Federal Highway Administration

National Long-Distance Passenger Model Documentation

Model Development

Exploratory Advanced Research Program

DTFH61-10-R-00036

June 2018

U.S.Department of Transportation
Federal Highway Administration

Technical Report Documentation Page

<table>
<tr><td>1. Report No.
FHWA-PL-18-022</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework: Model Documentation</td><td colspan="2">5. Report Date
June 2018</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)
Maren Outwater, Mark Bradley</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
RSG
55 Railroad Row
White River Junction, VT 05001</td><td colspan="2">10. Work Unit No. (TRAIS)</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTFH61-11-C-00015</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period Covered

Model Documentation</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes
The project was managed by Task Manager for the Federal Highway Administration, Wenjing Pu, who provided technical direction.</td></tr>
<tr><td colspan="4">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 and focused on enhancing 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. The modeling framework was implemented with application software that simulates long-distance travel for all households in the United States (rJourney). This model documentation reports on the data and modeling process applied to estimate long-distance passenger travel for the United States.</td></tr>
<tr><td colspan="2">17. Key Words
Long-Distance; National Passenger Models; Intercity Travel; Scheduling Models; Joint Mode and Destination</td><td colspan="2">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
195</td><td>22. Price
N/A</td></tr>
</table>

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

AADT	average annual daily traffic
ABA	American Bus Association
ACS	American Community Survey
ATS	American Travel Survey
BLS	Bureau of Labor Statistics
BPR	Bureau of Public Roads
BTS	Bureau of Transportation Statistics
CPS	Current Population Survey
CSV	comma-separated values
FAF	Freight Analysis Framework
FAF^3	Freight Analysis Framework Version 3
FHWA	Federal Highway Administration
FIPS	Federal Information Processing Standard
FRA	Federal Railroad Administration
GISDK	Geographic Information System Developer's Kit
GTFS	General Transit Feed Specification
HBNW	home-based non-work
HBW	home-based work
HH	household
HPMS	Highway Performance Monitoring System
IPF	iterative proportional fitting
LEHD	Longitudinal Employer Household Dynamics
LODES	LEHD Origin-Destination Employment Statistics
LOS	level-of-service
NHB	non-home-based
NHPN	National Highway Planning Network
NHTS	National Household Travel Survey
NPS	National Park Service
NUMA	National Use Model Area
O-D	origin-destination
ODME	origin-destination matrix estimation
OHPI	Office of Highway Policy Information
PUMA	Public Use Microdata Area
PUMS	Public Use Microdata Sample
QCEW	Quarterly Census of Employment and Wages
QRM	quick response methods
TAF	Traveler Analysis Framework
VMT	vehicle miles traveled

CHAPTER 1. INTRODUCTION

Intercity travel is increasingly important in the United States. The Federal government and many States are faced with improving mobility and reducing impacts for these travelers. FHWA has invested in several studies to better understand intercity travel; this study is an extension of that interest, which began with exploratory research to develop a long-distance passenger travel demand model framework and grew to include implementation of that 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 work-related travel (employer's business and commute) and nonwork travel (visiting friends and family, personal business and shopping, and leisure). The models are multimodal (auto, rail, bus, and air) and based on national networks for each mode. This provides opportunities to evaluate intercity transportation investments or test national economic, environmental, and pricing policies.

1.1 Overview of Related Products

This technical report documents the model development portion of the DTFH61-10-R-00036 Exploratory Advanced Research program to develop Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework. The original work included three phases: a design phase, a research phase, and an implementation phase focused on moving the research into practice and providing a model that State and Federal agencies interested in long-distance passenger travel can use. The original research concluded with the following products:

- Long-Distance Passenger Travel Demand Modeling Framework Final Report. Please visit the Long-Distance Passenger Travel Demand Modeling Framework Final Report: https://www.fhwa.dot.gov/policy/modelframework/model_framework.pdf.
- Long-distance passenger travel demand model framework, with models estimated from available data.
- rJourney software to implement the long-distance passenger travel demand models.
- Long-Distance Passenger Travel Demand Modeling Framework Implementation Report. Please visit the Long-Distance Passenger Travel Demand Modeling Framework Implementation Report: https://www.fhwa.dot.gov/policyinformation/analysisframework/docs/long-distance_model_implementation_report_final.pdf.

This report expands upon detail on data sources and mathematical formulations and synthesizes relevant portions of the Final Report and Implementation Report. This synthesis provides a comprehensive documentation of the model development, calibration, and validation of the long-distance passenger travel demand model. This report also presents results from the sensitivity tests and comparative data analysis. A companion user guide provides instruction on using rJourney for planning applications.

This long-distance passenger model research did not include any new data collection, so models were estimated based on long-distance surveys collected from several States (Ohio, Colorado, Wisconsin, California, and New York). A long-distance passenger travel survey for the United States is recommended to estimate these models using a comprehensive dataset.

1.2 Overview of the Model System

Methods for modeling long-distance passenger movements are in their infancy in the United States. Federal and State entities have recently become interested in modeling long-distance passenger movements as part of highway infrastructure planning; similarly, agencies studying high-speed rail, or those involved in airport planning, have also expressed interest due to their dependence on long-distance travel markets. This stronger interest at the Federal and State levels has created an intersection of policy needs for long-distance passenger modeling. In practice, some States and regions have expressed interest in long-distance passenger modeling for statewide models (e.g., California, Ohio, and Arizona) and for high-speed rail ridership studies (e.g., Florida, California, and the Northeast Corridor). However, these models rely on traditional travel demand forecasting methods rather than on a robust understanding of the underlying behavior and how and why it is different from other types of passenger travel.

The goal of this research was to develop a framework for a long-distance passenger travel demand model that was used to build a national model for the United States, one based on exploring new ways to simulate behavior of long-distance passenger movements. This national model was estimated, calibrated, and validated on currently available long-distance travel data in the United States. The types of planning applications suitable for the long-distance passenger model include the following:

- Testing national policies (e.g., modal investments, pricing, economics, environmental, livability, safety, and airport/rail planning).
- Measuring system performance.
- Evaluating the impacts of private sector decisions.
- Providing input to statewide and regional planning.
- Assessing regional differences.

The long-distance passenger travel demand forecasting modeling system (Figure 1) synthesizes long-distance travel for each household in the United States (117 million households and 309 million people based on the 2010 Census) using an annual scheduling of long-distance tours (round trips). Household and person characteristics are synthesized for the United States by census tract. The tour generation and joint mode and destination models are the centerpiece of the long-distance passenger models. Models for auto ownership, tour party size, and scheduling were developed to support the primary models.

This long-distance passenger travel demand forecasting modeling system is implemented using software called rJourney. For brevity, the long-distance passenger travel demand forecasting model is referenced in this report as rJourney. rJourney can produce long-distance travel for a specific date or for each day in a month or for each day in a year.

1.3 Contents of the Report

This report comprises 10 chapters. Chapter 1. Introduction includes the introduction and discusses the different products from multiple phases of the work and an overview of the modeling system.

Chapter 2. Data Sources presents the data sources used to develop, estimate, calibrate, and validate the modeling system. This includes information on zones, networks, socioeconomic data, land-use data, origin-destination (O-D) data, household surveys, and traffic counts.

Chapter 3. Long-Distance Model Development discusses the development of the modeling system and each component. Model estimation results are presented along with model estimation results.

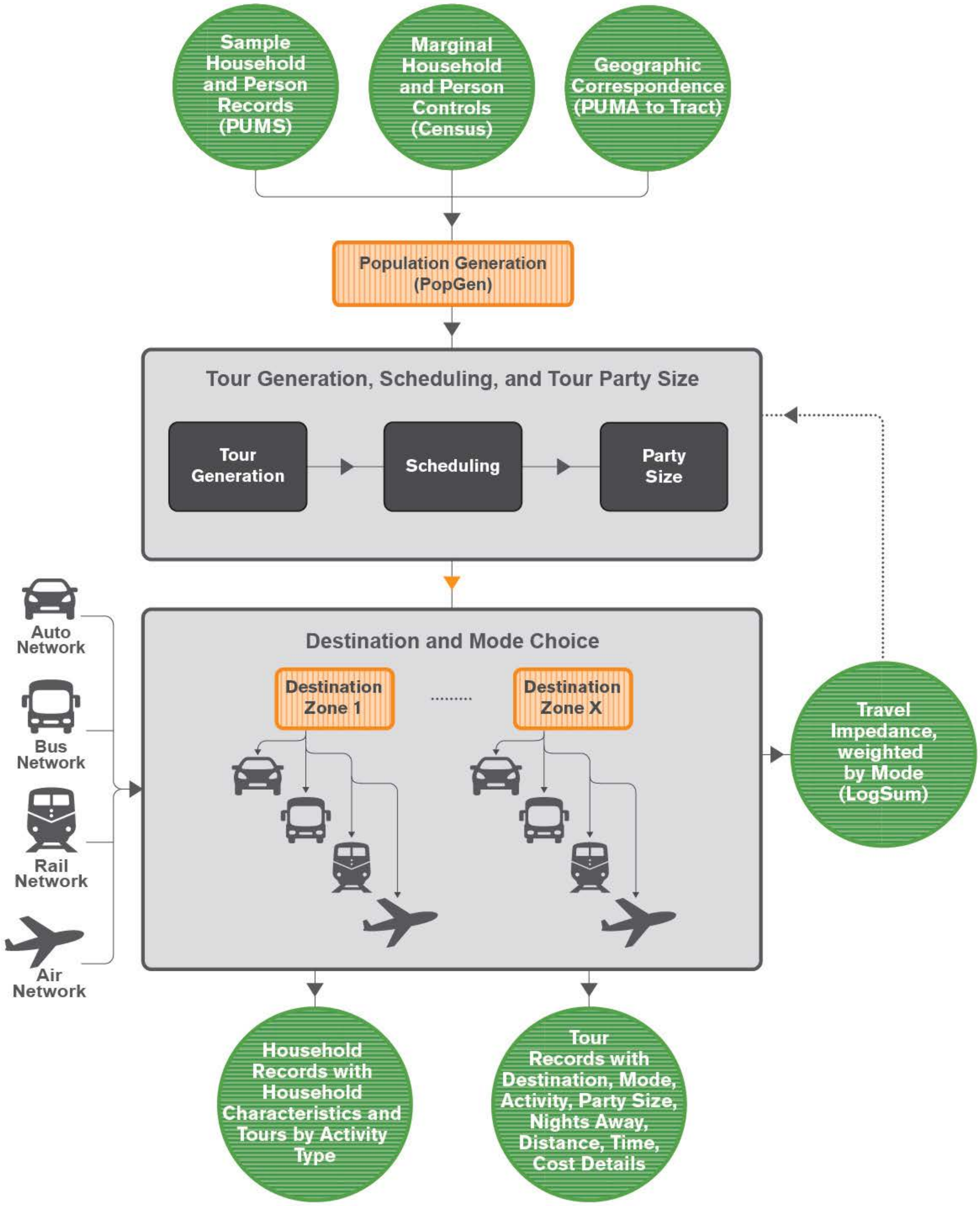
Chapter 4. Model Calibration discusses model calibration and reports the tour generation, destination choice, and mode choice model calibration results. It also includes a description of the preparation of the average daily long-distance passenger travel model trip tables.

Chapter 5. Highway Assignment describes the highway assignment parameters and the highway network. This chapter also includes a description of the background traffic estimation and the assignment application in TransCAD.

Chapter 6. Model Validation describes the trip table and highway performance validation tests. There were five sensitivity tests performed (discussed in Chapter 8. Sensitivity Tests) in addition to the validation tests. These tests were conducted to explore the reasonableness of the models to changes in various inputs.

Chapter 7. Performance Metrics discusses potential performance metrics that are producible with this new set of long-distance models. Five sensitivity tests were performed (discussed in Chapter 8. Sensitivity Tests) in addition to the validation tests. These tests explored the reasonableness of the models to changes in various inputs.

Chapter 9. Comparative Data Analysis presents a comparative data analysis of the long-distance passenger model outputs to the available national datasets. A summary of the report findings is presented in Chapter 10. **SUMMARY**.



Source: FHWA

Figure 1. National long-distance passenger travel demand modeling system.

CHAPTER 2. DATA SOURCES

Applying the long-distance passenger travel demand model to predict long-distance travel behavior of all the households in the United States required preparing several datasets. A summary of the datasets used is provided in Table 1. Each of these datasets is described in more detail in the following sections, along with the key steps in preparing the application datasets.

2.1 Zone Systems

This section summarizes the development of a new zonal system for forecasting long-distance travel at the national scale. This describes the sources the RSG team used to create a new zone system and how the RSG team developed zone connectors.

Zone System Creation

This project created and adopted a new geographical construct, termed the National Use Model Area (NUMA). NUMA-level geography is a composite representation of counties and U.S. 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 [ACS]). Using counties or PUMAs as zones for a national-level travel model is appropriate; both offer a reasonable geographic resolution 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 to enhance the level of detail in the zone system, but with approximately 75,000 census tracts, this was 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. This was done by building travel paths that connect a census tract at the origin to an origin station, connecting the origin station to the destination station, and 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. Long-Distance Model Development); however, it was not included in the demonstration system (Chapter 4. Model Calibration). To support this effort, the census tract was implemented for synthetic population generation.

Table 1. Summary of national long-distance passenger data sources.

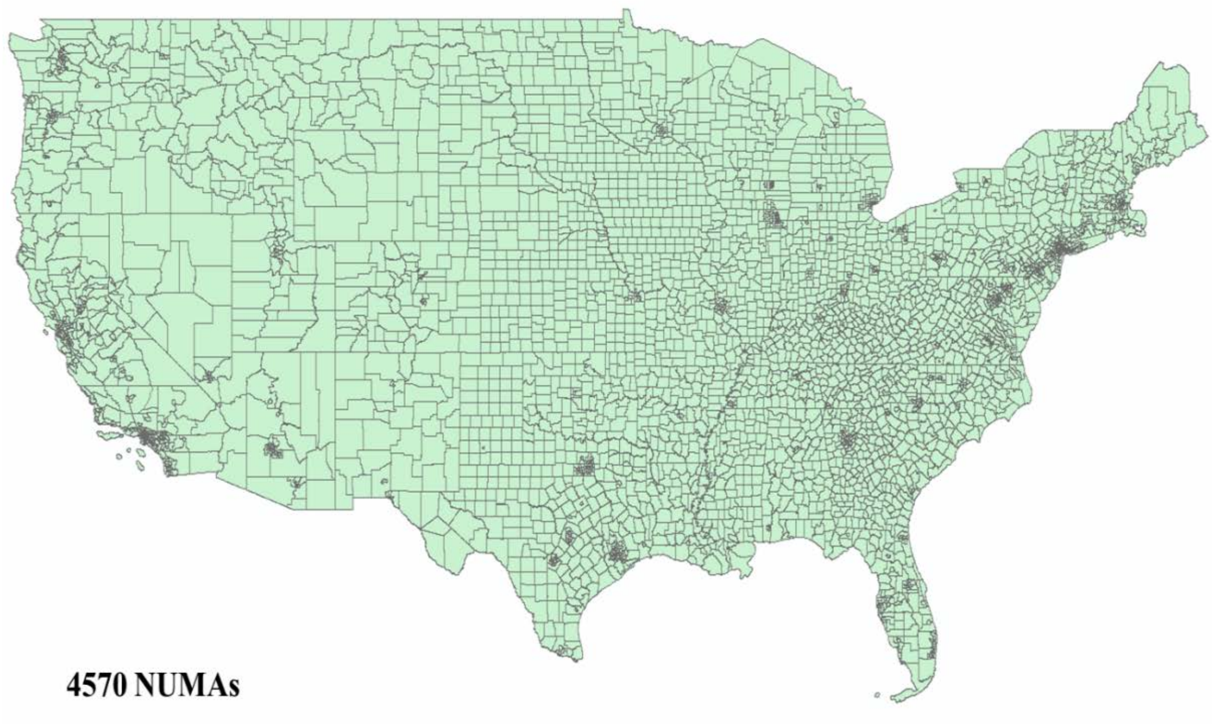
Data Product	Year	Source	Input	Estimation	Calibration	Validation	Comment
Zone System	2013	Census	Yes	No	No	No	Created by Arizona State University
Road System	2011	FHWA	Yes	No	No	No	Centroid connectors added
Toll Facilities	2016	FHWA	Yes	No	No	No	Toll facilities identified; tolls added
Rail System	2011	Amtrak	Yes	No	No	No	Access links added; GTFS data imported
Rail Fares	2004	Amtrak	Yes	No	No	No	Data factored to 2012 levels
Air System	2012	BTS	Yes	No	No	No	Airport connectors added
Bus System	2015	Bus Service Providers	Yes	No	No	No	Compiled from online schedules
Demographics	2010	Census	Yes	No	Yes	Yes	2010 PUMS and 2007-2011 ACS
Employment Data	2010	Census Bureau of Labor Statistics	Yes	No	No	No	Compiled from Census LEHD and BLS QCEW
Land Use Data	2010	Census	Yes	No	No	No	N/A
Park Data	2012	National Park Service	Yes	No	No	No	TomTom and ESRI data used to supplement NPS
Enrollment Data	2011	National Center for Education	Yes	No	No	No	N/A
Origin-Destination Data	2011	FHWA	No	No	No	Yes	Traveler Analysis Framework Interpolated from 2008 & 2040
Bus Ridership	2014	FHWA	No	No	No	Yes	Intercity Bus Ridership project
American Travel Survey	1995	BTS	No	Yes	No	No	12-month survey of long-distance travel in the U.S.
National Household Travel Survey	2001	FHWA	No	Yes	Yes	Yes	4-week survey of long-distance travel in the U.S.
California Household Travel Survey	2012	California DOT	No	Yes	Yes	No	8-week survey of long-distance travel in California
Colorado Front Range Travel Survey	2010	Colorado MPOs	No	No	Yes	No	2-week survey of long-distance travel in eastern Colorado
Ohio Household Travel Survey	2003	Ohio DOT	No	Yes	Yes	No	2-week survey of long-distance travel in Ohio
Traffic Counts	2007	FHWA	No	No	No	Yes	HPMS added to road system network
Vehicle Miles Traveled	2013	FHWA	No	No	No	Yes	Rural vehicle miles from the Highway Statistics Manual

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. 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 was untenable. 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. To define the geographic zone system for the national travel model developed in this study, the project team used the smaller of the two geographies 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 offers a reasonable geographic representation that is neither too large nor too small in its definition in the context of modeling long-distance travel.

Following the 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.

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 like 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 near one another, NUMAs were split through a manual process so that the rail stations were dispersed across multiple NUMAs to the extent possible. Because of this process, some NUMAs (particularly in the dense Northeast) may contain “pockets” of closely spaced 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 the network level-of-service (LOS) data for highway (auto and bus) modes follow this geographical resolution. The final NUMA map for the United States is shown in Figure 2. 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.



Source: FHWA

Figure 2. 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 U.S. highway network downloaded from the FHWA Highway Performance Monitoring System (HPMS) website[1] was also converted to a TransCAD network file. The NUMA centroid point file was overlaid on the U.S. 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.

2.2 Modal Networks and LOS

Road System

The model uses the National Highway Planning Network (NHPN) to generate estimates of travel time, distance, and cost in the form of highway skims. The NHPN, developed by FHWA, is a

[1] Please visit the HPMS Public Release of Geospatial Data in Shapefile Format website: https:/www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm.

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 for this work.[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 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 U.S. 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 U.S. highway network. In addition, a generalized cost skim was also generated for the auto mode. Generalized cost to traverse a link was computed as shown in Figure 3.

$$\begin{aligned} &Generalized\ cost \\ &\qquad = Value\ of\ Time * \ Travel\ Time + (Auto\ Operating\ Cost + Toll) * Length \end{aligned}$$

Figure 3. Equation. Generalized cost to traverse a link.

Where: Value of Time is dollars per hour, Travel Time is hours, Auto Operating Cost is dollars per mile, and Length is miles.

Value of time ($17 per hour) and auto operating cost ($0.18 per mile) were used to compute generalized cost. The user can change these values 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 U.S. 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).

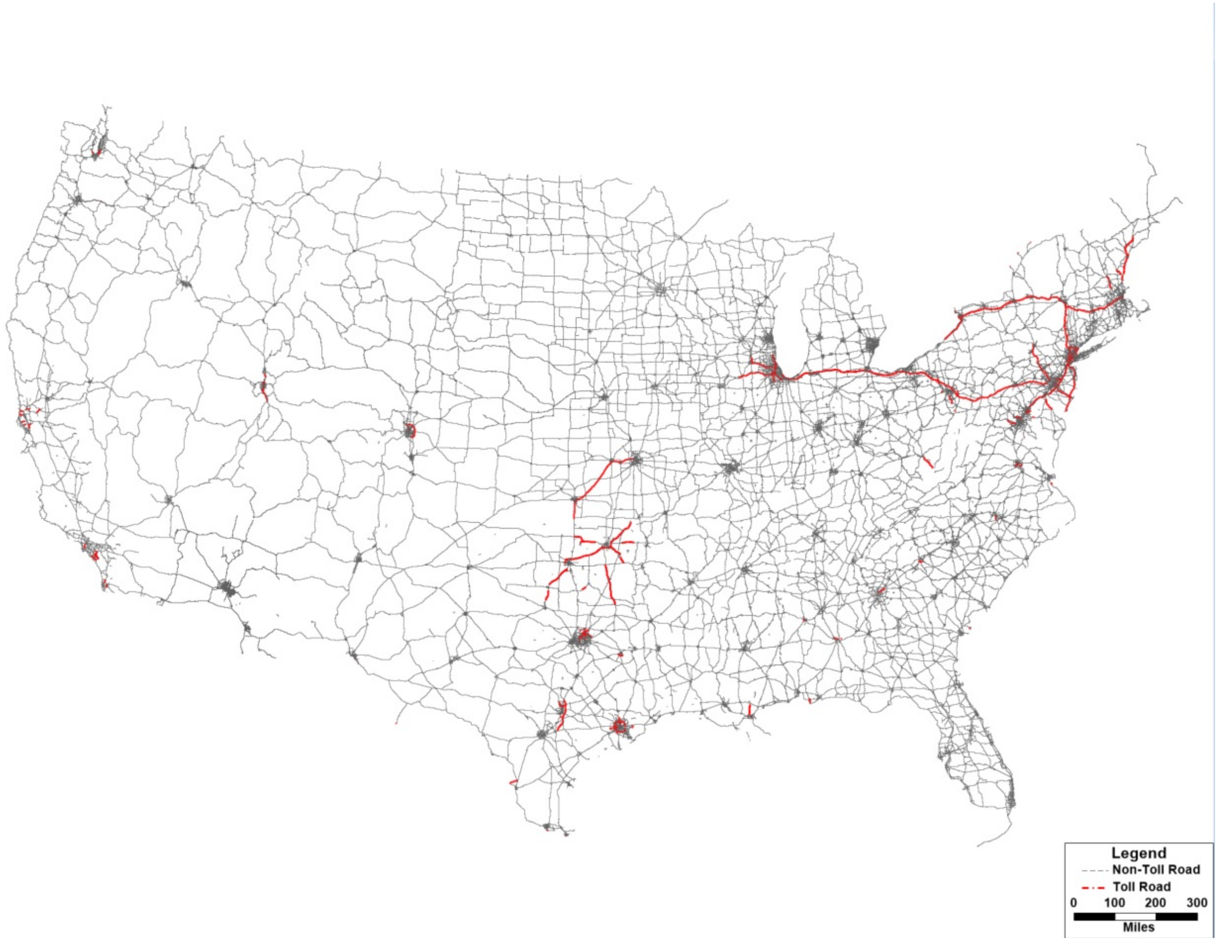
[2] For the most up-to-date U.S. highway network, published in 2011, please visit The NHPN (Version 14.05) website: https://www.fhwa.dot.gov/planning/processes/tools/nhpn/.

Toll Facilities

The model uses shapefiles containing information on the highway network attributes (at the link level) for the United States from the FHWA's HPMS 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 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 were also added manually after a visual inspection of the toll facilities in Google Earth. The toll roads shapefile was merged with the rest of the U.S. 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 4 presents the U.S. highway network with toll roads identified in red lines with dashes and dots.

[3] Please visit the FHWA HPMS website: https://www.fhwa.dot.gov/policyinformation/hpms.cfm.

[4] Please visit the FHWA Toll Facilities in the United States website: https://www.fhwa.dot.gov/policyinformation/tollpage/.



Source: FHWA

Figure 4. Toll roads in the United States.

Rail System

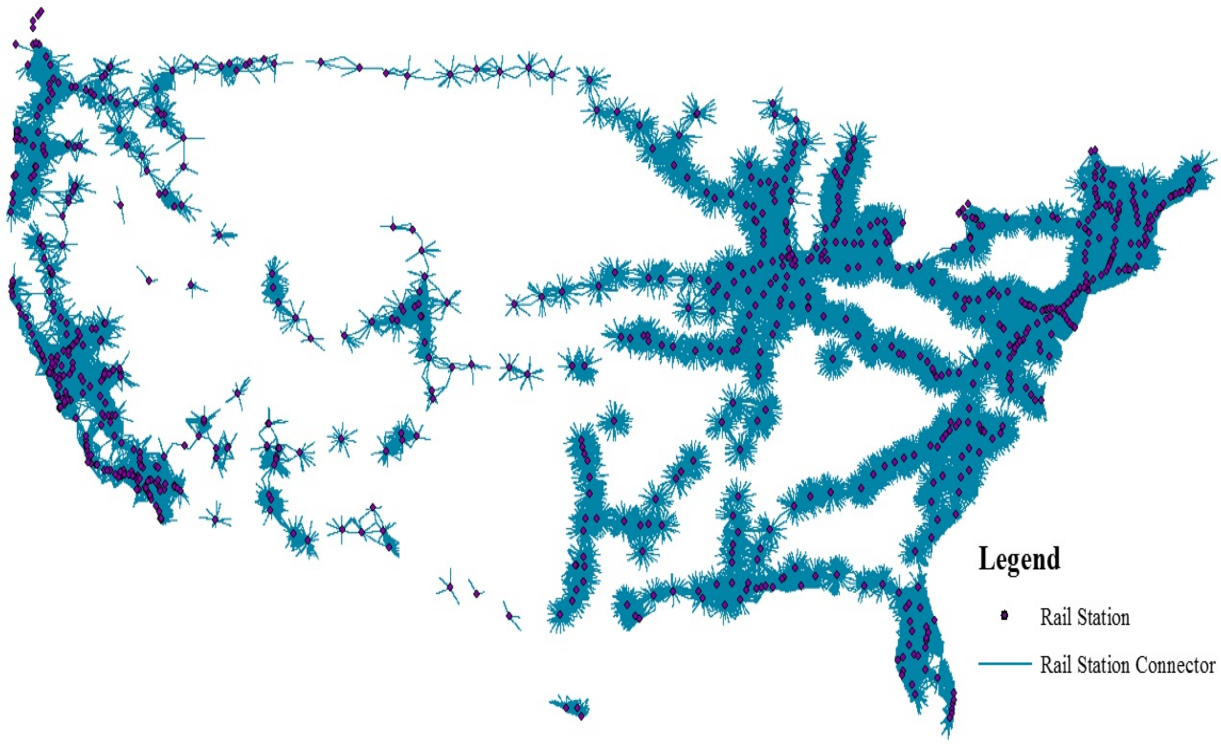
The rail network was developed from the Amtrak rail system (Figure 5). Additional commuter rail systems could be added, but these were not considered essential for this project.

Source: AMTRAK

Figure 5. 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 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 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 6. A census tract can have a connector to all rail stations within 50 miles from the location of its centroid.



Source: FHWA

Figure 6. 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 LOS measures for the national rail network. The GTFS data[5] comprise the following information on various services operated by Amtrak across the nation:

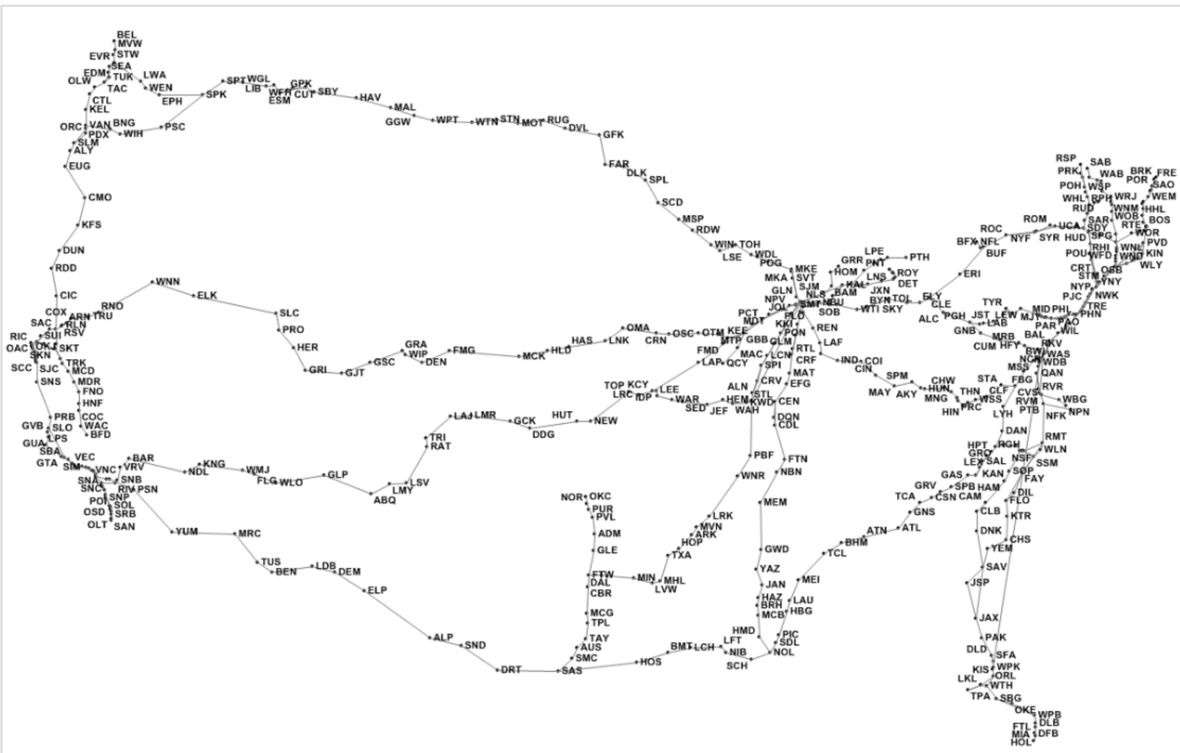
- **Agency**: Contains information on all the transit agencies that provided data in the transit feed.
- **Calendar**: Contains the dates on which a service operates; data regarding start and end times of the service, and the days of the week on which the service operates.
- **Routes**: Contains information 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.

---

[5] Please visit Google Transit APIs General Transit Feed Specification Reference Overview website: https://developers.google.com/transit/gtfs/reference/?csw=1

- **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 7.



Source: FHWA

Figure 7. 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

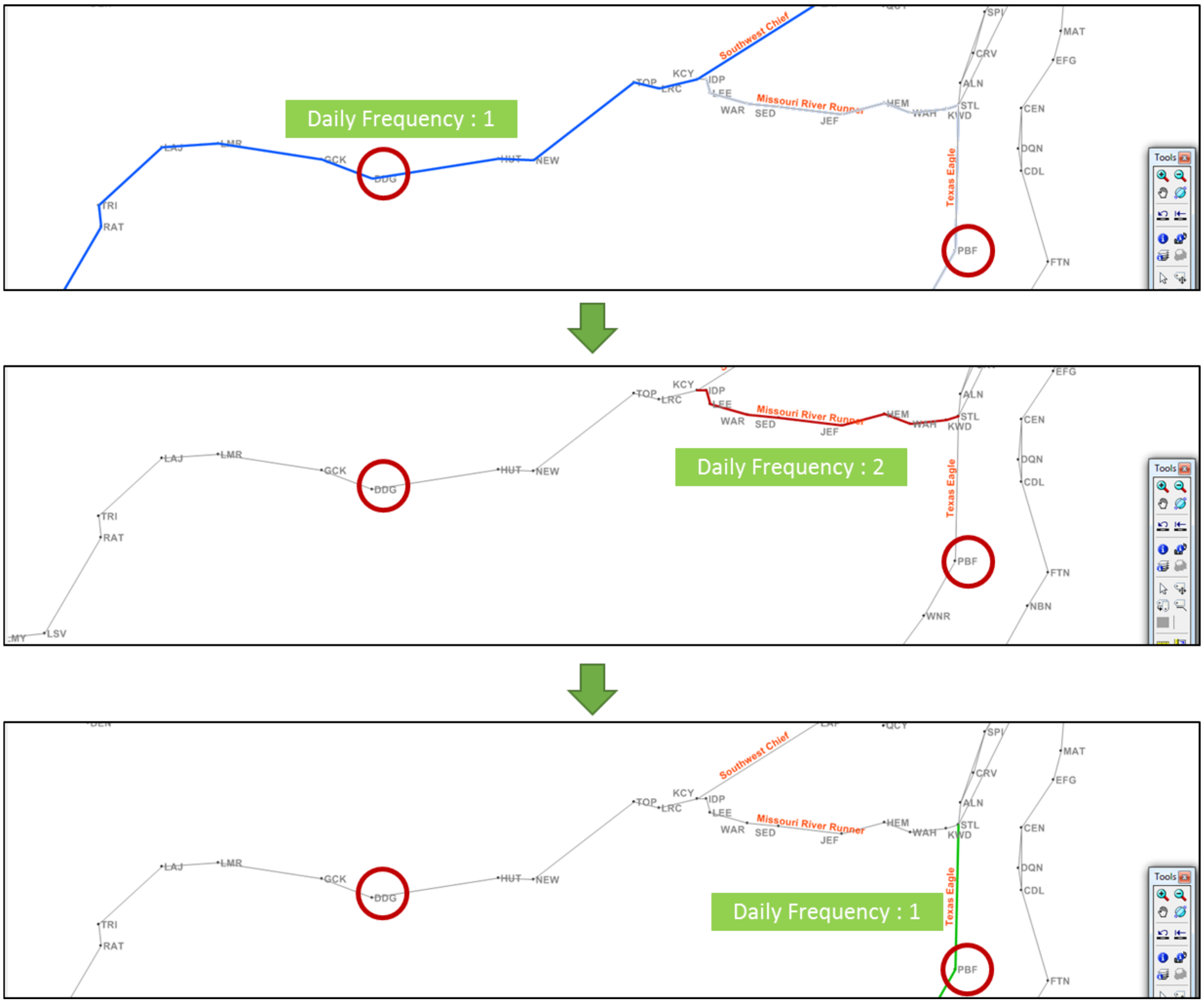
The transfer-frequency matrix defines the minimum number of transfers a traveler needs to make to travel from one Amtrak station to another. Two sets of travel time skims were generated in TransCAD, employing the procedures described in the previous section (i.e., one skim where transfer time is included 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

The project team developed an innovative methodology 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. A few links on the rail network 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 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 8). 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.



Source: FHWA

Figure 8. Rail frequency computation—illustrative example.

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, several 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.

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 LOS variables for other modes, 2004 rail fares were factored up to 2012 levels by using Consumer Price Index values for U.S. 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. Several 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; Figure 9 through Figure 20 summarize the final model results. 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. The model-predicted fares were compared against Amtrak's online reservation fares for a limited number of station pairs with mixed demand. The model performance was 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.

Rail Fare Model, by Region and Fare Class

California (First/Business Class)

$$fare = 16.89 + 0.39 \times distance - 0.07 \times \frac{distance^2}{1000}$$

Figure 9. Equation. California rail fare model (first/business class).

Sample size = 1,256
Adjusted R-squared = 0.85

California (Economy Class)

$$fare = 13.31 + 0.13 \times distance - 0.03 \times \frac{distance^2}{1000}$$

Figure 10. Equation. California rail fare model (economy class).

Sample size = 1,877
Adjusted R-squared = 0.90

Midwest (First/Business Class)

$$fare = 16.17 + 0.32 \times distance - 0.04 \times \frac{distance^2}{1000}$$

Figure 11. Equation. Midwest rail fare model (first/business class).

Sample size = 1,762
Adjusted R-squared = 0.81

Midwest (Economy Class)

$$fare = 7.76 + 0.16 \times distance - 0.04 \times \frac{distance^2}{1000}$$

Figure 12. Equation. Midwest rail fare model (economy class).

Sample size = 2,733
Adjusted R-squared = 0.88

Northeast (First/Business Class)

$$fare = 32.57 + 0.37 \times distance - 0.09 \times \frac{distance^2}{1000}$$

Figure 13. Equation. Northeast rail fare model (first/business class).

Sample size = 2,661
Adjusted R-squared = 0.77

Northeast (Economy Class)

$$fare = 20.02 + 0.23 \times distance - 0.11 \times \frac{distance^2}{1000}$$

Figure 14. Equation. Northeast rail fare model (economy class).

Sample size = 3,674
Adjusted R-squared = 0.72

Northwest (First/Business Class)

$$fare = 20.87 + 0.42 \times distance - 0.11 \times \frac{distance^2}{1000}$$

Figure 15. Northwest rail fare model (first/business class).

Sample size = 607
Adjusted R-squared = 0.81

Northwest (Economy Class)

$$fare = 10.51 + 0.17 \times distance - 0.05 \times \frac{distance^2}{1000}$$

Figure 16. Northwest rail fare model (economy class).

Sample size = 765
Adjusted R-squared = 0.88

South (First/Business Class)

$$fare = 36.32 + 0.33 \times distance - 0.06 \times \frac{distance^2}{1000}$$

Figure 17. Equation. South rail fare model (first/business class).

Sample size = 2,085
Adjusted R-squared = 0.61

South (Economy Class)

$$fare = 22.90 + 0.15 \times distance - 0.03 \times \frac{distance^2}{1000}$$

Figure 18. Equation. South rail fare model (economy class).

Sample size = 3,108
Adjusted R-squared = 0.65

West (First/Business Class)

$$fare = 40.21 + 0.33 \times distance - 0.06 \times \frac{distance^2}{1000}$$

Figure 19. West rail fare model (first/business class).

Sample size = 1,531
Adjusted R-squared = 0.58

West (Economy Class)

$$fare = 14.27 + 0.15 \times distance - 0.03 \times \frac{distance^2}{1000}$$

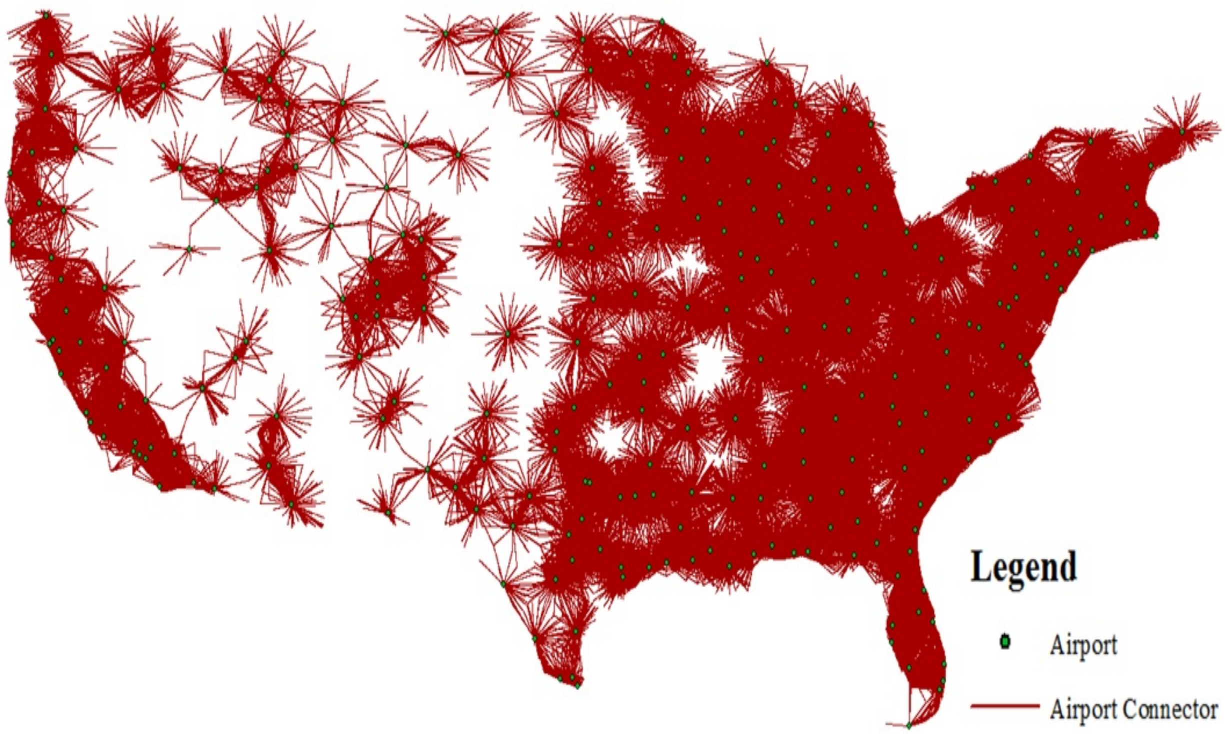
Figure 20. Equation. West rail fare model (economy class).

Sample size = 2,706
Adjusted R-squared = 0.78

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; 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 21. As in the case of rail station connectors, a census tract could have a connector to all airports within 100 miles from the location of its centroid.



Source: FHWA

Figure 21. 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 U.S. Bureau of Transportation Statistics (BTS): 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 percent 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).[6] The DB1B is a 10 percent 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.[7] Using these two databases, an airport-to-airport O-D matrix with the following air LOS 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 is <15 minutes.
    - A flight is considered on-time if arrival delay is <30 minutes.
- The average flight duration (not including transfers) in minutes.
- The average passenger-weighted fare, by season, for an O-D pair for:
    - Economy class.
    - First/business class.
- The number of passengers, by season, for trips between the airports with:
    - No stop.
    - One stop (summarized by stop locations).
    - Two or more stops (summarized by stop locations).
- The average coupon-mileage for trips with:
    - No stop.
    - One stop (summarized by stop locations).
    - Two or more stops (summarized by stop locations).

[6] Please visit the BTS Homepage: https://www.bts.gov/.

[7] Please visit the BTS Airline Origin and Destination (DB1B) Database website: https://www.transtats.bts.gov/Tables.asp?DB_ID=125.

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 LOS data file to use in the models. The fields in this resulting file are shown in Table 2, 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 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 2. Airport-to-airport LOS variables

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)

Source: BTS

Notes on Table 2:

- *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 that are either in the on-time database, 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

The initial model application uses zones (NUMAs) as the basic level of spatial aggregation for rail and air matrices. 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 LOS 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.

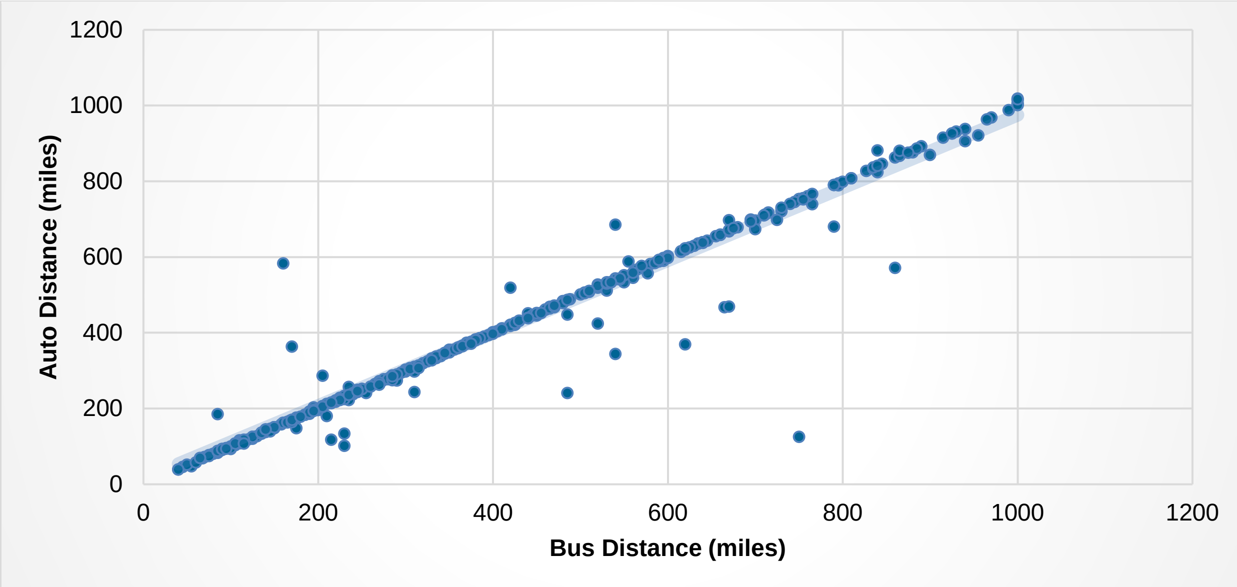
The same procedure was used to select rail routes, using all combinations of rail stations within 50 miles of the O-D census tracts.

Bus System

Travel time, distance, and fare skims were generated in this study for the long-distance bus network of the United States. The bus network was identified as a subset of the road network by identifying all zone-to-zone pairs that provide intercity bus service. The LOS characteristics were developed from observed data for a sample of routes, because it was beyond the scope of this effort to identify every bus route across the nation. These observed data sources were evaluated and efficient procedures were employed to estimate the LOS measures for the remaining routes in the bus network. Procedures to generate bus LOS 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 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 22.



Source: FHWA

Figure 22. Comparison between auto and bus distances.

The data points in Figure 22 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. Table 3 presents this information.

Table 3. 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 3, 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:

- Travel distance (miles).
- Number of transfers.
- Number of stops.
- Travel time (minutes).
- One-way fare ($).
- Frequency.
- Transfer point.
- 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:

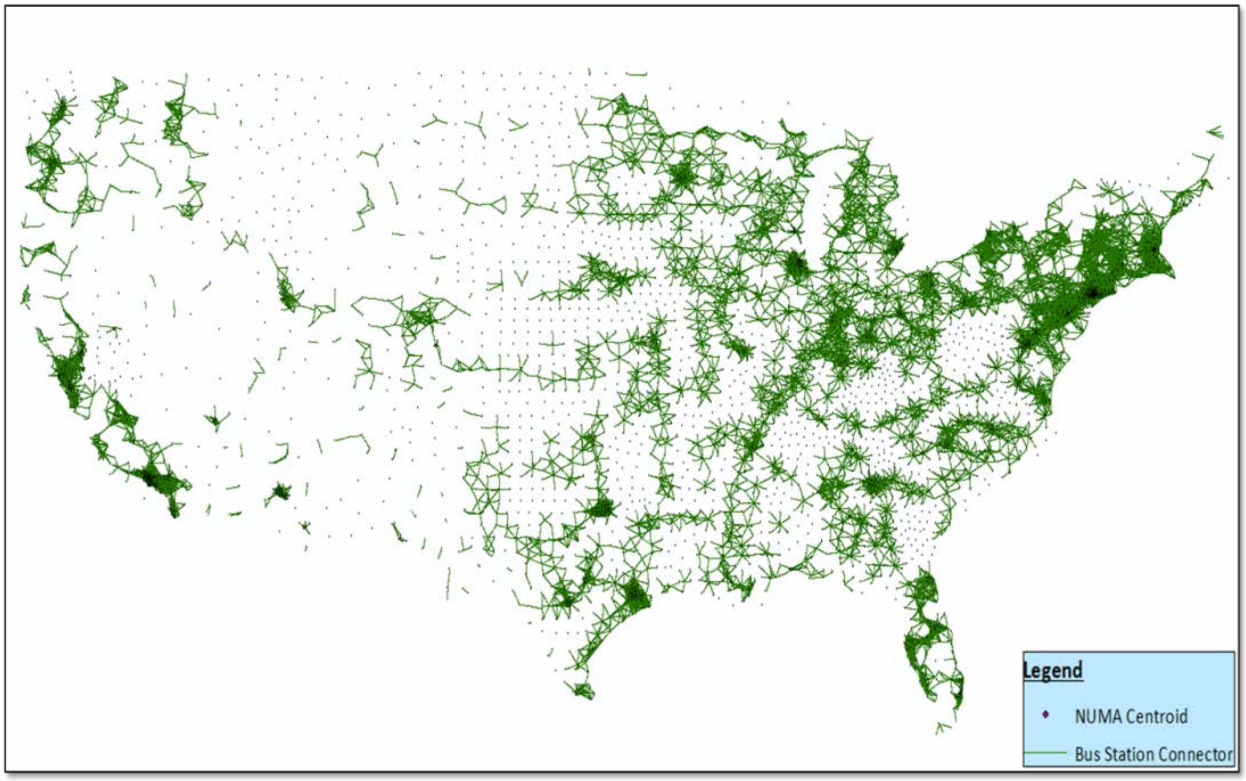
$$Fare = 9.65 + 0.107 * travel\ time$$

Figure 23. Equation. Bus fare regression model.

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, like 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 and 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 24. A NUMA can have a connector to all bus stations within 40 miles from the location of its centroid.



Source: FHWA

Figure 24. 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 3 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 from this exercise was used to compute a fare matrix (Figure 22).

2.3 Socioeconomic Data

Person and Household Characteristics

Person and household characteristics were derived from the Public Use Microdata Sample (PUMS) of the 2010 census and the 2007–2011 ACS 5-year estimates. These are used primarily as input to the synthetic population process, described in Chapter 3. Long-Distance Model Development.

The personal characteristics selected from the census data include the following:

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

The household characteristics selected from the census data include the following:

- 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.
- 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 is the person listed first. 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 (CPS) as noninstitutional living arrangements for groups not living in conventional housing units or groups living in housing units containing 10 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 CPS.

Employment Data

Employment data were 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

The 2010 LEHD Origin-Destination Employment Statistics (LODES) database was the primary source of employment data. Table 4 presents categories used in developing these data. The database contains private and public job numbers for all States but one 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 several subcategories (also shown in Table 4 through Table 6).

Table 4. 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

NAICS Employment Categories
(13) management of companies and enterprises
(14) administrative and support and waste management and remediation services
(15) educational services
(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 5. 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 6. 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), 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 data, which are 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 Quarterly Census of Employment and Wage

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. 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.

2.4 Land-Use Data

Several 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 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.
    - Number of permanent households and noninstitutionalized group quarters.
- **U.S. National Park Service (NPS) 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 NPS, TomTom, and Esri. Information available from these layers included park/forest name, type (e.g., national park, State park, regional park, national forest), 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 for 2012.
- **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 select number of State/regional models and it was not feasible to create a national-level dataset for this project.

2.5 Origin-Destination Data

2011 Traveler Analysis Framework

In 2013 and 2015, FHWA developed multimodal interregional passenger travel origin and destination data for the years 2008 and 2040.[8] These modal trip tables were developed from observed and forecast data to provide county-level geographic detail:

- Auto trips were developed from the 1995 American Travel Survey (ATS), the 2001 National Household Travel Survey (NHTS), the 2008 ACS.
- Bus trips were developed from the 2008 American Bus Association (ABA) Motorcoach Survey, the 2010-2011 ABA Member Origin-Destination Surveys, and the Russell's National Motor Coach Guide.
- Rail trips were developed from 2008 Amtrak station-to-station data combined with Federal Railroad Administration (FRA) data on California Thruway Bus Services and California High-Speed Rail Authority survey data on Amtrak riders.
- Air trips were developed from the 2008 Airline Origin and Destination Survey (DB1B) and T-100 data. These data describe air passenger trips between airports, and a collection of airport specific and regional airport ground access surveys that describe air passenger trips from trip origins (e.g., homes, offices, hotels) to airports and from airports to trip destinations. The combination of trip origin to airport, airport-to-airport, and airport to trip destination describes a complete air passenger trip from origin to destination.

The project team also aggregated these data to Traveler Analysis Framework (TAF) zones, which are groups of counties. Forecasts for 2040 were developed using socioeconomic data from Woods and Poole.

The TAF O-D estimates were based on the best available data at the time. The air and rail estimates are of higher reliability than the auto and bus estimates, based on the observed ridership data sources available for these modes. The 1995 ATS was 20+ years old when it was used to develop the 2008 auto and bus trip tables and there was no network-based volume data

[8] Please visit the Final Report to FHWA For Traffic Analysis Framework Part IIA -- Establishing Multimodal Interregional Passenger Travel Origin Destination Data: https://www.fhwa.dot.gov/policyinformation/analysisframework/docs/taf_final_report.pdf.

available for validation. The 2011 TAF was developed by interpolating trip tables between 2008 and 2040. Interpolation was based on a linear interpolation at the county-to-county level.

2014 Intercity Bus Ridership Table

The 2014 Intercity Bus Ridership Table was developed as part of FHWA's Developing Refined Estimates of Intercity Bus Ridership project.[9] The project included defining the top 200 markets (where a market is a pair of metropolitan areas defined using the U.S. Census Bureau's Core Based Statistical Areas [CBSAs]), identifying the characteristics of those markets, developing schedule data for those markets, and estimating bus ridership for those markets. The table utilized data from several sources, including GTFS data for intercity bus services compiled from several sources, intercity bus schedule data from Russell's Guide, and Northeast Corridor traveler survey. The 2014 bus ridership table was factored down to the 2011 level.

Intercity bus ridership estimates for the year 2008 were initially developed as part of FHWA's TAF Multimodal Interregional Passenger Travel Origin-Destination Data project. Those estimates were based on extrapolations from the 1995 ATS. However, the intercity bus market changed considerably in the interval between 1995 and 2008 (and has continued to change since 2008) and simple extrapolations apparently did not capture the full extent of those changes. A review of the estimates by ABA and its member companies indicated that these initial TAF estimates were likely too low.

2.6 Household Surveys

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

- 1995 ATS.
- 2001 NHTS.
- 2012 California Household Travel Survey.
- 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.

---

[9] Please visit the FHWA, Final Report: Developing Refined Estimates of Intercity Bus Ridership: https://www.fhwa.dot.gov/policyinformation/analysisframework/docs/rsg_bus_study.pdf.

National Travel Surveys

1995 ATS[10]

The "standard" dataset for modeling long-distance travel in the United States has long been the 1995 ATS. The U.S. BTS conducted the ATS periodically up until 1995, but has not performed it since, which is the main reason such a dated source of data is still in use. Table 9 outlines the attractive features of this dataset:

- 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 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 9, this survey was not entirely retrospective. Respondents were contacted before the yearlong 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.

The 1995 ATS collected information on the origin, destination, volume, and socioeconomic characteristics of long-distance travelers in the United States.[11] The survey consisted of four detailed interviews conducted approximately every three months from April 1995 to March 1996. These interviews were conducted primarily by telephone, with in-person interviews for some respondents unreachable by telephone.

The 1995 survey achieved an 85 percent response rate from those households that were eligible for interview. The survey gathered demographic characteristics of all household members regardless of age and information about their trips of 100 miles or more taken during 1995. Trip characteristics included such items as the origin and destination of the trip, stops along the way and side trips from the destination, the principal means of transportation, the access and egress modes to airports, train and bus stations, and information about the travel party. Some basic travel and tourism information was also collected including the reason for the trip, number of nights spent away from home, and the type of lodging. Route distances of all trips were calculated by Oak Ridge National Laboratory.

The 1995 ATS remains the only U.S. survey of long-distance travel for a 12-month period. It is a large dataset, with over one-half million long-distance trips (75 miles or more) reported by

---

[10] Please visit the BTS 1995 ATS Publications website: https://www.bts.gov/publications/1995_american_travel_survey/technical_documentation/entire.pdf.

[11] Please visit the BTS, Airline Origin and Destination (DB1B) Database website: https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=505.

almost 70,000 households randomly selected from across the United States. It contains one full year's worth of trips for each household. Respondents were contacted before the yearlong 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 valuable information about the trips they had reported. Also, CAPIs were performed with respondents who could not participate by telephone, reducing one potential source of nonresponse bias.

Besides these data being 20+ years old, several factors might limit their usefulness for some types of modeling. First, no geocode information exists for the trips, which precludes attachment of detailed mode-impedance information. Second, there is a high respondent burden associated with a 12-month survey with repeated interviews. Even though the reported response rate is high (85 percent), 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."

The 1995 ATS included 18 modes and the relevant modes were aggregated to the four modes of interest in this study, as shown in Table 7. Several of the ATS modes were not used, since they do not align with the four primary modes of the national long-distance passenger model.

Table 7. 1995 ATS modes.

2010 rJourney Mode	1995 ATS Mode
Auto	01=Car, Pickup Truck, or Van 02=Other Truck 03=Rental Car, Truck, or Van
Air	04=Commercial Airplane 05=Corporate/Personal Airplane
Bus	06=City to City Bus 07=Charter Bus or Tour Bus 08=School Bus
Rail	09=Train
Not Used	10=Taxi 11=Ship or Boat 12=Cruise Ship 13=Passenger Line or Ferry 14=Recreational Boat, Sailboat, Pleasure Boat or Yacht 15=Recreational Vehicle or Motor Home 16=Bicycle 17=Motorcycle, Moped or Motor Bicycle 18=Other

2001 NHTS[12]

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 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 percent 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).

Statewide Travel Surveys

2012 California Household Travel Survey

The California Department of Transportation (Caltrans) performed a major survey effort for the entire State in 2012. The design of the survey is like that used in the Colorado survey, but with the retrospective period extended from 2 weeks to 8 weeks. Extension of the retrospective period sought to provide more trips for modeling. 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) is likely to be good enough to use for modeling those other aspects of behavior.

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. Data on LOS were appended to this survey from modal network skim files provided by Caltrans in June 2013. These data include travel time, distance and cost for the California Statewide Travel Demand Model.[13]

---

[12] Please visit the 2001 NHTS User's Guide: https://nhts.ornl.gov/2001/usersguide/UsersGuide.pdf.

[13] Please visit the California Household Travel Survey page: http://www.dot.ca.gov/hq/tpp/offices/omsp/statewide_travel_analysis/chts.html

2010 Colorado Front Range Travel Survey[14]

Few regional planning agencies (MPOs) 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 taking the lead. This survey is recent and offers detailed geocoding information that is likely to be available for all trips (for researchers who sign a confidentiality agreement). Also, the retrospective period of two weeks is 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

Like Colorado, the Ohio Department of Transportation conducted a long-distance passenger travel survey as part of a larger household travel survey effort. A total of 2,094 households 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 two-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 the aforementioned United States datasets (summarized in Table 8) are similar in terms of the data items that they contain. These surveys often limited the amount of detail collected from each long-distance trip because they were (mainly) retrospective surveys and were time-constrained "add-ons" to standard household travel surveys. The common data items include the following:

- **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, 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.

[14] Please visit The 2010 Colorado Front Range Travel Survey page: https://nfrmpo.org/wp-content/uploads/2010-nfrmpo-household-survey.pdf.

- **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).
- **Scheduling at the destination**. Along with trip purpose, this information helps to define specific types of journeys for segmentation.

Table 8. Summary of long-distance travel survey characteristics.

National	ATS	NHTS (NY)	NHTS (WI)	California	Colorado	Ohio
Year	1995	2001	2001	2012–2013	2010	2001–2003
One-way trip length	75+ miles[15]	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[16]
Number of HHs reported long-distance trips	48,527 household (HH)	7,032 HH	11,027 HH	≅15,500 HH	3,000 HH	2,094 HH
Number of long-distance 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 long-distance trips?	Self-reported	Self-reported	Self-reported	Self-reported	Self-reported	Self-reported
Trip origin used to define long-distance trips	Home	Home	Home	Home	Home	Home
Trip destination used to define long-distance trips	Farthest destination	Farthest destination	Farthest destination	Any destination	Any destination	Any destination

- **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 the following:

---

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

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

- **Information on the trip planning process**. This may include how information was gathered, how reservations were made, and how far in advance planning was done. 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 conducted on the trip.
- **How often the destination had been visited in the past**. Differences can exist 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 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.

Data Preparation

1995 ATS

The 1995 ATS collected long-distance travel information from more than 70,000 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, tour 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.

Statewide Travel Surveys

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 one of the following:
    - Commute.
    - Business.
    - Visiting friends and relatives.
    - Leisure.
    - Personal business.
- Identifying the tour mode and, where necessary, grouping it as one of the following:
    - Auto.
    - Bus.
    - Rail.
    - 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 tour 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.

2.7 Traffic Counts

2007 HPMS

The HPMS is a national-level highway information system that includes traffic counts on the nation's highways. The HPMS contains traffic count data as a mix of universe and sample data for arterial and collector functional systems. Traffic counts are represented here as average annual daily traffic (AADT). For national traffic count data, the Freight Analysis Framework Version 3 (FAF^3) database was applied to the NHPN, which resulted in adding HPMS AADT for 2007 to 98 percent of functionally classified links within the NHPN.

2013 Office of Highway Policy Information (OHPI)

Rural vehicle miles were selected as a useful comparative statistic, given that most long-distance travel is on rural facilities. Annual vehicle miles traveled (VMT) data is available for urban and rural facilities published by FHWA's OHPI. Rural VMT was obtained from Table VM-3 from the Highway Statistics 2013 manual[17] and these data were aggregated from the State level into census divisions.

2.8 Data Used for Model Estimation

For model estimation purposes, each dataset was examined in detail and several descriptive statistics were generated.[18] It was clear from these analyses that the data from the sources listed previously varied in terms of the following:

- 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).
- Data collection schedule (e.g., all through the year or only a few months in one year).
- Spatial resolution of tour origin/destination.
- 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 9 lists the datasets used to estimate different long-distance model components. All model components were estimated using a combined dataset from the California, New York, Ohio, and Wisconsin surveys except that only the California data were used for estimation of the logit tour generation and tour-scheduling models and the mode generalized cost.

---

[17] Please visit the Policy and Governmental Affairs, OHPI, Highway Statistics 2013 website: https://www.fhwa.dot.gov/policyinformation/statistics/2013/vm3.cfm.

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

Table 9. Datasets used to estimate long-distance travel model components.

Household Travel Survey	Auto Ownership	Tour Generation	Scheduling	Tour party size	Mode and Destination Choice	Mode Generalized Cost
2001 NHTS (NY)	Yes	No	Yes	Yes	Yes	No
2001 NHTS (WI)	Yes	No	Yes	Yes	Yes	No
2012 California	Yes	Yes	Yes	Yes	Yes	Yes
2003 Ohio	Yes	No	Yes	Yes	Yes	No

2.9 Data Used for Model Calibration and Validation

The following data sources were used to obtain observed target values, rates, and distributions:

1. 2007–2011 ACS 5-year estimate.
2. 2001 New York NHTS add-on.
3. 2001 Wisconsin NHTS add-on.
4. 2003 Ohio Household Travel Survey.
5. 2010 Colorado Front Range Travel Survey.
6. 2012 California Household Travel Survey.

Target values and distribution from the ACS data were used for the household vehicle ownership model. For other models, target distributions and rates obtained from expanded household travel survey data were used. Expansion factors were not available for 2012 California Household Travel Survey, so this survey was not used for any expanded targets. Using these five statewide household travel surveys provided a range of target distributions and rates across the United States, but it does not represent a true national household travel survey for long-distance passenger travel. As a result, calibration of these models was not intended to achieve a tight comparison between the model results and the five-State observed dataset.

Table 10. Datasets used to calibrate long-distance travel model components.

Household Travel Survey	Auto Ownership	Tour Generation	Scheduling	Tour Party Size	Destination and Mode Choice
2007-2011 ACS	Yes	No	No	No	No
2001 NHTS (NY)	No	Yes	No	Yes	Yes
2001 NHTS (WI)	No	Yes	No	Yes	Yes
2012 California	No	Yes	No	Yes	Yes
2003 Ohio	No	Yes	No	Yes	Yes
2010 Colorado Front Range	No	Yes	No	Yes	Yes

Table 11 presents a summary of the datasets used to validate long-distance travel mode choice and highway assignment components. Model validation is typically performed only for assignment results, but the multimodal O-D data in the 2011 TAF, the 2011 Intercity Bus Ridership, and the 2001 NHTS provided an opportunity to compare modal volumes to these independent data sources. These data sources are considered independent because they were not used in the estimation or calibration efforts. The TAF and the Intercity Bus Ridership data are not solely observed data sources, although much of the underlying data in these data came from observed sources. Two datasets help validate the highway assignment component: the 2007 HPMS rural AADT and the 2013 OHPI VMT.

Table 11. Datasets used to validate long-distance travel model components.

Household Travel Survey	Mode Choice	Highway Assignment
2011 TAF	Yes	No
2011 Intercity Bus Ridership	Yes	No
2001 NHTS	Yes	No
2007 HPMS Rural AADT	No	Yes
2013 OHPI VMT	No	Yes

CHAPTER 3. LONG-DISTANCE MODEL DEVELOPMENT

This chapter details the development of models and datasets as described in Figure 1. The first section of this chapter describes the national synthetic population process; the second section discusses the model structure used for the long-distance travel demand passenger model. The third section documents the specific model components used in the long-distance passenger model.

3.1 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.[1] The PopGen system is a robust synthetic population generation software that can control 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 several 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 requiring matching in the synthesis process. The sample file is expanded such 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

---

[1] 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. Please visit [A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.537.723&rep=rep1&type=pdf](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.537.723&rep=rep1&type=pdf).

the sample file (often derived from the PUMS data of the U.S. 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. The first step determines the total number of households or persons required for each cell of the joint distribution matrix.

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 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[2] simulation process using the weights computed in the second step. This completes the synthetic population generation procedure. A few additional steps 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) drawn into the synthetic population are whole numbers (the weights at the end of the second step are likely to be fractional weights, which requires application of appropriate rounding methods 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

[2] Monte Carlo simulation produces distributions of possible outcome values. By using probability distributions, variables can have different probabilities of different outcomes occurring.

the entire population (at the agent level), the synthetic population can be expanded so that a unique record exists 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, nine commonwealths/territories, and six 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 12.

Table 12. 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
Hawai'i	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

State	Counties	Tracts	Block groups
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**

The project required selecting a geographic resolution for synthesis of a national synthetic population. 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.

To perform the synthetic population generation, the research team chose to use the 2007–2011 5-year ACS datasets for population synthesis. Thus, the marginal control data for several 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 census version of the MABLE GeoCorr geographic-correspondence files, developed by the Missouri Census Data Center,[3] are datasets that were used to map the census tracts to corresponding PUMAs.

[3] Please visit the Missouri Census Data Center website: http://mcdc.missouri.edu/websas/geocorr14.html.

Control Variables

PopGen can use any combination of controls for synthesizing a population for the nation. While using many control variables may sound appealing from a synthetic population representativeness standpoint, using myriad control variables comes with its own drawbacks. The presence of large numbers of control variables may generate thousands—or even millions—of constraints. Having so many 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 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 several small-scale trials to identify an appropriate set of controls and categories for a national-level synthetic population generation effort at the census tract resolution (should that occur).

Table 13 presents the control variables and categories used in the synthetic population generation process. At the household level, 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). A total of 4,480 constraints (cells in the joint distribution) exist 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, 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 several 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 (postsynthesis) 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 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 can match any other variables in the PUMS files 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

Level	Variable Description	Category Value	Category Description
		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 assessed the performance of the synthesis process for each State before proceeding to a subsequent State. At the end of the synthetic population generation process, PopGen produced 51 folders, with each folder containing the following:

- 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.
- Marginal tract level records for person attributes.

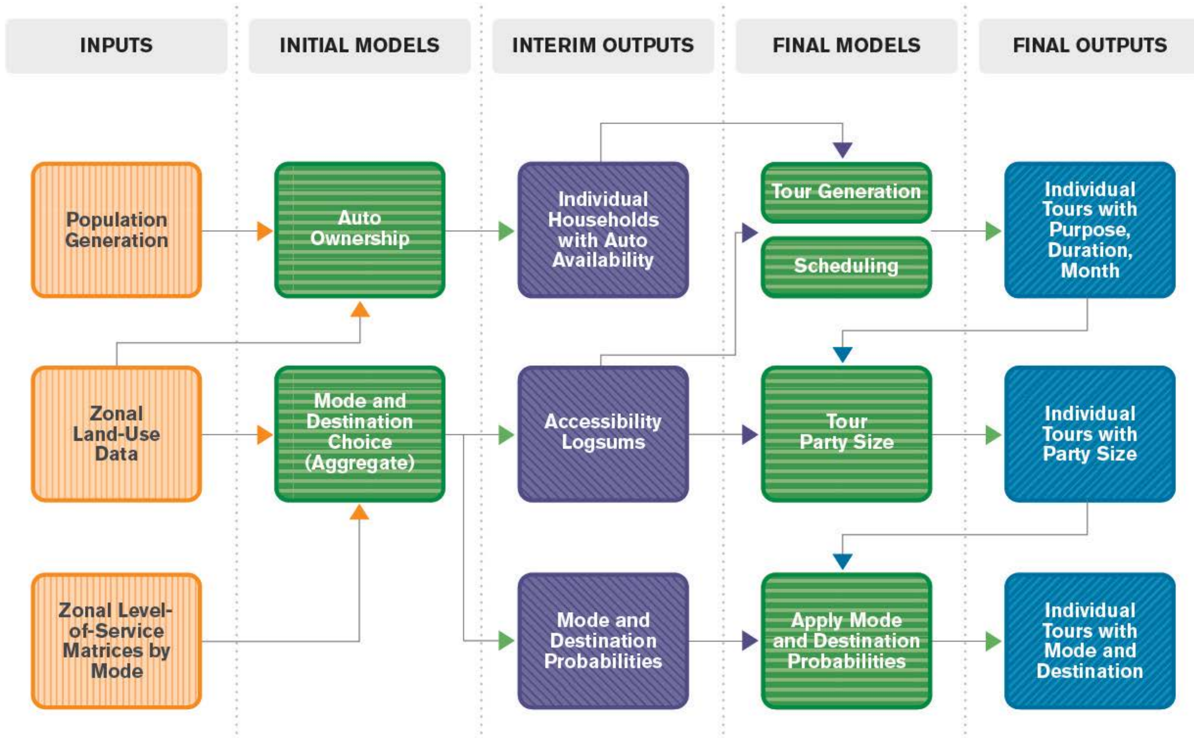
In PopGen, the number of households synthesized is always equal to the total number of households in the marginal control file. As long-distance travel choices may often involve household level negotiations and decision processes, it was considered important to exactly match the number of households 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 inconsistencies between household level marginal control distributions and person-level marginal control distributions. At the end of the synthetic population generation process, the project team integrated the synthetic population files in the 51 folders to form the national synthetic population files.

3.2 Structure of the Travel Modeling

The project team considered several different application structures for implementation of the national long-distance passenger model, ranging from a more aggregate structure to a fully disaggregate microsimulation model.

Figure 25 illustrates how the model is structured. This structure is a fully disaggregate structure, except that the last step to predict tour modes and destinations is aggregate. The final model is applied separately for each of the five trip purposes:

- Commute
- Business
- Visit Friends & Relatives
- Leisure
- Personal Business



Source: FHWA

Figure 25. rJourney model process.

The model stores the probabilities calculated from the destination and mode choice models while calculating aggregate logsums and uses those probabilities in the final step to perform the Monte Carlo microsimulation. This predicts a specific destination and mode combination for each tour. This structure results in a model system that runs quickly while still providing all the advantages of a fully disaggregate model system.

The disaggregate microsimulation structure has several advantages over the aggregate structure:

- The population does not require aggregation into market segments; each household in the synthetic population is simulated individually.
- The models include all household characteristics in the synthetic sample as explanatory variables, which accommodates more model specifications.
- 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. Users can aggregate these tour records up to O-D matrices along any desired dimensions, which provides more flexibility than in an aggregate model system that requires pre-specification of the number and definition of the output matrices. 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.

In this structure, it is still expedient to use a more aggregate version of the destination and mode 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 and destination choice models for every possible tour purpose for every household would be prohibitive in terms of run time.[4]

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 application of more choice model types (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 model application population size, the more that run time becomes an issue, while random simulation error may become less of an issue (because random simulation error is generally proportional to the square root of the sample size).

3.3 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 1 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.

[4] This method of using pre-calculated accessibility logsums is also used in most urban, activity-based [AB] microsimulation model systems.

Mathematical Formulation

The probability of a household choosing the number of vehicles available is described by the multinomial discrete choice logit model equation (Figure 26).

$$\Pr(i) = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)}$$

Figure 26. Equation. Multinomial discrete choice logit model.

Where:

- Pr(i) is the probability of the decision-maker choosing an alternative *i*.
- V_j is the systematic component of the utility of alternative *j*.

Four alternatives (i) exist for the auto ownership model: zero cars, one car, two cars and three or more cars per household. The utility component (V_i) is presented in Figure 27.

$$V_i = \in_i + a * HH_{1adult} + b * HH_{3adult} + c * HH_{4+adult} + d * \frac{HHworkers}{HHadults}$$

$$+ e * HH_{children} + f * HHhead_{65+} + g * HHhead_{<35} + h * Log(Density)$$

$$+ i * Log(\frac{HHincome}{1000})$$

Figure 27. Equation. Auto ownership utility.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a, b, c, d, e, f, g, h, i* are estimated coefficients (see Table 14)
- HH_{1adult} is the households with one adult.
- HH_{3adult} is the households with three adults.
- $HH_{4+adult}$ is the households with four or more adults.
- *HHworkers/HHadults* is the fraction of household workers compared to household adults.

- $HH_{children}$ is the households with children.
- $HHhead_{65+}$ is the households with the head of the household over 65 years of age.
- $HHhead_{<35}$ is the households with the head of the household less than 35 years of age.
- *Density* is the households plus jobs per square mile.
- *HHincome* is the total gross household income in dollars.

We assume that the errors $\in_i$ are independent and identically distributed such that $E[\in_i] = 0$ and $var[\in_i] = \sigma^2$. Typically, we assume $\in_i \sim N(0, \sigma^2)$ as a basis for inference (e.g., t-tests on parameters).

Estimation Results

The model estimation results are shown in Table 14. The base alternative in the model is two 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 that 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. Additional results include:

- Household income, used in logarithmic form, is an important variable, particularly for higher income households much less likely to own 0 or 1 cars.
- 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. An urban-regional model would use a more detailed variable for accessibility, 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 resist accurate measurement 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.

Table 14. Household car ownership model.

Alternative/Statistics	0 cars Coeff.	0 cars T-stat.	1 car Coeff.	1 car T-stat.	2 cars (base)	3 cars Coeff.	3 cars T-stat.	4+ cars Coeff.	4+ cars T-stat.
Constant	-2.98	-20.9	0.726	10.9	--	-1.4	-21.6	-2.8	-31.2
1 adult in HH	2.45	44	2.42	111.6	--	--	--	--	--
2 adults in HH	--	--	--	--	--	--	--	--	--
3 adults in HH	0.813	9.3	-0.189	-4.5	--	1.61	77.6	1.89	71.7
4+ adults in HH	1.14	8.1	--	--	--	1.95	47.9	3.68	93.5
Log(income)	-1.52	-50	-0.87	-55.4	--	0.114	7.7	0.276	13.8
Missing income data	-6.22	-41.8	-3.65	-50.1	--	0.487	6.7	1.24	12.6
Workers/adults ratio	-0.442	-7	-0.224	-7.8	--	0.464	16.9	1.09	27.5
HH has children	-0.877	-15.3	-1	-44	--	--	--	--	--
HH head age 65+	--	--	0.184	6.6	--	-0.218	-7.6	-0.229	-5.5
HH head age <35	0.269	4.4	0.112	4.1	--	-0.278	-12.3	-0.144	-4.9
Log (HH+Job density)	0.767	57.6	0.243	43.9	--	-0.103	-21.3	-0.234	-35.9
Observations	114103	--	--	--	--	--	--	--	--
Rho-squared (0 coeff.)	0.353	--	--	--	--	--	--	--	--
Rho-squared (constants)	0.197	--	--	--	--	--	--	--	--

As in most models presented in this section, a separate “nuisance” variable was estimated for those households with missing income data to facilitate inclusion 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.

Accessibility Logsums

The accessibility logsum captures LOS effects at the upper level of a nested model in a way that takes into consideration all lower level alternatives and avoids counter-intuitive effects.[23] The use of logsum measures have been highly regarded as perhaps the best available means of capturing composite effects that cannot be measured directly in a model.

The use of accessibility logsums comes from making the upper level models appropriately sensitive to variables that affect the upper level outcome but cannot be measured directly because they differ among the undetermined lower level model outcomes. In formal nested logit hierarchies, the upward integrity comes from the logsum, the composite measure of expected utility across the lower level alternatives. Unfortunately, the strength of the logsum variable as a composite measure rests in a feature that makes it computationally expensive, and essentially infeasible with very large and detailed hierarchical model systems: it requires the calculation of utility for every single alternative in the hierarchy below the level being modeled. To model the highest-level outcome, utilities of all alternatives in the entire hierarchy must be computed.

An approximate, or aggregate, logsum is calculated in the same basic way as a true logsum, by calculating the utility of multiple alternatives, and then taking expectation across the alternatives by calculating the log of the sum of the exponentiated utilities. In this context, the amount of computation is reduced, by calculating utility for a carefully chosen aggregation of the available alternatives. The approximate logsum is pre-calculated and used by the tour generation, scheduling, and tour party size models.

The categories of decisionmakers and the aggregation of alternatives are chosen so that in all choice cases an approximate logsum is available that closely approximates the true logsum. In essence, this is a sophisticated ad hoc measure that is intended to achieve most of the realism of the true logsum at a small fraction of the cost. The approximate tour mode-destination choice logsum is used in situations where information is needed about accessibility to activity opportunities in all surrounding locations by all available transport modes. Because of the large amount of computation required for calculating a true logsum for all feasible combinations in these three dimensions, an approximate logsum is used.

Mathematical Formulation

The mathematical formulation for the logsum is based on random utility theory. This utility is a function of distance and the opportunities in zone *j*.

$$LogSum = \log \sum_{j} x_j \exp(-\beta t_{ij})$$

Figure 28. Equation. Accessibility logsum.

[23] Ben-Akiva, M. and Lerman, S.R. (1985) Discrete Choice Analysis Theory and Application to Travel Demand. MIT Press, Cambridge.

Where:

- x_j = the supply of x across all zones *j*.
- t_{ij} = the travel time between *i* and *j*.
- β = estimated coefficient.

Tour Generation

The tour generation and scheduling logit models use a single day as the decision period. Although the various surveys have different lengths of retrospective recall for the long-distance surveys (e.g., eight weeks for the California statewide survey), breaking the data down into individual days has the advantage that a few household-days (approximately 0.04 percent) have more than one long-distance tour generated on that given day, meaning that the first step of tour generation can be modeled as a binary choice—no tour, or one tour for a given day. The second step is modeled as a binary choice between making a second tour or not.

Mathematical Formulation

The probability of a household choosing the number of tours per day is described by the multinomial discrete choice logit model (see Figure 26). Two models exist for each purpose within tour generation: the first model estimates whether the household will make one tour as a binary choice and the second model estimates whether the household will make a second tour as a binary choice for those households that made a first tour. Five purposes exist: commute, business, visit friends and relatives, leisure, and personal business. This produces 10 individual tour generation models. The alternatives (i) for the tour generation model: no tours or one tour per household per day. The utility component (V_i) is presented in Figure 29.

$$V_i = \epsilon_i + a_{1-12} * Month + b * AccessLogsum_{<50miles} + c * AccessLogsum_{50-150miles}$$

$$+ d * AccessLogsum_{150-500miles} + e * AccessLogsum_{>500miles}$$

$$+ f * NoAccessLogsum_{<50miles} + g * Log\left(\frac{HHincome}{1000}\right) + h * HH_{0cars}$$

$$+ i * HH_{carcompetition} + j * HH_{children} + k * \frac{HHworkers}{HHadults} + l * HH_{1person}$$

$$+ m * HHhead_{<35} + n * HHhead_{65+} + o * HHsize$$

Figure 29. Equation. Tour generation utility.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- $a_{1\text{-}12}$, *b, c, d, e, f, g, h, I, j, k, l, m, n, o* are estimated coefficients (see Table 16).
- *Month* is the month that the tour takes place, with each month as a separate dummy variable with a separate coefficient ($a_{1\text{-}12}$).
- $AccessLogsum_{<50miles}$ is the accessibility logsum from the destination and mode choice models for tours under 50 miles one-way.
- $AccessLogsum_{50\text{-}150miles}$ is the accessibility logsum from the destination and mode choice models for tours under 50 miles one-way.
- $AccessLogsum_{150\text{-}500miles}$ is the accessibility logsum from the destination and mode choice models for tours under 50 miles one-way.
- $AccessLogsum_{>500miles}$ is the accessibility logsum from the destination and mode choice models for tours under 500 miles one-way.
- $NoAccessLogsum_{<50miles}$ is the accessibility logsum from the destination and mode choice models for tours under 50 miles one-way.
- *HHincome* is the total gross household income in dollars.
- HH_{0cars} is the households with zero cars.
- $HH_{carcompetition}$ is the households with fewer cars than adults.
- $HH_{children}$ is the households with children.
- *HHworkers/HHadults* is the fraction of household workers compared to household adults.
- $HH_{1person}$ is the households with one person.
- $HHhead_{<35}$ is the households with the head of the household less than 35 years of age.
- $HHhead_{65+}$ is the households with the head of the household over 65 years of age.
- *HHsize* is the number of persons in a household.

We assume that the errors $\in_i$ are independent and identically distributed such that $E[\in_i] = 0$ and $var[\in_i] = \sigma^2$. Typically, we assume $\in_i \sim N(0, \sigma^2)$ as a basis for inference (e.g., t-tests on parameters).

Estimation Results

First Tour Generation

Table 15 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. 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.)

The fifth row of Table 15 shows that on any given survey day, only 0.07 percent of households made a long-distance commute tour, 0.44 percent a long-distance business tour, etc. Across the five purposes, these fractions sum to 2.25 percent, meaning that in 97.75 percent 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. Some larger rural zones (typically counties) have no other zones accessible within 0–50 miles. The dummy variable for these zones is negative, compensating for those zones that 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).

Table 15. 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 alternative is no tour in the day for the purpose										
Constant	-4.28	constr.	-7.21	constr.	-5.73	constr.	-6.54	constr.	-4.18	constr.
Log (income)	0.273	5.6	0.521	14	0.128	6.7	0.266	9	0.0915	4.7
Missing income data	1.46	6	2.5	14.4	0.509	5.6	1.12	8.1	0.339	3.5
HH has 0 car	-1.6	-2.8	-0.24	-2.1	-0.323	-2.9	--	--	-0.757	-7.4
HH has fewer cars than adults	0.0812	1	-0.106	-2.9	-0.108	-3.4	-0.242	-7.7	-0.114	-3.2
HH has children	0.195	3	-0.112	-3.8	-0.245	-6.8	0.0613	1.9	-0.0859	-2.6
HH workers/adults ratio	0.175	1.4	0.584	10.1	-0.111	-3.9	-0.134	-4.3	-0.456	-11.4
One-person HH	--	--	-0.134	-3.1	-0.0991	-2.9	-0.301	-8.5	-0.339	-8.1
HH head under age 35	-0.426	-3.6	-0.251	-5.1	0.0994	2.7	--	--	-0.478	-8
HH head age 65 or older	-0.365	-3.8	-0.21	-5.5	--	--	-0.0698	-2.6	-0.111	-3.3
Household size	--	--	--	--	-0.0425	-2.9	-0.0281	-2.1	--	--
Mode/dest. logsum 0-50 miles	-0.157	-12.8	-0.0909	-7.9	-0.0522	-6.7	-0.0682	-6.5	-0.218	-31.2
Mode/dest. no zones 0-50 miles	0.449	10.6	0.0468	2.5	0.0467	2.2	0.01	0.5	0.0329	1.9
Mode/dest. logsum 50-150 miles	--	--	--	--	0.08	2.1	0.167	4.8	--	--
Mode/dest. logsum over 150 miles	--	--	0.134	3.4	0.28	7.4	0.402	7.7	--	--
No logsum computed 0-50 miles	1.12	4.9	--	--	-0.267	-4	-0.159	-2.4	0.342	6
January	0.599	5.3	-0.125	-2.2	-0.456	-9.3	-0.494	-9.6	-0.256	-4.4
February	0.598	5.5	0.0945	2	-0.273	-6.5	-0.316	-7.1	--	--
March	0.629	6	0.242	5.5	-0.18	-4.5	-0.107	-2.7	--	--
April	--	--	--	--	--	--	--	--	--	--
May (base)	0.523	5	--	--	--	--	--	--	--	--
June	--	--	--	--	0.0556	1.6	0.161	4.6	--	--
July	--	--	-0.086	-1.8	0.0306	0.8	0.365	10.7	-0.134	-2.6

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
August	0.386	3.5	--	--	-0.0823	-2.2	0.21	6	-0.152	-3
September	--	--	--	--	-0.241	-6.1	--	--	-0.176	-3.6
October	--	--	0.13	3	-0.288	-6.9	-0.152	-3.8	-0.265	-5.1
November	0.288	2.3	-0.107	-2.1	--	--	-0.308	-6.8	-0.15	-2.8
December	0.277	2.1	-0.403	-6.4	--	--	-0.509	-9.8	-0.317	-5.2
No. of days before survey	--	--	-0.0076	-4.3	-0.013	-9.2	-0.0096	-6.9	-0.0141	-7.5
Log (no. days before survey)	-0.412	-15	-0.176	-6.2	-0.097	-4.2	-0.13	-5.7	-0.131	-4.4
Statistics										
Observations	1,478,748		--	--	--	--	--	--	--	--
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.)	0.9	--	--	--	--	--	--	--	--	--
Rho-squared (c constants)	0	--	--	--	--	--	--	--	--	--

Second Tour Generation

For the 33,000 or so household-days for which at least one long-distance tour was reported, there are about 2.3 percent where a second tour was also reported. As a result, a second model was estimated (Table 16) 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 tour-scheduling model in Table 15, 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 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.

Table 16. Household-day tour generation model—second tour in the day.

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 alternative is no second tour in the day for the purpose										
Constant	-7.17	-5.2	-8.21	-8.7	-6.29	-9.2	-6.94	-9	-4.59	-20.7
Log (income)	0.16	0.6	0.288	1.5	0.104	1	0.106	0.5	--	--
Missing income data	-0.0557	0	0.74	0.8	-0.0084	0	0.328	0.3	--	--
HH has fewer cars than adults	--	--	--	--	-0.427	-2	--	--	--	--
HH has children	1.03	2.7	--	--	-0.246	-1.5	0.28	1.8	0.274	1.8
HH workers/adults ratio	--	--	0.758	1.9	--	--	--	--	--	--
One-person HH	--	--	-0.351	-1	-0.61	-2.5	--	--	--	--
HH head under age 35	--	--	--	--	--	--	--	--	-0.64	-1.8
HH head age 65 or older	--	--	--	--	--	--	-0.422	-2	--	--
Mode/dest. logsum 0-50 miles	--	--	-0.0589	-1	--	--	-0.109	-1.5	-0.136	-3.8
Mode/dest. no zones 0-50 miles	--	--	0.221	1.5	0.207	2	0.0029	0	0.0461	0.6
Mode/dest. logsum 50-150 miles	--	--	0.355	1.3	--	--	0.108	0.4	--	--
Mode/dest. logsum over 150 miles	--	--	--	--	--	--	0.58	1.5	--	--
Log (no. days before survey)	-0.319	-2	-0.131	-1.5	--	--	--	--	-0.073	-1.1
Statistics										
Observations	33,307			--	--	--	--	--	--	--
No. of tours (% of HH-days)	28	0.08%	127	0.38%	212	0.64%	199	0.60%	214	0.64%
Rho-squared (0 coeff.)	0.918	--	--	--	--	--	--	--	--	--
Rho-squared (c constants)	0.011	--	--	--	--	--	--	--	--	--

Scheduling

The scheduling 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.

Modeling this aspect of the tour is important because it may influence the travel distance or mode use (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).

Mathematical Formulation

The probability of a household choosing the number of nights away from home is described by the multinomial discrete choice logit model (see Figure 26). One model exists for each purpose within the scheduling model: commute, business, visit friends and relatives, leisure, and personal business. This produces five individual scheduling models. Four alternatives (i) exist for the scheduling model: zero nights away, one to two nights away, three to six nights away, and seven or more nights away. The utility component (V_i) is presented in Figure 30.

$$V_i = \in_i + a * HHsize + b * Log\left(\frac{HHincome}{1000}\right) + c * HHhead_{65+} + d * HHhead_{<35}$$

$$+ e * Log(Density) + f * Month_{Jan-March} + g * Month_{June-Aug}$$

$$+ h * Month_{Nov-Dec}$$

Figure 30. Equation. Number of scheduling utility.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h* are estimated coefficients (see Table 17).
- *HHsize* is the number of persons in a household.
- *HHincome* is the total gross household income in dollars.
- $HHhead_{65+}$ is the households with the head of the household over 65 years of age.
- $HHhead_{<35}$ is the households with the head of the household less than 35 years of age.
- *Density* is the households plus jobs per square mile.

- $Month_{Jan\text{-}March}$ is when the tour is conducted during winter (January through March).
- $Month_{June\text{-}Aug}$ is when the tour is conducted during summer (June through August).
- $Month_{Nov\text{-}Dec}$ is when the tour is conducted during the holidays (November through December).

We assume that the errors ϵ_i are independent and identically distributed such that $E[\epsilon_i] = 0$ and $var[\epsilon_i] = \sigma^2$. Typically, we assume $\epsilon_i \sim N(0, \sigma^2)$ as a basis for inference (e.g., t-tests on parameters).

Estimation Results

The results of this model in Table 17 show that even for the tour purposes that tend to have the longest distances and durations (visiting friends and relatives and leisure tours), over 40 percent of tours are day tours, only 6 percent to 8 percent of tours stay away from home for seven nights or more. Some results shown in Table 17 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 older 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 to make the longer tours.
- The discretionary purposes tend to be of shorter duration in the winter months (January through March), except for 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 17. Tour scheduling 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	-1.26	-2.5	-2.34	-12.3	-0.127	-1.3	-2.12	-18.4	-2.19	-13.7
HH size	--	--	-0.0275	-1.5	-0.0592	-5.1	--	--	--	--
Log(income)	-0.152	-1.3	0.369	8.7	0.0445	1.9	0.275	10.7	0.155	4.4
Missing income	-3.19	-2.9	1.74	8.2	0.18	1.6	1.09	8.7	0.631	3.6
HH head age 65+	--	--	--	--	-0.358	-7.7	-0.354	-7.8	-0.288	-4.3
HH head age<35	--	--	--	--	0.361	9.3	0.141	3.8	0.419	6.7
Log(res+emp density)	--	--	--	--	--	--	0.0873	9.5	0.0963	7.5
June-Aug	--	--	--	--	--	--	--	--	-0.0917	-1.7
Jan-Mar	--	--	--	--	-0.102	-2.7	-0.105	-2.7	-0.348	-5.4
Nov-Dec	--	--	--	--			-0.224	-5.1	-0.263	-3.6
Alt = 3-6 nights										
Constant	-3.88	-5.4	-3.37	-14	-1.51	-10.6	-3.93	-27.5	-3.48	-14.7
HH size	--	--	-0.0794	-3.3	-0.142	-8.9	-0.0517	-3.9	--	--
Log(income)	0.391	2.4	0.437	8.2	0.099	3.3	0.433	13.6	0.167	3.2
Missing income	2.31	3	2.2	8.3	0.383	2.6	1.88	12.3	0.875	3.5
HH head age 65+	--	--	--	--	0.141	2.6	--	--	0.239	2.8
HH head age<35	--	--	--	--	0.15	2.8	--	--	0.217	2.2
Log(res+emp density)	--	--	0.058	3.7	0.0678	6.4	0.172	15.6	0.134	7.2
June-Aug	--	--	--	--	0.273	5.6	0.565	13.7	--	--
Jan-Mar	-0.4	-2.1	--	--	-0.152	-2.7	-0.151	-2.7	-0.458	-4.9
Nov-Dec	--	--	--	--	0.389	7.8	-0.178	-2.9	-0.357	-3.3

Alt = 7+ nights										
Constant	-3.5	-8.9	-4.85	-11.4	-2.25	-16.6	-5.13	-25.9	-4.7	-12.7
HH size	-0.165	-1.4	--	--	-0.218	-8.5	--	--	--	--
Log(income)	--	--	0.263	2.7	--	--	0.524	12	0.361	4.2
Missing income	--	--	1.59	3.4	--	--	2.29	11	1.65	4
HH head age 65+	--	--	--	--	0.32	4.3	0.334	5.2	0.313	2.4
HH head age<35	--	--	--	--	-0.275	-3.1	--	--	--	--
Log(res+emp density)	--	--	0.125	4.3	0.111	6.9	0.137	9.3	--	--
June-Aug	1.19	3.7	0.504	3.9	0.482	7	0.629	10.9	0.173	1.5
Jan-Mar	--	--	0.298	2.3	--	--	0.29	4.2	--	--
Nov-Dec	--	--	--	--	0.348	4.6	-0.373	-4	--	--
Statistics										
Total Observations	1967	--	9689	--	21829	--	25706	--	11932	--
Rho-squared(0 coeff)	0.491	--	0.229	--	0.15	--	0.157	--	0.352	--
Rho-square(constants)	0.018	--	0.01	--	0.014	--	0.021	--	0.014	--

Tour party size

The tour party size model predicts the number of members (including non-household participants) in the travel party. The base alternative is one person traveling alone, while the other alternatives are 2, 3, or 4+ persons.

Mathematical Formulation

The probability of a household choosing the number of people traveling together is described by the multinomial discrete choice logit model equation (see Figure 26). One model exists for each purpose within the tour party size model: commute, business, visit friends and relatives, leisure, and personal business. This produces five individual tour party size models. Four alternatives (i) exist for the tour party size model: one person traveling alone, two persons traveling together, three persons traveling together, and four or more persons traveling together. The utility component (V_i) is presented in Figure 31.

$$V_i = \in_i + a * \frac{HHworkers}{HHadults} + b * Log\left(\frac{HHincome}{1000}\right) + c * HH_{0cars} + d * HH_{carcompetition}$$

$$+ e * HHhead_{65+} + f * HHhead_{<35} + g * Log(Density)$$

$$+ h * Month_{Jan-March} + i * Month_{June-Aug} + j * Month_{Nov-Dec}$$

$$+ k * NightsAway_0 + l * NightsAway_{1-2} + m * NightsAway_{7+}$$

Figure 31. Equation. Tour party size utility.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a, b, c, d, e, f, g, h, i, j, k, l, m* are estimated coefficients (see Table 18)
- *HHworkers/HHadults* is the fraction of household workers compared to household adults.
- *HHincome* is the total gross household income in dollars.
- HH_{0cars} is the households with zero cars.
- $HH_{carcompetition}$ is the households with fewer cars than adults.
- $HHhead_{65+}$ is the households with the head of the household over 65 years of age.
- $HHhead_{<35}$ is the households with the head of the household less than 35 years of age.

- *Density* is the households plus jobs per square mile.
- $Month_{Jan\text{-}March}$ is when the tour is conducted during winter (January through March).
- $Month_{June\text{-}Aug}$ is when the tour is conducted during summer (June through August).
- $Month_{Nov\text{-}Dec}$ is when the tour is conducted during the holidays (November through December).
- $NightsAway_{0}$ is when the travel does not require any nights away.
- $NightsAway_{1\text{-}2}$ is when the travel requires one to two nights away.
- $NightsAway_{7+}$ is when the travel requires seven or more nights away.

We assume that the errors $\in_i$ are independent and identically distributed such that E $[\in_i] = 0$ and var $[\in_i] = \sigma^2$. Typically, we assume $\in_i \sim N(0, \sigma^2)$ as a basis for inference (e.g., t-tests on parameters).

Estimation Results

The model results, shown in Table 18, indicate the following:

- By far, the largest positive effect, applied to all alternatives, is when the tour 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 tour party size for business and commute, but it has no effect on the other purposes.
- The more workers in the household, the smaller the tour party size for all purposes except commute. This may be because one or more of the workers must stay home and work.
- In general, higher car ownership tends to increase tour 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 tour party sizes, but the opposite appears true for visit and personal business tours.
- Tours in the summer months tend to have larger tour party sizes for all discretionary purposes.

Table 18. 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
Base alternative is tour party size = 1 person										
All alternatives										
Tour party size = household size	0.264	5.2	0.544	23.9	1.75	101.5	1.28	90.7	1.01	49.4
Tour party size = household adults	-0.256	-5.7	--	--	-0.591	-28.4	-0.393	-22.2	-0.179	-7.8
Alternative—tour party size = 2 people										
Constant	0.0226	0.1	0.203	1.4	-0.089	-0.9	0.81	9.7	0.355	4.7
Workers/Household size	--	--	-0.501	-8	-0.266	-5.5	-0.113	-2.8	-0.187	-3.8
Log (Income)	-0.424	-7.6	-0.151	-5	0.0502	2.3	0.0818	4.2	--	--
Missing income data	-1.7	-6.4	-0.739	-4.9	0.0759	0.7	0.386	4.1	--	--
Head of HH age under 35	0.33	4.8	0.0619	1.3	--	--	--	--	-0.237	-4.5
Head of HH age 65 or over	-0.456	-2.9	0.411	7	0.115	2.5	--	--	0.107	2
HH has 0 vehicles	-0.616	-2.5	0.78	4.6	0.216	1.8	--	--	-0.467	-3.9
HH has few vehicles than adults	--	--	--	--	0.37	7.6	0.29	4.7	0.428	7.9
0 nights away	--	--	-0.419	-8.5	0.526	10	--	--	0.259	4
1 to 2 nights away	--	--	--	--	0.286	5.5	0.132	4.2	0.284	3.6
7 or more nights away	--	--	--	--	--	--	--	--	--	--
Missing duration data	--	--	-0.6	-13.4	-0.24	-4.6	-0.0538	-1.6	-0.264	-3.5
Summer (Jun-Aug)	--	--	--	--	--	--	--	--	0.118	2.2
Winter (Jan-Mar)	0.38	3.5	--	--	-0.121	-2.8	--	--	-0.122	-2.5
Holidays (Nov-Dec)	--	--	--	--	0.129	2.7	-0.121	-3.2	--	--
Missing month data	-0.054	-0.7	--	--	0.0125	0.2	-0.136	-2.2	-0.287	-4.5
Alternative—tour party size = 3 people										
Constant	-0.172	-0.3	-0.0635	-0.3	-0.715	-5.3	0.0434	1.5	-0.297	-4.4

Purpose	Commute		Business		Visit F&R		Leisure		Pers.Bus.	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Workers/Household size	--	--	--	--	-0.225	-3.5	--	--	-0.135	-2.1
Log (Income)	-0.548	-4.8	-0.439	-8.6	-0.074	-2.6	--	--	--	--
Missing income data	-3.53	-5	-2.14	-8.1	-0.401	-2.9	--	--	--	--
Head of HH age under 35	-0.745	-4	--	--	0.203	4.4	0.104	2.3	--	--
Head of HH age 65 or over	-0.987	-2.6	--	--	-0.109	-1.7	-0.232	-4.6	-0.193	-2.6
HH has 0 vehicles	1.6	6.5	0.948	3.8	0.612	4.5	0.526	4.7	--	--
HH has few vehicles than adults	0.674	3.9	--	--	0.501	8.6	0.67	9.9	0.477	7.4
0 nights away	--	--	-0.496	-5.9	0.689	10.5	--	--	--	--
1 to 2 nights away	--	--		--	0.447	6.8	--	--	--	--
7 or more nights away	--	--	--	--	--	--	--	--	-0.524	-3.1
Missing duration data	--	--	-0.636	-6.8	-0.17	-2.4	--	--	-0.675	-10.6
Summer (Jun-Aug)	-1.02	-3.1	--	--	0.136	2.7	0.177	4.7	0.303	4.4
Winter (Jan-Mar)	-0.981	-3.9	--	--	-0.141	-2.4	--	--	0.159	2.5
Holidays (Nov-Dec)	-1.43	-3.3	-0.221	-2	0.422	7	--	--	0.229	3.2
Missing month data	-1.13	-6.6	-0.461	-4.3	0.205	2.5	-0.172	-2.6	-0.181	-2.1
Alternative—tour party size = 4 or more people										
Constant	1.54	3.4	0.894	4.1	-0.845	-11.8	0.637	14.3	-0.0688	-0.5
Workers/Household size	0.898	3.1	-0.315	-2.9	-0.261	-4.5	-0.179	-3.9	--	--
Log (Income)	-1.5	-16.6	-0.594	-12.4	--	--	--	--	-0.0933	-3.3
Missing income data	-6.22	-12.8	-3.04	-11.7	--	--	--	--	-0.623	-4.2
Head of HH age under 35	0.575	4.2	--	--	0.36	9.2	0.287	8	0.325	5.9
Head of HH age 65 or over	-0.755	-2.1	--	--	-0.425	-6.2	-0.315	-6.8	-0.194	-2.7
HH has 0 vehicles	2.39	13	--	--	0.519	3.8	--	--	--	--
HH has few vehicles than adults	1.48	9.5	--	--	0.346	6.4	0.361	5.7	0.254	4
0 nights away	--	--	-0.716	-8.9	0.746	12.4	--	--	--	--

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	--	--	--	--	0.349	5.8	--	--	0.291	4.8
7 or more nights away	--	--	--	--	--	--	--	--	--	--
Missing duration data	--	--	-0.636	-7.3	-0.292	-4.5	--	--	-0.776	-11.7
Summer (Jun-Aug)	--	--	0.294	3.9	0.261	5.7	0.44	13.7	0.557	9.3
Winter (Jan-Mar)	-1.44	-3.9	--	--	-0.139	-2.6	0.0795	2.2	--	--
Holidays (Nov-Dec)	0.863	3.2	--	--	0.545	9.9	--	--	0.113	1.7
Missing month data	-0.491	-2.7	-0.941	-8.3	0.191	2.6	-0.597	-9.6	-0.0725	-0.9
Statistics										
Total Observations	9012	--	18626	--	31634	--	35998	--	18833	--
Rho-squared(0 coeff)	0.557	--	0.342	--	0.24	--	0.208	--	0.136	--
Rho-square(constants)	0.084	--	0.04	--	0.212	--	0.131	--	0.095	--

Mode and Destination Choice

The destination and mode choice models were initially estimated on the data from the California statewide model (zones and networks) and household travel data from the 2013 California statewide survey. These networks contained more spatial detail (roughly 4,500 zones for the entire United States vs. 5,700 zones just for California) and produced better results for the time and cost coefficients than the national-level networks. Following the estimation of the generalized cost function, the destination and mode choice models were re-estimated using the national-level zones and network data and the combined data from the four long-distance surveys (California, Ohio, New York and Wisconsin) to produce coefficients that represented a broader portion of the U.S.

Destination choice models are multinomial logit models used to choose the destination zones of the long-distance tours. These zones are based on the National Use Microdata Zones (NUMAs) established and described in the Section 2.1 for the research phase. For this set of models, all destinations that are 50+ miles away from origins were considered available. Five destination choice models exist, one for each tour purpose. The models are primarily functions of opportunities (represented by employment or households) and travel impedance. Opportunities that have significant effects on long-distance destination choices vary by tour purpose. In general, number of employment in accommodation, entertainment, medical, other services, retail, and wholesale industry; park areas; number of households; and college/university enrollment played a large role in determining the attractiveness of a destination. In this model, travel impedance (such as distance) was used to offset attractiveness of a destination zone. Other significant variables include logsum parameters from mode choice models, destination type (urban/rural), and tour duration.

Mathematical Formulation

The probability of a household choosing a destination is described by the multinomial discrete choice logit model equation (see Figure 26). The destination choice model represents the upper nest of a nested model, where the mode choice model represents the lower nest of the model (as shown in Figure 1). One model exists for each purpose within destination choice and another set of models for each purpose within mode choice. Five purposes exist: commute, business, visit friends and relatives, leisure, and personal business. This produces 10 individual destination and mode choice models. The alternatives (i) for the destination choice model are all destination zones and the alternatives (i) for the mode choice model are auto, bus, rail, and air.

Destination Choice

The utility component (V_i) for the destination choice model is presented in Figure 32.

$$V_i = \in_i + \ a * ModeChoiceLogsum + b * Log(Distance) + c * Distance^2$$

$$+d * NightsAway_0{}^2 + e * NightsAway_{1-2}{}^2 + f * Distance_{50-100}$$

$$+g * Distance_{100-150} + h * Distance_{150-250} + \ i * Distance_{250-500}$$

$$+j * Distance_{500-1000} + k * Distance_{1000-1500} + \ l * Distance_{>2000}$$

$$+ m * Destination_{Urban} \ + n * Destination_{Rural} \ + o * OD_{Urban}$$

$$+p * OD_{Rural} \ + q * Log(Size_0 + Exp(r) * Size_1$$

$$+Exp(s) * Size_2 + Exp(t) * Size_3 + Exp(u) * Size_4)$$

Figure 32. Equation. Destination choice utility.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- a[24], *q* are constrained (set=1) coefficients (see Table 19).
- *b, c, d, e, f, g, h, I, j, k, l, m, n, o, p, r, s, t, u* are estimated coefficients (see Table 19).
- *ModeChoiceLogsum* is the weighted value of generalized cost to each destination.
- *Distance* is the one-way distance in miles.
- $NightsAway_0$ is when the travel does not require any nights away.
- $NightsAway_{1-2}$ is when the travel requires one to two nights away.
- $Distance_{50-100}$ is the one-way distance in miles for trips that are 50-100 miles long.
- $Distance_{100-150}$ is the one-way distance in miles for trips that are 100-150 miles long.
- $Distance_{150-250}$ is the one-way distance in miles for trips that are 150-250 miles long.
- $Distance_{250-500}$ is the one-way distance in miles for trips that are 250-500 miles long.
- $Distance_{500-1000}$ is the one-way distance in miles for trips that are 500-1000 miles long.
- $Distance_{1000-1500}$ is the one-way distance in miles for trips that are 1000-1500 miles long.

[24] The commute tour purpose coefficient *a* is estimated, not constrained.

- $Distance_{1500\text{-}2000}$ is the one-way distance in miles for trips that are 1500-2000 miles long.
- $Distance_{<2000}$ is the one-way distance in miles for trips that are more than 2000 miles long.
- $Destination_{Urban}$ is whether the density of the destination zone is urban.
- $Destination_{Rural}$ is whether the density of the destination zone is rural.
- OD_{Urban} is whether the density of both the origin and the destination zones are urban.
- OD_{Rural} is whether the density of both the origin and the destination zones are rural.
- $Size_{1\text{-}4}$ is defined by the employment categories relevant to each tour purpose, as shown in Table 20.

Table 19. Destination choice models.

		Models by Tour Purpose…	Personal Business		Visit F&R		Leisure		Commute		Employer Business	
Coeff. #	Altern.	Variable	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
1	All	Mode choice logsum	1	constr	1	constr	1	constr	0.211	7.5	1	constr
2	All	Log (one-way distance)	-1.9	-39	-1.09	-30	-1.35	-38.2	-3.58	-46.7	-1.64	-34.3
3	All	One-way dist squared	0.006	11.7	0.0033	15.2	0.0045	16.5	0.0238	12.7	0.0035	15.5
4	All	Day trip*1-way dist squared	-0.0192	-15	-0.023	-19.9	-0.0269	-70	-0.0032	-3.8	-0.0084	-13.2
5	All	1-2 nights*1-way dist squared	-0.004	-7.1	-0.0104	-17.5	-0.012	-38.3	-7.30E-04	-0.8	-0.0022	-12.2
6	All	Data missing*1-way dist squared *	-0.003	-10.3	-0.0018	-15.5	-0.0021	-15.1	-0.0123	-7.5	-0.0017	-12.9
7	All	One-way dist 50-100 miles										
8	All	One-way dist 100-150 miles (calibrated)	-0.151 **(-0.101)**	-4.4	-0.185 (-0.185)	-7.2	-0.31 (-0.31)	-12.8	-0.464 **(-0.364)**	-7.6	-0.277 **(-0.207)**	-7.9
9	All	One-way dist 150-250 miles (calibrated)	-0.704 **(-0.604)**	-12.8	-0.719 (-0.719)	-17.8	-0.862 **(-0.702)**	-22.6	-0.784 **(-0.584)**	-8	-0.887 **(-0.807)**	-16.2
10	All	One-way dist 250-500 miles (calibrated)	-1.07 (-1.07)	-13.1	-1.21 (-1.21)	-20.2	**-1.41** **(-1.253)**	-24.2	-0.803 **(-0.603)**	-5.3	-1.12 **(-1.10)**	-14
11	All	One-way dist 500-1000 miles (calibrated)	0.808 **(0.408)**	7.1	0.229 **(-0.109)**	2.7	0.101 **(-0.101)**	1.2	1.61 **(-0.906)**	8.4	0.132 **(-0.232)**	1.2

		Models by Tour Purpose…	Personal Business		Visit F&R		Leisure		Commute		Employer Business	
Coeff. #	Altern.	Variable	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
12	All	One-way dist 1000-1500 miles (calibrated)	0.959 (0.959)	6.1	0.389 (0.389)	3.6	0.633 **(0.333)**	6	-0.581 **(-1.081)**	-1.6	0.151 **(0.051)**	1.1
13	All	One-way dist 1500-2000 miles (calibrated)	0.518 (0.518)	2.1	0.363 (0.363)	2.7	0.16 **(0.051)**	1.1	-2.44 **(-2.94)**	-3.6	0.235 **(0.041)**	1.4
14	All	One-way dist over 2000 miles (calibrated)	-0.037 (-0.037)	-0.1	0.184 (0.184)	1	-0.254 **(-0.604)**	-1.3	-12.4 **(-16.4)**	-6.9	0.376 **(-0.106)**	1.9
15	All	Dest zone has urban density	-0.162	-7.4	-0.448	-26.5	-0.344	-21.8	-0.108	-3	-0.239	-10.7
16	All	Dest zone has rural density	0.486	11.7	0.471	16.5	0.573	23.5	0.0175	0.3	0.573	14
17	All	O and D zones have urban density	-0.261	-6.1	0.0783	2.9	-0.0675	-2.6	0.0618	0.7	0.31	8.3
18	All	O and D zones have rural density	-0.569	-6	-0.306	-3.3	-0.555	-7.9	0.581	3.9	0.393	3.9
25	All	Log-size function multiplier	0.715	63.8	0.688	66.3	0.689	106.9	0.611	37.5	0.79	79.6
20	All	Size variable 0	1	constr	1	constr	1	constr	1	constr	1	constr
20	All	Size variable 1 (log of coeff.)	0.273	3.1	-1.35	-5	-0.68	-7.5	0.327	2	-1.2	-5.1
21	All	Size variable 2 (log of coeff.)	-11.6	-0.1	-0.615	-8.4	-37.3	0	-30	(*)	-30	(*)
22	All	Size variable 3 (log of coeff.)	-4.36	-15.9	-20	(*)	-30	(*)	-5.45	-4	-2.93	-24.9

		Models by Tour Purpose…	Personal Business		Visit F&R		Leisure		Commute		Employer Business	
Coeff. #	Altern.	Variable	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
23	All	Size variable 4 (log of coeff.)	-0.908	-11.9	-5.25	-6.4	1.31	24.1	-15.2	-0.1	-2.2	-18.9
		Model Fit Statistics										
		Observations	15130	--	27880	--	30865	--	6151	--	15987	--
		Final log-likelihood	-79405.8	--	-164121.7	--	-174552.1	--	-27130.8	--	-91013.5	--
		Rho-squared vs. 0	0.375	--	0.299	--	0.326	--	0.475	--	0.322	--
		Rho-squared vs. constants	0.14	--	0.118	--	0.078	--	0.119	--	0.088	--

Table 20. Definition of the size variables by purpose.

Tour Purpose	Size 0	Size 1	Size 2	Size 3	Size 4
Personal business	Medical employment	Entertainment employment	Other service employment	All other employment	University enrollment
Visit friends or relatives	Accommodation employment	Entertainment employment	Medical employment	All other employment	Households
Leisure/ vacation	Accommodation employment	Entertainment employment	Other service employment	All other employment	Square miles of public parks
Commuting	Other service employment	Entertainment employment	Retail/ wholesale employment	All other employment	University enrollment
Employers' business	Accommodation employment	Entertainment employment	Retail/ wholesale employment	All other employment	University enrollment

Mode Choice for Auto

The utility components (V_i) for the mode choice model are different for each mode. Auto is presented in Figure 33.

$$V_i = \in_i + \; a * ModeGC_{auto} + b * HH_{0car} + c * HH_{carcompetition} \; + d * PartySize_1$$

$$+e * PartySize_{3+} + f * NightsAway_0 + g * NightsAway_{7+}$$

$$+h * Distance_{>500}$$

Figure 33. Equation. Mode choice utility for auto.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a, b, c, d, e, f, g, h* are estimated coefficients (see Table 21).
- $ModeGC_{auto}$ is the weighted value of generalized cost to each destination (see Figure 34).
- HH_{0cars} is the households with zero car
- $HH_{carcompetition}$ is the households with fewer cars than adults.
- $PartySize_1$ is the trips with only one person traveling solo.
- $PartySize_{3+}$ is the trips with three or more people traveling together.
- $NightsAway_0$ is when the travel does not require any nights away.
- $NightsAway_{7+}$ is when the travel requires seven or more nights away.

- $Distance_{>500}$ is the one-way distance in miles for trips more than 500 miles in length.

The utility components (V_i) for the generalized cost component of the mode choice model applies to all modes. This mode generalized cost for auto is presented in Figure 34.

$$ModeGC_{auto} = a * Cost_{auto} + b * Time_{auto}$$

Figure 34. Equation. Mode generalized cost utility for auto.

Where,

- $Cost_{auto}$ is the auto operating cost and any tolls from the origin to the destination for the trip in cents.
- $Time_{auto}$ is the auto travel time from the origin to the destination in minutes.

Mode Choice for Bus

The utility components (V_i) for the bus mode in the mode choice model is presented in Figure 35.

$$V_i = \in_i + \ a * ModeGC_{bus} + b * Log\left(\frac{HHincome}{1000}\right) + c * Log(Density_{origin})$$

$$+d * Log(Density_{destination}) + e * Distance_{50-150}$$

Figure 35. Equation. Mode choice utility for bus.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a, b, c, d, e* are estimated coefficients (see Table 21).
- $ModeGC_{bus}$ is the weighted value of generalized cost to each destination (see Figure 36).
- *HHincome* is the total gross household income in dollars.
- $Density_{origin}$ is the density of the origin zone.
- $Density_{destination}$ is the density of the destination zone.
- $Distance_{50-150}$ is the one-way distance in miles for trips between 50 and 150 miles in length.

The mode generalized cost for bus is shown in Figure 36.

$$ModeGC_{bus} = a * Cost_{bus} + b * Time_{bus}$$

Figure 36. Equation. Mode generalized cost utility for bus.

Where,

- $Cost_{bus}$ is the bus fare from the origin to the destination for the trip in cents.
- $Time_{bus}$ is the bus travel time from the origin to the destination in minutes.

Mode Choice for Rail

The utility components (V_i) for the rail mode in the mode choice model is presented in Figure 37.

$$V_i = \in_i + a * ModeGC_{rail} + b * Log\left(\frac{HHincome}{1000}\right) + c * Log(Density_{origin})$$

$$+d * Log(Density_{destination}) + e * Distance_{50-150}$$

Figure 37. Equation. Mode choice utility for rail.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a, b, c, d, e* are estimated coefficients (see Table 21).
- $ModeGC_{rail}$ is the weighted value of generalized cost to each destination (see Figure 38).
- *HHincome* is the total gross household income in dollars.
- $Density_{origin}$ is the density of the origin zone.
- $Density_{destination}$ is the density of the destination zone.
- $Distance_{50-150}$ is the one-way distance in miles for trips between 50 and 150 miles in length.

The mode generalized cost for rail is shown in Figure 38.

$$ModeGC_{rail} = a * Cost_{rail} + b * Time_{rail} + c * Transfers + d * Frequency$$

$$+e * Access + f * \frac{Access}{Distance}$$

Figure 38. Equation. Mode generalized cost utility for rail.

Where,

- $Cost_{rail}$ is the rail fare from the origin to the destination for the trip in cents.
- $Time_{rail}$ is the rail travel time from the origin to the destination in minutes.
- *Transfers* is the number of transfers required for the rail trip.
- *Frequency* is the number of trains per week for the rail trip.
- *Access* is the distance for access to rail and egress from rail.
- *Distance* is the distance from the origin to the destination by rail.

Mode Choice for Air

The utility components (V_i) for the air mode in the mode choice model is presented in Figure 39.

$$V_i = \in_i + a * ModeGC_{rail} + b * Log\left(\frac{HHincome}{1000}\right) + c * Log(Density_{origin})$$

$$+d * PartySize_1 + e * NightsAway_0 + f * NightsAway_{1-2}$$

$$+g * Log(Density_{destination}) + h * Distance_{50-150}$$

Figure 39. Equation. Mode choice utility for air.

Where,

- $\in_i$ is the error term, also referred to as a constant for each alternative *i*.
- *a, b, c, d, e, f, g, h* are estimated coefficients (see Table 21).
- $ModeGC_{rail}$ is the weighted value of generalized cost to each destination (see Figure 38).
- *HHincome* is the total gross household income in dollars.
- $Density_{origin}$ is the density of the origin zone.

- $Density_{destination}$ is the density of the destination zone.
- $PartySize_1$ is the trips with only one person traveling solo.
- $NightsAway_0$ is when the travel does not require any nights away.
- $NightsAway_{1\text{-}2}$ is when the travel requires one or two nights away.
- $Distance_{50\text{-}150}$ is the one-way distance in miles for trips between 50 and 150 miles in length.

The mode generalized cost for air is shown in Figure 40.

$$ModeGC_{air} = a * Cost_{air} + b * Time_{air} + c * Transfers + d * Frequency$$

$$+e * Access + f * \frac{Access}{Distance}$$

Figure 40. Equation. Mode generalized cost utility for air.

Where,

- *a, b, c, d, e, f* are estimated coefficients (see Table 21).
- $Cost_{air}$ is the air fare from the origin to the destination for the trip in cents.
- $Time_{air}$ is the air travel time from the origin to the destination in minutes.
- *Transfers* is the number of transfers required for the air trip.
- *Frequency* is the number of planes per week for the air trip.
- *Access* is the distance for access to air and egress from air.
- *Distance* is the distance from the origin to the destination by air.

We assume that the errors $\in_i$ are independent and identically distributed such that $E[\in_i] = 0$ and $var[\in_i] = \sigma^2$. Typically, we assume $\in_i \sim N(0, \sigma^2)$ as a basis for inference (e.g., t-tests on parameters).

Table 21. Mode choice models.

		Models by Tour Purpose…	Personal Business		Visit F&R		Leisure		Commute		Employer Business	
Coeff. #	Altern.	Variable	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
1	All	Mode generalized cost	0.281	12.6	0.344	24.3	0.344	22.9	0.298	7.2	0.265	19.1
101	Car	HH has no cars	-2.03	-8.8	-2.2	-14.7	-1.22	-7.8	-0.538	-0.8	-1.59	-5.8
102	Car	HH has car competition	-0.571	-5.7	-0.269	-3.2	-0.313	-4.3	-0.182	-1.3	-0.161	-1.7
103	Car	Tour party size = 1	-0.821	-9	-0.894	-12.7	-0.467	-6.4	--	--	--	--
104	Car	Tour party size = 3 or more	--	--	0.539	6.8	--	--	--	--	-0.515	-5.2
105	Car	0 nights away from home	0.332	2.3	0.412	2.7	-0.856	-9.2	-0.678	-2.4	0.366	2.2
106	Car	7+ nights away from home	--	--	--	--	--	--	--	--	0.502	2.9
107	Car	Missing nights data *	0.401	2.8	0.164	1.4	-0.112	-1.1	-0.786	-2.9	0.115	0.8
112	Car	One-way dist over 500 miles	-1.07	-8.1	-1.47	-17.6	-0.993	-12.3	--	--	-1.21	-12.6
200	Bus	Estimated constant / Calibrated constant	-7.27/ -6.96	-14.8	-5.86/ -5.17	-12.3	-0.847/- 0.847	-3.1	-5.58 / -4.51	-6.3	-6.01/ -5.65	-7.9
207	Bus	Missing HH income data *	-0.0729	-0.1	-2.57	-5.2	-3.06	-12.7	-1.24	-1.3	-0.905	-1.2
208	Bus	Log of (HH income/1000)	0.101	1.2	-0.524	-6.3	-0.95	-17.5	-0.14	-0.8	-0.274	-1.8
209	Bus	Log of origin zone density	0.135	3.7	0.274	7.1	0.0425	1.7	0.129	1.9	0.175	3.3
210	Bus	Log of dest zone density	0.284	8.5	0.14	3.6	0.0416	2.3	0.336	6.7	0.239	4.9
215	Bus	One-way dist 50-150 miles	-0.236	-1.6	0	(*)	-0.41	-4.2	-2.08	-9.7	-0.682	-3.4
300	Rail	Estimated constant / Calibrated constant	-12.9/ -12.95	-15	-7.78/ -7.35	-13.7	-11.6/ -12.0	-16.1	-19.5/ -17.7	-18.7	-12.6/ -12.7	-16.2
307	Rail	Missing HH income data *	-0.13	-0.2	-0.92	-2.1	-1.84	-2.1	6.56	9.1	-0.416	-0.7
308	Rail	Log of (HH income/1000)	0.12	1	-0.213	-2.5	0.0498	0.5	1.28	8.5	-0.132	-1.2
309	Rail	Log of origin zone density	0.274	5.2	0.256	6.9	0.179	4.2	0.24	4.5	0.186	4.8
310	Rail	Log of dest zone density	0.802	12.8	0.371	8.6	0.757	16.3	1.08	22	1.05	19.6
315	Rail	One-way dist 50-150 miles	--	--	--	--	-0.348	-2.1	--	--	-0.449	-3
400	Air	Estimated constant / Calibrated constant	-5.6/ -5.12	-10	-6.18/ -6-18	-20.8	-4.17/ -3.93	-12.9	-8.51/ -7.61	-5.2	-8.04 /-8.94	-21.1
407	Air	Missing HH income data *	0.901	1.8	1.17	4.3	0.264	0.9	4.01	2.5	3.42	9.6

		Models by Tour Purpose…	Personal Business		Visit F&R		Leisure		Commute		Employer Business	
Coeff. #	Altern.	Variable	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
408	Air	Log of (HH income/1000)	0.197	2	0.0917	1.6	-0.0442	-0.7	0.745	2.2	0.65	9.2
409	Air	Log of origin zone density	0.153	4.6	0.151	8	0.0676	3.4	0.116	1.4	0.156	7.5
410	Air	Log of dest zone density	0.178	5.7	0.15	8.4	0.188	11.5	0.305	3.5	0.221	11.1
411	Air	0 nights away from home	-2.26	-8.3	-1.8	-7.7	-3.04	-16.9	-4.06	-5.8	-1.19	-6
412	Air	1-2 nights away from home	-1.01	-6.1	-1.03	-9.6	-1.57	-12.7	-1.57	-3.2	-0.219	-2
413	Air	Missing nights data *	-0.946	-4.8	-0.546	-3.9	-1.12	-8.9	-1.91	-4.6	-0.795	-4.9
414	Air	Tour party size = 1	--	--	--	--	--	--	--	--	0.626	7.6
415	Air	One-way dist 50-150 miles	-3	-7	-2.52	-9.9	-1.7	-8.5	-5.01	-4.8	-3.19	-16.3

Model Fit Statistics

Observations	14743	--	27602	--	30077	--	6076	--	15824
Final log-likelihood	-2620.7	--	-4614.8	--	-6478.3	--	-1604.3	--	-3940.5
Rho-squared vs. 0	0.852	--	0.863	--	0.816	--	0.783	--	0.797
Rho-squared vs. constants	0.354	--	0.525	--	0.385	--	0.4	--	0.542

Estimation Results for Destination Choice

The primary explanatory variables in the destination choice model produce an increasing disutility on one-way distance, an example of which is shown in Figure 41. This demonstrates that the one-way distance disutility will increase faster for shorter-distance trips.

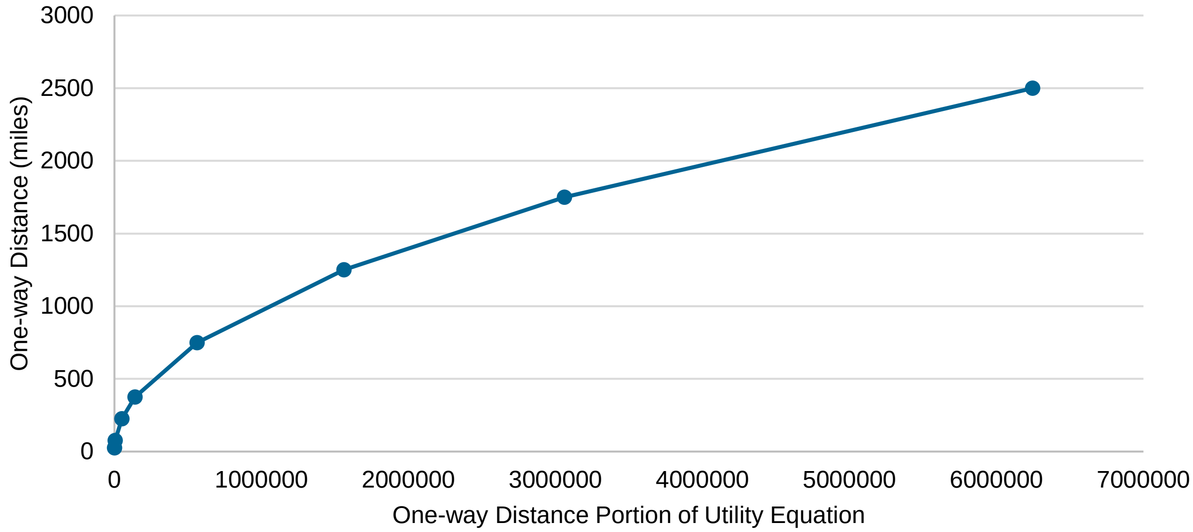


Figure 41. Example distance decay function.

The other results for destination choice are as follows:

- Day trips (with zero nights away) will be much less likely for longer distance trips and trips with one to two nights away will be slightly less likely for longer distance trips.
- Density has an impact on destination choice. Urban densities tend to discourage choosing a destination, except for employer business trips where it encourages choosing a destination in an urban area. Rural densities tend to encourage choosing a destination for all purposes, except when both the origin and destination zones are in rural areas for nonwork purposes.
- Size variables will encourage trips to zones with relevant employment and discourage trips to zones with nonrelevant employment. For example, size variables will encourage leisure trips to zones with accommodation and entertainment employment and public park acres and discourage trips to zones with other types of employment.

Estimation Results for Mode Choice

The generalized cost coefficients for mode choice were estimated from the California data and transferred to ensure consistency across purposes and modes (see Table 22). The data from the California networks were more spatially detailed than the national networks, providing improved estimates of the generalized cost coefficients. Employer's business purpose shows a smaller impact from the cost coefficient and a larger impact from transfers compared to other purposes, as expected, as well as a smaller overall impact from the mode generalized cost coefficient in the mode choice utility equation.

The mode choice model estimation results are shown in Table 21. These model estimation results include:

- Parties of one are more likely to take air for employer business trips and less likely to take car for nonwork purposes. Parties of three or more are more likely to take car for visiting friends and relatives and less likely to take car for employer business trips.
- Households without cars or with competition for cars are much less likely to choose car for all trip purposes.
- Travelers are less likely to choose air for short trips (two or fewer nights away) and more likely to choose car for all trip purposes. Employer business trips are more likely to choose car for longer trips (seven nights or more).
- Higher income travelers are more likely to choose air for all purposes, except for commute trips where higher income travelers are more likely to choose rail.
- Shorter trips (50-100 miles) are less likely to be on bus or rail modes while longer trips (more than 500 miles) are much less likely to be made by car.
- Travelers in higher density areas tend to favor bus, rail and air modes.

Table 22. Generalized cost coefficients for mode choice models.

		Models by Tour Purpose…	Personal Business		Visit F&R		Leisure		Commute		Employer Business	
Coeff. #	Altern.	Variable	Coefficient		Coefficient		Coefficient		Coefficient		Coefficient	
10	All	Cost	-0.006		-0.006		-0.006		-0.006		-0.0025	
11	Car	Time	-0.002		-0.002		-0.002		-0.002		-0.002	
21	Bus	Time	-0.0015		-0.0015		-0.0012		-0.0015		-0.0015	
31	Rail	Time	-0.002		-0.0015		-0.0012		-0.0015		-0.0015	
32	Rail	Transfers	-0.3		-0.3		-0.3		-0.3		-0.5	
33	Rail	Frequency/week	0.06		0.06		0.06		0.06		0.06	
34	Rail	Access+egress distance	-0.025		-0.015		-0.02		-0.025		-0.015	
35	Rail	Access+egress distance/car distance	-1.16		-3.04		-2.36		-1.16		-1.69	
41	Air	Time	-0.0015		-0.0015		-0.0015		-0.0015		-0.0015	
42	Air	Transfers	-0.3		-0.3		-0.15		-0.3		-0.5	
43	Air	Frequency/week	0.06		0.06		0.06		0.06		0.12	
44	Air	Access+egress distance	-0.005		-0.005		-0.009		-0.005		-0.006	
45	Air	Access+egress distance/car distance	-1.86		-3.3		-0.46		-1.86		-4.93	
46	Air	On-time percentage	0.015		0.03		0.015		0.03		0.03	

The approach used for the model application was 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 23 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, to match the observed shares closely when the models are applied to the four-State estimation dataset. This is discussed in Chapter 3. Long-Distance Model Development and Chapter 5. Highway Assignment.

Table 23. 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%
Business: 150-350 miles (1-way)	16.5%	0.2%	0.6%	1.2%
Business: 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%
Commute: 150-350 miles (1-way)	5.3%	0.4%	0.2%	0.0%
Commute: 350+ miles (1-way)	1.1%	0.0%	0.0%	1.3%
Visiting friends & relatives: 50-150 miles (1-way)	59.6%	0.4%	0.5%	0.1%
Visiting friends & relatives: 150-350 miles (1-way)	22.8%	0.2%	0.3%	0.3%
Visiting friends & relatives: 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%
Leisure: 150-350 miles (1-way)	21.3%	0.8%	0.1%	0.2%
Leisure: 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%
Personal business: 150-350 miles (1-way)	18.2%	0.6%	0.2%	0.2%
Personal business: 350+ miles (1-way)	4.9%	0.2%	0.0%	2.7%

CHAPTER 4. MODEL CALIBRATION

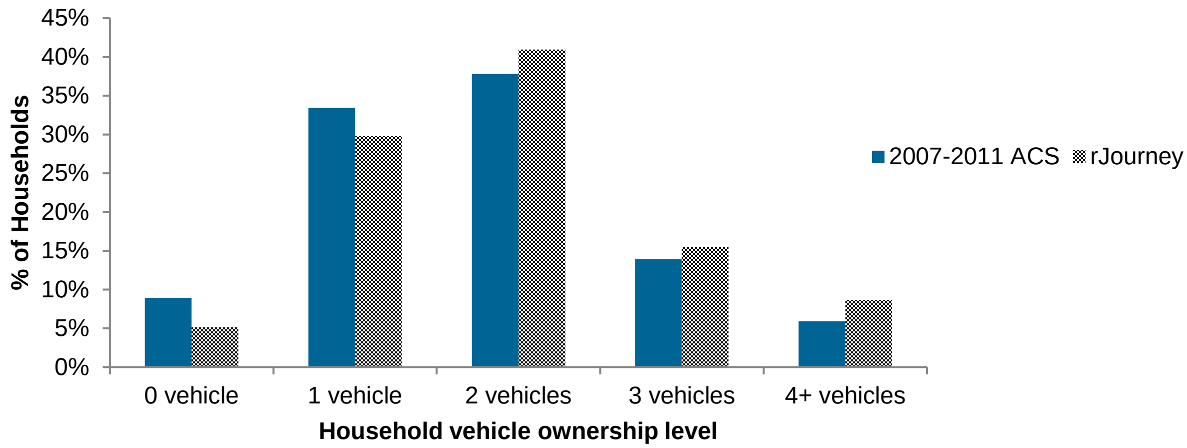
Model calibration is the process of applying the estimated models, comparing the results to observed values, and adjusting either the model specification or the alternative-specific constants. The various components of rJourney are vertically linked to ensure dependency between upper- and lower-level model components. As a result, calibrating one model component is likely to affect outcomes of other model components. In such cases, the general approach is to calibrate the model components in the order in which they are applied (i.e., the upper-level models are calibrated before the lower-level models). In this instance, the research team calibrated the tour generation-related model component first, followed by destination- and mode choice models. The calibration process was applied in an iterative manner until the model, performing as a system, converged to a stable set of parameter values for all of the model components and the observed travel patterns were well represented.

Table 24 summarizes rJourney model components in the order in which they were calibrated, if required. The population synthesizer is the first step in the modeling system. Calibration and validation involves checking the aggregate distributions against the observed distributions. Auto ownership[1] did not require any calibration since the model prediction matched ACS data reasonably well (Figure 42). And, after tour generation model was calibrated, it was not necessary to calibrate tour-scheduling, tour duration, and travel tour party size models (please see Figure 43, Figure 44, and Figure 45 for tour-scheduling, tour duration, and travel tour party size models, respectively). In these models, the five-State merged household travel survey dataset was used to represent the observed data. The calibration process of population synthesis, tour generation, tour destination, and tour mode choice models is discussed in subsequent sections.

Table 24. Model components that required calibration.

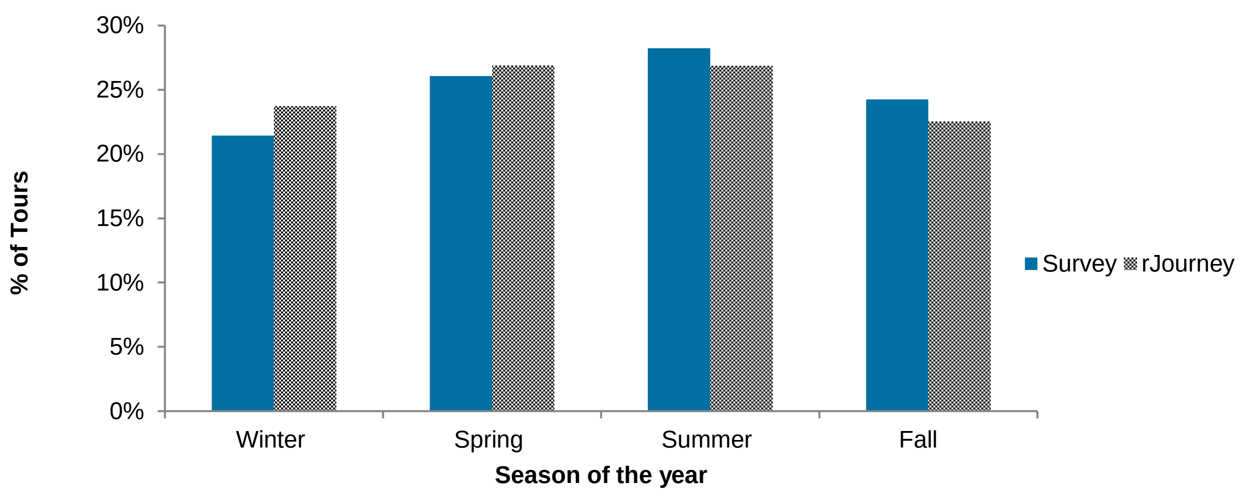
Long-Distance Model Components	Calibration Required
Population Synthesizer	No
Auto Ownership Model	No
Tour Generation Models	Yes
Scheduling Models	No
Party Size Models	No
Destination Choice Models	Yes
Mode Choice Models	Yes

[1] This report uses the terms "vehicle," "auto," and "car" interchangeably.



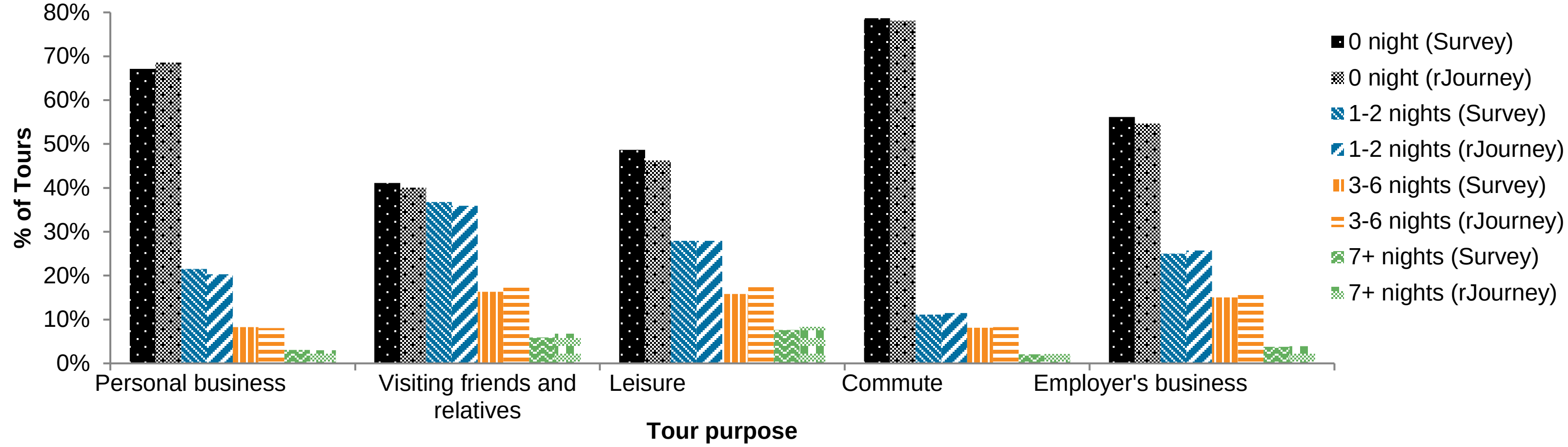
Source: FHWA

Figure 42. Percentage of households, by vehicle ownership level.



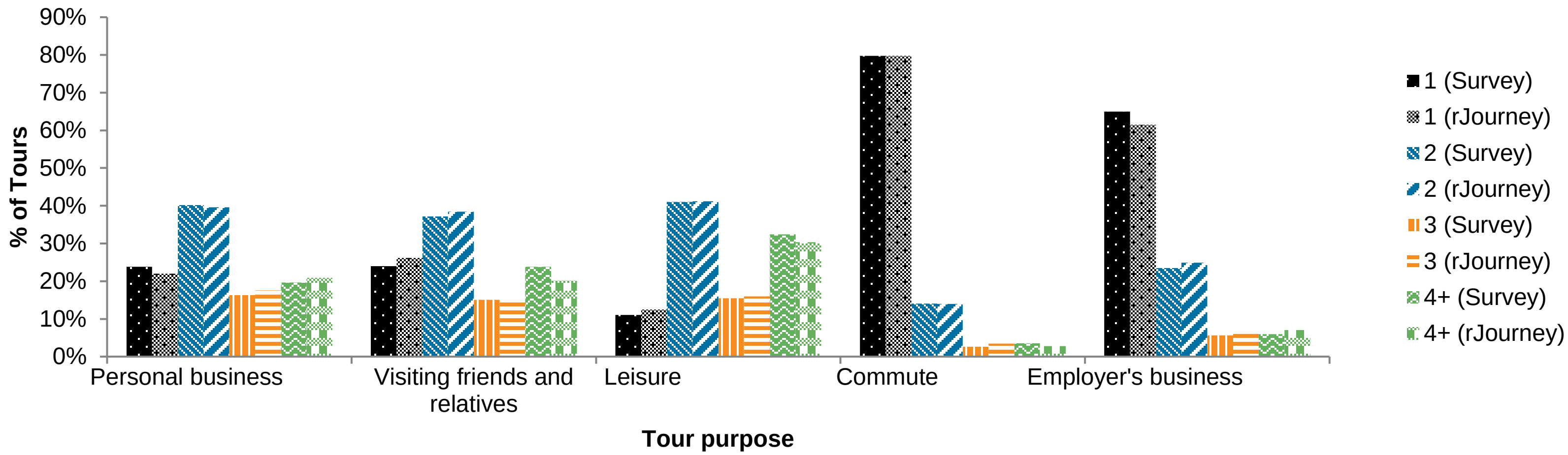
Source: FHWA

Figure 43. Percentage of tours, by season of the year.



Source: FHWA

Figure 44. Percentage of tours, by number of nights away from home.



Source: FHWA

Figure 45. Percentage of tours, by travel tour party size.

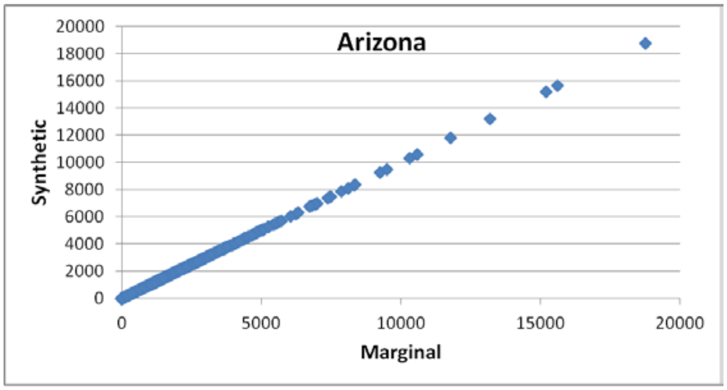
4.1 Population Synthesis

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 the population synthesis effort using 2007–2011 ACS datasets.

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 U.S. Census Bureau. Figure 46 through Figure 48 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 49 through Figure 51 show a comparison of the synthetic population versus the marginal control total at the person level.

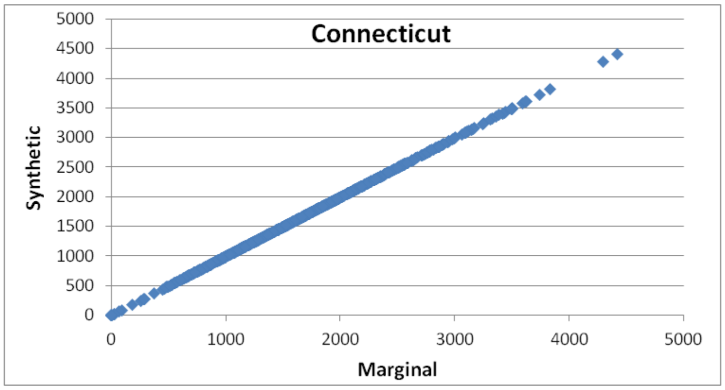
As mentioned previously, PopGen does not exactly match person totals in its attempt to control for the number of households. Slight inconsistencies exist between household- and person-level controls, and the Monte Carlo simulation process by which households are drawn into the synthetic population introduces 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. 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. Comparisons are possible 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.



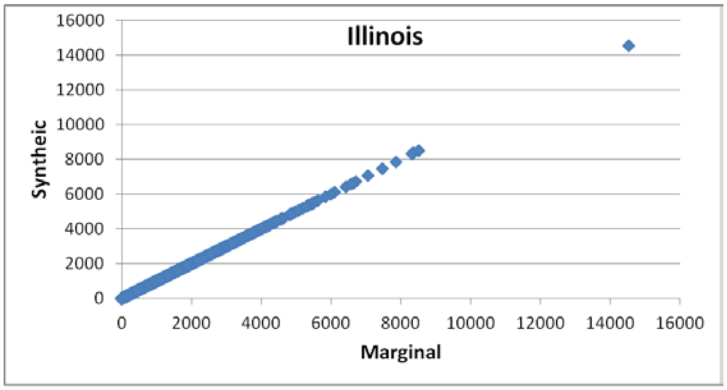
Source: FHWA

Figure 46. Comparison of number of households in synthetic population versus marginal control total for census tracts in Arizona.



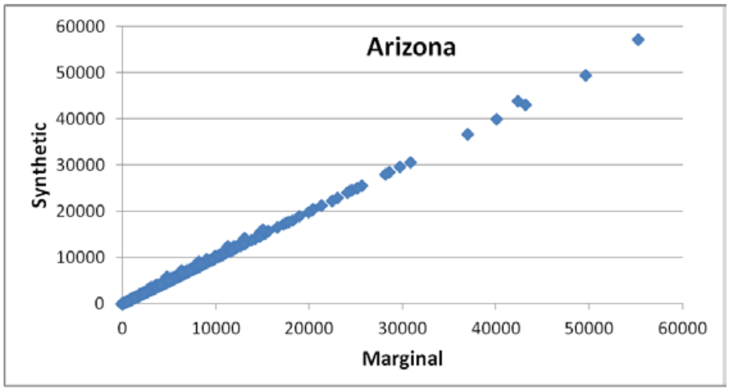
Source: FHWA

Figure 47. Comparison of number of households in synthetic population versus marginal control total for census tracts in Connecticut.



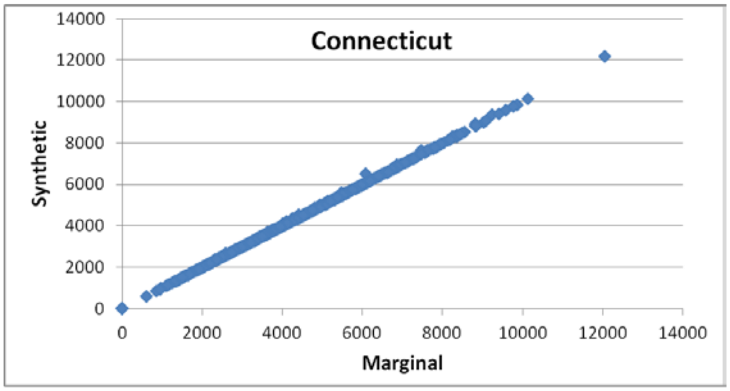
Source: FHWA

Figure 48. Comparison of number of households in synthetic population versus marginal control total for census tracts in Illinois.



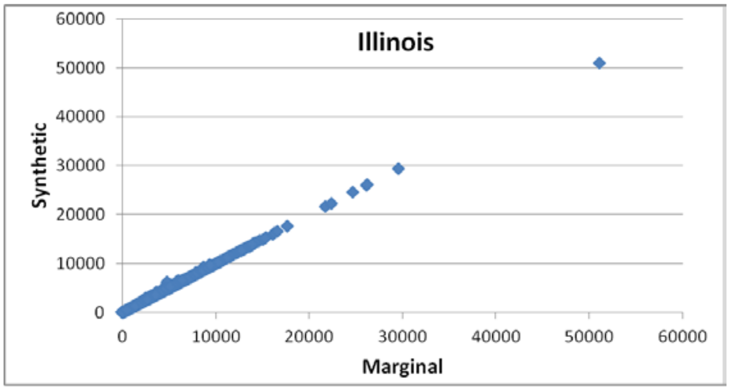
Source: FHWA

Figure 49. Comparison of number of persons in synthetic population versus marginal control total for census tracts in Arizona.



Source: FHWA

Figure 50. Comparison of number of persons in synthetic population versus marginal control total for census tracts in Connecticut.

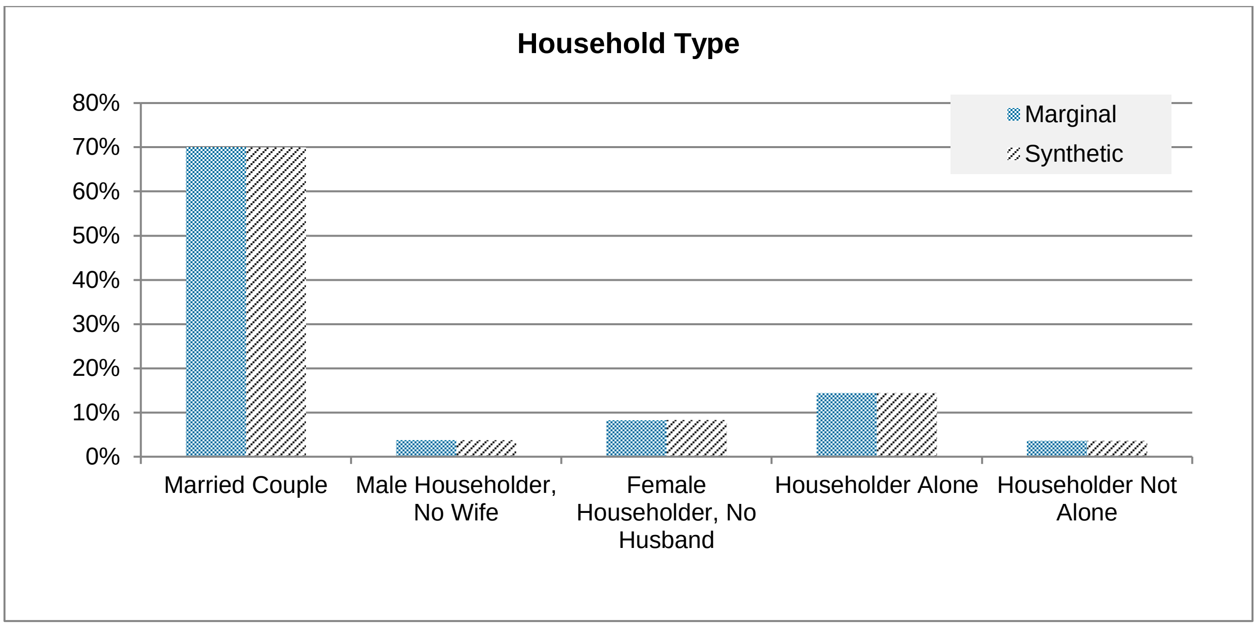


Source: FHWA

Figure 51. Comparison of number of persons in synthetic population versus marginal control total for census tracts in Illinois.

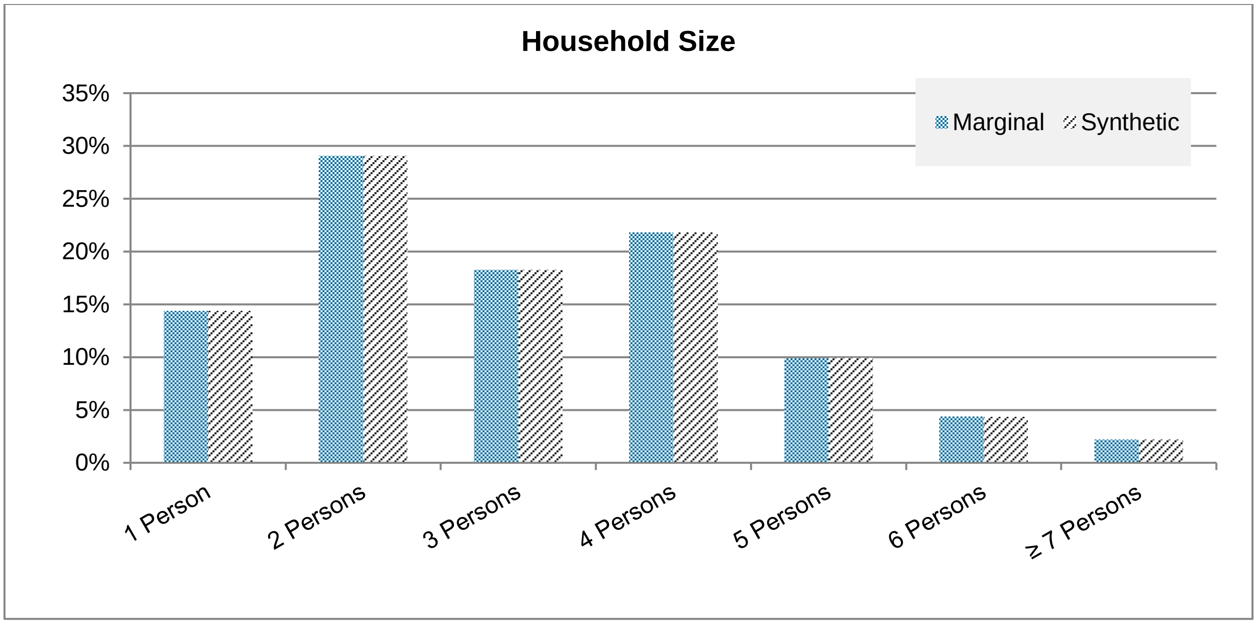
The set of graphs in Figure 52 through Figure 55 show a comparison of household and person attributes for one randomly chosen census tract in Maricopa County (Greater Phoenix metropolitan region) in Arizona. In the interest of brevity, such comparisons are not shown for other census tracts in the country, 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 56 through Figure 59, 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.



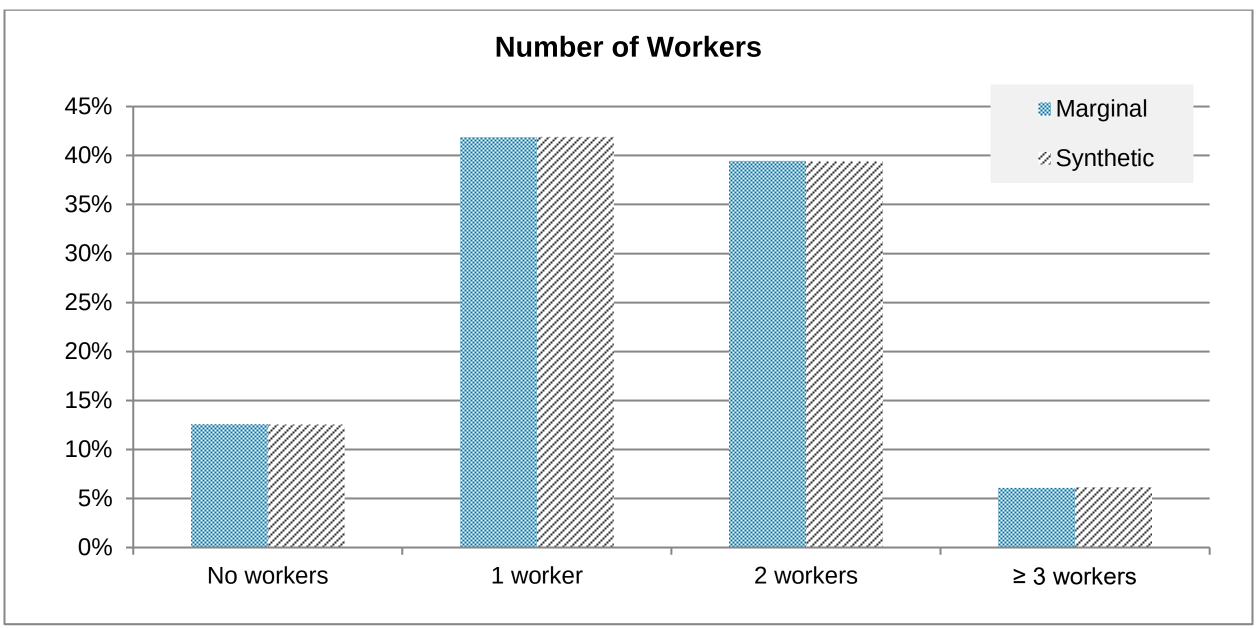
Source: FHWA

Figure 52. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (household type).



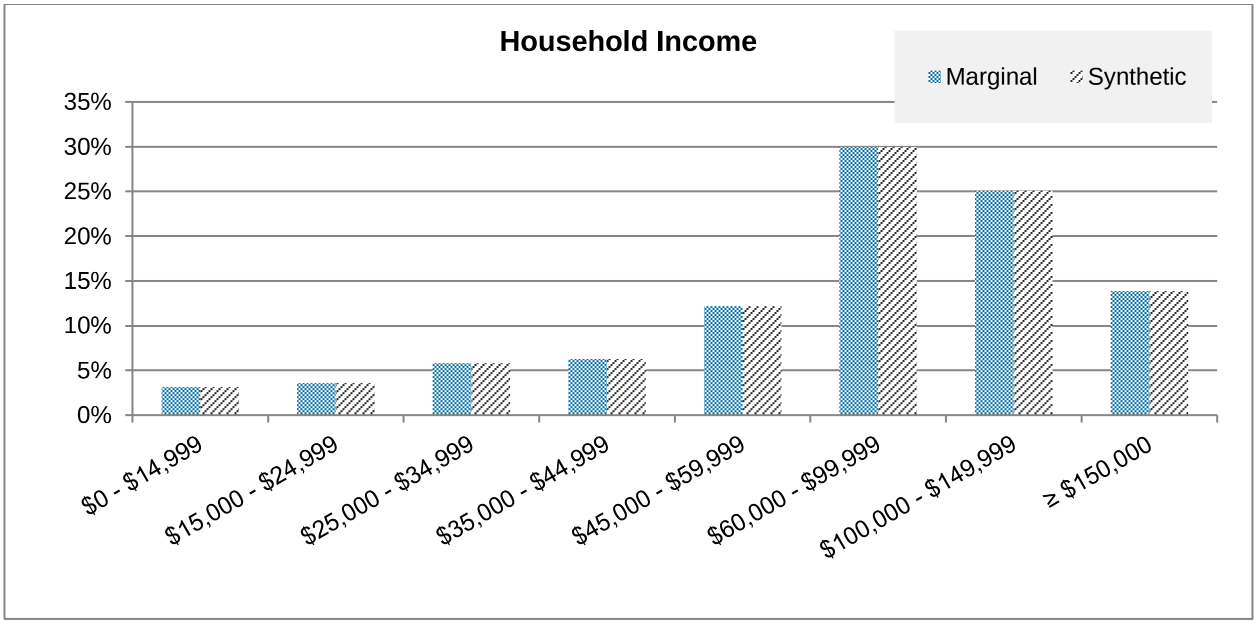
Source: FHWA

Figure 53. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (household size).



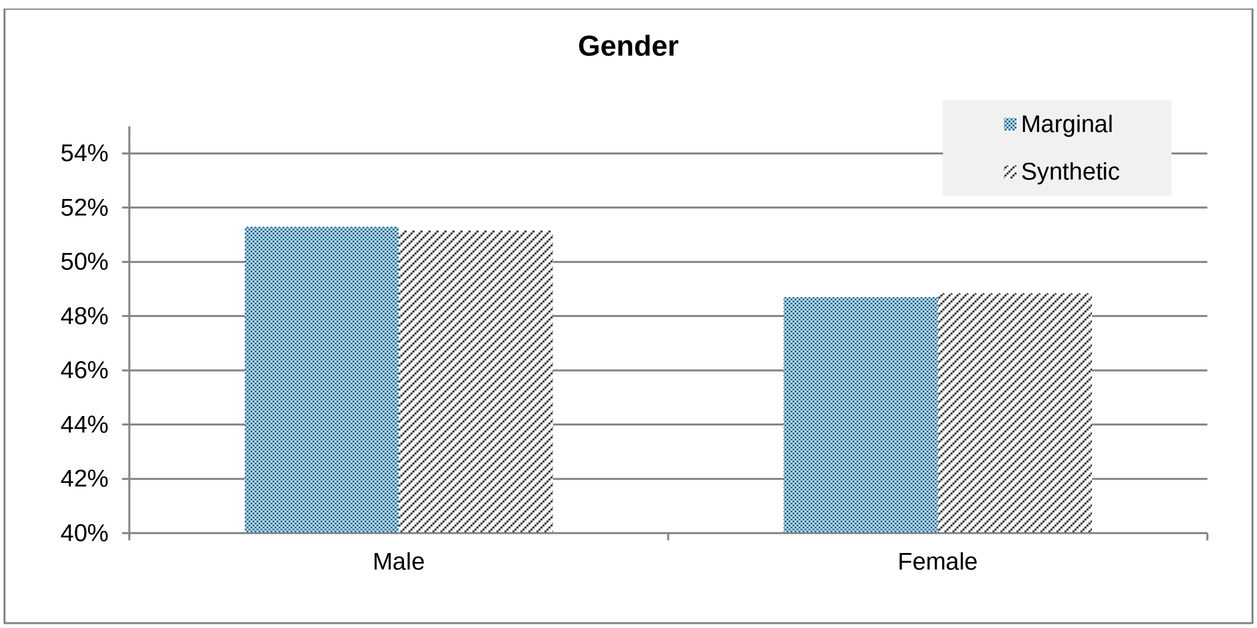
Source: FHWA

Figure 54. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (number of workers).



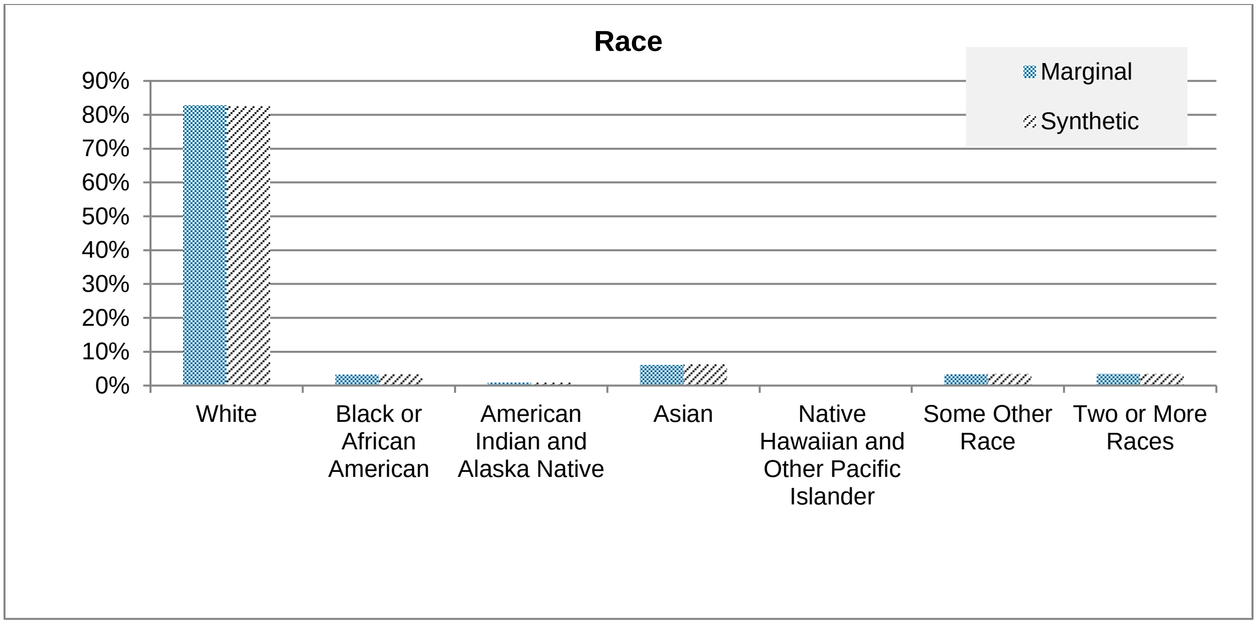
Source: FHWA

Figure 55. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (household income).



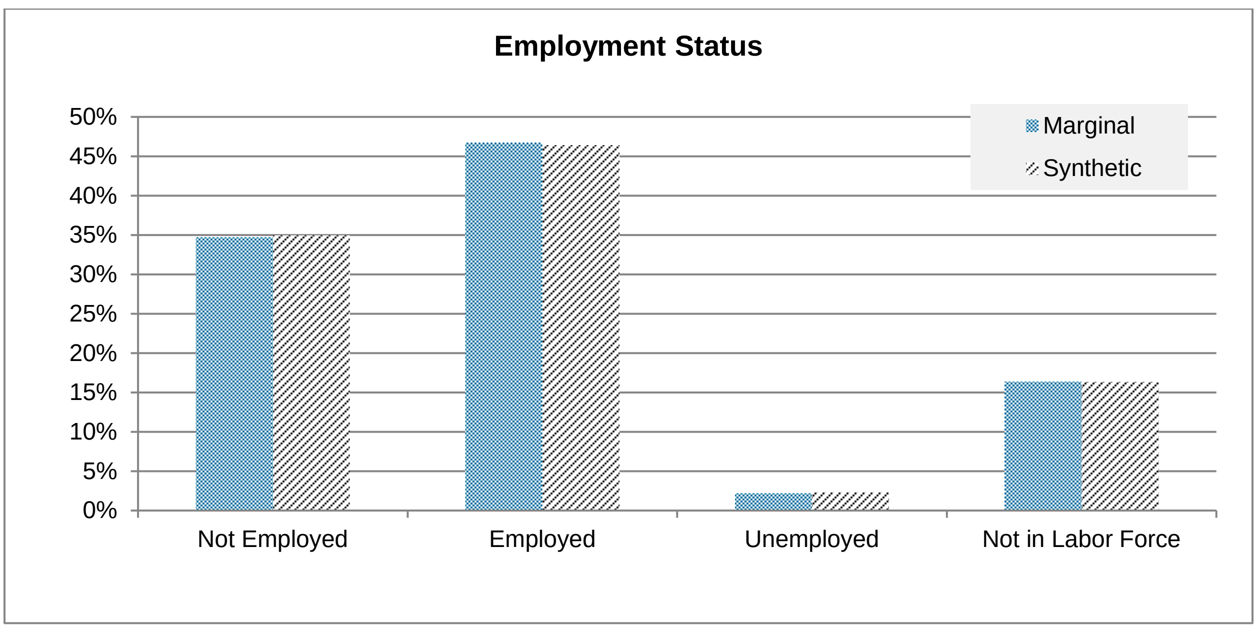
Source: FHWA

Figure 56. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (gender).



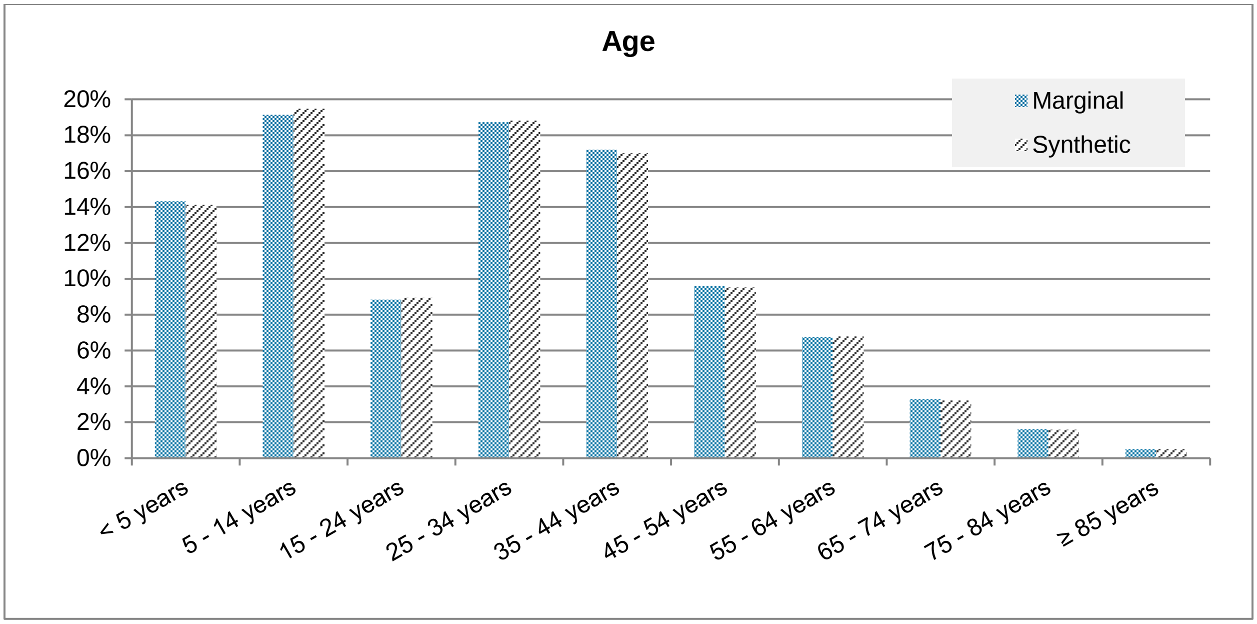
Source: FHWA

Figure 57. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (race).



Source: FHWA

Figure 58. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (employment status).



Source: FHWA

Figure 59. Comparison of control distributions between actual synthetic populations (census tract 522745 in Maricopa County, Arizona) (age).

4.2 Tour Generation Models

Tour generation models include two models applied sequentially: 1) for each tour purpose, the first model predicts whether or not a household undertakes a long-distance tour within a period of one week; and 2) the second model predicts whether or not a household undertakes more than one long-distance tour by purpose in one week. In application mode, these two models jointly predict number of tours by purpose generated by households over one year. The tour purposes are: personal business, visiting friends and relatives, leisure, commute, and employer's business. Many variables have significant effects on the likelihood of long-distance tour generation by purpose, including household size, presence of children, age of householder, household income, household auto ownership level relative to number of adults, distance between origin and primary destination, tour duration, and month of the year.

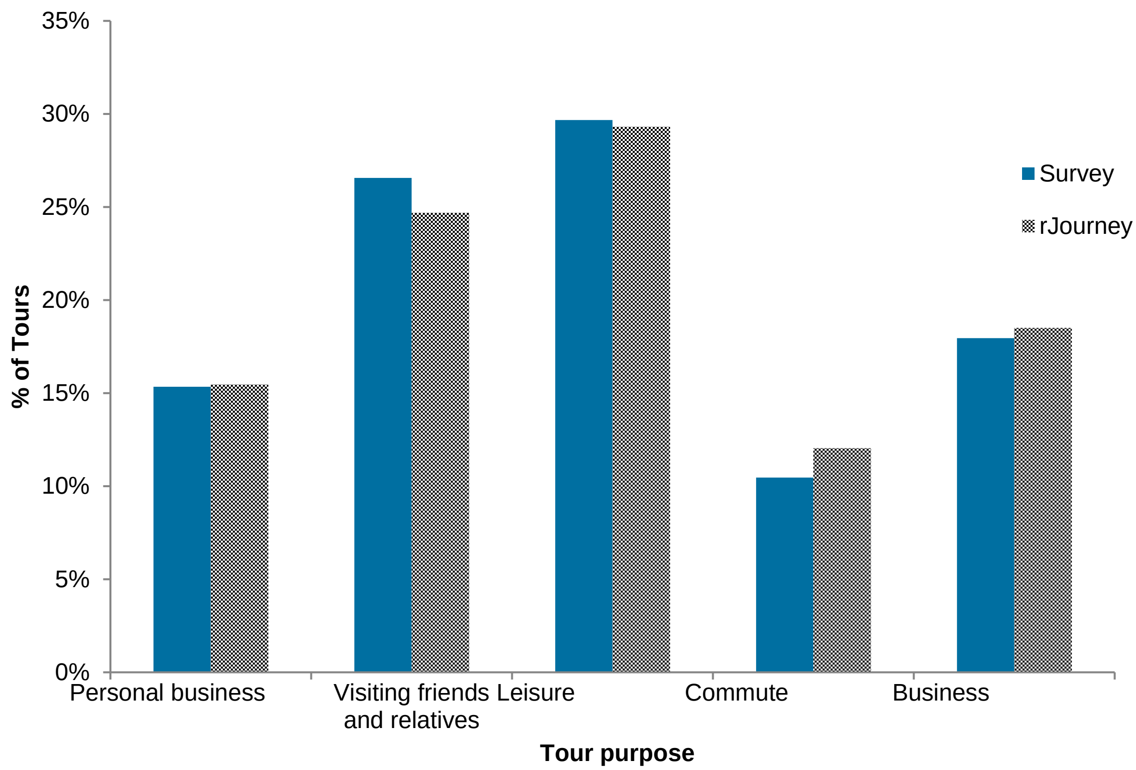
Calibration of the tour generation model involved the change of the alternative-specific constants to match observed tour rates by purpose with model prediction. Table 24 shows weekly tour rates by tour purpose from survey data and calibrated model prediction. Survey tour rates were calculated using data from the aforementioned five household travel surveys.[26] In general, tour rates predicted by rJourney closely match observed data. While the frequency models do not

[26] For brevity, all quantities that are derived using data from the household travel surveys will be referred to as "survey" instead of "five household travel surveys."

control for tour distribution by purpose, Figure 60 shows there is significant alignment between observed and model-predicted tour distribution by purpose.

Table 25. Weekly tour rate, by purpose.

Tour Purpose	Weekly Tours per Household		Difference
	Survey	rJourney	
Personal Business	0.034	0.031	-0.003
Visiting Friends and Relatives	0.057	0.049	-0.008
Leisure	0.066	0.058	-0.008
Commute	0.028	0.024	-0.004
Employer's Business	0.042	0.037	-0.005

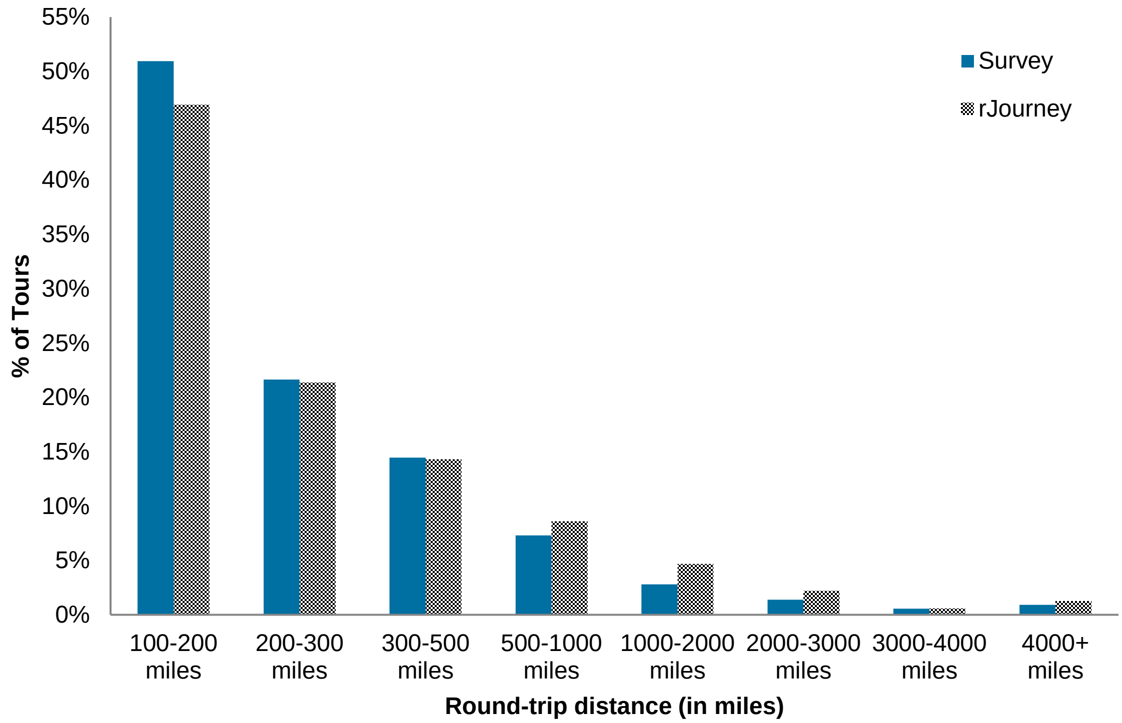


Source: FHWA

Figure 60. Percentage of tours, by purpose.

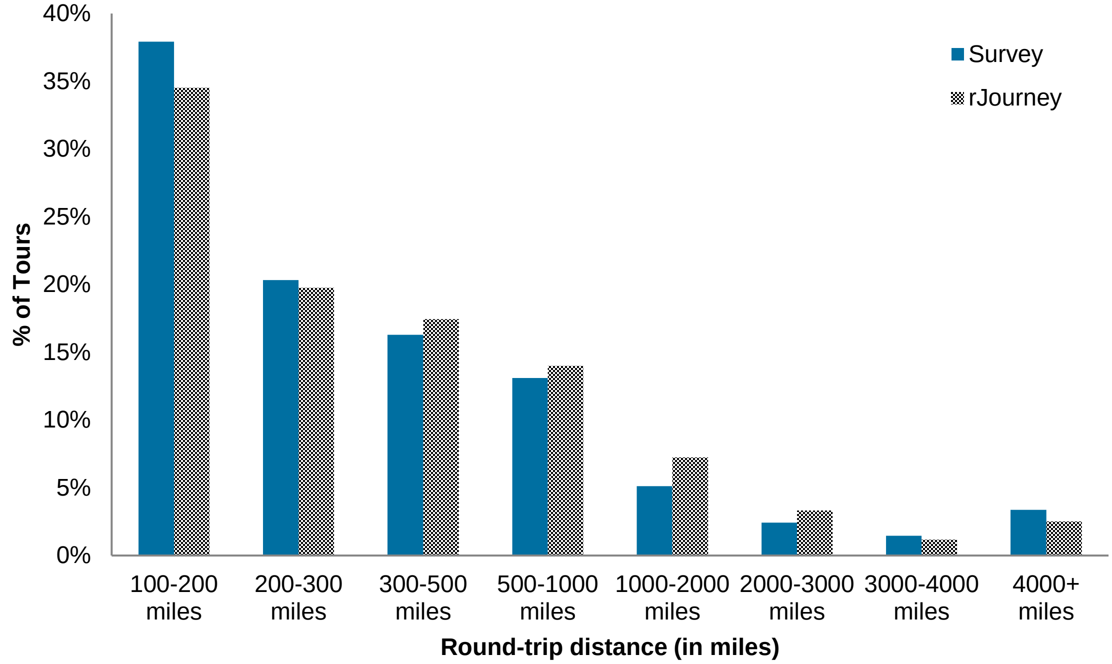
4.3 Destination Choice Models

The model calibration process revealed that for almost all purposes, there was some under-prediction of relatively short-distance tours and some over-prediction of relatively long-distance tours. To address this discrepancy, minor adjustments were made to relevant distance-related coefficients. Figure 61 to Figure 65 compare calibrated tour-length distribution for each purpose with survey data. In general, the model's predicted tour-length distributions are similar to observed tour-length distribution. Where there are divergences between two distributions, the differences are within 4 percent. Table 26 presents average person-miles traveled, by purpose. While predicted average person-miles traveled for commute and employer's business tours match survey data well, some variations between model prediction and survey data exist for non-work-related tours. These variations may be due to rJourney over-predicting tours within 1,000- to 2,000-mile tour lengths.



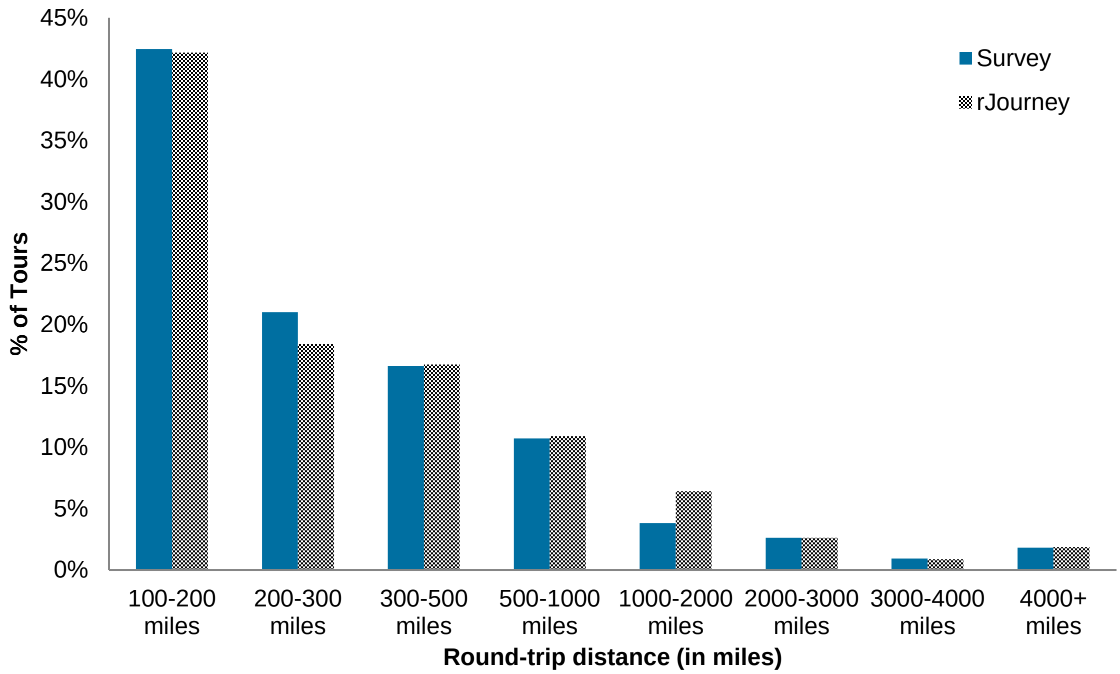
Source: FHWA

Figure 61. Round-trip Distance-band distribution, by purpose—personal business.



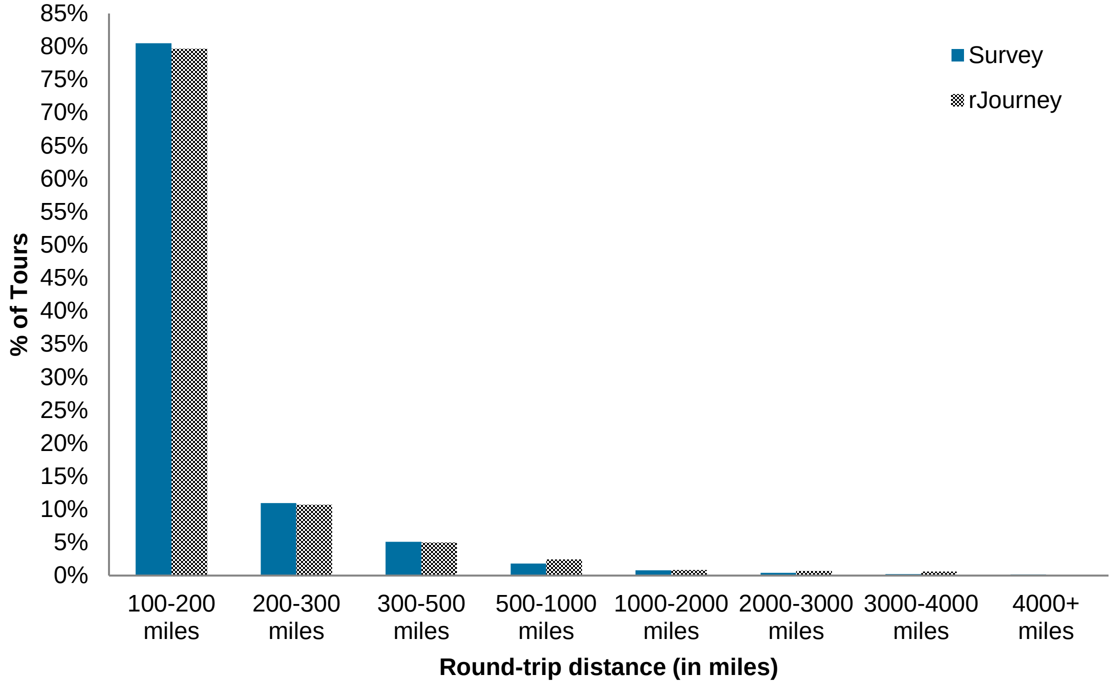
Source: FHWA

Figure 62. Round-trip distance-band distribution, by purpose—visiting friends and relatives.



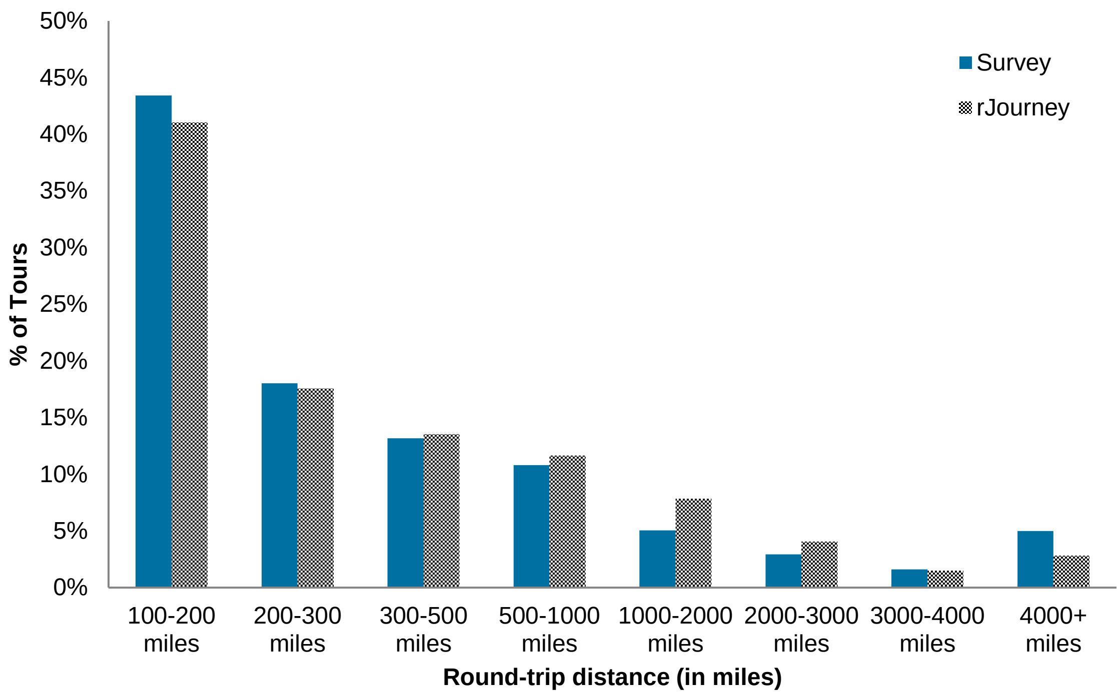
Source: FHWA

Figure 63. Round-trip distance-band distribution, by purpose—leisure.



Source: FHWA

Figure 64. Round-trip distance-band distribution, by purpose—commute.



Source: FHWA

Figure 65. Round-trip distance-band distribution, by purpose—employer's business.

Table 26. Average person-miles traveled.

Tour Purpose	Survey	rJourney	Difference	% Difference
Personal Business	396.48	441.01	44.53	11.2%
Visiting Friends and Relatives	464.70	578.36	113.66	24.5%
Leisure	478.25	531.75	53.5	11.2%
Commute	219.25	219.62	0.37	0.2%
Employer's Business	673.02	641.17	-31.85	-4.7%

4.4 Mode Choice Models

The tour mode choice model for each purpose is structured as a multinomial logit model with the following mode choices:

1. **Auto**: Available for all O-D destination combinations that are 50+ miles apart, except:

- From/to destinations within contiguous United States to/from destinations within Alaska and Hawai'i; and
- From/to destinations within Alaska to/from destinations within Hawai'i.

2. **Bus**: Available for all O-D destination combinations that are 50+ miles apart and are connected to bus network.

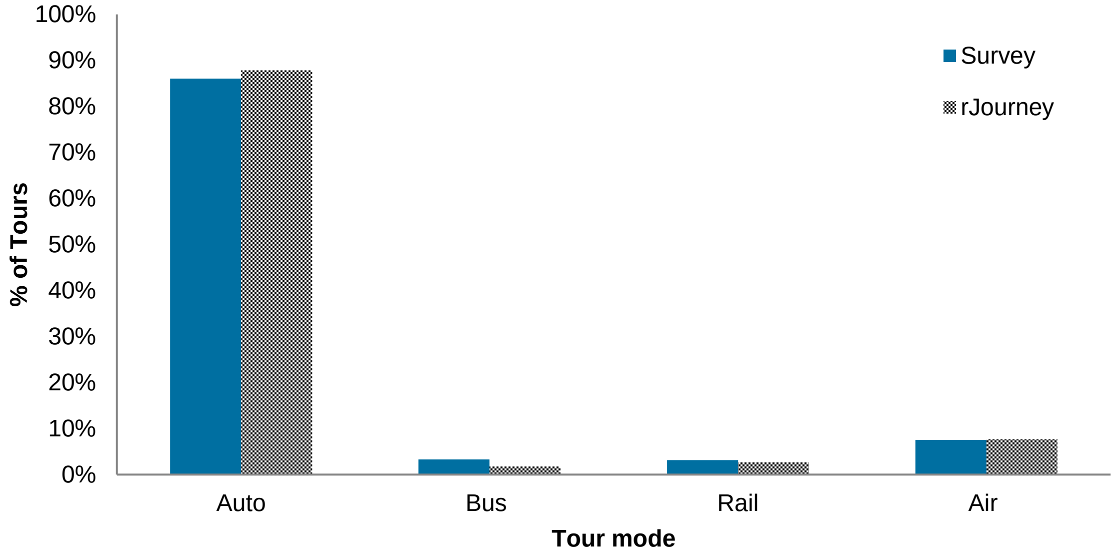
3. **Rail**: Available for all O-D destination combinations that are 50+ miles apart and are connected to rail network.

4. **Air**: Available for all O-D destination combinations that are 50+ miles apart and are connected to the air network.

The reader is referred to the Section 2.2 for further details on the development of the bus, rail, and air networks.

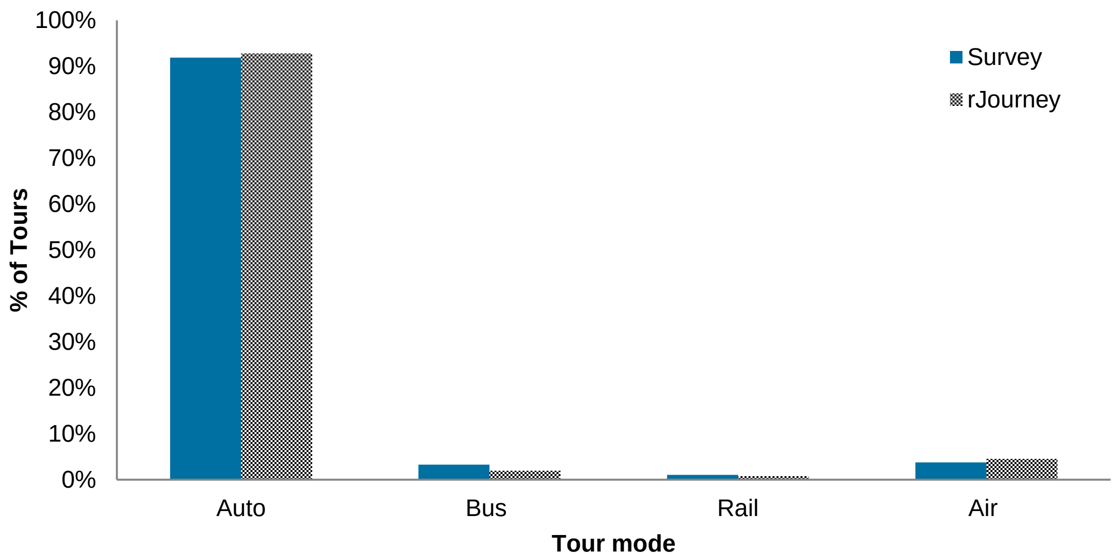
Several household, person, tour-level, and destination-related attributes were found to have significant effects on tour mode choices. The calibration task was undertaken by adjusting mode-specific constants. Similar to destination choice models, mode choice models were calibrated for each purpose.

Figure 66 to Figure 71 present the calibrated mode choice model results. Specifically, Figure 66 shows overall distribution of tour mode share for all purposes and Figure 67 to Figure 71 present tour mode share distribution for personal business, visiting friends and relatives, leisure, commute, and employer's business tour purpose, respectively. Regardless of tour purposes, the calibrated mode shares match observed mode shares reasonably well with a difference within 4 percent. Auto is the predominant mode for long-distance tours and has an overall mode share of 88 percent. Personal business tours have the highest auto share (92.8 percent) and employer's business tours have the lowest auto share (82.1 percent). The second most frequently used mode is air, with an overall share of about 8 percent. Air share is the highest for employer's business (14.6 percent) and the lowest for commute (0.9 percent). Compared to auto and air, bus and rail have relatively small mode shares, in most cases ranging from less than 1 percent to a little over 2 percent (exceptions are bus and rail shares for commute tours, these shares are 3.1 percent and 12 percent, respectively).



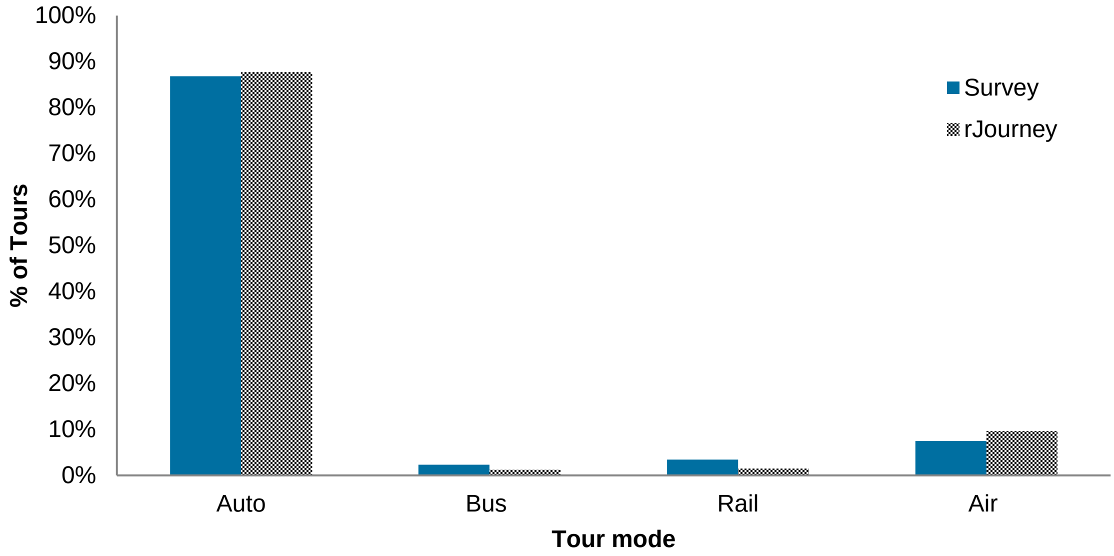
Source: FHWA

Figure 66. Overall tour mode share.



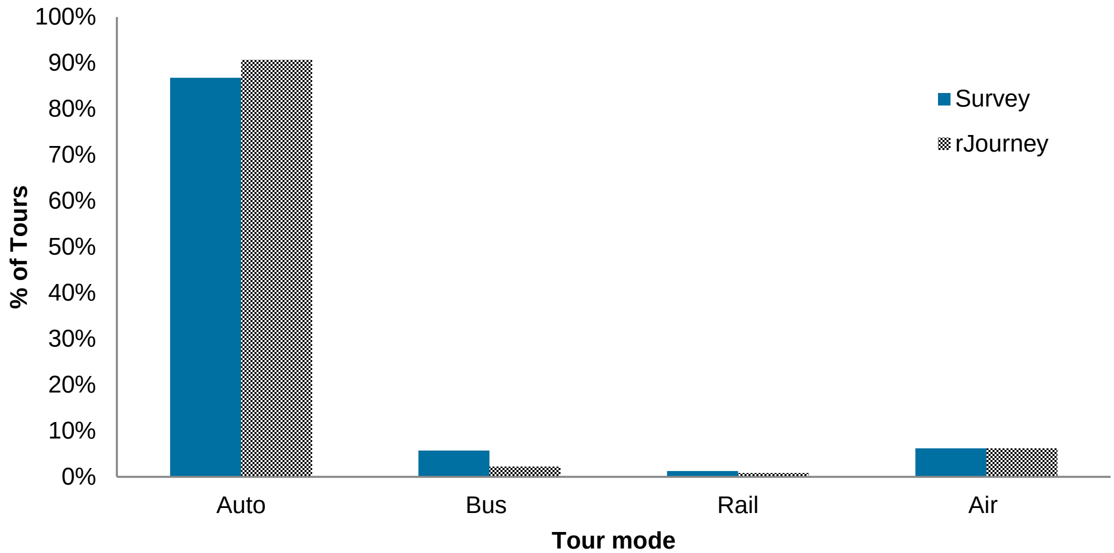
Source: FHWA

Figure 67. Tour mode share, by purpose—personal business.



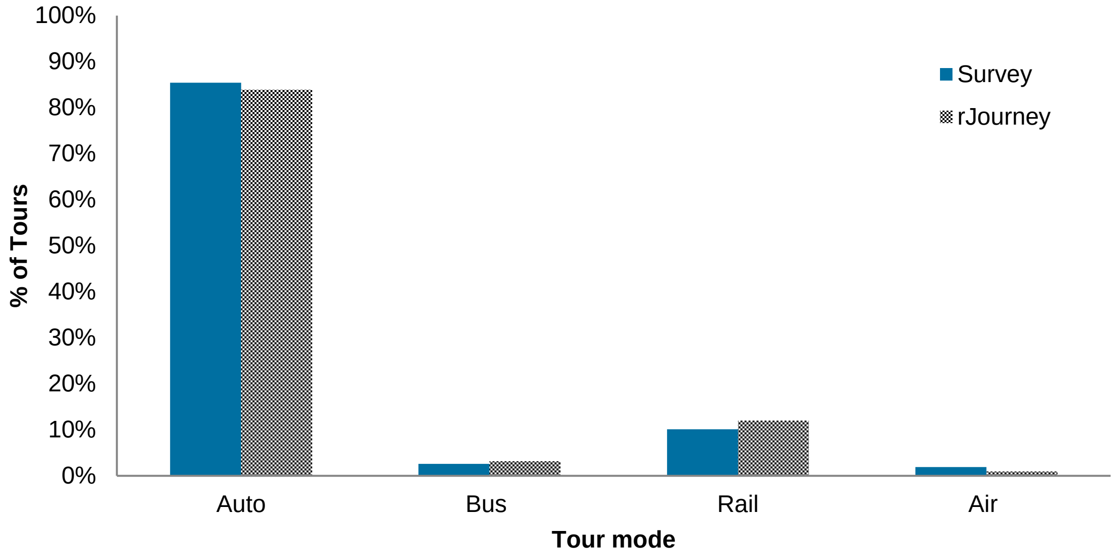
Source: FHWA

Figure 68. Tour mode share, by purpose—visiting friends and relatives.



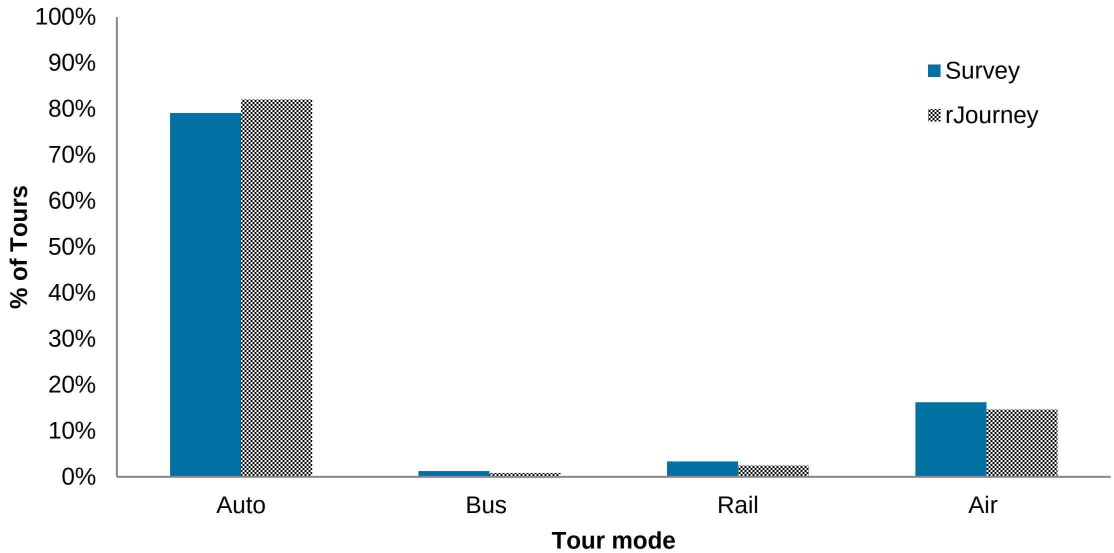
Source: FHWA

Figure 69. Tour mode share, by purpose—leisure.



Source: FHWA

Figure 70. Tour mode share, by purpose—commute.



Source: FHWA

Figure 71. Tour mode share, by purpose—employer's business.

4.5 Preparation of Average Daily Long-Distance Trip Tables

The final outputs generated by rJourney include a household file (includes household level information), a tour file (includes tour-level information), and trip matrices by mode. The trip matrices contain average daily long-distance trips and are derived from the tour file as follows:

- First, tours are converted to half-tours/trips using tour O-D zones. Information on mode, tour party size, distance, and expansion factors are extracted from each tour and are appended to the corresponding trip records.
- Second, expansion factors are applied to obtain an expanded trip record file. The file includes all the trips undertaken over one year. The trip records are divided by a factor of 365 to convert the annual vehicle-trip table to an average daily vehicle-trip table. Mode information is used to separate the trips into trip tables for auto, bus, rail, and air mode.
- Third, for person trip tables, the trip records are multiplied by tour party size to convert vehicle-trip tables to person trip tables.

CHAPTER 5. HIGHWAY ASSIGNMENT

5.1 Overview of Highway Network

Highway assignment was completed in TransCAD. The NHPN was the main source of the TransCAD network. 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 was downloaded from the FHWA's website. To build highway skims for the NUMA-level zonal system, centroid connecters were added to the NHPN network as additional links. Centroid connecters are not allowed to directly link to interstate facilities, since travelers have to access interstate facilities through other roads. The final highway network contains 198,634 links. TransCAD assigns long-distance and background traffic to this network to produce planning-level estimates of traffic volumes.

The key variables for building highway skims are speed and capacity. While speeds and capacities vary from facility to facility, the project team developed these based on the functional class of the highway links; this was due to a lack of facility-specific data. Table 27 and Table 28 are the lookup tables for the speed and capacity assumption.

Table 27. Urban roads' speed and capacity, by functional class.

Functional Classification	Urban	Posted Speed	Free-Flow Speed	Hourly Capacity Per Lane
Interstate	11	65	71.50	1,900
Other Freeway/Expressway	12	55	60.50	1,700
Principal Arterial	14	45	47.25	1,200
Minor Arterial	16	35	36.75	1,000
Collector	17	30	31.50	900
Local	19	25	26.25	600

Table 28. Rural roads' speed and capacity, by functional class.

Functional Classification	Rural	Posted Speed	Free-Flow Speed	Hourly Capacity Per Lane
Interstate	1	70	73.50	2,000
Other Freeway/Expressway	2	60	63.00	1,800
Principal Arterial	6	50	52.50	1,400
Minor Arterial	7	45	47.25	1,200
Collector	8	40	42.00	1,000
Local	9	35	36.75	700

Centroid connectors also need speed and capacity constraints. The project team assumed that the speed on centroid connectors was the same as that for local roads. However, we set their capacities at an arbitrarily high level (999,999) because all demands must flow through the centroid connectors.

A free-flow travel time highway skim was built for the NUMA zones. It is a 4486*4486 matrix—some NUMAs in Hawai'i and Puerto Rico were not directly connected to the continental United States.

5.2 Estimation of Background Traffic

Long-distance trips are a small portion of the total demand on the national highway network. To obtain better assignment results, one should estimate the other trips taking up capacity on the road system so that congestion is adequately represented. These other trips include short-distance passenger trips and truck trips. At the link level, the total traffic is defined in Figure 72.

$$Total\ Volume = Truck\ Volume + Long - Distance\ Passenger\ Volume + Short - distance\ Passenger\ Volume$$

Figure 72. Equation. Defining total traffic.

The original NHPN, while containing AADT data, does not have truck AADT. The Freight Analysis Framework (FAF) network is useful for this purpose. FAF estimates commodity movements by truck and weight for truck-only, long-distance moves over specific highways. It is also available from the FHWA website.[1] The greatest advantage of the FAF network was that it was also based on NHPN, which makes it relatively easy to correlate the average annual daily truck traffic with the highway links. A total of 176,231 matches were found in the FAF network. The links in Figure 73 represent those with FAF traffic counts.

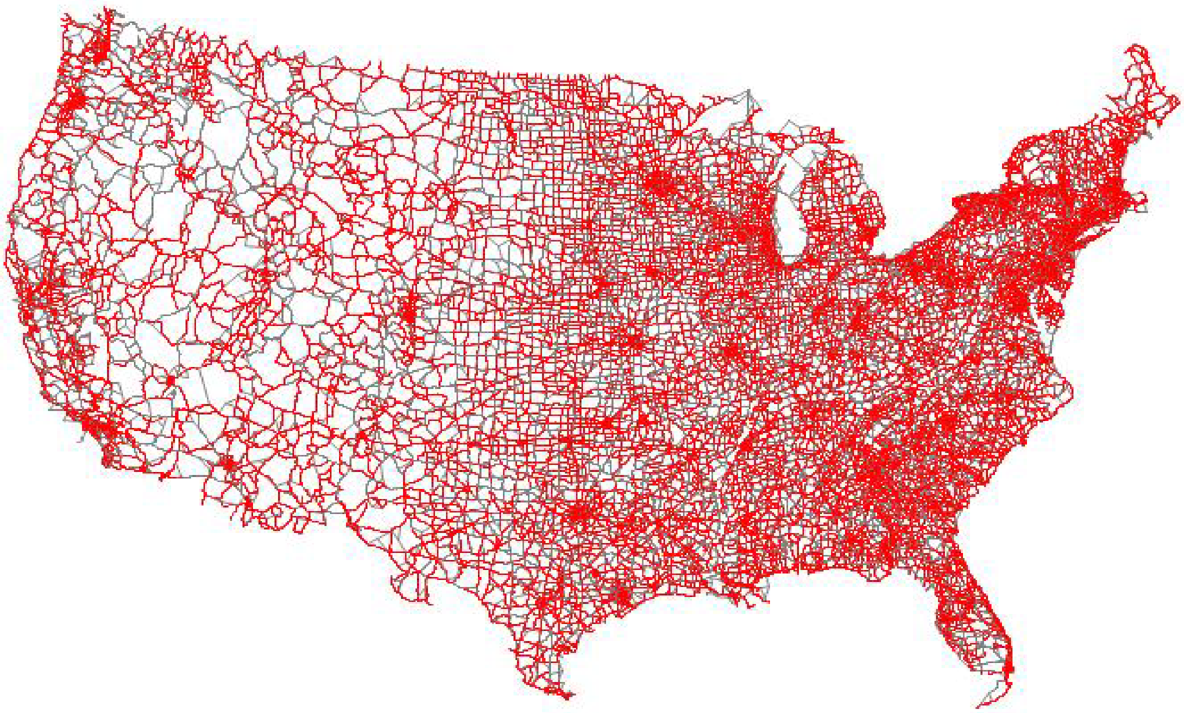
To estimate the background trip table, the long-distance passenger trip table was assigned with the truck trip table using the stochastic method and subtracted from total volumes to produce an estimate of short-distance passenger volumes. These volumes were used in combination with origin-destination matrix estimation (ODME) methods to produce a short-distance passenger trip table. The short-distance passenger trips, added to the truck trips, produced a "background" trip table.

This initial estimation of background trips did not produce a reasonable estimate of total volumes, because the "seed" matrix for the ODME process was not reasonable. The seed matrix is for initial assignment purposes and could take various values—as simple as a matrix of all ones. A more theoretically sound approach, which has been applied by the project team, was

---

[1] Please visit the Office of Operations, FAF[3] Network Database and Flow Assignment: 2007 and 2040 website: https://ops.fhwa.dot.gov/freight/freight_analysis/faf/faf3/netwkdbflow/index.htm.

generating a seed matrix using the quick response methods (QRM) for passenger travel. This method assumes trip rates (per household) for three purposes: home-based work (HBW), home-based non-work (HBNW), and non-home-based (NHB). The QRM approach uses a cross-classification table, segmented by the size of the urban area, household income, and auto ownership. For each purpose, separated trip production and trip attraction rates were applied, and a final trip table was created by balancing both. A total QRM matrix was created by combining all three purposes.



Source: FHWA

Figure 73. Highway network with FAF truck traffic data.

Since the background travel was focused only on short-distance travel, trips between any O-D pairs with greater than 50 miles of distance were eliminated from the QRM matrix. The final seed matrix contains 88,306 O-D pairs, as shown in Table 29.

Table 29. Statistics of the QRM seed matrices.

Matrix	Count	Mean	Std.	Pct. Diag.	Min.	Max
HBW	19731666	15.6	1019.0	45.6	0	2279468
HBNW	19731666	27.9	1994.5	55.7	0	4084150
NHB	19731666	11.9	836.4	53.6	0	1740189
Total	19731666	55.4	3835.0	52.4	0	8103807
Less50	88306	12289.5	55998.3	52.8	1.13	8103808

QRM also produces intrazonal trips. Although these trips were never assigned to the network, a uniform 10 minutes of travel time is added to the diagonal cells of the skim to avoid invalid computational errors. The background traffic (a zonal trip matrix and link volumes) was successfully estimated using TransCAD's ODME process to assign the QRM seed matrix onto the network.

5.3 Highway Assignment Parameters

Background trip and long-distance trip matrices produced are assigned to the NHPN. Background trips are assigned first using a biconjugate Frank-Wolfe method. The biconjugate Frank-Wolfe method is a user equilibrium assignment, which is an iterative process to achieve a convergent solution where route changes would not improve individual users' travel times. The traditional Bureau of Public Roads (BPR) volume-delay function is used to determine the change in travel as congestion occurs (see Figure 75). This equation (Figure 74) relates link travel times as a function of the volume/capacity ratio. The alpha and beta defined in the standard BPR function are globally assumed to be equal to their traditional values in rJourney. The background trip assignment is run with a relative gap of 0.003, with a maximum of 200 iterations.

$$t=t_f\,[1+\alpha\,(v/c)^{\beta}]$$

Figure 74. Equation. Volume-delay function.

Where	t	=	congested link travel time
	t_f	=	link free-flow travel time
	v	=	link volume
	c	=	link capacity
	α,β	=	0.15, 4.0

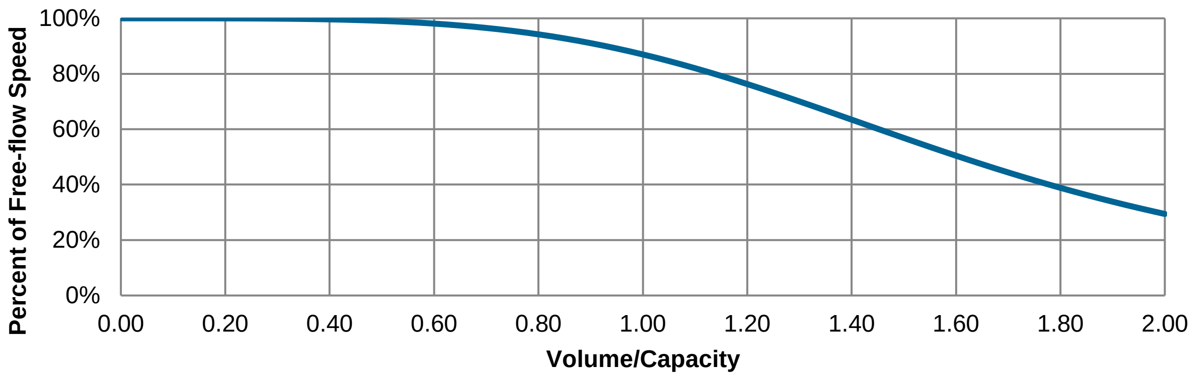


Figure 75. Volume-delay curve.

Source: FHWA

The resulting user equilibrium travel times from the background trips are applied to the network to provide congested travel times for long-distance trips. Due to the limited detail of the national network and the desire to utilize alternative routes, long-distance trips are assigned to the network using a stochastic assignment. A stochastic assignment distributes trips between multiple alternative paths that connect O-D pairs. The proportion of trips assigned to a path equals the choice probability for that path, which is calculated by a simple logit route choice model. Generally, a path with a lower overall travel time will have a higher choice probability. Only "reasonable" paths are considered in a stochastic assignment, which does not necessitate assigning every alternative path. A path is determined "reasonable" if it takes the traveler farther away from the origin or closer to the destination. The stochastic error parameter is set at 40 and runs for 98 iterations.

5.4 Application in TransCAD

The rJourney assignment was implemented in TransCAD Version 6.0, a GIS-based travel demand modeling software, using the software's scripting language, GISDK (Geographic Information System Developer's Kit). TransCAD was chosen due to its ease of use and ability to handle large-scale traffic assignment algorithms within reasonable run times.

Some preprocessing is needed prior to assignment within TransCAD. While background trips were estimated in TransCAD, conversion was needed to bring the long-distance trip table into TransCAD's matrix (.mtx) format. Long-distance tabular data was converted into a comma-separated values (CSV) file. Once processed, the CSV file was imported and converted using TransCAD import tools so that long-distance trips were in an appropriate O-D format for the national network.

A single GISDK script was created to complete the assignment approach detailed in Section 3.2. The process was broken into four parts, outlined in Figure 76. This includes the creation of the TransCAD highway network file (.net), the biconjugate Frank-Wolfe assignment of the background trips, updates to network attributes, and the stochastic assignment of long-distance trips.

Source: FHWA

Figure 76. Application in TransCAD process.

CHAPTER 6. MODEL VALIDATION

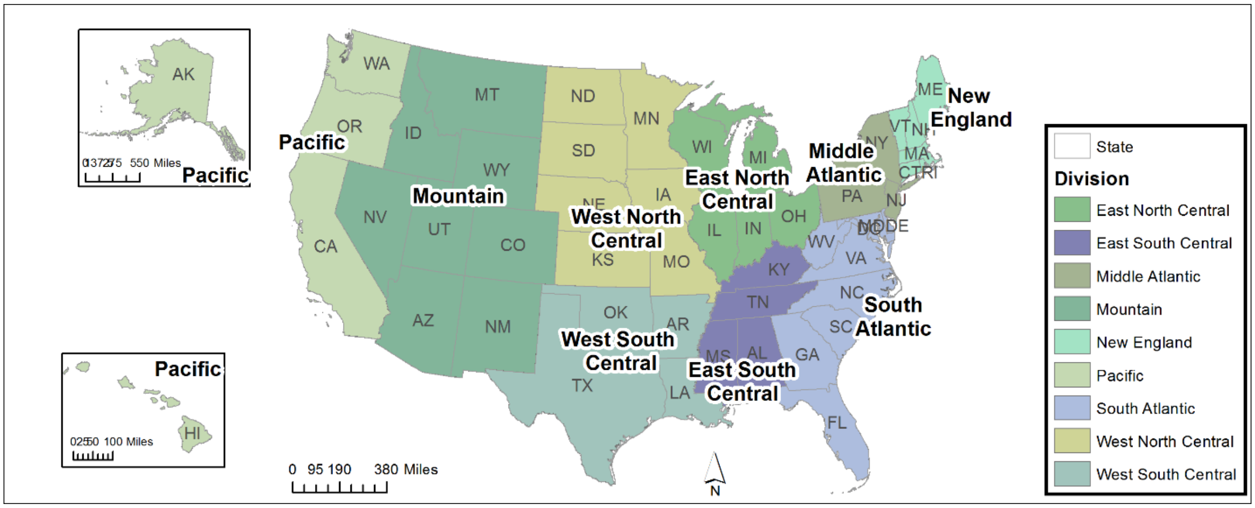
6.1 Trip Tables by Mode

As part of model validation, the research team compared model estimated trip tables by mode with mode-specific trip tables obtained from the following sources (see Section 2.5 for more detail):

- **2008 National O-D Trip Tables.** These are 2008 county-to-county person trip tables for auto, bus, rail, and air. The tables include trips that are 100+ miles in length. The trip tables were developed as part of FHWA's TAF Multimodal Interregional Passenger Travel Origin-Destination Data project.
- **2011 Intercity Bus Ridership Table.** This is a 2014 Core Based Statistical Area-to-Core Based Statistical Areas bus trip table for the top 200 markets. The 2014 bus ridership table was factored down to the 2011 level.

The 2008 national O-D tables and the 2014 Intercity Bus Ridership Table are not observed data and so are not used as conclusive sources for validation. The 2014 Intercity Bus Ridership Table also does not provide any information on the overall market share captured by the top 200 markets. Therefore, it is not feasible to treat these tables as benchmark values and use them for model validation. Rather, the research team compared the model estimated trip tables with the 2011 national O-D tables and the 2011 intercity bus ridership table to obtain a general overview on the performance of the model. For this, the trip tables were summarized by nine census regions[1] shown in Figure 77. The results are presented in Table 30 and Table 31. Overall, the model estimated auto and air trip tables align relatively well with national O-D tables. The variation is more pronounced for bus and rail modes. Relative to national O-D tables, the model underpredicts total daily bus trips and over-predicts total daily rail trips by approximately 25 percent. When the model-predicted bus ridership values are compared with the 2011 intercity bus ridership table, the over-prediction rate is 60 percent. Divergence between rJourney values and intercity bus ridership values may be because the spatial resolution and other information available on the definition of the top 200 markets were not detailed enough to enable a selection of the same bus markets from the model.

[1] The U.S. Census Bureau refers to these regions as Divisions, with larger aggregations of these Divisions as Regions.



Source: U.S. Census

Figure 77. Regions in the U.S. census.

Another potential data source for the current research is the long-distance component of the 2001 NHTS. Table 32 summarizes average daily long-distance trips by mode from rJourney, 2011 national O-D tables, and 2001 NHTS. The difference between the number of auto trips from rJourney and the NHTS data may be attributable to the following:

a) For consistency with the values in the national O-D tables, the project team selected only auto trips with a length ≥ 100 miles from rJourney, while the NHTS data includes all trips with a length ≥ 50 miles.

b) rJourney predicted values correspond to the year 2011 while NHTS data correspond to the year 2001.

Compared to NHTS data, the model over-predicts the number of air trips by more than 90 percent. This is not surprising since there was a significant decline in air travel in 2001 after September 11, 2001. Table 33, which shows overall mode share, also captures this decline. In contrast, mode share from rJourney and national O-D tables show similar distribution.

Table 30. Average daily person-trips, by region and by mode (trip length ≥ 100 miles).

Region	rJourney (includes only trips with a length ≥ 100 miles)				National O-D table (2011)				The top 200 bus ridership markets (2011)	
	Auto	Bus	Rail	Air	Auto	Bus	Rail	Air	rJourney[29]	Ridership table
New England	307,492	6,920	10,706	51,684	272,101	8,822	8,120	57,088	4,052	6,926
Mid-Atlantic	860,904	22,533	37,781	125,675	618,977	28,305	25,508	132,363	17,493	20,329
East-North Central	1,300,657	28,841	12,352	130,370	955,474	26,317	9,034	135,017	21,092	5,044
West-North Central	640,750	10,175	2,282	60,416	621,961	10,765	1,602	63,244	2,394	324
South Atlantic	1,487,693	30,254	18,221	200,458	1,267,450	37,227	14,629	281,225	17,598	7,435
East-South Central	591,437	10,243	736	37,496	482,329	9,235	317	37,640	1,009	148
West-South Central	856,572	16,474	3,899	116,627	1,053,825	22,906	934	125,255	8,597	4,839
Mountain	481,558	6,772	1,934	97,360	710,182	17,318	1,069	163,833	2,559	1,641
Pacific	694,852	13,053	13,956	191,769	1,003,079	34,321	20,065	270,918	10,692	6,156
Total	7,221,915	145,264	101,868	1,011,855	6,985,379	195,216	81,278	1,266,582	85,486	52,842

[29] Information available on the definition of the top 200 bus ridership markets were not detailed enough to select the corresponding 200 markets from rJourney.

Table 31. Model estimates over trip table values ratio.

Region	Ratio: rJourney/National O-D table				Ratio: rJourney/Bus ridership table
	Auto	Bus	Rail	Air	
New England	1.13	0.78	1.32	0.91	0.59
Mid-Atlantic	1.39	0.80	1.48	0.95	0.86
East-North Central	1.36	1.10	1.37	0.97	4.18
West-North Central	1.03	0.95	1.42	0.96	7.38
South Atlantic	1.17	0.81	1.25	0.71	2.37
East-South Central	1.23	1.11	2.32	1.00	6.81
West-South Central	0.81	0.72	4.18	0.93	1.78
Mountain	0.68	0.39	1.81	0.59	1.56
Pacific	0.69	0.38	0.70	0.71	1.74
Overall Ratio	1.03	0.74	1.25	0.80	1.62

Table 32. Average daily long-distance trips, by mode.

Travel Mode	rJourney (includes only trips with a length ≥ 100 miles)	2011 National O-D tables (trip length ≥ 100 miles)	2001 NHTS (trip length ≥ 50 miles)[30]	Ratio	
				rJourney/ National O-D table	rJourney/ NHTS
Auto/Personal Vehicle	7,221,915	6,985,379	6,400,274	1.03	1.13
Bus	145,264	195,216	151,781	0.74	0.96
Train	101,868	81,278	57,808	1.25	1.76
Air	1,011,855	1,266,582	529,589	0.80	1.91
Other	--	--	15,890	--	--
Total	8,480,901	8,528,455	7,155,342	0.99	1.19

Table 33. Overall mode share.

Transportation Mode	rJourney (includes only trips with length ≥ 100 miles)	2011 National O-D tables (trip length ≥ 100 miles)	2001 NHTS (trip length ≥ 50 miles)
Auto/Personal Vehicle	85.2%	81.9%	89.4%
Bus	1.7%	2.3%	2.1%
Train	1.2%	1.0%	0.8%
Air	11.9%	14.9%	7.4%
Other	--	--	0.2%
Total	100.0%	100.0%	100.0%

6.2 Highway Performance

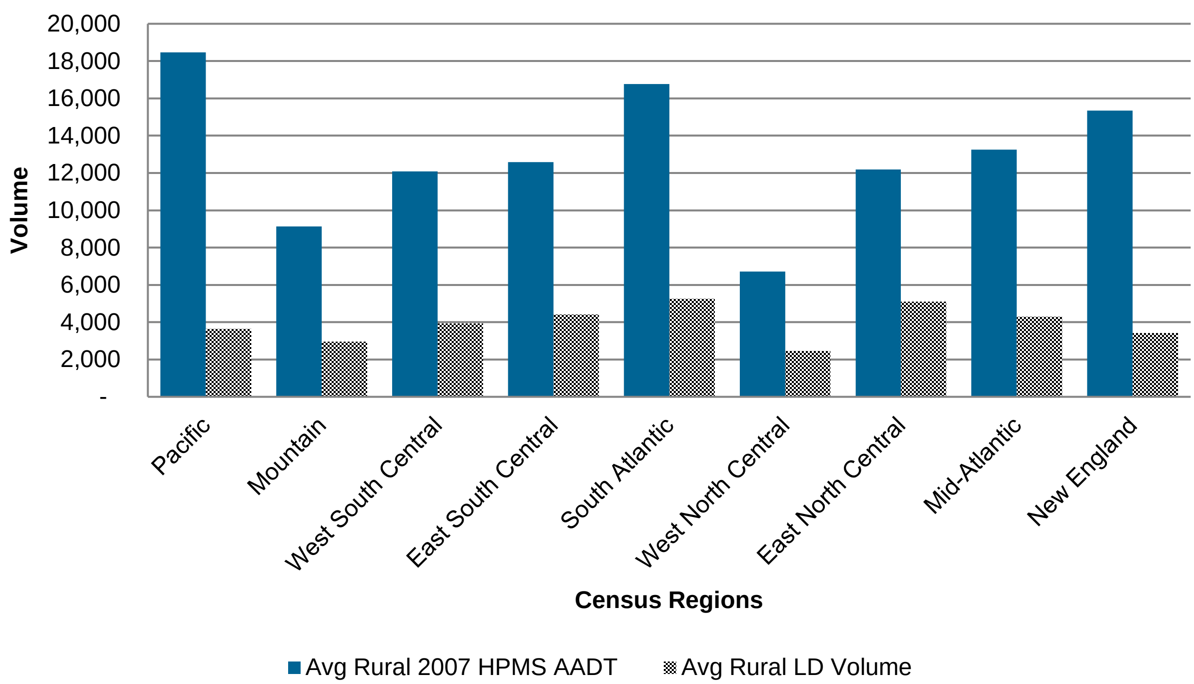
Highway validation of passenger long-distance trips was completed by studying rural functional classes at the census division level. The census divisions are nine subdivisions of the four census regions (Northeast, Midwest, South, West), which provide groupings of the United States and the District of Colombia (see Figure 77). Highway network validation is difficult at this national-level for several reasons. Of necessity, the model has limited spatial resolution. Short-distance trips or background traffic are treated in an extremely simplified fashion, and limited data were available for the calibration of the long-distance demand patterns. However, an effort was made to analyze long-distance passenger trips with national data currently available. For national traffic count data, the HPMS AADT for 2007 was used. For rural VMT data, the FHWA Highway Statistics 2013 manual was aggregated from the State level into census divisions. Table 34 presents the long-distance rural volumes and VMT from rJourney with the percent distributions of traffic counts and VMT counts from available sources.

[30] 2001 NHTS annual long-distance trips were divided by 365 to obtain daily long-distance trips.

Table 34. Highway model validation data, by region.

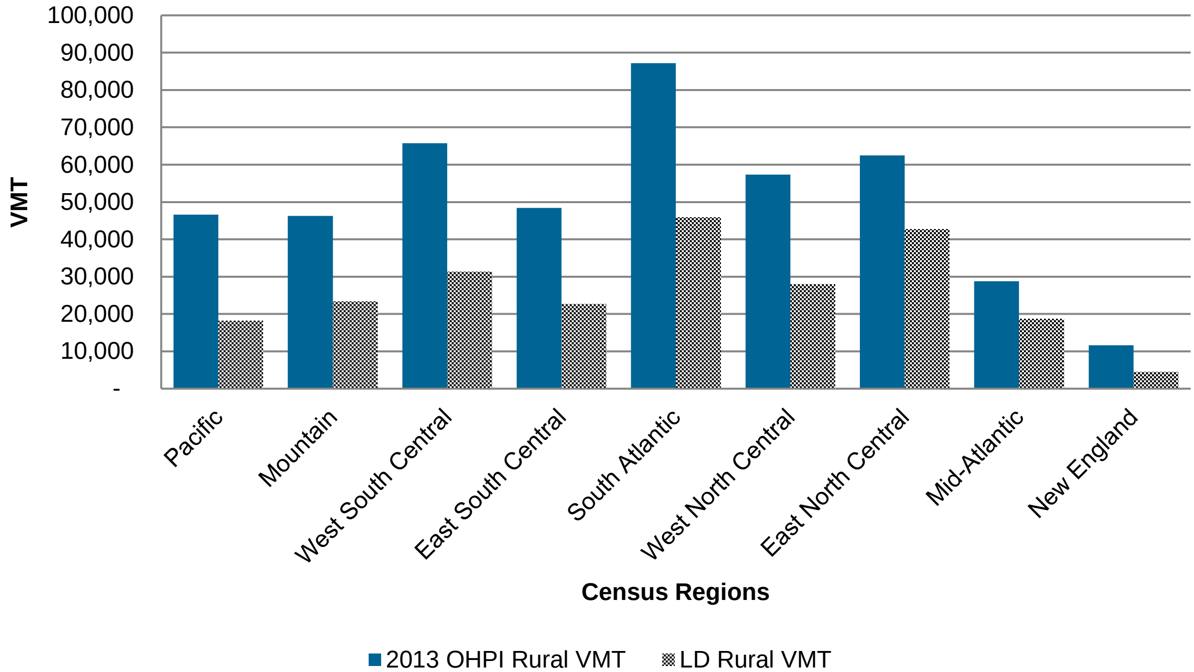
Region	rJourney Rural Avg. Volume	rJourney Rural Total VMT	2007 AADT	2013 OHPI VMT
Pacific	3,650	18,472	19.8%	39.0%
Mountain	2,966	9,142	32.4%	50.5%
West-South Central	3,928	12,091	32.5%	47.6%
East-South Central	4,424	12,581	35.2%	46.9%
South Atlantic	5,255	16,771	31.3%	52.7%
West-North Central	2,454	6,721	36.5%	48.8%
East-North Central	5,096	12,194	41.8%	68.4%
Mid-Atlantic	4,296	13,253	32.4%	65.2%
New England	3,427	15,352	22.3%	38.8%

Apart from the Pacific and New England regions, comparing these datasets illustrates that average long-distance passenger volumes are roughly 35 percent of the 2007 total traffic counts and 54 percent of the rural VMT (see Figure 78 and Figure 79). Looking closer at the Pacific and New England regions shows a decrease in both average count and VMT comparisons. This could be attributable to the small size and relatively fewer rural roadways of these regions.



Source: FHWA

Figure 78. Highway model validation volumes, by region.

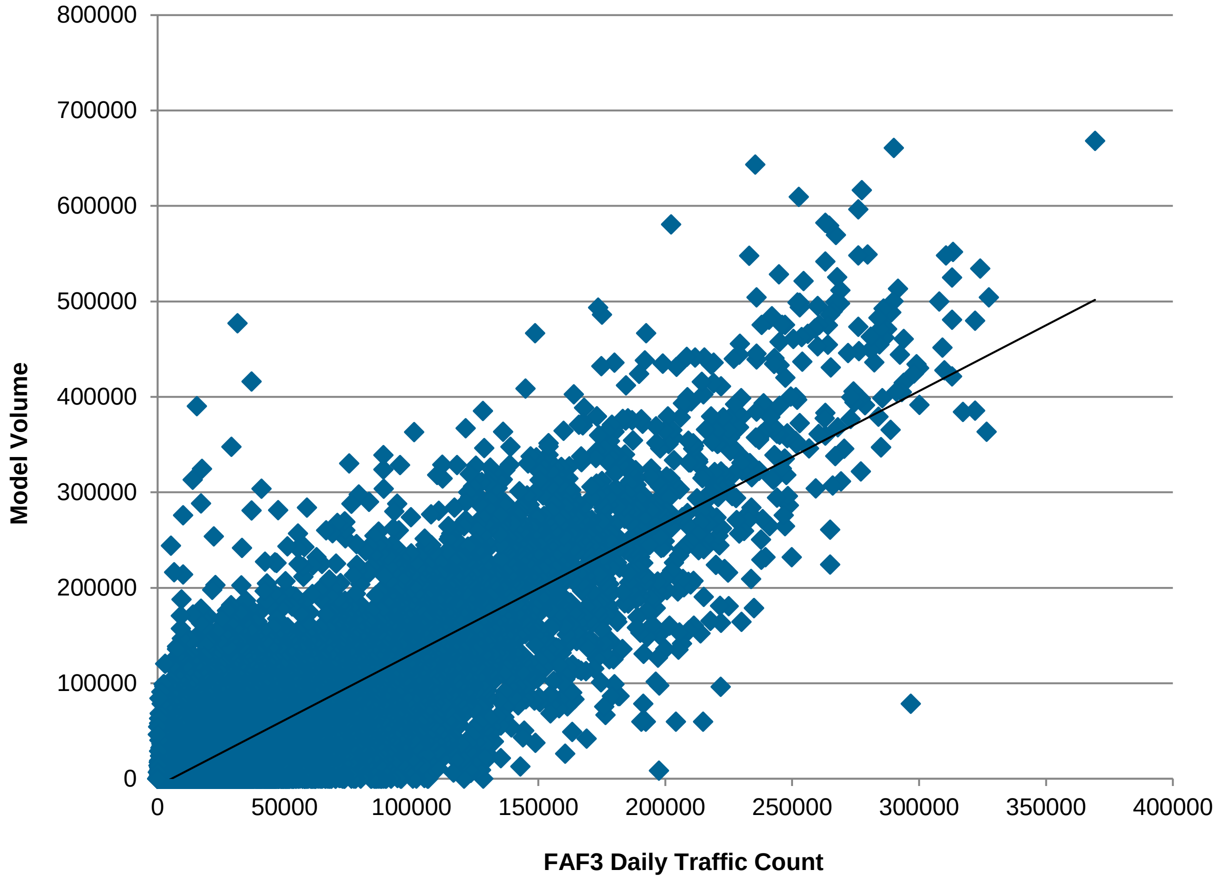


Source: FHWA

Figure 79. Highway model validation VMT, by region.

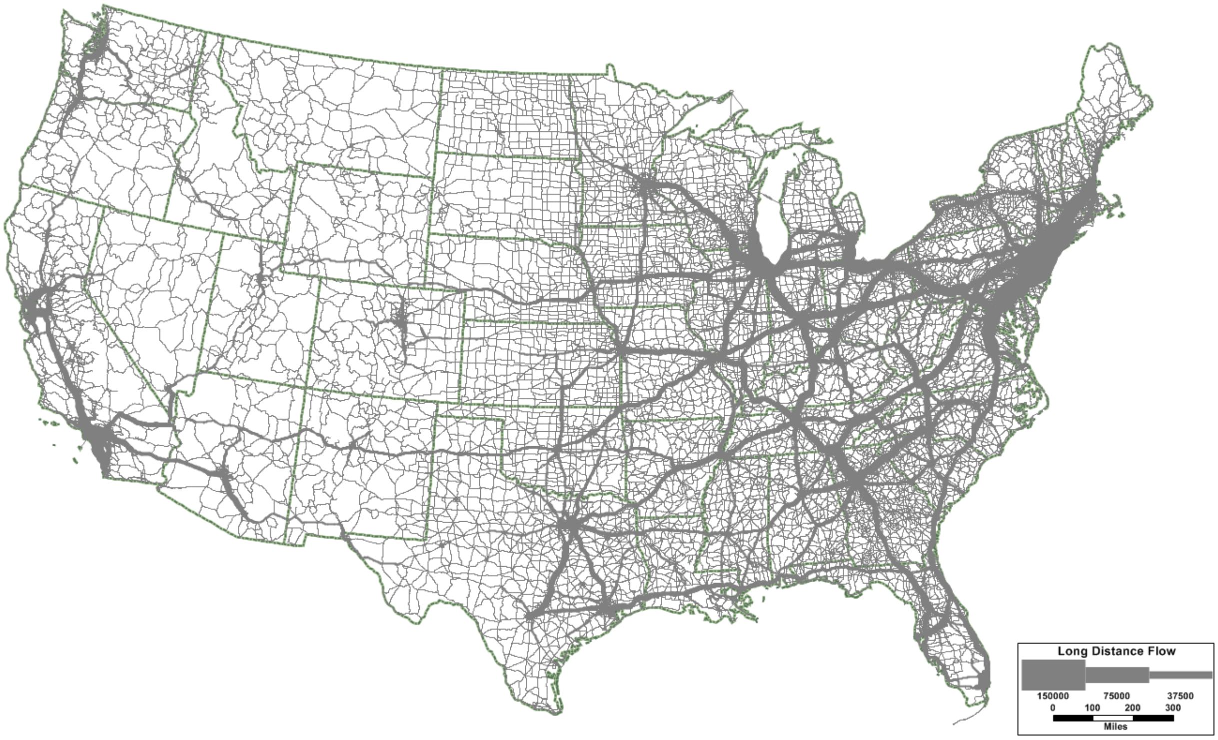
Improvement in assignment validation is possible with further investments. Network improvement is possible by addressing remaining connectivity issues, further adjusting centroid connectors, and improving assumptions regarding speeds and capacities. Improvements to handling short-distance trips or background traffic, and enhancements to the long-distance trip table estimates, are possible by incorporating additional data, including data from traffic counts or additional O-D data, if such data are available.

An overall view of the assignment of rJourney volumes on the national highway network confirms the reasonableness of the highway assignment (Figure 80 and Figure 81). These long-distance volumes are greater around metropolitan areas due to higher population concentrations; these volumes also represent smaller populations in rural areas who travel long distances.



Source: FHWA

Figure 80. rJourney total volumes, by count.



Source: FHWA

Figure 81. Long-distance rJourney volumes in the United States.

CHAPTER 7. PERFORMANCE METRICS

For the demonstration of the national long-distance passenger travel demand forecasting model, a sample of performance metrics helped demonstrate what types of data may be derived and how these may be interpreted for planning studies. The demonstration model was run initially to simulate travel for the month of October 2010; the result produced sample model results. The simulation model can also produce outputs for every month in the year, which permits aggregation 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.

7.1 Travel Metrics

Modes

Modal performance metrics support a wide variety of planning activities and are used to evaluate modal investments. The project team developed the travel metrics so that State, region, corridor, or zone summaries can be produced. These provide consistent evaluations of modal investments across the United States. Mode shares for person-tours and person-miles traveled are presented in Table 35. The auto mode has the highest mode share for both person-tours and person-miles traveled, but also tends to have more tours at shorter distances, 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 35. 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 36 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 trade-off 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 are 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 a bus tour.

Table 36. Average cost, travel time, and tours for October, by mode.

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 37 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 37. 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 38 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 trade-off with average travel time by mode, so air has the fastest travel times per tour.

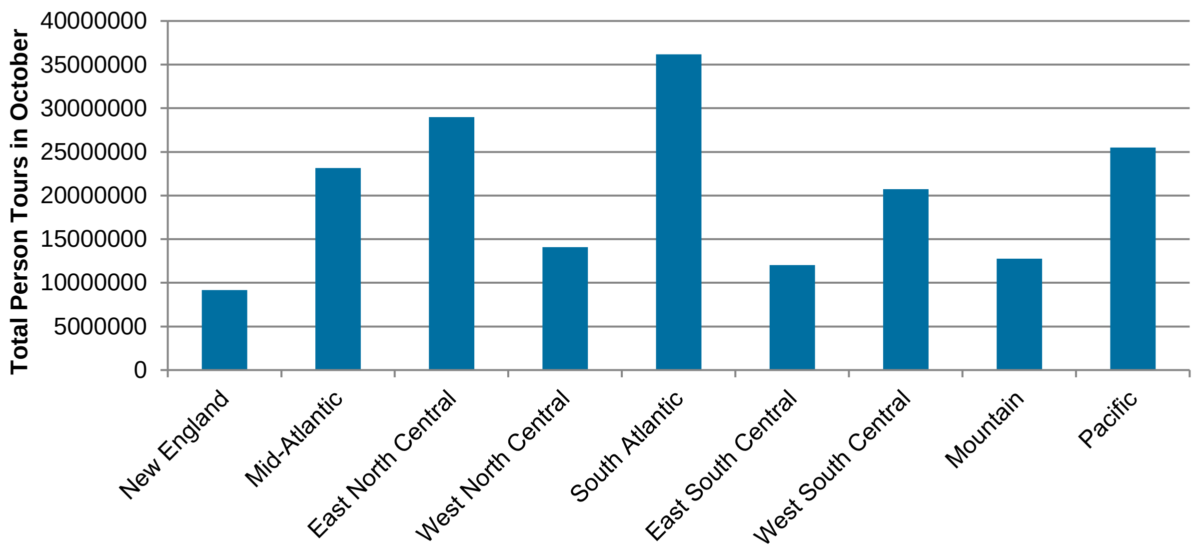
Table 38. Average cost, travel time, and tours for October, by purpose.

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 U.S. Census Bureau,[1] as shown in Figure 77. The simulation data output from the long-distance model is available to aggregate in many ways, so these regions are just one example of destination aggregation for reporting.

Figure 82 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.



Source: FHWA

Figure 82. Total person-tours in October, by region.

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

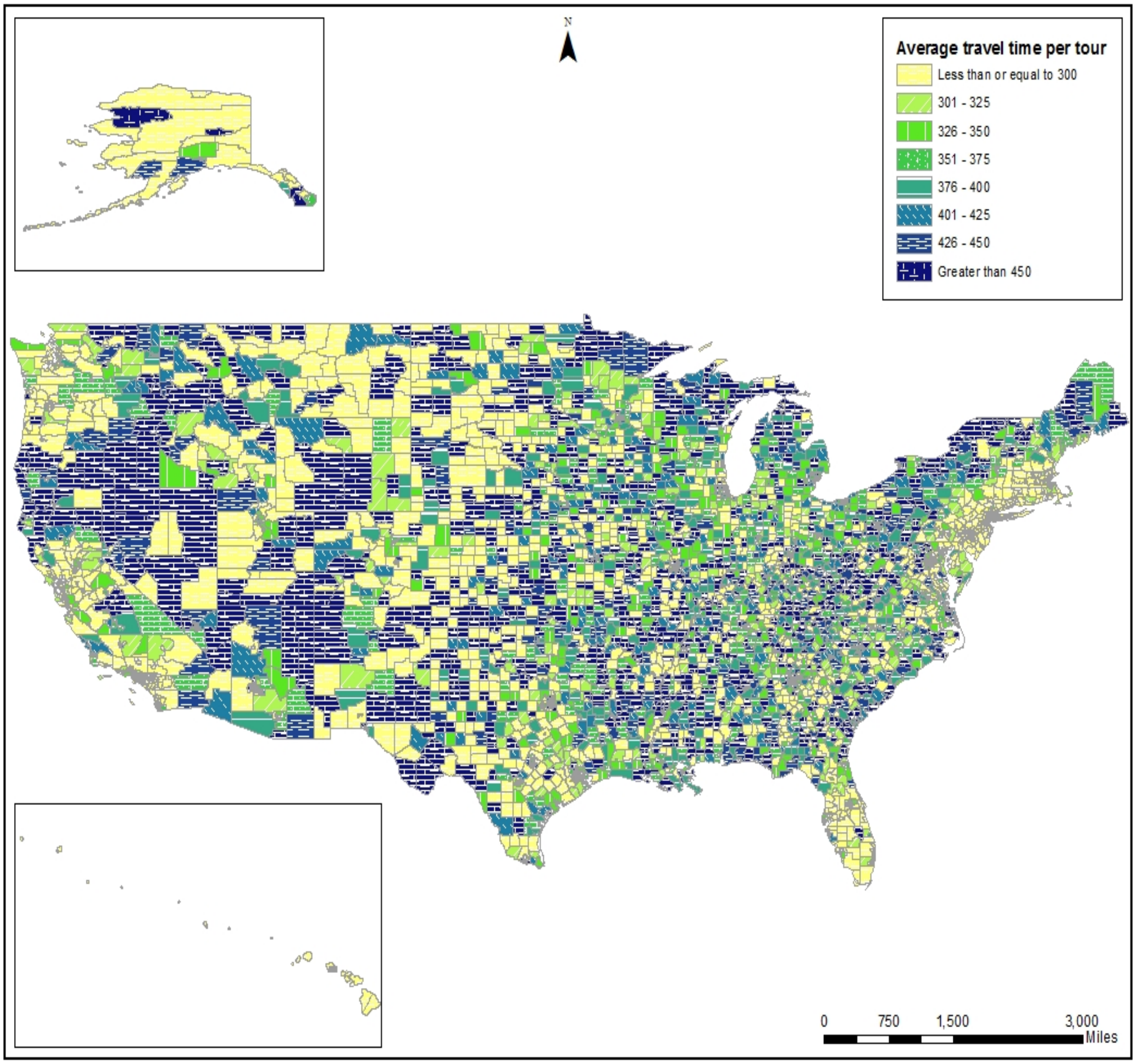
Table 39 presents an O-D matrix of person-tours in October to and from each region across the United States. This matrix demonstrates that most long-distance travel in the United States is within a single region, with the Pacific region retaining the highest percentage of long-distance travel (94 percent) and the East-South Central region retaining the least (50 percent).

Table 39. 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%

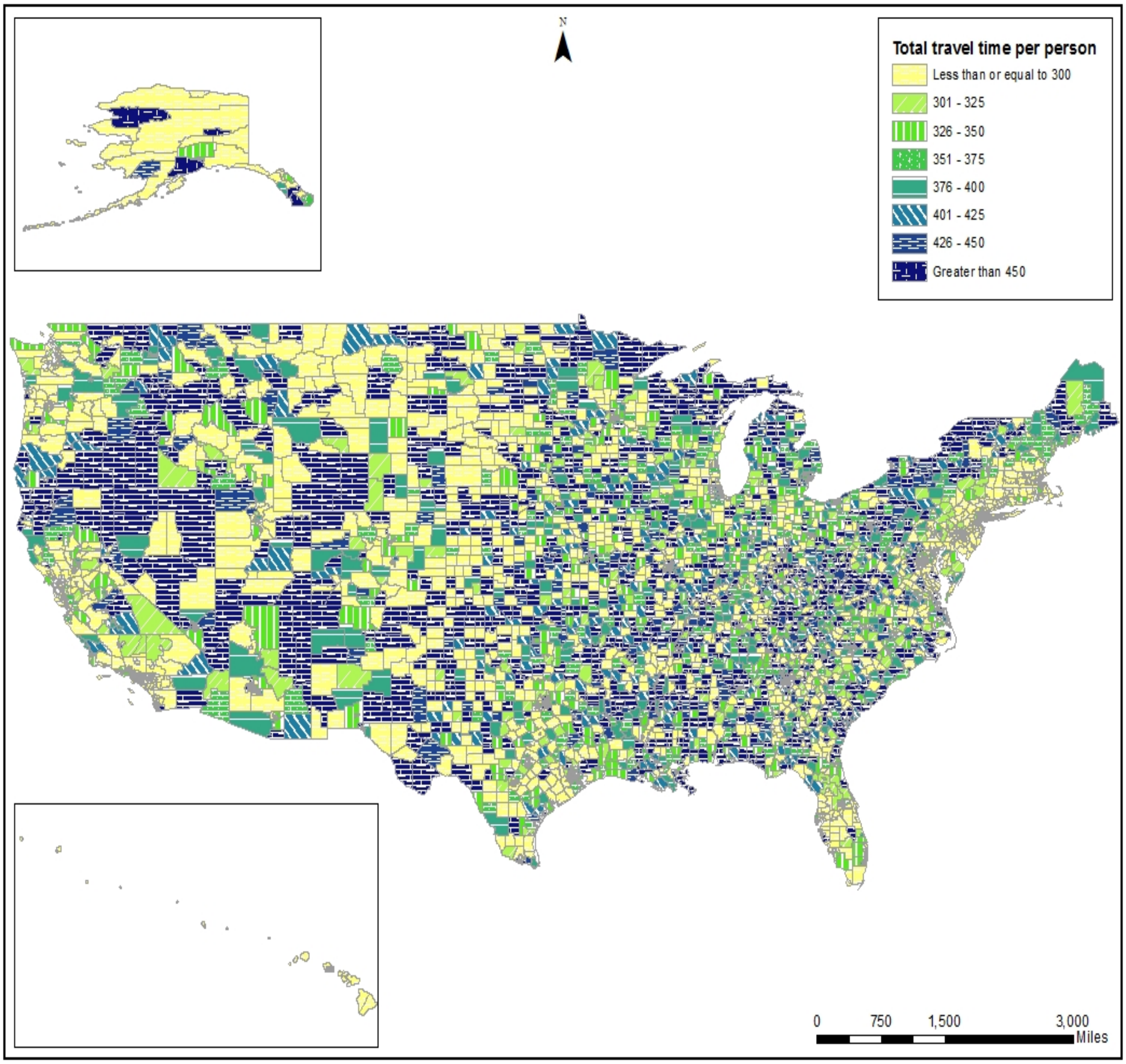
Travel Time

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 83 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 84, 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.



Source: FHWA

Figure 83. Travel time per tour, by origin NUMA.



Source: FHWA

Figure 84. Travel time per person, by origin NUMA.

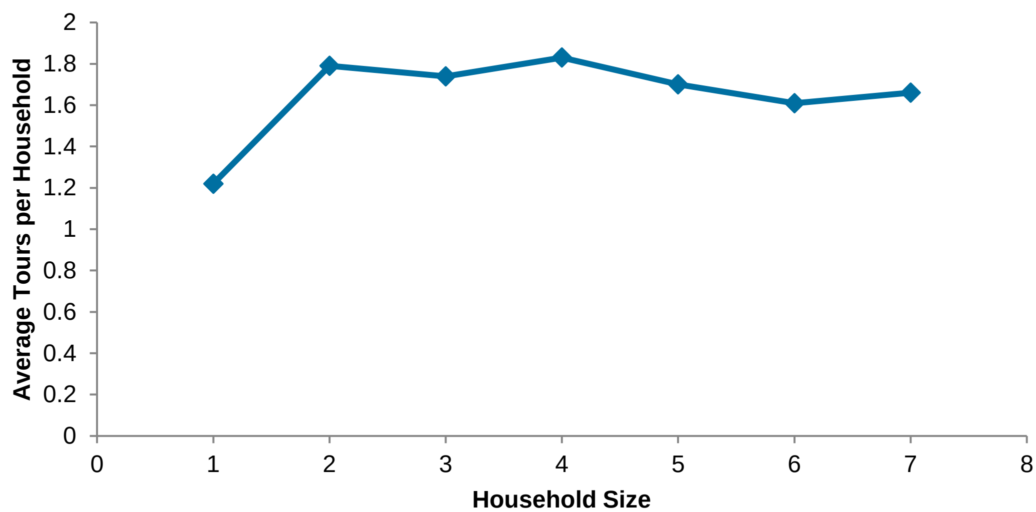
Table 40 presents the average travel time from region-to-region in October. Some O-D pairs do not have any person-tours represented and therefore have no travel times in this table (e.g., Pacific region to New England region). While there is some correlation between higher travel demand and lower travel 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 40. 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

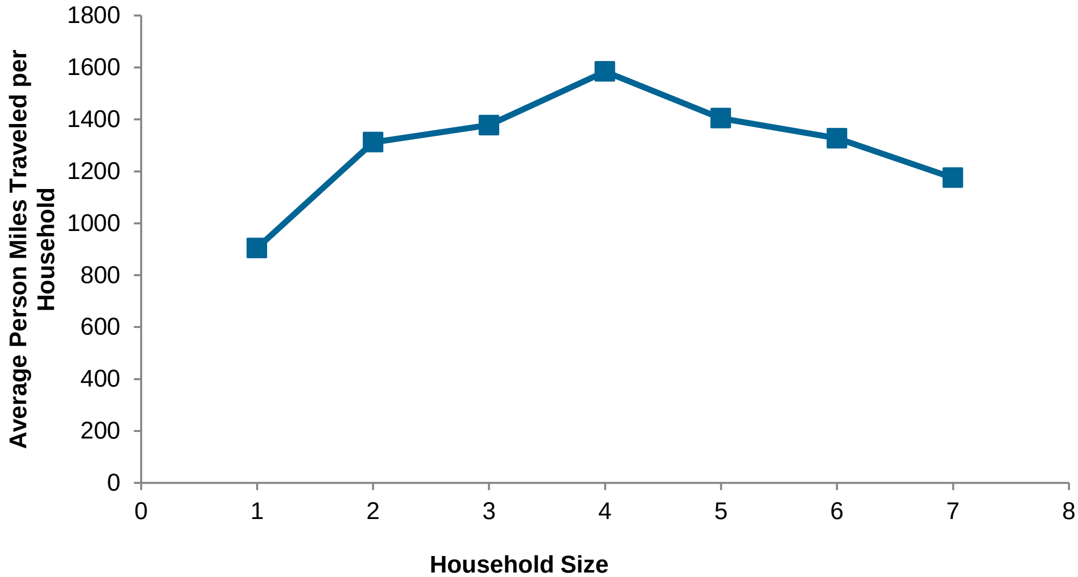
Demographics

Households that are larger or smaller in size tend to travel less in terms of overall travel and distance, as shown in Figure 85. The largest difference in travel metrics is seen in one-person households.



Source: FHWA

Figure 85. Long-distance travel metrics in October, by household size (average tours per household).



Source: FHWA

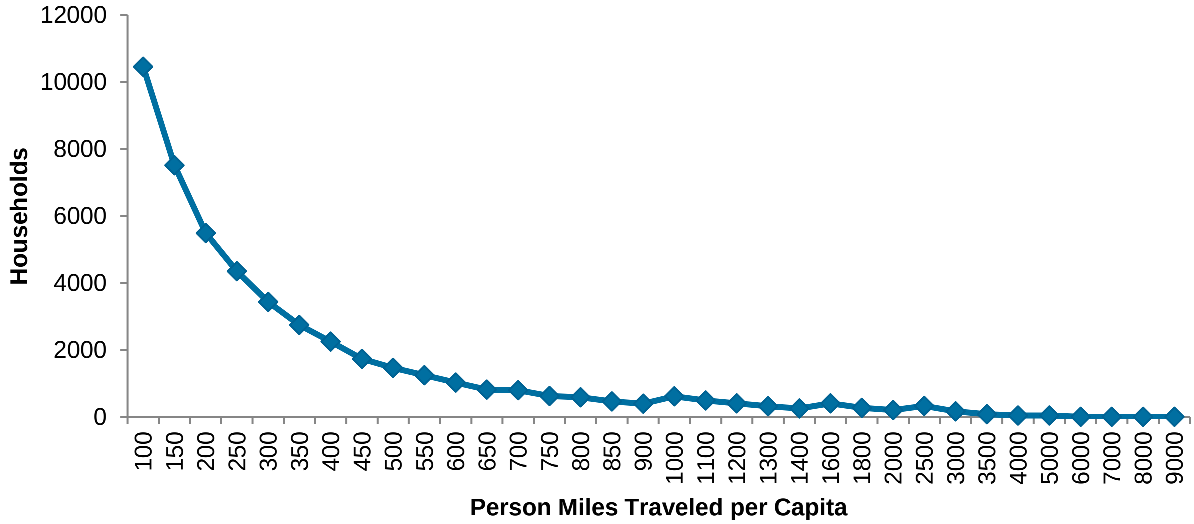
Figure 86. Long-distance travel metrics in October, by household size (average person-miles traveled per household).

7.2 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.

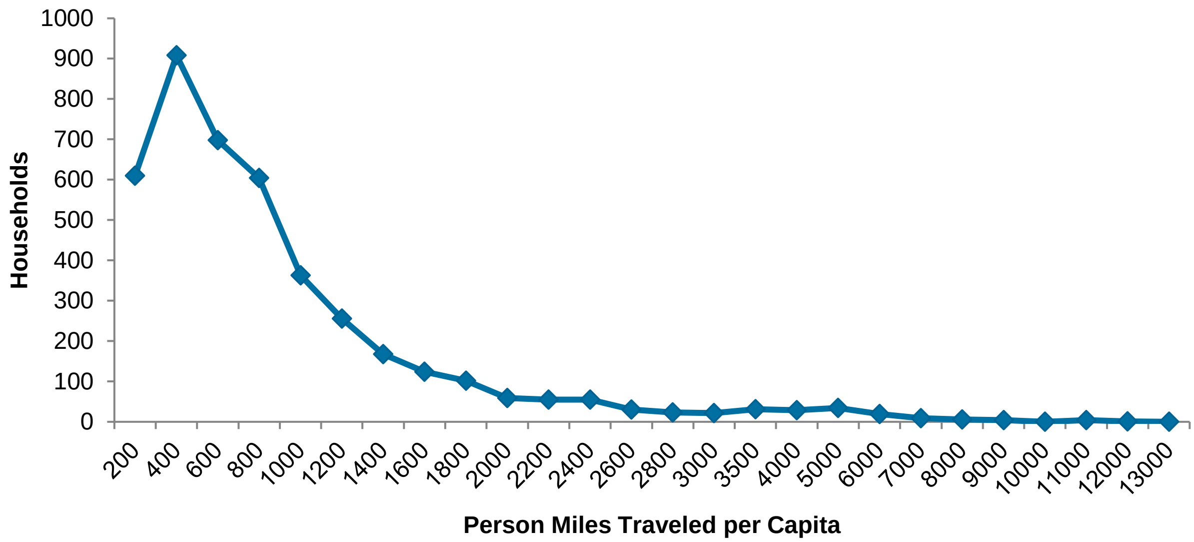
Distribution of Miles Traveled

One travel metric that provides insight into these additional metrics is the distribution of person-miles traveled by mode. Figure 87 through Figure 90 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 or multiple travelers making the same tour.



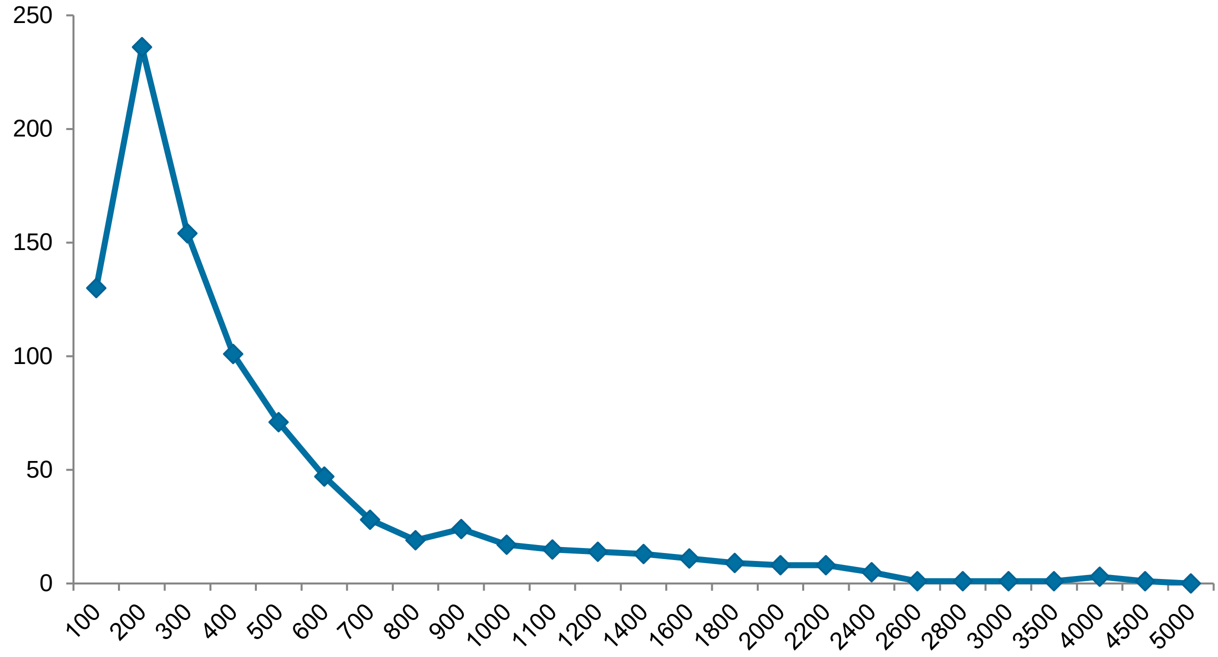
Source: FHWA

Figure 87. Distribution of person-miles traveled in October, by auto.



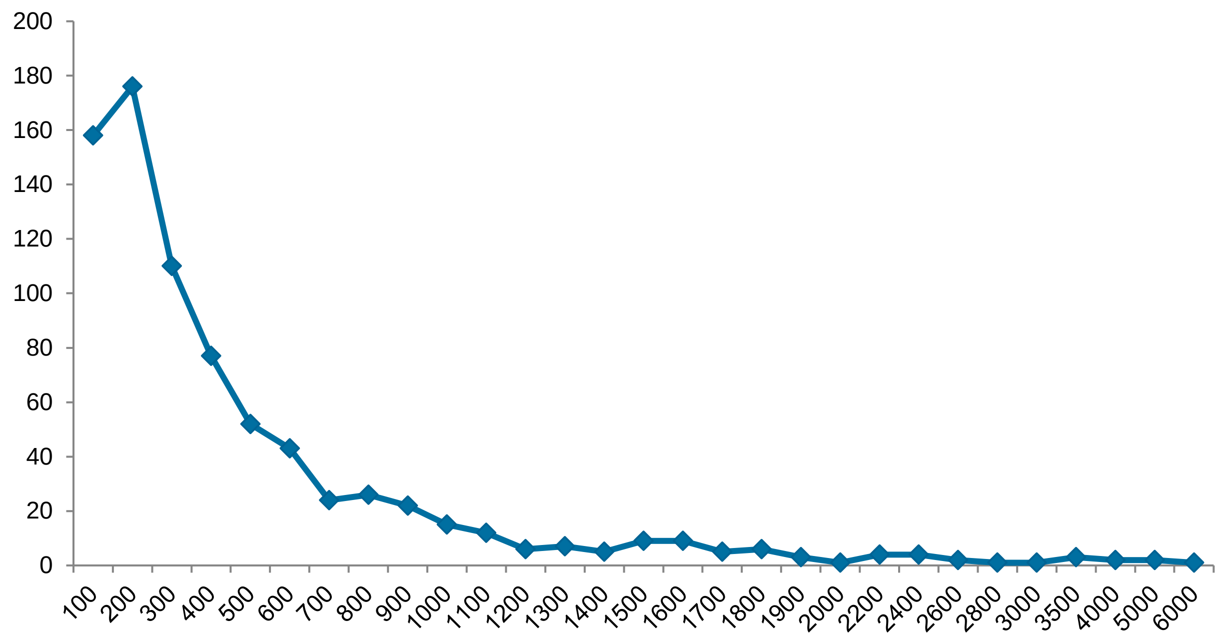
Source: FHWA

Figure 88. Distribution of person-miles traveled in October, by air.



Source: FHWA

Figure 89. Distribution of person-miles traveled in October, by rail.



Source: FHWA

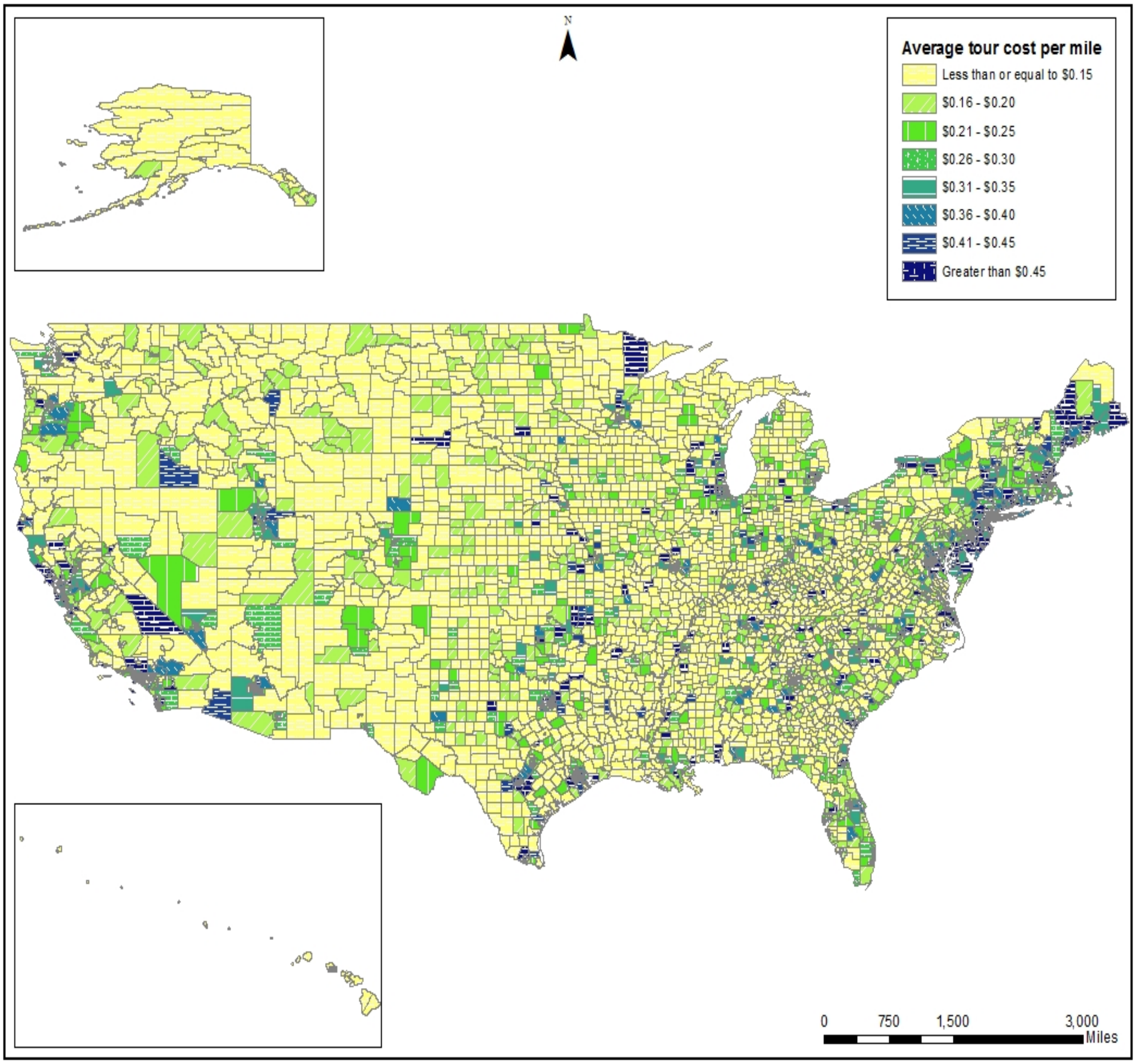
Figure 90. Distribution of person-miles traveled in October, by bus.

Tour party Size

More tours exist per household undertaken by single travelers, as shown in Figure 85, 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 91 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.



Source: FHWA

Figure 91. Average tour cost per mile, by origin NUMA.

7.3 Equity Metrics

The equity of public expenditures on transportation investments is an increasing concern for public agencies. Aspects of travel and household income correlate, so this is a useful metric to understand equity of a particular investment. Table 41 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 41. 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 8. SENSITIVITY TESTS

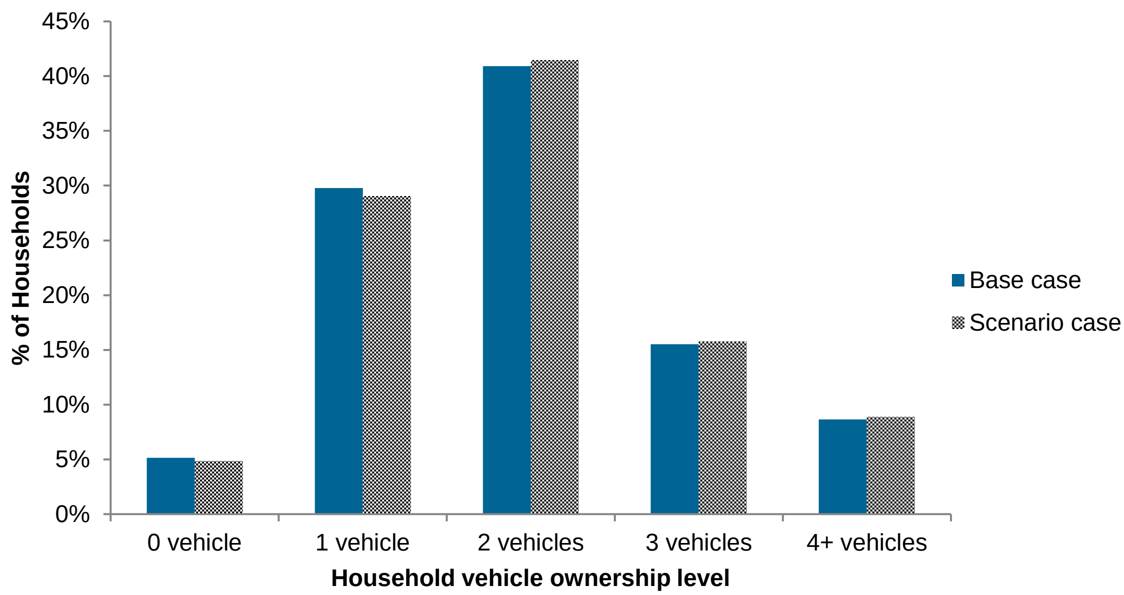
Five sensitivity tests were undertaken to assess the model's responsiveness to changes in policy sensitive variables. The policy sensitive variables and the changes tested included:

1. **Household income**: Increase all household incomes by 10 percent.
2. **Auto cost**: Increase all O-D car toll and operating costs by 50 percent.
3. **Auto travel time**: Increase all O-D car travel times by 25 percent.
4. **Air fare**: Increase all O-D air fares by 50 percent.
5. **Rail travel time**: Decrease all O-D rail travel times by 50 percent.

The sensitivity tests and key findings are discussed below.

8.1 Income Test

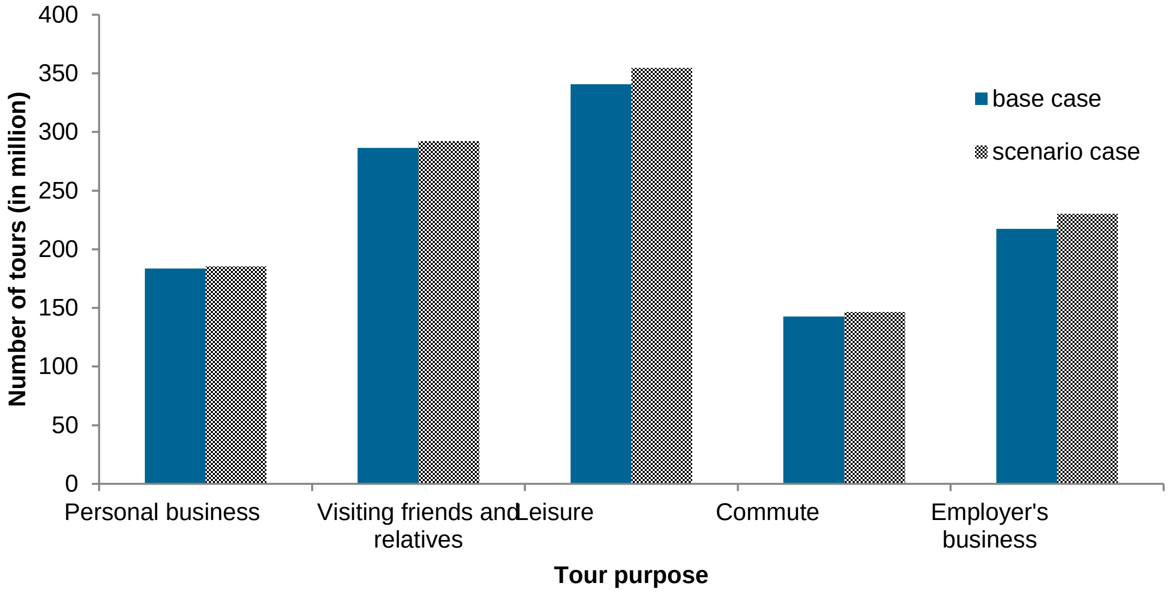
This test involved evaluating the impacts of changes in socioeconomic conditions on long-distance travel behavior. Specifically, this sensitivity test quantified changes in long-distance travel behavior due to a 10 percent increase in income. Figure 92 shows that a 10 percent increase in income is likely to increase household vehicle ownership level by shifting 0 and 1 vehicle households toward multivehicle households (income elasticities of vehicle ownership are -.58, -.25, .14, .17, and .26 for 0, 1, 2, 3, 4+ vehicles, respectively).



Source: FHWA

Figure 92. Percentage of households, by vehicle ownership level (scenario case: income test).

An increase in income is also expected to encourage more travel. The model results show a 3.2 percent increase in tour generation, a significant portion of which may be attributable to leisure and employer's business tours, as shown in Figure 93. Income elasticity of tour generation for leisure, employer's business, and other tour purposes are presented in the last column of Table 42. The table also shows that income increase is likely to cause an almost proportional increase in air mode (overall elasticity .84).



Source: FHWA

Figure 93. Number of tours, by purpose (scenario case: income test).

Table 42. Elasticity of tour mode, by purpose (scenario case: income test).

Purpose	Auto	Bus	Rail	Air	Total
Personal Business	0.09	0.06	0.10	0.57	0.11
Visiting Friends and Relatives	0.18	-0.17	-0.02	0.53	0.20
Leisure	0.39	-0.32	0.33	0.84	0.40
Commute	0.19	0.03	0.78	0.98	0.26
Business	0.49	0.26	0.35	1.17	0.58
Overall	0.28	-0.10	0.52	0.84	0.32

As expected, under this scenario, tours made by auto and rail are likely to increase as well, while tours by bus are likely to decrease. Unsurprisingly, similar proportional changes in total travel time, total travel cost, and total travel distance can be expected for each mode (Table 43). The table also shows that a 10 percent increase in income is likely to result in a 6.5 percent increase in travel expenditure. However, change in average person-miles traveled for each purpose and mode is expected to be none to minimal (Table 44 and Table 45).

Table 43. Elasticity of total travel time, cost, and distance, by mode (scenario case: income test).

Mode	Total Travel Time	Total Travel Cost	Total Travel Distance
Auto	0.28	0.31	0.28
Bus	-0.14	-0.15	-0.14
Rail	0.27	0.32	0.28
Air	0.79	0.86	0.79
Total	0.30	0.65	0.45

Table 44. Elasticity of average person-miles traveled, by purpose (scenario case: income test).

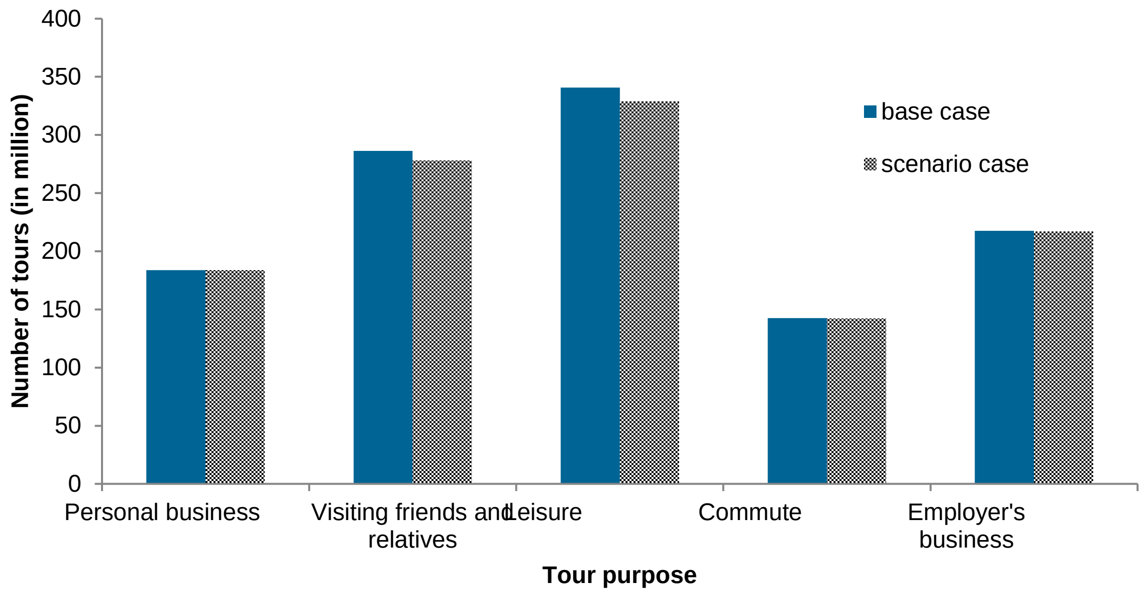
Purpose	Average Person-Miles Traveled
Personal Business	0.14
Visiting Friends and Relatives	0.11
Leisure	0.15
Commute	0.02
Business	0.25

Table 45. Elasticity of average person-miles traveled, by mode (scenario case: income test).

Mode	Average Person-Miles Traveled
Auto	0.04
Bus	0.04
Rail	-0.09
Air	0.00

8.2 Pricing Test (Auto Costs)

For the Pricing Test scenario, auto costs were increased by 50 percent to test the effect of pricing on a household's long-distance travel pattern. Such a change in auto costs is likely to result in an approximately 1.8 percent reduction in long-distance tour generation, mostly from leisure and visiting friends and relatives tour categories (Figure 94).



Source: FHWA

Figure 94. Number of tours, by purpose (scenario case: auto costs test).

The test indicated that households' long-distance travel behavior, in terms of mode choice, is fairly inelastic (Figure 94). Relative to base condition, a 50 percent increase in auto costs is likely to reduce auto tours by less than 2 percent (elasticity is -.04). This may be because for almost 90 percent of long-distance tours, auto is the only viable mode option.

Table 46. Change in mode share (scenario case: auto costs test).

Tour Mode	Base Case	Scenario Case	Difference	Elasticity
Auto	87.88%	87.58%	-0.30%	-0.04
Bus	1.80%	1.84%	0.04%	0.01
Rail	2.66%	2.72%	0.06%	0.01
Air	7.66%	7.86%	0.20%	0.02

To offset increase in travel costs by auto, in some instances households/individuals are likely to visit destinations that are closer to home. Table 47 and Table 48 demonstrate that a 50 percent increase in auto costs is likely to reduce total distance traveled, and average person-miles traveled by auto, by a little over 5 percent and just under 3 percent, respectively. A similar reduction can be expected in total travel time by auto (Table 49). On the other hand, travel cost by auto is likely to increase by approximately 55 percent (Table 50 and Table 51). This indicates that despite a 50 percent increase, from a total travel cost standpoint, auto is still the preferred mode for most long-distance tours.

Table 47. Change in distance traveled, total travel distance (in million miles) (scenario case: auto costs test).

Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	896,814	850,546	-5.16%	-0.10
Bus	22,114	22,086	-0.13%	0.00
Rail	15,305	15,295	-0.07%	0.00
Air	492,317	494,411	0.43%	0.01

Table 48. Change in distance traveled, average person-miles traveled (scenario case: auto costs test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	365	354	-2.91%
Bus	487	487	0.10%
Rail	300	300	-0.18%
Air	2,642	2,638	-0.15%

Table 49. Change in total travel time (scenario case: auto costs test).

Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	15,485	14,710	-5.00%	-0.10
Bus	546	545	-0.14%	0.00
Rail	324	323	-0.18%	0.00
Air	1,266	1,272	0.46%	0.01

Table 50. Change in travel cost, total travel cost (in thousand $) (scenario case: auto costs test).

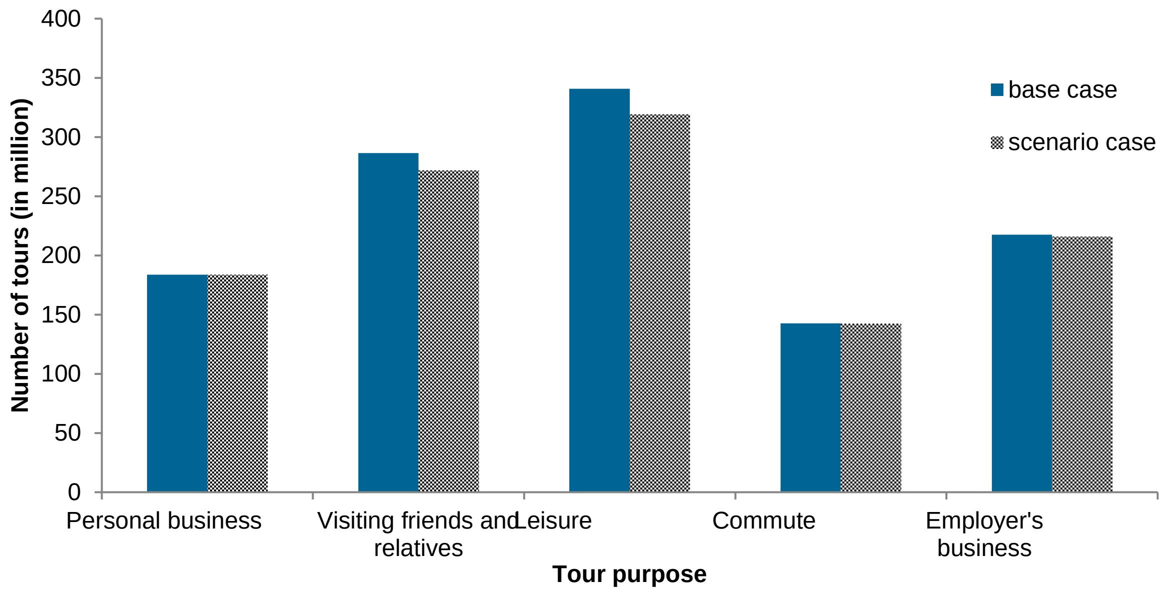
Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	98,408,206	140,515,046	42.79%	0.86
Bus	3,226,345	3,220,779	-0.17%	0.00
Rail	4,696,937	4,705,946	0.19%	0.00
Air	174,259,351	175,282,139	0.59%	0.01

Table 51. Change in travel cost, average travel cost per mile (in $/mile) (scenario case: auto costs test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	0.11	0.17	54.55%
Bus	0.15	0.15	0.00%
Rail	0.31	0.31	0.00%
Air	0.35	0.35	0.00%

8.3 Safety Test (Auto Times)

The Safety Test scenario indicated that long-distance travel is more sensitive to an increase in auto travel time than to an increase in auto travel cost. Under this scenario, travelers are likely to make 3.2 percent fewer long-distance tours—mostly fewer visiting friends and relatives and leisure tours—if auto travel time is increased by 25 percent (Figure 95). Such an increase in auto travel time is not expected to make any significant changes in long-distance travel mode share. Table 52 shows a 0.6 percent decrease in auto mode share and a 0.4 percent increase in air mode share under this scenario. Relative to base scenario, in some cases individuals are likely to travel to destinations closer to home by auto and to destinations that are farther afield by nonauto modes (Table 53 and Table 54). Despite switching destinations for some tours, total travel time by auto is likely to increase, though not proportionately (Table 55). A 10 percent increase in auto travel time is expected to increase total travel time by auto by approximately 5 percent (elasticity 0.46). However, driving to destinations closer to home may decrease total auto cost by a little less than 10 percent (Table 56 and Table 57).



Source: FHWA

Figure 95. Number of tours, by purpose (scenario case: auto times test).

Table 52. Change in mode share (scenario case: auto times test).

Tour Mode	Base Case	Scenario Case	Difference	Elasticity
Auto	87.88%	87.27%	-0.61%	-0.16
Bus	1.80%	1.88%	0.08%	0.05
Rail	2.66%	2.78%	0.12%	0.04
Air	7.66%	8.07%	0.41%	0.08

Table 53. Change in distance traveled, total travel distance (in million miles) (scenario case: auto times test).

Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	896,814	796,177	-11.22%	-0.45
Bus	22,114	22,449	1.52%	0.06
Rail	15,305	15,528	1.46%	0.06
Air	492,317	501,542	1.87%	0.07

Table 54. Change in distance traveled, average person-miles traveled (scenario case: auto times test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	365	338	-7.16%
Bus	487	490	0.68%
Rail	300	301	0.29%
Air	2,642	2,636	-0.24%

Table 55. Change in total travel time (scenario case: auto times test).

Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	15,485	17,254	11.43%	0.46
Bus	546	554	1.54%	0.06
Rail	324	328	1.43%	0.06
Air	1,266	1,291	1.98%	0.08

Table 56. Change in travel cost, total travel cost (in thousand $) (scenario case: auto times test).

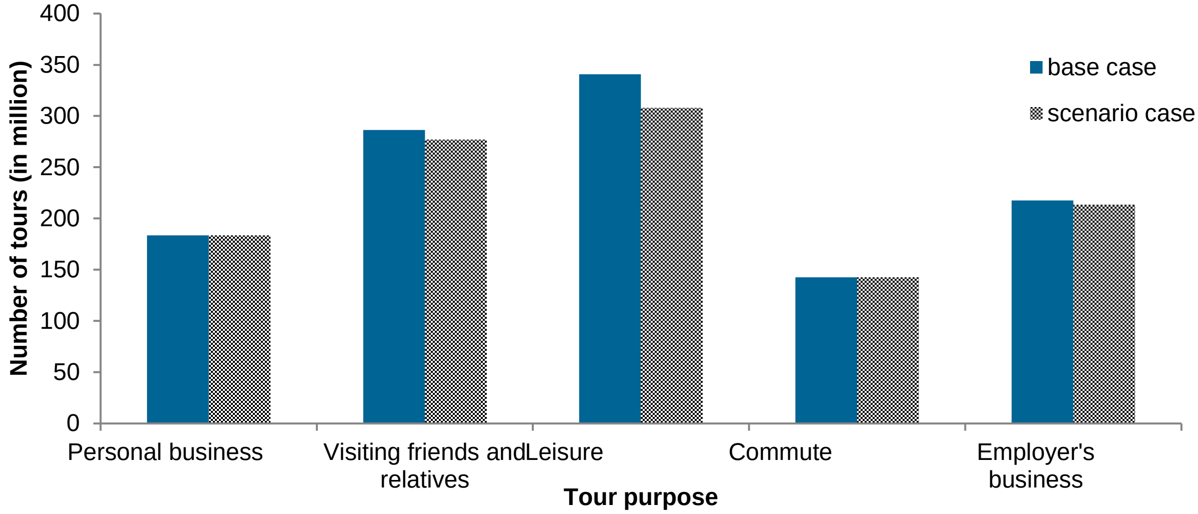
Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	98,408,206	89,045,193	-9.51%	-0.38
Bus	3,226,345	3,268,319	1.30%	0.05
Rail	4,696,937	4,772,735	1.61%	0.06
Air	174,259,351	178,276,796	2.31%	0.09

Table 57. Change in travel cost, average travel cost per mile (in $/mile) (scenario case: auto times test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	0.11	0.11	0.00%
Bus	0.15	0.15	0.00%
Rail	0.31	0.31	0.00%
Air	0.35	0.36	2.86%

8.4 Air Fare Test

A 50 percent increase in air fare is likely to suppress long-distance tours by 4 percent, mostly leisure tours, followed by visiting friends and relatives and employer's business tours (Figure 96). This scenario indicates a modal shift primarily from air to auto (1.6 percent, see Table 58).



Source: FHWA

Figure 96. Number of tours, by purpose (scenario case: air fare test).

Table 58. Change in mode share (scenario case: air fare test).

Tour Mode	Base Case	Scenario Case	Difference	Elasticity
Auto	87.88%	89.48%	1.60%	-0.04
Bus	1.80%	1.82%	0.02%	-0.05
Rail	2.66%	2.75%	0.09%	-0.01
Air	7.66%	5.95%	-1.72%	-0.51

As a result, total distance traveled by air is likely to drop by almost 30 percent, though expected reduction in average person-miles traveled by air is more modest, approximately 1.9 percent (Table 59 and Table 60). In line with total travel distance, total travel time by air is also likely to decrease significantly (Table 61). In addition, the results indicate that, far from being proportionate, a 10 percent increase in air fare is going to increase air expenditure by only 1.5 percent (elasticity of air travel cost with respect to air fare is 0.15, Table 62 and Table 63). This finding, together with other summary tables for this scenario, points to changes in long-distance travel patterns that are a combination of tour suppression, modal shift, and changes in destination choice.

Table 59. Change in distance traveled, total travel distance (in million miles) (scenario case: air fare test).

Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	896,814	875,697	-2.35%	-0.05
Bus	22,114	21,425	-3.12%	-0.06
Rail	15,305	15,161	-0.95%	-0.02
Air	492,317	349,727	-28.96%	-0.58

Table 60. Change in distance traveled, average person-miles traveled (scenario case: air fare test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	365	367	0.65%
Bus	487	493	1.16%
Rail	300	302	0.64%
Air	2,642	2,591	-1.92%

Table 61. Change in total travel time (scenario case: air fare test).

Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	15,485	15,116	-2.38%	-0.05
Bus	546	529	-3.08%	-0.06
Rail	324	321	-0.88%	-0.02
Air	1,266	898	-29.07%	-0.58

Table 62. Change in travel cost, total travel cost (in thousand $) (scenario case: air fare test).

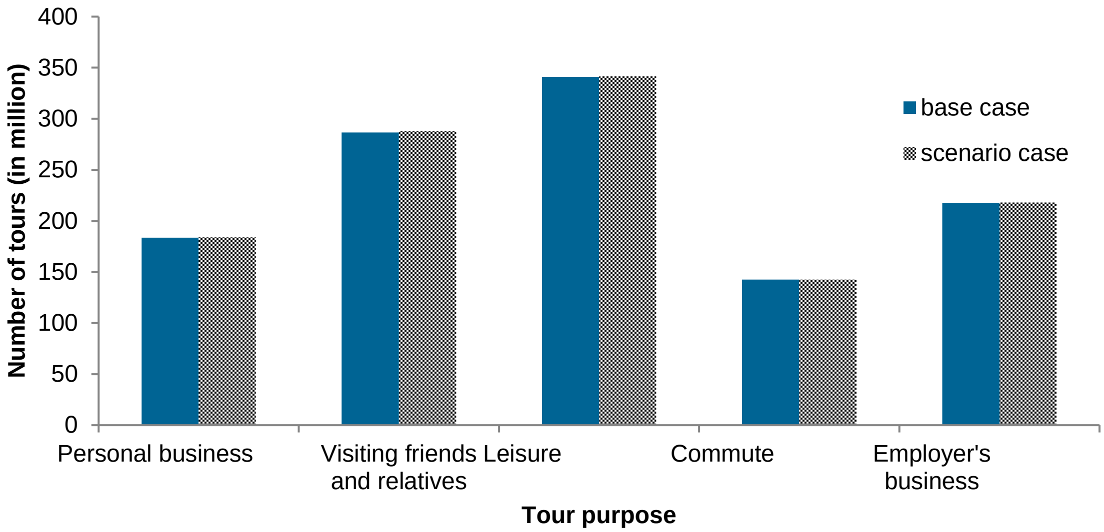
Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	98,408,206	96,545,001	-1.89%	-0.04
Bus	3,226,345	3,114,353	-3.47%	-0.07
Rail	4,696,937	4,649,209	-1.02%	-0.02
Air	174,259,351	187,384,824	7.53%	0.15

Table 63. Change in travel cost, average travel cost per mile (in $/mile) (scenario case: air fare test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	0.11	0.11	0.00%
Bus	0.15	0.15	0.00%
Rail	0.31	0.31	0.00%
Air	0.35	0.54	54.29%

8.5 Rail Time Test

The rail time test scenario measures the effect of a 50 percent reduction in rail travel time on long-distance travel. The results indicated that this scenario is likely to generate approximately 2.5 million more tours, mostly visiting friends and relatives, leisure, and employer's business tours (Figure 97). The results also indicate that a 50 percent faster rail system is likely to have no to a negligible effect on long-distance travel mode share (Table 67).



Source: FHWA

Figure 97. Number of tours, by purpose (scenario case: rail time test).

Table 64. Change in mode share (scenario case: rail time test).

Tour Mode	Base Case	Scenario Case	Difference	Elasticity
Auto	87.88%	87.70%	-0.18%	0.00
Bus	1.80%	1.79%	-0.01%	0.01
Rail	2.66%	2.88%	0.22%	-0.17
Air	7.66%	7.64%	-0.03%	0.00

This scenario is likely to encourage individuals to travel farther by rail, however. Table 65 and Table 66 shows that total distance traveled by rail is highly sensitive to rail travel time (elasticity -1.04). As a result, average person-miles traveled by rail can be expected to increase by almost 40 percent. Because of this significant increase in total travel distance, total travel time can be expected to result in an over 16 percent decrease (Table 67). Longer rail tours are also likely to contribute to higher travel costs (Table 68 and Table 69).

Table 65. Change in distance traveled, total travel distance (in million miles) (scenario case: rail time test).

Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	896,814	896,730	-0.01%	0.00
Bus	22,114	21,979	-0.61%	0.01
Rail	15,305	23,297	52.21%	-1.04
Air	492,317	491,230	-0.22%	0.00

Table 66. Change in distance traveled, average person-miles traveled (scenario case: rail time test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	365	364	-0.09%
Bus	487	486	-0.12%
Rail	300	415	38.24%
Air	2,642	2,640	-0.07%

Table 67. Change in total travel time (scenario case: rail time test).

Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	15,485	15,484	-0.01%	0.00
Bus	546	543	-0.60%	0.01
Rail	324	271	-16.35%	0.33
Air	1,266	1,263	-0.21%	0.00

Table 68. Change in travel cost, total travel cost (in thousand $) (scenario case: rail time test).

Tour Mode	Base Case	Scenario Case	% Difference	Elasticity
Auto	98,408,206	98,351,951	-0.06%	0.00
Bus	3,226,345	3,208,206	-0.56%	0.01
Rail	4,696,937	6,059,240	29.00%	-0.58
Air	174,259,351	173,982,734	-0.16%	0.00

Table 69. Change in travel cost, average travel cost per mile (in $/mile) (scenario case: rail time test).

Tour Mode	Base Case	Scenario Case	% Difference
Auto	0.11	0.11	0.00%
Bus	0.15	0.15	0.00%
Rail	0.31	0.26	-16.13%
Air	0.35	0.35	0.00%

8.6 Summary

A summary of the sensitivity test results is provided in Table 70. The test results indicate that:

- Higher incomes generate more tours, with some shift to longer distances and more expensive modes, mainly air.
- For auto, sensitivity to time changes is higher than sensitivity to cost changes—this may be because current auto costs are low.
- For auto trips, changing destinations is much more likely than changing mode or changing number of tours—this is because, for shorter distances, there is often no reasonable alternative to auto.
- The air fare elasticity is higher than car cost elasticity, with the largest mode shift effect.
- The rail time elasticity is higher than the car time elasticity, with substantial shifts in both mode and destination.

Table 70. Sensitivity test results summary.

Scenario	Income	Car Time	Rail Time	Car Cost	Air Fare
Change	Up 10%	Up 25%	Down 50%	Up 50%	Up 50%
Modes Included In Numbers Below	All	Car	Rail	Car	Air
Change in Total Tours Made	3.2%	-3.2%	0.2%	-1.8%	-4.0%
Change in Mode Share as a Percentage of Base Case Mode Share	n/a	-0.7%	8.2%	-0.3%	-22.4%
Change in Average Travel Distance per Tour	1.3%	-7.6%	40.4%	-3.2%	-4.7%
Change in Total Travel Distance in Mode(s)	4.5%	-11.2%	52.2%	-5.2%	-29.0%
Change in Average Travel Time per Tour	-0.2%	15.9%	-22.8%	-3.0%	-4.8%
Change in Total Travel Time in Mode(s)	3.0%	11.4%	-16.3%	-5.0%	-29.1%
Change in Average Travel Cost per Tour	3.2%	-5.8%	19.0%	45.9%	44.3%
Change in Total Travel Cost in Mode(s)	6.5%	-9.5%	29.0%	42.8%	7.5%
Elasticity of Travel Distance in Mode(s)	0.45	-0.45	-1.04	-0.10	-0.58
Elasticity of Travel Time Expenditure in Mode(s)	0.30	0.46	0.32	-0.10	-0.58
Elasticity of Travel Cost Expenditure in Mode(s)	0.65	-0.38	-0.58	0.86	0.15
Elasticity of travel distance by purpose					
Personal Business	0.23	-0.25	-1.11	-0.06	-0.53
Visit Friends or Relatives	0.31	-0.49	-1.72	-0.13	-0.66
Leisure/Vacation	0.54	-0.58	-1.15	-0.13	-0.75
Commuting	0.13	-0.13	-0.21	-0.05	-0.43
Employer's Business	0.75	-0.31	-1.06	-0.03	-0.26

CHAPTER 9. COMPARATIVE DATA ANALYSIS

The purpose of this chapter is to compare the 1995 ATS and the 2011 TAF data to the outputs from the long-distance passenger travel demand model (rJourney) for the year 2010. These comparisons provide insight on O-D patterns and modal shares among the various sources. To provide the most useful comparisons, shares of travel, by State or census division, are used to compare the data sources.

These data comparisons supplement the original implementation report where rJourney was compared to the original FHWA's Traffic Analysis Framework Multimodal Interregional Passenger Travel Origin-Destination Data results. The results contained herein are more detailed in terms of O-D patterns and mode shares.

Two national long-distance travel datasets permit comparison to the results of the National Long-Distance Passenger Model: the 1995 ATS and the 2011 TAF. The 1995 ATS is a survey and represents observed behavior from several decades ago. The 2011 TAF was developed from several observed sources, including the 1995 ATS, but does not represent a single observed data source.

The data comparisons contained herein focus primarily on patterns and shares rather than absolute values to provide the strongest comparative value for O-D patterns and mode shares. The O-D patterns are provided by census division and the origin patterns and mode shares are provided, by State. Origin-destination patterns are also provided by distance-band to compare one-way trip lengths. Comparing shares provides a direct comparison of results between rJourney and the TAF, which are representing the same timeframe (2010-2011), but may reflect some changes in shares between rJourney and the ATS, which represent a 15-year gap (1995-2010).

The rail and air modes in the 2008 TAF were based primarily on observed data, which allowed direct comparison of these modes. These ridership volumes are compared as average daily rail and air ridership, by State.

9.1 Origin-Destination Patterns

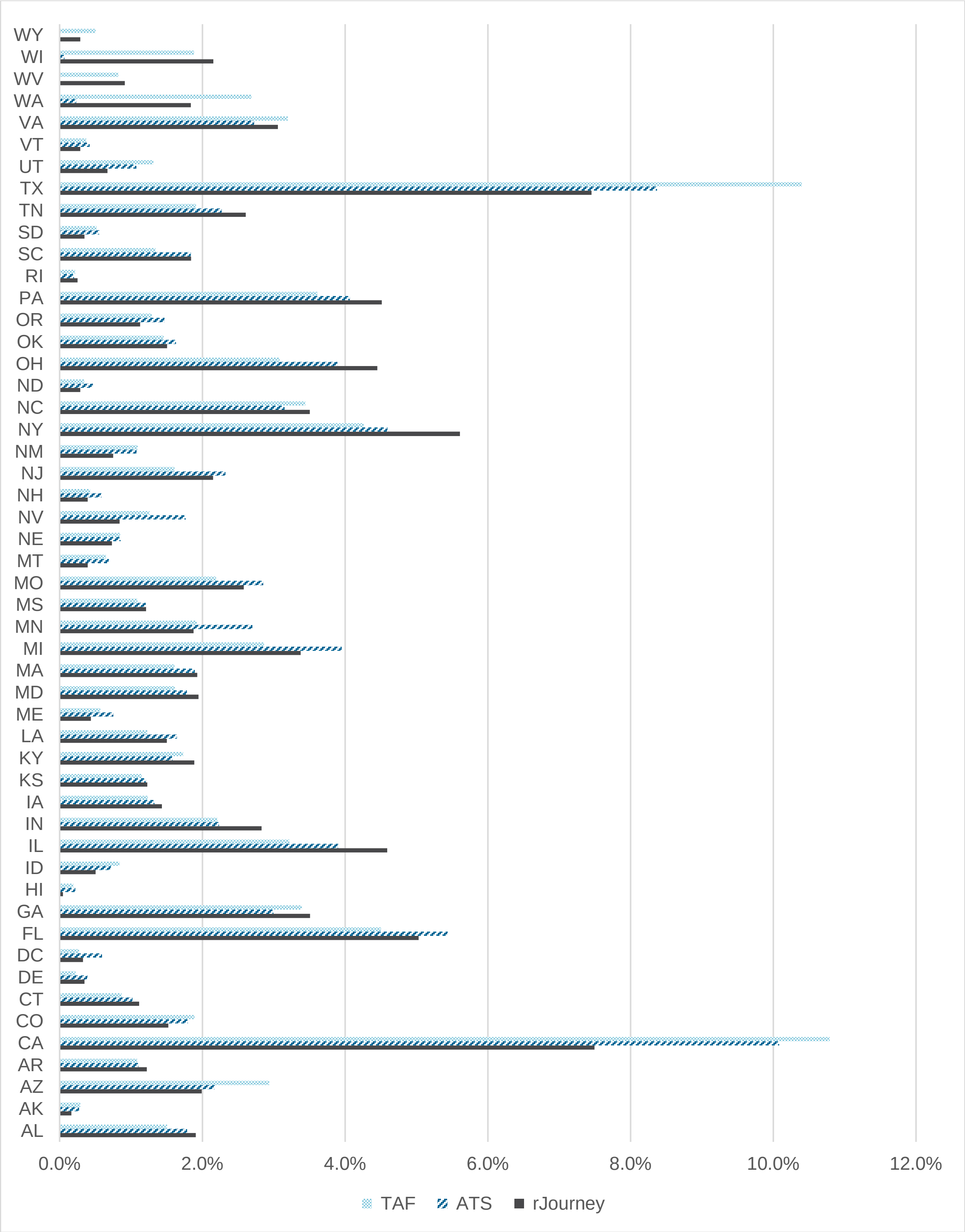
By State

Different ways exist to evaluate O-D patterns in the long-distance passenger travel context. Since one of the comparison datasets is from 1995, one useful way to compare the patterns of travel is by comparing the shares of trips by origin. Figure 98 presents the comparison of rJourney trip shares by origin State with the 1995 ATS and the 2011 TAF. Some observations of this comparison are:

- While the 1995 ATS was a large sample survey for long-distance travel, there are still some States with little or no travel originating (Wyoming, Wisconsin, and West Virginia).

- Since the 2011 TAF was built from the 1995 ATS, there are similarities between these data. One significant differences in Washington likely reflects a shortage of travelers from this State in the 1995 ATS.
- The largest States like Texas and California show higher shares of origin trips from the 1995 ATS and the 2011 TAF sources when compared to rJourney.
- Other large States (New York, Illinois, Pennsylvania, Ohio) show higher shares of origin trips from rJourney, when compared to the 1995 ATS and 2011 TAF data. Florida is an exception to this, with rJourney shares of origin trips in between the 1995 ATS and 2011 TAF data.

Overall, a strong correlation exists between the origin travel from each State in each of the three data sources.



Source: FHWA

Figure 98. Share of trips, by origin State.

By Census Division

The United States includes nine census divisions (Figure 77), which provides a means to consider O-D patterns among and between these regions. Average daily travel from rJourney, the 1995 ATS and 2011 TAF were aggregated to census divisions and reported for each O-D pair in Table 71 through Table 73, as a share of total travel. Table 74 and Table 75 present a comparison of rJourney trip shares with the 1995 ATS and 2011 TAF, respectively.

Table 74 and Table 75 show that rJourney is predicting higher shares of travel east of the Mississippi (East-North Central, East-South Central, Mid-Atlantic and South Atlantic) compared to both the 1995 ATS and 2011 TAF, except in New England, where rJourney predicts similar shares of travel. In addition, rJourney predicts lower shares of travel west of the Mississippi (West-North Central, West-South Central, Mountain and Pacific) compared to the 1995 ATS and 2011 TAF. In most cases, the differences between rJourney and the 2011 TAF are larger than the differences between rJourney and the 1995 ATS, but the underlying comparative patterns are similar (as expected since the 2011 TAF was derived from the 1995 ATS).

By Distance

A third comparison of these data sources to the rJourney output is possible by distance bands. This comparison can determine whether trip lengths are significantly different between the various sources. Figure 99 presents the trip shares, by distance-band. These are denoted in 100 mile increments up to 800 miles.

Table 71. Origin-destination patterns, by census division—rJourney.

Census Division	East-North Central	East-South Central	Mid-Atlantic	Mountain	New England	Pacific	South Atlantic	West-North Central	West-South Central	Grand Total
East-North Central	10.8%	1.2%	1.2%	0.2%	0.2%	0.2%	1.3%	1.7%	0.5%	**17.4%**
East-South Central	1.2%	2.9%	0.2%	0.1%	0.0%	0.1%	2.0%	0.3%	0.8%	**7.6%**
Mid-Atlantic	1.2%	0.2%	5.4%	0.1%	1.9%	0.4%	2.7%	0.2%	0.2%	**12.3%**
Mountain	0.2%	0.1%	0.1%	4.1%	0.1%	1.3%	0.2%	0.4%	0.5%	**7.0%**
New England	0.2%	0.0%	1.9%	0.1%	1.6%	0.2%	0.4%	0.0%	0.1%	**4.4%**
Pacific	0.2%	0.1%	0.4%	1.3%	0.2%	7.7%	0.4%	0.2%	0.3%	**10.7%**
South Atlantic	1.3%	2.0%	2.7%	0.2%	0.4%	0.4%	12.6%	0.3%	0.6%	**20.5%**
West-North Central	1.7%	0.3%	0.2%	0.4%	0.0%	0.2%	0.3%	4.5%	0.9%	**8.5%**
West-South Central	0.5%	0.8%	0.2%	0.5%	0.1%	0.3%	0.6%	0.9%	8.0%	**11.7%**
Grand Total	**17.4%**	**7.6%**	**12.3%**	**7.0%**	**4.4%**	**10.7%**	**20.5%**	**8.5%**	**11.7%**	**100.0%**

Source: FHWA

Table 72. Origin-destination patterns, by census division—1995 ATS.

Census Division	East-North Central	East-South Central	Mid-Atlantic	Mountain	New England	Pacific	South Atlantic	West-North Central	West-South Central	Grand Total
East-North Central	9.5%	0.8%	0.7%	0.3%	0.1%	0.3%	1.2%	0.8%	0.3%	**14.1%**
East-South Central	0.8%	3.5%	0.1%	0.1%	0.0%	0.1%	1.5%	0.2%	0.5%	**6.9%**
Mid-Atlantic	0.7%	0.1%	5.8%	0.2%	1.2%	0.3%	2.5%	0.1%	0.2%	**11.0%**
Mountain	0.3%	0.1%	0.2%	5.5%	0.1%	2.1%	0.2%	0.4%	0.5%	**9.3%**
New England	0.1%	0.0%	1.2%	0.1%	2.8%	0.1%	0.5%	0.0%	0.0%	**4.9%**
Pacific	0.3%	0.1%	0.3%	1.9%	0.1%	8.8%	0.3%	0.2%	0.2%	**12.3%**
South Atlantic	1.1%	1.5%	2.4%	0.2%	0.5%	0.3%	12.2%	0.3%	0.5%	**18.9%**
West-North Central	1.2%	0.2%	0.1%	0.4%	0.0%	0.2%	0.3%	6.9%	0.6%	**10.0%**
West-South Central	0.3%	0.5%	0.2%	0.5%	0.0%	0.2%	0.5%	0.6%	9.8%	**12.8%**
Grand Total	**14.5%**	**6.8%**	**10.9%**	**9.2%**	**4.9%**	**12.5%**	**19.2%**	**9.5%**	**12.7%**	**100.0%**

Source: FHWA

Table 73. Origin-destination patterns, by census division—2011 TAF.

Census Division	East-North Central	East-South Central	Mid-Atlantic	Mountain	New England	Pacific	South Atlantic	West-North Central	West-South Central	Grand Total
East-North Central	7.8%	1.0%	0.9%	0.3%	0.2%	0.2%	1.3%	1.4%	0.3%	**13.3%**
East-South Central	1.0%	2.3%	0.2%	0.1%	0.0%	0.1%	1.6%	0.3%	0.7%	**6.2%**
Mid-Atlantic	0.9%	0.2%	4.1%	0.2%	1.3%	0.2%	2.4%	0.1%	0.1%	**9.5%**
Mountain	0.3%	0.1%	0.2%	6.1%	0.1%	2.1%	0.3%	0.6%	0.9%	**10.5%**
New England	0.2%	0.0%	1.3%	0.1%	1.8%	0.1%	0.5%	0.0%	0.0%	**4.1%**
Pacific	0.2%	0.1%	0.2%	2.1%	0.1%	11.5%	0.3%	0.3%	0.4%	**15.2%**
South Atlantic	1.3%	1.6%	2.4%	0.3%	0.5%	0.3%	11.6%	0.3%	0.5%	**18.8%**
West-North Central	1.4%	0.3%	0.1%	0.6%	0.0%	0.3%	0.3%	4.5%	0.8%	**8.2%**
West-South Central	0.3%	0.7%	0.1%	0.9%	0.0%	0.4%	0.5%	0.8%	10.3%	**14.2%**
Grand Total	**13.3%**	**6.2%**	**9.5%**	**10.5%**	**4.1%**	**15.3%**	**18.8%**	**8.2%**	**14.2%**	**100.0%**

Source: FHWA

Table 74. Comparison of trip O-D patterns, by census division—rJourney vs. 1995 ATS.

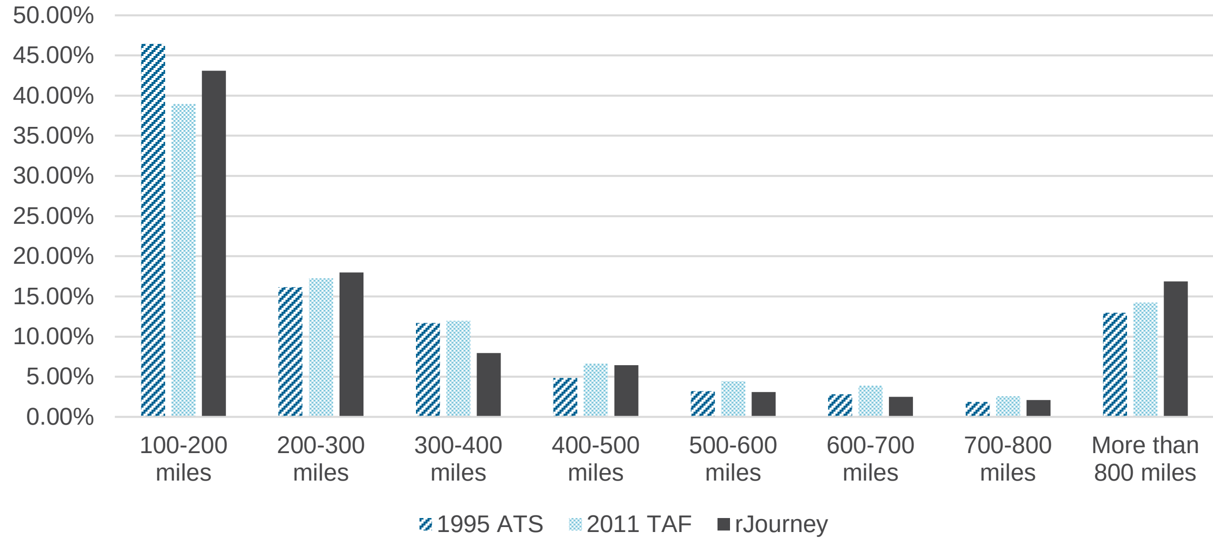
Census Division	East-North Central	East-South Central	Mid-Atlantic	Mountain	New England	Pacific	South Atlantic	West-North Central	West-South Central	Grand Total
East-North Central	1.3%	0.4%	0.6%	-0.1%	0.0%	-0.1%	0.2%	0.9%	0.1%	**3.4%**
East-South Central	0.4%	-0.7%	0.1%	0.0%	0.0%	0.0%	0.4%	0.2%	0.3%	**0.8%**
Mid-Atlantic	0.5%	0.1%	-0.5%	0.0%	0.7%	0.1%	0.2%	0.1%	0.0%	**1.3%**
Mountain	-0.1%	0.0%	0.0%	-1.4%	0.0%	-0.8%	0.0%	0.0%	0.0%	**-2.3%**
New England	0.0%	0.0%	0.7%	0.0%	-1.2%	0.0%	-0.1%	0.0%	0.0%	**-0.5%**
Pacific	-0.1%	0.0%	0.1%	-0.7%	0.0%	-1.1%	0.1%	0.0%	0.0%	**-1.6%**
South Atlantic	0.3%	0.5%	0.3%	0.0%	-0.1%	0.1%	0.4%	0.1%	0.1%	**1.6%**
West-North Central	0.5%	0.2%	0.1%	-0.1%	0.0%	-0.1%	0.1%	-2.4%	0.2%	**-1.5%**
West-South Central	0.1%	0.3%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	-1.9%	**-1.1%**
Grand Total	**3.0%**	**0.8%**	**1.4%**	**-2.2%**	**-0.4%**	**-1.8%**	**1.3%**	**-1.0%**	**-1.0%**	**0.0%**

Source: FHWA

Table 75. Comparison of trip O-D patterns, by census division—rJourney vs. 2011 TAF.

Census Division	East-North Central	East-South Central	Mid-Atlantic	Mountain	New England	Pacific	South Atlantic	West-North Central	West-South Central	Grand Total
East-North Central	3.0%	0.3%	0.4%	0.0%	0.0%	0.0%	0.1%	0.3%	0.1%	**4.1%**
East-South Central	0.3%	0.5%	0.0%	0.0%	0.0%	0.0%	0.3%	0.1%	0.1%	**1.4%**
Mid-Atlantic	0.4%	0.0%	1.3%	0.0%	0.6%	0.1%	0.3%	0.1%	0.0%	**2.8%**
Mountain	0.0%	0.0%	0.0%	-1.9%	0.0%	-0.8%	-0.1%	-0.2%	-0.4%	**-3.5%**
New England	0.0%	0.0%	0.6%	0.0%	-0.2%	0.1%	-0.2%	0.0%	0.0%	**0.3%**
Pacific	0.0%	0.0%	0.1%	-0.8%	0.1%	-3.7%	0.1%	-0.1%	-0.2%	**-4.6%**
South Atlantic	0.1%	0.3%	0.3%	-0.1%	-0.2%	0.1%	1.0%	0.0%	0.1%	**1.7%**
West-North Central	0.3%	0.1%	0.1%	-0.2%	0.0%	-0.1%	0.0%	0.0%	0.1%	**0.3%**
West-South Central	0.1%	0.1%	0.0%	-0.4%	0.0%	-0.2%	0.1%	0.1%	-2.3%	**-2.5%**
Grand Total	**4.1%**	**1.4%**	**2.8%**	**-3.5%**	**0.3%**	**-4.6%**	**1.7%**	**0.3%**	**-2.5%**	**0.0%**

Source: FHWA



Source: FHWA

Figure 99. Trip shares, by distance-band.

The differences are that rJourney predicts more trips that are shorter (less than 300 miles) and longer (more than 800 miles). Table 76 presents the average trip lengths by mode from each source, indicating that overall, the rJourney results have shorter trip lengths in total and for auto and air modes, while the bus and rail trip lengths for rJourney are in between the 1995 ATS and 2011 TAF data sources.

Table 76. Average trip lengths (miles), by mode.

Mode	1995 ATS	2011 TAF	rJourney
Auto	368	404	340
Bus	419	499	464
Rail	593	359	418
Air	1,199	1,293	905
Total	506	535	386

9.2 Mode Shares

Comparing mode shares across disparate data sources provides an equal comparison. These are compared for each of the four modes (auto, bus, rail, and air).

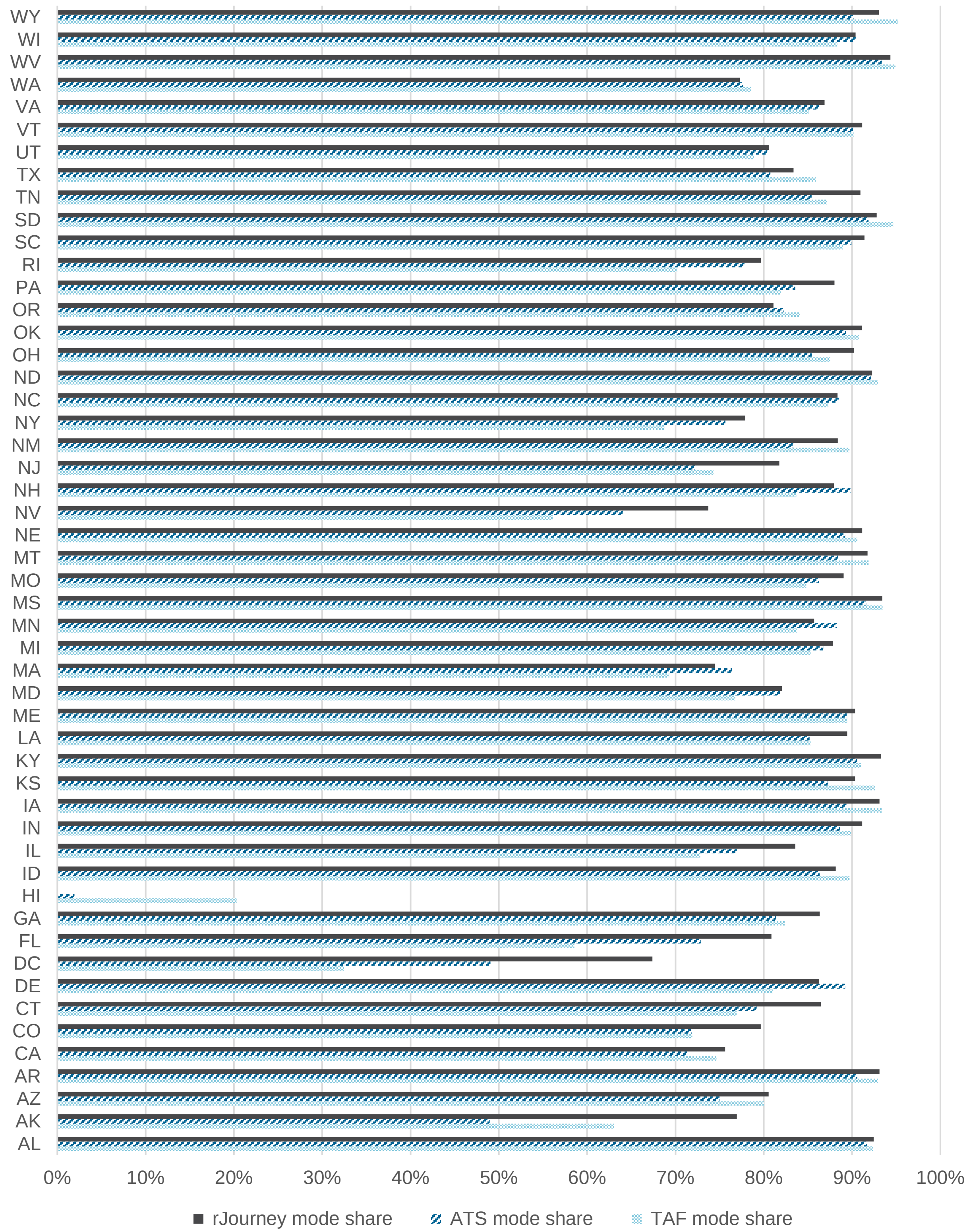
By State

Figure 100 presents a comparison of auto mode shares by source and State. A high correlation exists in most States among the three datasets, but some differences are worth noting:

- One discrepancy reflects different assumptions for Hawai'i, where rJourney assumes no auto travel more than 100 miles, ATS assumes a small proportion of auto travel and TAF assumes a large portion of auto travel. The rJourney assumption for no auto travel was a simplification, since there can be some travel >100 miles on the island of Hawai'i.
- In the District of Columbia (DC), the rJourney auto mode share is much higher (67 percent) than either the 1995 ATS 49 percent) or the 2011 TAF (32 percent). DC is an urban area, rather than a State, so a further evaluation of urban and rural patterns may provide insight on this comparison.
- In Nevada (NV), the rJourney auto mode share is higher (74 percent) than either the 1995 ATS (64 percent) or the 2011 TAF (56 percent).
- In Florida (FL), there is a similar trend with rJourney auto mode share higher (81 percent) than either the 1995 ATS (73 percent) or the 2011 TAF (59 percent).
- In Alaska (AK), there is a similar trend with rJourney auto mode share higher (77 percent) than either the 1995 ATS (49 percent) or the 2011 TAF (63 percent).

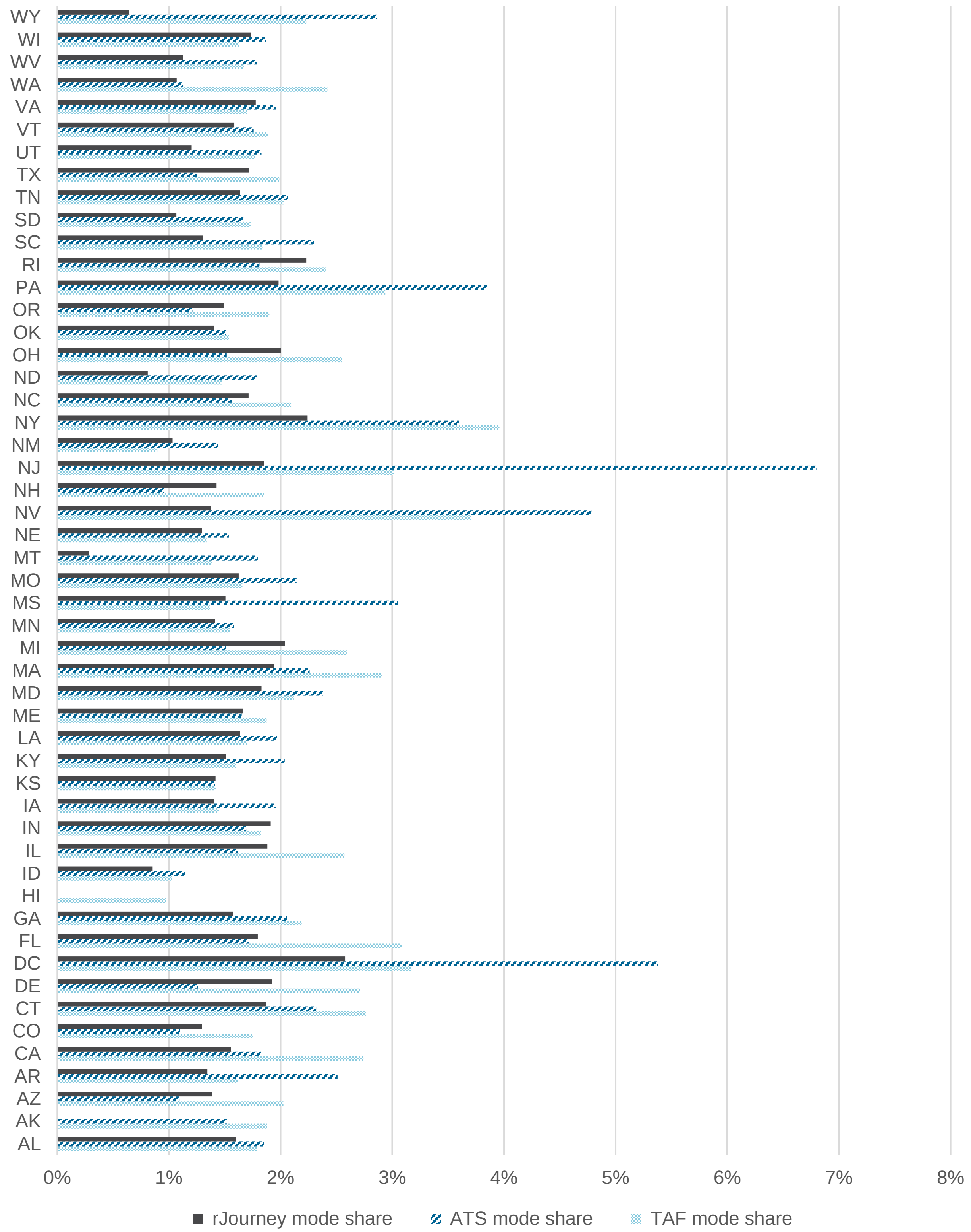
No States exist where the rJourney significantly underpredicts the auto mode share compared to the 1995 ATS or the 2011 TAF.

Figure 101 presents a comparison of the bus mode shares by source and State. These percent mode shares are quite small overall, with a maximum less than 7 percent and most States under 2 percent. rJourney does not produce any long-distance bus trips for Alaska or Hawai'i, as a simplifying assumption. Estimates for long-distance bus are lower in Wyoming (WY), Washington (WA), Pennsylvania (PA), New York (NY), Nevada (NV), Florida (FL), CA, and Arkansas (AK) than both the 1995 ATS and the 2011 TAF.



Source: FHWA

Figure 100. Auto mode shares, by origin State.



Source: FHWA

Figure 101. Bus mode shares, by origin State.

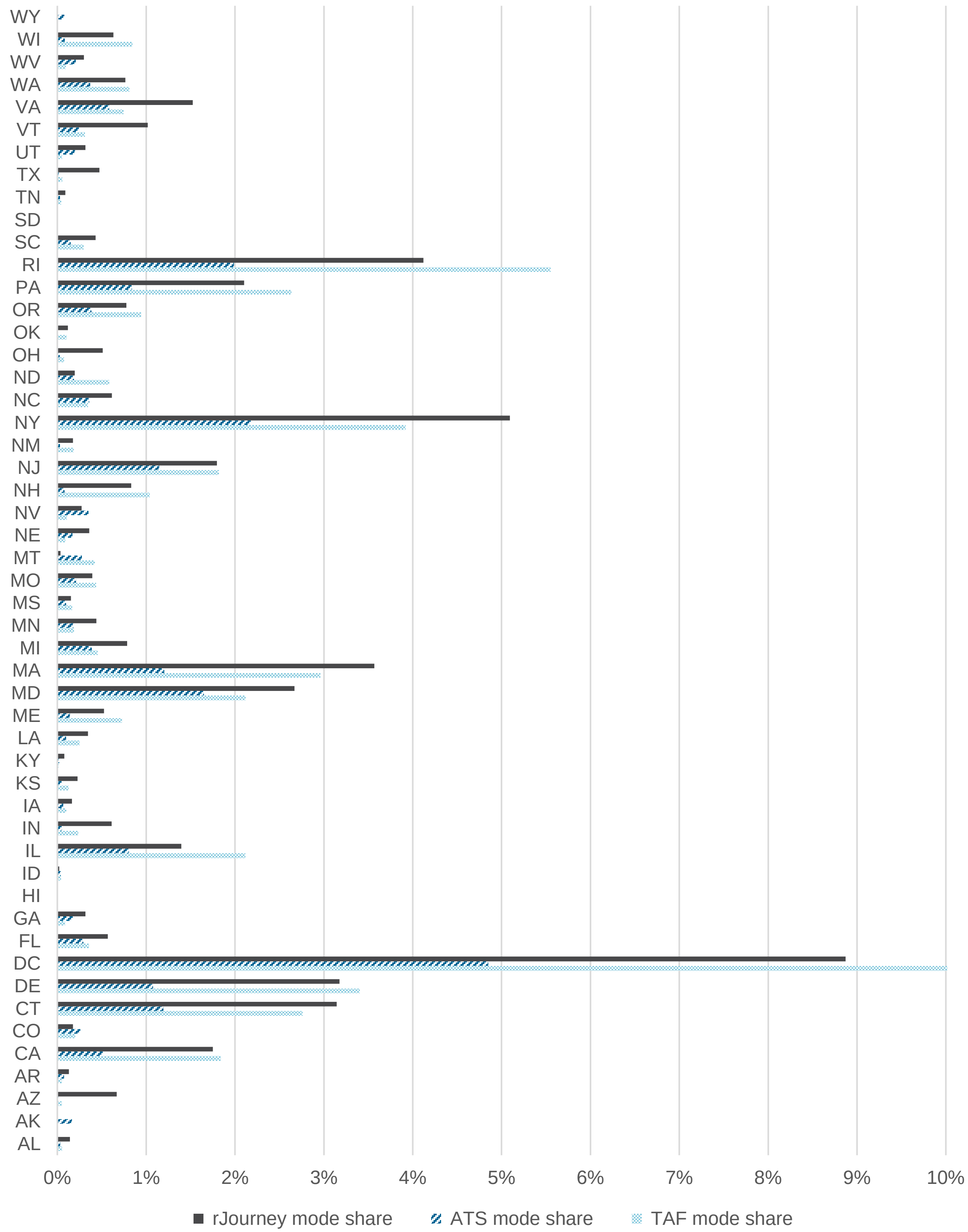
The rail mode shares are presented in Figure 102, by State and data source. Rail mode shares are reasonably consistent between the 2011 TAF and rJourney estimates, except in DC where the 2011 TAF shows a 25 percent rail mode share and rJourney shows a 9 percent rail mode share. The 1995 ATS rail mode shares are quite a bit smaller than either other source, in part because of the introduction of more long-distance rail service in the last 20+ years.

Observed data sources for rail facilitated development of the 2011 TAF, which allowed comparisons between average daily rail ridership and the rJourney rail ridership (Figure 103). Nationwide, 2011 rJourney produces 140,000 average daily rail riders while the 2011 TAF produces 81,000 average daily rail riders. The Figure 103 comparison shows that the over-estimation in rail ridership is primarily along the east coast and Midwest, while the western regions are more closely aligned. The discrepancy noted above for DC is less significant in this comparison because the absolute values are small. This can provide some direction to update the calibration and validation of the long-distance model for rail.

Air mode shares, by State and data source, are presented in Figure 104. Air mode shares are reasonably consistent across the States, except in previously mentioned States with mode share discrepancies:

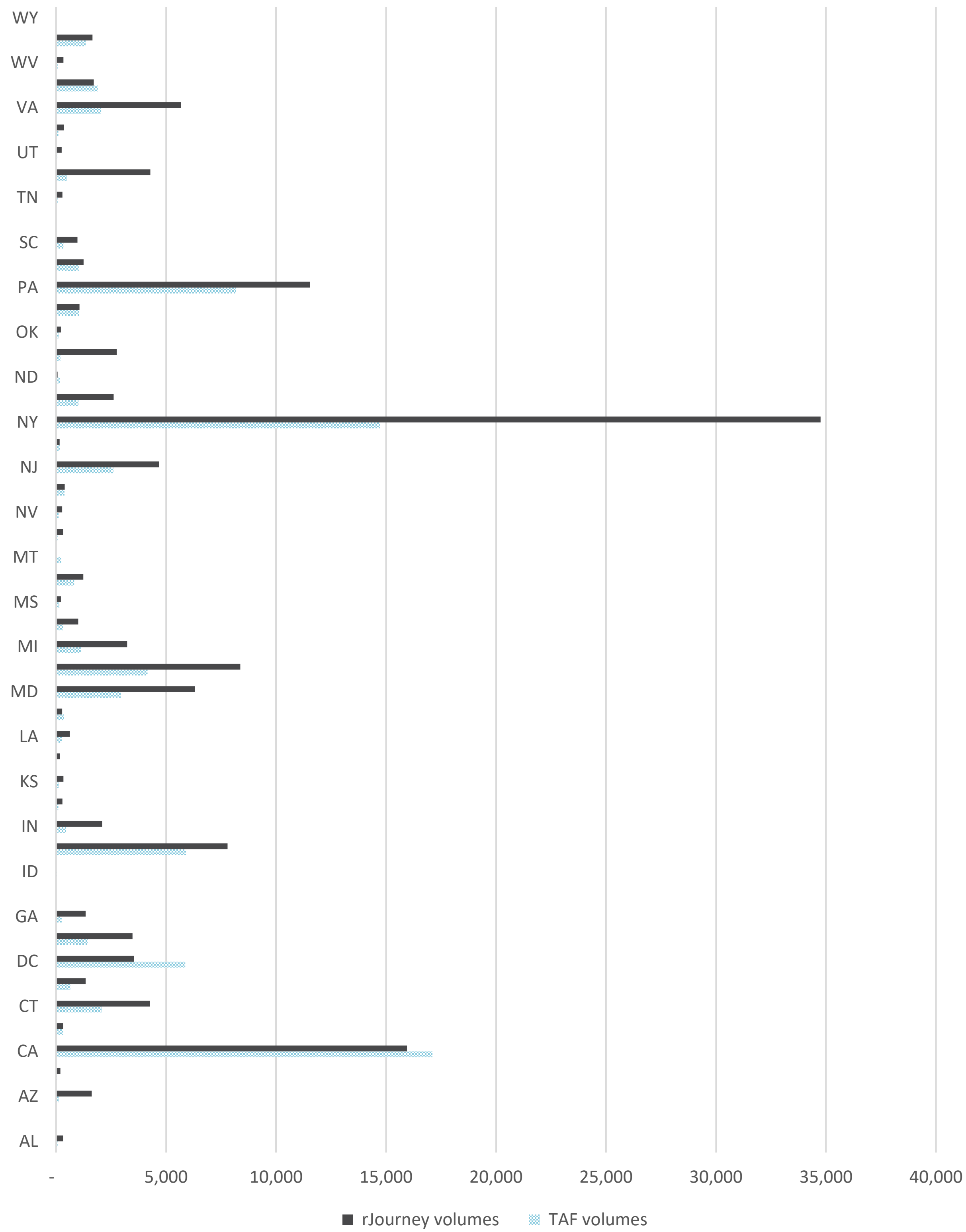
- In Hawai'i, where rJourney assumes all travel more than 100 miles is by air, ATS assumes most travel is by air (98 percent) and TAF assumes 79 percent by air. The rJourney assumption for all air travel was a simplification, since most travel over 100 miles is between islands or to and from the mainland.
- In the District of Columbia (DC), the rJourney air mode share is much lower (21 percent) than either the 1995 ATS (41 percent) or the 2011 TAF (40 percent).
- In NV, the rJourney air mode share is lower (25 percent) than either the 1995 ATS (31 percent) or the 2011 TAF (40 percent).
- In Florida (FL), there is a similar trend with rJourney air mode share lower (17 percent) than either the 1995 ATS (25 percent) or the 2011 TAF (38 percent).
- In Alaska (AK), there is a similar trend with rJourney air mode share lower (23 percent) than either the 1995 ATS (49 percent) or the 2011 TAF (35 percent).

The project team developed 2011 TAF for air travel directly from observed data sources to facilitate comparison between average daily air ridership and the rJourney air ridership (Figure 105). Nationwide, rJourney produces 1,464,000 average daily air riders and the 2011 TAF produces 1,251,000 average daily air riders. The Figure 105 comparison by State shows a reasonable comparison across most States and significant differences in only a few States (Texas and New York are overestimating air riders, while Florida is under-estimation air riders).



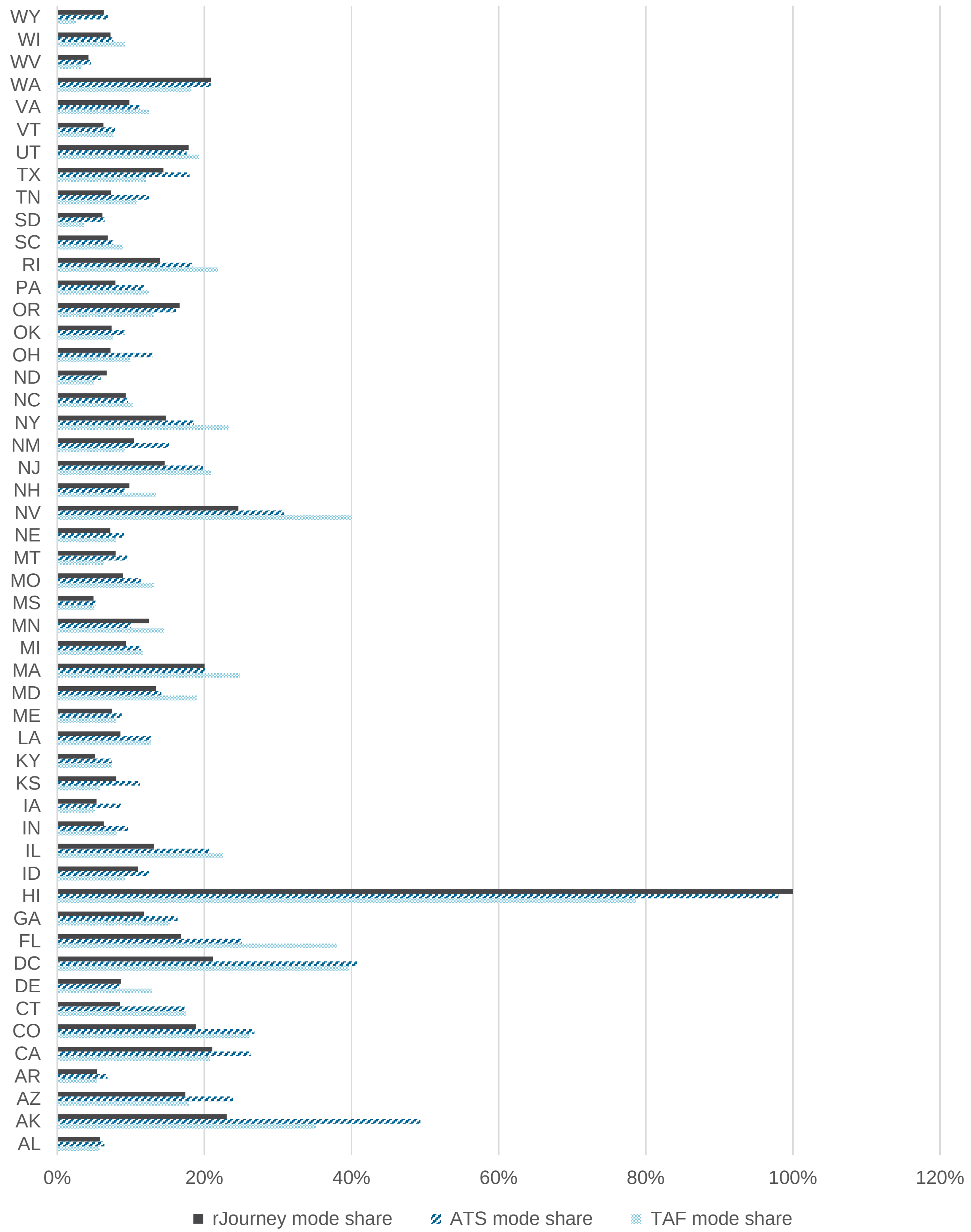
Source: FHWA

Figure 102. Rail mode shares, by origin State.



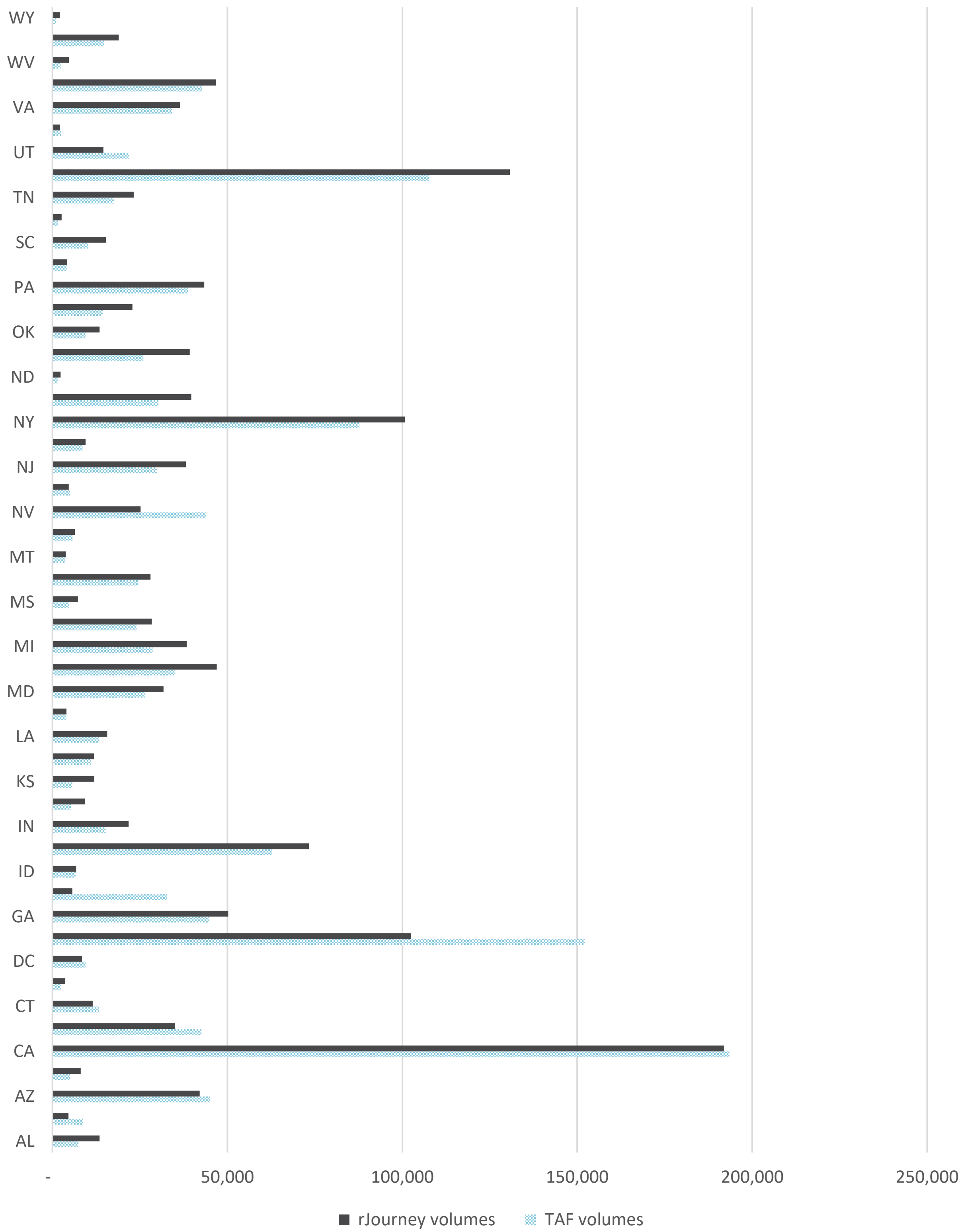
Source: FHWA

Figure 103. Average daily rail ridership, by origin State.



Source: FHWA

Figure 104. Air mode shares, by origin State.



Source: FHWA

Figure 105. Average daily air ridership, by origin State.

By Census Division

The nine census divisions presented in provide an opportunity to review the mode shares by origin and destination pairs. Figure 106 and Figure 107 show a comparison of auto mode shares from rJourney compared to the 1995 ATS and the 2011 TAF, respectively. In each chart, an orange line represents an exact match between the two data sources being compared. presents the auto mode shares by census division origin and destination pairs, comparing the rJourney with the 1995 ATS and the 2011 TAF sources. The overall auto mode share for rJourney is 85 percent, compared to the 1995 ATS of 81 percent and the 2011 TAF of 82 percent.

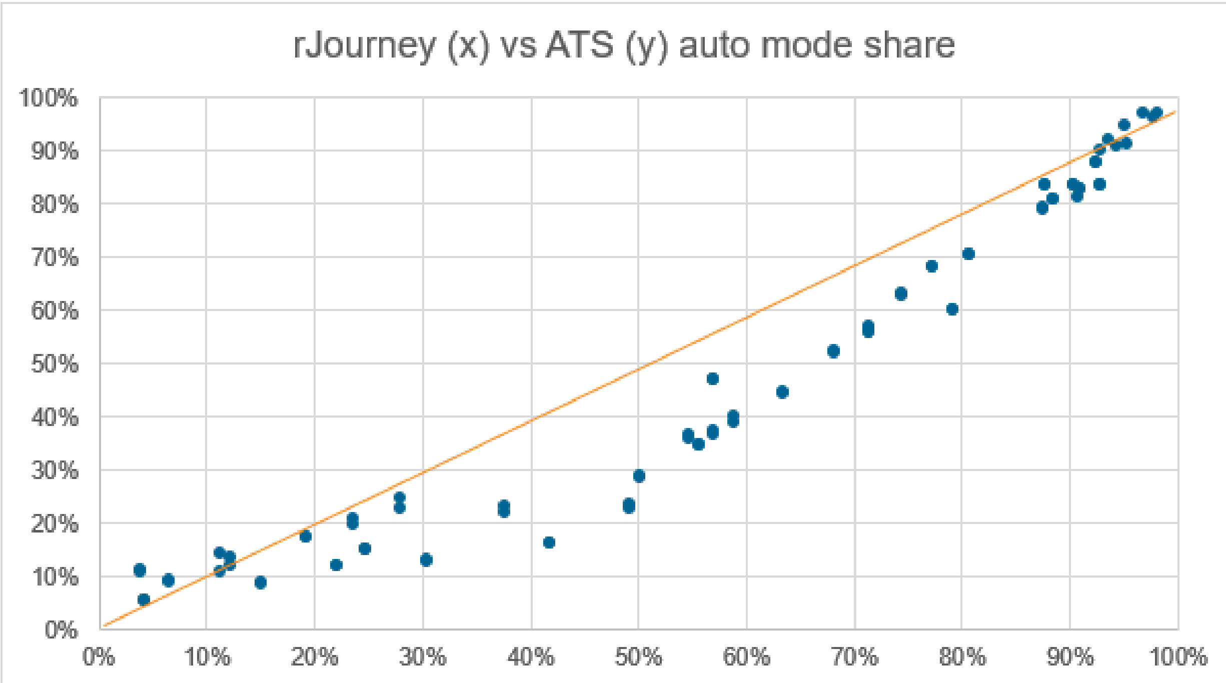
In the case of the 2011 TAF, there are several census division pairs with no auto mode shares:

- New England (1) to East-North Central (3).
- New England (1) to Mountain (8) and vice versa.
- Mid-Atlantic (2) to Mountain (8) and vice versa.
- Mid-Atlantic (2) to Pacific (9) and vice versa.
- East-South Central (6) to South Atlantic (5).

These data sources are likely under-representing auto mode shares for these O-D pairs.

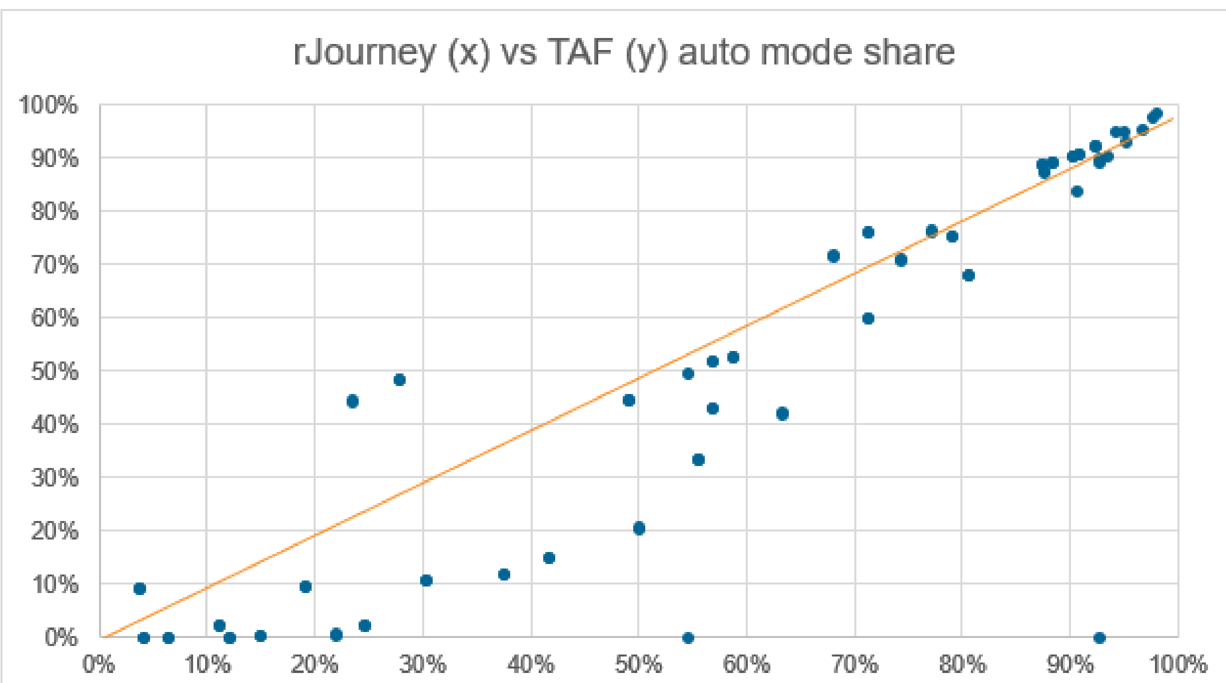
Figure 108 and Figure 109 present the bus mode shares from the 1995 ATS and 2011 TAF data compared to rJourney. The overall bus mode share for rJourney is 1.7 percent, compared to the 1995 ATS of 2.1 percent and the 2011 TAF of 2.3 percent. These comparisons do not show as much alignment as the auto mode shares, but there is still reasonable correlation across the O-D pairs. Again, there are a few places in the 2011 TAF with no bus mode shares, even though the other data sources show a bus mode share for these pairs:

- New England (1) to East-North Central (3).
- New England (1) to Pacific (9) and vice versa.
- East-South Central (6) to South Atlantic (5).



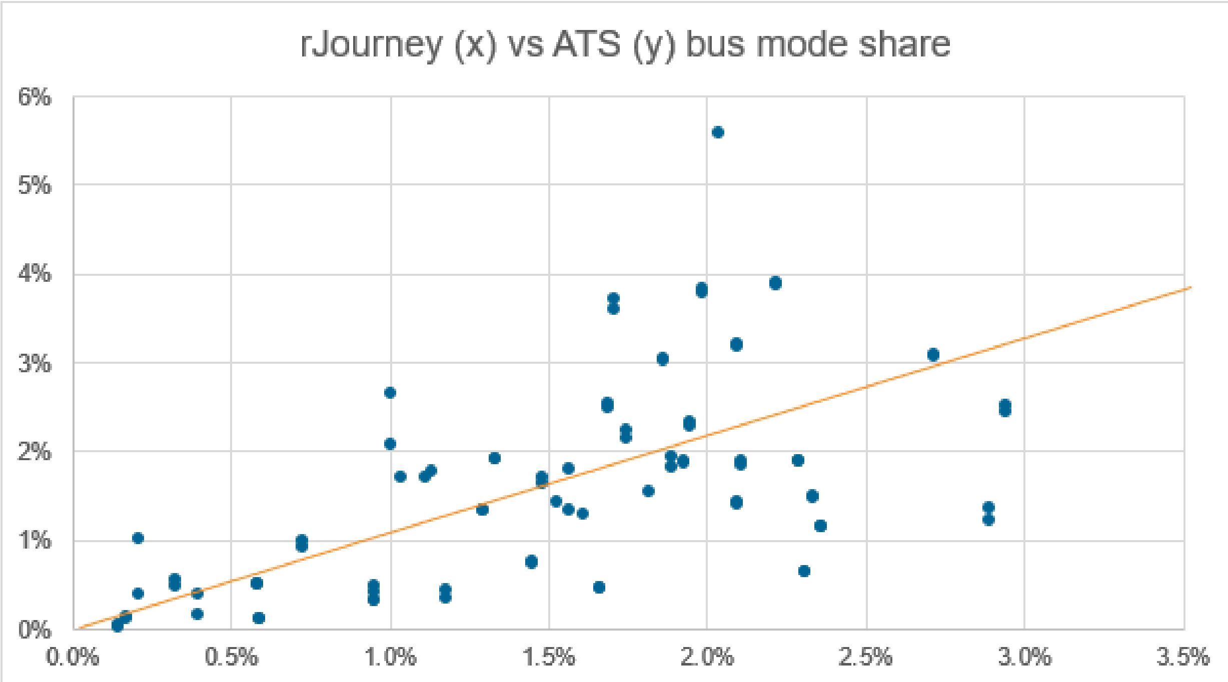
Source: FHWA

Figure 106. Auto mode shares compared to ATS by census division O-D pairs.



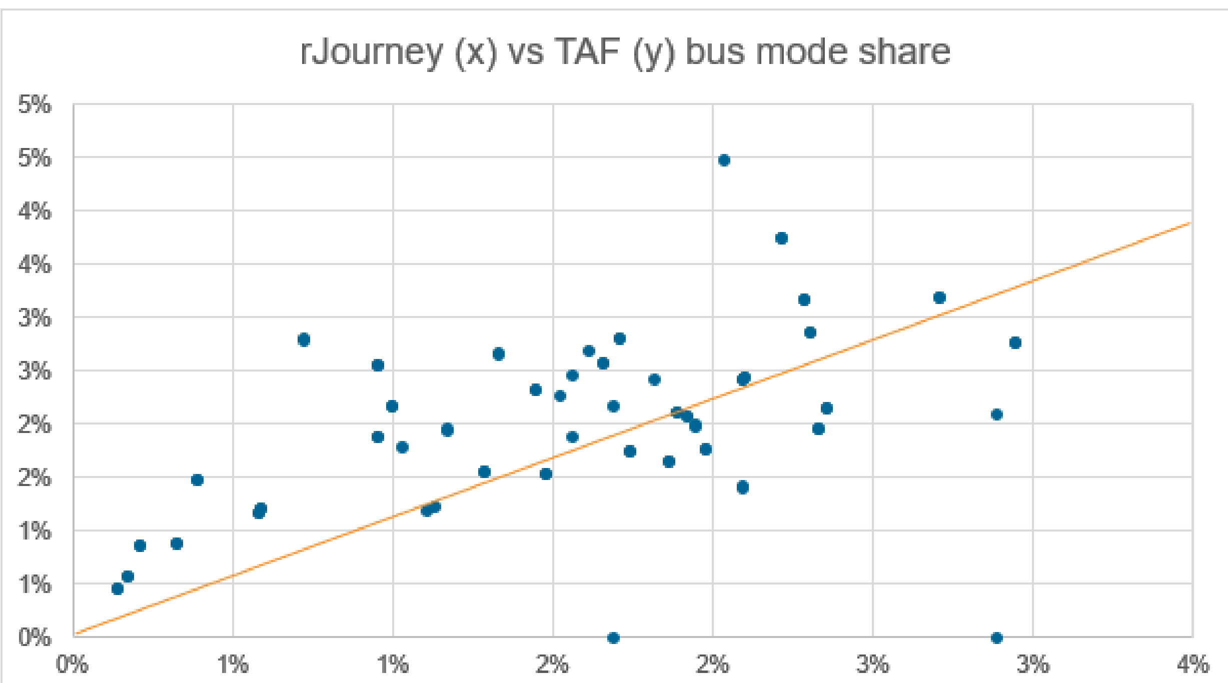
Source: FHWA

Figure 107. Auto mode shares compared to TAF by census division O-D pairs.



Source: FHWA

Figure 108. Bus mode shares compared to ATS by census division O-D pairs.

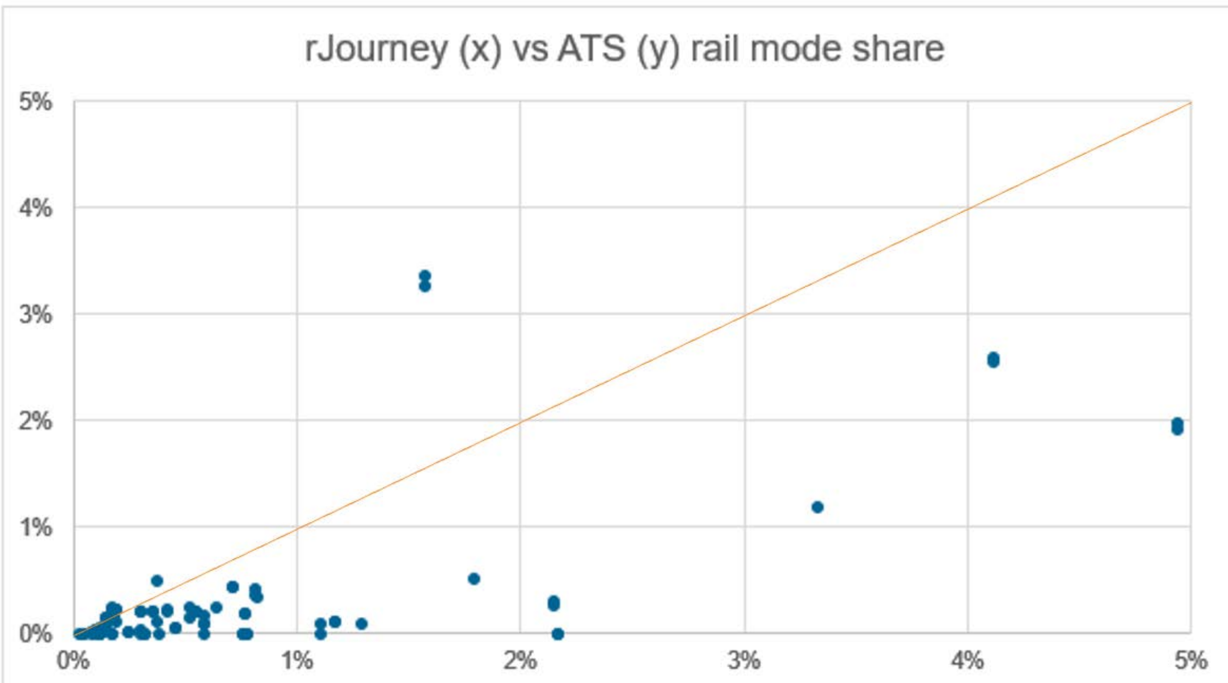


Source: FHWA

Figure 109. Bus mode shares compared to TAF by census division O-D pairs.

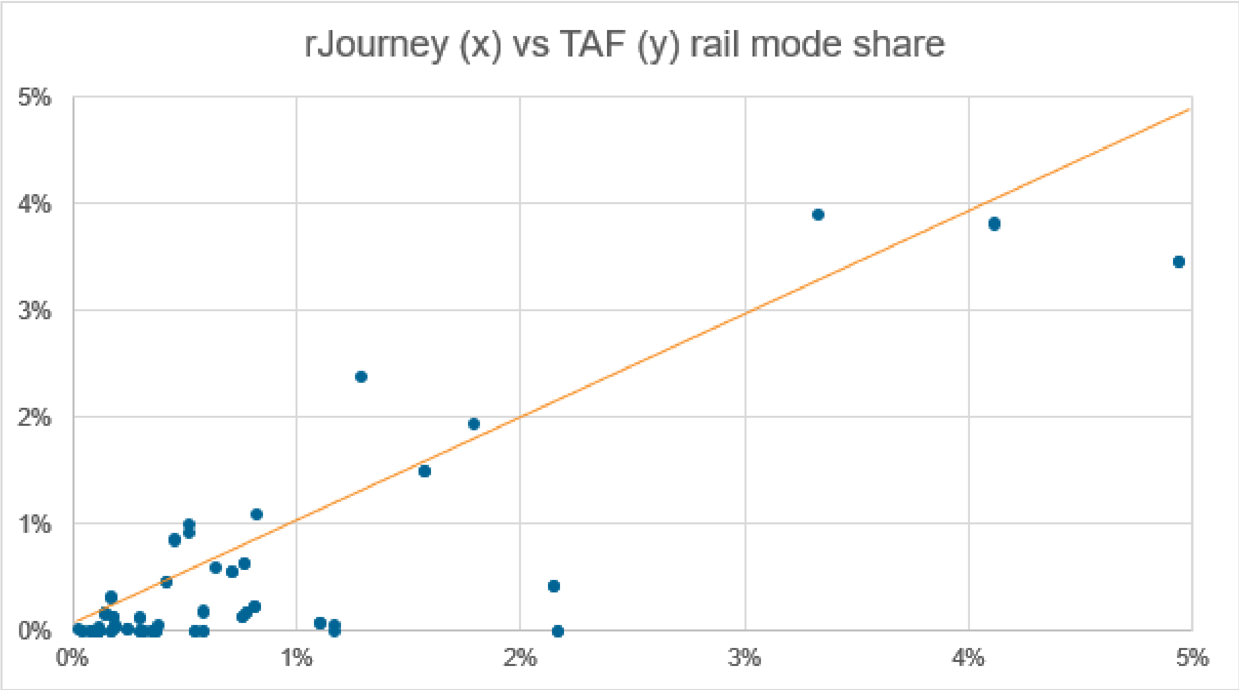
The rail mode shares by O-D census division compares the 1995 ATS, the 2011 TAF and the rJourney results in Figure 110 and Figure 111. The overall rail mode share for rJourney is 1.2 percent, compared to the 1995 ATS of 0.5 percent and the 2011 TAF of 1 percent. A strong correlation exists among the three datasets, with the TAF comparing closer to rJourney than the prior 1995 ATS for the larger rail markets.

Figure 112 and Figure 113 present the air mode shares from the 1995 ATS and 2011 TAF data compared to rJourney. The overall air mode share for rJourney is 12 percent, compared to the 1995 ATS of 17 percent and the 2011 TAF of 15 percent. A strong correlation exists between the data sources, with one exception where the 2011 TAF is showing a 0 percent mode share from New England (1) to East-North Central (3) and rJourney shows a 42 percent air mode share for this O-D pair.



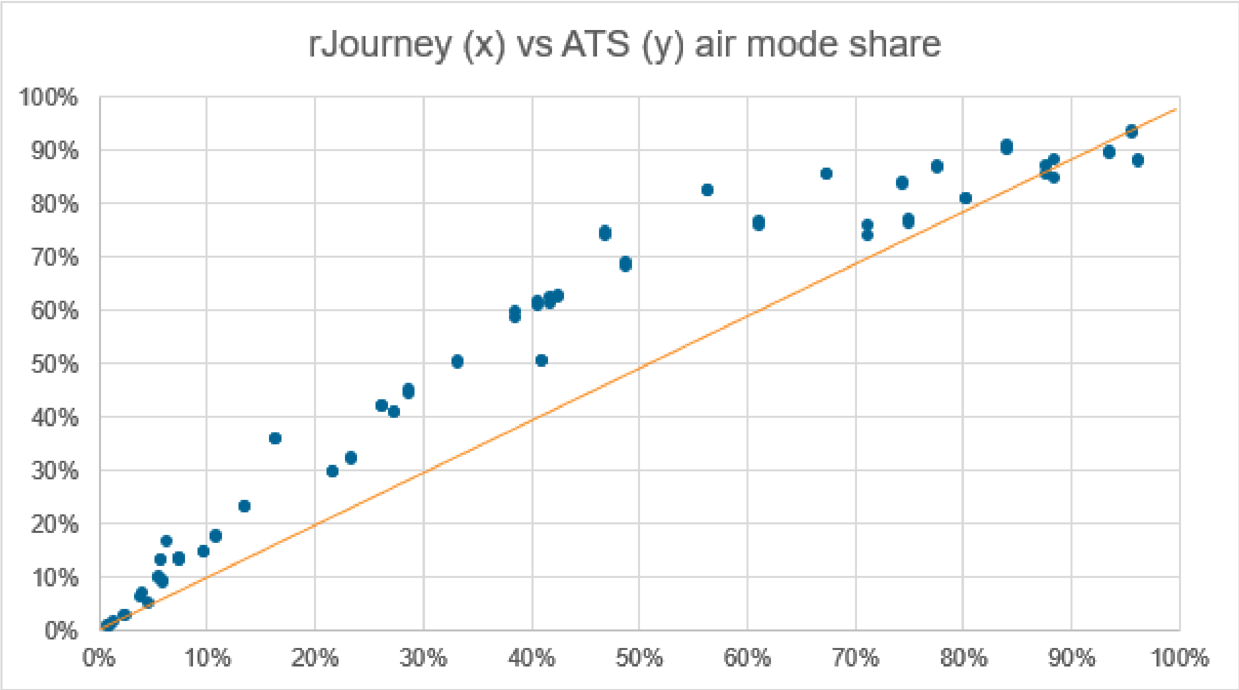
Source: FHWA

Figure 110. Rail mode shares compared to ATS by census division O-D pairs.



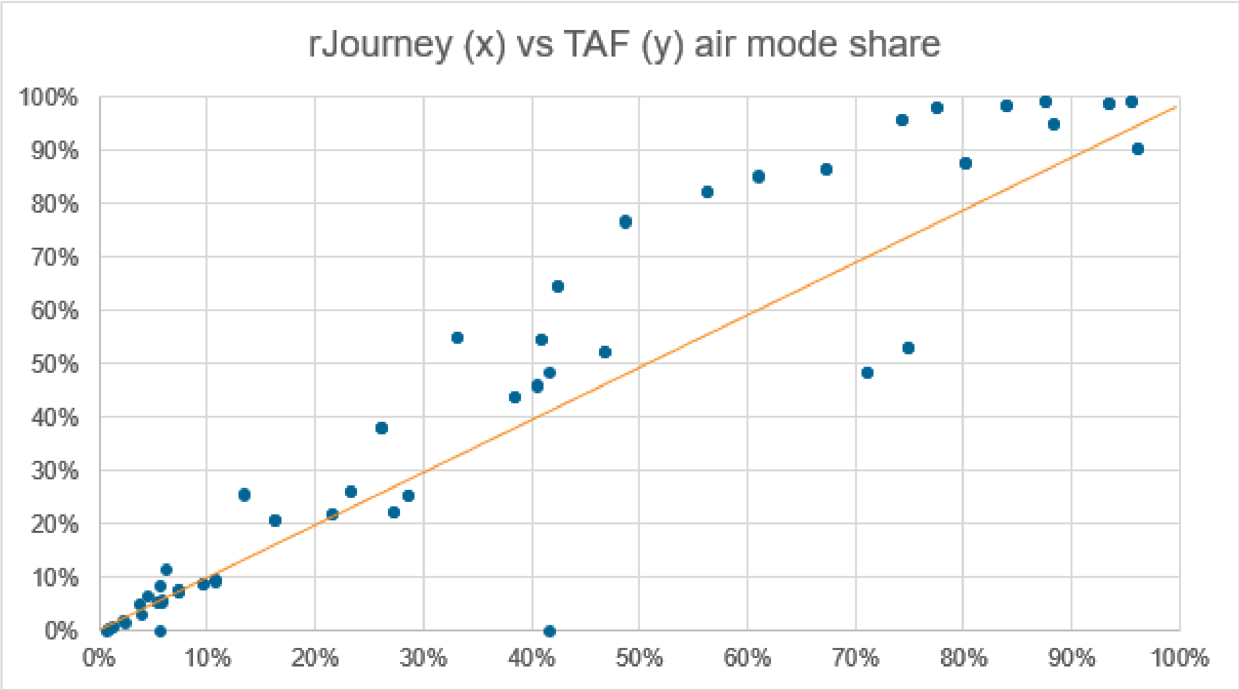
Source: FHWA

Figure 111. Rail mode shares compared to TAF by census division O-D pairs.



Source: FHWA

Figure 112. Air mode shares compared to ATS by census division O-D pairs.



Source: FHWA

Figure 113. Air mode shares compared to TAF by census division O-D pairs.

9.3 Comparison of Results

The comparison of the long-distance passenger travel demand model (rJourney) results to the 1995 ATS and 2011 TAF provide some insight on the reasonableness of the O-D patterns and mode shares. Neither of the comparison datasets provides an up-to-date observed assessment, so the comparisons serve to highlight anomalies in one or more of the datasets and to confirm reasonableness when these data align. FHWA has recently commissioned a study to develop an observed O-D trip matrix by mode from passively collected travel data, which can provide a useful dataset for future comparisons to calibrate and validate rJourney at the O-D level.

Overall, the comparisons show a reasonable alignment among the three data sources. Some differences are noteworthy given the focus on improving the rJourney model estimates. rJourney estimates higher shares of travel in larger States and lower shares of travel in the western United States (Mountain and Pacific census divisions). rJourney produces both more shorter trips (100-300 miles) and more longer trips (more than 800 miles) than the other datasets but has overall shorter trip lengths for auto and air.

The mode shares in rJourney are higher for auto and lower for air than in the other datasets. Bus and rail shares are both quite small and similar. In a few States, discrepancies indicate a simplifying assumption, which updates can address. rJourney slightly over-estimates rail and air ridership when compared to the 2011 TAF.

CHAPTER 10. SUMMARY

FHWA can use the long-distance passenger travel demand forecasting model, and adaptations to it will allow for its use by State and regional agencies across the United States. This modeling system (rJourney) is multimodal and may be useful to other Federal agencies (e.g., FRA, Federal Aviation Administration, or Federal Transit Administration).

The estimation of rJourney model components used the largest dataset that produced the most reasonable coefficients (in size, significance, and direction). This prompted using different datasets for different model components because not all datasets contained the necessary data for all model components. Recommendations to improve long-distance passenger travel demand datasets were provided in the original research report.

The calibration and validation of rJourney was completed at a national scale using available data sources. These available data sources were somewhat restricted in scope or detail, which limited comparisons of model outputs to these observed data sources. The household travel surveys for long-distance travel represent 5 of the 50 States in the United States where a national long-distance survey would have provided a more representative sample for model calibration. The traffic counts on the highway system include a large amount of short-distance passenger travel and truck travel. As a result, comparisons of long-distance traffic volumes with counts were compared for reasonableness rather than a more traditional model validation of the results. Recognizing these limitations, the models perform well compared to these available calibration and validation data sources.

rJourney is currently useful for testing national policies, based on the outcomes of the sensitivity testing conducted in the implementation phase. These sensitivity tests included changes to cost, time, and household income, and produced intuitively reasonable results. Additional sensitivity tests may be useful prior to evaluating national policies that may engage other aspects of the modeling system.

The implementation phase required additional effort to build multimodal national networks, with travel time and cost details, and a zone system, with land-use and demographic data, which may prove useful in other national planning activities. These data may also be useful to statewide or regional planning agencies that must look beyond their borders, with additional attention to areas surrounding the region or State of interest.

rJourney will also be helpful to regional and State agencies that want to represent long-distance passenger travel across their borders and test transportation investments that may affect these travelers. rJourney was designed with this objective in mind, but it does require a more detailed evaluation of the input data and a more detailed model validation surrounding the region or State of interest before these model outputs are ready to use.

GLOSSARY OF TERMS

Long-Distance Passenger Travel Demand Model: The model is comprised of the coefficients of each model component applied using the logit choice mathematic formulation.

Long-Distance Passenger Travel Demand Modeling Framework: A modeling system to predict long-distance passenger travel.

Long-Distance Passenger Travel Demand Software: The software, called rJourney, which is the programming code and graphical user interface to apply the model.

Model Calibration: The process to adjust model parameters, primarily constants, is called model calibration. This is performed to produce a better fit with observed behavioral data.

Model Estimation: Model estimation is a statistical process that produces model coefficients (or parameters) for each variable that influences the user's decision.

Model Validation: The process to compare model results to observed volume data by mode and adjust model parameters to produce a better fit with observed volume data.

Monte Carlo simulation: Monte Carlo simulation produces distributions of possible outcome values. By using probability distributions, variables can have different probabilities of different outcomes occurring.

Tour-based microsimulation model: A travel demand forecasting model that predicts travel behavior for individual people in the U.S. Tours are defined as a round trip, with one trip from the residence to the destination and a return trip from the destination to the residence of the person.

ACKNOWLEDGEMENTS

The development of the first national long-distance passenger model for the U.S. has been a collaborative effort between the project team members, the FHWA, and Volpe National Transportation Systems Center. Maren Outwater, of RSG, was the Project Manager 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 long-distance model and the development of the joint destination and mode choice models. Kaveh Shabani from RSG conducted the comparative data analysis. Ariel Oak and Aditya Gore, both from RSG, developed the graphical user interface and the user guide, respectively. Ram Pendyala and Venu Garikapati, of Arizona State University, developed and implemented the population synthesis and contributed significantly to the development of the modal networks and national zone system.

The quality of the long-distance passenger model system was enhanced by the direction and feedback provided by the FHWA leaders (Wenjing Pu and Tianjia Tang). The user guide and model documentation reports benefited from additional review of rJourney by Catherine Taylor and Matthew Keen of the Volpe National Transportation Systems Center.