonstration system (Chapter 4). To support this effort, the Census Tract was implemented for synthetic-population generation.

In comparing the relative sizes of counties and PUMAs, it was clear that these geographical units should not be used as zones without some additional transformation. The sizes of these geographical units vary widely throughout the country; in some instances, multiple counties constitute a single PUMA, and in other instances, multiple PUMAs constitute a single county. In order to define the geographic zone system for the national travel model developed in this study, it was decided that the smaller of the two geographies should be used to define the NUMAs. Thus, in a situation where multiple counties comprise a single PUMA, the county was selected as the NUMA (the smaller of the two); where multiple PUMAs comprise a single county, the PUMA was selected as the corresponding NUMA (again, the smaller of the two). In this way, the zone system adopted for this effort would offer a reasonable geographic representation that is neither too large nor too small in its definition in the context of modeling long-distance travel.

The NUMA generation exercise was performed in GIS software. The PUMA-level shapefile (a point file) was overlaid on top of the county-level shapefile (a polygon file). The number of PUMAs that fall within each county were then counted. If the total number of PUMAs in a county was greater than one, then the PUMA was selected as the geographical resolution; otherwise, the county was selected as the geographical resolution for defining NUMAs. The process followed is shown in Figure 1.

Figure 1: Procedure to Define NUMA-Level Geographical Resolution



In Figure 1, assume that the first panel (blue-colored blocks with diagonal lines) represents the counties in a (hypothetical) state. There are three counties in the state in this illustrative example: county A, county B, and county C. The second panel (orange-colored blocks with vertical lines) depicts the PUMA-level geographical layout of the same state (PUMA 1, PUMA 2, and PUMA 3). To arrive at the NUMA-level geographical representation, the PUMA-level polygon file is converted to a point file and overlaid on the county-level polygon file (third block). It can be observed from the figure that county A encompasses two PUMAs; therefore, the level of geographical representation that is adopted here is the PUMA. County B has zero PUMAs within its boundary and county C has a one-to-one correspondence between a county and PUMA. In both of these situations, the county is selected as the geographical resolution to define NUMAs. The NUMA-level geographical representation for the state is shown in the final block as a mix of counties and PUMAs. In this example, the NUMA-level representation for the state consists of four geographies (two counties and two PUMAs).

This methodology was applied across all states (and for each county and PUMA within a state) to select the preferred geographical resolution that would define the NUMA. Following this exercise, consistency checks were performed to ensure that no geography was left unrepresented. A few anomalies were identified in this process, due to minor inconsistencies in the alignment of PUMA- and county-level geographical boundaries.

Figure 2 depicts an example for the State of Arizona. This figure depicts the NUMA-level geographical representation for Arizona. The PUMA is the selected geographical level in Part A of the figure, and county is the selected geographical level in Part B of the figure. This leaves the white region neither selected as part of the PUMA representation in Part A nor as part of the county representation in Part B. These anomalies were resolved by manually selecting the geographical resolution in such a way (i.e., larger geography is selected) that no area was left unrepresented. In the instance depicted in Figure 3 shows that county is selected as the preferred geography to define a NUMA, even though the county encompasses multiple PUMAs. This ensured that no area was left unrepresented in the NUMA-level geographical representation. There were 19 such occurrences across the United States, where the larger geography (between county and PUMA) was manually selected to ensure completeness in areal coverage and geographical representation. The NUMA-level geographical file generated from this exercise comprised 4,381 NUMAs.

A key issue that arises in the context of the aforementioned procedure is that a NUMA may be quite large, encompassing a large area. The procedure involves the selection of the "larger" geography (county vs. PUMA) to ensure that no land area remains unaccounted for in the NUMA representation. In the 19 instances where this procedure was invoked, this procedure was particularly troublesome in only one case, where the resulting NUMA was a large county (Miami-Dade County) comprising more than 500 Census Tracts. Only 2% of the NUMAs have more than 50 Census Tracts. This one NUMA with 500+ Census Tracts has a population of 2.5 million, which is much larger than the population of any PUMA (PUMAs are defined as geographical areas with ~100,000 population). The procedure for developing NUMAs was found to be considerably robust, excepting this particular instance, where the size of the NUMA was quite large. In the future, such anomalies can be easily rectified by manually splitting large NUMAs into groups of Census Tracts. Rule-based heuristics would have to be established to implement such a procedure (e.g., rules defining the maximum number of Census Tracts per NUMA, and the maximum population per NUMA).

Figure 2: Inconsistencies in Geographical Representation—Panel A

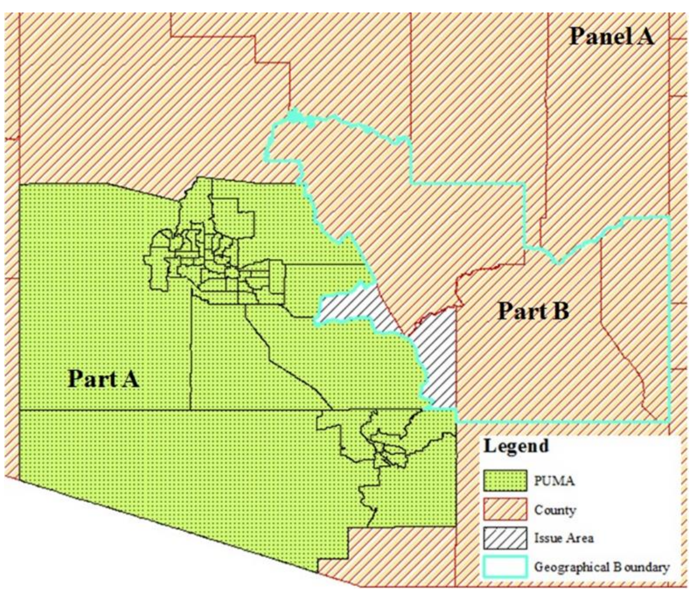
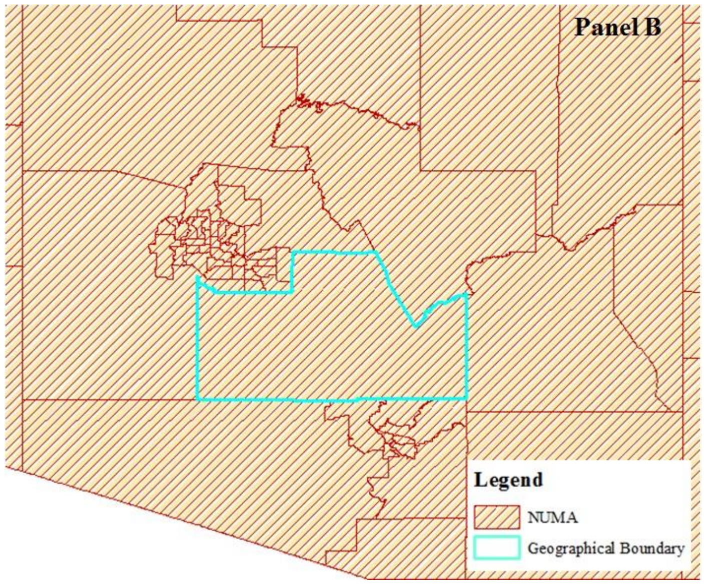


Figure 3: Inconsistencies in Geographical Representation—Panel B



Following the aforementioned initial NUMA generation exercise, the NUMAs were further split so that no NUMA had more than one airport. Major airports across the nation were converted to a GIS-point shapefile and overlaid on the NUMA polygon file. Only six NUMAs across the United States had more than one airport located within the NUMA boundary. In cases where a NUMA had more than one airport, and the Census Tract-level geography allowed for a clear demarcation, the NUMA was split along the Census Tract boundary (Figure 4). In the figure, the NUMA is shown with a border and the dots represent the airports in the NUMA under consideration. Since the Census Tract boundary allowed for a clear demarcation, the NUMA was split along the Census Tract boundary as shown in the figure. In cases where it was not possible for a clear delineation (because the NUMA split involved multiple Census Tracts), NUMAs were split as close to the Census Tract boundary as possible, as shown in Figure 5.

Figure 4: NUMA Split Exactly Along the Census Tract Boundary

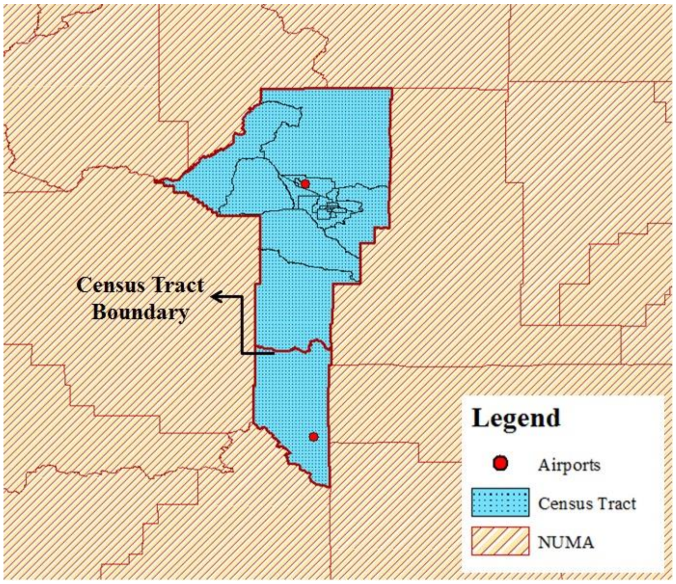
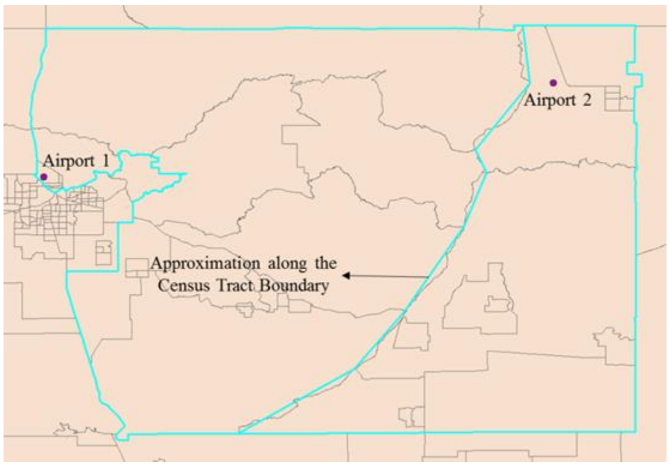
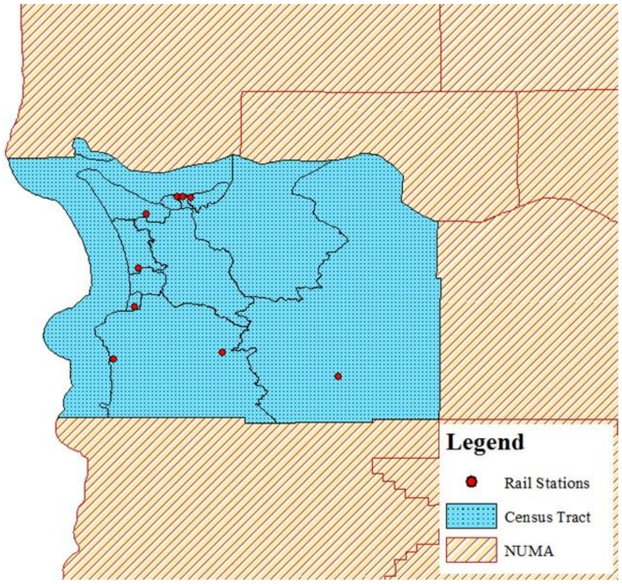


Figure 5: NUMA Split Approximately Along the Census Tract Boundary



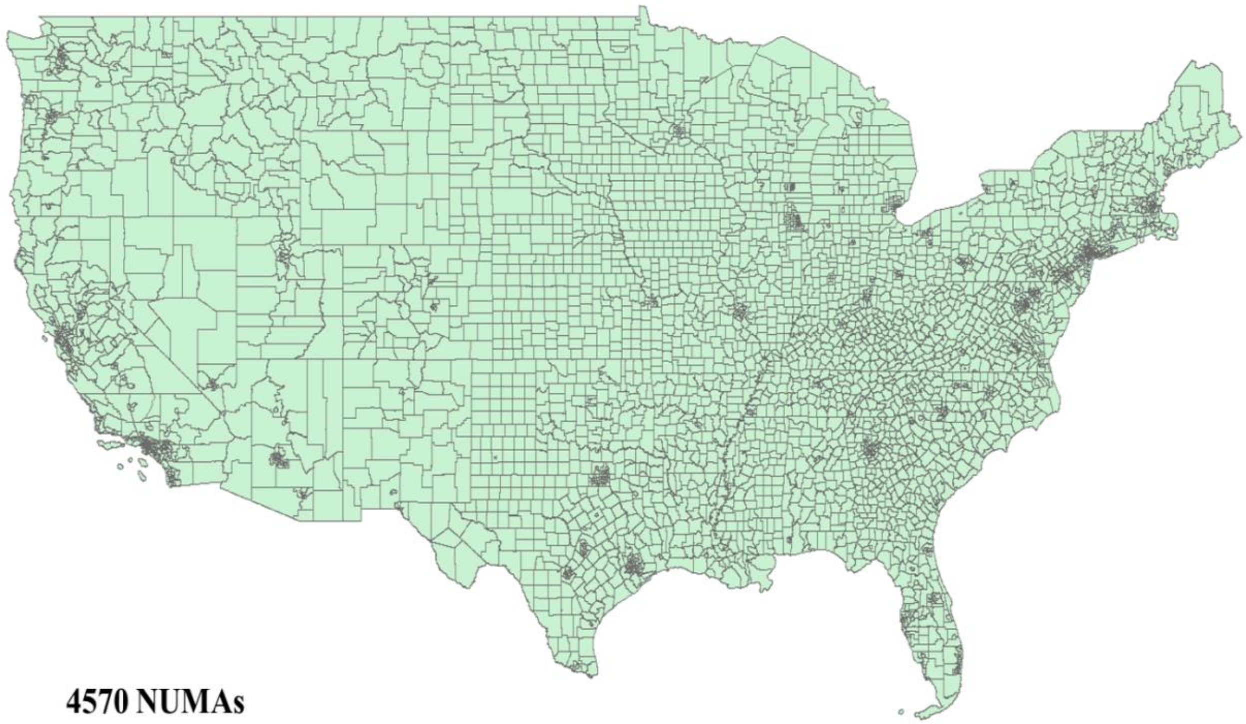
A similar exercise was performed for Amtrak rail stations. A total of 132 NUMAs had more than one Amtrak station within their respective boundaries. If a NUMA had multiple rail stations that were spatially separated, a process similar to the one previously outlined for airports was performed to split the NUMA into multiple NUMAs (such that each resulting NUMA had only one Amtrak rail station). However, for NUMAs with several rail stations located in close proximity to one another, NUMAs were split through a manual process so that the rail stations were dispersed across multiple NUMAs to the extent possible. Consider for example, the NUMA shown in Figure 6, which has nine rail stations. Given the level of geographical resolution, a prudent course was chosen, and this NUMA was not split into nine different NUMAs. Rather, a judgment was made to split this NUMA into three separate NUMAs. A similar inspection- and judgment-based approach was followed for all of the 132 NUMAs that had multiple Amtrak rail stations. As a result of this process, it is possible for some NUMAs (particularly in the dense Northeast) to contain "pockets" of closely spaced rail stations.

Figure 6: NUMAs with Multiple Rail Stations



After the NUMAs were split to account for multiple airports/rail stations, the final NUMA-level geographical file consisted of 4,570 NUMAs. All of the network level-of-service data for highway (auto and bus) modes follow this geographical resolution. The final NUMA map for the United States is shown in Figure 7. Following the creation of the NUMA polygon file, an equivalence table was generated between Census Tracts and NUMAs by overlaying the Census Tract point file on the NUMA polygon file.

Figure 7: Final NUMA Map

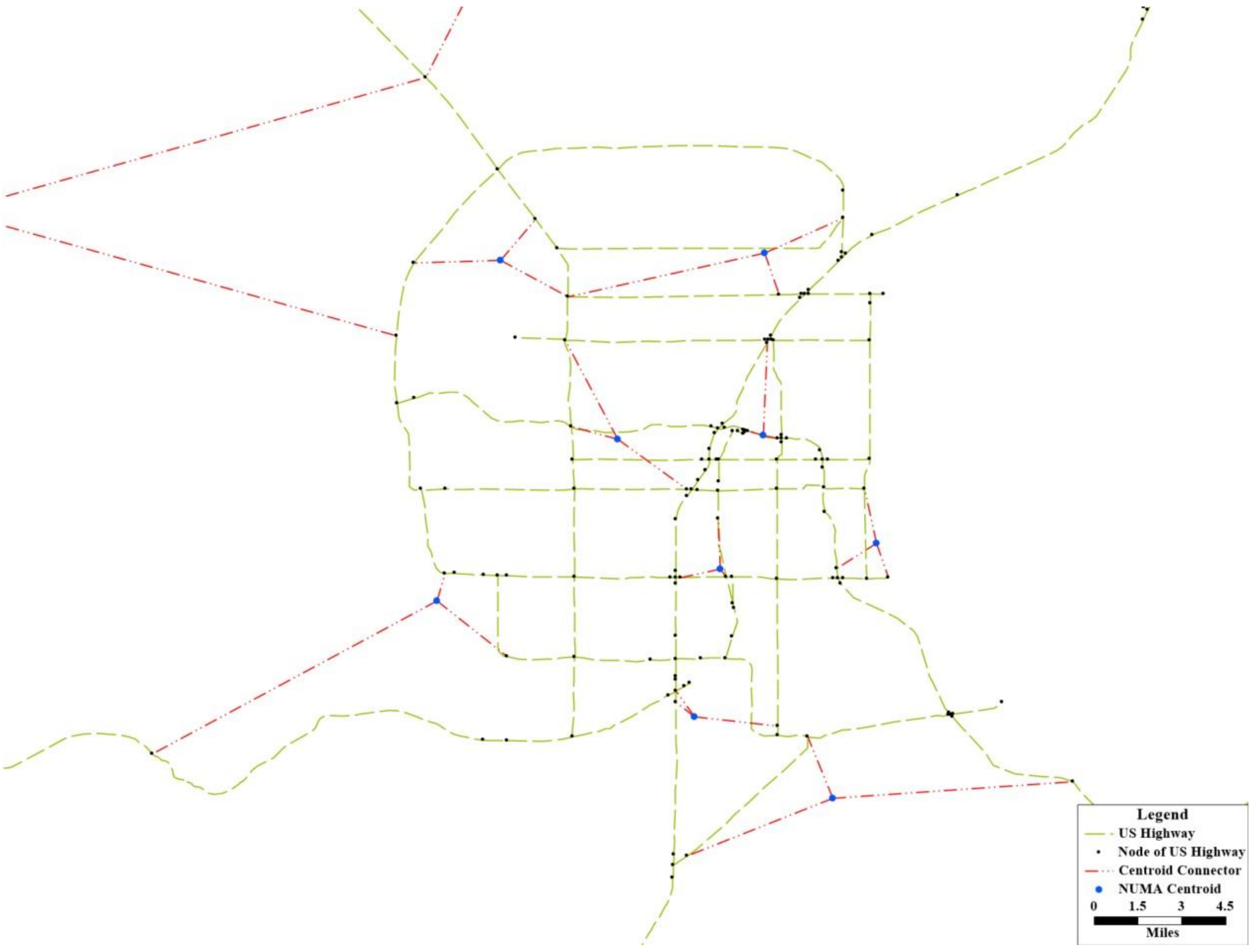


Zone Connectors

The NUMA polygon shapefile was imported into TransCAD and converted to a TransCAD geographic file. NUMA centroid locations (points) were generated from the NUMA polygon file automatically within TransCAD. The US highway network downloaded from the Federal Highway Administration (FHWA) Highway Performance Monitoring System website[1] was also converted to a TransCAD network file. The NUMA centroid point file was overlaid on the US highway network file and access connectors were generated from each NUMA centroid to the nearest highway link. Up to three highway connectors were generated for each NUMA, with an intent to mimic multiple entry points to a zone, subject to a distance threshold of 50 miles. Figure 8 shows an illustration of the centroid connectors generated using this procedure.

[1] Federal Highway Administration (FHWA) Highway Performance Monitoring System website

Figure 8: NUMA Centroid Connectors



3.2 Modal Networks and Level of Service

Road System

The National Highway Planning Network (NHPN) was used to generate estimates of travel time, distance, and cost in the form of highway skims. The NHPN, developed by FHWA, is a geospatial database that comprises interstates, principal arterials, and rural minor arterials (over 450,000 miles of existing and planned highways in the country). The most up-to-date highway network, which was published in 2011, was downloaded from the FHWA's website.[2] In addition, the network includes intermodal connectors that were linked with appropriate airports and rail stations.

The project team obtained distance and speed information for each highway link, along with toll information for different toll roads across the nation. This information was used to generate

[2] Most up-to-date US highway network, published in 2011, and is available from FHWA's website.

travel time, distance, and generalized cost skims for the NUMA-level zonal system. Procedures followed for each of these efforts are discussed in this section.

Auto Travel Time, Distance, and Cost

The network shapefile used to generate NUMA centroid connectors has information regarding distance (mile) and the posted speed limit (mph) for each link in the US highway network. This network file was imported to TransCAD and linked with the NUMA centroid file. Travel time to traverse a link was computed as distance divided by posted speed limit. Using built-in shortest-path computation methods in TransCAD, travel time and distance skims were generated for the US highway network. In addition, a generalized cost skim was also generated for the auto mode. Generalized cost to traverse a link was computed as follows:

Equation 1: Generalized Cost to Traverse a Link

$$\begin{aligned}&\textit{Generalized cost to traverse a link}\\&\quad= \textit{(value of time) x travel time (in hour) + (auto operating cost)}\\&\quad\textit{x link length (mile) + (toll per mile) x link length (mile)}\end{aligned}$$

Value of time ($17 per hour) and auto operating cost ($0.18 per mile) were used to compute generalized cost. These values may be adjusted during model calibration and validation as part of the implementation phase described in Chapter 9. These values may also be changed to assess sensitivity of travel demand to varying levels of value of time and auto operating costs. The toll per mile was computed based on the procedure described previously. The generalized cost value was computed for all links in the US highway network, and generalized cost skims were generated by minimizing the generalized cost across each NUMA pair. Travel time, distance, and generalized cost skim matrices were thus generated for the auto mode at the NUMA level (4570×4570 matrices).

Toll Facilities

Shapefiles containing information on the highway network attributes (at the link level) for the United States were obtained from the FHWA's Highway Performance Monitoring System website.[3] From these files, a subset of toll roads was extracted based on toll charge (>0) specified on the link. Supplementary information regarding toll facilities in the United States was obtained from FHWA's Toll Facility Information website.[4] Information from both of these sources was compared to ensure completeness of toll information data. The highway network shapefile did not designate several toll facilities that were reported in the supplementary toll information data. The missing toll facilities were manually digitized based on the supplementary information. The toll charge for missing facilities was imputed from the available data as the average of maximum and minimum toll charge for a passenger car. Directionality attributes for toll roads was also added manually after a visual inspection of the toll facilities in Google Earth. The toll roads

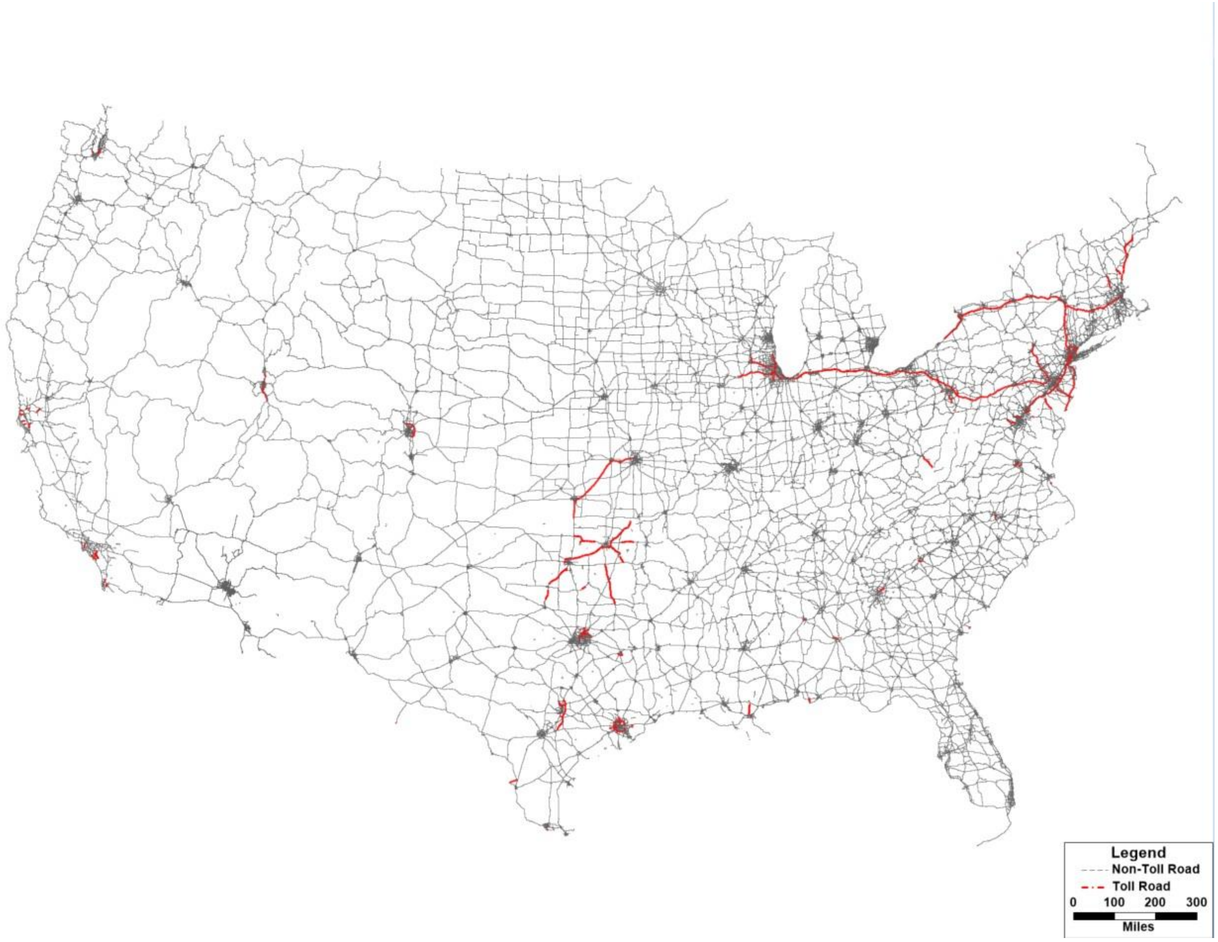
---

[3] FHWA's Highway Performance Monitoring System website.

[4] FHWA's Toll Facility Information website.

shapefile was merged with the rest of the US highway network shapefile to generate the highway network skims. The toll for each link on the highway network was represented on a per-mile basis (by dividing the toll cost by the length of the corridor). For links that did not have a toll associated with them, this value was set to zero. Figure 9 presents the US highway network with toll roads identified in red.

Figure 9: Toll Roads in the United States



Rail System

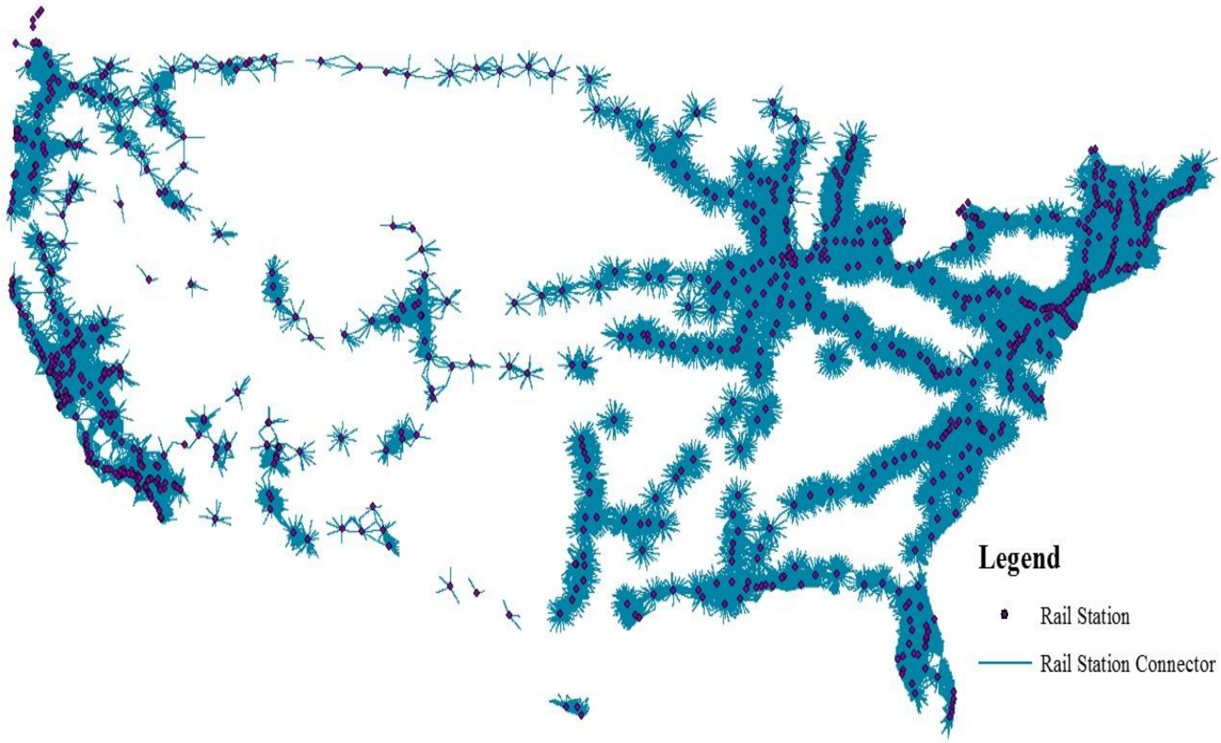
The rail network was developed from the Amtrak rail system (Figure 10). Additional commuter rail systems could be added, but these were not considered essential for this project.

Figure 10: Amtrak Rail Network

Rail Station Connectors

The project team generated access links for rail stations by creating connectors that linked each rail station to all Census Tracts that were within 50 miles of the station. To accomplish this, the rail station locations were first represented as points on the Census Tract (polygon) shapefile. Centroid locations were identified for all of the Census Tracts in the Census Tract polygon file. A circular buffer region, with a 50-mile radius, was created for each rail station. All of the Census Tract centroids that fell within the 50-mile buffer region of a rail station were selected, and a rail station connector was generated to each Census Tract within the buffer region. The spider network created from the generation of rail station to Census Tract connectors is shown in Figure 11. A Census Tract is allowed to have a connector to all rail stations within 50 miles from the location of its centroid.

Figure 11: Rail Station-to-Census Tracts Connectors



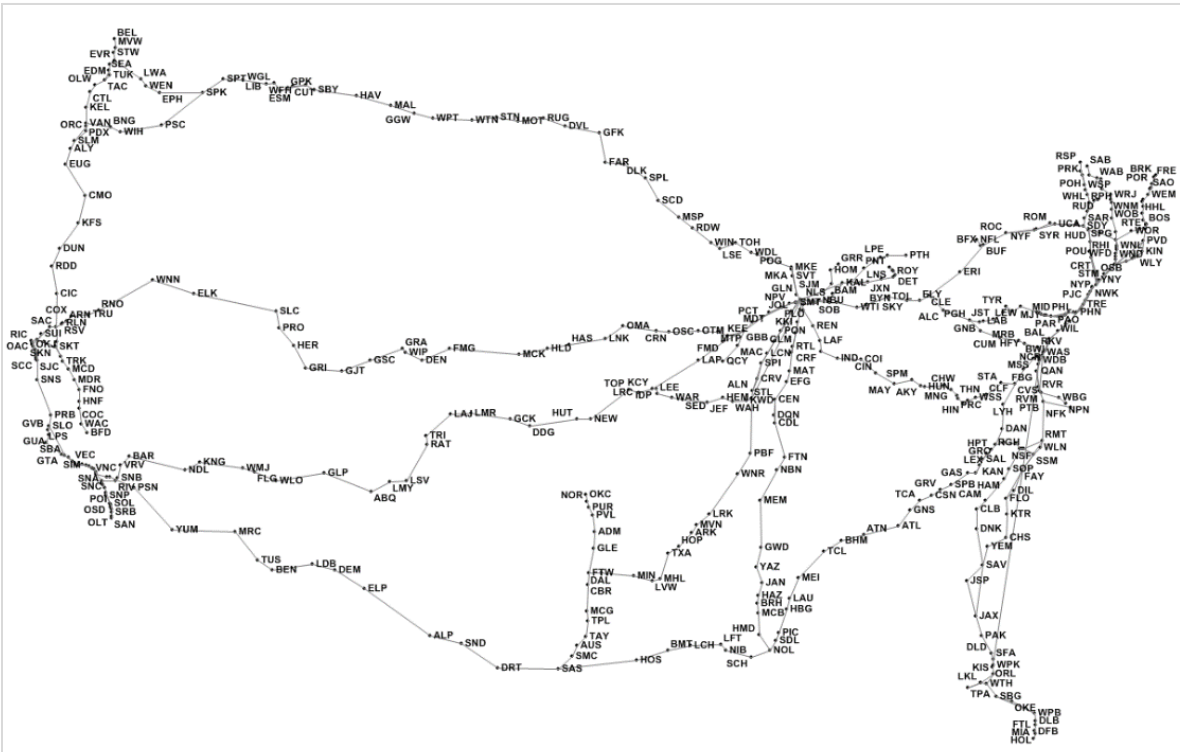
Rail Travel Time, Distance and Cost

Amtrak's General Transit Feed Specification (GTFS) data were processed and analyzed to construct level-of-service measures for the national rail network. The GTFS data comprise the following information on various services operated by Amtrak across the nation:

- **Agency**: Contains information on all of the transit agencies that provided data in the transit feed.
- **Calendar**: Contains the dates on which a particular service operates; data regarding start and end times of the service, and the days of the week on which the service operates.
- **Routes**: Contains information regarding transit routes; a route is defined as a group of trips (or consecutive stops) that are displayed as a single service.
- **Shapes**: Contains the rules for drawing lines on a map to represent routes.
- **Stop Times**: Contains arrival and departure times of the train at the stop level.
- **Stops**: Contains the geolocation of individual stops.
- **Transfers**: Defines the rules for making connections at transfer points between routes.
- **Trips**: Contains information at the trip level for each route; a trip is a sequence of two or more stops.

The GTFS data were imported to TransCAD using inbuilt functions in the software. TransCAD aggregates these files as inputs and generates node- (representing Amtrak stations) and link-level (representing Amtrak routes) geographical files. The Amtrak network generated by TransCAD is shown in Figure 12.

Figure 12: Amtrak Rail Network Generated from TransCAD



A manual inspection was performed to ensure that the Amtrak network was represented accurately by the output generated from TransCAD. The Amtrak network consists of a total of 43 rail routes and 518 rail stations.

From the Amtrak GTFS data, travel time and stop (dwell) time were extracted at the level of each individual link on the rail network. A transfer-time table, which defines the transfer times at all links where a route transfer is feasible, was also generated from the GTFS data. A network file was generated in TransCAD based on the link and node layers created from GTFS data. Each link on the network had three attributes assigned to it: 1) travel time; 2) stop time; and 3) transfer time. Travel time to traverse a link was computed as the sum of these three link attributes. Skims were generated for the rail network at the stop level by minimizing travel time between each station pair. TransCAD provides inbuilt functions to generate a distance skim corresponding to the travel-time skim. The travel-time and distance-skim matrices generated for Amtrak rail network were generated at the station level (518×518 matrices).

Transfer-Frequency

The transfer-frequency matrix defines the minimum number of transfers a traveler needs to make to travel from one Amtrak station to another. Two sets of travel-time skims were generated in TransCAD, employing the procedures described in the previous section (i.e., one skim where transfer time is included in the computation of total travel time, and another skim where transfer time is excluded). The difference between these two skim matrices provided the total transfer time between any Amtrak station pair. Based on a detailed analysis of the data, transfer times were defined as either short (one minute) or long (one hour) transfer times. Using a series of logic checks and count-calculation procedures, the number of short and long transfers was computed from the transfer-time matrix. The number of short and long transfers were then added together to obtain the total number of transfers between a station pair.

Rail Frequency

An innovative methodology was developed by the project team to obtain the operating (service) frequency between each Amtrak station pair. First, frequency lookup tables were created for all routes by manually parsing the Amtrak website. Information regarding frequency of operation on weekdays and weekends was collected for all 43 Amtrak routes. Using these data, average daily frequency and weekly frequency was computed for each route.

As part of the methodology, 43 Amtrak route variables (represented as columns) were created in the link files generated by TransCAD from Amtrak GTFS data. Each link on the Amtrak rail network was assigned to a unique route using a binary (0/1) indicator. There were a few links on the rail network that were common to multiple routes, and these links were assigned to the route with the highest daily frequency. For any given Amtrak station pair, if a route matrix has a nonzero entry, it implies that the specific route is used in computing the shortest travel time path between the station pair under consideration. For each station pair, a query was run across the 43 route skim matrices to identify all routes that were included in the shortest-path computation.

After all routes involved in the shortest-path computation were identified (for each station pair), the frequencies of all of these routes were obtained from the frequency lookup table. The route with minimum (lowest) frequency among those selected or included on the path defined the operational frequency for Amtrak services between a given station pair. For example, to travel from Dodge City in Kansas to Poplar Bluff in Missouri, the shortest path involves traveling on three different Amtrak routes: the Southwest Chief, the Missouri River Runner, and the Texas Eagle (shown in Figure 13). The operational frequency of Amtrak service between these two station pairs is one train per day, which is the minimum of the operating frequencies of the three routes involved in shortest-path computation between these stations. The aforementioned procedure systematically computes this frequency. Manual checks were performed to see how accurately this methodology was able to depict the operational frequencies for several station pairs and the results confirmed that the frequencies were accurate. Separate operating-frequency matrices were generated at the day and week level to account for differing temporal windows of interest.

Rail Fares

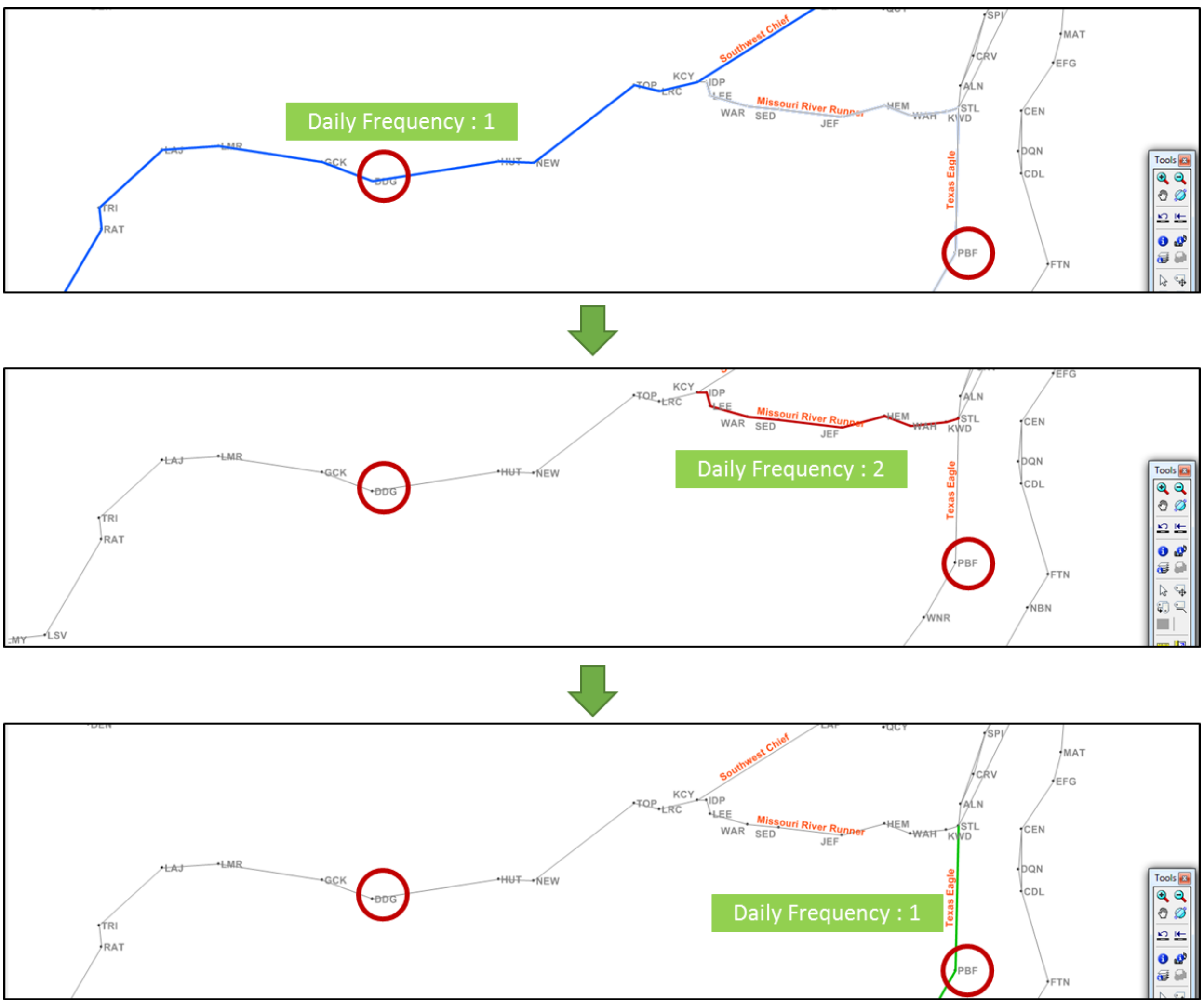
Generating station-to-station rail fare matrices involved two key steps:

1. Estimating models to predict one-way average rail fare, by class.

2. Applying estimated models to generate station-to-station O-D fare matrices.

For the first step, a number of linear regression models were estimated using 2004 rail fare data, obtained by the research team from Amtrak under a confidential agreement. This was a national dataset that included over 34,000 raw records and contained information on origin station, destination station, route, fare class, ridership, ticket revenues, and passenger miles traveled.

Figure 13: Rail Frequency Computation—Illustrative Example



For model estimation purposes, the average fare between an O-D pair was calculated from ticket revenues and ridership information. To be more consistent with base years that were used to derive level-of-service variables for other modes, 2004 rail fares were factored up to 2012 levels by using Consumer Price Index (CPI) values for US city averages for transportation between the years 2004 and 2012. Next, for each fare class, separate models were estimated for the following six regions to capture regional variation in rail fare:

- **California (CA)**
- **Midwest (MW)**: Includes Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Ohio, and Wisconsin.
- **Northeast (NE)**: Includes Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and West Virginia.
- **Northwest (NW)**: Includes British Columbia Oregon and Washington.
- **South (S)**: Includes Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia.
- **West (W)**: Includes Arizona, Colorado, Idaho, Kansas, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Texas, and Utah.

The estimated models describe the relationship between rail fare and rail-trip distance. A number of functional forms of the dependent variable (such as fare and logarithm of fare) and the independent variable (distance, distance square, and logarithm of distance) were tested; the final model results are summarized in Table 3. As shown, rail fare appears to have a polynomial relationship with trip mileage—the extent of this relationship varies by geographic region and fare class. Due to the polynomial specification of the model, it was necessary to impose a restriction to ensure that fare will only increase as the mileage increases. Figure 14 through Figure 25 present observed average rail fares and mileage of rail trips by region and fare class (in hollow blue circles). The figures also show model-predicted average rail fares (in hollow red circles), which track well with the observed data. In addition, the model-predicted fares were compared against Amtrak's online reservation fares for a limited number of station pairs with mixed demand. To conserve space, validation results for only California are presented in Table 2). Given that online rail fares show wide variation and depend, to a certain degree, on travel dates and how far in advance the reservations were made, the overall model performance was found to be within acceptable limits. The estimated models were applied to generate station-to-station average rail fare matrices, by class. These matrices were then converted to zone-to-zone matrices.

Table 2: Observed vs. Estimated Rail Fare—California

Origin State (Station Code)	Destination State (station code)	Fare Class	Predicted 2012 Fare (A)	Amtrak 2014 Fare (B)	Percent Difference in Fare (B-A)
California (SAN)	California (LAX)	First/business	60.96	56.00	9%
California (SAN)	California (LAX)	Economy	28.29	37.00	-24%
California (EMY)	California (SAC)	First/business	44.23	56.00	-21%
California (EMY)	California (SAC)	Economy	22.62	29.00	-22%
California (SAC)	Illinois (CHI)	First/business	509.02	792.00	-36%
California (SAC)	Illinois (CHI)	Economy	163.28	157.00	4%
California (MTZ)	Nevada (RNO)	First/business	81.85	140.00	-42%
California (MTZ)	Nevada (RNO)	Economy	35.35	75.00	-53%
California (LAX)	Arizona (MRC)	First/business	158.50	165.00	-4%
California (LAX)	Arizona (MRC)	Economy	60.96	55.00	11%

Table 3: Rail Fare Model by Region and Fare Class

Region	First/Business Class	Economy Class
California (CA)	$fare = 16.89 + 0.39 \times distance - 0.07 \times \frac{distance^2}{1000}$ Sample size = 1,256 Adjusted R-squared = 0.85	$fare = 13.31 + 0.13 \times distance - 0.03 \times \frac{distance^2}{1000}$ Sample size = 1,877 Adjusted R-squared = 0.90
Midwest (MW)	$fare = 16.17 + 0.32 \times distance - 0.04 \times \frac{distance^2}{1000}$ Sample size = 1,762 Adjusted R-squared = 0.81	$fare = 7.76 + 0.16 \times distance - 0.04 \times \frac{distance^2}{1000}$ Sample size = 2,733 Adjusted R-squared = 0.88
Northeast (NE)	$fare = 32.57 + 0.37 \times distance - 0.09 \times \frac{distance^2}{1000}$ Sample size = 2,661 Adjusted R-squared = 0.77	$fare = 20.02 + 0.23 \times distance - 0.11 \times \frac{distance^2}{1000}$ Sample size = 3,674 Adjusted R-squared = 0.72
Northwest (NW)	$fare = 20.87 + 0.42 \times distance - 0.11 \times \frac{distance^2}{1000}$ Sample size = 607 Adjusted R-squared = 0.81	$fare = 10.51 + 0.17 \times distance - 0.05 \times \frac{distance^2}{1000}$ Sample size = 765 Adjusted R-squared = 0.88
South (S)	$fare = 36.32 + 0.33 \times distance - 0.06 \times \frac{distance^2}{1000}$ Sample size = 2,085 Adjusted R-squared = 0.61	$fare = 22.90 + 0.15 \times distance - 0.03 \times \frac{distance^2}{1000}$ Sample size = 3,108 Adjusted R-squared = 0.65
West (W)	$fare = 40.21 + 0.33 \times distance - 0.06 \times \frac{distance^2}{1000}$ Sample size = 1,531 Adjusted R-squared = 0.58	$fare = 14.27 + 0.15 \times distance - 0.03 \times \frac{distance^2}{1000}$ Sample size = 2,706 Adjusted R-squared = 0.78

Figure 14: Relationship between average rail fare and trip mileage (CA first/business class)

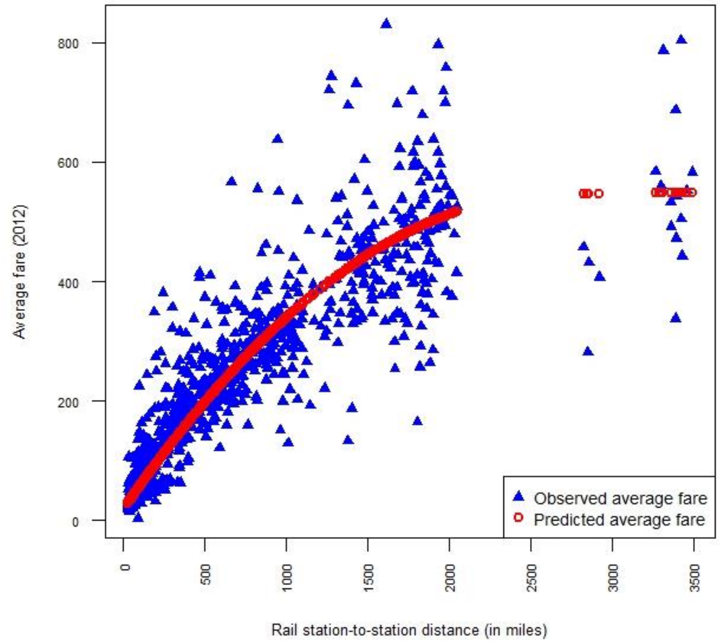


Figure 15: Relationship between average rail fare and trip mileage (CA economy class)

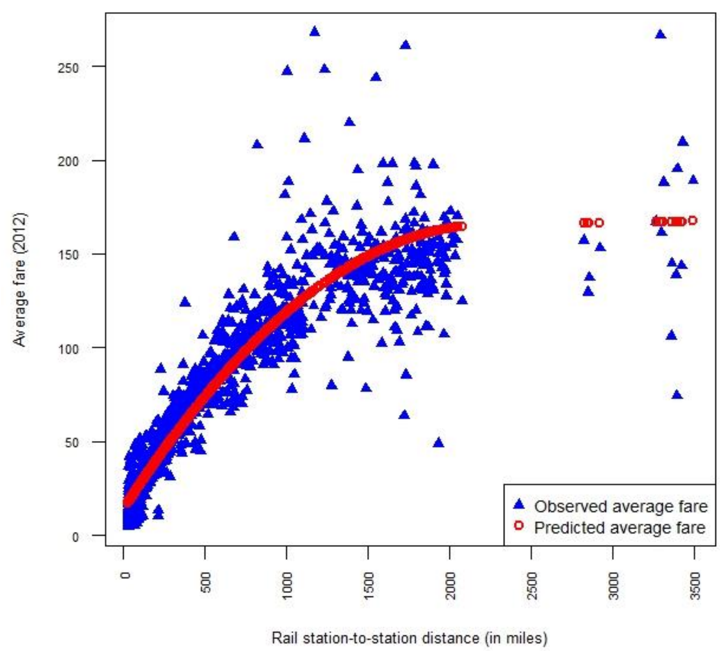


Figure 16: Relationship between average rail fare and trip mileage (Midwest first/business class)

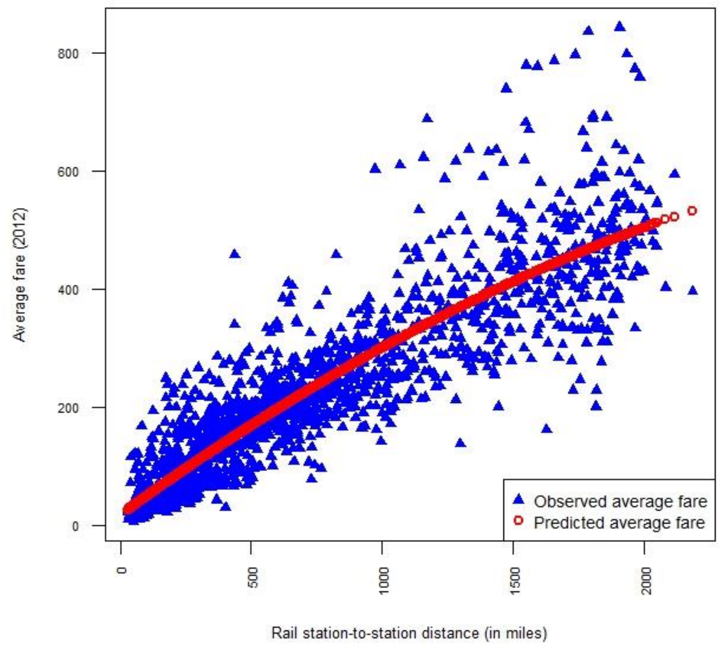


Figure 17: Relationship between average rail fare and trip mileage (Midwest economy class)

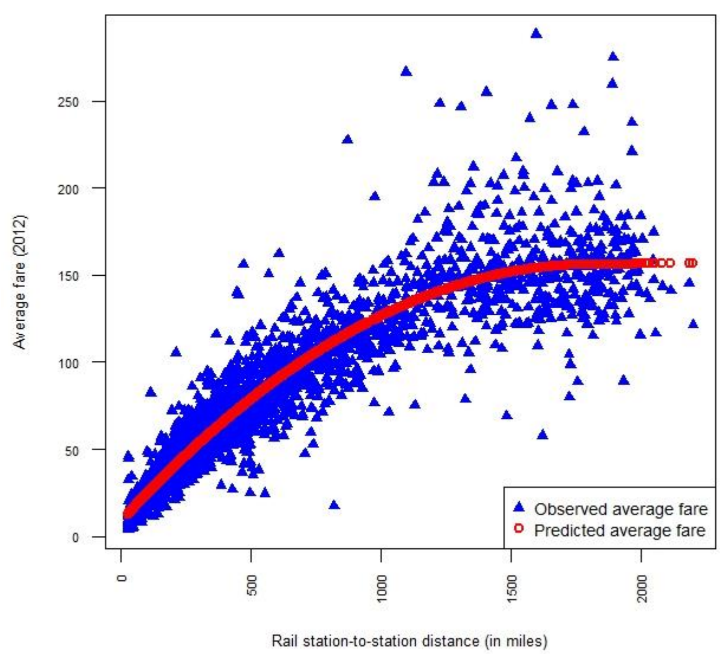


Figure 18: Relationship between average rail fare and trip mileage (Northeast first/business class)

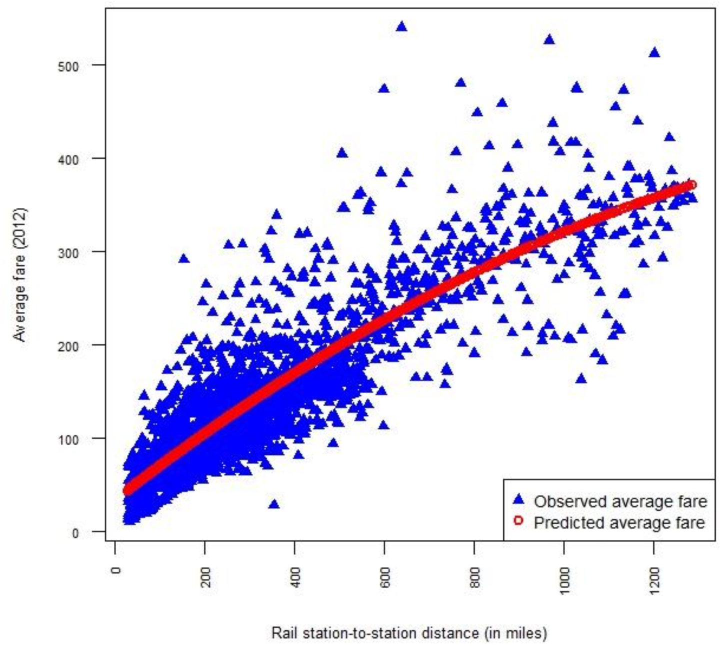


Figure 19: Relationship between average rail fare and trip mileage (Northeast economy class)

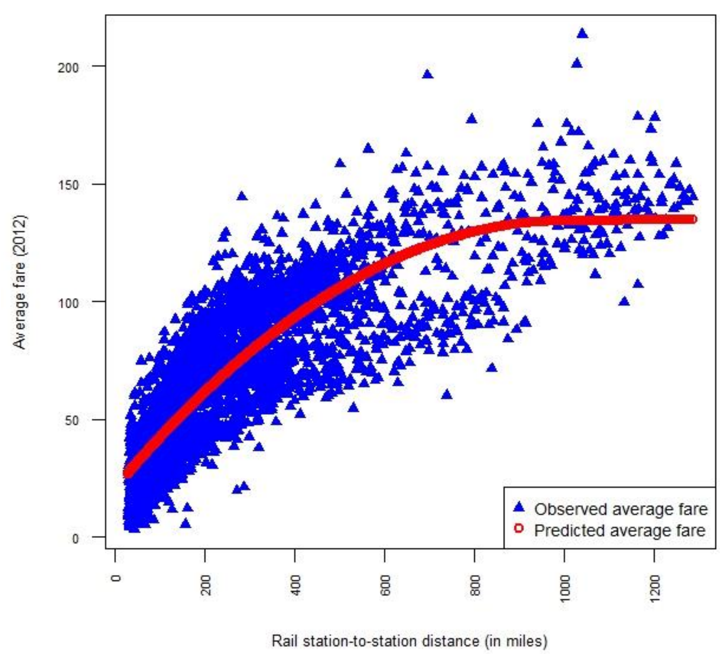


Figure 20: Relationship between average rail fare and trip mileage (Northwest first/business class)

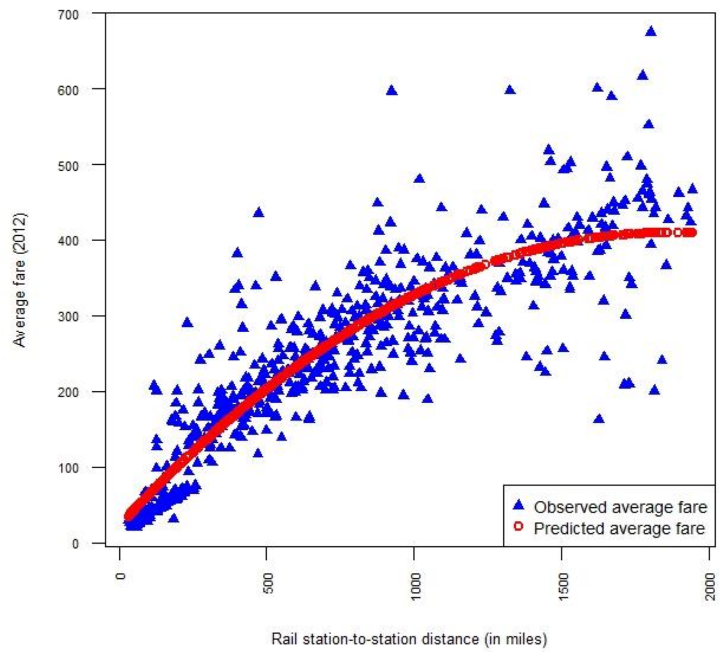


Figure 21: Relationship between average rail fare and trip mileage (Northwest economy class)

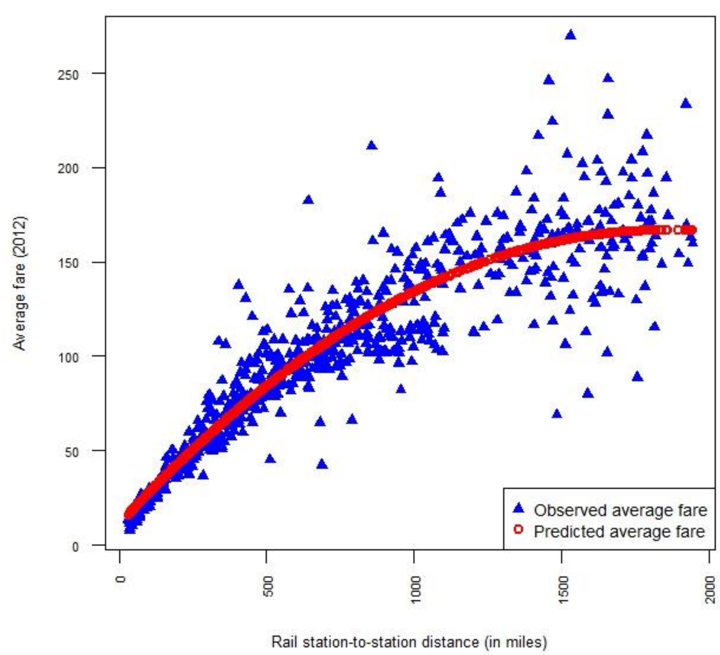


Figure 22: Relationship between average rail fare and trip mileage (South first/business class)

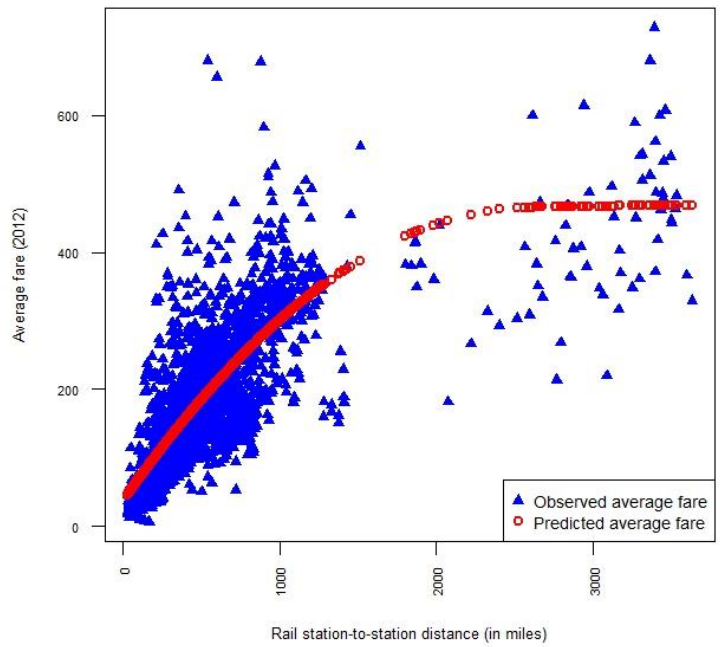


Figure 23: Relationship between average rail fare and trip mileage (South economy class)

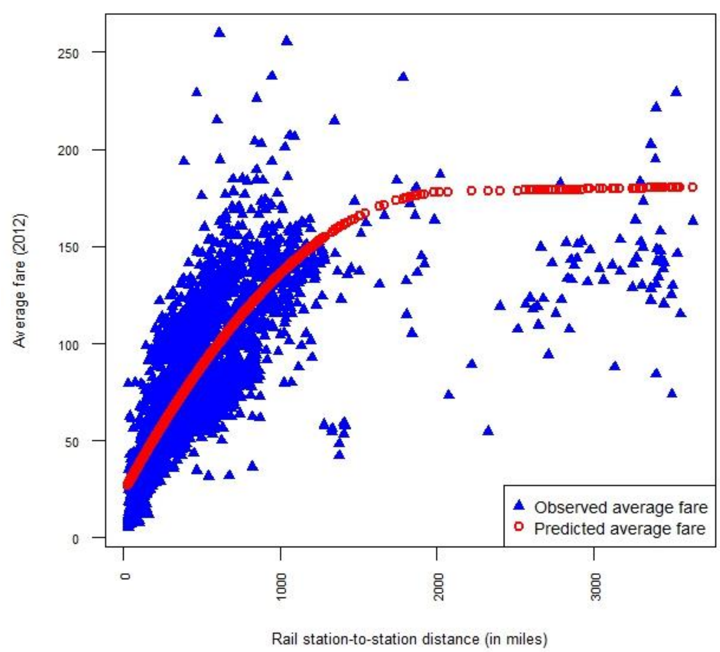


Figure 24: Relationship between average rail fare and trip mileage (West first/business class)

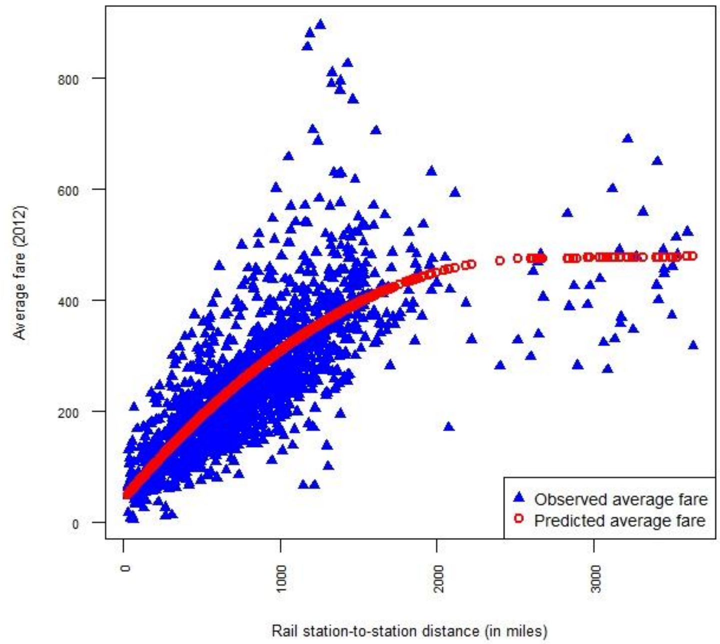
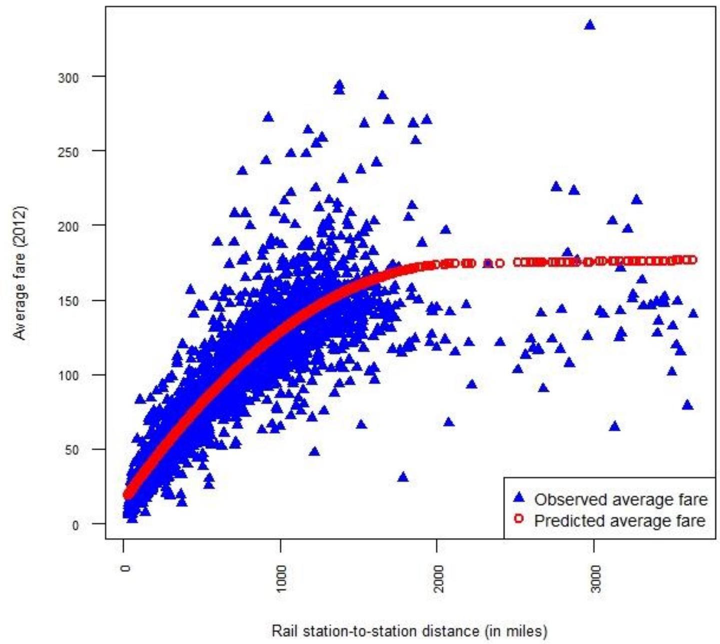


Figure 25: Relationship between average rail fare and trip mileage (West economy class)

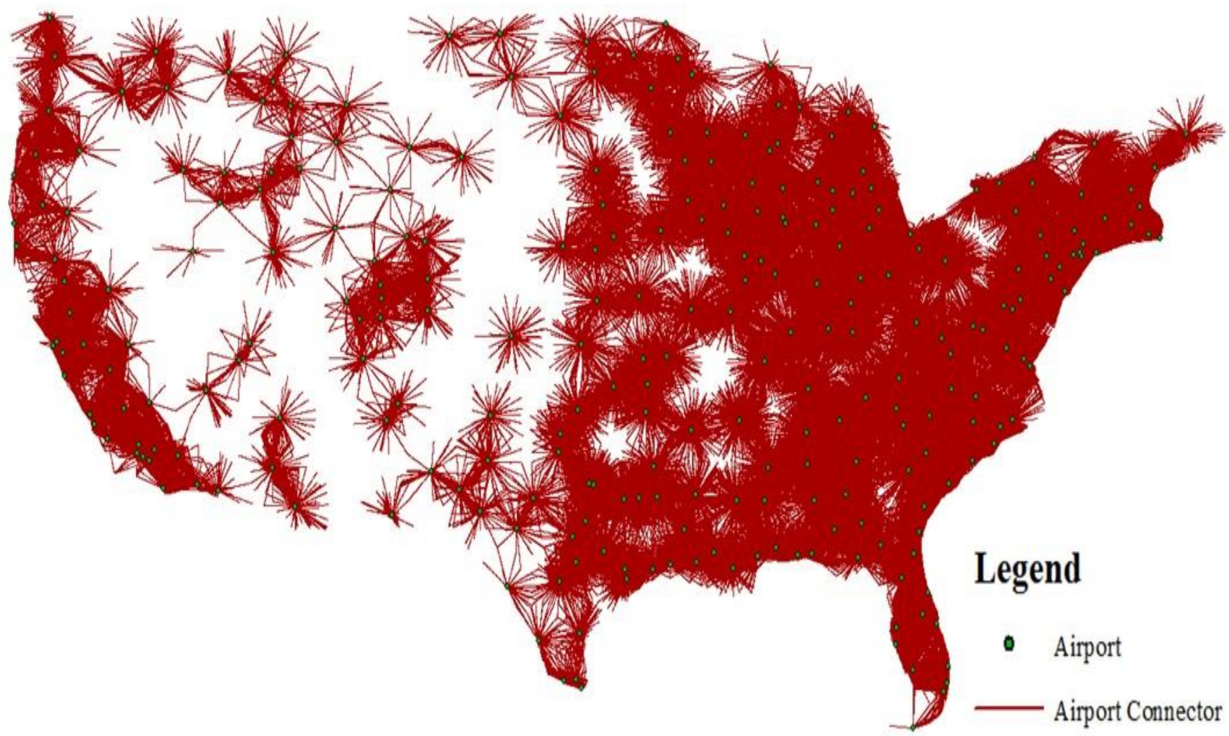


Air System

Airport Connectors

The same procedure used to generate rail station to Census Tract connectors was adopted to generate airport-to-Census Tract connectors. In the case of airports, the radius of the buffer region was set to 100 miles instead of 50 miles, as airports may draw travelers from a larger market area than rail stations. The spider network created from the generation of airport-to-Census Tract connectors is shown in Figure 26. As in the case of rail station connectors, a Census Tract was allowed to have a connector to all airports within 100 miles from the location of its centroid.

Figure 26: Airport-to-Census Tracts Connectors



Air Travel Time, Distance, and Cost

Air network characteristics for the year 2012 were obtained from two main databases provided by the Bureau of Transportation Statistics: the Airline On-Time Performance Data (on-time data hereafter) and the Airline Origin and Destination Survey (DB1B). The on-time data are published monthly and contain at least 1% domestic nonstop scheduled service flights information (i.e., air carrier, flight number, scheduled departure and arrival dates and times, actual departure and arrival times, canceled or diverted flights, taxi-out and taxi-in times, air time, and nonstop distance).[5] The DB1B is a 10% sample database of airline tickets from reporting carriers and includes the full itinerary information of domestic flights (i.e., air carrier, origin and destination airports, season, number of passengers, fare paid by each passenger, fare class, and distance). The DB1B data are published quarterly.[6] Using these two databases, an airport-to-airport origin-destination (O-D) matrix with the following air level-of-service and demand variables was derived:

- The number of flights serving a particular O-D pair over a period of one week (i.e., frequency per week).
- On-time performance (in percentage) across the flights serving a particular O-D pair over a period of one week when:
  - A flight is considered on-time if arrival delay <15 minutes; and
  - A flight is considered on-time if arrival delay <30 minutes.
- The average flight duration (not including transfers) in minutes.
- The average passenger-weighted fare, by season, for a particular O-D pair for:
  - Economy class; and
  - First/business class.
- The number of passengers, by season, for trips between the airports with:
  - No stop;
  - One stop (summarized by stop locations); and
  - Two or more stops (summarized by stop locations).
- The average coupon-mileage for trips with:
  - No stop;
  - One stop (summarized by stop locations); and
  - Two or more stops (summarized by stop locations).

The resulting files from the processing of the DB1B and on-time databases were further combined using a custom program to create the final airport-to-airport level-of-service data file to use in the models. The fields in this resulting file are shown in Table 4, with notes about how the variables are defined. These variables included: Average business class fare ($); average economy class fare ($)in the DB1B data for the O-D; average number of transfers; average total scheduled in-flight duration; average fraction of flights within 15 minutes of scheduled arrival; average fraction of flights within 30 minutes of scheduled arrival; number of direct flights per

---

[5] More information on on-time data and summary statistics is available at this website.

[6] More information on the DB1B database is available at this website.

week; frequency of one-stop flights per week (based on minimum of two flights); and frequency of two-stop flights per week (based on minimum of three flights).

Table 4: Airport-To-Airport Level-of-Service Variables, Based on DB1B and on-Time Databases

Field	Definition
OAIRPORT	3-letter code for origin airport
DAIRPORT	3-letter code for destination airport
BUSIPAX	Number of business class DB1B records
BUSIFARE	Average business class fare ($)
ECONPAX	Number of economy class DB1B records
ECONFARE	Average economy class fare ($)in the DB1B data for the O-D
NPAXVALID	Number of DB1B records with valid routes
AVGTRANSFERS	Average number of transfers
AVGDISTANCE	Average total route distance
AVGDURATION	Average total scheduled in-flight duration
AVGONTIME15	Average fraction of flights within 15 minutes of scheduled arrival
AVGONTIME30	Average fraction of flights within 30 minutes of scheduled arrival
NPAXDIRECT	Number of DB1B records with direct flight
FREQDIRECT	Number of direct flights per week
DISTDIRECT	Average distance of direct flights
DURADIRECT	Average flight duration of direct flights
OT15DIRECT	Average fraction of direct flights within 15 minutes of scheduled arrival
OT30DIRECT	Average fraction of direct flights within 30 minutes of scheduled arrival
NPAX1STOP	Number of DB1B records with one stop
FREQ1STOP	Frequency of one-stop flights per week (based on minimum of two flights)
DIST1STOP	Average total distance of 1-stop flights
DURA1STOP	Average total flight duration of 1-stop flights
OT151STOP	Fraction of 1-stop flights within 15 min of scheduled arrival (min of two flights)
OT301STOP	Fraction of 1-stop flights within 30 min of scheduled arrival (min of two flights)
NPAX2STOP	Number of DB1B records with two stops
FREQ2STOP	Frequency of 2-stop flights per week (based on minimum of three flights)
DIST2STOP	Average total distance of 2-stop flights
DURA2STOP	Average total flight duration of 2-stop flights
OT152STOP	Fraction of 2-stop flights within 15 min of scheduled arrival (min of three flights)
OT302STOP	Fraction of 2-stop flights within 30 min of scheduled arrival (min of three flights)

Notes on Table 4:

All fields are O-D specific, using only the 312 airports included in the on-time database.

All averages and fractions are passenger-weighted, where applicable, so that routes with more passengers using them weigh more heavily in the combined serviced levels.

*"Valid" routes are routes where that are either in the on-time data base, or where there are at least 10 DB1B records. Where no record is in the on-time database, the following default values are used: (a) frequency = 7 flights/week, (b) on-time percentage is the average of the overall on-time percentages of the departure airport and the arrival airport, and (c) the flight duration = 25.54 + 0.09 * distance + 1.509 * sq. rt. (distance); based on a regression equation estimated on valid records, where duration is in minutes and distance is in miles.*

For routes with two or more flights, the frequency is taken as the minimum scheduled frequency across the flights, and the on-time percentages are taken as the minimum on-time percentages across the flights.

Generating Zone-to-Zone Matrices for Rail and Air

Although the data were prepared so that, in the future, Census Tract-to-Census Tract level of service for air and rail could be used, the initial model application uses zones (NUMAs) as the basic level of spatial aggregation. This required using the station-to-station and airport-to-airport matrices along with the Census Tract-to-airport/station connectors to create zone-to-zone rail and air level-of-service matrices. This was done as follows:

- Within each zone, the Census Tract with the largest number of resident households was chosen as the representative origin tract within the zone, and the Census Tract with the largest total employment was chosen as the representative destination tract within the zone.

- Using estimates of value of time and relative travel time component weights from previous model estimations, generalized costs were calculated for all possible air routes via combinations of origin airports within 100 miles of the representative origin tract and destination airports within 100 miles of the destination tract. The tract-to-airport access and egress distances were also used in these calculations.

- The route via the least-generalized-cost airport pair was then selected as the representative air route for the zone pair.

The same procedure was used to select rail routes, using all combinations of rail stations within 50 miles of the O-D Census Tracts.

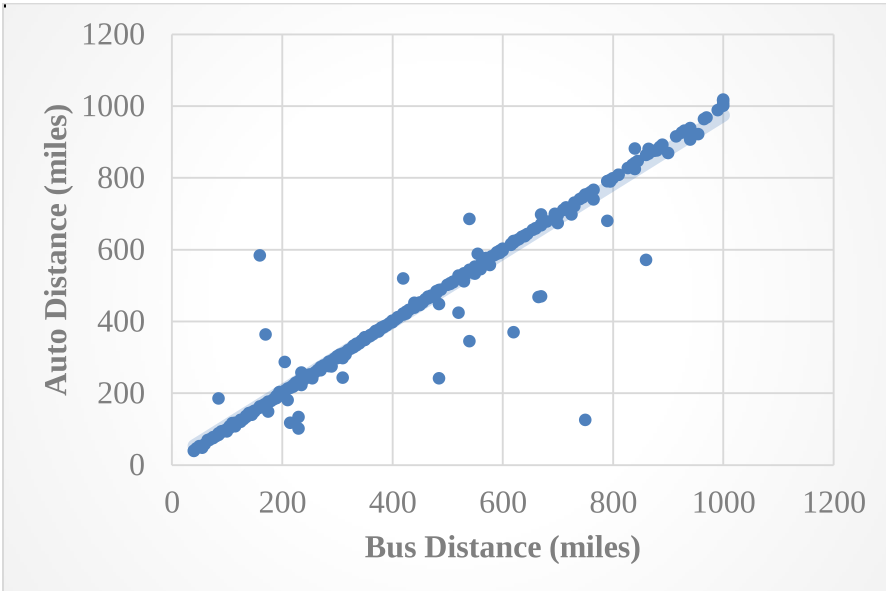
Bus System

Travel time, distance, and fare skims were generated in this study for the long-distance bus network of the United States. Because it is difficult to identify each and every bus route across the nation, efficient procedures were employed to arrive at the level-of-service measures for the bus network. Procedures followed for generating bus level-of-service measures are discussed in this section.

Bus Travel Time

The project team gathered a large amount of information on bus-service characteristics for several bus-service providers operating in a variety of markets across the country. The bus-service-attribute data collection effort corresponded to 447 unique city pairs. Information regarding distance and travel time by bus was available for each of the city pairs. The 447 city pairs were then geocoded in ArcGIS to obtain their spatial coordinates. A Python code was written to obtain the auto distances and travel times between these city pairs using Google's distance matrix application programming interface (API). Auto distances generated from Google's API and the corresponding bus distances that were collected manually were compared to ensure consistency between bus and auto distance data. The comparison is shown in Figure 27.

Figure 27: Comparison between Auto and Bus Distances



The data points in Figure 27 are heavily concentrated along the 45-degree line, which implies that the auto and bus distances between the city pairs in consideration are largely consistent with one another.

Next, a comparison was made between auto and bus travel times for different distance ranges and an auto-to-bus travel time conversion factor table was generated. This information is presented in Table 5.

Table 5: Auto-to-Bus Travel Time Conversion Factors

Distance (miles)	Factor
> 0 - 120	1.27
> 120 - 300	1.43
> 300 - 600	1.50
> 600	1.61

Using the information from Table 5, auto travel times (discussed in the road system section) were converted to bus travel times. (For example, if the distance between an O-D pair is 60 miles and the auto travel time is 60 minutes, the corresponding bus travel time for the O-D pair was found to be 60×1.27 = 76 minutes.) The difference between auto and bus travel times accounts for wait, transfer, and stop times that encumber bus travel.

Bus Fares

Bus fare was calculated through the estimation of a statistical regression model that related bus fare to various trip attributes. The fare-collection effort focused primarily on the popular bus routes (and carriers), while also ensuring that there is sufficient sample size for model estimation in different distance bands. A total of 1,000 data points were collected for 447 unique city pairs. The following information was collected from the carrier's website for each city pair, including:

- Travel Distance (miles);
- Number of transfers;
- Number of stops;
- Travel time (minutes);
- One-way fare ($);
- Frequency;
- Transfer point; and
- Interstate/intrastate travel.

The frequency information was missing for approximately one-third of the data collected. The missing frequency information was imputed using a cell mean-imputation approach. Several specifications were tested with a host of variables included to predict the fare between an O-D pair. However, only travel time was used in the final bus fare regression equation model owing to data limitations for other attributes in forecasting mode. A bus fare regression model with travel time as the independent variable was estimated and validated:

Equation 2: Bus Fare Regression Model

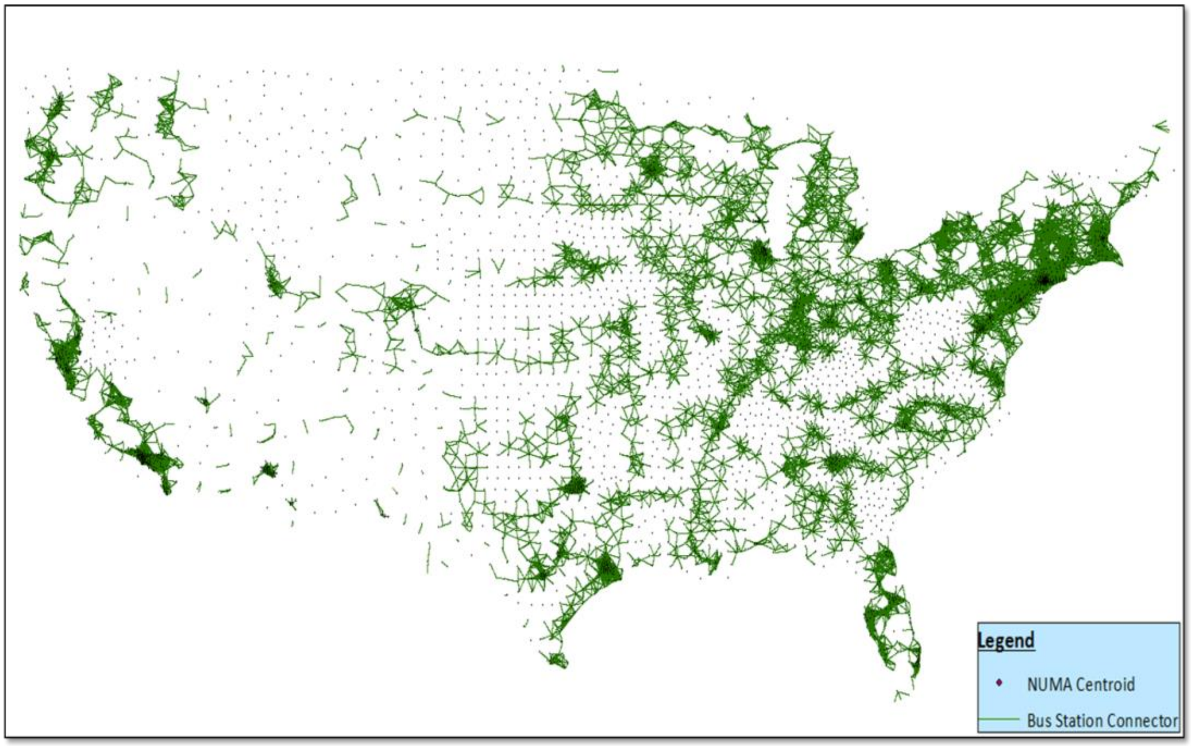
$$Fare = 9.65 + 0.107 * travel\ time$$

This model was used to generate a bus fare matrix from the bus-travel-time matrix.

Bus Feasibility Matrix

It was necessary to determine if bus is a feasible mode-choice option when considering long-distance travel between locations. To determine if bus travel was feasible or not for a given NUMA pair, a buffer-region approach, similar to that discussed in the airport-to-Census Tract connectors section, was adopted. Bus stop location information for the United States was obtained from bus GTFS data. The bus stop location (point) data was overlaid on the NUMA-level polygon file. A 40-mile buffer region was drawn from each NUMA centroid. The total number of bus stops that fall within the 40-mile buffer region of each NUMA was determined and stored. A binary (0/1) indicator was generated for each NUMA, where the NUMA would receive a "1" if there is at least one bus stop within the 40-mile buffer from the NUMA's centroid location. Otherwise, the NUMA received a "0." The NUMA-level information was converted to a feasibility matrix by multiplying the bus feasibility indicators for each NUMA-O-D pair. If both the origin and destination NUMAs had a value of "1" in their bus feasibility indicator column, bus travel was considered feasible between the NUMA pair under consideration. Otherwise, bus travel was considered infeasible between the NUMA pair. The spider network created from the generation of NUMA centroid-to-bus station connectors is shown in Figure 28. A NUMA is allowed to have a connector to all bus stations within 40 miles from the location of its centroid.

Figure 28: NUMA Centroid-to-Bus Station Connectors



Bus Travel Time and Fares

To obtain the bus-travel-time matrix, the auto-travel-time matrix was first generated at the NUMA level by minimizing travel time between each O-D pair using built-in skimming procedures in TransCAD. This process resulted in a complete 4570×4570 matrix of auto travel times. A corresponding distance matrix was automatically generated by TransCAD. The auto travel times between different O-D pairs were converted to bus travel times using the factors presented in Table 5 for different distance ranges. The bus-travel-time matrix was multiplied (cell-by-cell multiplication) by the feasibility matrix to obtain the final bus-travel-time matrix for O-D pairs (between which bus travel is deemed feasible). The bus-travel-time matrix obtained as a result of this exercise was used to compute a fare matrix (see Figure 27).

3.3 Socioeconomic Data

Person and Household Characteristics

Person and household characteristics were derived from the Public Use Microdata Sample (PUMS) of the 2010 Decennial Census and the 2007–2011 American Community Survey (ACS)

5-year estimates. These are used primarily as input to the synthetic-population process, described in Chapter 4.

The personal characteristics selected from the Census data include:

- Age of the person;
- Gender of the person;
- Race of the person; and
- Employment status of the person.

The household characteristics selected from the Census data include:

- Presence of children in the household;
- Household income level;
- Householder age;
- Household size;
- Type of household;
- Number of nonworkers in the household;
- Number of full-time workers in the household;
- Number of part-time workers in the household;
- Number of students in the household;
- Number of vehicles in the household; and
- Group quarter identifier.

The householder refers to the person (or one of the people) in whose name the housing unit is owned or rented (maintained) or, if there is no such person, any adult member, excluding roomers, boarders, or paid employees. If the house is owned or rented jointly by a married couple, the householder may be either a husband or a wife. The person designated as the householder is the "reference person" to whom the relationship of all other household members, if any, is recorded.

The household type is a function of whether members are related to the householder by birth, marriage, or adoption and whether the household is headed by a single householder (male or female) or a married couple. A nonfamily household consists of a householder living alone (a one-person household) or where the householder shares the home exclusively with people to whom he/she is not related.

As of 1983, group quarters were defined in the Current Population Survey as noninstitutional living arrangements for groups not living in conventional housing units or groups living in housing units containing ten or more unrelated people or nine or more people unrelated to the person in charge. Examples of people in group quarters include a person residing in staff quarters at a hospital, a halfway house, military housing, college dormitories, or retirement housing. Since 1972, inmates of institutions have not been included in the Current Population Survey.

Employment Data

Employment data was compiled from two sources:

- Longitudinal Employer Household Dynamics (LEHD).
- Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW).

A brief description of the steps that were undertaken to generate employment database for the current project is provided below.

Longitudinal Employer Household Dynamics (LEHD)

2010 LEHD Origin-Destination Employment Statistics (LODES) database was the primary source of employment data. Categories used in developing these data are presented in Table 6. The database contains private and public job numbers for all states and the District of Columbia (the only exception is Massachusetts, which has yet to join the LEHD program). For the private sector, employment numbers were summarized by 20 different industries. In addition, tourism and recreation-related industries, such as arts/entertainment/recreation, accommodations, and food services, were further divided into a number of subcategories (also shown in Table 6 through Table 8).

Table 6: National Employment Categories—NAICS Employment Categories

NAICS Employment Categories
(1) agriculture, forestry, fishing, and hunting
(2) mining quarrying, and oil and gas extraction
(3) utilities
(4) construction
(5) manufacturing
(6) wholesale trade
(7) retail trade
(8) transportation and warehousing
(9) information,
(10) finance and insurance
(11) real estate and rental and leasing
(12) professional scientific, and technical services
(13) management of companies and enterprises
(14) administrative and support and waste management and remediation services
(15) educational services

NAICS Employment Categories
(16) health care and social assistance
(17) arts, entertainment, and recreation
(18) accommodation and food services
(19) other services [except public administration]
(20) public administration

Table 7: National Employment Categories— Subcategories of Tourism and Recreation Employment

Subcategories of Tourism and Recreation Employment
(1) performing arts companies
(2) spectator sports
(3) promoters of performing arts, sports, and similar events
(4) agents and managers for artists, athletes, entertainers, and other public figures
(5) independent artists, writers, and performers
(6) museums, historical sites, and similar institutions
(7) amusement parks and arcades
(8) gambling industries
(9) other amusement and recreation industries

Table 8: National Employment Categories— Subcategories of Accommodation and Food Service Employment

Subcategories of Accommodation and Food Service Employment
(1) traveler accommodation
(2) RV (recreational vehicle) parks and recreational camps
(3) rooming and boarding houses
(4) full-service restaurants
(5) limited-service eating places
(6) special food services
(7) drinking places (alcoholic beverages)

This step was undertaken to create proxies for attraction variables (e.g., number of rooms/beds in hotel/motel/resort, number of employment in theme parks, etc.), which were not readily available. The LODES database, which includes data at block-level, provides job numbers by main industry only. To create a database that includes employment in tourism and recreation-related industries, broken down by subcategories, the LEHD Quarterly Workforce Indicators

(QWI) data, which is available at a spatial resolution larger than Census block, was employed. Finally, the private-sector data were aggregated at the appropriate level to produce a Census Tract-level file. For the private sector, the job numbers were available for the Federal, state, and local government. Here, the data-processing step involved aggregation of block-level public sector employment data to the Census Tract level.

Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW)

Census Tract-level QCEW data for the year 2010, published by the BLS, was used to generate an employment database for the Commonwealth of Massachusetts (MA). Since the QCEW is an essential input to the LEHD program, the assumption was that, though the employment dataset was compiled using multiple databases/sources, the final dataset contains consistent records.

3.4 Land-Use Data

A number of national-scale data sources that provide data free of charge were used to assemble a land-use file. To be consistent with the spatial unit applied to summarize the level-of-service (LOS) data, a Census Tract-level land-use file was compiled to facilitate both model estimation and application tasks. The land-use data and corresponding sources are listed below.

- **2010 Census**. National-level geographic files (i.e., shape files) include all the tracts in the Unites States that are available from the U.S. Census Bureau. For the current project, 2010 Census Tract-level geographic files with demographic profile information were downloaded to obtain the following land-use data:
  - Total land area; and
  - Number of permanent households and noninstitutionalized group quarters.
- **US National Park Service, 2012, TomTom data, and ESRI**. A group of layers containing the national, state, and regional parks were available in the ArcGIS software. The layers were created using data from several sources, including the National Park Service (NPS), TomTom, and ESRI. Information available from these layers included park/forest name, type (e.g., national park, state park, regional park, national forest, etc.), location, and size. In total, information on 3,355 parks/forests was used to create a database that provides total park/forest area, by Census Tract.
- **National Center for Education**. Information available from the National Center for Education was used to create an initial database of colleges and universities that offer a bachelor’s degree or higher. Variables included in this database were institution name, location (latitude and longitude), and total enrollment in 2011. Once this initial database was created, community colleges, vocational colleges, and online colleges were dropped from the list since these colleges are likely to attract fewer long-distance trips. Next, the data were aggregated to create a file that provides total college/university enrollment, by Census Tract.

The research team recognizes that, in addition to the land-use variables mentioned previously, other variables (e.g., parking costs) are likely to improve predictive capability of the model.

However, such variables are only available from a selected number of state/regional models and it was not feasible to create a national-level dataset for this project.

3.5 Household Surveys

Several datasets were identified both during and since the review of experience. The following surveys are discussed in more detail:

- 1995 American Travel Survey (ATS).
- 2001 National Household Travel Survey (NHTS).
- 2012 California Household Travel Survey (CHTS)
- 2010 Colorado Front Range Travel Survey
- 2003 Ohio Household Travel Survey

The datasets are described with reference to the model components that they might support development of, and discussion of their known limitations.

1995 American Travel Survey (ATS)[7]

The "standard" dataset for modeling long-distance travel in the United States has long been the 1995 ATS. The Census Bureau of Transportation Statistics (BTS) carried out the ATS periodically up until 1995, but has not carried it out since, which is the main reason such a dated source of data is still in use. The attractive features of this dataset can be seen in Table 10:

- It is a large dataset, with over one-half million long-distance trips (75 miles or more), reported by almost 70,000 households, randomly selected from all across the United States.
- It contains one full year's worth of trips for each household.
- In contrast to the other surveys listed in Table 10, this survey was not entirely retrospective. Respondents were contacted before the year-long reporting period began, and were sent a calendar/diary to record key details of every long-distance trip made by every household member. They were then contacted every three months to relay important information about the trips they had reported.
- Also, computer-assisted personal interviews (CAPIs) were performed with respondents who could not participate by telephone, reducing one potential source of nonresponse bias.

[7] The 1995 American Travel Survey (ATS)

Besides the fact that these data are 20 years old, there are several factors that might limit their usefulness for some types of modeling:

- There is no geocode information available for the trips, so it would not be possible to attach detailed mode-impedance information.
- There is a high respondent burden associated with a 12-month survey with repeated interviews. Even though the reported response rate is high (85%), there may have been some amount of "soft refusal," with respondents simply declining to report any more trips after reaching a certain level of "fatigue."

2001 National Household Travel Survey (NHTS)[8]

In 2001, instead of repeating the ATS for long-distance travel, a decision was made to combine the ATS with the periodic NHTS, which is a more typical travel diary survey of all trips made during one 24-hour period. A subset of NHTS households were given a separate log on which to retrospectively record all trips of 50 miles or more they had made during the four weeks before their survey travel day, and then report those trips during the same telephone call as they reported all trips made on their selected travel day. (This is essentially the same survey methodology that was also used for all of the other surveys described here.)

In retrospect, it may have been a questionable decision to combine the long-distance travel into the NHTS, as the resulting 46,000 long-distance trips comprise less than 9% of the number of trips obtained in the 1995 ATS. As a result, the 2001 NHTS long-distance data have not been used extensively for modeling or analysis, and the long-distance component was dropped from the 2009 NHTS survey altogether.

The NHTS data lack the main attractive features of the ATS data (large sample size and nonretrospective methodology), but these data share some of the key weaknesses of the ATS (i.e., older data and lack of geocodes and detailed spatial information).

2012 California Household Travel Survey (CHTS)

The California Department of Transportation (CalTrans) performed a major survey effort for the entire state in 2012. The design of the survey is similar to that used in the Colorado survey, but with the retrospective period extended from 2 weeks to 8 weeks. The rationale for extending the retrospective period was that it would provide more trips for modeling, and that even if the full period may not be useful for modeling trip frequency/generation (due to increasing recall nonresponse bias), if a respondent does remember the trip, their recall of the details of that trip (e.g., mode, destination, etc.) is likely to be good enough to use for modeling those other aspects of behavior.

[8] The 2001 National Household Travel Survey (NHTS)

With a sample size of over 40,000 households, plus the 8-week period, this survey yielded a large sample of trips to use in modeling. Also, full geocoding and spatial information will be available (for approved researchers who sign a confidentiality agreement).

2010 Colorado Front Range Travel Survey[9]

There are few regional planning agencies (MPOs) that have included a special long-distance travel component as part of their household travel survey. A recent example, however, is the 2010 survey carried out by a group of Colorado MPOs, with the Denver Regional Council of Governments (DRCOG) taking the lead. An attractive feature of this survey is that it is quite recent, and also that detailed geocoding information is likely to be available for all trips (for researchers who sign a confidentiality agreement). Also, the retrospective period of two weeks seems short enough to allow fairly accurate respondent recall. The short recall period, however, along with a limited sample size of just over 3,000 households, resulted in just over 6,100 long-distance trips.

2003 Ohio Household Travel Survey

Similar to Colorado, the Ohio Department of Transportation conducted a long-distance passenger travel survey as part of a larger household travel survey effort. There were 2,094 households who made 13,807 long-distance trips. This survey is biased for total demand, since the survey contained only households that made at least one long-distance trip over the 2-week assigned travel period. These data were collected only in the spring and fall seasons, and so no data were collected during the winter and summer or major holidays. There were no commute trips collected in this survey.

Content of the Long-Distance Household Travel Survey Datasets

All of the aforementioned United States datasets (summarized in Table 9) are similar in terms of the data items that they contain. The fact that they were (mainly) retrospective surveys, and were time-constrained "add-ons" to standard household travel surveys, has tended to limit the amount of detail that could be collected regarding each long-distance trip. The common data items include:

- **Main trip purpose**. This is the most important variable for model segmentation.
- **Journey leg**. Whether the trip is leaving home, returning home, or is part of a multidestination tour.
- **Trip origin and destination addresses**. This information is necessary to connect land-use information and travel network information for modeling mode choice and destination choice. The national-level datasets (ATS and NHTS) were collected primarily for descriptive analyses and not for modeling, and also are subject to strict privacy regulations, so detailed geocodes are not available. The California and Colorado surveys,

[9] The 2010 Colorado Front Range Travel Survey.

on the other hand, were designed to provide data for modeling, and used the most modern geocoding methods ("real time" online geocoding, using Google maps technology).

- **Travel group size (and composition)**. This is another key segmentation or explanatory variable.
- **Date (or day of week) of travel, and trip departure time of day**. These can also be important segmentation or explanatory variables (e.g., separating weekend from weekday travel).
- **Duration of stay at the destination**. Along with trip purpose, this information helps to define specific types of journeys for segmentation.
- **Main travel mode and access/egress modes and locations**. This is necessary information for modeling mode choice. For air, rail and bus trips, the extra information collected typically includes the boarding and alighting airport or station, and the modes used to/from those locations.

The surveys are also similar in terms of the types of information they do not include, such as:

- **Information on the trip planning process**. This may include how information was gathered, how reservations were made, how far in advance planning was done, etc. Data regarding "packaging" of travel, lodging, and activities may be especially useful.
- **More-detailed information on trip purpose(s)**. In addition to classifying the main purpose, it may be useful to have data on the range of different activities carried out on the trip.
- **How often the destination had been visited in the past**. There can be differences in decision-making for first-time versus repeat trips, and for frequent versus infrequent trips.
- **Type of lodging used**. This is an important consideration in terms of the cost of the trip, and can also influence mode choice.
- **The specific route (and operator) used**. This may be useful data for auto trips and air and bus trips.
- **Fares actually paid and subsidies received**. This may be useful for understanding air travel decisions, where different travelers can pay different prices, and many receive reimbursement.
- **Class of travel used**. This is important for air travel, and possibly for rail.
- **Type of auto used**. In terms of size class/body type, or else make/model. This has implications in terms of travel cost, comfort level, and accessibility to recreational areas.

Typically, these types of additional questions are only included in special-purpose surveys for market research purposes, and such data are often proprietary. While they could provide useful data for exploratory modeling, they are not strictly necessary for modeling long-distance travel, and may even be problematic to use in the context of longer-term predictive models, since future-year assumptions or predictions would need to be made for these factors.

Table 9: Summary of Long-Distance Travel Survey Characteristics

National	ATS	NHTS		California	Colorado	Ohio
		NY	WI			
Year	1995	2001	2001	2012-2013	2010	2001-2003
One-way trip length	75+ miles[10]	50+ miles	50+ miles	50+ miles	50+ miles	50+ miles
Timeframe for Data Collection - retrospective	1 year	4 weeks	4 weeks	8 weeks	2 weeks	2 weeks[11]
Number of HHs reported LD trips	48,527 HH	7,032 HH	11,027 HH	≅15,500 HH	3,000 HH	2,094 HH
Number of LD trips/tours reported	556,026 tours	28,021 tours	44,011 tours	≅58,500 trips	≅6,100 trips	≅13,807 trips
During the data collection process, how was it determined that the reported trips are LD trips?	Self-reported	Self-reported	Self-reported	Self-reported	Self-reported	Self-reported
Trip origin used to define LD trips	Home	Home	Home	Home	Home	Home
Trip destination used to define LD trips	Farthest destination	Farthest destination	Farthest destination	Any destination	Any destination	Any destination

[10] ATS data are available for 100+ miles trips only.

[11] The Ohio survey also included a 4-week prospective survey of nonhousehold travel survey households that were screened for a probability of making a long-distance trip.

Data Preparation

The 1995 ATS collected long-distance travel information from 80,000 households in the United States over the course of one year. The ATS survey gathered information on all tours to destinations 100 miles or more away from a respondent's home. For each home-to-home tour, households were asked to identify the main purpose of the tour, accompaniment type, party size, mode, and destination.

To generate the sample for analysis, the person-level tour information was first aggregated into household-level data. Several consistency checks were then performed and those households with missing or inconsistent information were deleted from the estimation sample. As a result, the final estimation sample included 47,931 households. To estimate the nonbusiness and business model structures, only those households that undertook at least one nonbusiness or business tour during the year were selected. Second, the tours that had a destination outside the United States (i.e., international tours) were eliminated from the analysis. The final nonbusiness and business samples included 40,794 and 14,664 households, respectively.

The remaining four household travel surveys were processed to allow merging of these surveys. The major data-processing steps included the following:

- Forming tours from trip-level data. This step was only applicable for the 2003 Ohio and the 2012 California surveys, and involved identifying the primary destination of the tour. To be consistent with other datasets that used tour as the unit of travel-related information, the farthest destination from home was used to identify the tour destination.
- Identifying the tour purpose and, where necessary, recoding it as:
  - Commute;
  - Business;
  - Visiting friends and relatives;
  - Leisure; or
  - Personal business.
- Identifying the tour mode and, where necessary, grouping it as:
  - Auto;
  - Bus;
  - Rail; and
  - Air.
- Appending O-D Census Tract and NUMA-zone identifications (IDs) to each tour record. These IDs were used to append appropriate network skims, land-use, and employment data.

While collected information was not uniform across all household survey datasets, the data were processed in such a way that the following variables were common across all estimation datasets:

- **Household Characteristics**: Household size, number of driving-age adults, number of workers, age of head of household, number of vehicles, income, residence location (longitude/latitude, Census Tract ID, NUMA-zone ID, county and state Federal Information Processing Standard [FIPS] codes), the date on which trip reporting period ended, and survey year.
- **Person Characteristics**: Age, gender, worker status, and student status.
- **Travel Characteristics**: Number of trips in the tour, the date on which the tour began, number of nights away from home, total travel party size, number of household members traveling together, tour origin (always home), tour origin and destination locations (longitude/latitude, Census Tract ID, NUMA-zone ID, county and state FIPS codes), primary tour purpose, outbound and inbound tour modes, and outbound and inbound access modes.

Data Used for Model Estimation

For model estimation purposes, each dataset was examined in detail and a number of descriptive statistics were generated.[12] It was clear from these analyses that the data from the sources listed previously varied in terms of:

- Trip length employed to identify long-distance travel (e.g., 50+ miles vs. 100+ miles);
- Geographic coverage of the study area (e.g., national, state, or regional);
- Duration of tour reporting period (e.g., one year, three months, four weeks, etc.);
- Data collection schedule (e.g., all through the year or only a few months in one year);
- Spatial resolution of tour origin/destination; and
- The type and the level of details of travel-related information collected.

As a result, some datasets were more suitable for estimation of a particular type of model(s) than others. Table 10 lists the datasets used to estimate different long-distance model components. As summarized in Table 10, combination household time budget/ annual tour-generation/tour-scheduling/tour-duration/tour-participation models were estimated using a multiple discrete-continuous extreme value (MDCEV) structure on the 1995 ATS data. This is because the ideal model framework is designed to predict a household's long-distance travel schedule for a period of one year as a single scheduling process, and ATS is the only dataset that had a tour reporting period of one full year. However, the age of the dataset and its coarse spatial resolution made ATS less suitable for estimating other models, particularly mode and destination choice. All other models—including simpler logit models of tour generation, scheduling, and duration—were estimated using a combined dataset from the California, New York, Ohio, and Wisconsin

[12] To conserve space, the descriptive statistics are not included in this report, but are available from the research team upon request.

surveys (with the exception that only the California data were used for the logit tour-generation and tour-scheduling models).

Table 10: Datasets Used to Estimate Long-Distance Travel Model Components

Household Travel Survey	Time Budget	Tour Generation	Tour Scheduling	Tour Duration	Travel Party Size	Destination and Mode Choice
1995 ATS	√	√	√	√	√	
2001 NHTS (NY)				√	√	√
2001 NHTS (WI)				√	√	√
2012 CHTS		√	√	√	√	√
2003 Ohio				√	√	√
2010 Colorado Front Range		√		√	√	√

CHAPTER 4. INTEGRATED MODELING SYSTEM

4.1 Overview

This overview (Figure 29) outlines the aspects of the Long-Distance Passenger Travel Demand Modeling Framework that will be important to frame the individual modeling components. The core model components, and those necessary to support the core model components in the model design, are the focus of this overview, but there are a few additional elements of the model design that are identified in the model design for comprehensiveness, but are reserved for future model development efforts due to data or resource limitations.

The Long-Distance Passenger Travel Demand Modeling Framework considered four aspects of modeling travel demand:

- **Macroeconomic and land-use models.** These produce socioeconomic forecasts of the population and economy as input to the modeling system. These models can be quite complex, were not the main focus of this effort, and were not intrinsic to the core system components. Data on socioeconomic forecasts were available for current- and future-year conditions and were used as inputs without developing integrated modeling components for these elements.
- **Population and long-term mobility models.** These produce synthesized populations with personal and household characteristics, and long-term choices such as work and school locations and vehicle availability for the population in the United States. These models were also not directly the focus of this study, but were more directly and necessarily integrated with the travel demand model components.
- **Long-distance travel demand models.** These produce schedules, destinations, and modes for long-distance travel across the United States. These comprise the core model components. Route choice is the one element within this category of models that was problematic to test, given the limited data that were available. Route choice is also the model component of least importance for policy and scenario testing and was therefore not included in the final framework.
- **Assignment models.** These produce volumes on highway, rail, bus, and air systems for the evaluation of performance on travel, environmental, economic, and other measures. These models are essential to the eventual use of the long-distance travel models and are included as essential elements of the framework. Existing technologies for aggregate, static assignments will be employed to implement assignment models. Disaggregate, dynamic-assignment methods are not practical or necessary to deploy at a national scale for long-distance travel. This is also the element that brings together commercial and short-distance passenger travel demand from other sources to produce comprehensive assessments of all travel in the United States.

The long-distance passenger travel demand forecasting model framework is designed to cover the contiguous United States, with external zones for Mexico, Canada, and other countries. Alaska and Hawaii were also incorporated.

One core recommendation for the framework is the use of a one-year timeframe for the passenger travel demand models. This recommendation deviates from available data sources and existing long-distance travel model practice. That said, this is a critical recommendation to adequately capture the behavioral elements of long-distance travel, which are often scheduled over a one-year timeframe and have typical seasonal fluctuations. Integration of this travel with other models (such as statewide or multistate corridor models) may require conversion to an average weekday, but the richness of the annual data can also provide input for seasonal or peak travel periods. This is described more fully in Chapter 7 on long-distance travel survey data.

4.2 Framework

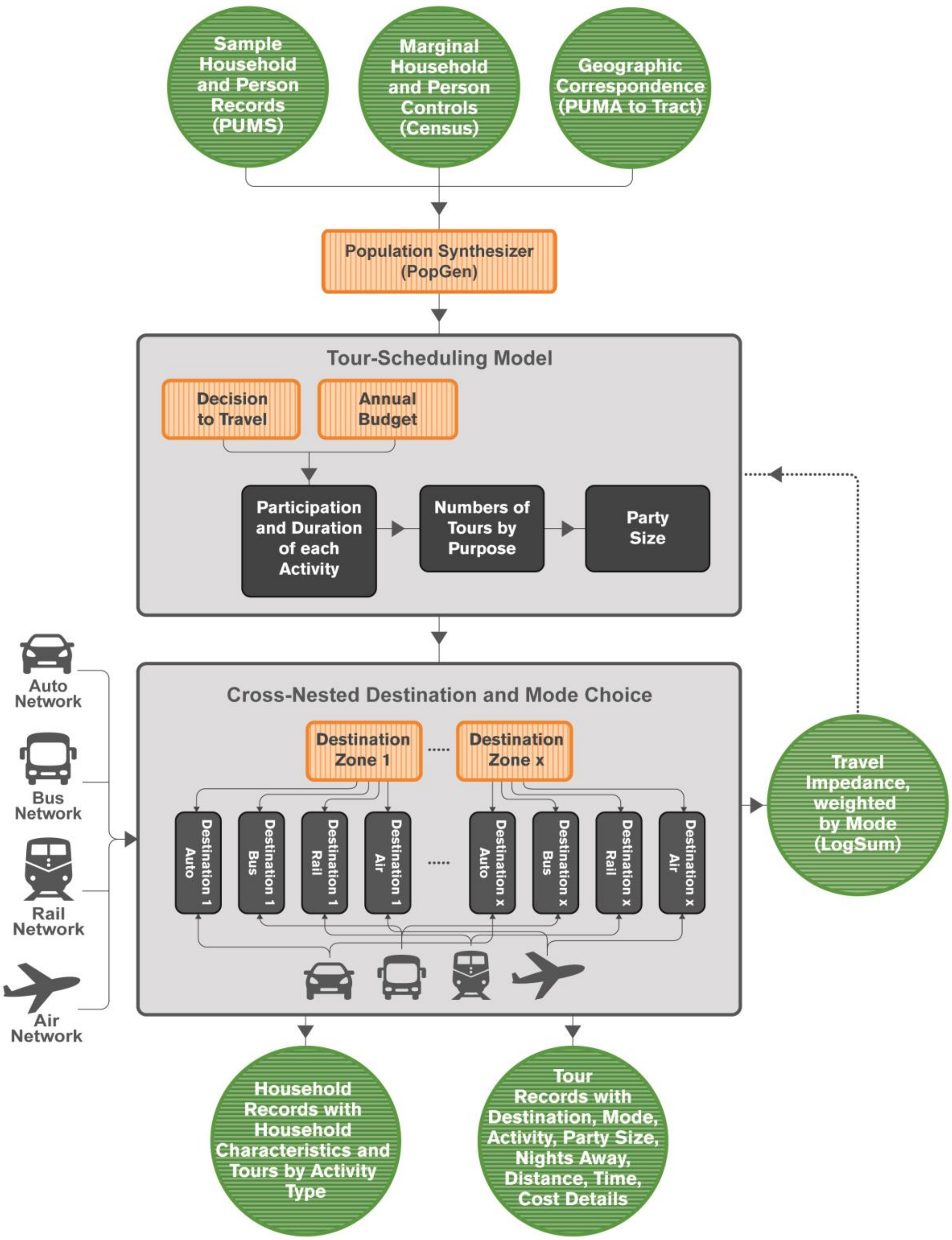
The integrated modeling system framework presented here reflects the long-term goal to develop a long-distance passenger travel demand model that achieves the objectives of the project. The shorter-term goal, to develop a long-distance travel demand model that is practical for current use following calibration and validation, is described in Chapter 5 of this report. Long-distance travel is defined as a tour—or round trip—whose two-way (home-to-home) length exceeds 100 miles.

The long-distance passenger travel demand modeling system is presented in Figure 29. There are three main elements to this system:

- **Population Synthesis** to synthesize persons and households across the United States with demographic and socioeconomic characteristics. The sample records include a comprehensive set of household characteristics for each PUMA in the PUMS. The marginal household and person control files include relevant household and person characteristics for each Census Tract.
- **Tour Generation, Scheduling, Duration, and Party Size** to estimate how many long-distance tours each household makes in each month of the year, based on sociodemographic characteristics and annual household budgets. This element also includes estimating the duration, purpose, and party size for each long-distance tour.
- **Destination and Mode Choice** to estimate the destination and mode choice for each long-distance tour for each household for the year, based on characteristics of each mode (i.e., auto, bus, rail, and air) and each destination. The four modal networks were analyzed to produce estimates of travel time, cost, and distance for each mode as input. These joint models produced a combined utility across all modes and destinations (referred to as LogSum) that can be used as accessibility measures for travel from specific residence zones. These measures provided feedback of long-distance accessibility to the generation and scheduling of tours, as represented by the dotted line in Figure 29.

The elements of the framework are described in more detail in the following sections.

Figure 29: Long-Distance Passenger Travel Demand Modeling Framework



4.3 National Synthetic-Population Generation

The generation of a national synthetic population is essential for modeling long-distance travel demand at the level of the individual traveler. In this study, a national synthetic population was generated using the procedures embedded in the PopGen software package (Ye et al., 2009). The PopGen system is a robust synthetic-population-generation software that is capable of controlling for both household- and person-level attributes in the synthetic-population generation process. Although the software is computationally efficient and capable of running in parallel (i.e., utilizing multiple cores in a computer) the process can be quite computationally burdensome and time consuming when attempting to synthesize a population for the entire nation. For this reason, the parameters and levels of spatial disaggregation adopted in the synthetic-population generation process were established to balance the desire for a synthetic population generated based on controls at a fine geographical resolution and the desire for rapid computational time.

Methodological Procedure

The methodological procedure embedded in the PopGen software allows the generation of a synthetic population using a variety of control variables at both the household and person levels. The key input datasets for the population synthesis process are as follows:

- A sample file that includes disaggregate household and person records for a sample of the population of interest. This sample file serves two key purposes: it provides the joint distribution among attributes of interest and households included in the synthetic population that are drawn from the sample.
- A marginal file that includes aggregate household- and person-level control totals for the geographic region of interest at the desired level of geographic resolution. This file provides the control totals that must be matched in the synthesis process. The sample file is expanded in such a way that the expanded sample mirrors the marginal control totals.
- A geographic-correspondence file that maps individual geographies (e.g., block groups or Census Tracts) to larger geographic areas (e.g., the PUMA). This file is important because the sample file (often derived from the PUMS data of the Census Bureau) is geocoded only to the PUMA level; thus, the joint distribution of attributes of interest for a specific PUMA is applied to all Census Tracts or block groups that belong to that particular PUMA in the geographic-correspondence file.

PopGen follows a three-step process in the synthesis of a population. First, the joint distribution of the attributes of interest is determined for each geography. The marginal control totals from the Census files are used to expand this joint-distribution matrix so that the marginal control totals are matched exactly. This procedure, known as iterative proportional fitting (IPF), is applied to both the household- and person-level attribute joint distributions. As a result of the first step, the total number of households or persons that need to be generated for each cell of the joint-distribution matrix is determined.

In the second step, every household in the sample is given a weight such that the weighted total of households (persons) matches the total number of households (persons) as calculated through the IPF procedure. This step is referred to as the iterative proportional updating (IPU) algorithm, wherein the weights associated with households are iteratively updated such that the weighted frequencies of households and persons match the expanded joint-distribution totals at both the household and person levels.

In the third step, households are drawn through a Monte Carlo simulation process using the weights computed in the second step. This completes the synthetic-population generation procedure. There are a few additional steps to ensure the process is robust and yields the best fitting synthetic population:

- Application of an appropriate rounding procedure so that the frequencies of households (in the sample) to be drawn into the synthetic population are whole numbers (the weights at the end of the second step are likely to be fractional weights and appropriate rounding methods need to be applied to determine whole numbers of households without introducing rounding errors).
- Repeated drawing of a synthetic population through the Monte Carlo simulation procedure with a goodness-of-fit check after each draw. The best draw from among 25 different draws is chosen as the synthetic population for the study.

In the procedure adopted for this study, the output of the synthetic-population-generation process was a sample of households with a frequency or weight variable that indicates the number of times the (sample) household is replicated in the synthetic population. In other words, the synthetic population was not expanded to comprise an exhaustive dataset of more than 300 million records. Instead, a sparse representation of the synthetic-population data files was used to achieve efficiency in data handling and storage. In addition, this format is consistent with the notion of computing “expected” travel demand using the weight variable, as opposed to simulating long-distance travel for each agent in the population (which would be vastly more computationally burdensome). To produce a microsimulation model of long-distance travel for the entire population (at the agent level), the synthetic population can be expanded such that there a unique record for each household and for each person in every household of the synthetic population. Processing and managing such large data highlights big-data challenges that require further study to identify the most efficient ways to process synthetic-population datasets.

Context

The United States includes 50 states, 9 commonwealths/territories, and 6 military states. For this project, the national synthetic-population-generation effort was limited to the 50 states plus the District of Columbia. No synthetic population was generated for the other eight commonwealths/territories (excluding the District of Columbia) and the six military states. According to the 2010 Census, the 50 states and the District of Columbia collectively had a population of 308.7 million people. Of this population, 300.8 million people resided in 116.7 million households, while the remaining 8 million people lived in group quarters. The nation had 3,143 counties, 73,057 Census Tracts, and 217,740 block groups in the 50 states plus the District of Columbia. The frequency distribution of counties, tracts, and block groups across the 51 entities is shown in Table 11.

Table 11: Number of Counties, Census Tracts, and Block Groups, by State

State	Counties	Tracts	Block groups
Alabama	67	1,181	3,438
Alaska	29	167	534
Arizona	15	1,526	4,178
Arkansas	75	686	2,147
California	58	8,057	23,212
Colorado	64	1,249	3,532
Connecticut	8	833	2,585
Delaware	3	218	574
District of Columbia	1	179	450
Florida	67	4,245	11,442
Georgia	159	1,969	5,533
Hawaii	5	351	875
Idaho	44	298	963
Illinois	102	3,123	9,691
Indiana	92	1,511	4,814
Iowa	99	825	2,630
Kansas	105	770	2,351
Kentucky	120	1,115	3,285
Louisiana	64	1,148	3,471
Maine	16	358	1,086
Maryland	24	1,406	3,926
Massachusetts	14	1,478	4,985
Michigan	83	2,813	8,205
Minnesota	87	1,338	4,111
Mississippi	82	664	2,164
Missouri	115	1,393	4,506
Montana	56	271	842
Nebraska	93	532	1,633
Nevada	17	687	1,836
New Hampshire	10	295	922
New Jersey	21	2,010	6,320
New Mexico	33	499	1,449
New York	62	4,919	15,464
North Carolina	100	2,195	6,155
North Dakota	53	205	572
Ohio	88	2,952	9,238
Oklahoma	77	1,046	2,965
Oregon	36	834	2,634
Pennsylvania	67	3,218	9,740
Rhode Island	5	244	815
South Carolina	46	1,103	3,059
South Dakota	66	222	654
Tennessee	95	1,497	4,125
Texas	254	5,265	15,811
Utah	29	588	1,690
Vermont	14	184	522
Virginia	134	1,907	5,332
Washington	39	1,458	4,783
West Virginia	55	484	1,592
Wisconsin	72	1,409	4,489
Wyoming	23	132	410
TOTAL	**3,143**	**73,057**	**217,740**

A decision had to be made regarding the geographic resolution that would be chosen for synthesis of a national synthetic population. The number of counties, at just 3,143, is quite modest and in line with the number of traffic analysis zones (TAZs) included in many regional travel demand models. In fact, several larger regional travel demand models have well over 3,000 TAZs and it has been computationally feasible to generate synthetic populations for such model regions at the level of the individual TAZ without having to utilize massively parallel computing infrastructure. While it would be computationally efficient to synthesize a national population at the level of the county (because there would be only 3,143 geographies), such a level of geographical resolution is less than ideal due to the aggregate and coarse nature of a county. A county can be extremely heterogeneous in nature; as a result, even though the synthetic population may closely match countywide marginal totals, it is likely that such a synthetic population will perform extremely poorly in matching tract- or block-group-level marginal totals.

Ideally, a synthetic-population-generation process at the level of the block group would be performed. The block group is a detailed level of geography for which the Census data provides a rich set of marginal control totals. In addition, a synthetic population that replicates population distributions at the level of the individual block groups would undoubtedly be representative of the population in a region. However, in the context of a national synthetic-population-generation effort, it can be computationally burdensome to synthesize a population for more than 217,000 geographies. Massively parallel computing architectures would have to be deployed to realize reasonably efficient computational times. As a compromise between the geographic detail offered by the block-group-level synthesis and the computational ease afforded by the county level, the research team conducted a Census Tract-level synthesis of the national population. The tract-level synthesis involved generating a population for just over 73,000 Census Tracts in the country; in this instance, the deployment of a modest parallel computer architecture provided reasonable computational times for such a synthesis effort.

In order to perform the synthetic population generation, the research team had to identify the Census datasets that had the richest amount of information at the desired level of geography. With the availability of the decennial 2010 Census data and ACS datasets, it is possible to synthesize a population using a set of datasets that reflect the current state of the population in the United States. Based on a number of considerations—the age of the long-distance travel survey datasets used for model development and estimation, the availability of complete data for a number of variables of interest available at the Census Tract-level, and the base year of the long-distance travel demand model development effort—the project team chose to use the 2007–2011 5-year ACS datasets for population synthesis. Thus, the marginal control data for a variety of household- and person-level attributes was derived from the ACS 2007–2011 5-year data compilation. Similarly, for all syntheses, the ACS PUMS 2007–2011 sample data were used. As a result, the sample and marginal control data are consistent. The latest 2010 decennial Census version of the Mable Geocorr geographic-correspondence files, developed by the Missouri

Census Data Center,[13] are datasets that were used to map the Census Tracts to corresponding PUMAs. Table 12 provides information on the sizes of the ACS marginal and PUMS datasets.

A first-round population synthesis effort was previously undertaken using 2005–2009 ACS data sets (both marginal and PUMS data). The second round of the population synthesis is being undertaken using the more up-to-date 2007–2011 ACS datasets. While the population totals corresponding to the marginal files in Table 12 are those for the 2007–2011 datasets, the sample sizes for PUMS files (in the latter two columns of the table) still correspond to those for the 2005–2009 ACS datasets. Numbers in these two columns will be updated as soon as the second round of population synthesis is complete. Given that the PUMS file sizes are generally consistent across years, the sample sizes shown in Table 12 for PUMS files are likely to also be reflective of sample sizes in the 2007–2011 PUMS files.

Control Variables

PopGen can use any combination of controls for synthesizing a population for the nation. While the use of many control variables may sound appealing from a synthetic population representativeness standpoint, the use of a large number of control variables comes with its own drawbacks. In the presence of large numbers of control variables, thousands—or even millions—of constraints may be generated. Having such a large number of constraints can greatly increase computational time and can lead to sparse matrices; this is because some of the cells in a multidimensional joint-distribution matrix may not have many (or any) observations in the sample file. In addition, several variables may be correlated with one another and it is not necessary to explicitly control for each and every household or person-level socioeconomic variable of interest. Rather, it is important to identify a set of largely uncorrelated dimensions that are key determinants of long-distance travel demand and that would adequately capture the heterogeneity of the population. By choosing a limited set of control variables, the synthetic-population-generation run time can be kept manageable while simultaneously obtaining a representative synthetic population. In addition to identifying an appropriate set of control variables, it is also necessary to specify the categories for each control variable. Once again, the number of categories should be set so that the joint-distribution matrix does not become too sparse while simultaneously retaining a richness of population representation, reflected in the synthetic population that is generated. The research team conducted a number of small-scale trials to identify an appropriate set of controls and categories that may be adopted in a national-level synthetic-population-generation effort at the Census Tract resolution.

[13] Missouri Census Data Center

Table 12: Marginal Population for PUMS and ACS 2007–2011 Data

State	Marginal Control Files			PUMS Files (2005-2009)	
	Total Households	Population in HH	Group Qtr Pop	Household Sample	Person Sample
Alabama	1,831,269	4,747,424	115,816	94,639	236,597
Alaska	252,920	700,703	26,352	14,063	35,156
Arizona	2,344,215	6,337,373	139,384	126,562	316,405
Arkansas	1,121,386	2,895,928	78,931	57,735	144,338
California	12,433,172	36,969,200	819,816	737,628	1,844,071
Colorado	1,941,193	4,966,061	115,878	99,578	248,945
Connecticut	1,360,115	3,558,172	118,152	70,767	176,918
Delaware	332,837	890,856	24,413	17,779	44,448
District of Columbia	260,136	593,955	40,021	11,914	29,785
Florida	7,140,096	18,688,787	421,709	372,266	930,665
Georgia	349,0754	9,600,612	253,199	191,816	479,539
Hawaii	445,513	1,346,554	42,880	26,934	67,335
Idaho	575,497	1,549,987	28,951	31,038	77,595
Illinois	4,773,002	12,790,182	301,773	254,047	635,116
Indiana	2,472,870	6,454,254	186,923	128,379	320,948
Iowa	1,219,137	3,032,266	98,112	60,318	150,795
Kansas	1,104,479	2,830,985	79,074	56,492	141,229
Kentucky	1,681,085	4,316,040	125,870	85,919	214,799
Louisiana	1,675,097	4,484,596	127,427	89,761	224,402
Maine	551,601	1,328,543	35,545	26,302	65,754
Maryland	2,128,377	5,736,545	138,375	114,316	285,791
Massachusetts	2,522,409	6,512,227	238,882	129,643	324,108
Michigan	3,825,182	9,920,621	229,068	195,696	489,240
Minnesota	2,094,265	5,278,190	135,395	105,018	262,544
Mississippi	1,085,062	2,956,700	91,964	58,752	146,881

State	Marginal Control Files			PUMS Files (2005-2009)	
	Total Households	Population in HH	Group Qtr Pop	Household Sample	Person Sample
Missouri	2,354,104	5,955,802	174,142	118,581	296,452
Montana	403,495	982,854	28,849	19,590	48,976
Nebraska	715,703	1,813,061	51,165	36,162	90,404
Nevada	986,741	2,673,396	36,154	53,471	133,677
New Hampshire	514,869	1,315,911	40,104	26,066	65,165
New Jersey	3,180,854	8,753,064	186,876	174,080	435,199
New Mexico	762,002	2,037,136	42,629	40,772	101,929
New York	7,215,687	19,302,448	585,678	383,686	959,216
North Carolina	3,664,119	9,418,736	257,246	188,803	472,006
North Dakota	278,669	666,783	25,056	13,317	33,293
Ohio	4,554,007	11,525,536	306,266	228,423	571,057
Oklahoma	1,432,735	3,714,520	112,017	74,277	185,692
Oregon	1,509,554	3,801,991	86,642	75,855	189,638
Pennsylvania	4,952,566	12,660,739	426,113	251,507	628,768
Rhode Island	410,475	1,053,959	42,663	20,841	52,102
South Carolina	1,758,732	4,575,864	139,154	91,582	228,956
South Dakota	318,466	807,697	34,050	16,121	40,302
Tennessee	2,457,997	6,297,991	153,472	125,653	314,132
Texas	8,667,807	24,774,187	581,139	497,882	1,244,705
Utah	871,358	2,715,379	46,152	54,725	136,812
Vermont	256,711	624,958	25,329	12,390	30,974
Virginia	2,991,025	7,926,192	239,834	158,420	396,051
Washington	2,602,568	6,652,845	139,375	133,146	332,865
West Virginia	740,080	1,846,372	49,382	36,689	91,723
Wisconsin	2,279,738	5,664,893	150,214	112,602	281,506
Wyoming	219,628	554,697	13,712	11,160	27,899
TOTAL	**114,761,359**	**306,603,772**	**7,987,323**	**6,113,162**	**15,282,904**

Table 13 presents the control variables and categories used in the synthetic-population-generation process. At the household level, it can be seen that the control variables include presence or absence of children, household size, age of householder, household income, number of workers in household, and type of household. At the person level, the control variables include age, gender, employment status, and race. The synthetic population also generates a group-quarter population, distinguishing between individuals in institutionalized and noninstitutional group-quarter settings (not shown in Table 13). There are 4,480 constraints (cells in the joint distribution) at the household level and 560 constraints at the person level. In addition, there are two group-quarter constraints. In general, these variables represent important socioeconomic and demographic characteristics that are known to affect travel demand in statistically significant ways. In addition, while a few variables are closely related to one another, they each contribute uniquely to the generation of a representative synthetic population.

The sociodemographic characteristics included in the synthetic-population files are not limited to the variables used as controls. Any uncontrolled variables that are available in the sample data can be added in a straightforward manner to the synthetic population generated by PopGen. The synthetic-population files generated in this project include a number of raw variables (corresponding to the controlled categorized variables) and uncontrolled variables so that a comprehensive set of information is available for model application.

The variables added to the household file (post-synthesis) from the raw PUMS file include the following raw variables, which refers to the original uncategorized variable available in the PUMS file:

- Raw household size.
- Raw household income.
- Number of own children in the household.
- Number of vehicles in the household.
- Raw householder age.
- Number of workers in the household.
- Number of nonworkers in the household.
- Number of full-time workers in the household.
- Number of part-time workers in the household.
- Number of students in the household.

At the person level, only one raw variable is added to the synthetic person file. The raw age variable is appended to the file. All of these variables are matched from the original PUMS records using the unique PUMS identifier associated with each household and person in the sample files. The unique PUMS identifier included in the synthetic-population files may be used to match any other variables in the PUMS files that may be desired for model application in forecasting mode.

Table 13: Household- and Person-Level Constraints for Generating Synthetic Population

Level	Variable Description	Category Value	Category Description
Household	Presence of children in the household	1	Presence of own children
		2	No own child presence
	Household income level	1	Annual household income $0 - $14,999
		2	Annual household income $15,000 - $24,999
		3	Annual household income $25,000 - $34,999
		4	Annual household income $35,000 - $44,999
		5	Annual household income $45,000 - $59,999
		6	Annual household income $60,000 - $99,999
		7	Annual household income $100,000 - $149,999
		8	Annual household income over $150,000
	Householder age	1	Householder age 64 years or less
		2	Householder age 65 years or more
	Household size	1	Household size = 1
		2	Household size = 2
		3	Household size = 3
		4	Household size = 4
		5	Household size = 5
		6	Household size = 6
		7	Household size = 7
	Type of household	1	Family: Married couple
		2	Family: Male householder, no wife
		3	Family: Female householder, no husband
		4	Nonfamily: Householder alone
		5	Nonfamily: Householder not alone
	Number of workers in the household	1	Household has no workers (coded as 1 in synthetic data)
		2	Household has 1 worker (coded as 2 in synthetic data)
		3	Household has 2 workers (coded as 3 in synthetic data)
		4	Household has 3 or more workers (coded as 4 in n. data)
Person	Age of the person	1	Person age under 5 years
		2	Person age 5 to 14 years
		3	Person age 15 to 24 years
		4	Person age 25 to 34 years
		5	Person age 35 to 44 years
		6	Person age 45 to 54 years
		7	Person age 55 to 64 years
		8	Person age 65 to 74 years
		9	Person age 75 to 84 years
		10	Person age 85 years or more
	Gender of the person	1	Male person
		2	Female person
	Race of the person	1	White alone
		2	Black or African American alone
		3	American Indian and Alaska Native alone
		4	Asian alone
		5	Native Hawaiian and Other Pacific Islander alone
		6	Some other race alone
		7	Two or more races
	Employment status of the person	1	Not employed (less than 16 years old)
		2	Employed
		3	Unemployed
		4	Not in labor force (over 64 years old)

PopGen was run for the entire nation, synthesizing the population for each state in a sequential manner. PopGen wrote out the synthetic-population files for each state and conducted an assessment of the performance of the synthesis process for each state before proceeding to a subsequent state. If there was an issue with the synthesis of a population for a particular state, the software paused and displayed an error message asking the operator if the process should be terminated (this was done so that the analyst could fix the error and relaunch the program from the point it stopped) or continued. The process was automated such that no human intervention was needed if there were no errors in the synthesis process, there was adequate storage and memory on the machine, and there was uninterrupted network connectivity. PopGen downloads Census datasets in real time, so it is important to have an uninterrupted and fast network connection that facilitates efficient downloading and processing of data. Census datasets also vary from one release to the next, so some minor modifications to the data download and processing steps need to be made in PopGen if it is desired to apply PopGen in the future using a different release of Census datasets. At the end of the synthetic-population-generation process, PopGen produced 51 folders, with each folder containing:

- Synthetic household and group quarter records;
- Synthetic person records;
- Sample household and group quarter records;
- Sample person records;
- Marginal tract-level records for household attributes; and
- Marginal tract-level records for person attributes.

In PopGen, the number of households synthesized is always equal to the total number of households in the marginal control file. As long-distance travel choices may often involve household-level negotiations and decision processes, it was considered important to exactly match the number of households to control totals. Due to some inconsistency between personal- and household-level controls, it is possible that the total population (number of persons) synthesized by PopGen will be slightly different from the marginal control total for the number of persons in each Census Tract. This modest difference generally arises due to the inevitable modest inconsistencies between household-level marginal control distributions and person-level marginal control distributions. At the end of the synthetic-population-generation process, the synthetic-population files in the 51 folders were integrated to form the national synthetic-population files. The process of concatenation of the 51 household- and person-level synthetic-population files has been automated, but it needs to be run separately as a post-processing step upon completion of the PopGen procedure.

Results of Population Synthesis Process

The synthetic-population-generation process was performed along with periodic checks that identified some issues related to the integrity and consistency of the Census datasets and geographic-correspondence files. The project team completed a first-round population synthesis effort using 2005–2009 ACS datasets and has embarked on a second-round population synthesis effort using 2007–2011 ACS datasets. A future-year synthetic population may be generated using a population-evolution model system or using future-year controls that may be available or inferred at the Census Tract level. It should be noted that model execution time is highly dependent on hardware configurations, availability of multiple processors, and availability of memory. Using a set of six workstations (each with a quad-core CPU) with reasonable memory and processing speed configurations should provide for the generation of a synthetic population for the entire nation in about one week. In general, it should be expected that the synthesis of a nationwide population is a computationally burdensome process that will involve substantial computation times. Moving to a county-level resolution may bring about some efficiencies, but at a great cost in terms of population representativeness.

The synthetic-population files were assessed for each state to ensure that the population synthesized for each Census Tract closely mirrored that in the marginal control datasets from the Census Bureau. Figure 30 through Figure 32 show an illustration of the total households generated for three sample states: Arizona, Connecticut, and Illinois. As expected, the points (each point represents a Census Tract) fall strictly along a 45-degree line, indicating that PopGen synthesizes the exact number of households as contained in the marginal control files. Figure 33 through Figure 35 show a comparison of the synthetic population versus the marginal control total at the person level.

As mentioned previously, PopGen does not exactly match person totals in its attempt to control for the number of households. This is because there will be slight inconsistencies between household- and person-level controls, and the Monte Carlo simulation process by which households are drawn into the synthetic population will introduce some noise; as a result, these graphs do not show perfect adherence to the 45-degree line. Nonetheless, the points are wrapped tightly around the 45-degree line, showing a good level of fit and representativeness of the synthetic population. It should be noted that the goodness of fit would have been less had the procedure not adequately controlled for person-level attributes. By controlling for both household- and person-level attributes, PopGen is able to generate a representative synthetic population where marginal control totals are matched perfectly at the household level and are exceptionally close at the person level.

In addition to ensuring that the population synthesis process generates the correct number of households and persons (in total), it is also useful to assess the performance of the synthesis process by comparing actual marginal control distributions against corresponding distributions in the synthetic population. These comparisons can be performed at various geographic levels, including state, county, and Census Tract level. As the population synthesis was undertaken at the level of the Census Tract, it may be appropriate to compare distributions at this geographic level. If the distributions match closely at this level of geographic resolution, then it implies that the distributions match at higher levels of aggregation (county and state). On the other hand, just because control distributions match at the county or state level, this does not mean that the control distributions would adequately match at the Census Tract level (which is a higher degree of spatial resolution). Comparisons at the block-group level may also be undertaken; but, given the spatial definition of the NUMA zonal system, validation at such a disaggregate spatial level appears unnecessary for the long-distance travel modeling context.

Figure 30: **Comparison of Number of Households in Synthetic Population versus Marginal Control Total for Census Tracts in Arizona**

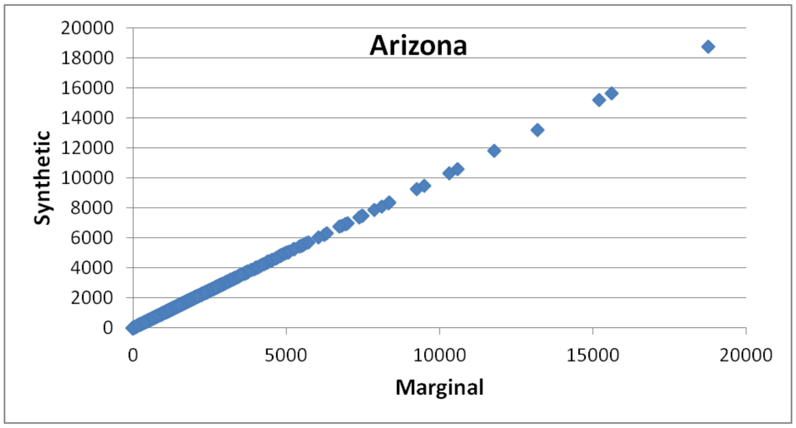


Figure 31: Comparison of Number of Households in Synthetic Population versus Marginal Control Total for Census Tracts in Connecticut

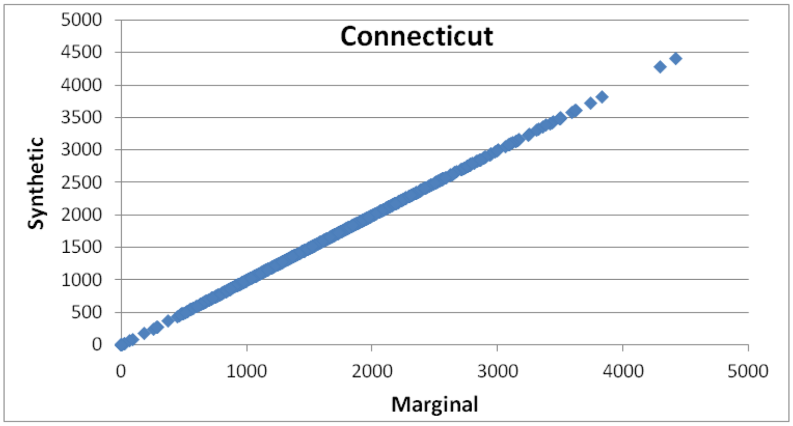


Figure 32: Comparison of Number of Households in Synthetic Population versus Marginal Control Total for Census Tracts in Illinois

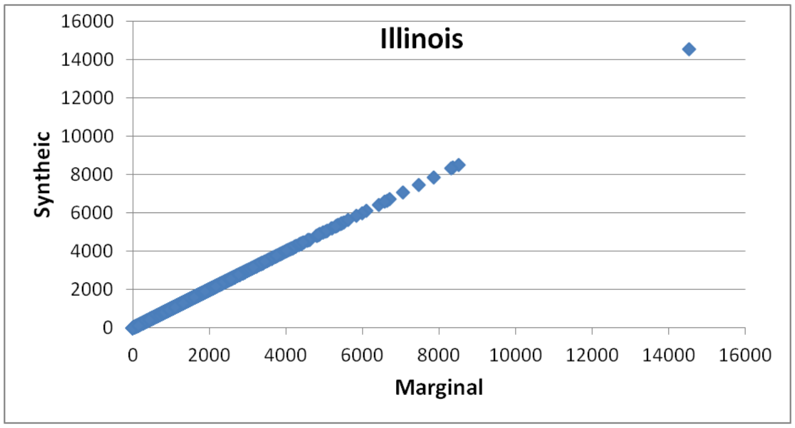


Figure 33: Comparison of Number of Persons in Synthetic Population versus Marginal Control Total for Census Tracts in Arizona

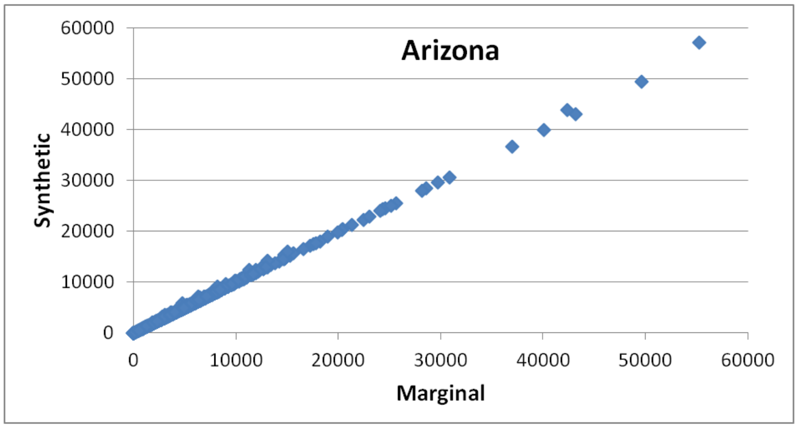


Figure 34: Comparison of Number of Persons in Synthetic Population versus Marginal Control Total for Census Tracts in Connecticut

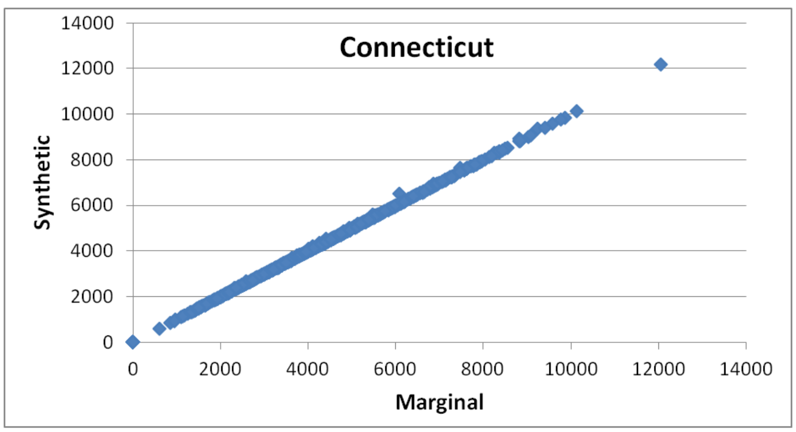
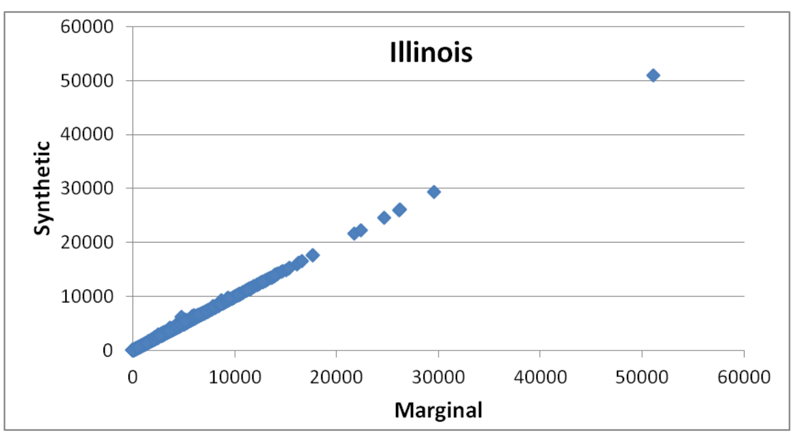


Figure 35: Comparison of Number of Persons in Synthetic Population versus Marginal Control Total for Census Tracts in Illinois



The set of graphs in Figure 36 through Figure 39 show a comparison of household and person attributes for one randomly chosen Census Tract in Maricopa County (Greater Phoenix metropolitan region) in Arizona. In the interest of brevity, such comparisons are not shown for other Census Tracts in the country, although the project team completed an extensive set of comparisons for Census Tracts across the nation to ensure that the population synthesis process is generating a representative population. The comparisons demonstrate the close match between actual population characteristics and synthetic-population characteristics. All of the distributions seen in Figure 40 through Figure 43, for example, show a high level of agreement between the actual marginal control distribution and the synthetic-population distribution. This pattern was found to repeat itself without exception for Census Tracts across the nation.

In sum, the national synthetic-population-generation effort was successful in producing a representative national synthetic population suitable for travel demand modeling and forecasting. An updated synthetic population, based on the 2007–2011 ACS datasets, is under development and will offer a more up-to-date and representative population of the nation.

Figure 36: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Household Type)

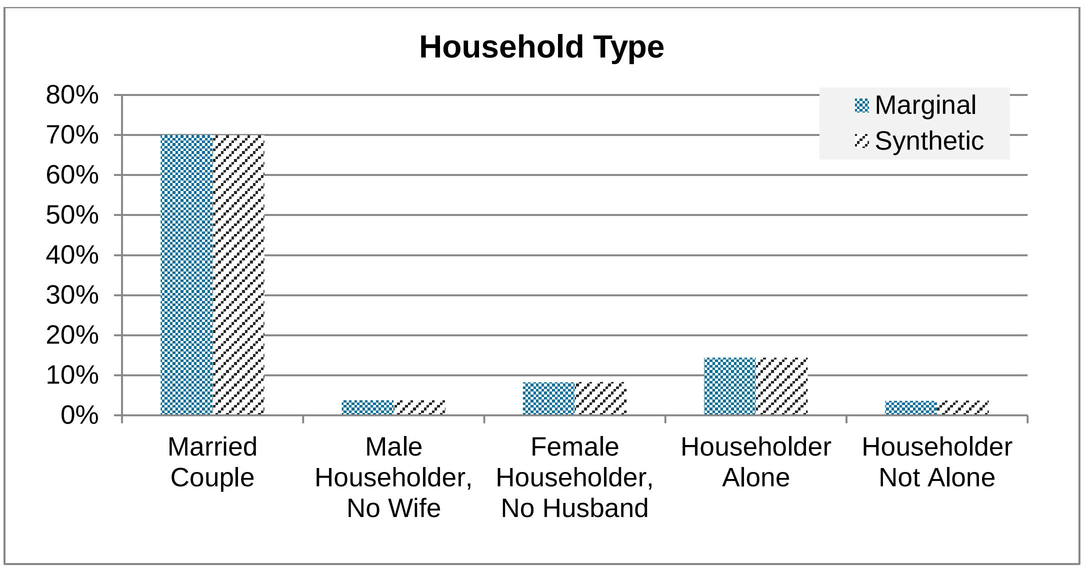


Figure 37: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Household Size)

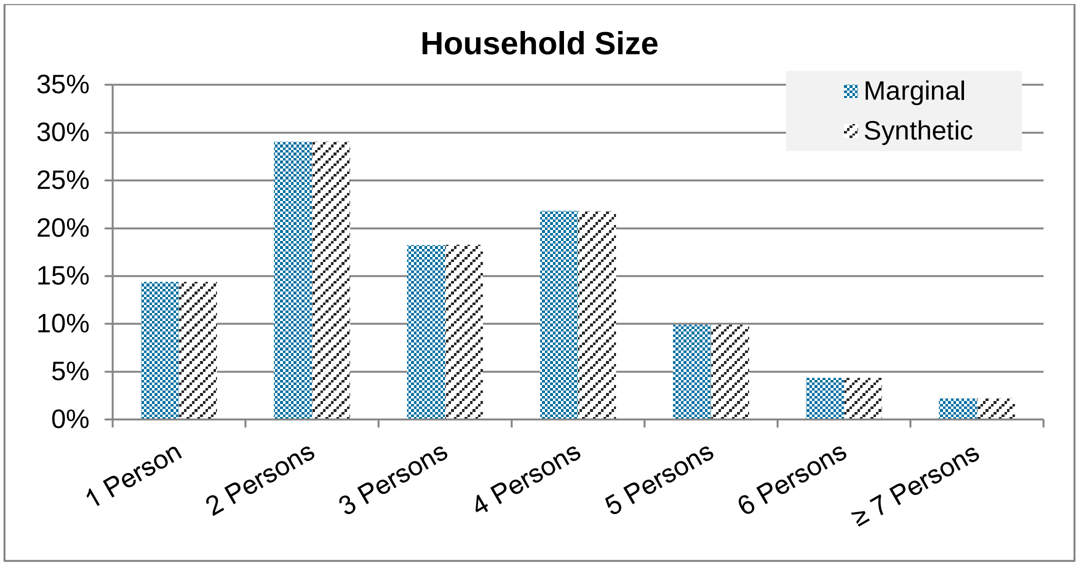


Figure 38: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Number of Workers)

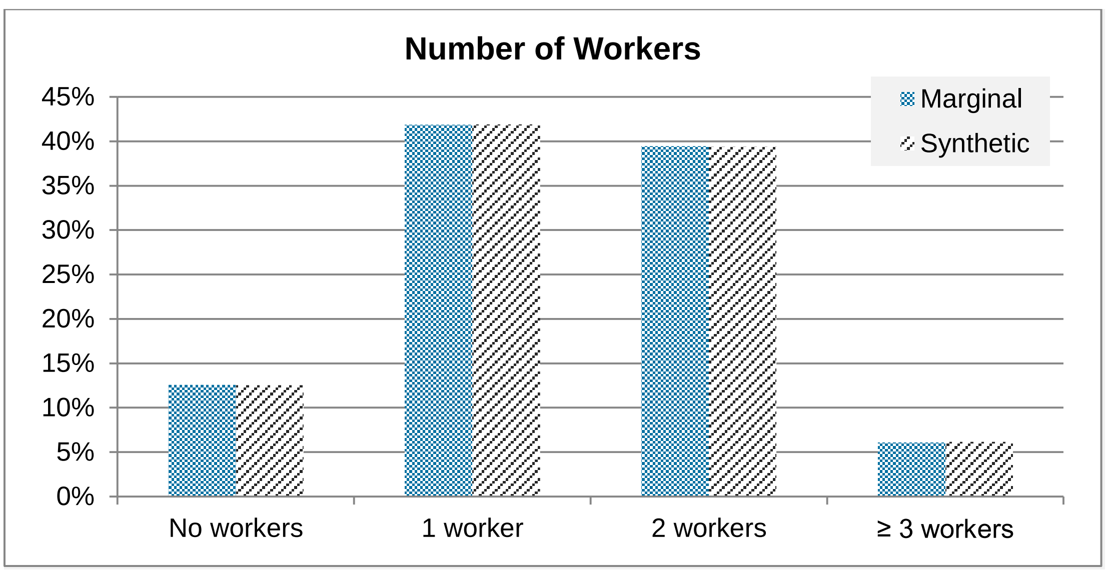


Figure 39: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Household Income)

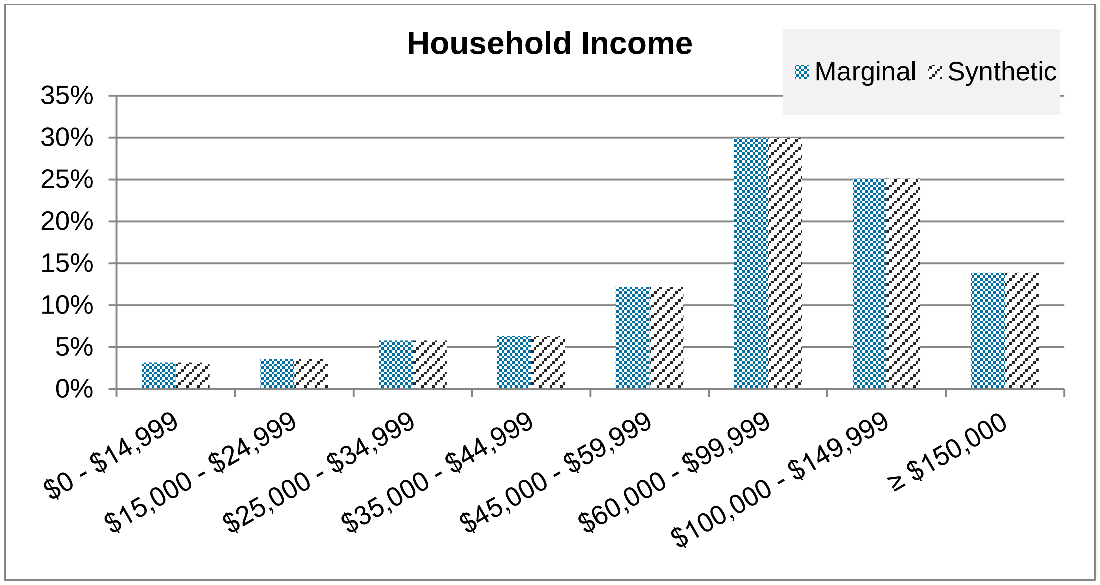


Figure 40: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Gender)

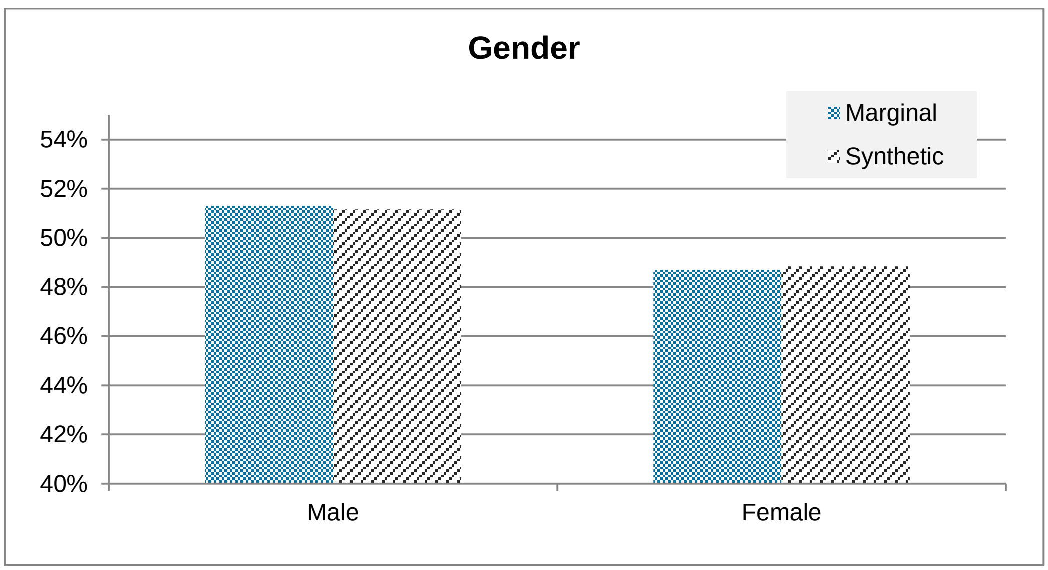


Figure 41: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Race)

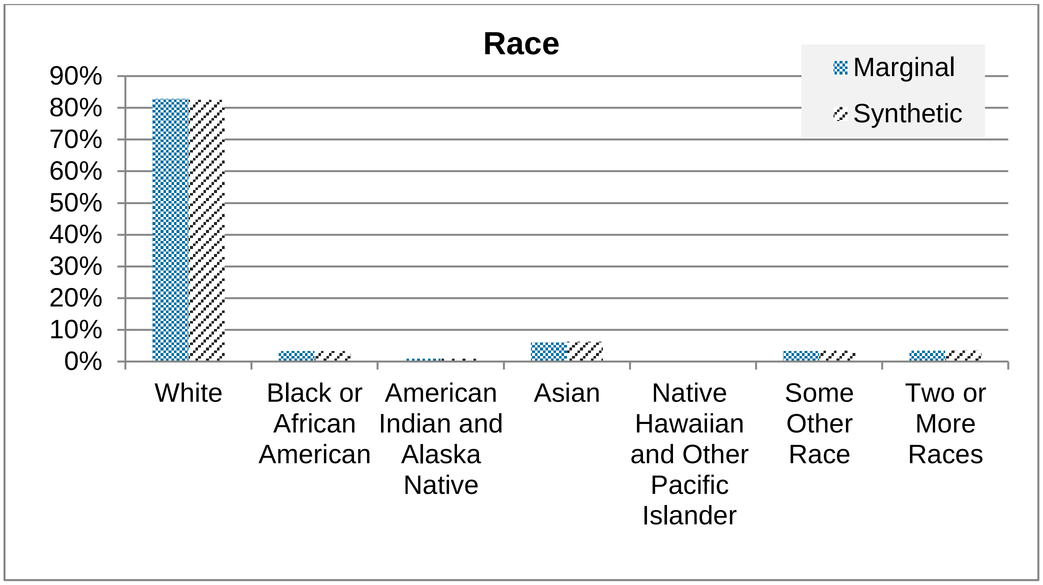


Figure 42: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Employment Status)

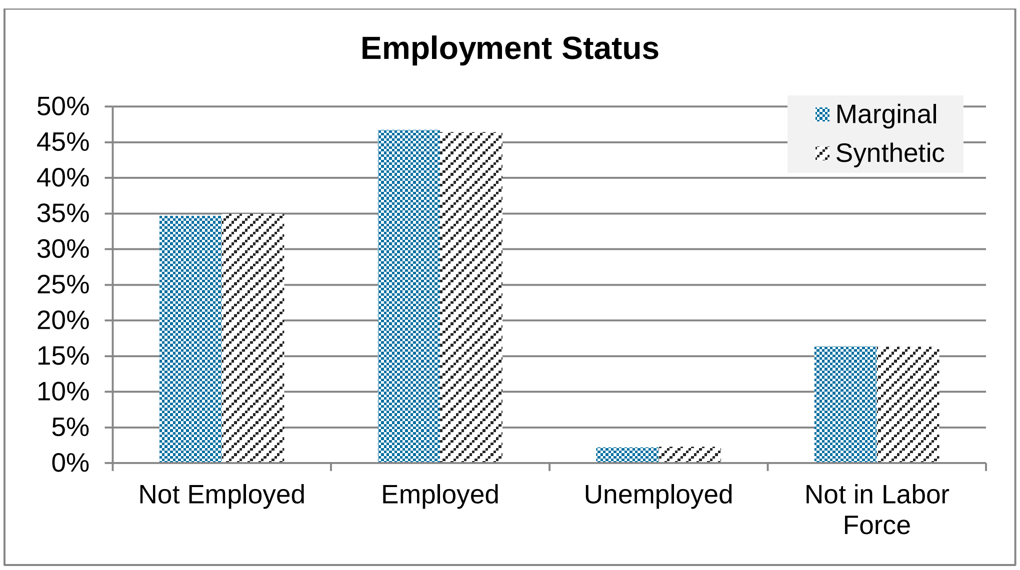
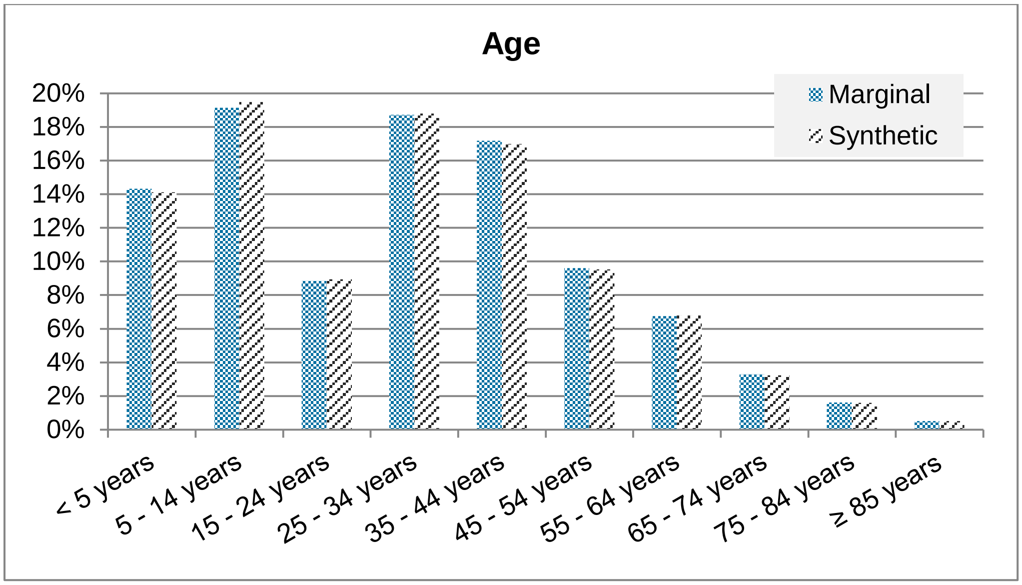


Figure 43: Comparison of Control Distributions between Actual Synthetic Populations (Census Tract 522745 in Maricopa County, Arizona) (Age)



4.4 Tour Generation, Scheduling, and Participation

Overview

This section presents the estimation results of a model structure designed to predict the tour-scheduling dimensions of long-distance travel for nonbusiness and business purposes. Nonbusiness travel is divided into six activity categories, including:

- Visiting friends;
- Relaxation;
- Sightseeing;
- Recreation, including sports, hunting, fishing, boating, camping, etc.;
- Entertainment, including attending the theater or sports events, etc.; and
- Personal business, including weddings, funerals, health treatments, family gatherings, and other personal matters.

Business travel is divided into three activity categories, including:

- Business;
- Business/pleasure; and
- Convention/conference/seminar.

The unit of analysis for the decision-making agent is the household for both types of tour purpose. The household was chosen as the unit of analysis because decisions regarding participation and scheduling of nonbusiness tours are likely to be made collectively at the household level, rather than by each individual separately (see Laesser et al., 2009). Also, some of the business purposes (e.g., business/pleasure and convention/conference/seminar) provide the opportunity for the person making the tour to include his/her family members in the tour; hence, the decision process becomes a collective one, as opposed to being purely based on the individual making the tour. Further, the data source used for the estimation (the 1995 ATS) does not provide socioeconomic details at the individual level; thus, it is impossible to use the individual (as opposed to the household) as the unit of analysis. Whether business or nonbusiness, the tour-scheduling process involves several decisions. The following sequential decision-making process was used in this analysis:

1. The household decides whether to make a tour (either business or nonbusiness).
2. If the decision is "yes," then the household allocates an annual budget for time spent on nonbusiness and business activities.
3. The household further splits the total annual budget into various tour purposes. To illustrate, assume a household decides to engage in nonbusiness travel (step 1) and allocates a 30-day budget (step 2) for it. Then, in step 3 (the current step), the household will further split the 30-day budget into various nonbusiness purposes. The same process is followed for business purposes.
4. The household decides the number of tours to make in a given year based on the budget allocated for various nonbusiness and business purposes. That is, if a household allocates eight days for recreational purposes, the household may make more than one tour to consume their total recreational budget.
5. The household decides the tour-party composition (i.e., number of people in the tour).

The above-described decision-making steps reflect a top-down approach, where the households first decide whether to make a specific kind of tour, followed by determination of tour-specific characteristics, such as duration, number of tours, party size, and composition.

Decision-Making Steps in the Process

This section provides an overview of the models used in the decision-making steps. Additional detail on the underlying methodology of these models can be found in the references provided.

- The first and second steps (decision to make a nonbusiness or business tour followed by the determination of total annual budget) are modeled using a sample selection model (see Greene, 2012, pp. 913–918).
- The third step (allocation of total annual budget in various nonbusiness and business purposes) is modeled using Bhat's (2008) MDCEV model. The MDCEV model simultaneously estimates the participation and duration of a tour. Further, the MDCEV structure allows the households to participate in more than one tour, which is more realistic in long-distance travel. This hypothesis can further be corroborated based on the statistics provided in Table 14 and Table 15 for nonbusiness and business purposes, respectively. As shown in Table 14 and Table 15, households who participate in one kind of tour also participate in other kinds of activities. For example, out of 7,149 households

(Table 14, first row) who participated in tour type “visiting friends” during the third quarter (July–September), 6,505 households (Table 14, last column) also participated in other kinds of activities. The same can also be observed for business purposes. Further, for each tour purpose, there is significant variation in the number of households across different times of the year, which is divided into quarters:

- Q1: January–March (winter).
- Q2: April–June (spring).
- Q3: July–September (summer).
- Q4: October–December (fall).

Thus, to capture seasonal variations in the MDCEV model, a 24-option (6 purposes * 4 quarters) was used to structure nonbusiness purposes and a 12-option (3 purposes * 4 quarters) structure was used for business purposes.

- The fourth step (number of tours by tour purpose) is modeled using a traditional zero truncated Poisson regression (see Greene, 2012, pp. 850–852).
- Finally, the fifth step (party size and composition) is modeled using a multinomial logit model (MNL).

Nonbusiness Travel Model Estimation Results

Decision to Make Nonbusiness Tours and Annual Budget Model

A sample selection model was estimated using the maximum likelihood estimation (MLE) technique to determine the household’s decision to make nonbusiness tours and the annual duration. (Note that for these models, the term “nonbusiness” also excludes long-distance work Commute tours, which were not included in the models presented in this section.)

The results of the first component (decision to conduct nonbusiness travel) are provided in Table 14 that the alternative-specific constants in this context do not have any tangible meaning. They simply represent the effect of all unobserved factors (explanatory variables not considered in the specification). The results indicate that the propensity to conduct nonbusiness travel increases as the household income increases. This result is intuitive, as households with higher income levels can afford to make nonbusiness tours, which may not be an option for low-income households, as long-distance discretionary tours can be expensive (see Mergoupis and Steuer, 2003 for similar results). The number of business tours is also found to have a significant effect on a household’s decision to engage in nonbusiness travel: households that do not make any business tours are more likely to make nonbusiness tours than are the households that make at least one business tour. This may indicate that households where individuals make business tours may face challenges in terms of finding an itinerary feasible for all the individuals in the household (nonbusiness travel is a collective decision made at the household level).

Table 14: Participation and Duration (Nonbusiness Purposes)

Activity Type	Total Number (%) of Households Participating	Participation Duration (Days)				Number of Households (% of Total Number Participating) Who Participate…	
		Mean	St. Dev.	Min.	Max.	Only in this Activity	In This and Other Activity Types
Visit friends/relatives (Q1)	7178 (17.60)	5.84	4.92	1	24	640 (8.92)	6538 (91.08)
Visit friends/relatives (Q2)	10036 (24.6)	5.28	4.11	1	20	1341 (13.36)	8695 (86.64)
Visit friends/relatives (Q3)	10772 (26.41)	5.87	4.85	1	24	1768 (16.41)	9004 (83.59)
Visit friends/relatives (Q4)	9919 (24.31)	5.20	3.72	1	18	1430 (14.42)	8489 (85.58)
Relaxation (Q1)	2570 (6.30)	6.61	6.55	1	35	297 (11.56)	2273 (88.44)
Relaxation (Q2)	4263 (10.45)	5.37	4.69	1	29	631 (14.80)	3632 (85.20)
Relaxation (Q3)	5620 (13.78)	5.36	3.89	1	20	1072 (19.07)	4548 (80.93)
Relaxation (Q4)	2264 (5.55)	4.41	3.31	1	17	257 (11.35)	2007 (88.65)
Sightseeing (Q1)	684 (1.68)	4.31	3.32	1	17	83 (12.13)	601 (87.87)
Sightseeing (Q2)	1637 (4.01)	3.97	2.92	1	14	258 (15.76)	1379 (84.24)
Sightseeing (Q3)	2543 (6.23)	3.83	2.76	1	13	472 (18.56)	2071 (81.44)
Sightseeing (Q4)	1167 (2.86)	3.42	2.38	1	11	171 (14.65)	996 (85.35)
Recreation (Q1)	1698 (4.16)	4.80	3.39	1	17	131 (7.71)	1567 (92.29)
Recreation (Q2)	2492 (6.11)	3.89	2.62	1	13	222 (8.91)	2270 (91.09)
Recreation (Q3)	3449 (8.45)	4.47	3.25	1	16	387 (11.22)	3062 (88.78)
Recreation (Q4)	1650 (4.04)	4.50	3.14	1	15	130 (7.88)	1520 (92.12)
Entertainment (Q1)	1353 (3.32)	3.30	2.24	1	11	130 (9.61)	1223 (90.39)
Entertainment (Q2)	2039 (5.00)	3.19	2.09	1	10	254 (12.46)	1785 (87.54)
Entertainment (Q3)	2283 (5.60)	3.00	2.03	1	10	318 (13.93)	1965 (86.07)
Entertainment (Q4)	1591 (3.90)	2.85	1.71	1	8	156 (9.81)	1435 (90.19)
Personal business (Q1)	2621 (6.42)	5.05	4.99	1	27	275 (10.49)	2346 (89.51)
Personal business (Q2)	4549 (11.15)	4.11	3.68	1	20	650 (14.29)	3899 (85.71)
Personal business (Q3)	4418 (10.83)	3.93	3.41	1	18	650 (14.71)	3768 (85.29)
Personal business (Q4)	3180 (7.80)	3.68	3.12	1	17	428 (13.46)	2752 (86.54)

Q1: January–March (winter), Q2: April–June (spring), Q3: July–September (summer), Q4: October–December (fall)

Table 15: Participation and Duration (Business Purposes)

Activity Type	Total Number (%) of Households Participating	Participation Duration (Days)				Number of Households (% of Total Number Participating) Who Participate…	
		Mean	St. Dev.	Min.	Max.	Only in this Activity	In This and Other Activity Types
Business (Q1)	4464 (30.44)	6.38	6.36	1	33	953 (21.35)	3511 (78.65)
Business (Q2)	5159 (35.18)	5.64	5.35	1	28	1212 (23.49)	3947 (76.51)
Business (Q3)	4472 (30.50)	5.62	5.43	1	29	1067 (23.86)	3405 (76.14)
Business (Q4)	3880 (26.46)	4.84	4.27	1	23	906 (23.35)	2974 (76.65)
Business/Pleasure (Q1)	761 (5.19)	5.20	4.06	1	20	210 (27.60)	551 (72.40)
Business/Pleasure (Q2)	1050 (7.16)	4.62	3.29	1	18	305 (29.05)	745 (70.95)
Business/Pleasure (Q3)	1016 (6.93)	5.11	3.98	1	21	297 (29.23)	719 (70.77)
Business/Pleasure (Q4)	713 (4.86)	4.27	3.04	1	15	220 (30.86)	493 (69.14)
Convention/Conference/Seminar (Q1)	591 (4.03)	3.37	2.00	1	11	166 (28.09)	425 (71.91)
Convention/Conference/Seminar (Q2)	982 (6.70)	3.33	1.89	1	9	372 (37.88)	610 (62.12)
Convention/Conference/Seminar (Q3)	883 (6.02)	3.91	2.45	1	12	393 (44.51)	490 (55.49)
Convention/Conference/Seminar (Q4)	586 (4.00)	2.93	1.66	1	8	204 (34.81)	382 (65.19)

Table 16: Nonbusiness Discretionary Tour Model (Decision to Make Nonbusiness Travel)

Variables	Coeff.	T-Stat
Alternative-Specific Constant	1.626	46.212
Income (Base: 25K-49K)		
Less Than 25K	-0.202	-9.662
50K-99K	0.189	9.068
100k and more	0.389	9.114
Business Tour (Base: Zero Tours)		
1 or more tours	-0.583	-33.784
Family Composition		
Presence of Children (less than 17 Years Old)	-0.215	-6.121
# of individuals between 17 and 49 years old	-0.087	-8.546
# of individuals >= 50 years old	-0.017	-1.483
Working Status		
# of full-time workers	0.017	1.989
# of part-time workers	0.033	2.315
Vehicle Ownership (Base: Three or More Vehicles)		
Zero Vehicle	-0.085	-2.842
One or Two Vehicles	-0.141	-7.071
Household Residential Location (Base: Mountain)		
New England	-0.174	-5.867
Atlantic	-0.163	-4.120
East-North Central	-0.165	-5.008
West-North Central	-0.095	-3.207
South Atlantic	-0.087	-3.011
East-South Central	-0.238	-7.231
West-South Central	-0.120	-3.335
Pacific	-0.062	-1.825

Table 17: Annual Budget Model (Annual Nonbusiness Budget)

Variables	Coeff.	T-Stat
Constant	2.459	122.281
Income (base: 25K-49K)		
Less than 25K	-0.201	-12.901
50K-74K	0.164	9.796
75k – 99K	0.342	13.271
100K and more	0.535	17.670
Working status		
# of full-time workers	-0.125	-28.013
# of part-time workers	-0.069	-6.851
Vehicle ownership (base: Three or more vehicles)		
One or Two vehicles	-0.090	-6.920
Household residential location (base: Mountain)		
New England	-0.126	-6.321
Atlantic	-0.060	-2.190
East-North Central	-0.134	-5.753
West-North Central	-0.159	-7.819
South Atlantic	-0.116	-6.036
East-South Central	-0.272	-11.053
West-South Central	-0.220	-8.473
Pacific	---	---

Note for Table 17: The dependent variable (annual budget) is transformed on logarithmic scale.

Table 17 Model fit:

Sample Size: 47,931 households. Log-likelihood value at convergence: -14999.30.

For the family composition and working-status variables, the coefficient on family composition provides the effect of nonworkers in the household and the coefficient on working status provides the differential impact between a nonworker and a worker in the household. The results indicate that households with more workers are more likely to engage in nonbusiness travel as compared to the households with fewer workers. This is intuitive, as individuals in the workforce have a relatively more hectic day-to-day schedule (long working hours, commute time, etc.) than do nonworkers; thus, these households have a higher likelihood of taking vacations to rest and recharge themselves. The presence of children in the household has a negative effect on the propensity to conduct nonbusiness travel, possibly due to the inherent expenses of child rearing (e.g., education, health care, apparel, etc.), which may limit the household's ability to direct resources to long-distance nonbusiness discretionary tours. The results also suggest that elderly dominated households (age 50 and above) have a high likelihood of conducting nonbusiness travel. Jang et al. (2004) also found that older travelers outspent younger travelers.

Vehicle ownership also has an impact on the decision to engage in nonbusiness travel. Intuitively, the availability of private vehicles in the household increases mobility for the residents, thus providing extra flexibility to the household in terms of scheduling a tour (see Wu et al., 2013 for similar findings).

Finally, location-specific variables (i.e., New England, Middle Atlantic, East-North Central, West-North Central, South Atlantic, East-South Central, West-South Central, Mountain, and Pacific) were added to the model to capture any location-specific effects that were not directly controlled for in the model. The availability of recreational sites in a specific region may induce more nonbusiness discretionary tours in that region's households, as compared to households in other regions. For example, one may expect a higher number of recreation sites (e.g., national parks, camping and hunting sites, etc.) in the Mountain region than in other regions; thus, a Mountain household has a higher likelihood of making a discretionary tour than a household in another region. In the current specification, the Mountain region is the base category; a negative sign on other region indicator variables indicates that households in those regions are less likely to engage in nonbusiness travel as compared to the households in the Mountain region.

For the annual budget model (see Table 17), the duration variable (indicating the number of nights away from home) was transformed on a logarithmic scale to avoid prediction of negative values. First, the alternative-specific constants do not have any tangible meaning, as discussed earlier. Second, with an increase in household income, the annual budget of nonbusiness travel increases (see, Wu et al. 2013 for similar findings). This result is intuitive, as households with higher income levels can afford to make longer-duration nonbusiness discretionary tours, which may not be an option for low-income households. Third, the effect of the number of workers (full- or part-time) aligns with the results, which indicate a small budget for worker-oriented households relative to non-worker-oriented households. The workers typically have only a few days off each year and thus cannot afford to take longer-duration nonbusiness discretionary tours. Fourth, the duration increases with the increase in vehicle ownership (see Nicolau and Mas, 2005, for similar findings). This is understandable, as a high number of vehicles owned may suggest an availability of personal vehicles for all eligible adults, resulting in an increased mobility that enables the individuals in the household to plan for long-duration nonbusiness discretionary tours without worrying about the traveling needs of other individuals in the household (in a household with low vehicle ownership, the daily traveling decisions generally tend to be joint, as opposed to the independent decisions made in households with high vehicle ownership). Similar to the first component of the model, the location-specific variables were added to capture any location-specific effects, which are not directly controlled for in the model.

Finally, the variance-covariance matrix (t-statistics in parenthesis) is presented in Equation 3:

Equation 3: Nonbusiness Annual Budget Model Variance-Covariance Matrix

$$\begin{bmatrix} 1.00(\mathit{fixed}) & 0.37(8.27) \\ & 1.19(72.95) \end{bmatrix}$$

The results indicate a significant correlation between two components (decision to conduct a nonbusiness travel and its duration) of the model, suggesting that some common unobserved factors impact the decision and duration in the same direction.

Nonbusiness Tour-Participation and Duration Model

Table 18 presents the estimation results for nonbusiness purposes. As established, the alternative-specific constants do not have any tangible meaning, but simply represent the effect of all unobserved factors (explanatory variables not considered in the specification). Among the set of explanatory variables considered in the specification, the age-related variables offer a number of interesting results:

- The results indicate that households with children (less than 17 years old) prefer relaxation, sightseeing, recreation, entertainment, and personal business over visiting friends. This is not surprising, as children prefer activities that offer participation, fun, and adventure (see Nickerson and Jurowski, 2001). At the same time, these activities (relaxation, sightseeing, recreation, and entertainment) provide a good opportunity for such households to spend quality time together.
- Young individuals (17–34 years old) seem to prefer recreation (spring and summer), entertainment (winter, spring, and summer), and visiting friends (winter, spring, and summer) over other nonbusiness purposes. These young individuals are likely to be in the early stages of career and family, resulting in long hours at the office in addition to an increase in family responsibilities. Thus, these individuals may prefer activities that potentially provide a quick break from their heavily scheduled lives, such as recreation, entertainment, and even visiting friends.
- Middle-aged individuals (35–49 years old) exhibit a somewhat similar preference, but are also more likely to allocate time to activities such as relaxation (winter, spring, and summer) and sightseeing (winter, spring, and summer) as compared to younger individuals. The increase in preference of middle-aged individuals for relaxation and sightseeing activities can be attributed to the increase in age (relaxation and sightseeing present more convenient forms of vacation, offering both mental and physical relaxation as compared to other activities) and the presence of children and partners in their lives.
- Baby boomers and empty nesters (50–64 years old) prefer visiting friends (winter, spring, and summer), relaxation (winter and spring), sightseeing (winter, spring, and summer), and entertainment (all four seasons) over other tour purposes. However, households with baby boomers and empty nesters are also more likely to spend more time on personal business as compared to households with young and middle-aged individuals. This is not surprising, as these households are in the stage where their children do not live with them anymore, allowing them to be more socially active. These households also are likely to spend more time visiting doctors and getting health treatments due to aging; as a result, it is unsurprising that such households spend more time on personal business than young and middle-aged households.

Seniors (65 and older) also have same preferences for activities (visiting friends, sightseeing, entertainment, and personal business) as baby boomers and empty nesters, which is understandable. Overall, the results suggest significant variation in tour participation across different times of the year by different groups of people (see Grigolon et al., 2014, for similar findings).

Table 18: Tour Duration and Participation Model (Nonbusiness Purposes)

<table>
<tr><th rowspan="4">Alternative
(Base : Visit friends/relatives (Q3))</th><th colspan="12">Explanatory Variables</th></tr>
<tr><th colspan="2" rowspan="2">Alternative-Specific Constant</th><th colspan="10">Family Composition</th></tr>
<tr><th colspan="2">Presence of children (< 17 years old)</th><th colspan="2"># of individuals between 17–34 years old</th><th colspan="2"># of individuals between 35–49 years old</th><th colspan="2"># of individuals between 50–64 years old</th><th colspan="2"># of individuals >= 65 years old</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Visit friends/relatives (Q1)</td><td>-0.448</td><td>-14.707</td><td>-0.142</td><td>-1.579</td><td rowspan="2">0.049</td><td rowspan="2">6.305</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Visit friends/relatives (Q2)</td><td>-0.047</td><td>-1.596</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Visit friends/relatives (Q4)</td><td>0.130</td><td>4.211</td><td>-0.470</td><td>-5.352</td><td>-0.071</td><td>-6.945</td><td>-0.071</td><td>-6.945</td><td>-0.071</td><td>-6.945</td><td>-0.071</td><td>-6.945</td></tr>
<tr><td>Relaxation (Q1)</td><td>-1.623</td><td>-37.27</td><td rowspan="3">0.215</td><td rowspan="3">3.815</td><td rowspan="3">-0.037</td><td rowspan="3">-4.743</td><td></td><td></td><td></td><td></td><td rowspan="2">-0.041</td><td rowspan="2">-2.468</td></tr>
<tr><td>Relaxation (Q2)</td><td>-1.128</td><td>-29.297</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Relaxation (Q3)</td><td>-1.135</td><td>-28.87</td><td></td><td></td><td rowspan="2">-0.051</td><td rowspan="2">-4.548</td><td rowspan="2">-0.128</td><td rowspan="2">-7.448</td></tr>
<tr><td>Relaxation (Q4)</td><td>-1.68</td><td>-37.929</td><td></td><td></td><td>-0.208</td><td>-10.549</td><td>-0.130</td><td>-10.374</td></tr>
<tr><td>Sightseeing (Q1)</td><td>-2.741</td><td>-49.837</td><td>0.215</td><td>3.815</td><td rowspan="3">-0.037</td><td rowspan="3">-4.743</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Sightseeing (Q2)</td><td>-2.130</td><td>-38.744</td><td></td><td></td><td>0.080</td><td>4.485</td><td>0.080</td><td>4.485</td><td>0.080</td><td>4.485</td></tr>
<tr><td>Sightseeing (Q3)</td><td>-1.703</td><td>-37.522</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Sightseeing (Q4)</td><td>-2.07</td><td>-38.31</td><td></td><td></td><td>-0.208</td><td>-10.549</td><td>-0.130</td><td>-10.374</td><td>-0.130</td><td>-10.374</td><td></td><td></td></tr>
<tr><td>Recreation (Q1)</td><td>-1.400</td><td>-24.765</td><td rowspan="2">0.298</td><td rowspan="2">3.637</td><td>-0.049</td><td>-3.388</td><td>-0.049</td><td>-3.388</td><td>-0.049</td><td>-3.388</td><td>-0.381</td><td>-8.745</td></tr>
<tr><td>Recreation (Q2)</td><td>-1.154</td><td>-27.387</td><td></td><td></td><td rowspan="2">0.064</td><td rowspan="2">7.173</td><td rowspan="2">-0.064</td><td rowspan="2">-4.709</td><td rowspan="3">-0.223</td><td rowspan="3">-11.920</td></tr>
<tr><td>Recreation (Q3)</td><td>-0.905</td><td>-19.808</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Recreation (Q4)</td><td>-1.430</td><td>-27.088</td><td></td><td></td><td>-0.049</td><td>-3.388</td><td>-0.049</td><td>-3.388</td><td>-0.049</td><td>-3.388</td></tr>
<tr><td>Entertainment (Q1)</td><td>-1.827</td><td>-44.213</td><td rowspan="3">0.353</td><td rowspan="3">4.989</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Entertainment (Q2)</td><td>-1.431</td><td>-40.567</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Entertainment (Q3)</td><td>-1.536</td><td>-23.99</td><td></td><td></td><td></td><td></td><td></td><td></td><td rowspan="2">0.027</td><td rowspan="2">1.227</td></tr>
<tr><td>Entertainment (Q4)</td><td>-1.688</td><td>-43.897</td><td></td><td></td><td>-0.116</td><td>-4.496</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Personal business (Q1)</td><td>-1.369</td><td>-43.281</td><td>0.114</td><td>1.675</td><td rowspan="4">-0.064</td><td rowspan="4">-7.519</td><td></td><td></td><td rowspan="3">0.065</td><td rowspan="3">7.688</td><td></td><td></td></tr>
<tr><td>Personal business (Q2)</td><td>-0.686</td><td>-17.805</td><td>0.349</td><td>3.839</td><td></td><td></td><td rowspan="2">0.077</td><td rowspan="2">5.219</td></tr>
<tr><td>Personal business (Q3)</td><td>-0.812</td><td>-26.233</td><td rowspan="2">0.114</td><td rowspan="2">1.675</td><td></td><td></td></tr>
<tr><td>Personal business (Q4)</td><td>-1.025</td><td>-30.508</td><td>-0.056</td><td>-4.950</td><td></td><td></td><td></td><td></td></tr>
</table>

Q1: January–March (winter), Q2: April–June (spring), Q3: July–September (summer), Q4: October–December (fall)

Table 18 (cont.): Tour Duration and Participation Model (Nonbusiness Purposes)

Alternative *(Base : Visit friends/relatives (Q3))*	Explanatory Variables							
	Working Status				Household Residential Location (Mountain is the base category)			
	# of full-time workers		# of part-time workers		England		M-Atlantic	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Visit friends/relatives (Q1)	-0.044	-7.637	-0.018	-1.332	0.153	4.653	0.122	2.608
Visit friends/relatives (Q2)								
Visit friends/relatives (Q4)								
Relaxation (Q1)	-0.033	-4.216	-0.035	-1.851	0.507	16.088	0.649	15.699
Relaxation (Q2)								
Relaxation (Q3)	0.066	7.943	0.069	4.254				
Relaxation (Q4)								
Sightseeing (Q1)	-0.033	-4.216					0.211	2.013
Sightseeing (Q2)					0.118	2.558	0.394	6.000
Sightseeing (Q3)	0.066	7.943	0.069	4.254				
Sightseeing (Q4)							0.211	2.013
Recreation (Q1)			0.031	1.619	0.157	2.206		
Recreation (Q2)			-0.042	-1.252	-0.259	-6.363	-0.318	-5.328
Recreation (Q3)	0.062	5.767	0.031	1.619				
Recreation (Q4)								
Entertainment (Q1)					-0.568	-7.889	-0.228	-4.037
Entertainment (Q2)					-0.282	-5.245		
Entertainment (Q3)	0.067	3.082	-0.040	-1.009				
Entertainment (Q4)					-0.568	-7.889		
Personal business (Q1)					-0.170	-5.141	-0.389	-7.939
Personal business (Q2)	-0.047	-4.300	-0.026	-1.008				
Personal business (Q3)								
Personal business (Q4)								

Table 18 (cont.): Tour Duration and Participation Model (Nonbusiness Purposes)

Alternative *(Base : Visit friends/relatives (Q3))*	Explanatory Variables											
	Household Residential Location (Mountain is the base category)											
	EN-Central		WN-Central		S-Atlantic		ES-Central		WS-Central		Pacific	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Visit friends/relatives (Q1)	0.180	5.418	0.121	4.224	0.034	1.175	0.141	4.187	0.136	4.362		
Visit friends/relatives (Q2)												
Visit friends/relatives (Q4)	0.055	1.255										
Relaxation (Q1)	0.480	14.292	-0.140	-3.335	0.192	3.249	0.204	2.494			0.379	11.515
Relaxation (Q2)					0.486	12.791	0.418	11.500	0.170	3.480		
Relaxation (Q3)					0.634	16.382						
Relaxation (Q4)			-0.140	-3.335	0.486	12.791						
Sightseeing (Q1)	-0.275	-1.786	-0.144	-1.866	-0.165	-2.766						
Sightseeing (Q2)			-0.331	-3.876			0.418	11.500				
Sightseeing (Q3)	0.298	5.003	0.174	2.854							0.166	2.164
Sightseeing (Q4)			-0.144	-1.866	0.304	3.955						
Recreation (Q1)			-0.193	-5.649	-0.513	-12.761	-0.574	-9.819				
Recreation (Q2)					-0.351	-5.385					0.344	8.297
Recreation (Q3)					-0.513	-12.761	-0.574	-9.819	-0.328	-4.094		
Recreation (Q4)												
Entertainment (Q1)	-0.212	-1.977			-0.394	-10.777			-0.349	-2.759	0.389	6.894
Entertainment (Q2)											0.179	3.335
Entertainment (Q3)	0.235	3.192	0.122	1.924								
Entertainment (Q4)											0.389	6.894
Personal business (Q1)	-0.196	-4.739	-0.052	-1.685	-0.283	-9.577						
Personal business (Q2)							-0.242	-5.614	-0.179	-4.316	-0.148	-3.906
Personal business (Q3)												
Personal business (Q4)	-0.196	-4.739										

Table 18 (cont.): Tour Duration and Participation Model (Nonbusiness Purposes)

Alternative *(Base : Visit friends/ relatives (Q3))*	Explanatory Variables					
	Vehicle Ownership (base: three or more vehicles)					
	Zero Vehicle		One Vehicle		Two Vehicles	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Visit friends/relatives (Q1)	-0.062	-2.585	-0.062	-2.585	-0.062	-2.585
Visit friends/relatives (Q2)						
Visit friends/relatives (Q4)						
Relaxation (Q1)	-0.136	-4.164	-0.407	-10.842	-0.260	-8.651
Relaxation (Q2)						
Relaxation (Q3)			-0.224	-6.128	-0.074	-2.516
Relaxation (Q4)						
Sightseeing (Q1)			-0.230	-5.116	-0.150	-4.935
Sightseeing (Q2)						
Sightseeing (Q3)	-0.215	-3.393	-0.230	-5.116		
Sightseeing (Q4)						
Recreation (Q1)	-0.457	-11.831	-0.823	-23.833	-0.391	-15.748
Recreation (Q2)						
Recreation (Q3)						
Recreation (Q4)						
Entertainment (Q1)	-0.280	-6.665	-0.434	-12.522	-0.309	-10.959
Entertainment (Q2)						
Entertainment (Q3)						
Entertainment (Q4)						
Personal business (Q1)	-0.124	-3.770	-0.364	-13.206	-0.166	-6.941
Personal business (Q2)						
Personal business (Q3)						
Personal business (Q4)						

Table 18 (cont.): Tour Duration and Participation Model (Nonbusiness Purposes)

Alternative *(Base : Visit friends/relatives (Q3))*	Explanatory Variables							
	Family Income (Base: 25K-49K)							
	Less Than 25K		50K-74K		75K-99K		100K and More	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Visit friends/relatives (Q1)	-0.090	-4.243						
Visit friends/relatives (Q2)								
Visit friends/relatives (Q4)								
Relaxation (Q1)	-0.374	-13.823	0.248	13.623	0.248	13.623	0.438	14.436
Relaxation (Q2)								
Relaxation (Q3)								
Relaxation (Q4)								
Sightseeing (Q1)	-0.394	-6.022	0.180	5.432				
Sightseeing (Q2)	-0.244	-5.425			0.166	2.509		
Sightseeing (Q3)							0.187	2.815
Sightseeing (Q4)	-0.394	-6.022			0.166	2.509		
Recreation (Q1)	-0.615	-14.025	0.163	6.456	0.296	5.997	0.340	9.784
Recreation (Q2)								
Recreation (Q3)	-0.268	-4.724						
Recreation (Q4)	-0.615	-14.025						
Entertainment (Q1)	-0.525	-6.492					0.148	3.525
Entertainment (Q2)	-0.272	-7.552						
Entertainment (Q3)								
Entertainment (Q4)								
Personal business (Q1)								
Personal business (Q2)								
Personal business (Q3)								
Personal business (Q4)								

Table 18 (cont.): Tour Duration and Participation Model (Nonbusiness Purposes)

Alternative *(Base : Visit friends/relatives (Q3))*	Explanatory Variables							
	Business Tour (base: no business tour)							
	One Tour		Two Tours		Three Tours		Four or more Tours	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Visit friends/relatives (Q1)					0.100	5.029	0.100	5.029
Visit friends/relatives (Q2)			0.100	5.029				
Visit friends/relatives (Q4)								
Relaxation (Q1)	0.102	3.163	0.102	3.163	0.102	3.163	0.102	3.163
Relaxation (Q2)								
Relaxation (Q3)								
Relaxation (Q4)			0.102	3.163	0.102	3.163		
Sightseeing (Q1)							0.236	3.032
Sightseeing (Q2)								
Sightseeing (Q3)	-0.137	-2.815	-0.137	-2.815	-0.137	-2.815	-0.137	-2.815
Sightseeing (Q4)			-0.137	-2.815	0.236	3.032	0.236	3.032
Recreation (Q1)			0.119	3.666			0.161	3.774
Recreation (Q2)	-0.076	-2.008					-0.114	-2.080
Recreation (Q3)					0.119	3.666		
Recreation (Q4)							0.161	3.774
Entertainment (Q1)			0.159	2.402				
Entertainment (Q2)			-0.141	-2.340	-0.141	-2.340	-0.195	-4.605
Entertainment (Q3)								
Entertainment (Q4)			0.159	2.402				
Personal business (Q1)			0.095	2.666	0.226	6.825	0.093	4.108
Personal business (Q2)								
Personal business (Q3)								
Personal business (Q4)			0.226	6.825				

Table 18 (cont.): Tour Duration and Participation Model (Nonbusiness Purposes)

Alternative	Translation Parameters	
	Coeff.	T-Stat
Visit friends/relatives (Q3)	5.869	38.841
Visit friends/relatives (Q1)	5.851	31.156
Visit friends/relatives (Q2)	5.013	35.979
Visit friends/relatives (Q4)	5.085	34.273
Relaxation (Q1)	8.261	19.404
Relaxation (Q2)	6.969	23.848
Relaxation (Q3)	7.583	25.828
Relaxation (Q4)	5.461	16.708
Sightseeing (Q1)	5.863	9.683
Sightseeing (Q2)	5.538	14.415
Sightseeing (Q3)	5.551	17.837
Sightseeing (Q4)	4.626	12.219
Recreation (Q1)	6.007	13.627
Recreation (Q2)	4.453	16.215
Recreation (Q3)	5.097	20.554
Recreation (Q4)	5.124	13.615
Entertainment (Q1)	3.919	12.663
Entertainment (Q2)	3.906	15.064
Entertainment (Q3)	3.753	16.26
Entertainment (Q4)	3.498	12.86
Personal business (Q1)	5.718	20.078
Personal business (Q2)	4.568	25.394
Personal business (Q3)	4.382	25.415
Personal business (Q4)	4.057	21.369

Table 18 Model Fit:

Sample Size: 40,794 households. Log-likelihood value at convergence: -369209.39.

Next, the coefficient on the working-status variable provides the differential impact between a nonworker and worker in the household. Thus, if the coefficient on working status is positive, it means that a household with more workers is more likely to participate in an activity type as compared to a household with more nonworkers. The results suggest that with an increase in the number of full-time workers (people working 35 or more hours per week), the household is more likely to pursue activities such as relaxation (summer and fall), sightseeing (summer and fall), recreation (all four seasons), entertainment (winter, spring, and fall), and personal business (winter, summer, and fall). This is not surprising, as full-time workers typically only get occasional getaways; when they do have vacation time, they prefer activities that offer more physical and mental relaxation than simply visiting friends can provide. The effect of part-time workers (people working fewer than 35 hours per week) is also similar to the effect of full-time workers, suggesting that—irrespective of the number of work hours—workers generally prefer relaxing vacations (relaxation, sightseeing, recreation, entertainment) over visiting family and friends (see LaMondia et al., 2008 for similar findings).

Location-specific variables (New England, Middle Atlantic, East-North Central, West-North Central, South Atlantic, East-South Central, West-South Central, Mountain, and Pacific) were also added to capture the effects of location on different nonbusiness purposes. This variable may capture the effect of the presence of various vacation-specific sites, which are not directly included in the model. The base category is Mountain; a negative sign on the location variable for a particular nonbusiness purpose suggests that households in the Mountain region are more likely to participate in that particular purpose than are households in that specified region. For example, for the recreation purpose (spring, summer, and fall), the coefficient corresponding to the New England region variable is negative (-0.254) suggesting that households in the Mountain region are more likely to make recreational tours in spring, summer, and fall than are households in the New England region. This outcome might be due to the abundance of recreational sites in the Mountain region as compared to the New England region, making it easy for Mountain residents to engage in hunting, fishing, boating, camping, etc. The other location-specific variables can be interpreted similarly.

Vehicle ownership is found to have a significant impact on participation and duration of all nonbusiness purposes. The results indicate that with an increase in vehicle ownership, the household is more likely to participate in various nonbusiness tour purposes (see the increase in magnitude of vehicle ownership coefficient across columns for all tour purposes) (see LaMondia et al., 2008 for similar findings). This is understandable, as a higher level of vehicle ownership may suggest an availability of personal vehicles for all eligible adults, resulting in an increased mobility that enables the individuals in the household to make various nonbusiness tours, as desired.

Household income also has the same effect as vehicle ownership; an increase in household income is associated with an increase in the likelihood of participation in various nonbusiness purposes (see the increase in magnitude of household income coefficient across the columns for all tour purposes) (see LaMondia et al., 2008 for similar findings). This result is intuitive, as households with higher-income levels can afford to participate in various nonbusiness purposes, which low-income households may not be able to afford.

Finally, the effect of the number of business tours on the likelihood of making nonbusiness tours is also intuitive. The results indicate that, in general, the households with more business tours are more likely to also participate in nonbusiness tours. The higher number of business tours may signify a more hectic work schedule for individuals in the households, which may necessitate nonbusiness discretionary tours in order to allow those workers to take a break. However, there are also several negative effects. For example, the households that do not make any business tours are more likely to participate in sightseeing during summer than are households that make at least one business tour. This result suggests that, given the flexibility in time (a household with no business tours may have a more flexible schedule than a household with at least one business tour), a household may schedule travel for some nonbusiness purposes to optimize their experience, such as vacations to spots that become more scenic during certain times of the year.

The translation parameters can be viewed as a measure of satiation. A large value for the translation parameter for a certain nonbusiness purpose indicates less satiation; as a result, households may invest more time into that tour, even increasing the number of such episodes. For example, the translation parameter for relaxation is larger than that of other purposes, indicating that households invest more time in relaxation-related activities.

Nonbusiness Tour-Frequency Model

Table 19 through Table 24 present the result for six nonbusiness purposes (i.e., visit friends/relatives, relaxation, sightseeing, recreation, entertainment, and personal business). In the next paragraph, the results of the tour-frequency model for different nonbusiness purposes are briefly discussed.

Table 19 presents the estimates for the visit friends/relatives purpose. The results indicate that a high-income household is likely to make more such tours than a low-income household. The presence of children in the household has a negative impact on number of tours because children prefer activities that offer fun, participation, and adventure. The propensity to make more of these tours decreases with an increase in the age of individuals in the household. Therefore, a young-individual-dominated (17–34 years old) household is likely to make more of these tours relative to a middle-aged, baby boomer and empty nester, or elderly household. This reflects the human desire of individuals to be more sociable during the early stages of life. Further, worker-oriented households are likely to make more of such tours relative to non-worker-oriented households; this is probably due to participation in social gatherings of workers’ peer groups. Overall, vehicle ownership has a negative impact on the frequency of tours. One would expect an increase in tour frequency with an increase in vehicle ownership. This is because accessibility to a personal vehicle increases mobility and reduces dependency on others for traveling. On the other hand, it is possible that in households with low vehicle ownership, the primary driver (individual who frequently uses the vehicle in the household) may cater to the traveling needs of other household members, such as dropping off kids at relatives house over the weekend and picking them up later, or drop-off/pick-up of other individuals, resulting in an overall increase in tour frequency. Location-specific variables were also included to capture any indirect effects that might be specific to a location. The base category is New England; as a result, a negative sign on a location variable suggests that households in that particular location are likely to make fewer such tours than other households from New England. Finally, as expected, the total budget allocated to the tour purpose has a positive impact on tour frequency.

Table 19: Tour-Frequency Model (Nonbusiness Purposes)

<table>
<tr><th rowspan="2">Variables</th><th colspan="2">Visit friends/relatives:Q1</th><th colspan="2">Visit friends/relatives:Q2</th><th colspan="2">Visit friends/relatives:Q3</th><th colspan="2">Visit friends/relatives:Q4</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-Specific Constant</td><td>-1.201</td><td>-22.589</td><td>-1.132</td><td>-23.262</td><td>-1.220</td><td>-24.869</td><td>-1.335</td><td>-28.339</td></tr>
<tr><td colspan="9">Income (base: less than 25K)</td></tr>
<tr><td>More than 25K</td><td></td><td></td><td></td><td></td><td>0.119</td><td>4.316</td><td></td><td></td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td>Presence of Children (less than 17 years old)</td><td>-0.301</td><td>-2.650</td><td>-0.419</td><td>-5.025</td><td>-0.400</td><td>-5.030</td><td>-0.392</td><td>-3.737</td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td>0.369</td><td>20.636</td><td>0.371</td><td>26.288</td><td>0.431</td><td>42.828</td><td>0.432</td><td>30.159</td></tr>
<tr><td># of individuals between 35 and 49 years old</td><td>0.303</td><td>16.984</td><td>0.321</td><td>24.89</td><td>0.389</td><td>44.159</td><td>0.398</td><td>30.027</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td>0.256</td><td>11.709</td><td>0.249</td><td>15.353</td><td>0.337</td><td>22.306</td><td>0.325</td><td>19.018</td></tr>
<tr><td># of individuals >= 65 years old</td><td>0.072</td><td>2.122</td><td>0.085</td><td>3.141</td><td>0.238</td><td>9.690</td><td>0.182</td><td>6.930</td></tr>
<tr><td colspan="9">Working status</td></tr>
<tr><td># of full-time workers</td><td rowspan="2">0.045</td><td rowspan="2">3.266</td><td rowspan="2">0.037</td><td rowspan="2">3.539</td><td></td><td></td><td rowspan="2">0.034</td><td rowspan="2">3.145</td></tr>
<tr><td># of part-time workers</td><td></td><td></td></tr>
<tr><td colspan="9">Vehicle ownership (base: Three or more)</td></tr>
<tr><td>Zero vehicle</td><td>-0.166</td><td>-3.180</td><td rowspan="2">0.034</td><td rowspan="2">0.985</td><td></td><td></td><td></td><td></td></tr>
<tr><td>One vehicle</td><td></td><td></td><td rowspan="2">0.127</td><td rowspan="2">5.255</td><td rowspan="2">0.105</td><td rowspan="2">4.239</td></tr>
<tr><td>Two vehicles</td><td></td><td></td><td>0.118</td><td>4.105</td></tr>
<tr><td colspan="9">Household residential location (base: New England)</td></tr>
<tr><td>Atlantic</td><td></td><td></td><td></td><td></td><td>-0.089</td><td>-1.720</td><td>-0.126</td><td>-1.977</td></tr>
<tr><td>East-North Central</td><td>0.132</td><td>2.495</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>West-North Central</td><td>0.171</td><td>3.862</td><td>0.203</td><td>6.385</td><td>0.152</td><td>4.374</td><td>0.226</td><td>7.416</td></tr>
<tr><td>South Atlantic</td><td>0.072</td><td>1.515</td><td></td><td></td><td>-0.118</td><td>-3.441</td><td></td><td></td></tr>
<tr><td>East-South Central</td><td>0.285</td><td>5.424</td><td>0.153</td><td>3.489</td><td></td><td></td><td>0.132</td><td>3.069</td></tr>
<tr><td>West-South Central</td><td></td><td></td><td>0.116</td><td>2.406</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mountain</td><td>-0.077</td><td>-1.599</td><td>-0.088</td><td>-2.583</td><td>-0.201</td><td>-6.190</td><td>-0.102</td><td>-3.154</td></tr>
<tr><td>Pacific</td><td>-0.300</td><td>-4.521</td><td>-0.268</td><td>-4.509</td><td>-0.333</td><td>-7.275</td><td>-0.323</td><td>-6.252</td></tr>
<tr><td>Natural Logarithm of Total Duration Allocated to the Alternative</td><td>0.558</td><td>34.178</td><td>0.438</td><td>30.306</td><td>0.304</td><td>23.771</td><td>0.423</td><td>26.288</td></tr>
<tr><td>Sample size</td><td colspan="2">7178</td><td colspan="2">10036</td><td colspan="2">10772</td><td colspan="2">9919</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="2">-11652.87</td><td colspan="2">-14163.08</td><td colspan="2">-14566.96</td><td colspan="2">-12838.00</td></tr>
</table>

Table 20: Tour-Frequency Model (Nonbusiness Purposes)

<table>
<tr><th rowspan="2">Variables</th><th colspan="2">Relaxation (Q1)</th><th colspan="2">Relaxation (Q2)</th><th colspan="2">Relaxation (Q3)</th><th colspan="2">Relaxation (Q4)</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.807</td><td>-14.944</td><td>-1.673</td><td>-20.740</td><td>-1.778</td><td>-22.582</td><td>-2.298</td><td>-14.904</td></tr>
<tr><td colspan="9">Income (base: less than 25K)</td></tr>
<tr><td>25K – 49K</td><td></td><td></td><td></td><td></td><td rowspan="2">0.123</td><td rowspan="2">1.875</td><td rowspan="4">0.440</td><td rowspan="4">3.437</td></tr>
<tr><td>50K – 74K</td><td rowspan="2">0.183</td><td rowspan="2">2.692</td><td rowspan="2">0.079</td><td rowspan="2">1.772</td></tr>
<tr><td>75K – 99K</td><td rowspan="2">0.138</td><td rowspan="2">1.913</td></tr>
<tr><td>100K and more</td><td>0.257</td><td>2.629</td><td>0.219</td><td>3.157</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td rowspan="2">0.310</td><td rowspan="2">10.452</td><td>0.397</td><td>16.856</td><td rowspan="2">0.413</td><td rowspan="2">32.881</td><td>0.449</td><td>13.011</td></tr>
<tr><td># of individuals between 35 and 49 years old</td><td>0.359</td><td>14.924</td><td>0.412</td><td>16.116</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td>0.238</td><td>6.680</td><td>0.319</td><td>11.687</td><td>0.345</td><td>17.040</td><td>0.335</td><td>10.333</td></tr>
<tr><td># of individuals >= 65 years old</td><td></td><td></td><td></td><td></td><td>0.258</td><td>7.577</td><td>0.123</td><td>1.771</td></tr>
<tr><td colspan="9">Working status</td></tr>
<tr><td># of full-time workers</td><td>0.061</td><td>2.309</td><td rowspan="2">0.042</td><td rowspan="2">2.116</td><td></td><td></td><td></td><td></td></tr>
<tr><td># of part-time workers</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="9">Vehicle ownership (base: Three or more)</td></tr>
<tr><td>Zero vehicle</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>One vehicle</td><td rowspan="2">0.179</td><td rowspan="2">2.869</td><td rowspan="2">0.129</td><td rowspan="2">2.909</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Two vehicles</td><td rowspan="2">0.052</td><td rowspan="2">1.291</td><td rowspan="2">0.141</td><td rowspan="2">1.521</td></tr>
<tr><td>Three or more vehicles</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="9">Household residential location (base: New England or Mountain)</td></tr>
<tr><td>Atlantic</td><td>0.137</td><td>1.219</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>East-North Central</td><td>0.143</td><td>1.365</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>West-North Central</td><td></td><td></td><td></td><td></td><td>0.096</td><td>1.860</td><td></td><td></td></tr>
<tr><td>South Atlantic</td><td>0.152</td><td>1.841</td><td>0.117</td><td>2.431</td><td></td><td></td><td></td><td></td></tr>
<tr><td>East-South Central</td><td></td><td></td><td></td><td></td><td>-0.133</td><td>-2.434</td><td></td><td></td></tr>
<tr><td>West-South Central</td><td>0.340</td><td>2.782</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Pacific</td><td>-0.149</td><td>-1.349</td><td>-0.124</td><td>-1.565</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Natural logarithm of total duration allocated to the alternative</td><td>0.488</td><td>15.213</td><td>0.427</td><td>20.627</td><td>0.519</td><td>24.574</td><td>0.496</td><td>12.331</td></tr>
<tr><td>Sample size</td><td>2570</td><td>4263</td><td>5620</td><td>2264</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Log-likelihood value at convergence</td><td>-3416.77</td><td>-5318.75</td><td>-7347.79</td><td>-2308.26</td><td></td><td></td><td></td><td></td></tr>
</table>

Table 21: Tour-Frequency Model (Nonbusiness Purposes)

Variables	Sightseeing (Q1)		Sightseeing (Q2)		Sightseeing (Q3)		Sightseeing (Q4)	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.331	-7.282	-1.998	-14.653	-1.627	-19.042	-1.846	-10.205
Income (base: greater than 75K)								
Less than 75K							-0.112	-1.205
Family composition								
# of individuals between 17 and 34 years old	0.430	10.990	0.546	21.290	0.511	26.407	0.465	14.746
# of individuals between 35 and 49 years old								
# of individuals between 50 and 64 years old	0.262	3.206	0.376	8.241	0.383	13.288	0.205	2.924
# of individuals >= 65 years old								
Working status								
# of full-time workers					0.025	1.757		
Vehicle ownership (base: Three or more)								
Zero vehicle					0.102	2.197		
One vehicle								
Two vehicles	-0.190	-1.425	0.296	3.459			0.285	2.119
Three or more vehicles	-0.456	-2.703	0.120	1.263				
Household residential location (base: New England)								
Atlantic	-0.466	-1.884						
East-North Central					0.129	1.967		
West-North Central			-0.200	-1.697				
South Atlantic	0.320	1.897						
East-South Central								
West-South Central								
Mountain	-0.185	-1.290						
Pacific	-0.278	-1.522						
Natural logarithm of total duration allocated to the alternative	0.302	3.578	0.190	4.532	0.082	3.102	0.273	5.503
Sample size	684		1637		2543		1167	
Log-likelihood value at convergence	-664.44		-1486.03		-2545.83		-839.52	

Table 22: Tour-Frequency Model (Nonbusiness Purposes)

Variables	Recreation (Q1)		Recreation (Q2)		Recreation (Q3)		Recreation (Q4)	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-Specific Constant	-1.436	-12.526	-1.363	-16.606	-1.205	-17.454	-1.458	-10.933
Income (base: less than 50K)								
50K – 74K	-0.164	-2.149					0.150	1.926
75K – 99K	-0.347	-4.375						
100K and more			0.142	1.780				
Family composition								
# of individuals between 17 and 34 years old	0.370	16.287	0.315	10.407	0.340	23.980	0.285	10.871
# of individuals between 35 and 49 years old			0.253	9.198				
# of individuals between 50 and 64 years old	0.295	7.572	0.170	5.008	0.239	9.527	0.230	5.696
# of individuals >= 65 years old								
Working status								
# of full-time workers			0.029	1.252				
Vehicle ownership (base: Three or more)								
Zero vehicle								
One vehicle								
Two vehicles					0.089	2.313		
Household residential location (base: New England or Mountain)								
Atlantic	-0.434	-3.244						
East-North Central	-0.350	-3.032			-0.102	-1.615		
West-North Central	-0.403	-3.957	0.177	2.423			-0.220	-2.087
South Atlantic					-0.158	-2.551		
East-South Central	-0.552	-4.449						
West-South Central	-0.212	-1.875	0.206	2.179				
Pacific	-0.245	-1.833					-0.169	-1.180
Natural Logarithm of Total Duration Allocated to the Alternative	0.812	20.417	0.661	18.802	0.465	16.686	0.591	11.014
Sample size	1698		2492		3449		1650	
Log-likelihood value at convergence	-2628.60		-3283.77		-4862.08		-2047.50	

Table 23: Tour-Frequency Model (Nonbusiness Purposes)

<table>
<tr><th rowspan="2">Variables</th><th colspan="2">Entertainment (Q1)</th><th colspan="2">Entertainment (Q2)</th><th colspan="2">Entertainment (Q3)</th><th colspan="2">Entertainment (Q4)</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-2.000</td><td>-13.022</td><td>-1.689</td><td>-11.550</td><td>-1.554</td><td>-16.593</td><td>-1.732</td><td>-13.158</td></tr>
<tr><td colspan="9">Income (base: 100K and more)</td></tr>
<tr><td>Less than 25K</td><td></td><td></td><td rowspan="3">0.115</td><td rowspan="3">1.373</td><td></td><td></td><td></td><td></td></tr>
<tr><td>25K – 49K</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>50K – 74K</td><td rowspan="2">0.108</td><td rowspan="2">1.286</td><td></td><td></td><td></td><td></td></tr>
<tr><td>75K – 99K</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td rowspan="2">0.452</td><td rowspan="2">13.557</td><td rowspan="2">0.368</td><td rowspan="2">16.153</td><td rowspan="2">0.424</td><td rowspan="2">22.134</td><td rowspan="2">0.413</td><td rowspan="2">15.692</td></tr>
<tr><td># of individuals between 35 and 49 years old</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td rowspan="2">0.395</td><td rowspan="2">7.863</td><td rowspan="2">0.239</td><td rowspan="2">5.917</td><td rowspan="2">0.350</td><td rowspan="2">10.487</td><td rowspan="2">0.264</td><td rowspan="2">4.753</td></tr>
<tr><td># of individuals >= 65 years old</td></tr>
<tr><td colspan="9">Vehicle ownership (base: Three or more)</td></tr>
<tr><td>Zero vehicle</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>One vehicle</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Two vehicles</td><td></td><td></td><td></td><td></td><td>0.071</td><td>1.278</td><td></td><td></td></tr>
<tr><td colspan="9">Household residential location (base: New England)</td></tr>
<tr><td>Atlantic</td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.205</td><td>1.377</td></tr>
<tr><td>East-North Central</td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.229</td><td>1.737</td></tr>
<tr><td>West-North Central</td><td>0.211</td><td>1.902</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>South Atlantic</td><td></td><td></td><td>-0.127</td><td>-1.298</td><td>-0.265</td><td>-3.202</td><td></td><td></td></tr>
<tr><td>East-South Central</td><td>0.314</td><td>1.963</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>West-South Central</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Mountain</td><td>0.262</td><td>2.560</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Pacific</td><td></td><td></td><td>-0.197</td><td>-1.971</td><td>-0.627</td><td>-4.966</td><td></td><td></td></tr>
<tr><td>Natural logarithm of total duration allocated to the alternative</td><td>0.572</td><td>10.824</td><td>0.599</td><td>10.556</td><td>0.463</td><td>12.677</td><td>0.508</td><td>8.130</td></tr>
<tr><td>Sample size</td><td colspan="2">1353</td><td colspan="2">2039</td><td colspan="2">2283</td><td colspan="2">1591</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="2">-1579.50</td><td colspan="2">-2167.82</td><td colspan="2">-2544.64</td><td colspan="2">-1618.74</td></tr>
</table>

Table 24: Tour-Frequency Model (Nonbusiness Purposes)

Variables	Personal Business (Q1)		Personal Business (Q2)		Personal Business (Q3)		Personal Business (Q4)	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-Specific Constant	-1.059	-11.989	-1.197	-17.256	-1.336	-19.279	-0.909	-8.621
Income (base: less than 25K)								
25K – 49K	0.116	2.110					-0.104	-1.456
50K – 74K								
75K – 99K							-0.185	-2.014
100K and more								
Family composition								
# of individuals between 17 and 34 years old	0.273	13.650	0.314	12.610	0.343	15.890	0.327	17.716
# of individuals between 35 and 49 years old								
# of individuals between 50 and 64 years old								
# of individuals >= 65 years old	0.173	3.983	0.208	5.594	0.308	8.440	0.231	6.331
Working status								
# of full-time workers			0.038	1.829	0.053	3.180		
# of part-time workers			0.062	2.215				
Vehicle ownership (base: Three or more)								
Zero vehicle					-0.104	-1.552	-0.137	-2.449
One vehicle					-0.208	-3.155		
Two vehicles								
Household residential location (base: New England)								
Atlantic			-0.228	-1.891				
East-North Central	-0.264	-2.343	-0.115	-1.235				
West-North Central	0.244	3.316	0.204	3.502	0.186	3.403	0.215	3.219
South Atlantic								
East-South Central	0.351	4.177	0.153	1.906	0.187	2.302	0.210	2.417
West-South Central							0.225	2.477
Mountain			0.151	2.811	0.167	3.189		
Pacific	-0.343	-3.247	-0.475	-5.131	-0.457	-5.113	-0.431	-3.911
Natural logarithm of total duration allocated to the alternative	0.479	20.228	0.380	17.540	0.390	15.237	0.337	12.257
Sample size	2621	4549	4418	3180				
Log-likelihood value at convergence	-4016.05	-5758.72	-5423.76	-3926.25				

The aforementioned explanatory variables (i.e., income, number of individuals in different age category, number of workers, vehicle ownership, total budget allocated to the tour, and location-indicator variables to capture any indirect location-specific effect) have an intuitive effect on remaining nonbusiness-quarter combinations. However, vehicle ownership, in particular, has a mixed effect. An increase in vehicle ownership is associated with both an increase (sightseeing and personal business) and decrease (visit friends, relaxation, recreation, and entertainment) in tour frequency, respectively. Two possible explanations for this could be: 1) the dependency on the primary driver in the household; and 2) the difference in level of satiation associated with different levels of vehicle ownership. That is, the impact of an additional vehicle on the tour frequency will be relatively less for high-vehicle-ownership households as compared to low-vehicle-ownership households.

Nonbusiness Tour-Party Composition Model

Table 25 through Table 48 present the result for tour-party composition models for six nonbusiness purposes (i.e., visit friends, relaxation, sightseeing, recreation, entertainment, and personal business) and quarters (i.e., January–March, April–June, July–September, and October–December). In the current study, five alternative options were considered:

- One adult and no children.
- Two adults and no children.
- Three or more adults with or without children.
- One adult with children.
- Two adults with children.

The one-adult-and-no-children option is the base category for all purpose-quarter combinations. In the model specification, family composition (i.e., presence of children and number of adults in different age groups), total budget allocated to the tour, number of episodes, and number of workers in the household are considered as explanatory variables.

Table 25 presents the estimates for visit friends/relatives tour purpose during the first quarter (January–March). First, the presence of children in the household has a positive impact on party composition, which includes children as one of the group members. Second, the effect of number of adults (decomposed in various age groups) has a positive impact on all the nonsingle party types, reinforcing the notion that individuals in these households jointly participate in nonbusiness discretionary tours. The results indicate that households with more adults for a given age category are more likely to have a tour-party composition that includes at least one child and one or more adults.

Table 25: Tour-Party Composition Model (Visit friends/relatives: January–March)

Alternatives (base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.669	-16.126	-8.071	-25.188	-5.795	-30.911	-6.267	-31.039
Family composition								
Presence of children in the household	-0.510	-2.114			0.964	4.773		
# of individuals between 17 and 34 years old	0.737	12.677	1.984	18.832	1.976	30.623	2.360	34.639
# of individuals between 35 and 49 years old	0.488	9.939	1.912	23.239	1.557	30.075	1.872	32.26
# of individuals between 50 and 64 years old	0.950	19.791	2.161	23.759	1.327	19.325	1.730	25.92
# of individuals >= 65 years old	1.216	23.101	2.388	22.029	1.411	13.756	1.668	16.38
Natural logarithm of total duration	-0.142	-4.624	-0.142	-4.624	-0.142	-4.624	-0.287	-5.648
# of episodes			0.017	1.960	0.017	1.960	0.017	1.960
# of workers (full-time or part-time)	-0.060	-1.919	-0.060	-1.919			0.092	2.702
Sample size	7178							
Log-likelihood value at convergence	-6845.29							

Table 26: Tour-Party Composition Model (Visit friends/relatives: April–June)

<table>
<tr><th rowspan="2">Alternatives
(Base: One Adult, No Children)</th><th colspan="2">Two Adults, No Children</th><th colspan="2">Three or More Adults with and Without Children</th><th colspan="2">One Adult with Children</th><th colspan="2">Two Adults with Children</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.785</td><td>-20.484</td><td>-8.384</td><td>-30.143</td><td>-5.687</td><td>-39.748</td><td>-6.281</td><td>-37.615</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td>Presence of children in the household</td><td>-0.900</td><td>-3.723</td><td></td><td></td><td>1.320</td><td>9.367</td><td></td><td></td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td rowspan="2">0.603</td><td rowspan="2">14.211</td><td rowspan="2">2.108</td><td rowspan="2">29.838</td><td rowspan="2">1.746</td><td rowspan="2">36.668</td><td>2.196</td><td>39.811</td></tr>
<tr><td># of individuals between 35 and 49 years old</td><td rowspan="2">1.949</td><td rowspan="2">37.144</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td>1.069</td><td>26.056</td><td rowspan="2">2.414</td><td rowspan="2">30.422</td><td rowspan="2">1.395</td><td rowspan="2">24.200</td></tr>
<tr><td># of individuals >= 65 years old</td><td>1.273</td><td>29.146</td><td>1.429</td><td>13.439</td></tr>
<tr><td>Natural logarithm of total duration</td><td>-0.166</td><td>-6.502</td><td>-0.166</td><td>-6.502</td><td></td><td></td><td>-0.166</td><td>-6.502</td></tr>
<tr><td># of episodes</td><td>0.031</td><td>6.007</td><td>0.031</td><td>6.007</td><td></td><td></td><td></td><td></td></tr>
<tr><td># of workers (full-time or part-time)</td><td>-0.082</td><td>-3.169</td><td>-0.082</td><td>-3.169</td><td>-0.082</td><td>-3.169</td><td>0.059</td><td>1.743</td></tr>
<tr><td>Sample size</td><td colspan="8">10036</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="8">-9710.91</td></tr>
</table>

Table 27: Tour-Party Composition Model (Visit friends/relatives: July–September)

<table>
<tr><th rowspan="2">Alternatives
(Base: One Adult, No Children)</th><th colspan="2">Two Adults, No Children</th><th colspan="2">Three or More Adults with and Without Children</th><th colspan="2">One Adult with Children</th><th colspan="2">Two Adults with Children</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.701</td><td>-19.602</td><td>-8.722</td><td>-34.236</td><td>-6.057</td><td>-40.747</td><td>-6.659</td><td>-40.026</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td>Presence of children in the household</td><td>-0.844</td><td>-3.097</td><td></td><td></td><td>1.715</td><td>10.812</td><td>0.341</td><td>1.780</td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td rowspan="2">0.692</td><td rowspan="2">17.113</td><td rowspan="4">2.315</td><td rowspan="4">34.278</td><td rowspan="3">1.750</td><td rowspan="3">35.964</td><td rowspan="4">2.173</td><td rowspan="4">53.987</td></tr>
<tr><td># of individuals between 35 and 49 years old</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td rowspan="2">1.223</td><td rowspan="2">32.641</td></tr>
<tr><td># of individuals >= 65 years old</td><td></td><td></td></tr>
<tr><td>Natural logarithm of total duration</td><td>-0.246</td><td>-8.297</td><td></td><td></td><td>0.130</td><td>2.837</td><td>-0.152</td><td>-3.632</td></tr>
<tr><td># of episodes</td><td>0.050</td><td>3.768</td><td>0.050</td><td>3.768</td><td></td><td></td><td>0.050</td><td>3.768</td></tr>
<tr><td># of workers (full-time or part-time)</td><td>-0.156</td><td>-6.943</td><td>-0.156</td><td>-6.943</td><td>-0.053</td><td>-1.59</td><td></td><td></td></tr>
<tr><td>Sample size</td><td colspan="8">10772</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="8">-10248.48</td></tr>
</table>

Table 28: Tour-Party Composition Model (Visit friends/relatives: October–December)

<table>
<tr><th rowspan="2">Alternatives
(Base: One Adult, No Children)</th><th colspan="2">Two Adults, No Children</th><th colspan="2">Three or More Adults with and Without Children</th><th colspan="2">One Adult with Children</th><th colspan="2">Two Adults with Children</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-2.019</td><td>-21.748</td><td>-8.81</td><td>-34.038</td><td>-6.086</td><td>-36.027</td><td>-6.67</td><td>-40.455</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td>Presence of children in the household</td><td>-0.682</td><td>-2.593</td><td>-2.267</td><td>-2.244</td><td>1.365</td><td>7.512</td><td></td><td></td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td rowspan="2">0.759</td><td rowspan="2">21.25</td><td rowspan="3">2.349</td><td rowspan="3">33.837</td><td>1.766</td><td>29.444</td><td rowspan="3">2.249</td><td rowspan="3">44.715</td></tr>
<tr><td># of individuals between 35 and 49 years old</td><td rowspan="2">1.730</td><td rowspan="2">31.078</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td rowspan="2">1.319</td><td rowspan="2">36.942</td></tr>
<tr><td># of individuals >= 65 years old</td><td>2.797</td><td>31.311</td><td>1.151</td><td>9.531</td><td></td><td></td></tr>
<tr><td>Natural logarithm of total duration</td><td>-0.202</td><td>-6.438</td><td>-0.202</td><td>-6.438</td><td>-0.202</td><td>-6.438</td><td>-0.202</td><td>-6.438</td></tr>
<tr><td># of episodes</td><td>0.051</td><td>3.417</td><td>0.051</td><td>3.417</td><td>0.051</td><td>3.417</td><td>0.051</td><td>3.417</td></tr>
<tr><td># of workers (full-time or part-time)</td><td></td><td></td><td>0.125</td><td>4.044</td><td>0.125</td><td>4.044</td><td>0.125</td><td>4.044</td></tr>
<tr><td>Sample size</td><td colspan="8">9919</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="8">-9099.12</td></tr>
</table>

Table 29: Tour-Party Composition Model (Relaxation: January–March)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.340	-9.580	-7.440	-10.144	-4.943	-15.28	-5.744	-19.656
Family composition								
Presence of children in the household					0.771	2.004		
# of individuals greater than 17 years old	0.600	10.644	1.486	7.810	1.093	11.696	1.672	20.622
Natural logarithm of total duration	0.195	4.651	0.195	4.651				
# of episodes							0.047	2.962
# of workers (full-time or part-time)	-0.065	-1.665					-0.065	-1.665
Sample size	2570							
Log-likelihood value at convergence	-2606.93							

Table 30: Tour-Party Composition Model (Relaxation: April–June)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.362	-12.109	-7.601	-18.077	-5.999	-23.625	-5.649	-25.932
Family composition								
Presence of children in the household					1.170	4.666		
# of individuals between 17 and 64 years old	0.526	14.418	1.733	17.053	1.367	20.214	1.565	24.007
# of individuals >= 65 years old							0.333	1.579
Natural logarithm of total duration	0.236	6.630			0.236	6.630	0.236	6.630
# of episodes			0.033	1.014	0.033	1.014	0.033	1.014
# of workers (full-time or part-time)							0.103	2.630
Sample size	4263							
Log-likelihood value at convergence	-4451.20							

Table 31: Tour-Party Composition Model (Relaxation: July–September)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.584	-15.832	-9.072	-28.369	-6.905	-31.181	-7.255	-36.013
Family composition								
Presence of children in the household					1.479	6.839	0.527	2.744
# of individuals greater than 17 years old	0.832	20.844	2.247	26.237	1.909	31.385	2.279	39.021
Natural logarithm of total duration			0.331	6.747	0.331	6.747	0.331	6.747
# of episodes	0.057	3.659					-0.087	-3.587
# of workers (full-time or part-time)			0.214	6.795			0.214	6.795
Sample size	5620							
Log-likelihood value at convergence	-5759.25							

Table 32: Tour-Party Composition Model (Relaxation: October–December)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.431	-9.268	-7.974	-12.758	-6.502	-17.155	-6.215	-19.151
Family composition								
Presence of children in the household								
# of individuals greater than 17 years old	0.780	12.488	1.697	10.267	1.606	14.699	1.675	17.931
Natural logarithm of total duration	0.226	3.867	0.226	3.867	0.226	3.867	0.226	3.867
# of episodes					0.065	1.324		
# of workers (full-time or part-time)	-0.096	-2.065	0.191	3.100	-0.096	-2.065	0.191	3.100
Sample size	2264							
Log-likelihood value at convergence	-2247.72							

Table 33: Tour-Party Composition Model (Sightseeing: January–March)

<table>
<tr><th rowspan="2">Alternatives
(Base: One Adult, No Children)</th><th colspan="2">Two Adults, No Children</th><th colspan="2">Three or More Adults with and Without Children</th><th colspan="2">One Adult with Children</th><th colspan="2">Two Adults with Children</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.195</td><td>-5.198</td><td>-6.755</td><td>-5.595</td><td>-6.205</td><td>-8.21</td><td>-5.422</td><td>-11.339</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td>Presence of children in the household</td><td></td><td></td><td></td><td></td><td>0.916</td><td>1.813</td><td>0.916</td><td>1.813</td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td rowspan="2">0.654</td><td rowspan="2">7.272</td><td></td><td></td><td rowspan="2">1.628</td><td rowspan="2">8.064</td><td rowspan="2">1.782</td><td rowspan="2">14.587</td></tr>
<tr><td># of individuals greater than 35 years old</td><td>1.724</td><td>5.758</td></tr>
<tr><td>Natural logarithm of total duration</td><td></td><td></td><td>-0.525</td><td>-2.345</td><td>-0.525</td><td>-2.345</td><td></td><td></td></tr>
<tr><td># of episodes</td><td></td><td></td><td></td><td></td><td>0.357</td><td>2.394</td><td></td><td></td></tr>
<tr><td># of workers (full-time or part-time)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Sample size</td><td colspan="8">684</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="8">-676.34</td></tr>
</table>

Table 34: Tour-Party Composition Model (Sightseeing: April–June)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.802	-9.190	-8.288	-14.257	-5.644	-13.268	-6.981	-18.778
Family composition								
Presence of children in the household					1.368	3.303		
# of individuals greater than 17 years old	0.775	11.068	2.189	14.868	1.491	12.271	2.096	18.882
Natural logarithm of total duration					-0.238	-1.762		
# of episodes	0.136	1.814	0.136	1.814	0.136	1.814		
# of workers (full-time or part-time)					0.222	3.672	0.222	3.672
Sample size	1637							
Log-likelihood value at convergence	-1649.55							

Table 35: Tour-Party Composition Model (Sightseeing: July–September)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.355	-8.683	-8.503	-19.406	-6.289	-19.459	-6.823	-22.054
Family composition								
Presence of children in the household								
# of individuals greater than 17 years old	0.856	14.789	2.305	19.673	1.728	17.401	2.331	28.019
Natural logarithm of total duration	-0.089	-1.400	-0.089	-1.400	-0.089	-1.400	-0.089	-1.400
# of episodes							-0.137	-1.427
# of workers (full-time or part-time)			0.250	5.476	0.250	5.476	0.250	5.476
Sample size	2543							
Log-likelihood value at convergence	-2551.31							

Table 36: Tour-Party Composition Model (Sightseeing: October–December)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-2.186	-8.620	-8.948	-9.174	-5.827	-11.206	-6.73	-13.423
Family composition								
Presence of children in the household					3.282	6.564		
# of individuals greater than 17 years old	1.143	14.935	2.277	9.698	1.297	7.587	2.024	13.997
Natural logarithm of total duration					0.201	1.386	0.201	1.386
# of episodes	0.207	1.443	0.207	1.443				
# of workers (full-time or part-time)					0.321	3.450	0.321	3.450
Sample size	1167							
Log-likelihood value at convergence	-1042.20							

Table 37: Tour-Party Composition Model (Recreation: January–March)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.001	-5.092	-3.236	-15.113	-5.122	-11.715	-5.148	-13.366
Family composition								
Presence of children in the household					1.243	3.673		
# of individuals greater than 17 years old	0.190	2.484	-0.766	-2.992	1.008	9.588	1.447	14.808
Natural logarithm of total duration	0.261	3.587	0.261	3.587	-0.292	-2.461	-0.292	-2.461
# of episodes					0.169	3.847	0.169	3.847
# of workers (full-time or part-time)	-0.166	-3.272					-0.166	-3.272
Sample size	1698							
Log-likelihood value at convergence	-1790.80							

Table 38: Tour-Party Composition Model (Recreation: April–June)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.145	-12.806	-8.554	-13.412	-5.275	-17.18	-5.034	-18.042
Family composition								
Presence of children in the household					0.620	1.878		
# of individuals between 17 and 49 years old			1.415	10.339	1.111	15.673	1.245	18.357
# of individuals greater than 50 years old	0.343	8.224						
Natural logarithm of total duration			0.490	2.722	-0.142	-1.374		
# of episodes	0.157	4.399	0.157	4.399	0.157	4.399	0.157	4.399
# of workers (full-time or part-time)								
Sample size	2492							
Log-likelihood value at convergence	-2586.10							

Table 39: Tour-Party Composition Model (Recreation: July–September)

<table>
<tr><th rowspan="2">Alternatives
(Base: One Adult, No Children)</th><th colspan="2">Two Adults, No Children</th><th colspan="2">Three or More Adults with and Without Children</th><th colspan="2">One Adult with Children</th><th colspan="2">Two Adults with Children</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.103</td><td>-11.363</td><td>-7.543</td><td>-19.819</td><td>-5.246</td><td>-21.089</td><td>-5.265</td><td>-22.444</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td>Presence of children in the household</td><td></td><td></td><td></td><td></td><td>1.017</td><td>3.903</td><td></td><td></td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td>0.285</td><td>8.416</td><td></td><td></td><td rowspan="3">1.174</td><td rowspan="3">18.102</td><td rowspan="3">1.479</td><td rowspan="3">26.697</td></tr>
<tr><td># of individuals between 35 and 49 years old</td><td></td><td></td><td rowspan="2">1.523</td><td rowspan="2">17.319</td></tr>
<tr><td># of individuals greater than 50 years old</td><td>0.285</td><td>8.416</td></tr>
<tr><td>Natural logarithm of total duration</td><td>0.235</td><td>5.030</td><td>0.235</td><td>5.030</td><td></td><td></td><td>0.235</td><td>5.030</td></tr>
<tr><td># of episodes</td><td>0.095</td><td>3.517</td><td>0.095</td><td>3.517</td><td>0.095</td><td>3.517</td><td></td><td></td></tr>
<tr><td># of workers (full-time or part-time)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Sample size</td><td colspan="8">3449</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="8">-3612.00</td></tr>
</table>

Table 40: Tour-Party Composition Model (Recreation: October–December)

<table>
<tr><th rowspan="2">Alternatives
(Base: One Adult, No Children)</th><th colspan="2">Two Adults, No Children</th><th colspan="2">Three or More Adults with and Without Children</th><th colspan="2">One Adult with Children</th><th colspan="2">Two Adults with Children</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.462</td><td>-14.298</td><td>-7.539</td><td>-10.201</td><td>-5.551</td><td>-12.087</td><td>-4.335</td><td>-10.912</td></tr>
<tr><td>Family composition</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Presence of children in the household</td><td></td><td></td><td></td><td></td><td>1.412</td><td>3.573</td><td></td><td></td></tr>
<tr><td># of individuals between 17 and 49 years old</td><td></td><td></td><td rowspan="2">1.231</td><td rowspan="2">7.320</td><td rowspan="2">0.875</td><td rowspan="2">5.473</td><td rowspan="2">0.962</td><td rowspan="2">10.345</td></tr>
<tr><td># of individuals greater than 50 years old</td><td>0.413</td><td>8.302</td></tr>
<tr><td>Natural logarithm of total duration</td><td></td><td></td><td></td><td></td><td>-0.310</td><td>-3.014</td><td>-0.310</td><td>-3.014</td></tr>
<tr><td># of episodes</td><td>0.089</td><td>2.438</td><td></td><td></td><td>0.089</td><td>2.438</td><td>0.089</td><td>2.438</td></tr>
<tr><td># of workers (full-time or part-time)</td><td></td><td></td><td></td><td></td><td>0.239</td><td>1.992</td><td></td><td></td></tr>
<tr><td>Sample size</td><td colspan="8">1650</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="8">-1613.83</td></tr>
</table>

Table 41: Tour-Party Composition Model (Entertainment: January–March)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-0.855	-5.358	-6.676	-8.81	-5.095	-10.878	-6.076	-11.972
Family composition								
Presence of children in the household					1.031	2.240		
# of individuals greater than 17 years old	0.346	5.703	1.349	7.107	1.088	8.587	1.398	10.630
Natural logarithm of total duration							0.409	3.245
# of episodes								
# of workers (full-time or part-time)							0.110	1.376
Sample size	1353							
Log-likelihood value at convergence	-1412.98							

Table 42: Tour-Party Composition Model (Entertainment: April–June)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-0.923	-7.059	-7.079	-11.077	-4.995	-13.652	-5.198	-15.92
Family composition								
Presence of children in the household					1.630	4.985	0.726	2.098
# of individuals greater than 17 years old	0.304	6.195	1.341	8.703	0.899	6.142	1.269	17.360
Natural logarithm of total duration							0.350	3.603
# of episodes								
# of workers (full-time or part-time)					0.184	1.641		
Sample size	2039							
Log-likelihood value at convergence	-2130.95							

Table 43: Tour-Party Composition Model (Entertainment: July–September)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.102	-7.42	-7.547	-13.859	-5.112	-16.791	-5.136	-18.626
Family composition								
Presence of children in the household					0.755	3.275	0.755	3.275
# of individuals greater than 17 years old	0.325	6.328	1.588	12.290	1.198	14.977	1.440	22.068
Natural logarithm of total duration	0.231	3.516	0.231	3.516	0.231	3.516	0.231	3.516
# of episodes								
# of workers (full-time or part-time)								
Sample size	2283							
Log-likelihood value at convergence	-2443.84							

Table 44: Tour-Party Composition Model (Entertainment: October–December)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.405	-8.18	-6.957	-10.231	-5.31	-11.357	-5.47	-14.154
Family composition								
Presence of children in the household					1.826	4.672		
# of individuals greater than 17 years old	0.499	9.019	1.347	7.852	0.756	2.445	1.433	15.622
Natural logarithm of total duration	0.290	3.703	0.290	3.703			0.290	3.703
# of episodes								
# of workers (full-time or part-time)					0.414	1.601		
Sample size	1591							
Log-likelihood value at convergence	-1598.34							

Table 45: Tour-Party Composition Model (Personal Business: January–March)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-0.937	-6.981	-7.306	-13.163	-4.854	-17.963	-5.356	-18.738
Family composition								
Presence of children in the household					1.135	3.797		
# of individuals greater than 17 years old	0.518	8.999	1.565	11.430	0.974	13.172	1.357	19.627
Natural logarithm of total duration	-0.143	-3.398					-0.143	-3.398
# of episodes	0.085	4.586	0.085	4.586	0.085	4.586	0.085	4.586
# of workers (full-time or part-time)	-0.240	-6.618	-0.240	-6.618				
Sample size	2621							
Log-likelihood value at convergence	-2730.00							

Table 46: Tour-Party Composition Model (Personal Business: April–June)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.265	-11.820	-6.893	-19.47	-5.199	-23.022	-5.842	-24.882
Family composition								
Presence of children in the household					1.265	6.131		
# of individuals greater than 17 years old	0.672	15.229	1.732	18.484	1.197	20.279	1.537	27.005
Natural logarithm of total duration	-0.140	-3.934	-0.140	-3.934	-0.140	-3.934	-0.140	-3.934
# of episodes	0.182	8.552	0.182	8.552	0.182	8.552	0.182	8.552
# of workers (full-time or part-time)	-0.297	-10.75	-0.297	-10.75				
Sample size	4549							
Log-likelihood value at convergence	-4819.28							

Table 47: Tour-Party Composition Model (Personal Business: July–September)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.271	-11.676	-7.197	-20.095	-6.546	-27.001	-6.435	-26.249
Family composition								
Presence of children in the household					1.662	7.253	0.490	1.894
# of individuals greater than 17 years old	0.563	14.230	1.755	18.775	1.678	25.831	1.716	23.622
Natural logarithm of total duration	-0.086	-2.448					-0.086	-2.448
# of episodes	0.063	4.461	0.063	4.461	0.063	4.461	0.063	4.461
# of workers (full-time or part-time)							0.205	4.900
Sample size	4418							
Log-likelihood value at convergence	-4667.34							

Table 48: Tour-Party Composition Model (Personal Business: October–December)

Alternatives (Base: One Adult, No Children)	Two Adults, No Children		Three or More Adults with and Without Children		One Adult with Children		Two Adults with Children	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.234	-9.925	-7.547	-16.634	-6.173	-21.196	-5.984	-21.961
Family composition								
Presence of children in the household					1.528	6.070		
# of individuals greater than 17 years old	0.665	11.791	1.826	15.676	1.610	19.283	1.797	23.608
Natural logarithm of total duration	-0.076	-1.849					-0.076	-1.849
# of episodes	0.117	4.859	0.117	4.859	0.117	4.859	0.117	4.859
# of workers (full-time or part-time)	-0.284	-7.576	-0.284	-7.576	-0.284	-7.576	-0.284	-7.576
Sample size	3180							
Log-likelihood value at convergence	-3340.06							

The total budget of the tour also has a negative impact on nonsingle party types, suggesting that nonsingle households may spend less time visiting friends, relative to single-person households, and allocate more time to activities such as relaxation, recreation, and entertainment (see effect of budget on tour-party composition for relaxation, recreation, and entertainment purposes [Table 29–Table 32; Table 37–Table 40; and Table 41–Table 44]) to spend more time with their family. This effect is manifested through a higher likelihood of a single-person tour for households that allocate a high budget for the visit friends/relatives tour purpose. The number of episodes also has a positive impact on nonsingle party types, which include at least one child, suggesting that households who visit their friends/relatives quite frequently prefer going with their family members or, perhaps, accompany friends. Finally, households with more workers have a higher likelihood of having a tour-party composition of two adults with children than households with more nonworkers while visiting their friends/relatives.

The aforementioned explanatory variables (i.e., number of individuals in different age categories, number of workers, total budget allocated to the tour, and number of episodes) intuitively affect other nonbusiness-quarter combinations. The variables' effects on various tour-party compositions can be interpreted similarly to the methods discussed previously. Households that allocate a large budget to relaxation, recreation, and entertainment purposes choose party composition with one or more adults with children. The households with more workers relative to nonworkers have a high likelihood of choosing a party composition that includes children.

Business Travel Model Estimation Results

Decision to Make Business Tours and Annual Budget Model

A sample selection model was estimated using the MLE technique to determine the household's decision to conduct business travel and its annual duration. The results of the first component (decision to conduct business travel) are provided in Table 49. As mentioned, the alternative-specific constants do not have any tangible meaning in this context, but simply represent the effect of all unobserved factors (explanatory variables not considered in the specification).

With an increase in household income, the propensity to make business tours increases. This result is intuitive, as household income level can be viewed as a proxy for job status. That is, it can be assumed that individuals belonging to a high-income household generally have high-status white collar jobs, for which long-distance business meetings are more common. For the family composition and working-status variables, the coefficient on family composition provides the effect of nonworkers in the household and the coefficient on working status provides the differential impact between a nonworker and a worker in the household. The results indicate that a household with more workers is more likely to engage in business travel as compared to households with fewer workers—an intuitive result. Finally, the location-specific variables were added in the model to capture any location-specific effects, which are not directly controlled for in the model. This variable may capture the effect of the presence of various employment opportunities that may exist in the region and are not directly included in the model.

Table 49: Business-Tour Model (Decision to Make Business Travel)

Variables	Coeff.	T-Stat
Alternative-Specific Constant	-0.375	-15.425
Income (base: 25K-49K)		
Less than 25K	-0.374	-19.596
50K-74K	0.291	16.137
75K-99K	0.606	21.769
100k and more	0.910	27.627
Family composition		
Presence of Children (less than 17 years old)	-0.263	-7.643
# of individuals between 17 and 49 years old	-0.023	-2.685
# of individuals between 50 and 64 years old	-0.070	-6.752
# of individuals >= 65 years old	-0.197	-18.147
Working status		
# of workers (full-time or part-time)	0.110	14.764
Household residential location (base: Mountain)		
New England	-0.328	-13.033
Atlantic	-0.416	-12.391
East-North Central	-0.337	-11.805
West-North Central	-0.122	-4.858
South Atlantic	-0.165	-6.920
East-South Central	-0.184	-6.345
West-South Central	-0.103	-3.326
Pacific	-0.097	-3.554

For the annual duration model (Table 50), the duration variable (indicating the number of nights away from home) was transformed on a logarithmic scale to avoid prediction of negative values. First, as noted, the alternative-specific constants do not have any tangible meaning. Second, with the increase in household income, the annual budget for business travel increases. This result corroborates the finding discussed earlier: high-income households are generally employed in white collar jobs that commonly require business travel. Third, the effect of the number of workers (full- or part-time) and nonworking adults is intuitive: the total business budget increases with the increase in the number of workers relative to increase in number of nonworkers. Finally, the lack of a vehicle in the household has a positive impact on the business-tour duration. This result likely arose because workers in households with high vehicle ownership may take their own vehicles on business tours, which allows them to minimize their travel time for relatively short business tours (300–500 miles). Finally, the variance-covariance matrix (t-statistics in parenthesis) is presented in Equation 4:

Equation 4: Business Annual Budget Model Variance-Covariance Matrix

$$\begin{bmatrix} 1.00(\mathit{fixed}) & -0.23(-4.17) \\ & 1.43(44.15) \end{bmatrix}$$

The results indicate a significant correlation between two components (decision to conduct business travel and its duration) of the model, suggesting that some common unobserved factors impact the decision and duration in the opposite direction.

Table 50: Annual Business Budget Model

Variables	Coeff.	T-Stat
Constant	2.268	24.319
Income (base: less than 50K)		
50K-74K	0.169	5.556
75K-99K	0.348	7.491
100k and more	0.449	8.000
Working status		
# of workers (full-time or part-time)	-0.065	-6.562
# of nonworking adults	-0.081	-5.655
Vehicle ownership (base: one or more vehicles)		
No vehicle	0.137	3.920
Note: The dependent variable (annual budget) is transformed on logarithmic scale		

Table 50 Model Fit:

Sample Size: 47,931 households. Log-likelihood value at convergence: -21821.10.

Business-Tour Participation and Duration Model

Table 51 presents the estimation results for business tour participation and duration. As mentioned earlier, the alternative-specific constants do not have any tangible meaning. Among the set of explanatory variables considered in the specification, the first is presence of children (less than 17 years old) in the household. The results indicate that, in general, households with children prefer to avoid business-related tours. However, they may prefer to schedule business-related travel by the time of year: business-only in winter, spring, and summer; business/pleasure in winter; and convention/conference/seminar travel in winter and summer. This is intuitive, as such households are likely to have childcare responsibilities that may discourage them from making long-distance/duration business tours. The interesting point to observe here is that households with children do not prefer any kind of business-related tours during fall (October to December), perhaps because they want to spend holidays with their family.

In general, the coefficients on family composition and working-status variables indicate a higher likelihood of participating in pure business tours (business and conferences) for the households with more workers. On the other hand, households with relatively more nonworkers are more inclined toward business/pleasure tour purposes. Also, baby boomers and empty nesters (50–64 years of age) and seniors (above 65 years of age) prefer the business/pleasure combination, indicating that workers older than 50 may like to take their spouses on business tours.

Table 51: Tour Duration and Participation Model (Business Purposes)

<table>
<tr><th rowspan="4">Alternatives
(Base : Business (Q1))</th><th colspan="12">Explanatory Variables</th></tr>
<tr><th colspan="2" rowspan="2">Alternative-Specific Constant</th><th colspan="10">Family Composition</th></tr>
<tr><th colspan="2">Presence of children (< 17 years old)</th><th colspan="2"># of individuals between 17–34 years old</th><th colspan="2"># of individuals between 35–49 years old</th><th colspan="2"># of individuals between 50–64 years old</th><th colspan="2"># of individuals >= 65 years old</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Business (Q2)</td><td>0.187</td><td>6.618</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business (Q3)</td><td>-0.040</td><td>-1.425</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business (Q4)</td><td>-0.246</td><td>-6.125</td><td>-0.253</td><td>-2.362</td><td>-0.035</td><td>-2.219</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business/Pleasure (Q1)</td><td>-1.666</td><td>-29.109</td><td></td><td></td><td rowspan="2">-0.129</td><td rowspan="2">-6.265</td><td rowspan="2">-0.077</td><td rowspan="2">-5.228</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business/Pleasure (Q2)</td><td>-1.347</td><td>-26.043</td><td rowspan="3">-0.444</td><td rowspan="3">-3.380</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business/Pleasure (Q3)</td><td>-1.804</td><td>-30.628</td><td></td><td></td><td></td><td></td><td>0.143</td><td>5.092</td><td rowspan="2">0.322</td><td rowspan="2">9.404</td></tr>
<tr><td>Business/Pleasure (Q4)</td><td>-1.84</td><td>-31.696</td><td>-0.129</td><td>-6.265</td><td>-0.077</td><td>-5.228</td><td></td><td></td></tr>
<tr><td>Convention/Conference /Seminar (Q1)</td><td>-2.398</td><td>-31.892</td><td></td><td></td><td></td><td></td><td rowspan="2">0.055</td><td rowspan="2">3.556</td><td rowspan="3">0.171</td><td rowspan="3">9.141</td><td rowspan="3">0.505</td><td rowspan="3">21.082</td></tr>
<tr><td>Convention/Conference /Seminar (Q2)</td><td>-1.862</td><td>-32.953</td><td>-0.254</td><td>-1.632</td><td></td><td></td></tr>
<tr><td>Convention/Conference /Seminar (Q3)</td><td>-1.859</td><td>-35.441</td><td></td><td></td><td rowspan="2">-0.189</td><td rowspan="2">-6.392</td><td></td><td></td></tr>
<tr><td>Convention/Conference /Seminar (Q4)</td><td>-2.203</td><td>-23.365</td><td>-0.254</td><td>-1.632</td><td>-0.115</td><td>-3.516</td><td></td><td></td><td>0.388</td><td>6.232</td></tr>
</table>

Q1: January-March (winter), Q2: April-June (spring), Q3: July-September (summer), Q4: October-December (fall)

Table 51 (cont.): Tour Duration and Participation Model (Business Purposes)

<table>
<tr><th rowspan="4">Alternatives
(Base : Business (Q1))</th><th colspan="8">Explanatory Variables</th></tr>
<tr><th colspan="4">Working Status</th><th colspan="4">Household Residential Location (Mountain is the base category)</th></tr>
<tr><th colspan="2"># of full-time workers</th><th colspan="2"># of part-time workers</th><th colspan="2">England</th><th colspan="2">M-Atlantic</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Business (Q2)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business (Q3)</td><td></td><td></td><td rowspan="2">0.035</td><td rowspan="2">1.753</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business (Q4)</td><td>0.024</td><td>1.957</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Business/Pleasure (Q1)</td><td rowspan="2">-0.040</td><td rowspan="2">-2.476</td><td></td><td></td><td rowspan="4">-0.165</td><td rowspan="4">-3.145</td><td rowspan="2">-0.161</td><td rowspan="2">-1.662</td></tr>
<tr><td>Business/Pleasure (Q2)</td><td></td><td></td></tr>
<tr><td>Business/Pleasure (Q3)</td><td></td><td></td><td></td><td></td><td>-0.464</td><td>-2.573</td></tr>
<tr><td>Business/Pleasure (Q4)</td><td>-0.040</td><td>-2.476</td><td></td><td></td><td>-0.161</td><td>-1.662</td></tr>
<tr><td>Convention/Conference /Seminar (Q1)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Convention/Conference /Seminar (Q2)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Convention/Conference /Seminar (Q3)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Convention/Conference /Seminar (Q4)</td><td>0.116</td><td>2.954</td><td>0.087</td><td>2.304</td><td></td><td></td><td></td><td></td></tr>
</table>

Table 51 (cont.): Tour Duration and Participation Model (Business Purposes)

Alternatives *(Base : Business (Q1))*	Explanatory Variables											
	Household Residential Location (Mountain is the base category)											
	EN-Central		WN-Central		S-Atlantic		ES-Central		WS-Central		Pacific	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Business (Q2)					0.083	2.041						
Business (Q3)							0.124	1.885	0.102	1.973		
Business (Q4)												
Business/Pleasure (Q1)					-0.200	-2.310			-0.189	-1.762	0.223	3.806
Business/Pleasure (Q2)							-0.110	-1.180				
Business/Pleasure (Q3)	-0.167	-1.753							-0.189	-1.762		
Business/Pleasure (Q4)					-0.200	-2.310	-0.110	-1.180			0.223	3.806
Convention/Conference /Seminar (Q1)	0.264	2.626	0.182	2.720	0.196	2.522	0.204	1.385			0.285	3.037
Convention/Conference /Seminar (Q2)												
Convention/Conference /Seminar (Q3)					0.196	2.522						
Convention/Conference /Seminar (Q4)	0.264	2.626					-0.314	-1.760			0.285	3.037

Table 51 (cont.): Tour Duration and Participation Model (Business Purposes)

Alternatives *(Base : Business (Q1))*	Explanatory Variables					
	Vehicle Ownership (base: three or more vehicles)					
	Zero Vehicle		One Vehicle		Two Vehicles	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Business (Q2)						
Business (Q3)	-0.089	-1.952				
Business (Q4)						
Business/Pleasure (Q1)	-0.214	-3.246	-0.229	-1.701		
Business/Pleasure (Q2)						
Business/Pleasure (Q3)			0.182	2.779	0.182	2.779
Business/Pleasure (Q4)						
Convention/Conference /Seminar (Q1)	-0.182	-1.972	-0.164	-1.847	-0.164	-1.847
Convention/Conference /Seminar (Q2)			0.126	2.432		
Convention/Conference /Seminar (Q3)						
Convention/Conference /Seminar (Q4)						

Table 51 (cont.): Tour Duration and Participation Model (Business Purposes)

Alternatives *(Base : Business (Q1))*	Explanatory Variables							
	Family Income (base: 25K-49K)							
	Less than 25K		50K-74K		75K-99K		100K and More	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Business (Q2)	0.179	2.762					-0.092	-1.732
Business (Q3)								
Business (Q4)								
Business/Pleasure (Q1)	0.440	6.119	-0.215	-5.478	-0.371	-7.162		
Business/Pleasure (Q2)							-0.371	-7.162
Business/Pleasure (Q3)								
Business/Pleasure (Q4)								
Convention/Conference /Seminar (Q1)	0.518	7.450	-0.525	-12.641	-0.795	-13.674	-0.824	-11.598
Convention/Conference /Seminar (Q2)								
Convention/Conference /Seminar (Q3)								
Convention/Conference /Seminar (Q4)								

Table 51 (cont.): Tour Duration and Participation Model (Business Purposes)

Alternatives	Translation Parameters	
	Coeff.	T-Stat
Business (Q1)	6.424	22.906
Business (Q2)	5.108	25.408
Business (Q3)	5.537	24.576
Business (Q4)	5.144	22.641
Business/Pleasure (Q1)	10.406	8.990
Business/Pleasure (Q2)	8.989	10.495
Business/Pleasure (Q3)	10.366	10.967
Business/Pleasure (Q4)	8.222	9.139
Convention/Conference/Seminar (Q1)	6.695	7.379
Convention/Conference/Seminar (Q2)	7.941	9.361
Convention/Conference/Seminar (Q3)	13.034	7.973
Convention/Conference/Seminar (Q4)	6.459	7.164

Table 51 Model Fit:

Sample Size: 14,664 households. Log-likelihood value at convergence: -79801.20.

Location-specific variables (i.e., New England, Middle Atlantic, East-North Central, West-North Central, South Atlantic, East-South Central, West-South Central, Mountain, and Pacific) were also added to capture the effect of location on different business purposes. This variable may capture the effect of various employment opportunities in a region that are not directly included in the model. For example, if a specific region has a greater percentage of the population employed in white collar jobs, then that region may see a higher number of business tours. The base category is Mountain; thus, a negative sign on the location variable for a particular nonbusiness purpose suggests that households in the Mountain region are more likely to participate in that particular nonbusiness purpose than are households from other regions. For example, for the business/pleasure (all four seasons) purpose, the coefficient corresponding to the New England region variable is negative (-0.175). This suggests that households in the Mountain region are more likely to take business/pleasure tours (all four seasons) than are households in the New England region. The other location-specific variables can be interpreted in a similar fashion.

Vehicle ownership is found to have a significant impact on participation in and duration of all business purposes. The results indicate that with an increase in vehicle ownership, the household is more likely to participate in various business-only purposes (see the increase in magnitude of vehicle ownership coefficient across the columns for all tour purposes) over the course of the year. This is understandable, as a high level of vehicle ownership may suggest an availability of personal vehicles for all eligible adults in the household, granting them increased mobility. Vehicle ownership can also be viewed as a proxy for income. That is, it can be assumed that individuals belonging to high-income households generally have high-status white collar jobs, for which long-distance business meetings are quite common. However, for the business/pleasure purpose in the third quarter, the households with one or two vehicles are more likely to participate than are households with three or more vehicles.

The signs on the household income coefficients suggest a negative impact of the increase in household income on business tours. However, the negative direction does not mean that high-income households are less likely to participate in various business tours than are low-income households. Instead, it indicates that business tours for such households may simply be shorter than for low-income households—while high-income individuals make more business tours, at the same time they are able to minimize their tour duration because they can afford faster modes of transportation.

Furthermore, the translation parameters can be viewed as a measure of satiation. A large value for the translation parameter for a certain nonbusiness purpose indicates less satiation; households may invest more time in that tour, even increasing the number of such episodes. For example, the translation parameter for business/pleasure is greater than that of the business-only purpose for all seasons and conference purposes for three seasons, indicating that people invest more time in business/pleasure tours than in business-only and conference-purpose tours. This result is understandable, as a business/pleasure tour provides an opportunity for the individual making the business tour to take family members along, or include relaxation in addition to the business meeting.

Business-Tour Frequency Model

Table 52 through Table 54 present the results for three business purposes (i.e., business, business/pleasure, and convention/conference/seminar). In the next paragraph, the results of the tour-frequency model for different business purposes are discussed.

Table 52 presents the estimates for the business purpose. The results indicate that high-income households are likely to make fewer business tours (except in the third quarter) relative to low-income households. This result is slightly surprising, as one might expect high-income households to make more business tours in all the quarters, relative to low-income households, due to their white collar job status, as noted. This is an area that warrants further evaluation with a more current dataset, which was not available in this project. The coefficients on family composition and working-status variables provides the impact of nonworkers and the differential impact between a nonworker and a worker on tour frequency. The results suggest an increase in tour frequency for worker-dominated households in the second quarter, while in other quarters the difference is not significant. The results also suggest that young and middle-aged households are likely to have a high tour frequency relative to baby boomers, empty nesters, and elderly dominated households. Vehicle ownership has a mixed effect. Increase in vehicle ownership is associated with both an increase (third quarter) and decrease (first, second, and fourth quarter) in tour frequency. Increases in tour frequency with an increase in vehicle ownership is intuitive, as people may take their own vehicle for short business tours (200–300 mile tour and 1–2 day duration) to reduce their journey time. On the other hand, decreases in tour frequency in the first, second, and fourth quarters for high-vehicle-ownership households could be due to unobserved corporate scheduling/decision factors. Location-specific variables were also included to capture any indirect effects (effect of various employment opportunities in a region that are not directly included in the model, which may increase or decrease tour frequency) that might be specific to a location. The base category is New England; a negative sign on a location variable suggests that households in that particular location are less likely to make such tours than other households from New England. Finally, as expected, the total budget allocated to the tour purpose has a positive impact on tour frequency.

All of the aforementioned explanatory variables (i.e., income, number of individuals in different age category, number of workers, vehicle ownership, total budget allocated to the tour, and location-indicator variables to capture any indirect location-specific effect) have intuitive effects for other business purpose quarter combinations. In particular, income and vehicle ownership each have a mixed effect for other business purposes, as noted. These mixed effects highlight the variation in tour making that exists across various quarters for a given tour purpose; thus, quarter-specific models (as developed here) should be preferred over the models that do not include time period as a modeling dimension.

Table 52: Tour-Frequency Model (Business Purposes)

Variables	Business (Q1)		Business (Q2)		Business (Q3)		Business (Q4)	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-Specific Constant	-0.451	-8.713	-0.445	-8.476	-0.396	-5.832	-0.634	-10.339
Income (base: less than 25K)								
25K – 49K								
50K – 74K	0.108	2.777	0.137	3.389	0.251	5.973	0.144	2.992
75K – 99K	0.153	3.281	0.212	4.458			0.273	4.613
100K and More	0.319	6.764	0.376	8.279	0.520	9.368	0.398	7.287
Family composition								
# of individuals between 17 and 34 years old	-0.037	-3.370	-0.023	-2.165	-0.062	-2.365	-0.045	-2.471
# of individuals between 35 and 49 years old					-0.033	-1.496		
# of individuals between 50 and 64 years old	-0.037	-3.370	-0.023	-2.165	-0.071	-2.779		
# of individuals >= 65 years old			-0.094	-2.600				
Working status								
# of full-time workers	0.036	3.410	0.043	4.102	0.064	3.270	0.053	4.270
Vehicle ownership (base: Three or more)								
Zero vehicle	0.133	2.731	0.149	2.920				
One vehicle					-0.165	-2.596		
Two vehicles					-0.066	-1.693	0.099	1.743
Household residential location (base: New England)								
Atlantic								
East-North Central	0.139	2.605	0.188	3.219	0.170	2.528	0.181	2.668
West-North Central					0.123	2.214	0.117	2.092
South Atlantic								
East-South Central	0.150	2.611	0.143	2.336	0.225	3.958	0.166	2.485
West-South Central	0.158	2.680	0.120	1.895	0.239	3.544	0.174	2.537
Mountain			0.107	2.632				
Pacific	-0.330	-5.821	-0.286	-4.979	-0.175	-2.686	-0.327	-4.570
Natural logarithm of total duration allocated to the alternative	0.659	43.027	0.549	33.140	0.505	28.199	0.599	30.107
Sample size	4464	5159	4472	3880				
Log-likelihood value at convergence	-8794.08	-9234.61	-7915.44	-6440.80				

Table 53: Tour-Frequency Model (Business Purposes)

<table>
<tr><th rowspan="2">Variables</th><th colspan="2">Business/Pleasure (Q1)</th><th colspan="2">Business/Pleasure (Q2)</th><th colspan="2">Business/Pleasure (Q3)</th><th colspan="2">Business/Pleasure (Q4)</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-Specific Constant</td><td>-1.687</td><td>-8.624</td><td>-1.705</td><td>-9.963</td><td>-1.758</td><td>-11.686</td><td>-1.849</td><td>-9.792</td></tr>
<tr><td colspan="9">Income (base: less than 50K)</td></tr>
<tr><td>50K – 74K</td><td></td><td></td><td>-0.207</td><td>-1.735</td><td rowspan="3">0.131</td><td rowspan="3">1.338</td><td rowspan="3">-0.358</td><td rowspan="3">-2.717</td></tr>
<tr><td>75K – 99K</td><td rowspan="2">-0.344</td><td rowspan="2">-2.986</td><td rowspan="2">-0.369</td><td rowspan="2">-2.913</td></tr>
<tr><td>100K and More</td></tr>
<tr><td colspan="9">Family composition</td></tr>
<tr><td># of individuals between 17 and 34 years old</td><td>0.359</td><td>6.025</td><td>0.417</td><td>6.116</td><td rowspan="2">0.405</td><td rowspan="2">14.066</td><td>0.485</td><td>8.394</td></tr>
<tr><td># of individuals between 35 and 49 years old</td><td>0.274</td><td>7.078</td><td>0.317</td><td>6.033</td><td>0.408</td><td>10.122</td></tr>
<tr><td># of individuals between 50 and 64 years old</td><td>0.203</td><td>3.368</td><td>0.244</td><td>3.472</td><td rowspan="2">0.310</td><td rowspan="2">7.209</td><td>0.288</td><td>3.254</td></tr>
<tr><td># of individuals >= 65 years old</td><td></td><td></td><td>0.311</td><td>3.137</td><td></td><td></td></tr>
<tr><td colspan="9">Working status</td></tr>
<tr><td># of full-time workers</td><td></td><td></td><td>0.050</td><td>1.221</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Vehicle ownership (base: Two or more)</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Zero vehicle</td><td>-0.357</td><td>-1.656</td><td></td><td></td><td rowspan="2">0.162</td><td rowspan="2">1.514</td><td></td><td></td></tr>
<tr><td>One vehicle</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="9">Household residential location (base: New England)</td></tr>
<tr><td>Atlantic</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>East-North Central</td><td></td><td></td><td></td><td></td><td>0.290</td><td>1.369</td><td></td><td></td></tr>
<tr><td>West-North Central</td><td></td><td></td><td></td><td></td><td>0.130</td><td>1.115</td><td>0.335</td><td>1.732</td></tr>
<tr><td>South Atlantic</td><td></td><td></td><td>-0.232</td><td>-1.536</td><td></td><td></td><td></td><td></td></tr>
<tr><td>East-South Central</td><td>0.412</td><td>2.166</td><td></td><td></td><td>0.166</td><td>1.340</td><td></td><td></td></tr>
<tr><td>West-South Central</td><td>0.502</td><td>2.617</td><td>0.240</td><td>1.553</td><td></td><td></td><td>0.376</td><td>2.246</td></tr>
<tr><td>Mountain</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Pacific</td><td></td><td></td><td></td><td></td><td></td><td></td><td>-0.625</td><td>-2.355</td></tr>
<tr><td>Natural Logarithm of Total Duration Allocated to the Alternative</td><td>0.771</td><td>11.862</td><td>0.592</td><td>12.453</td><td>0.312</td><td>6.985</td><td>0.501</td><td>6.823</td></tr>
<tr><td>Sample size</td><td colspan="2">761</td><td colspan="2">1050</td><td colspan="2">1016</td><td colspan="2">713</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="2">-1034.96</td><td colspan="2">-1207.50</td><td colspan="2">-1127.76</td><td colspan="2">-691.61</td></tr>
</table>

Table 54: Tour-Frequency Model (Business Purposes)

Variables	Convention/Conference/ Seminar (Q1)		Convention/Conference/ Seminar (Q2)		Convention/Conference/ Seminar (Q3)		Convention/Conference/ Seminar (Q4)	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.555	-10.143	-2.165	-11.890	-3.299	-10.732	-1.846	-7.042
Income (base: less than 50K)								
50K – 74K	-0.416	-2.354	-0.241	-1.748				
75K – 99K	-0.353	-1.659			-0.243	-1.209	-0.546	-1.729
100K and more								
Family composition								
# of individuals between 17 and 34 years old	-0.221	-2.231	0.273	5.322	0.524	7.455	0.143	1.704
# of individuals between 35 and 49 years old					0.418	7.628	0.249	5.042
# of individuals between 50 and 64 years old			0.177	1.952	0.496	6.352		
# of individuals >= 65 years old								
Working status								
# of full-time workers	0.170	3.237						
Vehicle ownership (base: Three or more)								
Zero vehicle	0.542	3.131						
One vehicle					0.300	1.742		
Two vehicles							0.242	1.266
Household residential location (base: New England)								
Atlantic			-1.233	-2.375				
East-North Central					0.815	3.126		
West-North Central	-0.357	-1.296						
South Atlantic					0.337	1.781		
East-South Central								
West-South Central	-0.340	-1.113	-0.516	-1.639				
Mountain					0.462	2.425		
Pacific							-0.697	-1.919
Natural logarithm of total duration allocated to the alternative	0.740	15.829	0.824	26.766	0.714	9.070	0.536	3.513
Sample size	591	982	883	586				
Log-likelihood value at convergence	-455.07	-697.22	-609.27	-334.02				

Business Tour-Party-Composition Model

Table 55 through Table 63 present the results for tour-party composition models for three business purposes (business, business/pleasure, and convention/conference/seminar) and quarters (January–March, April–June, July–September, and October–December). The current study considered three alternative options, including:

- Single person;
- Two people; and
- Three or more people.

The one-person option is the base category for all the purpose-quarter combinations. The model specification considered family composition (number of workers), vehicle ownership, total budget allocated to the tour, number of episodes, and number of workers in the household as explanatory variables. Further, for the convention/conference/seminar purpose (all quarters), the two-person and three-or-more-persons categories were combined into one category (two or more persons) due to an insufficient number of observations corresponding to the three-or-more-persons category (Table 63).

Table 55 presents the estimates for business-tour purpose during the first quarter (January–March). The results indicate that households with more full-time workers are likely to travel in a group of three or more people for their business tours; however, households with more part-time workers are likely to travel in a group of two. The positive impact of number of workers on tour-party size is unclear. In general, for business-only tours, the party size may depend on workers' job descriptions (lower-level employee or higher-level employee) and the nature of meeting (sales meeting vs. product design discussion, etc.). Since such information is unavailable for individuals in the current data sample, such effects can only be captured indirectly through working status (full- or part-time) of the worker and vehicle ownership of the household (especially for combined business and pleasure purpose).

The results also indicate that a worker belonging to a high-vehicle-ownership household is likely to travel in a group of two, since it allows the individual to take his/her own private vehicle for the tour without worrying about the travel needs of other individuals in the household. Overall, the total budget of the tour has a negative impact (see sign on the "natural logarithmic of total duration" variable in Table 55 to Table 59) on nonsingle party types, suggesting that individuals travel alone for long-duration business tours, which is not surprising. The number of episodes also has a negative impact on nonsingle party types, suggesting that frequent tour makers (possibly sales agents, business representative, etc.) may also travel alone. The effect of explanatory variables on the tour-party size for other business purposes and quarter combinations can be interpreted similarly, as noted previously. However, caution must be exercised while making inferences. This is because the current model specifications do not include key components (e.g., nature of workers' jobs, corporate decision-making process, etc.), which are the real drivers of business-tour-party size.

Table 55: Tour Size Model (Business: January–March)

<table>
<tr><th rowspan="2">Alternatives
(base: Single Person)</th><th colspan="2">Two People</th><th colspan="2">Three or more People</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.675</td><td>-19.254</td><td>-4.022</td><td>-19.115</td></tr>
<tr><td colspan="5">Working status</td></tr>
<tr><td># of full-time workers</td><td></td><td></td><td>0.383</td><td>7.384</td></tr>
<tr><td># of part-time workers</td><td>0.177</td><td>3.041</td><td></td><td></td></tr>
<tr><td colspan="5">Vehicle ownership (base: One or no vehicle)</td></tr>
<tr><td>Two vehicles</td><td></td><td></td><td rowspan="2">-0.540</td><td rowspan="2">-2.670</td></tr>
<tr><td>Three or more vehicles</td><td>0.179</td><td>1.955</td></tr>
<tr><td>Natural logarithm of total duration</td><td>-0.103</td><td>-2.097</td><td></td><td></td></tr>
<tr><td># of episodes</td><td>-0.032</td><td>-2.78</td><td></td><td></td></tr>
<tr><td>Sample size</td><td colspan="4">4464</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="4">-2330.64</td></tr>
</table>

Table 56: Tour Size Model (Business: April–June)

<table>
<tr><th rowspan="2">Alternatives
(base: Single Person)</th><th colspan="2">Two People</th><th colspan="2">Three or more People</th></tr>
<tr><th>Coeff.</th><th>T-Stat</th><th>Coeff.</th><th>T-Stat</th></tr>
<tr><td>Alternative-specific constant</td><td>-1.588</td><td>-13.931</td><td>-4.099</td><td>-21.408</td></tr>
<tr><td colspan="5">Working status</td></tr>
<tr><td># of full-time workers</td><td>-0.080</td><td>-2.976</td><td rowspan="2">0.384</td><td rowspan="2">8.139</td></tr>
<tr><td># of part-time workers</td><td>0.105</td><td>1.902</td></tr>
<tr><td colspan="5">Vehicle ownership (base: One or no vehicle)</td></tr>
<tr><td>Two vehicles</td><td rowspan="2">0.321</td><td rowspan="2">3.344</td><td></td><td></td></tr>
<tr><td>Three or more vehicles</td><td>-0.283</td><td>-1.650</td></tr>
<tr><td>Natural logarithm of total duration</td><td>-0.101</td><td>-2.487</td><td>-0.101</td><td>-2.487</td></tr>
<tr><td># of episodes</td><td>-0.043</td><td>-2.938</td><td></td><td></td></tr>
<tr><td>Sample size</td><td colspan="4">5159</td></tr>
<tr><td>Log-likelihood value at convergence</td><td colspan="4">-2796.66</td></tr>
</table>

Table 57: Tour Size Model (Business: July–September)

Alternatives (base: Single Person)	Two People		Three or more People	
	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.863	-16.592	-5.231	-18.341
Working status				
# of full-time workers	-0.110	-4.040	0.580	10.271
# of part-time workers				
Vehicle ownership (base: One or no vehicle)				
Two vehicles	0.616	5.567		
Three or more vehicles				
Natural logarithm of total duration			0.121	1.404
# of episodes	-0.050	-2.965		
Sample size	4472			
Log-likelihood value at convergence	-2377.58			

Table 58: Tour Size Model (Business: October–December)

Alternatives (base: Single Person)	Two People		Three or more People	
	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific con: stant	-1.724	-12.189	-4.704	-16.601
Working status				
# of full-time workers	-0.067	-2.238	0.477	6.678
# of part-time workers				
Vehicle ownership (base: One or no vehicle)				
Two vehicles	0.411	3.530		
Three or more vehicles				
Natural logarithm of total duration	-0.099	-1.811	-0.099	-1.811
# of episodes	-0.050	-2.296	-0.050	-2.296
Sample size	3880			
Log-likelihood value at convergence	-1949.50			

Table 59: Tour Size Model (Business/Pleasure: January–March)

Alternatives (base: Single Person)	Two People		Three or more People	
	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-2.001	-7.413	-6.098	-10.891
Working status				
# of workers (full-time or part-time)	0.567	6.971	1.349	11.558
# of nonworking adults	0.719	6.005	1.492	9.279
Vehicle ownership (base: One or no vehicle)				
Two vehicles	1.018	4.608	1.573	3.975
Three or more vehicles				
Natural logarithm of total duration				
# of episodes				
Sample size	761			
Log-likelihood value at convergence	-647.70			

Table 60: Tour Size Model (Business/Pleasure: April–June)

Alternatives (base: Single Person)	Two People		Three or more People	
	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.521	-6.244	-6.126	-13.354
Working status				
# of workers (full-time or part-time)	0.523	7.608	1.512	13.37
# of nonworking adults	0.741	7.293	1.730	11.361
Vehicle ownership (base: One or no vehicle)				
Two vehicles	0.737	4.454	0.737	4.454
Three or more vehicles				
Natural logarithm of total duration	-0.205	-2.568		
# of episodes				
Sample size	1050			
Log-likelihood value at convergence	-868.18			

Table 61: Tour Size Model (Business/Pleasure: July–September)

Alternatives (base: Single Person)	Two People		Three or more People	
	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.691	-7.421	-4.195	-14.741
Working status				
# of workers (full-time or part-time)	0.349	5.225	1.173	15.497
# of nonworking adults	0.482	5.022	1.264	13.009
Vehicle ownership (base: One or no vehicle)				
Two vehicles	1.014	5.583		
Three or more vehicles				
Natural logarithm of total duration				
# of episodes				
Sample size	1016			
Log-likelihood value at convergence	-903.60			

Table 62: Tour Size Model (Business/Pleasure: October–December)

Alternatives (base: Single Person)	Two People		Three or more People	
	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-1.391	-5.331	-5.116	-11.95
Working status				
# of workers (full-time or part-time)	0.200	2.778	0.991	9.902
# of nonworking adults	0.387	3.317	1.234	9.943
Vehicle ownership (base: One or no vehicle)				
Two vehicles	1.045	5.441	1.045	5.441
Three or more vehicles				
Natural logarithm of total duration				
# of episodes	0.129	1.826	0.129	1.826
Sample size	713			
Log-likelihood value at convergence	-627.44			

Table 63: Tour Size Model (Convention/Conference/Seminar)

Alternatives (base: Single Person)	Two and more People							
	January–March		April–June		July–September		October–December	
	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat	Coeff.	T-Stat
Alternative-specific constant	-2.347	-6.988	-1.738	-8.248	-2.579	-8.956	-1.900	-6.530
Working status								
# of workers (full-time or part-time)	0.324	4.373	0.159	2.876	0.361	5.736	0.188	2.554
# of nonworking adults	0.361	3.358	0.307	3.784	0.738	7.842	0.512	4.265
Vehicle ownership (base: One or no vehicle)								
Two vehicles	0.601	2.531	0.546	3.126	0.783	4.113	0.734	3.202
Three or more vehicles								
Natural logarithm of total duration	0.278	2.188						
# of episodes					0.274	2.010		
Sample size	591		982		883		586	
Log-likelihood value at convergence	-346.80		-573.00		-505.20		-344.65	

4.5 Destination and Mode Choice

Mode choice has generally been considered a key component of long-distance modeling and, where it is included, destination choice is typically modeled jointly with mode choice.

Destination Choice

The primary issue for the development of a long-distance model framework is whether destination choice should form part of the model. Of the 24 non-United States models reviewed for this study, just nine included a destination choice (or distribution) phase. The main arguments that have been made in the development of the models are the following:

- Social and business destinations appear fixed to the traveler and the only issue is how often they are visited. That is, this argument would suggest that changes in trip length and distribution should be handled by differential frequency rates to represent the differential impact on demand of services serving specific corridors.
- However, the evidence is that, in the long term, trip length distributions have a form that can only be explained by some sort of choice process. In the United Kingdom, redistribution effects are recognized as the most relevant behavioral change resulting from transport infrastructure (WebTAG, 3.10.3, para. 1.4.8).
- In model estimation, there is experience that estimating a joint model of mode and destination choice gives better results than separate estimates of, for example, a mode choice model; such results have been obtained over many years, most recently in the context of long-distance travel by Rohr et al. (2010). This is not a priori a surprising finding; there are other examples where focusing on just one dimension within a multidimensional choice process leads to inferior model performance and potentially very different substantive findings; for example, see the discussions on joint modeling of airport, airline, and access mode choice in Hess & Polak (2006a, 2006b).

When destination choice is included, the alternatives represented in the model have been defined as discrete, nonoverlapping areas, usually called zones. More detailed spatial definition could be considered (e.g., in the form of blocks or parcels), though this does not seem to have been done for long-distance modeling. The choice between more aggregate and less aggregate definitions turns on whether averaging or sampling is a more appropriate procedure for the specific context of the model being developed, as averaging must be applied when large spatial alternatives are used and sampling must be used for smaller and more numerous alternatives.

When spatial alternatives cover more than the smallest area, it is likely that the land use within the alternative will be mixed. A description of the attractive power of the area will then involve multiple variables, such as employment in different categories or areas devoted to specific activities, such as hotels or recreation areas. Procedures have long existed to handle such multiple attractors in modeling (Daly, 1982).

A possible improvement to this process, which has also not been applied in long-distance modeling, is to introduce a process of choice set formation. This would naturally be associated with the use of a high level of spatial definition for the elementary alternatives, relying on the choice set formation to reduce the number of alternatives. However, the formation of choice sets needs to be modeled in a way that will be intuitively convincing, statistically well founded and suitable for forecasting. These are challenging requirements, and estimation and application complexity is obviously increased.

When choice set formation is not modeled, it is useful to understand that there is no assumption that travelers consider all the destinations. The model represents a choice probability without being explicit about the process that leads to the formation of choices.

An issue in modeling destination choice is that the tours made by travelers can visit several destinations. Procedures of varying levels of complexity have been applied in general travel demand forecasting. For long-distance modeling, it is unusual that multiple long trips are made in the course of a tour. More often, a long trip will be made to a primary destination and then short trips will be made to other destinations in the vicinity of the primary destination before taking the long trip back home. Other common causes of multiple destinations are incidental stops along the route for an overnight stay, a meal, or simply to refuel; these are generally (and reasonably) neglected in modeling mode and destination choice. For long-distance modeling, therefore, the general approach has been to model an out-and-back tour to a primary destination and to neglect short-distance travel and the small number of people who make tours with three or more long-distance legs. However, models have represented the fact that people spending more time at their destination are more likely to make side trips and are therefore more likely to take their cars to facilitate the side trips than are people making shorter stays.

An important issue for destination choice modeling is that it is difficult, or in some cases perhaps impossible, to design stated preference data that are relevant. The central issue is that much destination choice results from long-term processes: business and social relationships are more easily developed at shorter than longer distances, but stated preferences require respondents to consider short-term behavioral changes, which are unlikely to be appealing. This means that modeling must be based on revealed preferences. There is potential for exploring the benefits of stated preference techniques other than stated choice, notably stated intentions-type data, though attempts at this have not yet been entirely convincing.

The models that have been used for destination choice have been of the logit form, treating the alternatives symmetrically so that the model of destination choice alone is multinomial logit. Sometimes these are described as gravity models, but the gravity formulation can be rewritten as a logit model without changing the essence of the model. The logit choice formulation is more suitable for statistical analyses.

A further possibility that has not been adopted in destination choice models that have been used in practice is to represent correlations between alternatives. This could be useful where there are similarities between alternatives (e.g., that they are close in space or have similar local characteristics, but this has not yet been exploited in long-distance modeling). The introduction of such correlations, together with those involving mode choice, might mean that cross-nested logit or even more sophisticated model forms, such as mixed logit, would be required.

Nonlinearity has been found to be necessary in many practical models (Ben-Akiva et al., 1987; Daly, 2010, which also gives a literature review), as maintaining linearity of the utility functions in a MNL formulation will often lead to the model predicting unreasonably high elasticities. It can be shown that a more sophisticated model formulation can represent these effects at least as well as a nonlinear formulation, but such models will usually involve an increase in run time (see Daly and Carrasco, 2009). It may be necessary to introduce nonlinearity in more than one attribute to maintain model accuracy over the full range of attributes.

Mode Choice

The mode choice alternatives defined in long-distance models have been:

- **Car**, which is sometimes split into car driver and car passenger to allow car occupancy to be a function of policy and network conditions—also, toll roads can be recognized as separate alternatives in some cases;
- **Train**, which is sometimes split into high-speed and "classic" trains;
- **Air**; and
- **Bus or coach**, not always included due to perceptions that it is a low-income mode.

The formulation for the car mode can vary, as it is possible to split cars into single- and multi-occupied vehicles, with the potential for further splits depending on the level of occupancy (e.g., two people vs. three-plus). The split into driver and passenger is more common outside the United States, where HOV lanes are rare and the improved forecasting of the total number of cars is important for modeling congestion. In the United States, the occupancy can be important for determining access to HOV lanes (see the discussion by de Jong et al., 1998). However, if the model represents travel by parties, rather than by individuals, these distinctions are treated as exogenous rather than as choice and a single-car alternative can be modeled.

For train, as well as the HSR/classic split mentioned above, there are possibilities for modeling:

- The class of travel (i.e., first vs. standard class [using the UK nomenclature], perhaps generalized to ticket type).
- Choice of access mode.
- Choice of access station.

The last two choices need to be made in all models, of course, to determine levels of service for the train mode, but in some models, positive probabilities are calculated for several access modes and/or access stations. In these cases, it needs to be noted that egress modes are different from access modes in that driving one's own car is not available as an egress mode, so that car driver will have a lower share on egress. The distinction between "park-and-ride" and "kiss-and-ride" (i.e., drop-off or pick-up) can be important if information is required about car parking at stations or simply to improve the measures of level of service.

For air, there are subchoices of class/ticket, access mode and airport, along exactly the same lines as for train. The extensive literature on airport access modeling provides useful guidance that may also be applicable to station access.

For bus, few subchoices have been modeled in existing models and the difficulty is rather to obtain good measures of levels of service. The bus network is much less fixed than train or even air networks, and highway distances are often used as a proxy, though these do not give good indications of levels of service for bus.

Determining fares for train, air, and bus modes is difficult, as fare systems can be complicated, with fares depending on person types, while yield management systems are often difficult or deliberately impossible to understand. Generally, average or proxy variables are used, with a corresponding loss of accuracy. This is especially true in an air travel context, where reliance on average fares (generally the only available information), despite the dynamic nature of fare systems, often leads to an inability to estimate meaningful and significant fare effects in revealed preference data (cf. Hess & Polak, 2006a, 2006b). Note, however, that these variables will generally be more reliable than using reported values, which would only give results for a small number of modes, probably only for the chosen destination, and which are subject to severe biases such as self-justification.

For mode choice, it is possible to work with stated preference data; stated choice has been used in many studies for high-speed rail and for other infrastructure studies such as tunnels and bridges, as well as being used extensively for air travel behavior analyses (cf. Hess et al., 2007, and the review therein). However, for forecasting, stated preference data would generally not be used unsupported, as the error in the model is a function of the interview process and cannot be reproduced for forecasting (Daly and Rohr, 1998).

The nonlinearities noted for destination choice modeling should be included, or at least tested, in mode choice modeling when destination choice is not a part of the model system.

Joint Destination and Mode Choice

Model systems with both mode and destination choice components have generally been of the tree-nested logit form. This form of model gives some freedom for the relationship of the various components of the model, without imposing an undue penalty on run time. In this context, it should be noted that the relationship of mode and destination choice is determined by the relative error in these two model components, which is a function of context-dependent aspects of the model, such as zone size and the measurement of socioeconomic variables. Thus, the nesting needs to be tested in every case. Typically, international models have found destination choice to be modeled with lower error than mode choice, whereas it seems that domestic models have made the opposite finding.

To determine model structure, it is important that the estimations should be made simultaneously rather than sequentially. In a statistical sense, sequential estimation, while known to be consistent, is quite inefficient. In practical terms, sequential estimation is liable to give wrong results and this has been found to be particularly true when the model structure is estimated. Sequential estimation is also much more time consuming and error prone.

Further nesting within the tree logit framework can be used to represent correlation within different destination classes (e.g., geographical or in terms of trip length) or between modes and submodes. For example, it is usual that station (or airport) access modes are nested below main mode choice and that station choice is at a still lower level. However, in cases where structure is uncertain, cross-nested logit models can be considered (cf. Hess & Polak, 2006b and Hess et al., 2011). Such models are considerably more difficult to estimate; however, once estimated, these models impose little extra run time penalty compared with tree-nested models. It is also worth noting that where there is a possibility that the optimal ordering of nesting levels varies across respondents in the data, a mixture approach can be used, such as discussed recently in the mode-destination context by Ishaq et al. (2012).

Moving beyond the GEV framework, which includes MNL and both types of nested logit model, is also possible. The most obvious model type to consider is mixed logit, which has not been applied for long-distance models, to the research team's knowledge, but which can be applied on the required scale (Daly and Carrasco, 2009). The use of mixed logit models would allow correlations between alternatives to be flexibly modeled, as well as allowing for variation in individual preferences such as values of time.

Current Practice

Mode and submode choice models form components of the majority of current and recent forecasting systems that include long-distance travel. While destination choice is not always included, it forms a natural component of model systems that cover more than a single corridor. Current practice represents these choices by tree-nested logit models.

The scope of the models includes choice among the key long-distance modes: air, rail, and car, with bus added in some contexts. Submodes could include choices such as:

- Choice of different types of train service; in particular, the use of high-speed rail and conventional services;
- Choice of class of train travel, ticket type, or even of operator;
- Choice of access modes for train or air travel; and
- Choice of major route options, such as the use of tunnels, bridges, and ferry services to cross major barriers.

When destination choice is modeled, the alternative destinations are usually modeled using geographical zones.

The "utility" functions used in the logit models typically comprise measures of generalized time (i.e., time including separately measured components such as "in-vehicle" [line haul] time) and, for modes other than car, access time and some measures of frequency (e.g., an estimate of waiting time) and of interchange penalty. The weighting of these time components may be obtained from stated preference analyses, particularly when new mode or submode alternatives are being considered; alternatively, standard weightings may be taken from government advice (e.g., the UK government's WebTAG), from meta-analysis, or from analysts' experience. The use of revealed preference is also possible in this context, though less common. It is not common in current practice that these weightings would vary between individual travelers, either based on measured characteristics of the traveler or as a random distribution. However, different models are often used for different travel purposes and these may have different weightings.

The price of the alternatives is converted to a time equivalent value by applying a value of time. Most often, time is used as the numéraire (i.e., dividing cost by the value of time). In this instance, it is more common to consider variation over the travelers—in particular, allowing variation by tour length (i.e., giving a nonlinear function) or by income, as well as by travel purpose. Again, estimation of the value of time may be made by stated preference analyses or by reference to standard values. The use of random variation of values of time may sometimes be considered where price differentials are large, but this remains uncommon.

Nonlinear functions may be used in the model, applied either to the components of generalized time or to the entire generalized time function, sometimes termed "cost damping." Often, these transformations are applied to bring the forecasting properties of the model in line with an a priori expectation (e.g., that the elasticity would lie in an expected range).

Often, the prediction of mode and submode choice within the logit model is applied with a tree-nesting structure. Here again, the nesting parameters may be obtained by stated preference analyses or based on values imported from previous studies.

When destination choice forms part of the model, the generalized time functions would be applied for each combination of mode and destination, but the form of the function would be consistent across the destinations. Additionally, however, it is necessary to include in the model a measure of the attractive power, or "size," of each destination, because the number of elementary destinations in a zone is not uniform and the number of tours choosing each destination is a function of the attractiveness of the destination and the difficulty of reaching it.

All else being equal, the probability of choosing a destination zone is proportional to the number of elementary destinations in that zone. As shown by Daly (1982), this proportionality can be achieved by including, in the utility function for each destination, a log function of the number of attractions in the zone; this formulation is widely adopted.

When destination choice is being modeled, it is desirable to model the choices of mode and destination simultaneously. This is usually done with a tree-nested structure, in which the relative sensitivity of choice between modes can be different from the sensitivity between destinations (i.e., a given change in level of service can have a different impact on mode choice than on destination choice). The relative strength of the two effects is represented by their positioning in the tree-nested model, with the higher choice being less sensitive and the lower choice being more sensitive. The positioning of the modes and destinations, and the "structural" parameter that indicates the relative sensitivity, is determined by estimation on local data—in this case, using revealed preference (as stated preference over destinations is difficult) or by import from earlier studies. For example, the UK WebTAG advice indicates that modes should be less sensitive (i.e., modeled at a higher level) than destinations, and provides structural parameter values.

In any case, a simultaneous model is recommended (i.e., the parameters should be estimated in a structure considering the alternative modes and alternative destinations at the same time). Not only does this help in obtaining a suitable nesting coefficient, but it is often found that the additional variance given by considering alternative destinations helps in obtaining more robust parameter estimates.

Advanced Approaches

In this section, three possible departures from current practice for mode and destination choice models are examined.

Cross-Nested Logit

Multinomial (MNL) and simple tree-nested (i.e., without cross-nesting) logit (NL) models are standard approaches for modeling the joint choice of destination and mode choice in large-scale applications. However, the specific case of joint mode and destination choice is often characterized by complex error structures along both the mode and destination dimensions, and a simple NL model cannot capture this. A cross-nested logit (CNL) structure presents an obvious solution in this context.

A key feature of models of destination choice is the incorporation of size variables. These variables capture attributes of the alternatives that are specific to the size of the destination (i.e., rather than the quality of the destination or the journey for the mode-destination pair), and which are independent of the characteristics of the journey from the origin (e.g., travel time and travel cost). To enable the use of CNL in this context, the existing framework needs to be extended by incorporating size variables. In the standard specification of CNL, a suitable model for a mode-destination context is one in which each alternative falls into one mode nest and one destination nest, as shown in Equation 5:

Equation 5: Standard Specification of CNL for Destination and Mode Choice Probabilities

$$p_d \propto \left(\textstyle\sum_m (\alpha_m \exp(V_{md}))^{1/\theta_d}\right)^{\theta_d}$$

$$p_m \propto \left(\textstyle\sum_d (\alpha_d \exp(V_{md}))^{1/\theta_m}\right)^{\theta_m}$$

$$\text{and} \qquad p_{d|m} \propto (\alpha_m \exp(V_{md}))^{1/\theta_m}$$

$$p_{m|d} \propto (\alpha_d \exp(V_{md}))^{1/\theta_d},$$

where V_{md} gives the modeled utility of mode m to destination d; p_d and p_m are unconditional destination and mode-choice probabilities; $p_{d|m}$ and $p_{m|d}$ are conditional choice probabilities; and $0 < \alpha < 1$ are nest membership parameters, subject to a normalization constraint such as $\alpha_m + \alpha_d = 1$.

To permit the inclusion of size variables in a general CNL, the membership of nest k for alternative j is parameterized, as shown in Equation 6:

Equation 6: Inclusion of Size Variables in a General CNL for a Destination or Mode Nest

$$\alpha_{jk} = \alpha_d R_j, \text{ if } k \text{ is a destination nest}$$

$$\alpha_{jk} = \alpha_m {R_j}^{\theta_k}, \text{ if } k \text{ is a mode nest}$$

where R is an adjusted size measure;

$$R_j = \left(\frac{S_j}{\overline{S}}\right)^{\gamma}$$

where S_j is the size of destination j;

$\overline{S}$ is the average destination size, simply normalizing these values[14]; and

γ is the 'log-size multiplier' with a standard value of 1.

What is then obtained, within a destination nest d is shown in Equation 7:

Equation 7: Conditional Probability of Choosing Mode Nest *m* within a Destination *d*

$$p_{m|d} \propto (\alpha_d R_d \exp(V_{md}))^{1/\theta_d} \propto (\alpha_1 \exp(V_{md}))^{1/\theta_d}$$

in which the factor ${R_d}^{1/\theta_d}$ cancels out, as it is constant across modes.

Within a mode nest m, the conditional probability of choosing destination d is shown in Equation 8:

Equation 8: Conditional Probability of Choosing Destination *d* within a Mode Nest *m*

$$p_{d|m} \propto \left(\alpha_m {R_d}^{\theta_m} \exp(V_{md})\right)^{1/\theta_m} = R_d(\alpha_2 \exp(V_{md}))^{1/\theta_m}$$

which is obviously proportional to the adjusted size R_d.

The unconditional choice probability of destination d is displayed in Equation 9:

14 This means that the values of α_m and α_d are not affected by the units of S.

Equation 9: Unconditional Choice Probability of Destination *d*

$$p_d \propto \left(\sum\nolimits_m (\alpha_d R_d \exp(V_{md}))^{1/\theta_d}\right)^{\theta_d} = R_d \left(\sum\nolimits_m (\alpha_d \exp(V_{md}))^{1/\theta_d}\right)^{\theta_d}$$

which is obviously proportional to the adjusted size R_d.

Finally, the unconditional choice probability of mode m is displayed in Equation 10:

Equation 10: Unconditional Choice Probability of Mode *m*

$$p_m \propto \left(\sum\nolimits_d \left(\alpha_m {R_d}^{\theta_m} \exp(V_{md})\right)^{1/\theta_m}\right)^{\theta_m} = \left(\sum\nolimits_d R_d (\alpha_m \exp(V_{md}))^{1/\theta_m}\right)^{\theta_m}$$

in which the weighting of the contributions of each destination is appropriately moderated by the size.

Mixed Logit

While all types of choice models allow for variations in sensitivities across decision-makers through interactions with characteristics of the respondents or the journey, there are limits to the amount of heterogeneity that can be captured in this way. The Mixed Multinomial Logit (MMNL) model has become the standard tool in academic research for capturing such heterogeneity, and it is also widely used in applied work.

The MMNL model accommodates taste heterogeneity in a continuous specification, through integration of MNL choice probabilities over the assumed multivariate random distribution of the vector of taste coefficients β. In particular, let $P_{n,t}(\beta)$ be the MNL probability of the observed choice for respondent *n*, conditional on a vector of taste coefficients β.

The log-likelihood (LL) function of the corresponding MMNL model is shown in Equation 11:

Equation 11: Log Likelihood (LL) Function of the Mixed Multinomial Logit Model

$$LL(\Omega) = \sum_{n=1}^{N} \int_{\beta} ln\left(P_n(\beta)\right) f(\beta|\Omega) \mathrm{d}\beta$$

where *N* is the number of respondents and the vector of taste coefficients β follows a random distribution $f(\beta|\Omega)$ with a vector of parameters Ω. This LL function has no closed-form solution, and the typical approach to estimation is to replace the LL by the simulated log-likelihood (SLL), with:

Equation 12: Simulated Log Likelihood (SLL) Function of the Mixed Multinomial Logit Model

$$SLL = \sum_{n=1}^{N} \sum_{r=1}^{R} \frac{ln\left(P_n(\beta_{r,n})\right)}{R}$$

where $\beta_{r,n}$ now represents one of *R* draws for respondent *n* from the distribution $f(\beta|\Omega)$. Increasing the number of draws from the multivariate vector β, and especially their coverage of the multivariate domain of β, reduces the error introduced in estimation.

Random-Regret-Minimization Approach

The field of choice modeling has seen a recent surge in interest in what can broadly be termed as models that depart from a purely compensatory decision-making approach. In a standard random utility maximization (RUM) framework, the probability of choosing a given alternative is a function of the relative value of that alternative's utility, compared to all the other available options. The utility is a function of only the attributes of that alternative and any disadvantages for one characteristic (e.g., cost) can be compensated by advantages in another characteristic (e.g., time). In models that depart from this compensatory framework, a situation could exist where disadvantages are penalized more than advantages. An example of such a framework is the random-regret-minimization approach, where an individual does not choose the option that maximizes his/her utility, but chooses one that minimizes the regret experienced when choosing an alternative that is outperformed by another alternative for one or more characteristics. Such model structures tend to favor compromise alternatives.

Factors Limiting Implementation of Advanced Models

The use of any model for the analysis of large-scale problems is affected by the size of the choice set (i.e., number of alternatives), the number of individuals in the data, and the number of model parameters. All three of these arise in the analysis of long-distance travel, and the resulting increase in complexity accentuates the already existing differences in computational cost between simple and advanced models. Sampling of individuals can help, as can sampling of alternatives. However, an important issue in this context is that the latter introduces bias, and correction approaches are not straightforward for advanced models.

The estimation of more parameters clearly leads to further difficulties for both theoretical and empirical identification. Moreover, the key factor in the computational cost of a modeling approach is the number of calculations involved. This affects the three departures from standard modeling approaches above in the following ways:

- The CNL model involves the calculation not only of utilities for individual alternatives (like a MNL model), but the calculation of the number of LogSums and probabilities for nests of alternatives; this is substantially higher than in a NL model given that each alternative can now fall into multiple nests.

- The MMNL model requires simulation-based estimation techniques, which means that utilities and choice probabilities need to be calculated not just once per alternative and individual, but once per draw used in the simulation process. This complexity increases further when making use of more flexible distributions in the specification of the model.
- For non-RUM models that exploit the choice context, the increase in computational complexity stems from the fact that the measure related to the appeal of an alternative (i.e., the counterpart of a utility) is a function of comparisons with other alternatives, which—with large numbers of alternatives—leads to a substantial increase in the number of computations.

Indirect Inference as a Possible Solution

Indirect inference (II) is a technique that can be used in a situation where the *true* model desired is difficult or impossible to estimate, but where it is possible to simulate data using this model with reasonable computational cost.

The process used in II can be explained in a few sequential steps:

1. The *true* model is the model that one desires to use on the data, but that cannot be estimated due to computational reasons.
2. This *true* model uses a vector of parameters β, and a set of *R* different sets of values are defined for β drawn from a reasonable range, which is not too wide and not too narrow, with β_r referring to the r^{th} such set.
3. *R* sets of choices are simulated using the *true* model, where the choices Y_r for set r are simulated using the vector of parameters β_r.
4. A simpler model is then defined, known as the *auxiliary* model, which is easier to estimate, but has at the least the same number of parameters as the *true* model; the *auxiliary* model is estimated on each of the *R* sets of simulated choices, yielding parameter vector α_r in estimation on Y_r. While this is not a theoretical requirement, it seems wise for the *auxiliary* model to bear some resemblance to the *true* model.
5. A relationship between the parameters of the auxiliary model, α, and the parameters of the true model, β, is then formulated. This relationship is known as the *binding* function, and it measures the diversion in parameters when estimating the *auxiliary* model on data simulated using the *true* model. A typical approach would be to use regression of α on β, where, with sufficiently large *R*, this binding function converges to a nonstochastic limit.
6. Finally, the *auxiliary* model is estimated on the real (nonsimulated) data—say Y_o—to yield estimates, α_o. The inverse of the preceding binding function is then applied to obtain inferred estimates, $\beta_{o,}$ for the true model on the real data without needing to estimate it.

Under (fairly) standard regularity assumptions, the II approach yields estimates that are consistent and asymptotically normal. In practice, the assumptions in step two have a certain impact on the performance of the approach; as a result, an iterative approach may be needed, first simulating for a wider range for β, and, after completing steps 3–6, narrowing the range of β and repeating steps 3–6 one or more times.

II was introduced by Smith (1993) and applications in choice modeling have been put forward by Keane and Smith (2003), and more recently by Karlström et al. (2013), for route choice modeling, and by Wang et al. (2013) for estimation of mixed logit on large samples of choices. In the current application, II can be exploited primarily by using an *auxiliary* model that is structurally the same as the *true* model, but which makes use of a much smaller sample of alternatives, and, in the context of MMNL, also a limited set of draws in simulation-based estimation.

What follows is a simple illustration of the potential benefits of the II framework to large-scale modeling of the type addressed in this project. In particular, the case of simulated data based on the California survey is examined. The reason for using simulated data is that in this case, the *true* values of the model parameters are known, permitting testing of the performance of the method without any outside effects. The California survey was selected over the national data for several reasons, including timing (in terms of data availability) and size (the smaller size of the California dataset made it more applicable to this initial testing work).

The specific scenario that was examined made use of a sample of 6,635 respondents, with four modes and up to 58 destinations. This produced a total maximum choice set size of 232; II was then applied to test the effect of sampling destinations.

The following steps were used in the empirical work:

1. Choices were first simulated from this data for a CNL model in which each alternative falls into one mode nest and one destination nest, using α_m=0.4, and as a result α_d=0.6. The corresponding nesting parameters, λ_{mode} and $\lambda_{dest,}$ were set to 0.3 and 0.7, respectively. Two size variables, relating to leisure/hospitality, were used and for which the log-size factor was set to 0; for other services, the log-size factor, γ_{OS}, was set to 0.65. Included were constants for bus (δ_{bus}=-4.5), rail (δ_{rail}=-3.5), and air (δ_{air}=-2.8), along with time (β_{time}=-0.0038) and cost (β_{cost}=-0.0142). These parameters are, to some extent, informed by estimation work on the actual real-world data.
2. For the II work, the100 simulated datasets were then produced, where the parameters used in simulation varied widely, covering approximately 50% on each side of the above true values.
3. A sample of up to 20 destinations per respondent was produced, using with-replacement sampling and simple utility functions from an MNL model estimated on the real data.
4. For each of the 100 datasets simulated above, the CNL model was then estimated using the sample of 20 destinations, adding the chosen destination when it was not already included in the sample. This process produced the parameters for the auxiliary model, which is a CNL model with sampling of destinations, so that the bias in the results for the auxiliary model is thus caused by sampling. In these estimations, the simple McFadden "positive conditioning" correction was included for sampling in MNL, which should address some, but not all of, the sampling bias given the use of CNL instead of MNL.
5. Multivariate regression was then used, with the auxiliary estimates as the dependent variables, and the true values as the explanatory variables. This yielded the estimates (with t-ratios in brackets) presented in Table 64.

6. Finally, the auxiliary model was then estimated on the true data, using the binding function inside the estimation. This implies that the parameters being estimated relate to the true model (i.e., the explanatory variables in the above regression), while the parameters used in calculating the choice probabilities (and hence the likelihood function, which is maximized) relate to the auxiliary model and are obtained through the binding function above. The crucial factor in this process is that the same sample of 20 destinations per individual is used as in the estimation on the 100 II datasets. This means that the bias introduced by sampling should be the same as in the II runs, allowing the binding function to correct for it.

The results of three models are presented in Table 65. The first set details the results for the model estimated without sampling of alternatives. This recovered the true values used in simulation quite closely; without any significant bias for rail, it approximates significance. The second set of results is for a CNL model estimated with sampling of alternatives, using only the McFadden correction and not II. Significant bias is seen in comparison with the estimates on the full data (i.e., not the values used in simulating that data) for six out of nine parameters, with some large biases. The third set of results is for a CNL model estimated with sampling and the II binding function. In this instance, the bias is no longer statistically significant for any of the nine parameters, and it decreases in magnitude for seven of them. The drop in significance for the one parameter that still has a high percentage bias (λ_{mode}) was a result of the correction to the standard errors that was also obtained through the use of the binding function. These standard errors now relate to the parameters one would estimate on the full sample, rather than those of the *auxiliary* model.

As a final step, implications in terms of model fit were also examined. The estimation of the *true* model on the *true* data gives a LL of -20,118.5. Using the parameters obtained with the second (i.e., sampled) model on the full dataset gives a LL of -20,194.1 (i.e., noticeably more negative). On the other hand, using the estimates obtained from II in the computation of the LL on the full sample of destinations gives a LL of -20,121.2, which is much closer to that of the *true* model, suggesting that II manages to correct the bias introduced by the sampling of alternatives.

Table 64: Multivariate Regression Model Estimation Results

		True Values									
		intercept	δ_{bus}	δ_{rail}	δ_{air}	β_{time}	β_{cost}	γ_{OS}	λ_{mode}	λ_{dest}	α_m
Auxiliary Values	δbus	-2.207 (-5.2)	1.1825 (26.74)	-0.0182 (-0.39)	0.0344 (0.78)	-40.6378 (-1.3)	-6.6977 (-0.74)	-0.047 (-0.16)	-0.6111 (-1.37)	2.5978 (8.55)	1.0884 (3.59)
	δrail	-1.7449 (-8.45)	0.0318 (1.48)	1.1597 (51.58)	0.0611 (2.83)	-76.7806 (-5.06)	-4.8799 (-1.11)	-0.0031 (-0.02)	0.0973 (0.45)	1.3958 (9.45)	1.07 (7.26)
	δair	-0.6647 (-0.87)	0.0053 (0.07)	-0.0977 (-1.17)	1.3254 (16.55)	-77.8717 (-1.38)	32.8946 (2.02)	-1.1899 (-2.19)	1.5398 (1.91)	1.0566 (1.93)	0.2974 (0.54)
	βtime	-0.0005 (-2.35)	0 (0.92)	-0.0001 (-2.86)	-0.0001 (-5.13)	1.1536 (71.4)	0.0062 (1.33)	-0.0001 (-0.68)	0.0008 (3.31)	0.0002 (1.35)	-0.0002 (-1.49)
	βcost	-0.005 (-3.78)	-0.0004 (-3.06)	-0.0001 (-0.93)	0 (0.25)	0.3466 (3.54)	1.0555 (37.39)	0.0008 (0.84)	-0.0005 (-0.34)	0.0054 (5.63)	0.0022 (2.3)
	γOS	-0.195 (-6.86)	0.0034 (1.15)	0.0032 (1.02)	-0.0039 (-1.3)	11.0706 (5.29)	0.7964 (1.32)	1.0053 (49.88)	-0.0791 (-2.65)	-0.0261 (-1.28)	0.0243 (1.2)
	λmode	0.2022 (7.39)	-0.0015 (-0.54)	-0.0013 (-0.44)	-0.0042 (-1.47)	-21.0553 (-10.46)	-1.7311 (-2.98)	-0.0023 (-0.12)	0.3017 (10.49)	0.0254 (1.29)	-0.1611 (-8.24)
	λdest	0.3079 (4.51)	-0.0133 (-1.87)	0.0098 (1.31)	-0.0125 (-1.76)	10.3609 (2.07)	1.401 (0.97)	0.0076 (0.16)	-0.0784 (-1.09)	0.7802 (15.97)	-0.085 (-1.74)
	αm	0.3642 (7.29)	-0.0037 (-0.72)	-0.0179 (-3.3)	-0.0272 (-5.21)	50.4103 (13.73)	4.2542 (4.01)	-0.0021 (-0.06)	0.2813 (5.36)	0.2026 (5.66)	0.267 (7.48)

Table 65: Results of Three Models

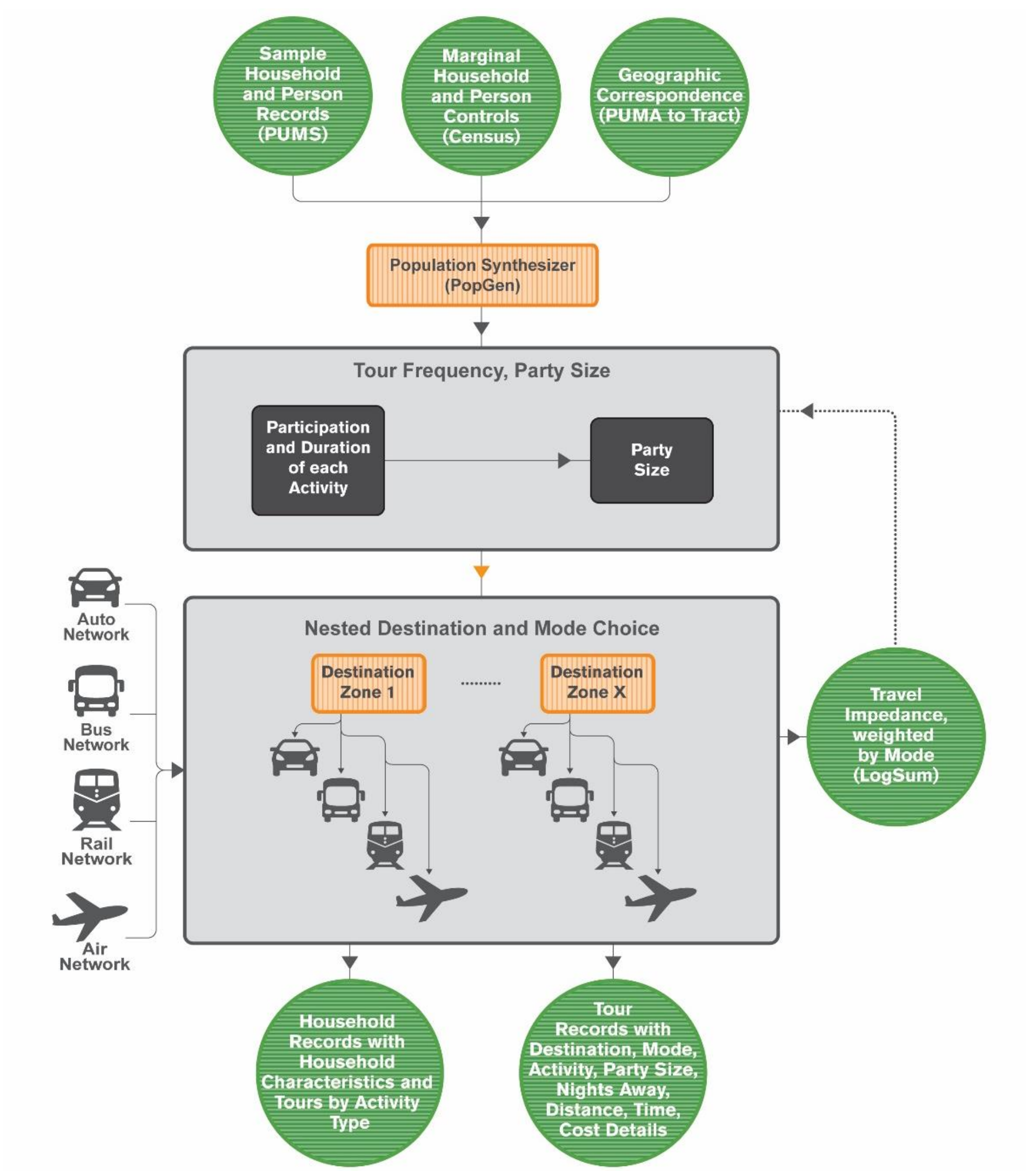
	δ_{bus}	δ_{rail}	δ_{air}	β_{time}	β_{cost}	γ_{OS}	λ_{mode}	λ_{dest}	α_m
True Values	-4.5	-3.5	-2.8	-0.0038	-0.0142	0.65	0.3	0.7	0.4
Estimate On Full Data	-4.4244	-4.2295	-2.7715	-0.0039	-0.0132	0.6349	0.2662	0.7045	0.4266
std err	0.36	0.38	0.21	0.00	0.00	0.06	0.07	0.07	0.07
t-ratio against zero	-12.33	-11.10	-13.52	-19.68	-9.21	11.22			
t-ratio against 1							-11.08	-4.12	
t-ratio against 0.5									-0.99
bias	-1.68%	20.84%	-1.02%	2.17%	-7.39%	-2.33%	-11.27%	0.65%	6.66%
significance of bias	0.21	-1.91	0.14	-0.42	0.74	-0.27	-0.51	0.06	0.36
Estimate Using Sampling	-5.1867	-5.1771	-3.3708	-0.0042	-0.0135	0.3521	0.3996	0.7955	0.6544
std err	0.28	0.36	0.22	0.00	0.00	0.06	0.02	0.08	0.06
bias	17.23%	22.40%	21.62%	7.48%	2.49%	-44.53%	50.10%	12.91%	53.40%
significance of bias	-2.72	-2.67	-2.76	-1.06	-0.23	-4.84	6.82	1.13	3.72
Estimate Using Sampling With II	-4.4458	-4.2915	-2.7683	-0.0038	-0.0138	0.6549	0.4041	0.7148	0.4304
std err	0.41	0.35	0.24	0.00	0.00	0.06	0.09	0.09	0.11
bias	0.48%	1.47%	-0.12%	-2.06%	5.24%	3.15%	51.78%	1.45%	0.88%
significance of bias	-0.05	-0.17	0.01	0.36	-0.46	0.34	1.48	0.11	0.04

While most work in a long-distance-travel context relies on simple nested logit structures, this work has made the case that complex error structures may exist along multiple dimensions of choice and that a cross-nesting structure might be more appropriate. This is, however, only one departure from “standard” techniques and another big gap between applied work and academic research is in the lack of applications of models allowing for random heterogeneity in preferences in a large-scale context. It is without a doubt the case that accommodating complex error structures and random heterogeneity would lead to major gains in model performance and likely prediction accuracy. It would, however, also result in large increases in computational complexity and empirical identification of the models. For such techniques to gain widespread exposure necessitates more work along the lines of the indirect inference explorations in the present report, or methods such as Bhat’s Maximum Approximate Composite Marginal Likelihood (MACML).

CHAPTER 5. LONG-DISTANCE MODEL DEMONSTRATION

This section details the preliminary implementation of models and datasets created during this project. The modeling process is closely aligned with the approach discussed in Chapter 3, but contains some differences in model form and structure, as described in Figure 44. The preliminary implementation will be further refined and calibrated during the next phase of this project. The first section of this chapter discusses options for the overall model implementation structure, and describes the specific structure used for the demonstration model implementation. The second section documents the specific model components used in the preliminary implementation. The third section discusses software performance issues and options for further software development.

Figure 44: Long-Distance Passenger Travel Demand Model Demonstration



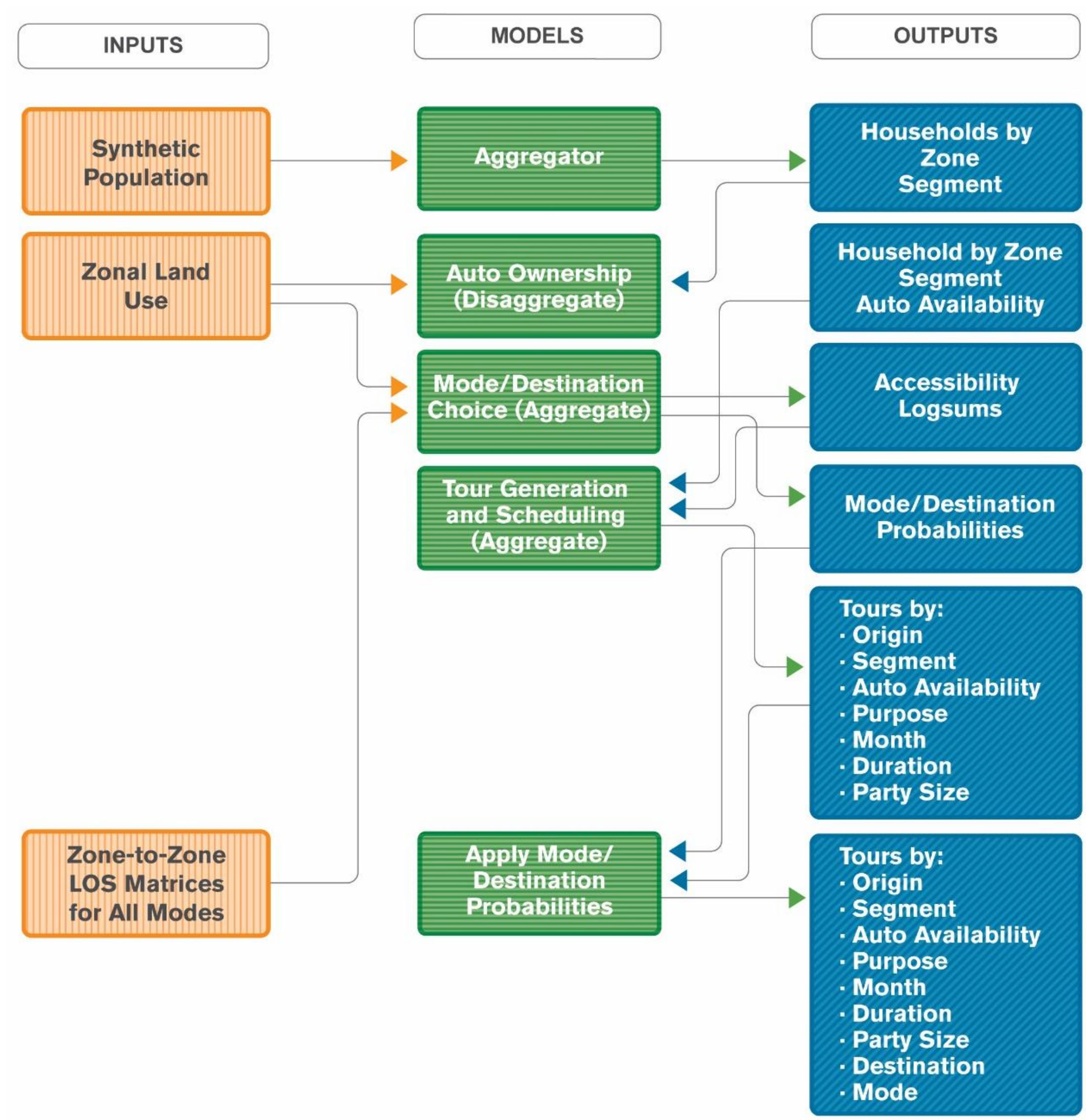
5.1 Structure of the Preliminary Model Implementation

A variety of different application structures were considered for implementation of the national long-distance passenger model, ranging from a more aggregate structure to a fully disaggregate microsimulation model.

Aggregate Structure

Figure 45 illustrates how an aggregate model could be structured. An example of such a model system is the California Statewide Long-Distance Model, first created to forecast demand for the proposed high-speed rail (HSR) system in California (Outwater, et al. 2011, JOCM).

Figure 45: Aggregate Model Structure



The main reason such a system is termed “aggregate” is that the households in the synthetic population are aggregated into sociodemographic segments, and the models are run for each residence zone/sociodemographic segment, rather than for individual households. Also, rather than using Monte Carlo microsimulation to simulate individual long-distance tours for each household, the model probabilities are used directly at each stage in the model system to add additional detail to the output. In this regard, the model structure shown in Figure 45 is analogous to an advanced “4-step” zonal model structure that might be used to forecast regional travel demand. Specifically, the steps in the model system would work as follows:

1. The “Aggregator” aggregates the households in each residence zone into demographic segments that are combinations of specific household attributes. For example, one might use the attributes of household size, number of workers, and income group—and about three or four categories of each—to split the households into roughly 50 different household segments (e.g., approximately 4 x 4 x 3 categories).
2. An auto-ownership model is then applied to each segment in each zone, and the probabilities are used to further divide each segment into auto-availability subsegments (e.g., zero-vehicle households, households with one or more vehicles per adult, and car-competition households [i.e., HH that own one or more vehicles but have fewer vehicles than adults]). This model is termed “aggregate” in Figure 45 because the household variables are limited to those variables used in defining the sociodemographic segments, as contrasted to a “disaggregate” model that could use any household characteristic variables that are in the synthetic-population file. (Note that auto ownership is a variable in the synthetic-population file, based on PUMS data. However, it is not a variable that is used as a control in drawing the synthetic population, nor is it typically available in future-year socioeconomic forecasts. For those reasons, it is better to include an auto-ownership model as part of the forecasting system, as is typically done for advanced urban-regional models.)

3. Aggregate Mode/Destination Choice models are used to calculate measures of the accessibility of long-distance travel from each origin zone to all possible destinations by all available modes, which can then be used as an input variable to the subsequent tour-generation and scheduling models. The LogSum across all modes and destinations is calculated as the relevant accessibility measure, and is calculated for every combination of origin zone, sociodemographic segment, auto-availability level, and long-distance tour purpose. The LogSum can also be calculated within specific distance bands, with the expectation that greater accessibility to destinations within 50 miles will tend to decrease the frequency of long-distance travel, while accessibility to destinations farther than 50 miles will tend to increase the frequency of long-distance travel, with travel for different tour purposes showing different sensitivity to accessibility within different distance bands. (For example, long-distance Commute and Personal Business tours are usually in the range of 50 to 150 miles, so the frequency of those tours may be most sensitive to accessibility and attractiveness of destinations within that distance range.)

4. Aggregate Tour Generation and Scheduling models are applied for each origin zone/demographic segment/auto-availability segment combination to predict the number of long-distance tours generated for each tour purpose during each period of the year. (The period shown in Figure 45 is the month of the year.) If desired, further aggregate models can

be used at this stage to further subdivide the predicted tours by duration of stay (e.g., day tours, 1–2 night tours, 3–6 night tours, and 7+ night tours), and also by travel party size (e.g., 1 person, 2 persons, 3 persons, 4+ persons). These models are sensitive to the mode/destination accessibility variables from the previous step in order to predict some level of tour induction or suppression as accessibility levels and attractions change.

5. The final step of the system is to apply the probabilities from the Mode/Destination Choice models in order to "distribute" each available long-distance destination/mode combination. Because this is typically the most computationally intensive step of the model system, the structure in Figure 45 leverages the fact that these models have already been used to calculate accessibility LogSums (step three). Instead of repeating those same calculations, the model probabilities (which must be calculated anyway in order to calculate the LogSums) are stored in memory and used for this step, avoiding significant potential run time.

The final output of such an aggregate model is tour origin-destination matrices, by mode. If desired, the model can also produce such O-D matrices for every combination of mode, purpose, month, demographic segment, car-availability segment, etc. However, this would require significant memory and disc space to save this large number of matrices, so it is more typical to just aggregate and write out the O-D matrices along a few key dimensions, such as mode, purpose, and time period.

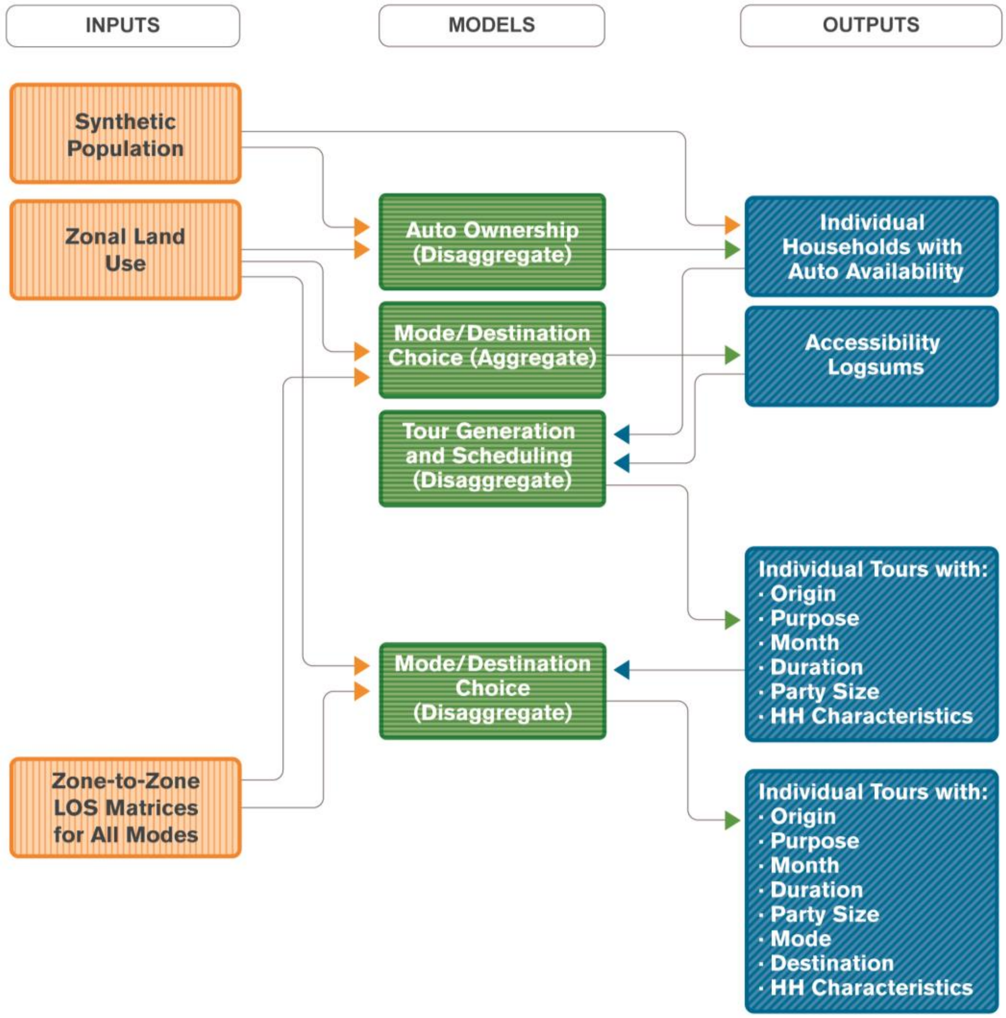
Disaggregate Structure

The other end of the spectrum for model application is a fully disaggregate microsimulation structure, as depicted in Figure 46. Compared to the aggregate structure shown in Figure 45, there are many similarities. The key differences are as follows:

- There is no population "Aggregator" needed, as each household in the synthetic population is simulated individually.
- The same model components are used, although each component is labeled as "disaggregate" instead of "aggregate." This means that the models can include all household characteristics in the synthetic sample as explanatory variables. It also means that more model specifications can be used. For instance, a model structure, such as the MDCEV model recommended for Tour Generation and Scheduling, requires a disaggregate microsimulation framework for application, as it does not allow calculation of closed-form probabilities that are required for aggregate-model-application structures.
- In this structure, it is still expedient to use a more aggregate version of the Mode/Destination Choice models to precalculate accessibility LogSums for use in the Tour Generation and Scheduling models. This is because applying the fully disaggregate version of Mode/Destination Choice models for every possible tour purpose for every household would be prohibitive in terms of run time. (Note that this method of using precalculated aggregate accessibility LogSums is also used in most urban, activity-based [AB] microsimulation model systems.)
- Rather than producing O-D matrices, this structure produces a separate output record for each predicted tour, with all relevant aspects of the tour on the file. These tour records can then be aggregated up to O-D matrices along any desired dimensions, providing more

flexibility than in an aggregate model system, where the number and definition of the output matrices need to be prespecified. Such a model can also produce an output record for each simulated household, indicating the predicted auto ownership, the number of tours predicted for each long-distance purpose, and, perhaps, other output variables summarizing each household's predicted long-distance tours.

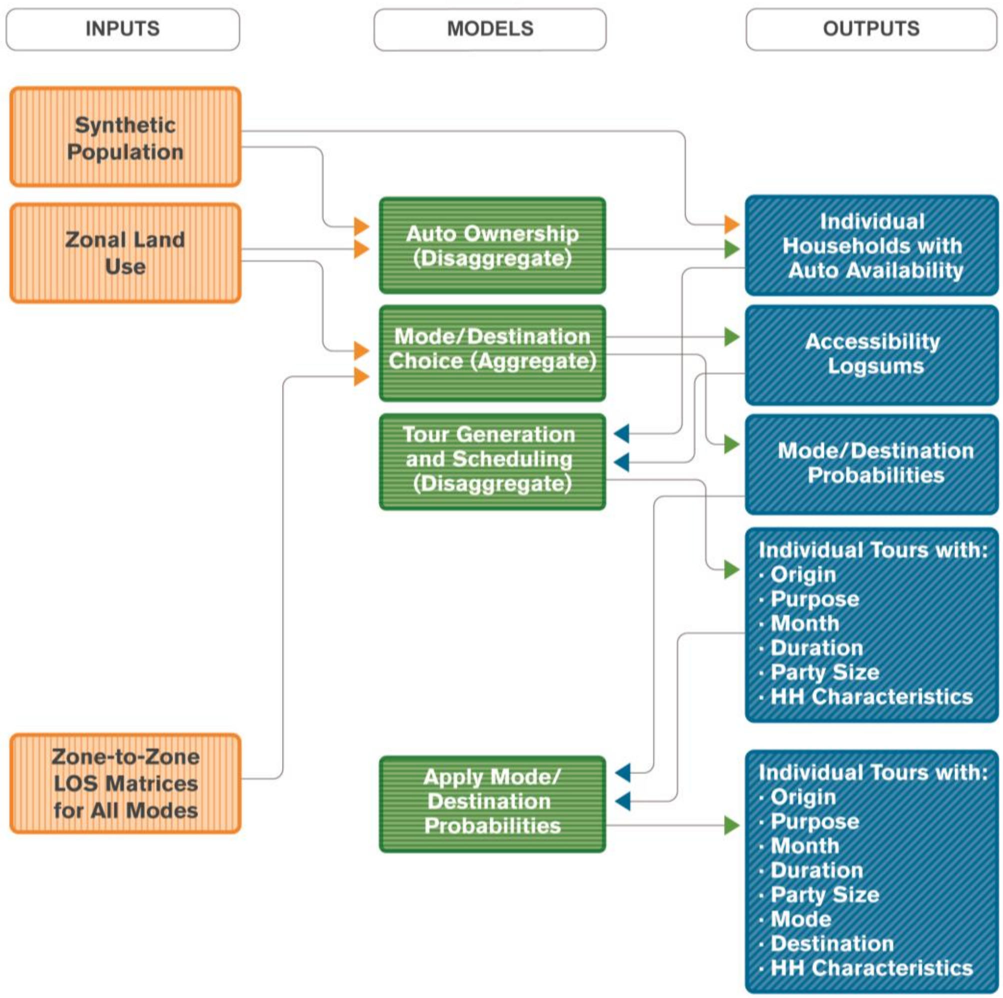
Figure 46: Disaggregate Model Structure



It is clear that the disaggregate model structure provides several advantages over an aggregate structure in terms of the variety of different variables that can be used in the models and written to the output files; this structure may also allow more choice model types to be applied (i.e., those that sample from distributions of parameters rather than having deterministic probability equations). The two potential disadvantages of a disaggregate structure include: 1) longer model run times; and 2) random simulation error from using Monte Carlo simulation rather than applying choice probabilities directly. The larger the population size that the model will be applied to, the more that run time becomes an issue, while random simulation error may become less of an issue (due to fact that random simulation error is generally proportional to the square root of the sample size).

Given the aforementioned advantages and disadvantages, for the preliminary model system, a modified disaggregate structure (Figure 47) was used. This structure is identical to the fully disaggregate structure shown in Figure 46, except for the last step to predict tour modes and destinations. It stores the probabilities calculated from the Mode/Destination Choice models while calculating aggregate LogSums, and uses those probabilities in the final step to perform the Monte Carlo microsimulation to predict a specific mode/destination combination for each tour. This structure results in a model system that runs quickly while still providing all of the advantages of a fully disaggregate model system.

Figure 47: Modified Disaggregate Model Structure



Currently, the Mode/Destination models used in this system are sensitive to household income and household car availability, as well as tour purpose. In future implementations, it would be possible to add additional dimensions to the aggregate Mode/Destination models—such as party size, month or year, and/or nights away from home— while still maintaining a significant run-time benefit versus a fully disaggregate model system implementation.

Note that it would also be possible to use the probabilities to distribute each simulated tour across all available modes and destinations in the final step, thus producing O-D matrices as output, similar to the aggregate model structure. This would partially reduce random simulation error at the final step, but it would have some of the disadvantages of the aggregate structure in that the model outputs would be more cumbersome and less flexible to work with, and would probably require some amount of pre-aggregation. A better approach to reducing random simulation error, especially with regard to spatial distribution, would be to do multiple mode/destination draws per simulated tour (all from the same probabilities), instead of just a single draw. For example, one could perform 25 Monte Carlo draws per tour, with the expansion factor on each of the output tours reduced by a factor of 25.

5.2 Logit Models Used for Initial Model System Implementation

In this section, the logit discrete choice models that are used for the preliminary model system implementation, as depicted in Figure 47, are presented and briefly described. The models described in this section were estimated using data records combined from four different surveys:

- The 2012–2013 California Statewide Travel Survey long-distance survey data.
- The 2001 New York NHTS add-on sample long-distance survey data.
- The 2001 Wisconsin NHTS add-on sample long-distance survey data.
- The 2003 Ohio Statewide Travel Survey long-distance survey data.

Auto Ownership

Although it is possible to observe household car ownership from the PUMS records in the synthetic population, this variable is not used as a control variable in drawing the population, and it is typically not available for future-year demographic forecasts. Therefore, the model has been estimated based on household characteristics from the households in the combined sample from the four aforementioned long-distance surveys.

The model estimation results are shown in Table 66. The base alternative in the model is 2 cars in the household, with utility = 0, and coefficients are estimated for the other four alternatives: 0 cars, 1 car, 3 cars, and 4+ cars. The main implications of the estimated coefficients are the following:

- One-adult households are most likely to own 0 or 1 cars, while households with 3 or 4+ adults of driving age are most likely to own 3 or 4+ cars.

Household income, used in logarithmic form, is an important variable, particularly for higher-income households much less likely to own 0 or 1 cars. (As in most models presented in this section, a separate “nuisance” variable was estimated for those households with missing income data, so that they can be included in the estimation without biasing the other income-related coefficients. Such variables for missing data are not used in model application, because the synthetic sample households do not have missing data.)

Table 66: Household Car Ownership Model

Alternative	0 cars		1 car		2 cars	3 cars		4+ cars	
	Coeff.	**T-stat.**	**Coeff.**	**T-stat.**	**(base)**	**Coeff.**	**T-stat.**	**Coeff.**	**T-stat.**
Constant	6.92	16.4	6.3	29.3	0	-2.69	-13	-4.54	-16.4
1 adult in HH	2.51	36.3	2.53	92.2	0				
2 adults in HH					0				
3 adults in HH					0	1.67	65.9	2	62.3
4+ adults in HH					0	1.93	38.3	3.61	74.1
Log(income)	-1.52	-36.7	-0.85	-41.7	0	0.162	8.5	0.275	10.8
Missing income data	-16.5	-36.4	-9.31	-41.2	0	1.92	8.9	3.15	10.9
Workers / adults ratio	-0.582	-7.1	-0.265	-7.3	0	0.458	13.4	0.98	20.2
HH has children	-0.903	-12.0	-1.01	-34.9	0				
HH head age 65+			0.265	7.5	0	-0.238	-6.5	-0.23	-4.4
HH head age <35	0.219	2.7	0.0762	2.2	0	-0.287	-9.8	-0.157	-4.2
Log(emp+res density)	0.842	48.2	0.261	36.0	0	-0.109	-17.1	-0.247	-29.0
Statistics									
Observations	**72737**								
Rho-squared (0 coeff)	0.361								
Rho-squared (constants)	0.202								

- Beyond the income effect, having a higher fraction of a household's adults working also favors higher auto ownership.
- Households with children are less likely to own 0 or 1 cars.
- Households with young heads (under age 35) or older heads (age 65 or older) tend to have lower car ownership.
- Households that live in zones with higher residential and employment density also tend to have lower car ownership. This last effect is quite strong in terms of t-statistics, and the logarithmic form for density gives a stronger effect than using a linear effect. In an urban-regional model, a more-detailed variable for accessibility would be used, such as the increase in an aggregate mode-destination choice LogSum that derives from car availability. However, such accessibility effects are mainly related to accessibility for local everyday tours, which cannot be accurately measured with the national-level zones and networks used in this model system. Thus, the density within the residence zone (PUMA or county) provides a strong proxy for local accessibility.

Tour Frequency

In contrast to the MDCEV model of annual generation and scheduling of long-distance tours, which treats all long-distance tours made by a household during the year in a joint-choice framework, these simpler logit models used a single day as the decision period. Although the various surveys have different lengths of retrospective recall for the long-distance surveys (8 weeks for the California statewide survey), breaking the data down into individual days has the advantage that only a few household-days (only about 0.04%) have more than one long-distance tour generated on that given day, meaning that tour generation can be modeled as a 0/1 choice—no tour, or one tour for a given day. Also, because the time lapse between each survey date and the date that the person actually completed the long-distance survey is known, a nonresponse bias function, predicting how the reported tour generation decreases with the amount of time passed between each travel day and the survey, can be estimated.

Table 67 presents a model estimated using the data from the California 2012–2013 Statewide long-distance survey. (Although the data from the Ohio, New York and Wisconsin long-distance surveys can be used to further calibrate the model, they have not yet been used in estimation.) The base alternative in the model is to begin no long-distance tours during the specific day. The five other alternatives shown in the columns are to make a tour for any of the long-distance purposes. Note that these were not estimated as separate models—the household has the choice of making a tour for any one of the tour purposes, but not more than one. (A separate model, described below, was used to predict the small number of household-days with two or more tours.)

Table 67: Household-Day Tour-Generation Model

Purpose alternative	Commute		Business		Visit F&R		Leisure		Pers. Bus	
	Coeff.	**T-stat**	**Coeff.**	**T-stat**	**Coeff.**	**T-stat**	**Coeff.**	**T-stat**	**Coeff.**	**T-stat**
Base alt. is no tour in the day										
Constant	-15.1	-19.9	-11.7	-38.2	-6.05	-35.6	-9.06	-48.3	-3.12	-11.6
Log (income)	0.41	8.4	0.619	28.5	0.200	13.5	0.445	27.0	0.0954	4.9
Missing income data	4.94	8.7	7.19	28.3	2.22	13;0	5.00	26.2	1.02	4.5
HH has 0 car	-1.8	-3.1	-0.222	-2.0	-0.519	-6.8	-0.151	-1.8	-0.701	-6.8
HH has fewer cars than adults	0.343	4.2	-0.122	-3.4	-0.188	-7.0	-0.327	-11.6	-0.0844	-2.4
HH has children	0.272	4.2	-0.099	-3.4	-0.299	-12.1	0.0257	1.1	-0.0897	-2.7
HH workers/adults ratio	1.08	9.9	0.837	19.4	-0.0951	-3.4	-0.0966	-3.3	-0.465	-11.7
One-person HH			-0.221	-5.5			-0.253	-7.9	-0.343	-8.2
HH head under age 35	-0.419	-3.3	-0.302	-5.6					-0.498	-7.8
HH head age 65 or older	-0.512	-5.2	-0.244	-6.8					-0.11	-3.3
Mode/dest LogSum 0-50 miles	-0.299	-15.2	-0.0684	-8.0	-0.0547	-6.2	-0.0423	-4.7	-0.243	-28.4
Mode/dest no zones 0-50 miles	-2.37	-8.7	-0.582	-5.6	-0.604	-6.6	-0.455	-4.8	-1.75	-20.1
Mode/dest LogSum 50-150 miles	0.611	12.1	0.0303	1.6					0.0407	2.3
Mode/dest LogSum over 150 miles					0.0401	3.0	0.0416	3.1		
January					-0.456	-9.6	-0.471	-9.3	-0.246	-4.3
February			0.164	3.7	-0.275	-6.8	-0.291	-6.7		
March			0.295	7.1	-0.177	-4.7				
April										
May (base)										
June	-0.379	-3.3					0.192	5.8		
July	-0.448	-3.6					0.396	12.1	-0.127	-2.4
August							0.234	6.9	-0.142	-2.8
September	-0.35	-3	0.0921	2.1	-0.249	-6.6			-0.163	-3.3

Purpose alternative	Commute		Business		Visit F&R		Leisure		Pers. Bus	
October	-0.472	-3.7	0.184	4.4	-0.297	-7.5	-0.126	-3.2	-0.253	-4.8
November							-0.282	-6.4	-0.135	-2.5
December			-0.348	-5.8			-0.484	-9.4	-0.298	-4.9
No. of days before survey			-0.0072	-4.3	-0.0131	-9.3	-0.0094	-6.8	-0.0141	-7.5
Log(no.days before survey)	-0.409	-15.0	-0.177	-6.4	-0.0973	-4.2	-0.13	-5.8	-0.131	-4.4
Statistics										
Observations	**1,479,150**									
No. of tours (% of HH-days)	1,074	0.07%	6,575	0.44%	9,857	0.67%	10,193	0.69%	5,619	0.38%
Rho-squared (0 coeff.)	0.924									
Rho-squared (c constants)	0.027									

The fifth row of Table 67 shows that on any given survey day, only 0.07% of households made a long-distance commute tour, 0.44% a long-distance business tour, etc. Across the five purposes, these fractions sum to 2.25%, meaning that in 97.75% of cases, the chosen alternative is the base alternative (no tour). The key results for the model include the following

- The long-distance tour rates for all purposes increase with the logarithm of household income, with the effects strongest for Business and Leisure tours, and weakest for Personal Business.
- Over and above the effect of income, the tour rates for all purposes also increase with car ownership, with zero-vehicle households making fewer tours, particularly for the Commute purpose. Car competition (fewer cars than driving-age adults) also has a negative effect for most purposes, but not for Commute.
- Households with children tend to make more Commute and Leisure tours, but fewer tours for the other purposes.
- The higher the fraction of household adults that work, the more Commute and Business tours are made, and the fewer long-distance tours for the other purposes, particularly Personal Business.
- One-person households tend to make fewer Business, Leisure, and Personal Business tours.
- Households with the head age under 35 or over 65 tend to make fewer Commute, Business, and Personal Business tours.
- The accessibility LogSums from the aggregate mode/destination models generally show the results one would expect. The greater the accessibility to zones within 50 miles, the fewer long-distance tours are made to zones greater than 50 miles away, all else being equal. There are also larger rural zones (typically counties) for which there are no other zones accessible within 0–50 miles. The dummy variable for these zones is negative, compensating for the fact that those zones do not have the negative effect from the accessibility LogSum. (In future versions of this model, it may also be useful to test density variables for the residence zone.)
- The accessibility LogSum to all zones within the 50–150 mile range is positive and large for the Commute purpose, and positive with much smaller values for the Business and Personal Business LogSums. In contrast, it is the accessibility LogSum to all zones farther than 150 miles that have the positive effects for visit friends/relatives and Leisure, as those are the two purposes that tend to have the longest tours. The LogSum coefficients are typically about 0.04, which indicates some tour induction/suppression effect would be predicted in response to changes in accessibility, but this is not a major effect.
- The next set of variables capture higher tour rates for certain purposes in certain months, relative to the "base" month of May. Leisure (vacation) tours are higher in the summer months and lower in the fall and winter, while Visit tours are lower in the winter and fall (but not in November or December, presumably due to holiday visits). Commute-tour rates are somewhat lower in the summer months, while Business tours are highest between February and March and September and October, and lowest in December.

- The final effects in the models are shown for the lag time between the travel day and the time the respondent took the survey. The greater the number of days before the survey, the lower the tour rates for all purposes, presumably due to recall bias. For most purposes, both logarithmic and linear variables are significant in combination, while for Commute, only a logarithmic variable was significant. In model application, these variables will not be applied, assuming that the tour rates reported for the day immediately prior to the survey are the most accurate (having the least recall bias).

For the 33,000 or so household-days for which at least one long-distance tour was reported, there are about 2.3% where a second tour was also reported. As a result, a second model was estimated (Table 68) and used to predict what household-days for which one tour is predicted also make a second tour. (The number of household-days with three or more tours was negligible, so no more than two tours per day were modeled.) Compared to the main tour-generation and scheduling model in Table 67, there are fewer significant variables in the model of the second tour. One of the most significant variables for all purposes was a dummy variable indicating whether or not the first tour was for that same purpose, as most people who reported multiple tours tended to report them all for the same purpose. (Additional data checking may eliminate duplicate tour records in the data.) For all purposes except Commute and Leisure, the accessibility LogSum variables have a positive—and even stronger—effect for making a second tour. For Business and Leisure, higher income is related to making multiple tours in the day. Multiple Visit and Personal Business tours are related to the number of adults in the household, while multiple Commute tours are related to the number of workers in the household. In this model, the recall bias is only (marginally) significant for the Commute purpose, as there may have been a nonresponse bias against people reporting the same long-distance Commute multiple times.

Duration of Stay

The duration of stay model predicts which of the following four categories each tour falls into:

- 0 nights away (day tour, the base alternative with utility 0).
- 1–2 nights away.
- 3–6 nights away.
- 7+ nights away.

This aspect of the tour is modeled because it may influence the amount of distance that can be traveled and/or the mode used (e.g., day tours will tend either to be short distance, or to go by air for medium distances, and are rarely for longer-distance ranges over 1,500 miles one way). (Note that in the MDCEV version of the tour-generation/scheduling model, the tour duration will also be determined by each household’s long-distance tour time budget, and will therefore be part of that model.)

Table 68: Household-Day Tour-Generation Model—Second Tour in the Day

Tour Purpose	Commute		Business		Visit F&R		Leisure		Pers. Bus.	
	Coeff.	**T-stat**	**Coeff.**	**T-stat**	**Coeff.**	**T-stat**	**Coeff.**	**T-stat**	**Coeff.**	**T-stat**
Base alt. is no tour in the day										
Constant	-7.935	-14.5	-12.31	-6.0	-8.997	-8.5	-10.07	-7.6	-5.99	-6.4
Same purpose as first tour	3.751	9.6	3.930	11.3	1.473	10.2	2.20	12.3	1.487	10.5
Mode/dest LogSum 0-50 miles			-0.0579	-0.9	-0.0519	-0.9			-0.0742	-1.7
Mode/dest no options 0-50 miles			-0.621	-0.9	-0.764	-1.1			-0.859	-1.7
Mode/dest LogSum 50-150 miles			0.259	1.9					0.0637	0.7
Mode/dest LogSum over 150 miles					0.3507	2.8				
HH no of adults					0.175	2.2			0.272	3.6
HH no. of workers	0.356	1.5								
Log (income)			0.169	1.1			0.327	2.9		
Missing income data			2.10	1.2			3.556	2.6		
No. of days before survey	-0.0235	-1.9								
Statistics										
No. of tours (% of HH-days)	28	0.08%	127	0.38%	212	0.64%	199	0.60%	214	0.64%
Observations	**33,318**									
Rho-squared (0 coeff.)	0.926									
Rho-squared (constants)	0.093									

The results of this model in Table 69 show that even for the tour purposes that tend to have the longest distances and durations (Visit and Leisure, over 40% of tours are day tours, only 6–8% of tours stay away from home for seven nights or more. Some results shown in Table 69 are detailed below:

- Those with higher incomes tend to make longer tours away from home for all purposes, but particularly for Business and Leisure.
- Larger households tend to make shorter tours for Business, Visits, and Leisure.
- Those with a head of household age 65 or over tend to make fewer 1–2 night stays for all discretionary purposes, but make more 3–6 and 7+ night tours, presumably because they are not as constrained by weekday work schedules.
- Those with head of household age under 35 tend to make more 1–2 and 3–6 night tours.
- Those living in higher-density zones (based on the logarithm of jobs plus households per square mile), tend to make longer tours for all purposes except commuting. This may because they do not have to make as many long-distance day tours because they already have adequate opportunities within 50 miles, so they tend more to make the longer tours.
- The discretionary purposes tend to be of shorter duration in the winter months (January through March), with the exception of Leisure tours, which may be more likely to be 7+ nights in the winter.
- Visit and Leisure tours are more likely to be 3–6 nights in the summer months (June through August), and all purposes are more likely to be 7+ nights away in the summer.
- During the holiday months (November through December), Leisure tours tend to be somewhat shorter in duration, but visit friends/relatives tours tend to be longer, with positive effects on both 3–6 and 7+ nights.

Table 69: Tour Duration of Stay Models

Tour Purpose	Commute		Business		Visit F&R		Leisure		Pers. Bus.	
	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat	Coeff.	T-stat
Base alternative is day tour (0-nights)										
Alt = 1-2 nights										
Constant	-2.01	-27.1	-5.15	-10.8	0.0366	0.9	-4.05	-14.5	-3.42	-8.9
HH size			-0.038	-2.0	-0.0503	-4.4				
Log(income)			0.393	9.3			0.277	11.0	0.165	4.7
Missing income			4.59	9.4			2.99	10.4	1.77	4.5
HH head age 65+					-0.39	-8.4	-0.361	-8.1	-0.29	-4.4
HH head age<35					0.338	8.8	0.173	4.6	0.477	7.6
Log(res+emp density)							0.0904	10.0	0.0989	7.8
Jan-Mar					-0.0831	-2.2	-0.108	-2.8	-0.333	-5.6
Nov-Dec							-0.234	-5.5	-0.224	-3.3
Alt = 3-6 nights										
Constant	-6.91	-3.9	-6.89	-11.9	-2.12	-6.7	-6.77	-20.1	-4.52	-7.8
HH size			-0.0922	-3.9	-0.139	-8.7	-0.0595	-4.6		
Log(income)	4.79	2.6	5.73	9.4	1.02	3.1	4.65	13.2	1.84	3.1
Missing income	0.404	2.5	0.483	9.2	0.0888	3.1	0.418	13.5	0.149	2.9
HH head age 65+					0.135	2.5			0.238	2.7
HH head age<35					0.182	3.4			0.341	3.4
Log(res+emp density)			0.0648	4.2	0.0695	6.6	0.179	16.5	0.142	7.6
June-Aug					0.295	6.1	0.614	15.1		
Jan-Mar					-0.171	-3	-0.135	-2.5	-0.487	-5.2
Nov-Dec					0.379	7.6	-0.176	-2.9	-0.306	-2.9
Alt = 7+ nights										
Constant	-3.99	-19	-8.83	-9	-2.46	-18.4	-9.4	-20.9	-6.76	-7.4
HH size			-0.0936	-2.3	-0.195	-7.8	-0.0752	-4.3		
Log(income)			5.96	5.8			6.75	14.6	3.26	3.5
Missing income			0.494	5.5			0.596	14.6	0.281	3.4

Tour Purpose	Commute		Business		Visit F&R		Leisure		Pers. Bus.	
HH head age 65+					0.3	4.1	0.243	3.9	0.378	3
HH head age<35					-0.213	-2.4				
Log(res+emp density)			0.146	5.6	0.134	8.6	0.169	12.5	0.0817	2.8
June-Aug	1.13	3.5	0.302	2.8	0.562	8.4	0.661	12.2	0.295	2.7
Jan-Mar							0.353	5.5		
Nov-Dec					0.359	4.8	-0.25	-3.0		
Statistics										
Total Observations	1933	%	9899	%	21922	%	26748	%	12235	%
Away 0 nights	1543	79.8	5528	55.8	9005	41.1	12838	48.0	8228	67.2
Away 1-2 nights	207	10.7	2423	24.5	8025	36.6	7352	27.5	2631	21.5
Away3-6 nights	143	7.4	1498	15.1	3561	16.2	4234	15.8	989	8.1
Away 7+ nights	40	2.1	450	4.5	1331	6.1	2324	8.7	387	3.2
Rho-squared(0 coeff)	0.504		0.22		0.148		0.147		0.354	
Rho-square(constants)	0.007		0.013		0.014		0.024		0.015	

Party Size

The party-size model predicts the number of members (including nonhousehold participants) in the travel party. The base alternative is one person traveling alone, while the other alternatives are 2, 3, or 4+ persons for Commute and Business, and 2, 3, 4, 5, or 6+ persons for the other purposes. These differences exist because Commute and Business tours were rarely observed to have a party size greater than four.

The model results, shown in Table 70, indicate the following:

- By far, the largest positive effect, applied to all alternatives, is when the party size is equal to the household size, indicating that many tours are made by all household members. This effect is smallest for Commute and Business, but it is still significant.
- Most purposes (except Business) have a counteracting negative effect when the household size equals the number of adults. This variable only has an effect when the household has children (otherwise it is identical to the previous variable), so it indicates that households with children are not as likely to have the adults travel without the children.
- A higher income tends to result in tours with smaller party size for Business and Commute, but has no effect on the other purposes.
- The more workers in the household, the smaller the party size for all purposes except Commute. This may be because one or more of the workers has to stay home and work.
- In general, higher car ownership tends to increase party size for most purposes—presumably because the marginal travel cost per person is lowest by car—but this effect does not appear to hold for the Business and Commute purpose.
- For Business and Leisure, tours of longer duration away from home tend to have larger party sizes, but the opposite appears true for Visit and Personal Business tours.
- Tours in the summer months tend to have larger party sizes for all discretionary purposes.

Table 70: Tour-Party-Size Models

Purpose	Commute		Business		Visit F&R		Leisure		Pers.Bus.	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
All alternatives (base is Party size = 0)										
Party size = household size	0.127	2.39	0.594	24.40	1.870	118.03	1.390	103.90	1.100	54.79
Party size = household adults	-0.115	-2.36			-0.580	-28.70	-0.319	-17.49	-0.142	-6.07
Alternative - party size = 2										
Constant	3.380	5.26	2.290	6.11	0.859	19.47	1.070	10.07	-0.086	-0.70
Workers / Household size	0.941	8.25	-0.281	-4.78	-0.668	-15.09	-0.288	-5.62	-0.477	-8.59
Log (Income)	-0.549	-9.50	-0.232	-7.49						
Missing income data	-5.810	-8.93	-2.650	-7.46						
HH has 1 vehicle			-0.697	-4.17			0.360	3.50	0.824	6.77
HH has 2 vehicles	0.385	3.31	-0.857	-5.25			0.261	2.56	0.688	5.64
HH has 3+ vehicles	1.040	9.25	-0.801	-4.89			0.457	4.46	0.805	6.59
1 to 2 nights away			0.479	8.97	-0.240	-4.98	0.163	4.28		
3 to 6 nights away			0.351	5.28	-0.526	-8.97			-0.145	-1.95
7 or more nights away			0.535	4.74	-0.586	-7.12			-0.372	-3.05
Missing duration data					-0.738	-16.75	-0.409	-10.05	-0.465	-9.43
Spring (Apr-Jun)	-0.470	-3.22	0.116	2.31			-0.098	-2.77		
Summer (Jul-Sep)	-0.434	-3.03							0.285	5.67
Fall (Oct-Dec)			0.125	2.64	0.157	4.05	-0.234	-5.24	0.191	3.99
Missing month data	-0.380	-4.77	-0.297	-5.82					-0.259	-4.44
Alternative - party size = 3										
Constant	2.880	2.29	4.730	7.87	0.075	1.32	0.697	11.16	-0.194	-2.68
Workers / Household size					-1.010	-16.52	-0.411	-6.58	-0.545	-7.37
Log (Income)	-0.456	-3.90	-0.559	-10.94						
Missing income data	-6.070	-4.38	-6.520	-11.11						

Purpose	Commute		Business		Visit F&R		Leisure		Pers.Bus.	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
HH has 1 vehicle	-2.200	-5.83	-1.060	-4.06						
HH has 2 vehicles	-1.370	-5.47	-0.758	-3.05			-0.191	-3.78	0.410	6.02
HH has 3+ vehicles	-1.100	-4.37	-0.845	-3.37			-0.227	-4.19	0.277	3.78
1 to 2 nights away			0.569	6.40	-0.181	-3.18				
3 to 6 nights away			0.436	3.87	-0.623	-8.55				
7 or more nights away			0.509	2.57	-0.801	-7.36	-0.205	-2.90	-0.669	-3.76
Missing duration data					-0.827	-13.71	-0.403	-8.58	-0.664	-10.40
Spring (Apr-Jun)									-0.274	-4.61
Summer (Jul-Sep)			-0.345	-3.82	0.162	3.37	0.173	4.31	0.203	3.30
Fall (Oct-Dec)			-0.204	-2.43	0.349	6.83	-0.228	-4.24		
Missing month data			-0.721	-7.76	0.283	3.94			-0.370	-4.57
Alternative - party size = 4										
Constant	12.400	13.62	4.120	7.77	0.000	0.01	0.343	2.28	-0.807	-10.24
Workers / Household size			-0.246	-2.31	-1.260	-18.40	-0.699	-10.96	-0.516	-5.99
Log (Income)	-1.380	-16.55	-0.549	-11.17						
Missing income data	-15.500	-16.31	-6.610	-11.64						
HH has 1 vehicle	-1.340	-6.77					0.316	2.13	0.261	3.11
HH has 2 vehicles	-1.960	-10.94	-0.390	-4.22			0.473	3.27	0.417	6.59
HH has 3+ vehicles	-2.980	-12.44	-0.467	-4.78			0.396	2.72		
1 to 2 nights away			0.563	6.51	-0.284	-4.86	0.157	3.39	0.209	3.06
3 to 6 nights away			0.790	8.12	-0.689	-9.07	0.144	2.96		
7 or more nights away			0.880	5.38	-0.941	-8.11			-0.652	-3.22
Missing duration data					-1.000	-15.44	-0.549	-10.25	-0.881	-13.53
Spring (Apr-Jun)	1.240	2.90					-0.097	-1.88		
Summer (Jul-Sep)	1.640	4.23	0.151	2.04	0.207	4.09	0.203	4.47	0.510	7.60

Purpose	Commute		Business		Visit F&R		Leisure		Pers.Bus.	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Fall (Oct-Dec)	2.280	6.31			0.428	8.01	-0.343	-5.78	0.132	1.89
Missing month data	1.130	3.43	-1.010	-10.16	0.339	4.31	-0.358	-5.04		
Alternative - party size = 5										
Constant					-0.300	-4.27	0.250	3.90	-1.560	-13.39
Workers / Household size					-1.640	-18.62	-1.260	-15.79	-1.200	-9.70
Log (Income)										
Missing income data										
HH has 1 vehicle							-0.331	-4.57		
HH has 2 vehicles									0.452	4.12
HH has 3+ vehicles									0.477	4.08
1 to 2 nights away					-0.254	-3.55	0.144	2.32	0.430	4.49
3 to 6 nights away					-0.687	-7.17	0.238	3.60		
7 or more nights away					-0.934	-6.22	0.166	1.95		
Missing duration data					-1.070	-12.69	-0.543	-8.39	-0.559	-6.34
Spring (Apr-Jun)										
Summer (Jul-Sep)					0.232	3.58	0.362	7.20	0.653	7.17
Fall (Oct-Dec)					0.272	3.97	-0.280	-4.14	0.303	3.11
Missing month data					0.278	2.58				
Alternative - party size = 6+										
Constant					-0.490	-5.99	0.349	5.58	-1.040	-9.99
Workers / Household size					-2.170	-19.34	-1.100	-14.25	-1.150	-10.39
Log (Income)										
Missing income data										
HH has 1 vehicle										
HH has 2 vehicles									0.491	4.68
HH has 3+ vehicles									0.750	6.96

Purpose	Commute		Business		Visit F&R		Leisure		Pers.Bus.	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
1 to 2 nights away					-0.448	-5.11	0.109	1.79	0.358	4.01
3 to 6 nights away					-0.912	-7.76	-0.190	-2.56	0.262	2.01
7 or more nights away					-1.040	-5.75	0.264	3.32		
Missing duration data					-1.310	-12.27	-0.807	-10.85	-1.180	- 10.00
Spring (Apr-Jun)										
Summer (Jul-Sep)					0.364	4.45	0.290	5.47		
Fall (Oct-Dec)					0.447	5.26	-0.229	-3.38	-0.231	-2.57
Missing month data					0.393	2.77	-0.229	-2.23	0.348	2.62
Statistics										
Total observations	8905		18985		31763		37097		19166	
Adjusted rho-square	0.56		0.34		0.30		0.22		0.21	

Destination and Mode Choice

After all of the national-level network zone-to-zone data were created and processed by all modes, the mode/destination choice models that had previously been estimated on the data from the California statewide model mode zones and networks and data from the 2013 California statewide survey were re-estimated using the national-level zones and network data and the combined data from the four long-distance surveys listed earlier. These models, using less-detailed spatial data (roughly 4,500 zones for the entire United States vs. 5,700 zones just for California) have thus far not provided satisfactory estimates of time and cost coefficients, values of time, and other parameters. While work is continuing on refining those models, and on estimating CNL versions of those models, the approach used for the initial model application was simply to apply the mode/destination choice models estimated on the California statewide model data, while rescaling them and calibrating them to match the choice shares in the larger survey dataset. Table 71 gives the mode shares and distance-band distribution of the tours in the larger dataset, by tour purpose. Scale factors were applied to the utilities from the previous models, and additional mode-specific calibration constants and distance calibration terms, in order to match the observed shares fairly closely when the models are applied to the four-state estimation dataset. This work will be refined as the calibration and validation of the model system goes forward, as discussed in Chapter 6.

Table 71: Mode-Choice and Distance-Band Distribution by Tour Purpose

Distance-band	Auto	Bus	Rail	Air
Business				
50-150 miles (1-way)	60.3%	0.4%	1.2%	0.2%
150-350 miles (1-way)	16.5%	0.2%	0.6%	1.2%
350+ miles (1-way)	4.7%	0.2%	0.1%	14.4%
Commute				
50-150 miles (1-way)	80.9%	1.1%	9.8%	0.0%
150-350 miles (1-way)	5.3%	0.4%	0.2%	0.0%
350+ miles (1-way)	1.1%	0.0%	0.0%	1.3%
Visiting friends and relatives				
50-150 miles (1-way)	59.6%	0.4%	0.5%	0.1%
150-350 miles (1-way)	22.8%	0.2%	0.3%	0.3%
350+ miles (1-way)	8.5%	0.1%	0.1%	7.2%
Leisure				
50-150 miles (1-way)	63.0%	1.7%	0.5%	0.1%
150-350 miles (1-way)	21.3%	0.8%	0.1%	0.2%
350+ miles (1-way)	6.9%	0.2%	0.1%	5.0%
Personal business				
50-150 miles (1-way)	71.0%	1.1%	0.8%	0.0%
150-350 miles (1-way)	18.2%	0.6%	0.2%	0.2%
350+ miles (1-way)	4.9%	0.2%	0.0%	2.7%

Software Performance

The initial implementation, using the modified disaggregate structure of Figure 47, was programmed in Delphi Pascal, which is an efficient language similar to C++. Running on a desktop computer and using only a single processor and about 4 GB of RAM, the application simulates long-distance tours for one representative day in all 12 months of the year for the full synthetic population of 107 million households in a run time of 45 minutes. That run is shown as Run 1 in Table 72. (This run produced about 75 million individual tour records, with an average expansion factor of about 30, or about 2.42 billion tours when expanded.) Therefore, run time does not appear to be a limiting factor with the current structure, although it may become so if more complex model components or structures are added into the system.

One way to expedite the simulation is to subsample households. For example, Run 2 in Table 72 uses a 1 in 10 random sample of households (simulating about 10.65 million individual household records, but about the same number of expanded households), which reduces the run time to approximately eight minutes. Run 3, using a 1 in 100 subsample (1.06 million households records) takes roughly four minutes to run. A potential problem with the subsampling approach is that there will be fewer O-D pairs that are represented in the forecast. The bottom of the Table 72 shows that when a 1 in 100 sample is used, only 459,000 zone pairs are represented on the output tour records, or an average of only about 100 different destinations for each of approximately 4,500 origin zones. This is only about one tenth of the number of zone pairs that are covered when the full synthetic population is used.

Table 72: Statistics from Example Model Application Runs

Example Run	1	2	3	4	5
Sampling rate	Full	1 in 10	1 in 100	1 in 100	1 in 10
Months simulated	12	12	12	12	12
Day simulated/month	1	1	1	All	All
Households simulated	106,419,862	10,641,986	1,064,198	1,064,198	10,641,986
Expanded households	112,589,291	112,588,010	112,548,700	112,548,700	112,588,010
Tour records output	75,229,535	7,521,817	752,488	22,875,293	228,799,586
Expanded tour records out	2,421,098,862	2,420,381,460	2,422,324,200	2,420,152,600	2,41,349,490
Tour records output/HH	0.71	0.71	0.71	21.50	21.50
Expanded tours per HH	21.50	21.50	21.52	21.50	21.51
Unique tour O-D zone pairs	4,094,744	1,705,799	459,045	2,710,514	5,647,454
Run time	45 minutes	8 minutes	4 minutes	12 minutes	95 minutes

An efficient way of generating more tours and more spatial variation is to run the tour-generation model separately for each day of each simulated month, rather than just running it for one representative day in each month and expanding the tours to represent the entire month. This means that the tour-generation model is run 365 times for each simulated household, once for each day of the year. Run 4 is identical to Run 3, except that it uses the approach of running the tour-generation model for each day of each month. Compared to Run 3, the number of tour records generated is about 30 times larger (22.9 million tour records), and the number of different O-D pairs represented is about six times larger (2.7 million distinct O-D zone pairs). The run time only increases from 4 to about 12 minutes. Finally, Run 5 uses this approach with a 1 in 10 sample of households, producing 228.8 million output tour records, and covering 5.65 million different O-D pairs (about 1,250 different destination zones for each of the 4,500 origin zones. The run time is approximately 95 minutes, which is lower than expected for a microsimulation of long-distance travel in the United States.

One could also use the approach of Runs 4 and 5 with the full synthetic sample. At that point, however, the program would be generating over 2 billion individual tour records, and would take approximately 10 times as long as Run 1, or 15 hours. Not only would an output file with over 2 billion records be difficult to work with in analyzing the results, but it would require over 100 GB of disk space. It is also not clear that simulating 2.2 billion tours would provide drastically different expanded output versus simulating over 220 million tours, which is already a large number. There are also programming options to store and access the output tour records more efficiently using binary file formats, although that would make the output more difficult to use for general users as opposed to using a standard text-based output format.

In the near future, the model application code will be translated into C#, which is a more common modern language. C# allows relatively simple programming of multithreading across multiple processors, which will help to minimize run time, and also allows efficient management of memory to use larger amounts of RAM when available. A GUI can also be programmed to make use of the model system fairly simple for novice users.

CHAPTER 6. PERFORMANCE METRICS

For the demonstration of the national long-distance passenger travel demand forecasting model, a sample of performance metrics have been developed to show what types of data may be derived and how these may be interpreted for planning studies. These results are not based on a calibrated or validated model and do not, therefore, represent model output and should not be interpreted as accurate or final results.

The demonstration model was run initially to simulate travel for the month of October 2010; a sample of model results has been produced. The simulation model can also produce outputs for every month in the year, which can then be aggregated to produce annual results. The annual scheduling models described in Chapter 3 will simulate tours across the entire year in a more simultaneous manner, rather than simulating each month separately.

6.1 Travel Metrics

Modes

Modal performance metrics support a wide variety of planning activities and are used to evaluate modal investments. These can be produced by state, region, corridor, or zone and provide consistent evaluations of modal investments across the United States. Mode shares for person-tours and person-miles traveled are presented in Table 73. The auto mode has the highest mode share for both person-tours and person-miles traveled, but also tends to have more and shorter distance tours, resulting in a reduction in mode share for person-miles traveled. As expected, the person-miles traveled for the air mode increase significantly over the person-tours mode share for air. Bus and rail person-miles traveled mode shares also increase over person-tours mode shares for these modes, but to a lesser degree than air.

Table 73: Person-Tours and Person-Miles Traveled by Mode for October

Mode	Person-Tours	Tour Shares	Person-Miles Traveled	PMT Shares
Auto	162,942,200	89.3%	110,656,651,400	78.5%
Bus	2,548,200	1.4%	2,366,378,800	1.7%
Rail	3,031,800	1.7%	2,532,631,800	1.8%
Air	14,030,600	7.7%	25,391,824,100	18.0%

Table 74 presents cost, travel time, and tours by mode as a function of distance, tours, and households, respectively. These metrics allow a more direct comparison across modes of cost, time, and travel. Average cost per mile metrics show that air is the most expensive mode, approximately three times as expensive as rail and five times as expensive as auto. This cost is a tradeoff with average travel time by mode, so air has the fastest travel times per tour. (Air, rail, and bus times do not include access and egress times to/from the station or airport, or the time in the airport or station waiting for the first departure, but they do include an estimate of transfer time for routes). Bus and rail tend to be competitive for longer tours, so their travel times per tour are longer than either auto or air. Travel times are reported as tours, so auto tours average 360 minutes (6 hours), or three hours each way. In October 2010, households took an average of 1.45 tours by auto; only 1 in 8 households took an air tour; only 1 in 33 households took a rail tour; and only 1 in 50 households took at bus tour.

Table 74: Average Cost, Travel Time, and Tours by Mode for October

Mode	Average Cost per Mile	Average Time per Tour (minutes)	Average Tours per Household
Auto	$0.15	360	1.45
Bus	$0.16	581	0.02
Rail	$0.26	523	0.03
Air	$0.76	192	0.12

Tour Purpose

The purpose of activities undertaken on a long-distance tour is a significant driver for travel behavior and is therefore important when trying to understand the source of long-distance travel on the national scale. Table 75 presents the person-tours and person-miles traveled for October. In October, personal business was the largest portion of travel, with significant person-tours for visiting friends and relatives and leisure/vacation purposes. Leisure/vacation and employer's business tours are longer tours, evidenced by the increase in person-miles traveled shares for these purposes, and personal business tours tend to be shorter tours.

Table 75: Person-Tours and Person-Miles Traveled by Purpose for October

Tour Purpose	Person-Tours	Tour Shares	Person-Miles Traveled	PMT Shares
Personal Business	73,420,400	40.2%	44,028,726,500	31.2%
Visit Friends and Relatives	39,906,300	21.9%	27,913,280,400	19.8%
Leisure/Vacation	37,534,800	20.6%	37,469,228,800	26.6%
Commute	11,931,900	6.5%	9,865,204,400	7.0%
Employer's Business	19,759,400	10.8%	21,671,046,000	15.4%

Table 76 presents cost, travel time, and tours by mode as a function of distance, tours, and households, respectively. Average cost per mile metrics show that employer's business is the most expensive purpose, but only slightly higher than personal business. Leisure/vacation is the lowest cost per mile, possibly because these tours tend to be longer and travelers may be cost conscious for this type of discretionary travel. This cost is a tradeoff with average travel time by mode, so air has the fastest travel times per tour.

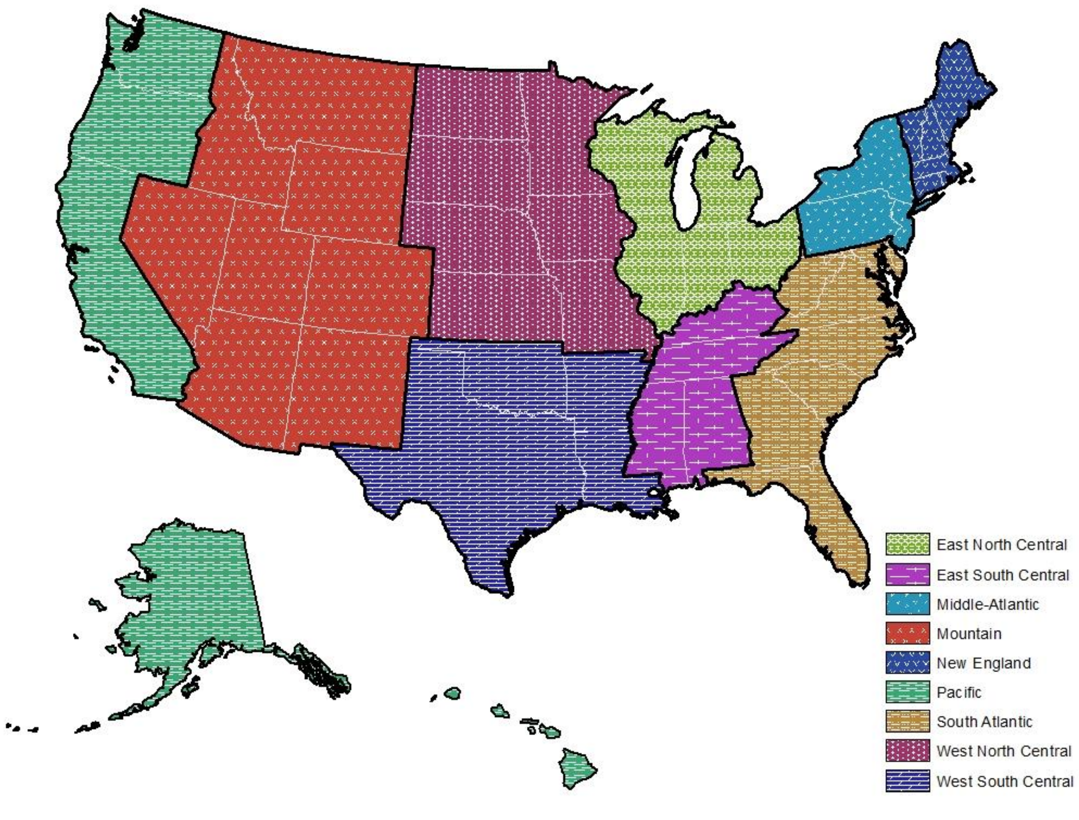
Table 76: Average Cost, Travel Time, and Tours by Purpose for October

Tour Purpose	Average Cost per Mile	Average Time per Tour (minutes)	Average Tours per Household
Personal Business	$0.29	338	0.65
Visit Friends and Relatives	$0.23	388	0.35
Leisure/Vacation	$0.20	367	0.33
Commute	$0.22	312	0.11
Employer's Business	$0.34	345	0.18

Destinations

Destinations are an important aspect of national long-distance travel. These are represented in this context by regions established by the US Census Bureau,[15] as shown in Figure 48. The simulation data output from the long-distance model is available to aggregate in many ways, so these regions are just one example of how destinations can be aggregated for reporting.

Figure 48: Regions in the US Census



[15] The Census Bureau refers to these regions as Divisions, with larger aggregations of these Division as Regions.

Figure 49 shows the total person-tours in October, by region. In this example, the South Atlantic region has the highest travel demand for long-distance travel and New England has the lowest travel demand. This travel demand may vary by month, but it is also likely affected by a combination of density of attractions and population.

Figure 49: Total Person-Tours in October, by Region

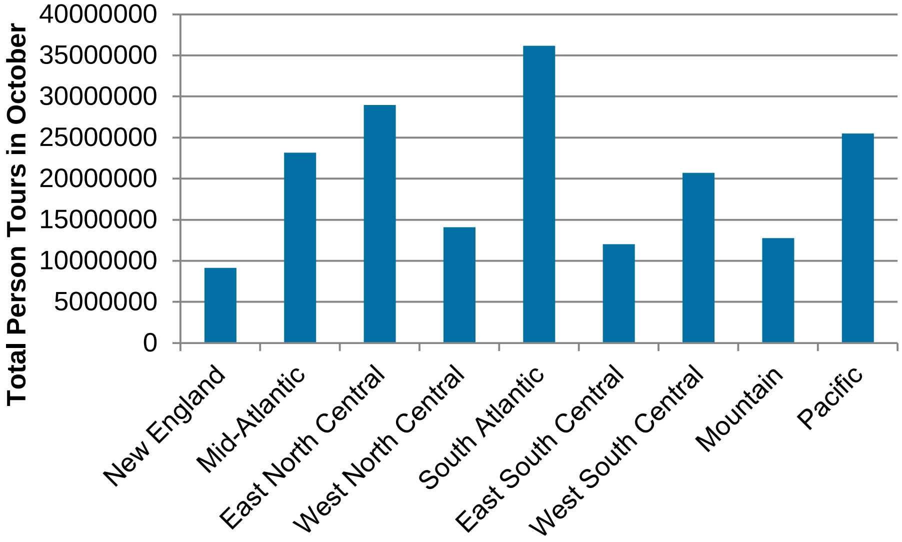


Table 77 presents an O-D matrix of person-tours in October to and from each region across the United States. This matrix demonstrates that the vast majority of long-distance travel in the United States is within a single region, with the Pacific region retaining the highest percentage of long-distance travel (94%) and the East-South Central region retaining the least (50%).

Table 77: Region-to-Region Distribution of Person-Tours in October

Home Destination Region	New England	Mid-Atlantic	East-North Central	West-North Central	South Atlantic	East-South Central	West-South Central	Mountain	Pacific
New England	57%	39%	1%	0%	4%	0%	0%	0%	0%
Mid-Atlantic	13%	62%	7%	0%	17%	1%	0%	0%	0%
East-North Central	0%	5%	77%	7%	4%	6%	0%	0%	0%
West-North Central	0%	0%	19%	68%	1%	3%	6%	2%	0%
South Atlantic	1%	10%	3%	0%	78%	7%	1%	0%	0%
East-South Central	0%	1%	15%	3%	23%	50%	8%	0%	0%
West-South Central	0%	0%	1%	5%	1%	7%	84%	2%	0%
Mountain	0%	0%	0%	3%	0%	0%	4%	73%	20%
Pacific	0%	0%	0%	0%	0%	0%	0%	6%	94%

Travel Time

Travel times for long-distance passenger travel offer a means to understand accessibility of households across the United States. In areas where there are ample opportunities for Business and Leisure activities, one would expect travel times per tour to be less than in areas where there are fewer opportunities nearby for these activities. Figure 50 presents the travel time per tour by origin zone and demonstrates that shorter travel times per tour are associated with higher-density areas and more opportunities for activities, and longer travel times per tour are associated with lower-density areas and fewer opportunities for activities. As expected, total travel time per person, presented in Figure 51, also shows similar trends. That is, individuals living in areas where there are more opportunities for activities spend relatively less time making long-distance tours.

Figure 50: Travel Time per Tour by Origin NUMA

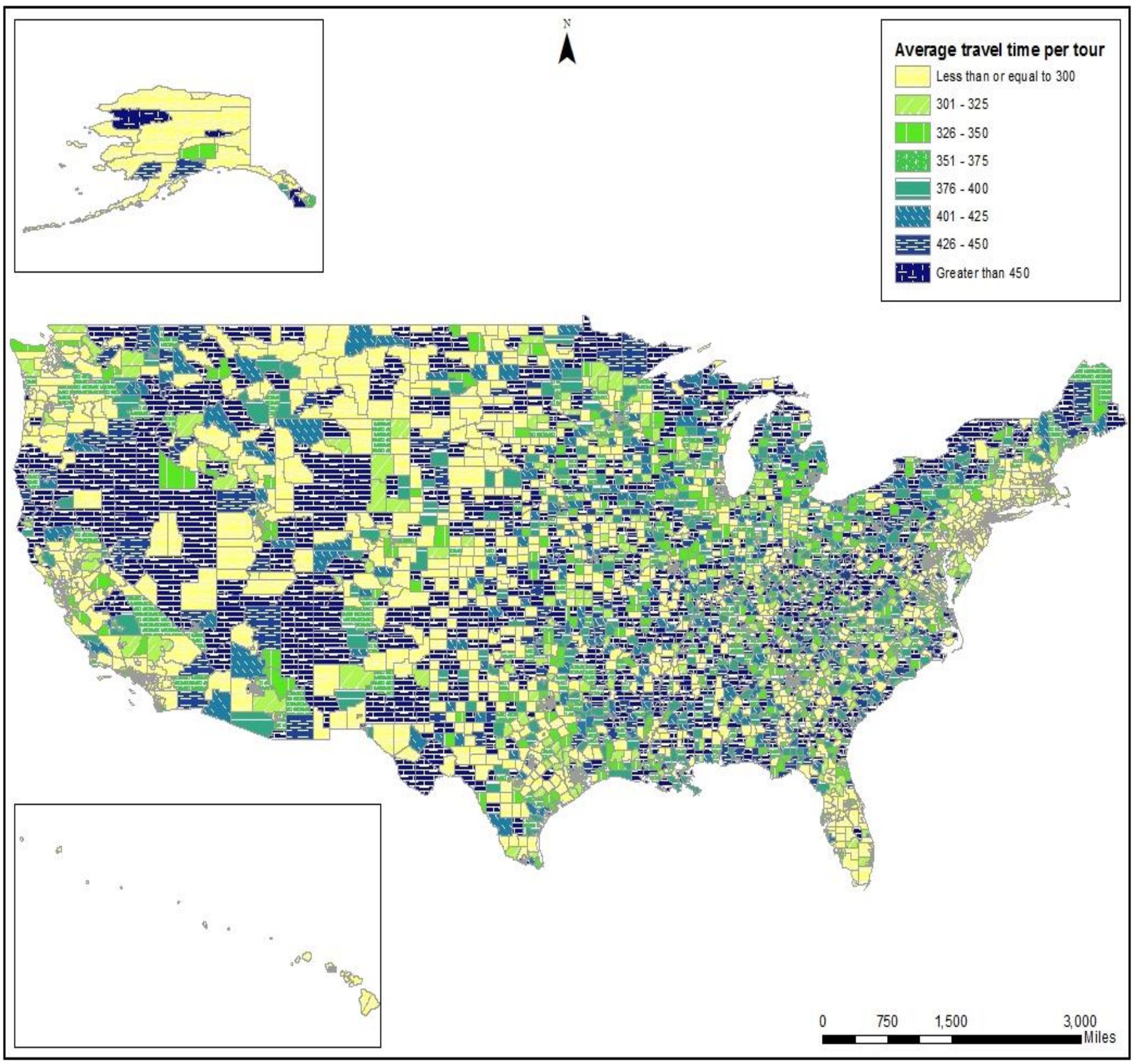


Figure 51: Travel Time per Person by Origin NUMA

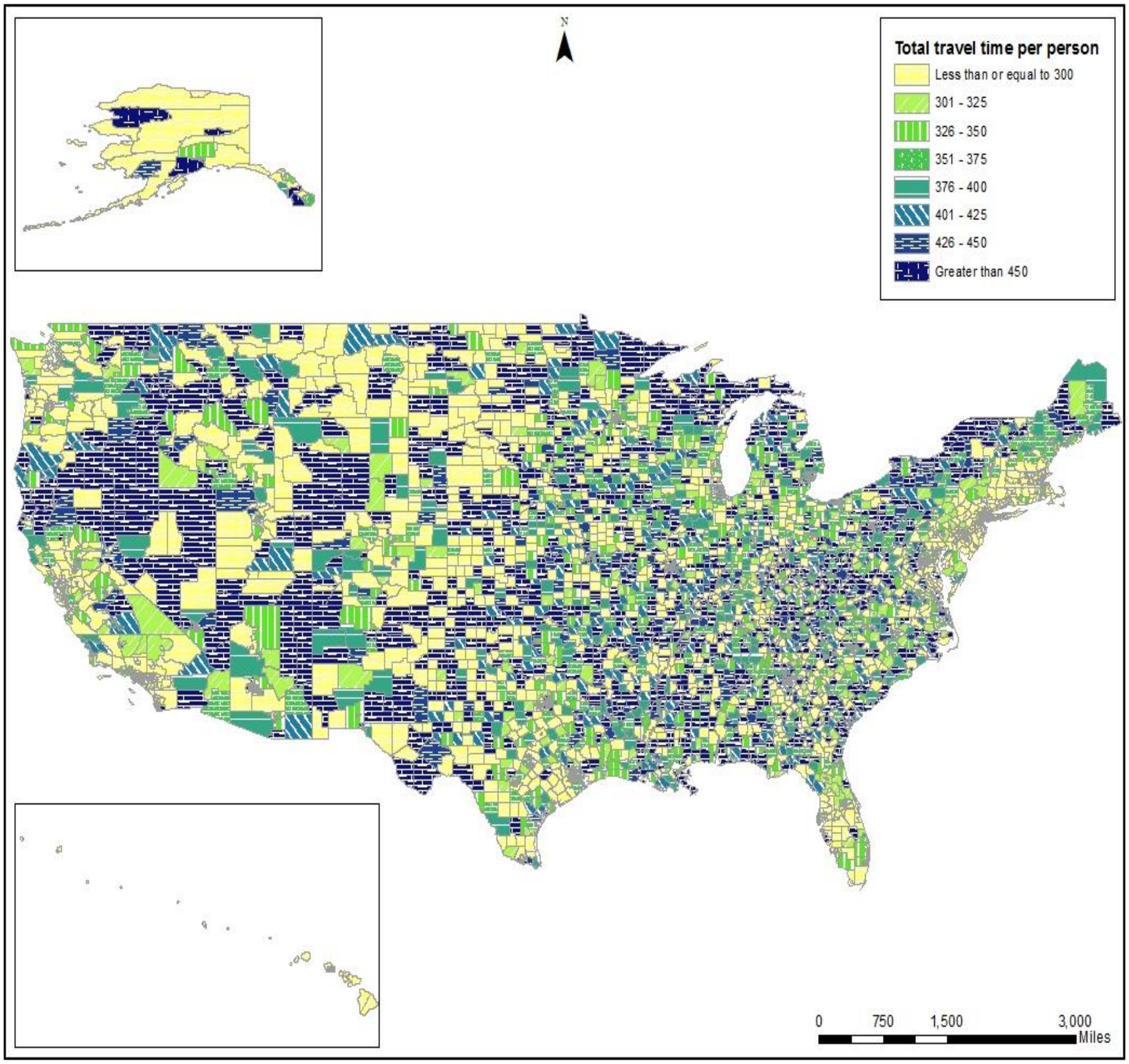


Table 78 presents the average travel time from region to region in October. Some O-D pairs do not have any person-tours represented and therefore have no travel times in this table (e.g., Pacific region to New England region). While there is some correlation between higher travel demand and lower travel times, there are also some destinations that have a higher demand with relatively long travel times. For example, the Mid-Atlantic region is closer to New England, but has a higher demand to the South Atlantic region. (Note that this example simulation was performed for just one day, on a 1 in 100 subsample of households. A more extensive simulation that simulated more days and covered more O-D pairs, as discussed earlier, avoids the issue of zero tours in some cells.)

Table 78: Average Travel Time from Region to Region in October

	New England	Mid-Atlantic	East-North Central	West-North Central	South Atlantic	East-South Central	West-South Central	Mountain	Pacific
New England	226	347	630	473	360	426			
Mid-Atlantic	335	262	525	727	391	813	447		
East-North Central	647	477	329	507	621	522	648	332	671
West-North Central	459	1401	491	364	506	635	531	768	
South Atlantic	469	393	618	823	321	499	686		
East-South Central	1009	871	498	591	471	326	522	451	666
West-South Central	539	460	763	567	643	524	352	509	573
Mountain		514	546	661	524		603	353	478
Pacific				545			435	442	266

Demographics

Households that are larger or smaller in size tend to travel less in terms of overall travel and distance, as shown in Figure 52. The largest difference in travel metrics is seen in one-person households.

Figure 52: Long-Distance Travel Metrics in October, by Household Size (Average Tours per Household)

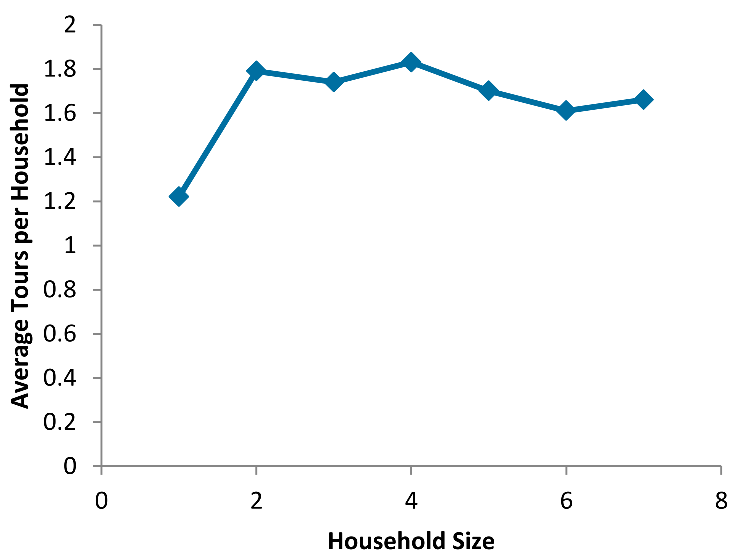
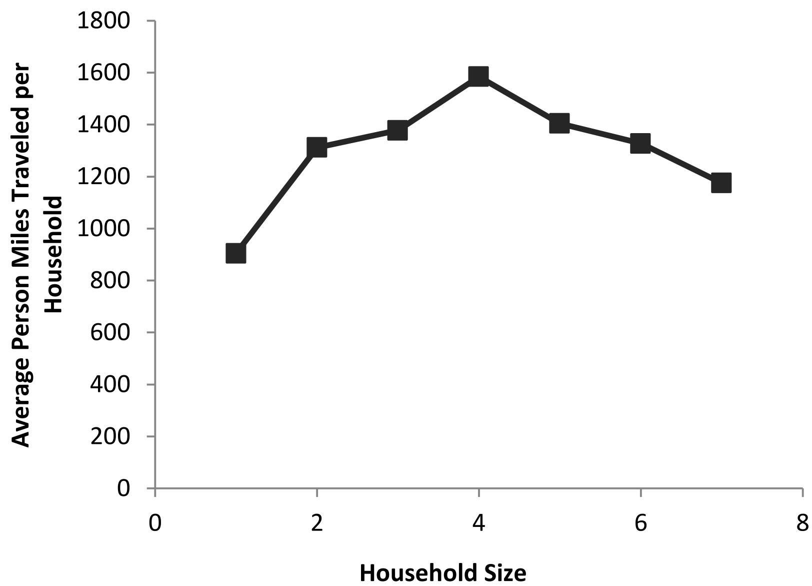


Figure 53: Long-Distance Travel Metrics in October, by Household Size (Average Person Miles Traveled per Household)



6.2 Environmental, Economic, Livability, Safety Metrics

The majority of environmental, economic, livability, and safety metrics require an additional method or model that processes the travel outputs from the long-distance passenger travel demand model. These additional methods have not been deployed for this demonstration project, but they include air quality models, economic impact, benefit-cost analyses, safety models, and health impact models.

Distribution of Miles Traveled

One travel metric that provides insight into these additional metrics is the distribution of person-miles traveled by mode. Figure 54 through Figure 57 present the distribution of person-miles traveled in October for auto, air, rail, and bus, respectively. The number of households traveling by rail and bus modes peak at a distance of approximately 200 miles, and the number of households traveling by air modes peak at a distance of approximately 400 miles. This is in contrast to the number of households traveling by auto, which peaks at the minimum distance of approximately 100 miles. These person-miles traveled represent a household's travel over one full month and could include multiple tours and/or multiple travelers making the same tour.

Figure 54: Distribution of Person-Miles Traveled in October, by Auto

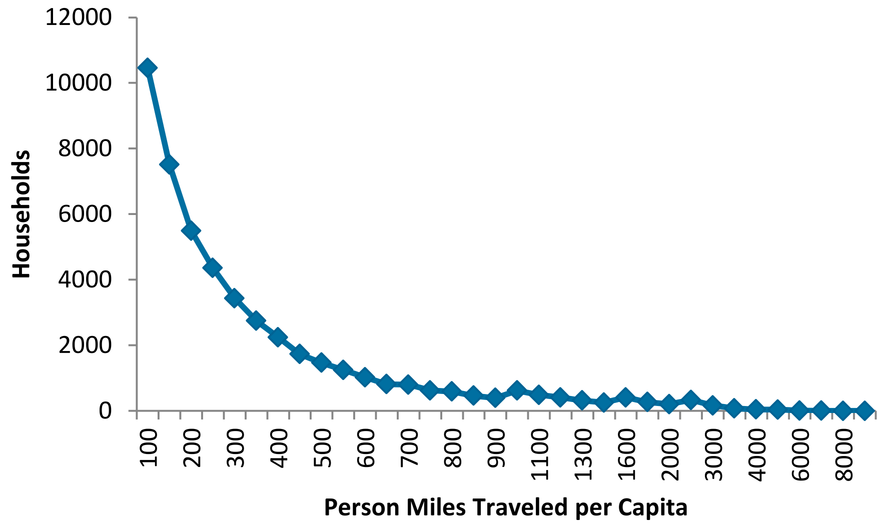


Figure 55: Distribution of Person-Miles Traveled in October, by Air

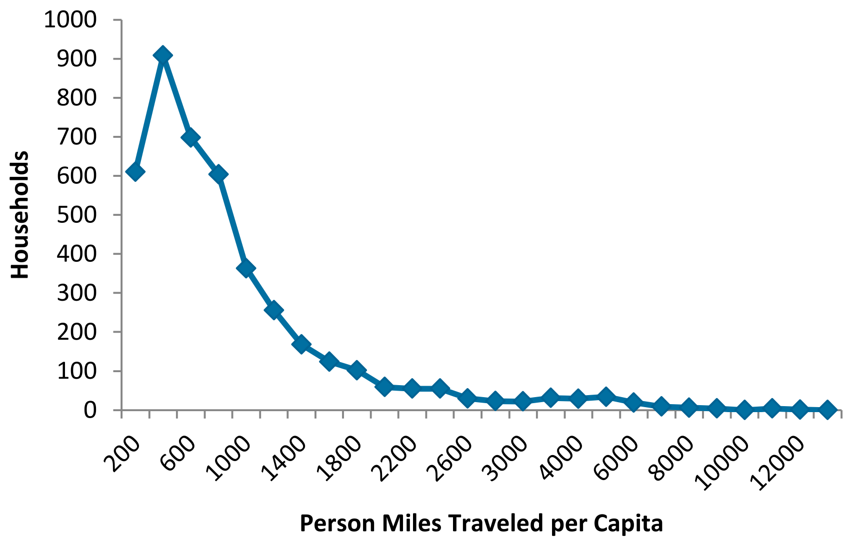


Figure 56: Distribution of Person-Miles Traveled in October, by Rail

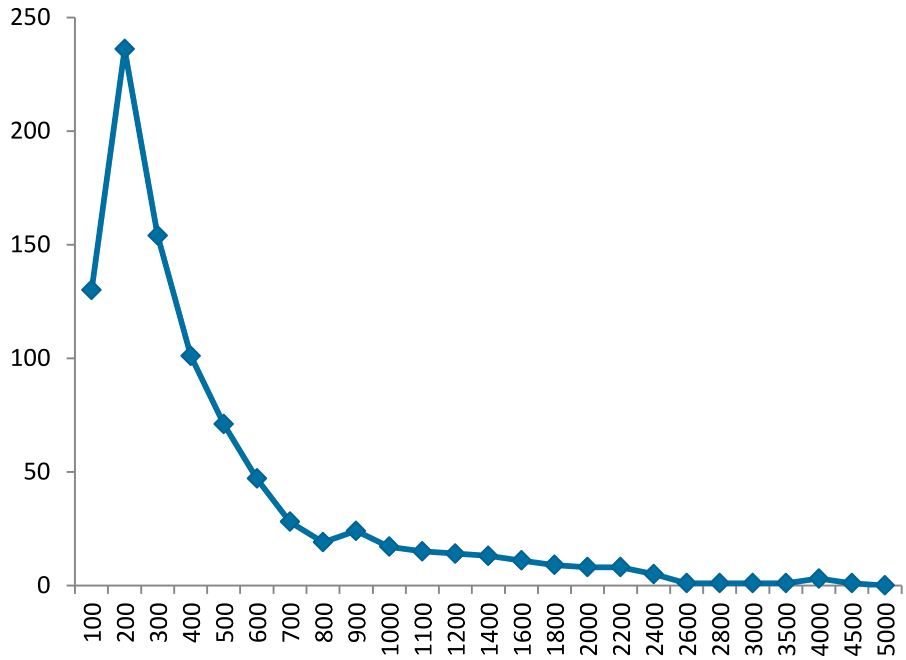
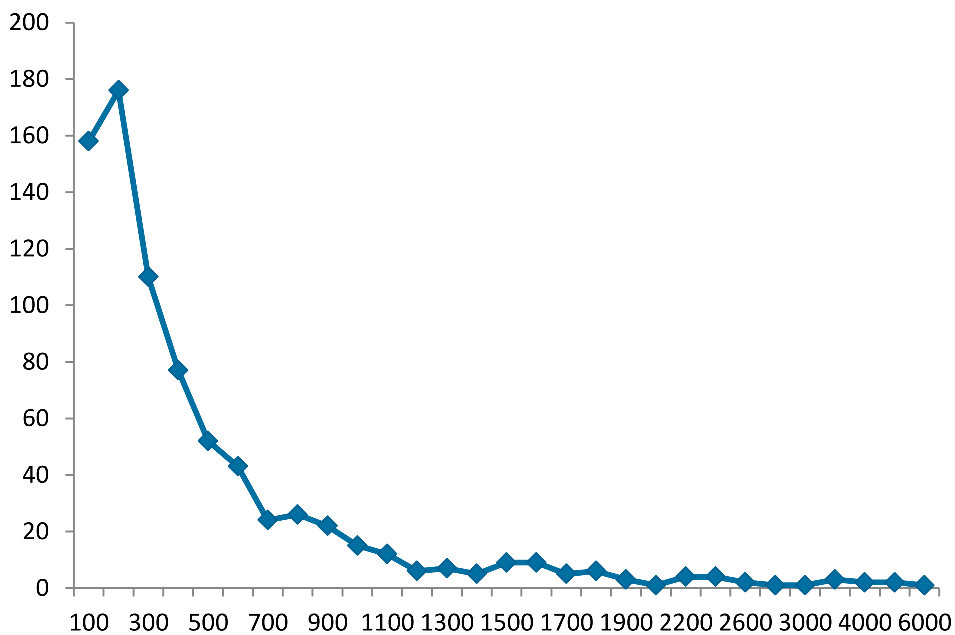


Figure 57: Distribution of Person-Miles Traveled in October, by Bus



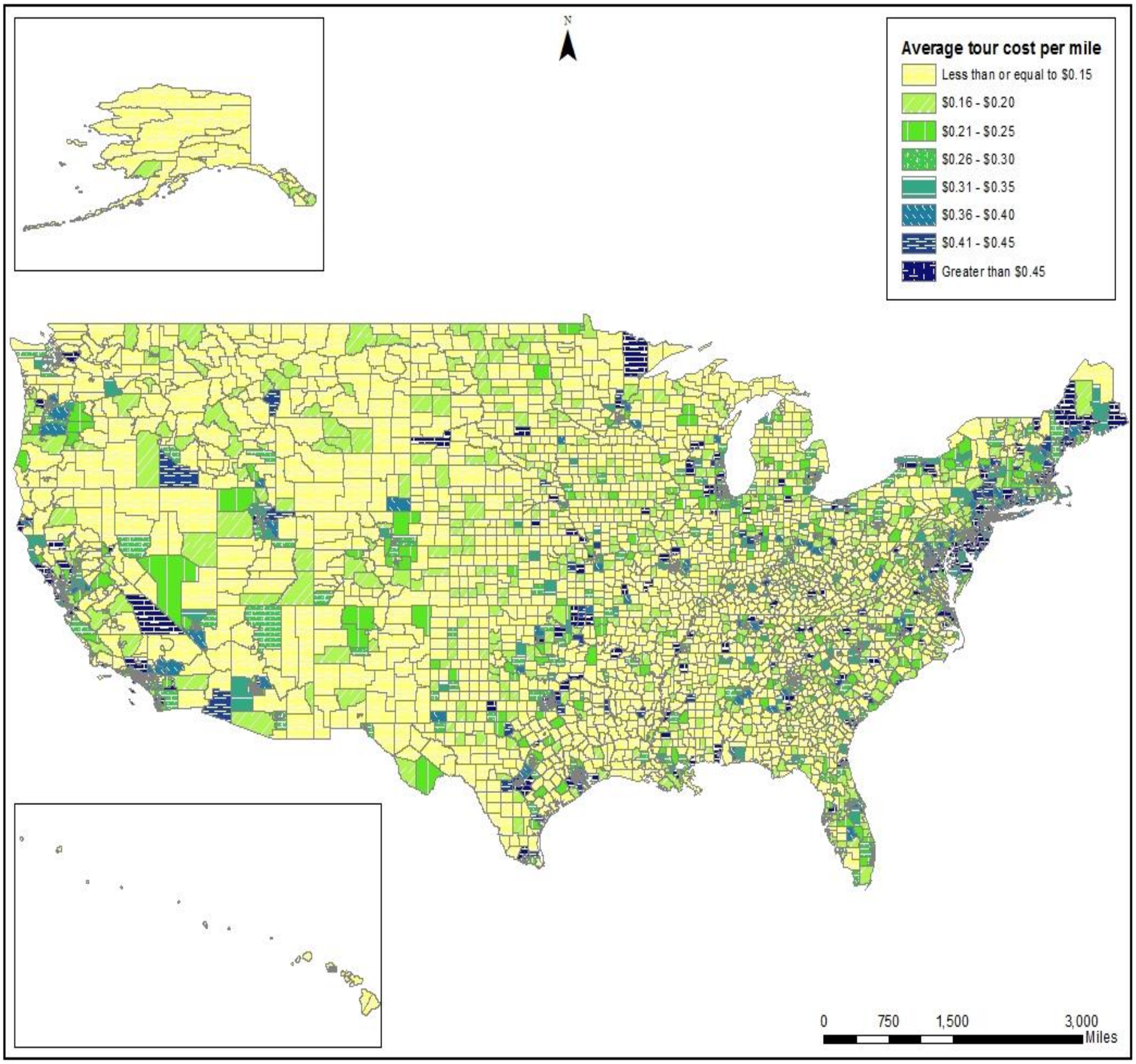
Party Size

There are more tours per household undertaken by single travelers, as shown in Figure 52, but parties of two or four travelers covered more miles than single travelers in October.

Cost

Cost is a useful means to understand the economics of travel demand and the potential for pricing policies to be effective. Figure 58 presents the average tour cost per mile by origin state. The higher costs per mile are in the Northeast (although Vermont, New Hampshire and Pennsylvania are lower cost) and in California.

Figure 58: Average Tour Cost per Mile, by Origin NUMA



6.3 Equity Metrics

The equity of public expenditures on transportation investments is an increasing concern for public agencies. There is a high correlation between aspects of travel and household income, so this is a useful metric to understand equity of a particular investment. Table 79 shows an increase in average tours per household with higher-income groups; this has a logarithmic relationship. The average person-miles traveled per household also increases with household income; this relationship is linear. The average cost per mile also increases with household income, although it is relatively flat for low- and medium-income households before it increases. Average travel time per tour decreases with household income, although only for households with more than $80,000 in annual household income.

Table 79: Long-Distance Travel Metrics in October, by Household Income

Average Household Income	Average Tours per Household	Average Person-Miles Traveled per Household	Average Cost per Mile	Average Travel Time per Tour (minutes)
$0-14,999	0.93	905	$0.18	358
$15,000-24,999	1.23	945	$0.17	363
$25,000-34,999	1.4	1,008	$0.19	357
$35,000-44,999	1.51	1,005	$0.19	351
$45,000-59,999	1.67	1,239	$0.24	361
$60,000-99,999	1.85	1,368	$0.24	361
$100,000-149.999	2.11	1,734	$0.35	346
$150,000 and over	2.23	1,812	$0.40	323

CHAPTER 7. DATA RECOMMENDATIONS

This chapter describes the types of data that—under the ideal conditions—would be available to support a national Long-Distance Passenger Travel Demand Modeling Framework. The main types of data considered here include the following:

- Long-distance travel survey data (for model estimation and calibration).
- Sociodemographic data (for creation of synthetic population and long-term modeling).
- Land-use and employment data (for use as attraction variables in model estimation and application).
- Mode-specific network-based data (for use as impedance variables in model estimation and application).
- Aggregate demand and flow data (for model validation).

For each type of data, the existing data sources considered for model estimation and the type of information currently available are discussed. Also discussed are limitations of existing data (if any) and the type of data that would be available in the ideal situation. In addition, ideal data are compared to existing data, and, when relevant, recommendations are provided for obtaining data that is closer to the ideal.

7.1 Long-Distance Travel Survey Data

Existing Datasets and Their Limitations (If Any)

Currently, the most significant data limitation is the lack of survey data on actual long-distance tours and trips, which would be used for estimating the models that are recommended for the long-distance framework. Table 80 provides a comparison between the key characteristics of existing long-distance travel survey datasets and an ideal dataset. The table indicates that an ideal data source would sample residents of every state and region in the United States, and would record one full year of long-distance travel for each household, so that detailed models of scheduling and seasonality could be estimated. It would also include sufficient geographic detail, at the Census Tract level, or finer, so that details on accessibility and attractiveness of various destinations and LOS by competing modes can be used as explanatory factors in modeling. In addition, it would include sufficient temporal detail to model the peak periods of travel, in terms of periods of the day and week, and specific weeks and months of the year.

Currently, the 1995 ATS is the only domestic survey that contains the first two of these key attributes (geographic and temporal coverage), but it does not contain the last two (geographic and temporal detail). The more-recent CHTS long-distance data provide adequate spatial detail and nearly adequate temporal detail, but these data are only for residents of one state, and only for an 8-week retrospective period. None of the available surveys listed in Table 80 meet all key criteria.

Table 80: A Comparison between an Ideal Dataset and Existing Datasets

Characteristic	Ideal dataset	Existing datasets					
		1995 ATS	2001 NHTS	2003 Ohio	2004/2009 Michigan	2010 Colorado	2012 California
Geographic coverage/sample area	Entire US	Entire US	Entire US	Entire state	Entire state	NFRMPO, DRCOG, PPACG, and PACOG MPO regions	Entire state
Temporal coverage/tour reporting period	Entire year	Entire year	4 weeks	Phase I and II: 2 weeks, Phase III: 4 weeks	3 months	2 weeks	8 weeks
Geographical resolution of travel data	Census Tract or finer	Metropolitan area/state	Metropolitan area/state	Latitude and longitude	Latitude and longitude	Latitude and longitude	Latitude and longitude
Temporal resolution of travel data	Date and nearest hour	Quarter	Date	Departure/arrival date and time	Departure/arrival day of the week	Date	Departure date and time
Temporal immediacy/range of trip recall period	Real time (GPS) or very recent recall	Approximately 3 months	4+ weeks	Phase I and II: 2+ weeks, Phase III: 4+ weeks	3+ months	2+ weeks	8+ weeks

Characteristic	Ideal dataset	Existing datasets					
		1995 ATS	2001 NHTS	2003 Ohio	2004/2009 Michigan	2010 Colorado	2012 California
One-way trip length	All trips of 50+ miles	All trips of 100+ miles	All trips of 50+ miles	All trips of 50+ miles	All trips of 100+ miles	All trips of 50+ miles	All trips of 50+ miles
Tours reported for…	Entire household	Entire household	Entire household	Entire household	Entire household	Entire household	1 person/household

Table 80 (cont.): A Comparison between an Ideal Dataset and Existing Datasets

Key characteristic	Ideal dataset	Existing datasets					
		1995 ATS	2001 NHTS	2001-2003 Ohio	2004/2009 Michigan	2010 Colorado	2012 California
Selected tour details: Purpose, Main mode, Access/egress mode/stations, Length of stay, Party size and composition, And, intermediate destinations	√ √ √ √ √ √	√ √ Mode only √ Limited √	√ √ Mode only √ Limited √	Excl. commute √ X √ (Phase III only) √ X	√ √ X X X X	√ √ Mode only √ Limited X	√ √ X X Limited X
Selected sociodemographic details: HH size and composition, Household/person income, Car ownership Age, Employment status, And, Usual work location	√ √ √ √ √ √	√ √ √ √ √ √	√ √ √ √ √ √	√ √ √ √ √ √	√ √ √ √ √ √	√ √ √ √ √ √	√ √ √ √ √ √

Another key characteristic of the existing long-distance surveys that is of concern is temporal immediacy/trip recall period. Except for the 2003 Ohio Phase III survey, all of the long-distance household surveys in Table 80 are retrospective, with recall periods ranging from 2 weeks to more than 3 months. In general, a long-distance survey with a relatively long recall period is likely to underrepresent "short" long-distance trips and underestimate overall trip rates. This relationship between the temporal immediacy of trip reporting and trip rates is also indicated by Figure 59, which presents trip rates by purposes and retrospective weeks for the CHTS data. The figure shows that a longer recall period results in a smaller trip rate, regardless of trip purposes.

Figure 59: Average Tours per HH-Week by Purpose and Retrospective Week (Data Source: 2012 CHTS)

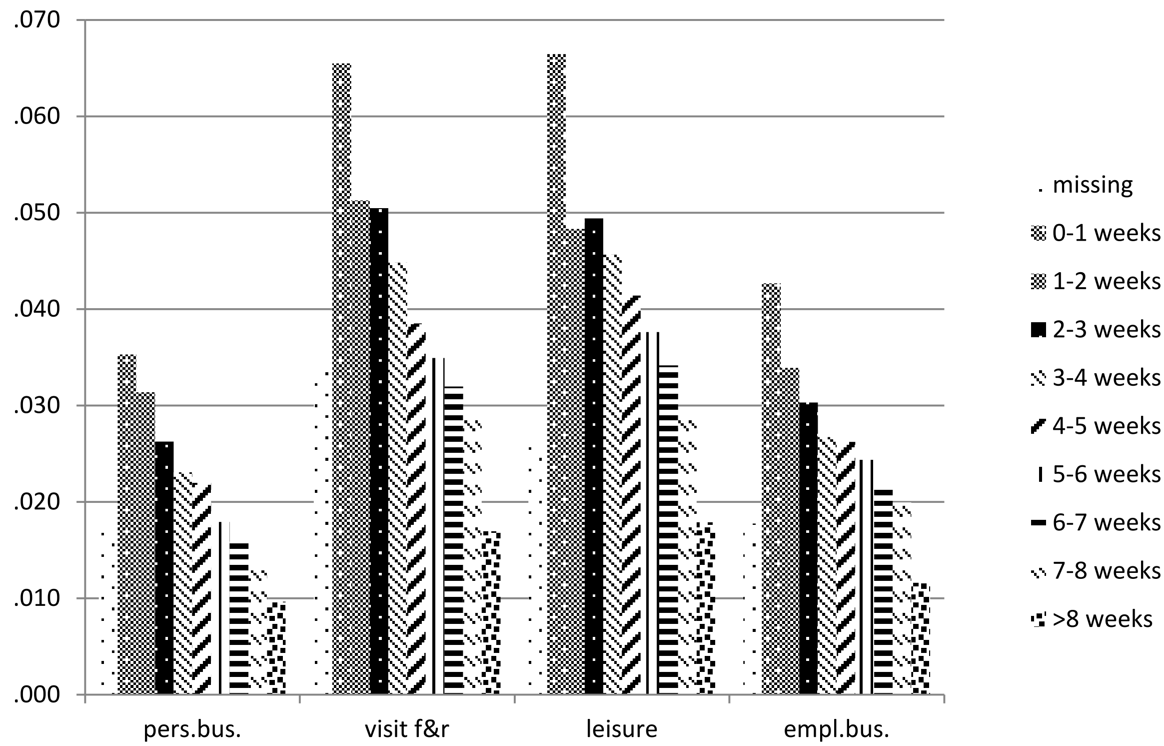


Table 80 also shows that none of the existing household surveys collect detailed tour information that are necessary to fully capture individuals' long-distance travel behavior. However, almost all the surveys collect tour information for the entire household along with relevant household- and person-level sociodemographics.

There was a small sample survey conducted for research purposes in 2013 that contains much of the recommendations for data collection in this study. This is a 12-month, long-distance travel survey of 1,200 participants. The Longitudinal Study of Overnight Travel (LSOT) was conducted monthly online between February 2013 and January 2014. The goal was to measure planned and executed overnight trips for all purposes by individuals over age 24 years. While each monthly survey focused on planned and completed overnight trips, the introductory survey data, asked about participation in long-distance trips of all types. Respondents described how often they make a trip for work or leisure/personal reasons that is: 1) overnight; 2) includes air travel; 3) includes intercity train travel (e.g., Amtrak); 4) includes intercity bus travel (e.g., Greyhound, Megabus); and 5) includes an international destination. This survey could be evaluated to understand survey procedures, attrition, or other data challenges with collecting data over a 12-month period.

Recommendations

The research team recommends that the Federal government conduct a new long-distance travel survey that meets the criteria described in Table 80. Such a survey is most likely to be successful using passive data collection from respondents' smartphones. Although survey technology using smartphones is under development, it has not yet been used for a major household travel survey in the United States. There are a number of reasons why it may be particularly useful and feasible for surveys of long-distance travel.

- Because long-distance trips are of long duration and not as frequent as shorter trips, the respondent's phone does not need to be monitored and tracked for location continuously, as it would be for an urban trip diary survey that collects *all* trips and activities. For example, it may be adequate to check the phone's location just once every few hours to see if it has reached a new location that is more than 50 miles from the home location or the previous long-distance stop location. (This discussion assumes that the respondent has installed a survey application on her or his phone that has access to the phone's location services—GPS, cell, and possibly wireless—whenever the phone is turned on, and has given permission to monitor her or his location. Because there is not a need to know about local trips within 50 miles of their home, data points within that range do not need to be recorded, which might ease respondents' privacy concerns, and reduce the need for data storage.
- Whenever respondents reach a (new) long-distance trip destination, the survey application can automatically prompt them to answer a number of related questions, regarding travel purpose, party size, mode of travel, and other key questions. Questions related to location, date and time of travel, and duration of stay are not as critical, because that information is recorded automatically from the phone location trace data. (When the person is observed in the course of a long-distance tour, the application could automatically switch to monitor the location somewhat more often than otherwise.) Questions regarding each long-distance trip (identified back to the respondent by place and time information) can be "queued" and the respondent prompted periodically until the questions are answered. (The application could also check for the motion of the phone before asking questions, so as not to prompt respondents to answer questions when it appears that they may be driving.)
- Respondents are not likely to forget to bring their phones on long-distance trips, and are likely to have battery power for at least some of the time at their destination location. Thus, two of the most common reasons for missing trips in an urban passive GPS survey—leaving the phone/device at home or running out of battery life—would not be as large of an issue for long-distance surveys.

- With a passive data collection strategy like this, participating in the survey for an entire year does not entail as much respondent burden as with a diary-based or retrospective recall-based survey of an entire year (such as 1995 ATS). It may be an issue that many respondents will replace or lose their smartphone within the period of one year, but that would simply require that respondents install the same survey application on their new phones. It is also important to note that only some of the models in this framework (i.e., the scheduling models for nonbusiness travel) require a full year's worth of data from each respondent. This means that the data provided by respondents who do not participate for one full year will still be useful for modeling several aspects of long-distance travel.
- It would be possible to ask each adult in the household to install the same survey application once a household is recruited to participate, helping to ameliorate the issue of missed trips that are made separately from other household members. The survey application could be intelligent enough to recognize when two or more phones from the same household are taking part in the same trip, and to ask questions about that trip of only one of the participants.
- Since 56% of adults in the United States own a smartphone (Pew Center, 2013), it would be possible to provide a small subsample of such respondents with a smartphone in order to take part in the survey. (The ability to make calls using the phone can be disabled in order to reduce the cost of providing the phone.) Smartphone ownership is increasing rapidly over time, so this will be less of an issue in the future, but it will nevertheless still be important to include nonowners in order to avoid obtaining a biased sample (especially as there is likely to be a correlation between smartphone ownership and the propensity to make long-distance trips).
- It would be important to also recruit a subsample of university students living away from home (either in on- or off-campus housing). University students tend to generate many long-distance trips on their visits to parents and relatives, and they are typically underrepresented in most household travel surveys. Once permission is obtained, it can be cost effective to recruit students through university e-mail systems.
- Another special population that may be worth sampling separately are frequent business travelers, who may be difficult to recruit through standard methods because they are less likely to be at home. Such travelers could possibly be recruited via their employers (i.e., contacting a sample of companies that generate such travel).

It would also be important to recruit a sample of international visitors to the United States as they both make domestic long-distance trips within and have distinct travel preferences (e.g., mode choices) from United States residents. The smartphone-based approach may be less feasible for an international visitor sample due to the difficulty of recruiting them and the possibility of high costs of using their phones. A simpler option may be to interview such respondents at airports when they are departing, asking them relevant questions about trips made while visiting the United States. It is expected that an application of this type could be developed and implemented successfully within the next two or three years, providing long-distance data at the same or less cost as past long-distance surveys. As with household travel surveys of any type, there could still be a significant cost to recruit the sample to take part in the survey.

A large national survey of this type would allow re-estimation of all of the types of models proposed in our long-distance model framework. A sample size of at least 50,000 households seems necessary to cover the variety of demographics and accessibility across the entire country. This would be comparable to the size of the NHTS base sample and the recent California statewide household travel survey.

7.2 Sociodemographic Data

Existing Datasets and Their Limitations (If Any)

This section focuses on sociodemographic data that were required to generate synthetic population for all 50 states and the District of Columbia. For this project, a synthetic-population-generation exercise was undertaken at the Census Tract level. While it would have been preferable to perform the synthetic-population-generation process at the most disaggregate level of spatial resolution for which data are available (such as Census block group), such an effort would have involved generating synthetic population for over 210,000 Census block groups instead of over 70,000 Census Tracts, increasing model run time considerably. Thus, considering the tradeoff between gain in prediction accuracy and increase in computational time, applying a Census Tract-level spatial resolution to synthesize national population was determined to be a reasonable compromise.

Open-source software, PopGen, was used in this project to generate population in permanent households and noninstitutionalized group quarters. For this portion of the process, 2011 ACS 5-year estimates summary file and PUMS file were used. The ACS datasets provided a rich source of information that allowed the research team to identify a set of largely uncorrelated dimensions/control variables that are generally considered key determinants of long-distance travel demand and would adequately capture the heterogeneity of the population.

By using current population synthesis methodologies, and software such as PopGen, it is already possible and quite efficient to create a full, representative base-year population for the entire United States at the Census Tract level. However, a method to create a similar synthetic population for a forecast year has not yet been determined, although population evolution is presented as a possibility in the next section.

Recommendations

PopGen can create a synthetic population for a future year if forecasts of the key sociodemographic variables are available at a reasonable geographic level. For the entire country, however, it is not likely that forecasts will be available for every key variable for every state, and what forecasts are available may only be available at county or state level, rather than the Census Tract level. It could therefore be challenging to satisfactorily compile forecasts of all of the input variables.

An alternative to this method would be to develop a “demographic evolution model” to begin with the base-year population and evolve each household over time to simulate events such as births, deaths, marriages, divorces, children “leaving the nest,” workers entering and leaving the workforce, and people moving to other locations (perhaps including foreign immigration and emigration). A variety of such models have been developed (Kazimi, 1994), (Sundararajan, 2003), and (Transportation Research Board, 2014 (forthcoming)) in the past, although none of them are yet used regularly in a transportation planning context in the United States.

The most challenging part of developing such a model is to predict net migration between different areas of the country, and what types of households will most likely move between what types of new locations. The purely demographic transitions (e.g., births, deaths, household formation, etc.) can be observed from panel datasets, such as the Panel Survey on Income Dynamics (University of Michigan, 2013). There are less data available on migration, however. Although the Census has a great deal of data on net in-migration rates and out-migration rates for different geographies, it does not provide much detail regarding where people have moved from. However, it may be possible to use such marginal data to estimate a full migration origin-destination matrix, using similar matrix estimation methods as are used to approximate trip O-D matrices in transportation planning.

One additional issue are the data available on international visitors. It may be possible to obtain reasonably accurate data on the number of visitors from different countries who are arriving at different airports during each season of the year. For example, such data are available in an extended, nonpublic version of the DB1B air ticket database that may be made available for federally funded research projects. However, it is not likely that one can obtain reliable sociodemographic data, such as household income or household size, for foreign visitors. For this reason, any models estimated to represent the long-distance travel behavior of foreign visitors while in the United States should avoid including variables that require such sociodemographic details.

7.3 Land-Use and Employment Data

Existing Datasets and Their Limitations (If Any)

Census Tracts were also used as the spatial units for summarizing land-use and employment data. National-scale surveys and data collection efforts listed in Chapter 3 were good sources of land-use data. The research team was able to assemble the relevant land-use information from these sources with relatively modest efforts. On the other hand, the LEHD database, which is the primary source of employment data for the current study, has several limitations:

- Massachusetts is yet to join the LEHD program. Thus, employment data are not available for this state.
- No employment data are available for uniformed military personnel, self-employed individuals, sole proprietors, railroad workers, and workers who are exempt from the Federal/state unemployment insurance laws. Areas (specifically small geographical areas) with a relatively high concentration of these excluded groups of workers may require careful consideration.
- Employers with multiple work sites/offices are not required by law to report actual work location of employees (the only exception is Minnesota). As a result, the LEHD data may contain inaccurate work location information and show employees located only at primary employer addresses.

- The data are subject to noise infusion by the Census Bureau to meet disclosure avoidance requirements, so values in the data are perturbed from actual values.

Employment data available from the BLS QCEW dataset was used for Massachusetts. This source provided data at a coarser county-level geographical resolution.

Recommendations

As indicated previously, the data already available at the national level support the typical type of land-use attraction variables used in most travel demand models. The Census and ACS data provide information on numbers of households and persons and key population distributions, while the LEHD data provide suitable attraction employment for various categories—even providing 3- or 4-digit NAICS classifications for certain types of employment that are geared toward serving tourists and visitors. There are also data available on the amount of land area in public parks, including national, state, and local parks. While one could imagine more direct measures of attraction for some types of nonresidential visitors (e.g., number of hotel/motel beds), the related measures of employment are a good proxy for this, and are consistent across the country. As a result, no specific recommendations have been included for improving this type of data, other than to continually refine the coverage, accuracy, and consistency of the LEHD employment data over time.

For forecasts of future-year land use and employment, it may not be a simple matter to provide forecasts of all input variables for the entire country. While some regions and/or states will have detailed population and employment forecasts, others may not. As a result, the process for generating future-year land-use scenario data may require the following:

- Start from any national-level forecasts that are available. For example, the Census Bureau provides 2015–2060 national population projections (U.S. Census Bureau, 2012), while the BLS produces 2010–2020 national employment projections by industry (U.S. Bureau of Labor Statistics, 2013).
- Possibly adjust those forecasts with more-detailed regional forecasts for key states and metropolitan areas that may be of interest.
- Apply scenario-based variations or assumptions to test the effects on long-distance travel of specific types of future land-use development patterns in different parts of the United States.

7.4 Network-Based Data

Existing Datasets and Their Limitations (If Any)

In the current project, network data attributes (i.e., travel time, cost, and distance) were used as impedance variables to estimate joint mode and destination choice models, described in Chapter 4. Datasets that cover the entire country (or the contiguous part of the country) were used in the model application described in Chapter 5. Several sources were used to assemble the following mode-specific network data attributes.

- **Highway Network**. The NHPN is a relatively sparse road network. Thus, the travel distance matrix generated from this network likely includes some entries that deviate considerably from the actual travel distance between a particular O-D pair. In addition, (relative to state/regional travel demand models) it was necessary to create relatively long connectors to connect zonal centroids with the nearest network links, which were typically either arterials or interstates. This network simplification is also likely to introduce significant error in travel distance calculation. For the travel-time matrix, although the research-team-applied average operating speeds varied by roadway functional class,[16] the adopted speeds do not necessarily reflect actual network travel conditions. Further, there is currently no national-level database that provides toll and other travel cost-related information.
- **Rail Network**. Amtrak's GTFS data were used to build the rail network. These data for a typical week were used to create a national dataset of station-to-station (and subsequently a TAZ-to-TAZ) O-D matrices for the number of trains serving a particular O-D pair over a period of one week (i.e., frequency per week), the average journey time (including transfers), the number of transfers, and the average journey distance. No national-level database exists on rail fare. The only fare information that is available in the public domain is through "Buy Ticket" service on Amtrak's website.
- **Air Network**. Airline on-time performance data (on-time data hereafter) and airline origin and destination survey (DB1B) were used to build the air network. These data, available from the BTS, were used to create a national dataset of the number of flights serving a particular O-D pair over a period of one week (i.e., frequency per week); on-time performance (in percentage) across the flights serving a particular O-D pair over a period of one week; the average flight duration (including transfers) in minutes; the average passenger-weighted fare by season for a particular O-D pair; the number of passengers, by season, for trips between the airports; and the average coupon-mileage. Merging on-time data with DB1B data is not straightforward and requires significant data processing effort/making a number of assumptions. For example, on-time data do not indicate whether or not a direct flight is nonstop or contains one or more stops. Thus, linking LOS information (e.g., frequency, flight duration, etc.) from on-time data with demand data from DB1B database—which does not include flight information—requires careful consideration.
- **Bus Network**. In addition to the limitations of the existing datasets discussed previously, this is currently no national-level database available for the long-distance bus network.

Recommendations

For the air and rail modes, there are considerable existing data available that can be assembled into networks, albeit with a significant amount of data processing required. For rail, it may be useful to add the few non-Amtrak long-distance routes into the network (e.g., the train from Albuquerque to Santa Fe in New Mexico). For the long-distance bus network, additional work is

[16] In this study, midpoints of anticipated operating speeds by roadway functional class, published in "NCHRP Report 504 Design Speed, Operating Speed, and Posted Speed Practices" were used.

required to code the available routes, frequencies, and fares. This work is made challenging by the fact that this market is so dynamic, with companies going out of business and new companies starting business quite frequently. Work in this area should continue and could be concentrated in the Northeast Corridor and in California, where these services are most competitive. The long-distance road network could be further improved in terms of the number of different roads included and the amount of detail for each road link. It is likely that more-detailed networks will also be needed for research on the freight side, so the research team recommends that such efforts be supported by Federal projects where possible.

Additional data that might be useful in some contexts are data on local road network level of service and local transit services. For example, if one wanted to use this model framework to also predict which mode of access and egress people use to get to and from airports for air travel, then it would require more-detailed data on local travel options and service levels, as well as details on parking cost and availability at the airport. This would be an expensive, and perhaps infeasible, process to create such a detailed local network database for the entire country. However, it may be possible to use the national long-distance travel model as a starting point for a model that adds additional network detail in a specific region or corridor in order to obtain more-detailed forecasts for that region.

7.5 Model Validation Data

Existing Datasets and Their Limitations (If Any)

Long-Distance Trip Origin-Destination Movements by Mode

This is to validate the main output of the long-distance model, which will be in the form of O-D trip tables by long-distance mode. For the air and rail modes, the model outputs can be compared to observed passenger demand in each season—from the DB1B ticket data for air, and from Amtrak ridership statistics for rail. For the Amtrak system, station-level boardings and alightings are published, and Amtrak has been willing to share confidential O-D data for use in other research studies. The O-D movements for bus will be more difficult to validate on a national level, but there may be some ridership statistics for comparison for certain routes in the Northeast Corridor and in California (from Amtrak/Caltrans), where much of the bus demand is located.

For auto travel, obtaining O-D data is somewhat more problematic. Although there are trip O-D tables for many specific regions, and link counts and screenline counts, nearly all of those data sources include both short- and long-distance trips; as a result, these data would not be comparable to matrices, including only long-distance car trips. This gap in the available data can be partially filled by aggregate trip-tables available from cellphone network providers, based on cellphone records from millions of customers. Although privacy issues prevent such firms from releasing any individual-level data or any sociodemographic information about the travelers, they can identify each person’s home location and work location (if applicable), and can provide aggregate trip tables for any zone system provided by the data purchaser, with the trips broken down into type—home-based work, home-based other, work-based other, and other non-home-based. Further, these firms can distinguish between trips made by residents of a particular region or corridor from those made by nonresidents.

Aggregate cellphone data (and aggregate GPS or Bluetooth data) data have been used for this specific purpose in several recent long-distance corridor studies, and could be used to validate the auto O-D flows from a long-distance model for several selected key corridors around the nation, even if not for every O-D pair. NCHRP is underway with a project (Transportation Research Board, 2013) that is assessing the usefulness of aggregate cell phone data for purposes such as this and preparing guidelines as to how it should be used. Also, it is likely that the accuracy and availability of such aggregate cellphone-based data will increase over time.

This type of data might also be used in model operation to develop rules on what proportion of auto trips on the roads are short- versus long-distance trips, as a function of geography (urbanization levels of the origin and destination), and the time period of the day, week, and year (weekdays, Friday evenings, weekends, holidays, etc.). Such a function would permit estimation of the “background” level of short-distance trips on the network so that the predicted long-distance auto trips can be added for route assignment and congestion analysis. (This is a process analogous to preloading commercial vehicle traffic and through-trips on the highway network in an urban travel model. Commercial vehicles would also need to be preloaded in a long-distance model).

Highway Speeds and Travel Times by Highway Segment and Time of Day/Week/Year

These data are important, both for model inputs (so that the model is using realistic travel times for auto trips) and model output. Even if the outputs of the model are not used explicitly for traffic assignment at the national level, one would expect the highest levels of congestion to occur in corridors and time periods for which the most long-distance auto trips are predicted. This is particularly true for weekend and holiday travel, and for key highway corridors that are mostly outside of urban areas and have the smallest proportion of local traffic.

Fortunately, although FHWA has not made the dataset publicly available, it is possible to obtain permission to use the National Performance Management Research Data Set (NPMRDS).[17] This dataset provides travel time data for segments of every interstate highway and other key highways and arterials in the entire country, for five-minute time period “bins” (e.g., 5:00 a.m. to 5:05 a.m., 5:05 a.m. to 5:10 a.m., etc.) continuously through time. The data are based on NAVTEQ/Nokia probe vehicle observations, with separate averages for heavy-goods vehicles and passenger vehicles. Because the length of each highway segment is known, the average speed on the segment can also be imputed from the travel time data. These data will be ideal for identifying weekly and seasonal congestion patterns on interstates and in other key long-distance corridors around the country; these data will also be useful for comparing against model outputs—in the form of modeled link speeds if the long-distance model includes assignment.

Highway Link Volumes by Highway Segment and Time of Day/Week/Year

Neither the aggregate cell phone data nor the NPMRDS data can provide accurate measures of hourly flow volumes on particular links, as their sample sizes are not large enough to provide

---

[17] National Performance Management Research Data Set (NPMRDS)

reliable estimates. These data are available in the Highway Performance Management System (HPMS) at the FHWA (Federal Highway Administration, n.d.). The HPMS is a national-level highway information system that includes data on the extent, condition, performance, use, and operating characteristics of the nation's highways. These data are available in shape files with average annual daily traffic (AADT), facility type, high-occupancy-vehicle lanes, tolls charged, and percent of traffic that are trucks as attributes. The NHPN is a geospatial network with a linear referencing system that links it to the HPMS.

Recommendations

In summary, the scarcity of observed data on long-distance auto trips is steadily being remedied by the availability of various types of aggregate passive datasets (from cell phones and GPS probe vehicles), in addition to the automated flow and speed data that are collected on many of the major highway links in the nation. FHWA should encourage and fund the use of these data in establishing reliable evidence and modeling of long-distance travel patterns.

7.6 Summary

The most critical need for modeling long-distance travel behavior at the level of detail recommended in this model framework is obtaining more complete survey data on households' actual long-distance travel characteristics. This report provides recommendations geared toward a smartphone-based survey methodology. Such a methodology would have a low respondent burden and could provide the necessary spatial and temporal detail for models, and it is possible that a substantial proportion of the respondents would provide data for one full year. There is also the need to consider special subpopulations in designing such a survey approach, including households without smartphones, university students living away from home, frequent business travelers, and foreign visitors from outside the United States.

For the current year, the existing data sources and methods for creating a synthetic population for the country are adequate for this report's purposes, as are the data for representing long-distance travel attractions, including employment in specific sectors, households, open/space parks, and universities. The most pressing question remains how to forecast these important input data for future years. On the population side, the most promising approach may be to use household "evolution" models that start with a base-year population and evolve it over time. Further research in this area is recommended.

The current data sources for air and rail network services (e.g., routes, frequencies, fares, etc.) are quite extensive and adequate for the level of modeling recommended in this report, although some complex data processing is involved in extracting the needed information. The existing representation of a national-level road network exists, but greater detail and coverage would be useful, and the research team recommends adding such detail in future Federal research projects. A national network representation for long-distance bus services does not currently exist, and it is recommended that the creation of such a network be the goal of related Federal projects.

Data exist to validate base-year model results, including the NPMRDS and HPMS data available at the national level, and O-D matrix data that can be created from aggregate cellphone records by providers. The availability and accuracy of such data is increasing steadily over time.

CHAPTER 8. SUMMARY AND IMPLEMENTATION PHASE

8.1 Summary

The development of the Long-Distance Passenger Travel Demand Modeling Framework included research into new methods for estimating long-distance passenger model components (Chapter 4) and implementation of selected methods to produce long-distance passenger travel demand on a national scale (Chapter5). The parallel paths allowed research to include methods that—while perhaps not immediately implementable—should be considered for future efforts. In the case of the MDCEV tour generation, scheduling, and participation models, the research methods are implementable and will be compared to the selected models during the implementation phase.

The disaggregate tour-based modeling structure was selected for the framework based on technical advantages this structure has over a more aggregate, trip-based structure:

- Tracks individual households' travel behavior (not averages).
- Connects long-distance travel choices for one full year.
- Allows for greater spatial and temporal detail.
- Allows greater household/person attribute detail.
- Integrates mode choice, destination choice, and Tour Generation and Scheduling with causal effects in both directions.

In addition, the structure is more intuitive and understandable to nontechnical audiences, providing more credibility to the results. Some of the challenges in the research phase have been based on limitations in available data for model estimation. The focus of this initial research was on developing a framework that could be re-estimated with more robust and comprehensive data sources when these data are collected. These data sources were described in Chapter 7.

The extension of the project to include an implementation phase has benefited the research phase in several ways. The development of the modal networks and level of service was a more robust effort than would have been necessary or possible for the research phase. These national modal networks have been developed with a more refined zone system that supports long-distance passenger travel, and additional data sources for bus and rail modes were obtained to improve the travel time, distance, and cost matrix development for these modes. The model estimation of the destination and mode-choice models benefited from a merged long-distance passenger travel survey dataset (four states) instead of a single-state dataset. This allowed the destination and mode-choice models to reflect travel behavior from the western, central, and eastern parts of the United States, instead of solely focusing on one part of the country.

8.2 Implementation Phase

The implementation phase is focused on moving the research into practice and providing a model that can be used by state and Federal agencies interested in long-distance passenger travel. The current research has produced a long-distance passenger travel demand model framework, with models estimated from available data, recommendations for future data collection, and a demonstration of the implementation framework. This third phase would produce a working model for a base year, adding trip assignment models, and would be calibrated and validated to available national data sources. This phase would also include sensitivity testing. This testing would provide assurances that the calibrated and validated models produce reasonable results under a select set of policy scenarios. This phase would also include additional software development to ensure stability and reasonable performance for the application software beyond the original demonstration software in the research phase.

The current research estimated models based on long-distance surveys collected from several states (Ohio, Wisconsin, California, and New York) because these states offered promising sample sizes for the destination and mode-choice models. Since there is no comprehensive, detailed survey in the recent past that can support all model estimation, the resulting long-distance passenger travel models will require additional calibration and validation efforts to ensure reasonable estimates of long-distance passenger travel. Fortunately, there are recent O-D trip tables developed by FHWA (with CDM Smith and RSG) that can be used for this purpose. The 2001 NHTS could also be used for calibration of individual model components.

The following tasks and subtasks of the next phase are itemized and briefly described. This is intended as a broad overview of the work expected in this phase.

Task 12: Build the Model for the Base Year

Enhance the Networks

This task is complete and has been described in this memo in Section 2.3 of this report. During the calibration phase, there will be some minor manual adjustments to the networks and zone systems to correct anomalies identified in the underlying national datasets.

Calibrate Model Components

Model calibration is the process of applying the estimated models, comparing the results to observed values, and adjusting either the model specification or the alternative-specific constants. The process is complicated by the fact that the various model components in the long-distance passenger travel demand model are not isolated: long-term decisions influence how and, to a certain extent, where people travel; lower-level decisions also can affect the higher-level choices through the LogSum, an explanatory variable in the higher-level choice models. As a result, a change in the share of one model is likely to influence the outcome of other models. Therefore, the general approach is to calibrate model components in the order in which they are applied, which generally means that the higher-level models are calibrated before the lower-level models. In addition, the calibration process will be applied in an iterative manner in order to incorporate all the interactions between models until the model, performing as a system, converges to a stable set of parameter values for all of the model components.

The MDCEV tour generation, scheduling, and participation models will be implemented during this task. These results will be compared to those from the logit-tour-frequency models in order to determine tradeoffs in terms of processing time and capabilities for implementation purposes.

The preferred approach will be selected and used in the calibration and validation of the final model. Calibration of all models will consider use of the combined four-state household survey dataset (California, Ohio, New York, and Wisconsin) and the 2001 NHTS. These will be compared to understand similarities and differences between the larger 4–8 week sample from the four-state merged dataset and the smaller one-day sample from the national survey. The party size and destination and mode-choice models will also be calibrated with these datasets.

One key aspect of the calibration will be to use the most reasonable values of time for all purposes and to match observed mode shares and distance distributions as closely as possible. Work conducted to date indicates that it may be beneficial to segment each tour purpose into "short" and "long" long-distance tours—for example, 50–150 miles versus over 150 miles. This is because most of the tours are in this shorter 50–150 mile band, and these tend to reveal quite different choice behavior and tradeoffs as compared to the longer-distance tours, which can be up to 3,000 miles. This type of purpose/distance segmentation was used for the mode/destination choice models for the California Statewide Long-Distance Model, and seemed to work quite well in model application.

Develop Highway-Assignment Model

The highway-assignment model will be developed using the daily calibrated trip table for the auto mode and the highway network developed for the project. This assignment will require estimation of background traffic to represent congestion on the national highway system. This background traffic will be estimated by subtracting long-distance volumes from the HPMS volumes to produce short-distance passenger volumes, which will be combined with truck traffic.

The average daily long-distance passenger trip tables will be assigned—using TransCAD—with the background traffic in order to produce average daily volumes on the highway system. Performance metrics from this system will be produced to evaluate the impact of long-distance travel separate from the background traffic. These performance metrics include vehicle miles traveled by region and state and average speeds and volumes by facility type and area type.

Validate Model System

The long-distance passenger travel demand models will be compared to observed sources to determine the reliability of the base-year (2010) estimates. There are several sources of observed data:

- National origin-destination trip tables by mode developed by FHWA.
- Traffic link volumes in the HPMS data.
- Airport-to-airport air passenger O-D counts from the DB1B 10% ticket database.
- Station-to-station rail passenger counts from the Amtrak proprietary data.
- External traffic surveys performed for large urban regions, where the through traffic are all long-distance trips and can be compared against the model's prediction of similar through-trips.

In this case, observed data have been developed from multiple sources and will be compared as the best source available, with the understanding that both the observed and modeled data have a margin of error. System performance will be determined for each mode according to the available data (e.g., air or rail passenger miles traveled, rural freeway vehicle miles traveled, etc.).

Conduct Sensitivity Tests

A series of sensitivity tests will be developed to demonstrate how the long-distance passenger travel demand modeling system can compare potential future alternatives or test policies of interest. A set of potential policy and alternatives of interest was provided in the Objectives Memorandum (October 2011). A set of six sensitivity tests will be conducted; potential tests include:

- Modal alternatives for rail;
- Modal alternatives for air;
- Pricing—system tolling;
- Economy—employment and incomes increase/decrease;
- Environmental—gas tax increase; and
- Safety—reduce freeway speeds.

The results from these sensitivity tests will be evaluated for reasonableness and elasticities produced by variations to the base-year inputs will be compared against demand elasticities reported elsewhere in the literature.

Task 13: Enhance Application Software Performance

Software enhancements will be used to enhance the demonstration software for use by state and Federal agencies, as follows:

- Improve run times with software engineering.
- Conduct tests to ensure stability.
- Compile and deliver for use.

Task 14: Develop User's Guide and Final Report

A user's guide will be developed to document the use of the software and a final report will be developed to combine the various technical memoranda developed over the course of the project and document the final models included in the long-distance passenger travel demand modeling system. This final report will include a summary of the design and research phases of the project and details from the implementation phase.

Schedule and Deliverables

Work on the network enhancements has already been completed; the remaining work will be performed in the implementation phase, which runs through June 30, 2015. The following products for the base year (2010) can be expected from the implementation phase:

- Aggregated annual and daily-trip tables, by mode.
- Loaded daily highway networks.
- Long-distance passenger travel demand models complete with input and output data and a user's guide.
- Final Report for the three phases of the project, documenting the design, research, and implementation of the model.

CHAPTER 9. REFERENCES

Alegre, J. and L. Pou. (2006) An Analysis of the Microeconomic Determinants of Travel Frequency. Department of Applied Economics, Universitat de les Illes Balears.

Algers, S. and Beser, M. (2000) SAMPERS – the new Swedish national travel demand forecasting tool, presented to IATBR Conference, Gold Coast, Australia.

Algers, S., Lindqvist Dillén, J. and Widlert, S. (1995) The national Swedish value of time study, presented to 'Traffic Days', Aalborg University.

Arentze T., H.J.P. Timmermans, and F. Hofman. Creating Synthetic Household Populations: Problem and Approach., Transportation Research Record: Journal of the Transportation Research Board, 2014, 2007, pp. 85-91.

Auld, J. and A. K. Mohammadian (2010) Efficient Methodology for Generating Synthetic Populations with Multiple Control Levels, Transportation Research Record: Journal of the Transportation Research Board, 2183, pp. 19–28.

Auld, J.A., A. K. Mohammadian, and K. Wies (2009) Population Synthesis with Subregion-level Control Variable Aggregation. ASCE Journal of Transportation Engineering 135(9), 632–639.

Axhausen, K.W. (2003) Social networks and travel: Some hypotheses. Arbeitsbericht Verkehrsund Raumplanung, 197, Institut fur Verkehrsplanung und Transportsysteme, ETH Zurich, Zurich.

Bakker, D., Mijjer, P. Daly, A. and Hofman, F. (2000) Updating the Netherlands National Model, presented to ETC Conference.

Bargeman, B., van der Poel, H. (2006) The role of routines in the vacation decision-making process of Dutch vacationers. Tourism Management 27: 707-720.

Bar-Gera, H., K.C. Konduri, B. Sana, X. Ye, and R.M. Pendyala (2009) Estimating Survey Weights with Multiple Constraints Using Entropy Optimization Methods. Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Barthelemy, J.and P.L. Toint (2013) Synthetic Population Generation without a Sample. Transportation Science (in press), Published online before print April 5, 2012, doi: 10.1287/trsc.1120.0408.

Beckman, R.J., Baggerly, K.A., and McKay, M.D. (1996) Creating synthetic baseline populations. Transportation Research Part A, 30(6), 415-429.

Beerli, A., Martin, J.D. (2004) Tourists' characteristics and the perceived image of tourist destinations: a quantitative analysis- a case study of Lanzarote, Spain. Tourism Management 25: 623-636.

Ben-Akiva, M., Cascetta, E., Coppola, P., Papola, A. and Velardi, V. (2010), High speed rail demand forecasting: Italian case study, paper presented at the European Transport Conference, Glasgow, October 2010.

Ben-Akiva, M., Daly, A. J. and Gunn, H. F. (1987) Destination Choice Models: Design and Appraisal, presented to P.T.R.C. Summer Annual Meeting.

Bernardin, Lochmueller and Associates, Inc. and Cambridge Systematics, Inc. (2004) Indiana Statewide Travel Demand Model Upgrade. Technical Memorandum. Prepared for Indiana Department of Transportation.

Bhat, C. R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological 42*(3), 274-303.

Bhat, C.R. (2000) Incorporating Observed and Unobserved Heterogeneity in Urban Work Mode Choice Modeling, Transportation Science, Vol. 34, No. 2, pp. 228-238.

Bhat, C.R. (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7), 923-939.

Bhat, C.R., C. Varin, N. Ferdous (2010). A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered response model. *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications* 26, 65-106, edited by W. Greene and R.C. Hill, Emerald Group Publishing Limited

Bhat, C.R., M. Castro, M. Khan (2013a). A new estimation approach for the multiple discrete-continuous probit (MDCP) choice model. Forthcoming, *Transportation Research Part B*.

Bhat, C.R., S.K. Dubey, R. Sidharthan, and P.C. Bhat (2013b). A multivariate hurdle count data model with an endogenous multiple discrete-continuous selection system. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, July 2013.

Bok, M. de, Costa, A., Melo, S. and Palma, V. (2010) Estimation of a choice model for long-distance travel in Portugal, presented to WCTR, Lisbon.

Bostrom, R. (2006) Kentucky Statewide Travel Model (KYSTM). Presentation. Combined Kentucky-Tennessee Model Users Group Meeting, Bowling Green KY. October 26, 2006.

Bowman, J. and M. Ben-Akiva (1996) Activity Based Travel Forecasting. Activity Based Travel Forecasting Conference Proceedings.

Bowman, J.L. (2004) A Comparison of Population Synthesizers Used in Microsimulation Models of Activity and Travel Demand. Unpublished Working Paper. Retrieved February 28, 2013 from http://jbowman.net/papers/2004.Bowman.Comparison_of_PopSyns.pdf

Brey, E. and X. Lehto. (2007) The relationship between daily and vacation activities. Annals of Tourism Research. 34(1): 160-180

Brög, W., Erl, E. and Schulze, B. (2003) DATELINE: Design and Application of a Travel Survey for Long-distance Trips Based on an International Network of Expertise – Concept and Methodology, available at http://www.fcsm.gov/03papers/Brog_Erl_Schulze.pdf.

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.

Burnett-Wolle, S., & Godbey, G. (2005). Activity aging 101. Parks & Recreation, 40(3), 30–40.

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.

Castro, C.B., Armario, E.M., Ruiz, D.M. (2007) The influence of market heterogeneity on the relationship between a destination's image and tourists' future behaviour. Tourism Management 28: 175-187.

Castro, M., R. Paleti, C.R. Bhat (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B* 46(1), 253-272.

Chan, F., C. Lim, and M. McAleer. (2005) Modelling multivariate international tourism demand and volatility. Tourism Management. 26: 459–471.

Chib, S., R. Winkelmann (2001). Markov Chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics* 19(4) 428–435.

Chung, P.-J. (2006) The determinants of consumers' outbound tourism mode. Master's Thesis,

Daly, A. (2010) Cost Damping in Travel Demand Models, RAND Europe report for UK Department for Transport, available at

Daly, A. and Carrasco, J. (2009) The influence of trip length on marginal time and money values, in The Expanding Sphere of Travel Behaviour Research: Selected papers from the proceedings of the 11th Conference on Travel Behaviour Research, Kitamura, R. et al. (eds.), Emerald Books.

Daly, A., Fox, J. and Rohr, C. (2002) Advanced modeling to overcome data limitations in the Norwegian Transport Model, presented at European Transport Conference, Cambridge.

Daly, A.J. (1982) Estimating Choice Models Containing Attraction Variables, Transportation Research Vol. 16B, 1.

Daly, A.J. (1998) Prototypical Sample Enumeration as a Basis for Forecasting with Disaggregate Models. In: PTRC/AET Conference.

Daly, A.J. and Miller, S. (2006), Advances in modeling traffic generation, European Transport Conference, Strasbourg.

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.

Daly, A.J., Rohr, C., Jovicic, G. (1999) Application of Models based on Stated and Revealed Preference Data for Forecasting Passenger Traffic between East and West Denmark. In: Proceedings of 8th World Conference on Transport Research, Antwerp, Belgium, pp. 121-134.

Dargay, J. (2010) The prospects for longer-distance coach, rail, air and car travel in Britain, report to International Transport Commission, published by Institute for Transport Studies, University of Leeds, available from http://www.theitc.org.uk/docs/6.pdf.

de Lapparent, M., Frei, A. & Axhausen, K. W., Choice of Mode for Long Distance Travel: Current SP-based Models from Three European Countries, paper presented at the European Transport Conference, October 2009

Decrop, A., and Snelders, D. (2004) Planning the summer vacation: an adaptable process. Annals of Tourism Research 31(4): 1008-1030.

Decrop, A., and Snelders, D. (2005) A grounded typology of vacation decision-making. Tourism Management 26:121-132.

Department for Transport (2009), National Transport Model: High Level Overview, available from http://www2.dft.gov.uk/pgr/economics/ntm/highleveloverview.pdf

Devall, B. (1973) The Development of Leisure Social Worlds. Humboldt Journal of Social Relations 1:53–59.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. American Psychologist, 55(1), 34–43.

Driver, B. L., Tinsley, H. E. A., & Manfredo, M. J. (1991). The paragraphs about leisure and recreation experience preference scales: Results from two inventories designed to assess the breadth of the perceived psychological benefits of leisure. In: B. L. Driver, P. Brown, & G. L. Peterson (Eds.), Benefits of leisure (pp. 263–286). State College, PA: Venture Publishing Inc.

Dunn Ross, E.L. and S.E. Iso-Ahola. (1991) Sightseeing Tourists’ Motivation and Satisfaction. Annals of Tourism Research. 18: 226-237.

Dupuis, L. (2004) Winter Tourism in Protected Pleasure Peripheries: Time-Space Use among Cross-Country Skiers in Abisko (Sweden)and Vercors (France). Scandinavian Journal of Hospitality and Tourism. 4(2): 129-153.

Eluru, N., A.R. Pinjari, J.Y. Guo, I.N. Sener, S. Srinivasan, R.B. Copperman, and C.R. Bhat (2008), Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems. Transportation Research Record: Journal of the Transportation Research Board, 2076, pp. 171-182.

Erhardt, G.D, J. Freedman, A. Stryker, H, Fujioka, and R. Anderson. (2007) Ohio Long-Distance Travel Model. Transportation Research Record: Journal of the Transportation Research Board. Washington DC. 2003: 130-138.

Eugenio-Martin, J. (2003). Modelling determinants of tourism demand as a five stage process: A discrete choice methodological approach. *Tourism and Hospitality Research* 4(4), 341-354.

Eugenio-Martin, J.L. (2008) Modelling determinants of tourism demand as a 5-stage process: a discrete choice methodological approach. Technical Paper, Universidad de Las Palmas

European Travel Commission. (2006) Tourism Trends for Europe. Bruxelles, Belgique. <www.etc-corporate.org>

Eymann, A., Ronning, G. (1997) Microeconometric models of tourists' destination choice. Regional Science and Urban Economics 27: 735-761.

Federal Highway Administration (2012) The Energy and Emissions Reduction Policy Analysis Tool: Model Documentation, available at http://www.planning.dot.gov/FHWA_tool/EERPAT_Model_Documentation_21.pdf

Federal Highway Administration. (n.d.). Office of Highway Policy Information. Retrieved December 20, 2013, from Highway Performance Management System (HPMS): http://www.fhwa.dot.gov/policyinformation/hpms.cfm

Fish, M. and D. Waggle. (1996) Current Income versus Total Expenditure Measures in Regression Models of Vacations and Pleasure Travel. Journal of Travel Research. 35: 70-74.

Fosgerau, M., Hjorth, K. & Lyk-Jensen, S.V. (2007), The Danish Value of Time Study, Final Report, DTU Copenhagen.

Frazier,C., R. Picado and C. Willison. (2009) The Tahoe Basin Region Visitor Model: an Activity-Based Approach. Technical Manuscript, Submitted to the TRB Transportation Research Board 2009 Conference.

Frei, A., Kuhnimhof, T. G. & Axhausen, K. W. (2010), Long-Distance Travel in Europe Today: Experiences with a New Survey, paper presented at the 89th Annual Meeting of the Transportation Research Board, Washington, D.C.

Frei, A., T. Kuhnimhof and K.W. Axhausen (2009) Long distance travel in Europe today: Experiences with a new survey, paper presented at the 89th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2010.

Fuijiwara, A. and J. Zhang. (2005) Development of Car Tourists' Scheduling Model for One-day Tour. Transportation Research Record 1921: 100-111.

Gibson, H. and A. Yiannakis. (2002) Tourist Roles: Needs and the Lifecourse. Annals of Tourism Research. 29(2): 358-383.

Gokovali, U., O. Bahar, and M. Kozak (2007) Determinants of length of stay: A practical use of survival analysis. Tourism Management. 28: 736–746

Gonzalez, P. and P. Mora. (1995) An analysis of the international tourism demand in Spain. International Journal of Forecasting 11: 233-251.

Green, W. H. (2012). Econometric Analysis (7th edition). Pearson Education Limited.

Greenidge, K. (2001) Forecasting tourism demand: an STM approach. Annals of Tourism Research 28(1): 98-112.

Grigolon, A. B., Borgers, A. W., Kemperman, A. D., & Timmermans, H. J. (2014). Vacation length choice: A dynamic mixed multinomial logit model. *Tourism Management 41*, 158-167.

Gunn H.F., van der Hoorn A.I.J.M. and Daly, A. J. (1987) Long-Range, Country-Wide Travel Demand Forecasts from Models of Individual Choice presented to Fifth International Conference on Travel Behaviour, Aix-en-Provence.

Gunn, H. and Petersen, E. (2007) Alternative strategies to estimating large-scale travel demand models: the UK National Transport Model, presented to European Transport Conference.

Gunn, H.F., Bradley, M.A. and Hensher, D.A. (1992) "High Speed Rail Market Projection: Survey Design and Analysis", Transportation, 19, 117-139.

Guo, J. Y., and C.R. Bhat (2007) Population Synthesis for Microsimulating Travel Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2014, pp. 92-101.

Hackney, J. (2004) Discrete choice models for long-distance travel based on the DATELINE survey. 4th Swiss Transport Research Conference, Monte Verita/Ascona, March 25-26, 2004.

Hackney, J.K. (2006) Results of discrete choice models for long-distance travel based on the DATELINE survey, Working papers of the Transportation Planning and -Systems Group, 396, IVT, ETH Zürich, Zürich.

Hague Consulting Group (1996) Overview and evaluation of methodologies for the forecasting of induced traffic on new transport infrastructure, European Commission, X I / 653 / 96.

Hague Consulting Group (1998) Value of Dutch Travel Time Savings in 1997: Final Report, to Adviesdienst Verkeer en Vervoer, Rijkswaterstaat, HCG Report 6098, The Hague.

Hague Consulting Group and Accent (1994), The Value of Travel Time on UK Roads. Hague Consulting Group, The Hague

Hammadou, K. & Jayet, H. (2002), La valeur du temps pour les déplacements à longue distance: une évaluation sur données françaises, les Cahiers Scientifiques du Transport, 42/2002, pp. 3-23

Haque, M.M., H.C. Chin, H. Huang (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis & Prevention* 42(1) 203-212.

Harland, K., A. Heppenstall, D. Smith, and M. Birkin (2012) Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques. Journal of Artificial Societies and Social Simulation, 15 (1). pp. 1-15.

Hellstrom, J. (2006) A Bivariate Count Data Model for Household Tourism Demand. Journal of Applied Economics, 21: 213-226.

Henderson, J. and A. Trani. (2008)Temporal Transferability of Long-Distance Trip Rates. Transportation Research Board 2008 Annual Meeting Conference Proceedings.

Hensher, D.A. (1997) A Practical Approach to Identifying the Market for High Speed Rail in the Sydney-Canberra Corridor, Transportation Research, 31 A(6). 431-446.

Hensher, D.A. (2001), Value of Travel Time Savings for Inter-City Travel with Special Reference to Air Travel: A Review of the Empirical Evidence, Institute of Transport Studies, The University of Sydney.

Herriges, J.A. and D.J. Phaneuf. (2000) Controlling for Correlation Across Choice Occasions and Sites in a Repeated Mixed Logit Model of Recreation Demand. Technical Chapter. Iowa State University.

Hess, S. & Polak, J.W. (2006a), Airport, airline and access mode choice in the San Francisco Bay area, Papers in Regional Science, 85(4), pp. 543-567.

Hess, S. & Polak, J.W. (2006b), Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area, Transportation Research Part E, 42, pp. 63-81.

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.

Hess, S., Fowler, M., Adler, T. & Bahreinian, A. (2011), The use of cross-nested logit models for multi-dimensional choice processes: the case of the demand for alternative fuel vehicles, Transportation, accepted for publication.

Hong, S.-K., Kim, J.-H., Jang, H., Lee, S. (2006) The roles of categorization, affective image and constraints on destination choice: An application of the NMNL model. Tourism Management. 27: 750-761.

Hu, B. and A.M. Morrison. (2002) Tripography: Can destination use patterns enhance understanding of the VFR market? Journal of Vacation Marketing. 8(3): 201-220.

Hubert, J. P. and F. Potier (2003) What is known? In K. W. Axhausen, J.-L. Madre, J. W. Polak and Ph. L. Toint (eds.) Capturing Long Distance Travel, 45-70, Research Studies Press, Baldock.

Hulkrantz, L. (1995) On the determinants of Swedish Recreational Domestic and Outbound Travel, 1989-1993. Tourism Economics. 1:119-145.

Hwang, Y. and D.R. Fesenmaier. (2003) Multidestionation Pleasure Travel Patterns: Empirical Evidence from the American Travel Survey. Journal of Travel Research. 42:166-171.

Hyde, K.F. and R. Lawson. (2003) The Nature of Independent Travel. Journal of Travel Research. 42: 13-23.

Ishaq, R., Bekhor, S. & Shiftan, Y. (2012), Flexible Model Structure for Discrete Choice Models, working paper.

Iso-Ahola, S.E. (1983). Towards a social psychology of recreational travel. *Leisure Studies* 2, 45-56.

Jang, S. C. S., Bai, B., Hong, G. S., & O'Leary, J. T. (2004). Understanding travel expenditure patterns: A study of Japanese pleasure travellers to the United States by income level. *Tourism Management* 25(3), 331-341.

Jara-Diaz, S.R., M.A. Munizaga, P. Greeven, R. Guerra, K. Axhausen. (2008) Estimating the value of leisure from a time allocation model. Transportation Research Part B. Forthcoming.

Jin, X. and A Horowitz. (2008) Time of Day Choice Modeling for Long Distance Trips. Transportation Research Board 2008 Annual Meeting Conference Proceedings.

Johnson, P., Ashworth, J. (1990) Modelling tourism demand: a summary review. Leisure Studies 9:145-160.

Joint Research Centre (2008), documentation of Trans-Tools models, available at http://energy.jrc.ec.europa.eu/transtools/TT_model.html.

Jones, C.A., Y. Sung, and H.Z. Chen. (1997) Modeling Demand for Recreational Trips of Differing Lengths: A Competing Risks Duration Model with Censoring and Time-varying Covariates. Technical Paper. NOAA Damage Assessment Center, MD.

Jong, G. de and Gunn, H. (2003) A meta-model for passenger transport in Europe integrating existing models, European Transport Conference.

Jong, G. de, Daly, A.J., Gunn, H.F., Blom, U. (1998) Forecasting Car Occupancy: literature Review and Model Development. In: J. de D. Ortúzar, D Hensher, S. Jara-Diaz (ed.) Travel Behaviour Research: Updating the State of Play, Elsevier.

Julien, H. and Morellet, O. (1990) M.A.T.I.S.S.E., un modèle intégrant étroitement induction et partage modal fin du trafic, Rapport INRETS No. 129. (in French)

Kandelaars, P. (1997) A Dynamic Simulation Model of Tourism and Environment in the Yucatán Peninsula. INTERIM REPORT IR-97-18/April. International Institute for Applied Systems Analysis.

Karlström, A., Sundberg, M. & Wang, Q. (2013), Consistently estimating flexible route choice models using an MNL lens, KTH Royal Institute of Technology

Kazimi, C. (1994, December). Hazard models of changing household demographics. Retrieved December 31, 2013, from Institute of Transportation Studies, University of California, Irvine: http://www.its.uci.edu/its/publications/papers/ITS/UCI-ITS-WP-94-11.pdf

Keane, M., Smith, A.A. (2003), Generalized indirect inference for discrete choice models, Manuscript (Yale University).

Kemperman, A., A. Borgers, H. Oppewal, and H. Timmermans. (2003) Predicting the Duration of Theme Park Visitors' Activities: An Ordered Logit Model Using Conjoint Choice Data. Journal of Travel Research. 41: 1-10.

König, A., Abay, G. and Axhausen, K.W. (2004), Zeitkostenansaetze im Personenverkehr, Schriftenreihe, 1065, final report for SVI 2001/534, Bundesamt für Strassen, UVEK, Bern.

Kouwenhoven, M. and Kroes, E. (2006) Welfare effects of capacity constraints at Schiphol airport – a new model to forecast air demand, European Transport Conference.

Kouwenhoven, M., Rohr. C., Miller S., Siemonsma H., Burge P., and J. Laird (2007) Isles of Scilly: Travel Demand Study, RAND Technical Report, available on http://www.rand.org/pubs/.

Kroes, E. and Fox, J. (2001) HSL- Zuid spreadsheet simulation system, European Transport Conference.

Kurri, J. & Pursula, M. (1995) Finnish preliminary value of time studies. KFB & VTI forskning/research 14, del 4. p. 50—70.

Laesser, C., P. Beritelli, T. Bieger (2009). Solo travel: Explorative insights from a mature market (Switzerland). *Journal of Vacation Marketing* 15(3), 217-227.

LaMondia, J., and C.R. Bhat, (2009) "A Conceptual and Methodological Framework of Leisure Activity Loyalty Accommodating the Travel Context: Application of a Copula-Based Bivariate Ordered-Response Choice Model," Technical chapter, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin, July .

LaMondia, J., C.R. Bhat, and D.A. Hensher (2008) An Annual Time Use Model for Domestic Vacation Travel. Journal of Choice Modelling 1(1): 70-97.

LaMondia, J., T. Snell, and C.R. Bhat (2010), "Traveler Behavior and Values Analysis in the Context of Vacation Destination and Travel Mode Choices: European Union Case Study,"Transportation Research Record, Vol. 2156, pp. 140-149.

LaMondia, J.J., and C.R. Bhat (2011), "A Study of Visitors' Leisure Travel Behavior in the Northwest Territories of Canada," Transportation Letters: The International Journal of Transportation Research, Vol. 3, No. 1, pp. 1-19

Larsen, J. (2008) De-exoticizing leisure travel. Leisure Studies 27(1): 21-34.

Larsen, J., J. Urry, and K.W. Axhausen. (2007) Networks and Tourism: Mobile Social Life. Annals of Tourism Research. 34 (1): 244-262.

Lee, D.-H., Fu, Y., 2011. Cross-Entropy Optimization Model for Population Synthesis in Activity-Based Microsimulation Models. Transportation Research Record: Journal of the Transportation Research Board 2255, pp. 20-27.

Lehto, X.Y., O'Leary, J.T., Morrison, A.M. (2004) The effect of prior experience on vacation behavior. Annals of Tourism Research 31(4): 801-818.

Leung, L., & Lee, P. S. N. (2005). Multiple determinants of life quality: The roles of Internet activities, use of new media, social support, and leisure activities. Telematics and Informatics, 22(3), 161–180.

Lew, A. and B. McKercher. (2006) Modeling Tourist Movements: A Local Destination Analysis. Annals of Tourism Research. 33(2): 403-423.

Lim, C. (1997) Review of international tourism demand models. Annals of Tourism Research 24(4): 835-849.

Limatanakool, N., M. Dijst and T. Schwanen. (2006) On the Participation in Medium- and Long-Distance Travel: A Decomposition Analysis for the UK and the Netherlands. Tijdschrift voor Economische en Sociale Geografie. 97(4): 389–404.

Littrell, M.A., R.C. Paige, and K. Song. (2004) Senior Travelers: Tourism Activities and Shopping Behaviours. Journal of Vacation Marketing. 10(4): 348-362.

Lloyd, K. M., & Auld, C. J. (2002). The role of leisure in determining quality of life: Issues of content and measurement. Social Indicators Research, 57, 43–71.

Mallett, W.J. (1999) Long-Distance Travel by Low-Income Households. TRB Transportation Research Circular E-CO26 –Personal Travel: The Long and Short of It. 169-175.

Mason K.J. (2005) Observations of fundamental changes in the demand for aviation services. *Journal of Air Transport Management.* 11(1): 19-25.

Mason, K.J. (2002) Future trends in business travel decision-making. *Journal of Air Transportation.* 7(1):47-68.

Melin, R., Fugl-Meyer, K. S., & Fugl-Meyer, A. R. (2003). Life satisfaction in 18- to 64-year-old Swedes: In relation to education, employment situation, health and physical activity. Journal of Rehabilitation Medicine, 35(2), 84–90.

Menec, V. H. (2003). The relation between everyday activities and successful aging: A 6-year longitudinal study. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58(2), S74– S82.

Mergoupis, T., & Steuer, M. (2003). Holiday taking and income. *Applied Economics 35*(3), 269-284.

Møller, L., Wätjen, W., Pedersen, K.S., Daly, A.J. (1999) Traffiken på Storebælt (Traffic across the Great Belt, published in Danish), Dansk Vejtidsskrift; (1999) Traffic across the Great Belt, English translation presented to International Road Federation regional conference, Lahti, Finland.

Morley, C.L. (1992). A microeconomic theory of international tourism demand. Annals of Tourism Research 19, 250-267.

Mottiar, Z. and D. Quinn. (2004) couple dynamics in household tourism decision making: Women as the gatekeepers? Journal of Vacation Marketing. 10(2): 149-160.

Müller, K. and K. W. Axhausen (2011) Population synthesis for microsimulation: State of the Art. Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, D.C.

Müller, K. and K. W. Axhausen (2012) Multi-level Fitting Algorithms for Population Synthesis. Paper presented at the 1st European Symposium on Quantitative Methods in Transportation Systems, Lausanne, September 2012.

Nickerson, N.P. and C. Jurowski. (2001) The Influence of Children on Vacation Travel Patterns. Journal of Vacation Marketing, 7: 19-30.

Nicolau, J. L., & Más, F. J. (2005). Heckit modelling of tourist expenditure: evidence from Spain. *International Journal of Service Industry Management 16*(3), 271-293.

Nicolau, J.L. and F.J. Mas. (2004) Analysing Three Basic Decisions of Tourists: Going Away, Going Abroad and Going on Tour. Instituto Valenciano de Investigaciones Economicas, S.A., February.

Nicolau, J.L., and F.J. Mas. (2005) Stochastic modeling: a three-stage tourist choice process. Annals of Tourism Research 32(1): 49-69.

Oppermann, M. (1995) A Model of Travel itineraries. Journal of Travel Research. Spring: 57-61

Ortúzar, J. de D. and Gonzalez, R.M. (2001) Inter-island travel demand response with discrete choice models: functional form, forecasts and elstaicities, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile.

Ory, D.T., Mokhtarian, P.L. (2007) Structural equation models of long-distance travel attitudes and behavior. Transportation Research Board 2008 Annual Meeting Conference Proceedings.

Outcrop. (2007) Strategic Overview of Current and Potential Visitor Markets of the Northwest Territories. Report prepared for the Northwest Territories. December.

Outwater, Maren; Tierney, Kevin; Bradley, Mark; Sall, Elizabeth; Kuppam, Arun; Modugula, Vamsee; California Statewide Model for High-Speed Rail, published Journal of Choice Modeling, Volume 3, No 1, March 2010.

Papatheodorou, A. (2001) Why people travel to different places. Annals of Tourism Research 28(1): 164-179.

Parsons Brinckerhoff, HBA Specto Incorporated, and EcoNorthwest. (2010) Oregon Statewide Integrated Model (SWIM2) Model Description. Draft Report Version 2.5. Submitted to the Oregon Department of Transportation.

Parsons Brinckerhoff, Northwestern University, Mark Bradley Research & Consulting, University of California at Irvine, Resource Systems Group, University of Texas at Austin, Frank Koppelman, Geostats (2013) Understanding Our Understanding of How Highway Congestion and Pricing Affect Travel Demand, Strategic Highway Research Program 2 Capacity Research, Report S2-C04-RW-1, Washington, DC.

Parsons, G.R. and M.J. Kealy. (1995) A Demand Theory for Number of Trips in a Random Utility Model of Recreation. Journal of Environmental Economics and Management. 29: 357-367.

Pearce, P.L. and U. Lee. (2005) Developing the Travel Career Approach to Tourist Motivation. *Journal of Travel Research*. 43:226-237.

Pedersen, K., Denver, L., Daly, A. and Rohr, C. (2001) Great Belt link – traffic before and after, presented to European Transport Conference, Cambridge.

Pew Center. (2013, June 5). Smart Phone Ownership 2013. Retrieved December 20, 2013, from Pew Internet: http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013/Findings.aspx

Pritchard, D.R. and E.J. Miller (2012) Advances in Population Synthesis: Fitting Many Attributes per Agent and Fitting to Household and Person Margins Simultaneously. Transportation 39(3), pp. 685-704. Retrieved February 28, 2013 from http://www.rti.org/rtipress.

Ramjerdi, F., Flügel, S., Samstad, H. & Killi, M. (2010) Value of time, safety and environment in passenger transport – Time. Summary in English. Available: http://www.toi.no/article29726-29.html.

Rodriquez, A., P. Latvoka and Y. Sun. (2008) The relationship between leisure and life satisfaction: application of activity and need theory. Social Indicator Research. 86:163–175.

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 (accepted for publication and expected publication in December 2013).

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., Tsang, F. (2010) Modelling long-distance travel in the UK, presented to European Transport Conference, Glasgow.

Schlich, R. S. Schonfelder, S. Hanson, and K.W. Axhausen. (2002) Leisure travel in a historical perspective – Changes in the structures of time and space use. Arbeitsbericht Verkehrsund Raumplanung, 197, Institut fur Verkehrsplanung und Transportsysteme, ETH Zurich, Zurich.

Schlich, R., S. Schonfelder, S. Hanson, and K.W. Axhausen. (2004) Structures of Leisure Travel: Temporal and Spatial Variability. Transport Reviews, 24(2): 219-237, March.

Seddighi, H.R., Theocharous, A.L. (2002) A model of tourism destination choice: a theoretical and empirical analysis. Tourism Management 23: 475-487.

Siderelis, C. and L. Gustke. (2000) Influence of On-Site Choices on Recreation Demand. Leisure Sciences. 22:123-132.

Simma, A., R. Schlich, and K.W. Axhausen. (2002) Destination Choice Modelling for Different Leisure Activities. Arbeitsberichte Verkehrs- und Raumplanung, 99, Institut für Verkehrsplanung und Transporttechnik, Strassen- und Eisenbahnbau, ETH Zürich, Zürich.

Simma, A., Schlich, R., Axhausen, K.W. (2001) Destination choice of leisure trips: the case of Switzerland, Arbeitsberichte Verkehrs- und Raumplanung, 99, Institut fur Verkehrsplanung und Transporttechnik, Strassen- und Eisenbahnbau, ETH Zurich, Zurich.

Simonetti, C. and Ortúzar, J. d. D. (2001) Modeling air travel and deriving values of time with the mixed data estimation method, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile.

Sirakaya, E., Woodside, A.G. (2005) Building and testing theories of decision making by travelers. Tourism Management 26: 815-832.

Smith, A.A. (1993), Estimating nonlinear time-series models using simulated vector autoregressions. Journal of Applied Econometrics 8, Supplement: Special issue on Econometric Inference using Simulation Techniques, S63–S84.

Srinivasan, S., L. Ma, and K. Yathindra (2008) Procedure for Forecasting Household Characteristics for Input to Travel-demand Models. Florida Department of Transportation, Final Report TRC-FDOT-64011-2008, Tallahassee, FL.

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

Sundararajan, A. G. (2003). Demographic microsimulation with DEMOS 2000: Design, validation, and forecasting. In K. e. Goulias, Transportation systems planning: Methods and applications (pp. Chapter 14, 1-23). CRC Press.

Tideswell, C. and B. Faulkner (1995) Multidestination travel patterns of international visitors to Queensland. Journal of Travel Research. 37: 364-374

Train, Kenneth E. (2009) Discrete Choice Methods with Simulation, Second Edition, University of California Berkeley, Cambridge University Press, New York.

Transportation Research Board. (1999) Statewide Travel Demand Forecasting Conference Proceedings. Transportation Research Circular. E-C011. September 1999.

Transportation Research Board. (2013, November 7). National Highway Cooperative Research Program. Retrieved December 20, 2013, from NCHRP 08-95 RFP: http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=3637

Transportation Research Board. (2014 (forthcoming)). Effects of Socio-Demographics on Travel Demand (Project 20-83(06)). Washington DC: National Highway Cooperative Research Program .

U.S. Bureau of Labor Statistics. (2013, December 19). EP Tables. Retrieved December 20, 2013, from Employment Projections: http://www.bls.gov/emp/tables.htm

U.S. Census Bureau. (2012). 2012 National Population Projections: Summary Tables. Retrieved December 20, 2013, from Population Projections: http://www.census.gov/population/projections/data/national/2012/summarytables.html

ULTRANS, Institute of Transportation Studies, UC Davis and HBA Specto Incorporated (2011) CSTDM09 California Statewide Travel Demand Model: Model Development Long Distance Person Travel Model, Final System Documentation: Technical Note, available at http://www.dot.ca.gov/hq/tsip/smb/documents/tdm/CSTDM09_LDPTM_Final.pdf

University of Michigan. (2013). Panel study of income dynamics. Retrieved December 31, 2013, from PSID Data on Families since 1968: http://psidonline.isr.umich.edu/Guide/Brochures/PSID.pdf

Urry, J. (2003) Social networks, travel and talk. British Journal of Sociology. 54(2): 155-175.

US Department of Transportation (2003). America on the Go: US Business Travel. *Special Topic Analytical Report.* Bureau of Transportation Statistics: October.

van Middlekoop, A. Borgers, and H.J.P. Timmermans. (2004) Merlin: Microsimulation System for Predicting Leisure Activity-Travel Patterns. Transportation Research Record, 1894: 20-27.

Wang, Q., Karlström, A. & Sundberg, M. (2013), Bias correction via Indirect Inference for Mixed Logit estimation under sampling of alternatives, KTH Royal Institute of Technology

WebTAG (2012), http://www.dft.gov.uk/webtag/documents/index.php, UK Department for Transport.

Wheaton, W.D., J.C. Cajka, B.M. Chasteen, D.K. Wagener, P.C. Cooley, L. Ganapathi, D.J. Roberts, and J.L. Allpress (2009) Synthesized Population Databases: A US Geospatial Database for Agent-Based Models. RTI Press publication No. MR-0010-0905. Research Triangle Park, NC: RTI International.

Williamson, P., M. Birkin, P.H. Rees (1998) The Estimation of Population Microdata by Using Data from Small Area Statistics and Samples of Anonymised Records. Environment and Planning A, 30(5), pp. 785–816.

Witt, S.F., Witt, C.A. (1995) Forecasting tourism demand: a review of empirical research. International Journal of Forecasting 11:447-475.

Woodside, A.G., and R. MacDonald. (1994) General System Framework of Customer Choice Processes of Tourism Services. In Spoilt for Choice. Decision Making Processes and Preference Change of Tourists: Intertemporal and Intercountry Perspectives, R. Gasser and K. Weiermair, eds., pp. 30–59. Thaur: Kulturverlag.

Wu, L., Zhang, J., & Fujiwara, A. (2013). Tourism participation and expenditure behaviour: Analysis using a scobit based discrete–continuous choice model. *Annals of Tourism Research 40*, 1-17.

Wu, Y. (2008) The Integrated Florida Statewide Model. Presentation. SE Florida DSUTMS Users Group Meeting. February 8, 2008.

Ye, X., K. Konduri, R.M. Pendyala, B. Sana, and P. Waddell (2009) A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations. Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, D.C. Available at http://urbanmodel.asu.edu/popgen/papers/PopulationSynthesizerPaper_TRB.pdf.