TECHNICAL APPENDIX

The Air/Rail Diversion Model

Introduction

This technical appendix is an addendum to Chapter 11 of this report and is intended to provide supplementary technical information about the Air/Rail Diversion model ("the model") for those who are interested in demand models and the model development process. Chapter 11 is presented in two parts. Part One explores how the model fits into the context of the policy analysis, and the need for better tools discussed in Chapter 10. Several early examples of how the model could be used in policy analysis are presented. In Part Two, Chapter 11 reviews how the model works, and how it was developed, and summarizes the use of eight steps in model application. It presents a summary of the model at a level of detail appropriate for transportation managers, planners and analysts. This technical appendix expands significantly upon Part 2 of Chapter 11, adding a detailed explanation of the overall model structure; descriptions of the model specifications, input data, and outputs from each model component; and an overview of the software application.

Model Overview

The general structure of the model is shown in Figure 1 and is introduced in this section of the technical appendix. This structure includes a set of data inputs, represented by the blue boxes; a set of policy inputs (future service descriptions), represented by the green boxes; two primary statistical models, represented by the purple ovals; and two sets of outputs, represented by the orange boxes.

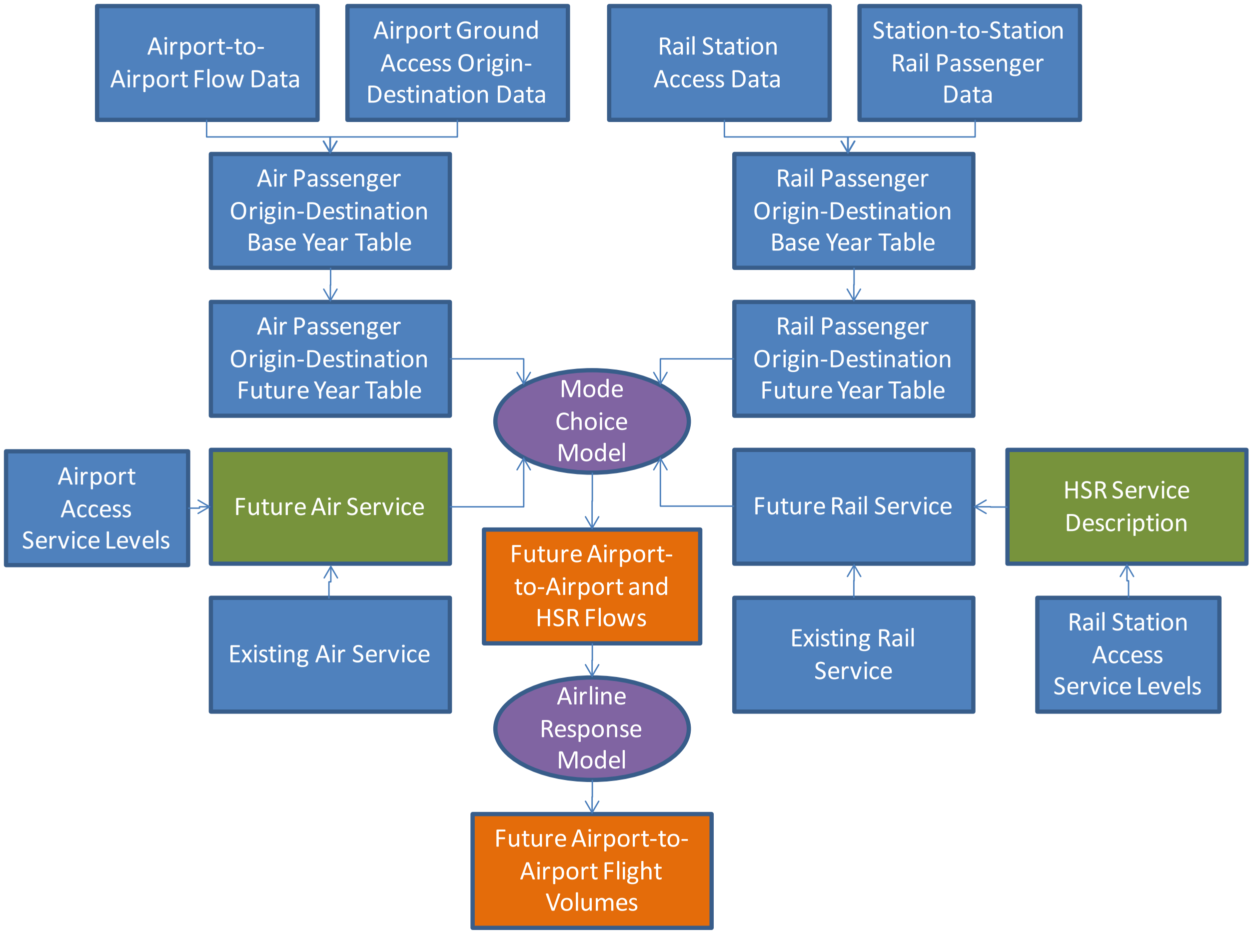

Figure 1. Overall model conceptual structure.

Representation of Travel Demand

Input travel demand is provided at a county-level resolution, while much of the model's simulation uses a more detailed spatial resolution of Census Tracts. This represents a tradeoff between the desire for additional resolution versus the reality of data availability. The observed travel demand data is currently only available at the county level for these large corridors. For example, in the output of the United States Department of Transportation's (U.S. DOT) Travel Analysis Framework project, modal O-D tables for long-distance trips are at the county level, while certain model inputs are available at more detailed spatial resolutions.

The primary purpose of the model is to understand the competition between air and improved passenger rail. The largest market of interest, and the primary focus of the model, is the short-haul market with BOTH an origin and destination in one of the corridors that form the study areas. The model compares one air service alternative with one rail alternative, for each O-D pair with a choice.

The analysis years for the model are 2008 and 2040. These years were chosen because they are consistent with the analysis years for the person-trip tables developed as part of the U.S. DOT Travel Analysis Framework project. The 2008 trip tables by air and rail were combined to reflect overall travel demand by these two modes, and they are used as a basis for base-year model calibration. The trip tables do not include any market segmentation (e.g., trip purpose or demographic data), and therefore the Research Team has used a simulation approach to characterize travel parties by drawing from observed distributions by trip purpose and demographics. While the trip tables are published in O-D format and do not explicitly identify the home end of the trip, the tables have been processed using assumptions described in later sections to infer the origin location and derive segmentation attributes based on the origin location.

Representation of Transportation Supply

The Research Team considered two options for representing the air and rail networks in the model. The first option is akin to a traditional travel model network depicting the physical linkages and service characteristics between stations and between airports. This first approach would have required that a network be included with the model that the model user could modify for scenario testing, and this network would represent the relative ease of traveling from place to place by each mode.

The second option, which the Research Team selected for model implementation, is a simpler approach whereby the network characteristics are embedded in the model in the form of predefined tabular inputs. In this approach, while it is clear what the travel time or cost is from place to place, there is no network included with the model. The predefined travel times and costs and frequencies by mode that are used as inputs to the model are viewable by the model user in the GUI; in addition, the model user has the ability to efficiently modify these inputs (rather than modifying networks) in order to develop and evaluate scenarios. This second approach was more straightforward because air networks are complex to represent and visualize, and although transit networks are clear when considering a single rail line, the inclusion of local transit access adds an unmanageable amount of complexity given the intent of the model.

Travel time and cost data were derived from several available sources. Published Amtrak schedules were used for the base condition (i.e., rail frequencies and travel times). Auto access times were derived from national roadway networks available from FHWA. Transit access times to both rail and airports were identified as a particularly difficult detail to represent, primarily because there is no corridor-wide transit network to rely on for this purpose. Instead, the Research Team used data from the Census journey-to-work data program (now the American Community Survey) to develop a comprehensive representation of the quality of transit service at a Census Tract level of spatial detail.

Airport, Station, and Mode Choice Models

The model framework makes use of three choice models to select a preferred airport pair, a preferred station pair, and the main mode for the trip, comparing air versus rail.

For the airport and station choice models, the Research Team used models developed for the West of Hudson Regional Transit Access Study, a study sponsored by the Port Authority of New York and New Jersey and MTA’s Metro-North Railroad to evaluate improved transit connections to Stewart Airport. These models were applied to consider a set of possible airport pairs for the trip being modeled and a set of stations pairs. The model also permitted the selection of a preferred pair in each case, based on the characteristics of the access and egress trips from the airports and stations at either end of the trip and the relative levels of service on the air or rail service between the airports and stations.

For the mode choice models, the team identified two existing models that have similar structures and whose coefficients have been estimated using recent data: a model developed by RSG for the most recent Toronto-Montreal-Quebec City corridor HSR study under contract to Transport Canada and the Ministries of Transport for Ontario and Quebec (2009/2010), and the California HSR model developed by Cambridge Systematics, modified to represent the air versus improved passenger rail trade-off elements. The mode choice model that was implemented is derived from the California HSR model.

Airline Response Model

The airline response model predicts how airline schedules will respond to the introduction of improved passenger rail service on the Northeast and California Corridors. The response may include several elements. First, to the extent that improved passenger rail captures traffic that previously traveled by air, airlines may reduce seat capacity on affected segments. Second, either as a result of offering reduced seat capacity, or as a direct response to competition from improved passenger rail, the distribution of capacity among different types of airlines—mainline network carriers, regional affiliates, and low cost carriers—may change. Finally, as either as a direct response to improved passenger rail competition, or as a secondary response to diminished traffic, carriers may change their fleet mixes in corridors with improved passenger rail service.

Model Implementation Approach

The model was developed and implemented using a combination of Microsoft Excel and custom scripts developed in the R open-source statistical analysis software. The following is a step-by-step sequence that describes the model flow as implemented in the model's code; each step is expanded upon in subsequent sections with discussion of data sources, model structure and specification, outputs, and sensitivity to changes in the inputs.

1. Combine the rail and air county-to-county flows (2008 and 2040 trip tables are preprocessed as round trips from county to county) to create total demand for rail and air service for each county pair.
2. Enumerate a population of travelers from the total demand data, sample a selection of travel parties to simulate in the remaining models, and simulate a trip purpose and party size for each traveler in the list.
3. Allocate each travel party to specific origin and destination Census tracts using an allocation model.
4. Simulate an income category and vehicle availability for the travel party based on their trip purpose for the travel party and their origin Census Tract.
5. Choose the best airport and rail station for the origin and destination ends of the trip for each party using an airport choice and a station choice model.
6. Apply the main mode choice model to choose between traveling by air or rail, based on the characteristics of the travel party, the accessibility of the airports and rail stations at origin and destination end of the trip, and the level of air and rail service between them.
7. Apply the airline response model to calculate a likely response to the rail service scenario in terms of air service changes by the airlines.
8. Generate data in spreadsheet summary format.

Study Areas

The model has been applied in two study areas—the Northeast Corridor ("East Coast") and California ("West Coast")—where there is considerable availability of both air and rail modes, meaning that many long-distance trip travelers have a reasonable choice between the modes. Figure 2 and Figure 3 show the East and West coast study areas respectively; they also show the county geography and the airport and rail station locations that are represented in the model. The data summaries presented in this technical appendix relate to these two study areas. Scenario test results for these two study areas are presented in Chapter 11. The geographic extent of the two study areas includes counties close to the major north east and California rail corridors since competition will be greatest for travelers starting and ending their trips near to existing or planned future rail lines. There are some notable differences between the two study areas in terms of the relative locations of population centers and in current rail service and ridership. In the Northeast Corridor, there is a series of large metropolitan areas from Boston to Washington, including New York, Philadelphia, and Baltimore, served by relatively fast and frequent rail service that competes with, and for some city pairs carries more passengers than, air service. In California, the long distance between the two largest metropolitan areas (Los Angeles and San Francisco) and comparatively slow and low frequency rail service means that air service has a more dominant share of the overall air and rail market.

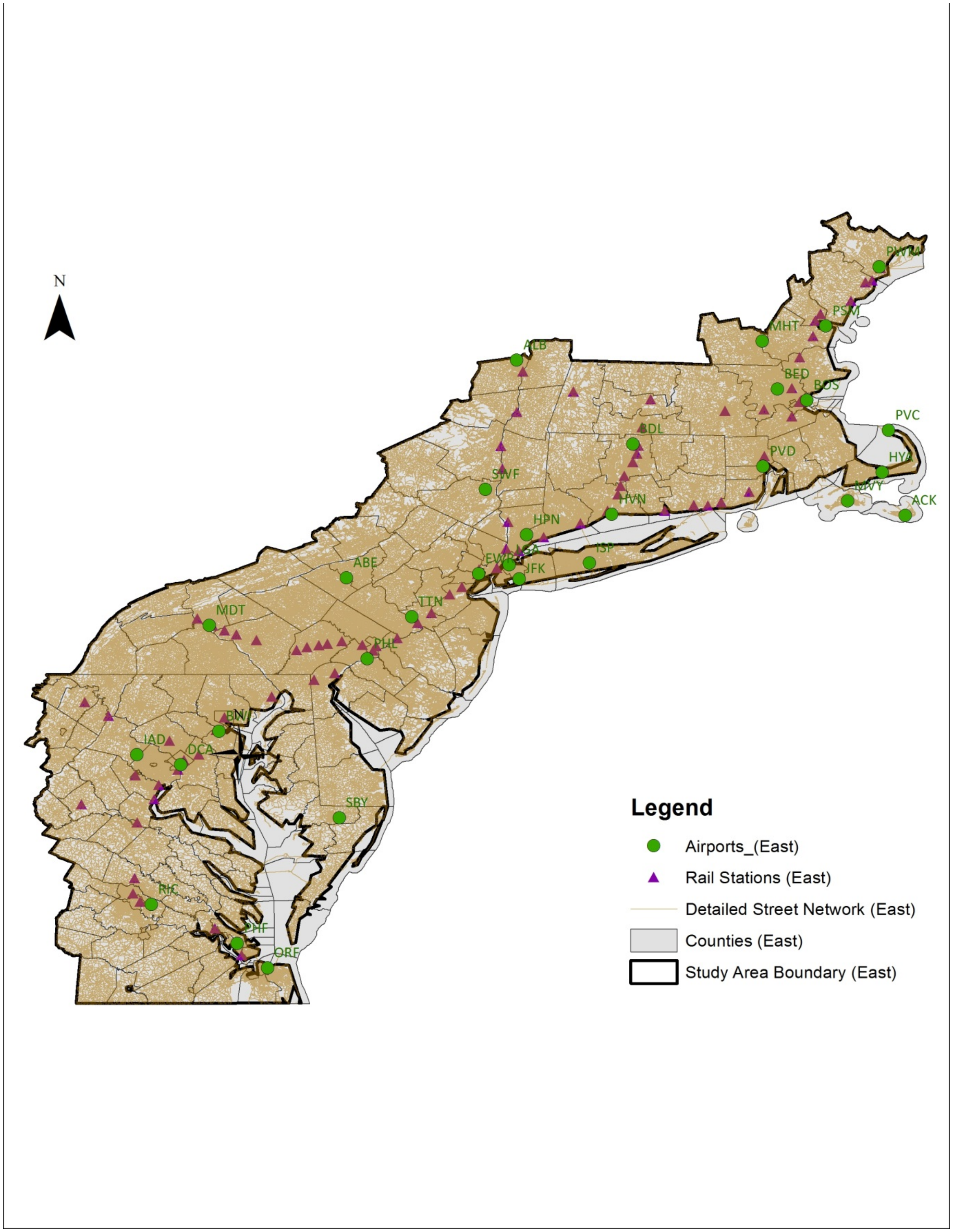

Figure 2. East coast study area.

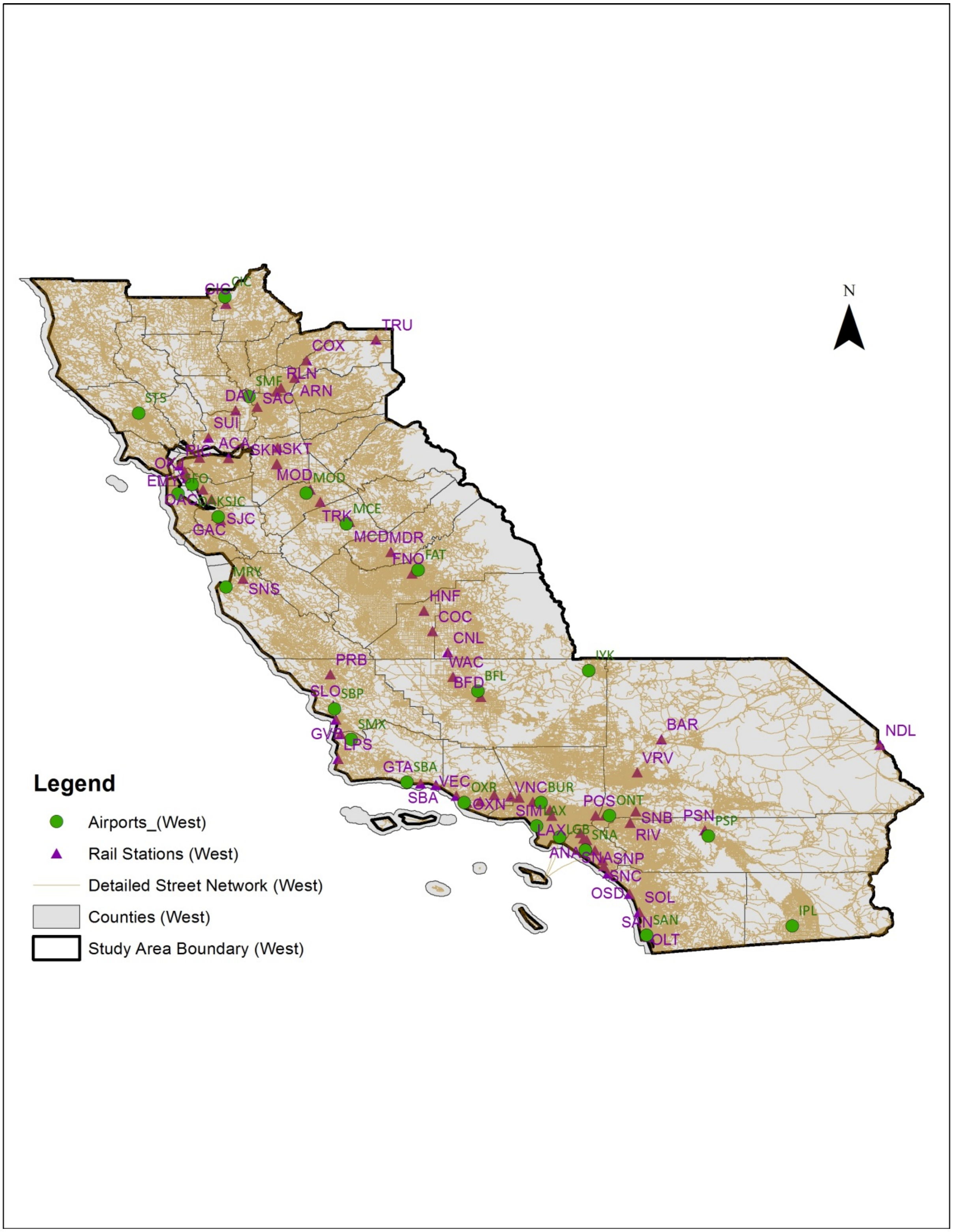

Figure 3. West coast study area.

Model Components

The following sections describe each of the models components in the sequence that they are applied in the model. For each model components, there is a discussion of data sources, model structure, model specification, and outputs.

Combine Rail and Air County-to-County Flows

Overview

The model's first step uses preprocessed county-to-county O-D flows for air and rail to create the total demand for air and rail travel in the two study areas for both 2008 and 2040.

Data Sources

The data source that describes county-to-county demand for travel by air and rail in the two study areas (the East Coast study area, centered on the Northeast Corridor, and the West Coast study area, centered on California between San Diego and the Bay Area/Sacramento) is taken from work conducted for the U.S. DOT's Travel Analysis Framework to develop 2008 and 2040 county-to-county O-D trip tables for air, rail, auto, and bus—although the auto and bus tables are not being used in this research.

The data used to develop the trip tables were taken from the U.S. DOT DB1B airport-to-airport O-D data, Amtrak passenger station-to-station flows provided by Amtrak, and survey data describing airport ground access (from regional and single airport air passenger surveys) and rail station access (from surveys collected in California by California High Speed Rail Authority.

The data for trips within each of the two study areas from the 2008 and 2040 tables for air and rail were extracted. The 2040 air tables incorporate forecasts produced by FAA at the airport-to-airport level of detail. These unpublished forecasts were provided by FAA in November 2012 as part of the U.S. DOT Travel Analysis Framework project and add additional detail beyond the published Terminal Area Forecasts that were available at that time, The rail tables have been augmented to more accurately account for transfers to Amtrak's connecting bus services in California, using data provided by the Federal Railroad Administration (FRA).

For use in the model, the trip table data was converted from one-way trips to roundtrip itineraries (i.e., origin county to destination county to origin county). This involves inferring round trip directionality for each pair of one-way trips.

In the case of air, the majority of the DB1B coupon data describes roundtrips. The Research Team has asserted the assumption that the itinerary starts in the traveler's home area, with the middle break in the itinerary occurring at the traveler's destination and the itinerary ending at home. An example itinerary included in the DB1B data can be found in Table 1. This itinerary represents a roundtrip from Asheville Regional Airport (AVL) in North Carolina to Ted Stevens Anchorage International Airport (ANC) in Alaska. Each direction of the trip had two intermediate stops—Douglas Airport (CLT) in Charlotte, North Carolina, and McCarran International (LAS) in Las Vegas, Nevada. The Break value of X indicates the directional O-D break in the itinerary.

The remaining one-way trips in the DB1B data have been matched with a one-way trip in the opposite direction to create roundtrips. These have been given a direction randomly, in proportion to the directional split of the roundtrips. The results of this data processing exercise are shown in Table 2.

Table 1. Example itinerary from the airline origin and destination survey.

COUPON	ORIGIN	DEST	BREAK
1	AVL	CLT	
2	CLT	LAS	
3	LAS	ANC	X
4	ANC	LAS	
5	LAS	CLT	
6	CLT	AVL	X

Since the Amtrak station-to-station flows do not capture directionality like the DB1B ticket data, a split between visitors and residents of 50% for each O-D pair has been applied to the rail data. In the two study areas, most O-D pairs in the air data are relatively balanced, and this appears to be a reasonable assumption.

Approach

This step of the model imports the county-to-county demand data files for 2008 and 2040 and adds the air and rail round trip demand together to create the total demand for air and rail travel in 2008 and 2040.

Table 2 summarizes the number of roundtrips by origin state (for the East Coast study area) and region within California (for the West Coast study area) in 2008, while Table 3 shows a similar tabulation for 2040. Note that the round trips are those that are within each study area and that the study area includes partial states (i.e., the coastal areas closer to, and including, the major metropolitan areas). On the East Coast, the states in the center of the region—New York, New Jersey, and Pennsylvania—generate large numbers of rail trips relative to air trips; Figure 4 shows the low air mode shares in those states for trips within the study area compared to other states. Table 4 shows the growth by state/region between 2008 and 2040 for air and rail and total trips. The growth rates for total trips vary between 36% and 74% by state/region.

Table 2. Air, rail, and total roundtrips by state within the study areas (2008).

State/Region*	Air	Rail	Total
Connecticut	186,755	406,908	593,663
Delaware	12,737	104,874	117,611
District of Columbia	154,698	853,280	1,007,978
Maine	96,752	55,334	152,086
Maryland	596,054	606,369	1,202,423
Massachusetts	1,032,723	700,014	1,732,737
New Hampshire	150,059	56,536	206,595
New Jersey	159,937	522,226	682,163
New York	735,790	1,856,841	2,592,631
Pennsylvania	225,271	1,375,768	1,601,039
Rhode Island	142,715	211,785	354,500
Virginia	674,518	254,002	928,520
West Virginia	3,123	1,979	5,102
East Coast Total	**4,171,132**	**7,005,916**	**11,177,048**
Los Angeles Region	2,721,573	1,019,987	3,741,560
Sacramento Region	733,654	300,976	1,034,630
San Diego Region	604,345	384,345	988,690
San Francisco Region	2,769,256	427,140	3,196,396
Remainder of California	233,278	385,292	618,570
West Coast Total	**7,062,106**	**2,517,740**	**9,579,846**

* The study area includes partial states, i.e. generally the coastal areas closer to and including the major metropolitan areas.

Table 3. Air, rail, and total round trips by state within the study areas (2040).

State/Region*	Air	Rail	Total
Connecticut	268,048	562,465	830,513
Delaware	25,922	152,559	178,481
District of Columbia	280,908	1,265,506	1,546,414
Maine	157,657	79,027	236,684
Maryland	992,693	907,647	1,900,340
Massachusetts	2,046,045	968,468	3,014,513
New Hampshire	221,086	80,237	301,323
New Jersey	311,659	737,029	1,048,688
New York	1,008,999	2,569,400	3,578,399
Pennsylvania	424,709	1,933,877	2,358,586
Rhode Island	192,162	291,717	483,879
Virginia	1,074,605	413,035	1,487,640
West Virginia	5,164	3,223	8,387
East Coast Total	**7,009,657**	**9,964,190**	**16,973,847**
Los Angeles Region	4,323,228	1,530,303	5,853,531
Sacramento Region	1,100,066	557,214	1,657,280
San Diego Region	971,976	587,270	1,559,246
San Francisco Region	4,375,406	866,480	5,241,886
Remainder of California	339,415	739,936	1,079,351
West Coast Total	**11,110,091**	**4,281,203**	**15,391,294**

* The study area includes partial states (i.e., the coastal areas closer to, and including, the major metropolitan areas)

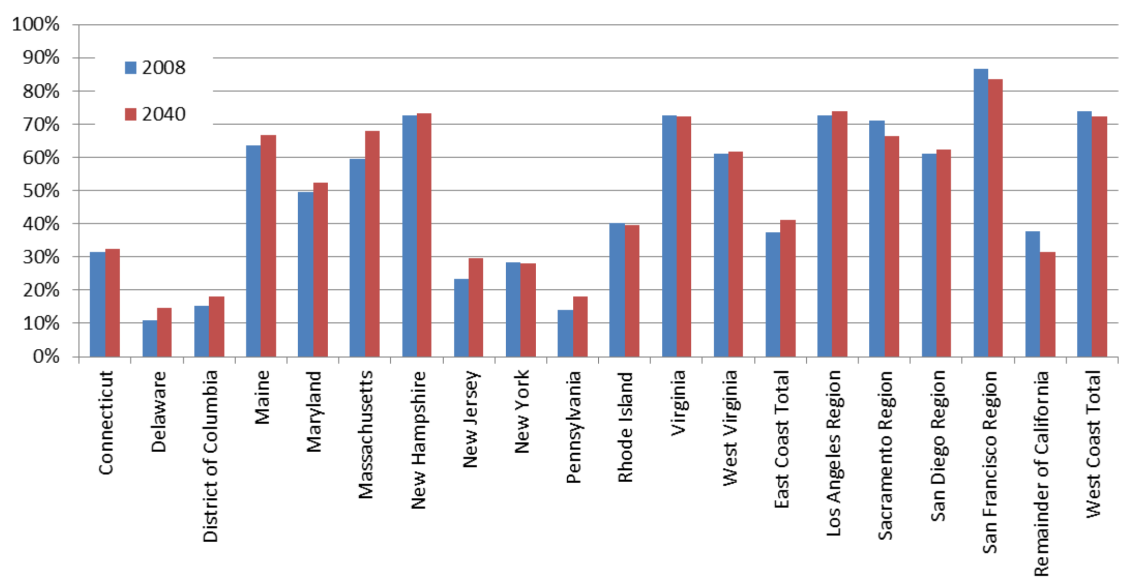

Figure 4. Air mode share by state/region for 2008 and 2040.

Table 4. Air, rail, and total round trips by state within the study areas (growth from 2008 to 2040).

State/Region*	Air	Rail	Total
Connecticut	44%	38%	40%
Delaware	104%	45%	52%
District of Columbia	82%	48%	53%
Maine	63%	43%	56%
Maryland	67%	50%	58%
Massachusetts	98%	38%	74%
New Hampshire	47%	42%	46%
New Jersey	95%	41%	54%
New York	37%	38%	38%
Pennsylvania	89%	41%	47%
Rhode Island	35%	38%	36%
Virginia	59%	63%	60%
West Virginia	65%	63%	64%
East Coast Total	**68%**	**42%**	**52%**
Los Angeles Region	59%	50%	56%
Sacramento Region	50%	85%	60%
San Diego Region	61%	53%	58%
San Francisco Region	58%	103%	64%
Remainder of California	45%	92%	74%
West Coast Total	**57%**	**70%**	**61%**

* The study area includes partial states, i.e. generally the coastal areas closer to and including the major metropolitan areas

Enumerate a Population of Travelers, Select a Sample, and Simulate Trip Purpose and Party Size

Overview

This step of the model converts the summary of demand from each county to total trips and then to an enumerated list of travelers. The model then selects a sample of them for further simulation. For each selected traveler, the model simulates a trip purpose and party size.

Data Sources

A joint trip purpose and party size distribution was derived for air travelers from a sample of air passenger surveys, with the intention that during a model run both trip purpose and party size would be simulated by drawing from the joint distribution. The air passenger surveys are those in the study areas with trips that have both ends within a study area from the set used for the development of the air trip table data. In the East Coast study area, the regional survey in New England included coded information on the destination airport, and so trips could be identified as being within the study area, while in the West Coast study area, surveys in the Bay Area and at LAX and ONT in the Los Angeles had similarly data, as shown in Table 5.

Prior to deriving the distributions, the data from each survey were weighted using the DB1B data up to the passenger volumes traveling on each of the represented O-D pairs and then normalized by the survey sample size for each airport so as not to over represent smaller surveys samples in the analysis. The business and non-business distributions derived from the surveys are shown in Table 6 for both the raw survey data and following weighting. The weights were calculated as follows:

Survey Record Weight = $(DB1B_{ij}/n_{ij})*n/N$

Where:

$DB1B_{ij}$ is the passenger volume traveling on the O-D pair according to the DB1B data,

n_{ij} is the number of survey records for that O-D pair

n is the total sample size for that airport

N is the total sample size for all of the ground access surveys in the study area

Table 5. Air passenger survey metadata.

SURVEY	YEAR	AIRPORT(S)	SAMPLE SIZE*
New England airports survey	2004	BOS	818
		MHT	598
		PVD	589
		BDL	404
		PWM	196
		BED	62
		HVN	46
		PSM	1
MTC Airline Passenger survey	2002	OAK	933
		SJC	876
		SFO	292
Los Angeles International Airport passenger survey	2006	LAX	2,337
Ontario International Airport passenger survey	2001	ONT	1,400

*Sample sizes are of survey records describing air passenger trips with both trip ends in the study area

Table 6. Party size distributions derived from air and rail survey data.

	BUSINESS		NONBUSINESS	
Partysize	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
1	71%	73%	50%	50%
2	20%	19%	33%	32%
3	4%	3%	9%	9%
4	2%	2%	5%	6%
5 or more	2%	2%	3%	4%
Total	100%	100%	100%	100%

The Amtrak station-to-station flows do not capture party size or trip purpose; as a result, rail survey data from California was used to identify a joint trip purpose and party size distribution for Amtrak trips. In comparison with the air data, the distributions are relatively similar for business passengers, with the most noticeable difference being slightly more large parties (three or more passengers) traveling on the train for non-business purposes.

Approach

The first aspect of this model step is the conversion of the summary of demand from county-to-county total trips to an enumerated list of travelers. This results in a list with one row for each traveler containing just the origin county and the destination county (the midpoint of the round trip). This list is used as the basis for the disaggregate simulation approach used in the model, where the travel choices of each individual traveler are simulated.

Running a simulation that includes all of the travelers can result in unreasonably long model run times without necessarily adding precision to the results. Therefore, the model provides users with the opportunity to choose a sampling fraction; this allows users to sample only a proportion of the total list of travelers for simulation (with the results later weighted). The model user can choose any percentage from 1% to 100%. The approach used is to scale and then round to integers the county-to-county demand using the sampling fraction prior to enumeration of the list.

This model step then simulates a trip purpose and party size for each traveler by drawing from a joint distribution of trip purpose and party size derived from air passenger survey data and rail survey data. During a model run, an overall joint trip purpose and party size distribution is created by combining the input distributions for air and rail, weighted based on the observed number of trips by mode. A random draw from the uniform distribution is made for each traveler, and this is used to select a value from the joint cumulative distribution of trip purpose and party size.

Once a party size has been assigned to all of the travelers in the enumerated list, a further sampling takes place to choose representative records for each party of more than one person. Since the unit of simulation is travel parties, samples of 50% of a party of 2 travelers, 33.3% of a party of 3 travelers, 25% of a party of 4 travelers, and 20% of a party of 5 or more travelers are added to all of the enumerated party of 1 traveler. The rate of 20% was chosen for a party of 5 or more travelers for simplicity due to the lack of information about the actual average party size of that category.

The observed joint distribution of party size and trip purpose that is constructed by combining the distributions observed for air and rail trips is shown in Table 7. The joint distribution is split 41% business and 59% non-business, with smaller party sizes for business trips than non-business trips.

Table 7. Joint distribution of trip purpose and party size (percentage of passengers).

Purpose	Party Size	Percentage
Business	1	31.0%
Business	2	6.4%
Business	3	1.7%
Business	4	0.8%
Business	5 or more	0.8%
Non Business	1	21.3%
Non Business	2	20.7%
Non Business	3	6.2%
Non Business	4	6.5%
Non Business	5 or more	4.6%

The following series of tables shows extracts from the data tabulations produced by the model during this step to illustrate the process of enumerating a list of travelers, selecting a sample of them for simulation, and assigning a party size and purpose to them. Table 8 shows six example county pairs (with origins in different New York state counties and a destination of District of Columbia). In this example, the model user has selected a 2% simulation rate, and the input number of air and rail round trips is combined, and the number of trips to be simulation is calculated as 2% of that total number of round trips. Overall, a 2% simulation rates is 222,000 round trips for the East Coast application in 2008.

Table 8. Simulation sample size for six example origin destination county pairs, using a 2% simulation rate.

Origin County	Destination County	Air Trips (from Trip Tables)	Rail Trips (from Trip Tables)	Simulation Sample (2% rate)
Kings	DC	8,930	4,051	260
Nassau	DC	9,514	717	205
New York	DC	29,790	343,826	7,472
Orange	DC	1,048	335	28
Putnam	DC	964	736	34
Queens	DC	9,737	4,791	291

Table 9 shows 10 records from the synthetic population of individual travelers, extracted from the 7,472 records for the O-D pair of round trips from 36,061 (Manhattan) to 11,001 (District of Columbia). Following enumeration to create a table with the correct number of rows, a purpose and party size are drawn for each record from the joint distribution shown above and entered into the Purpose and Party Size fields respectively. A draw from the uniform distribution is then assigned to each record (the "Draw" field). Finally, the "Keep" field is populated to select the correct number of records so that the records represent parties instead of individual passengers. All passengers with a party size of one are kept. For passengers with a party size of 2, records with a draw of greater than 0.5 are kept in order that half of the records are retained because there are half the number of size 2 parties as there are passengers in size two parties. Similarly, one third records for a party size of 3 (those with a draw greater than 0.67); one quarter of party size of 4 records (those with a draw greater than 0.75); and one fifth of party size of 5 records (those with a draw greater than 0.8) are retained.

Table 9. Synthetic population of individual travelers with simulated trip purpose and party size for New York County to District of Columbia.

Sequence	Purpose	Party Size	Draw	Keep
1	NonBusiness	5 or more	0.46376043	0
2	Business	1	0.75538148	1
3	Business	1	0.6910377	1
4	NonBusiness	2	0.77858352	1
5	Business	3	0.14426643	0
6	NonBusiness	2	0.3747489	0
7	NonBusiness	4	0.69179226	0
8	NonBusiness	2	0.62371052	1
9	NonBusiness	4	0.05444775	0
10	NonBusiness	2	0.26284101	0

Table 10 shows tabulations from the complete synthetic population for the East Coast using a 2% simulation rate. In this case, the synthetic population is comprised of 159,000 rows representing 159,000 parties and 222,000 individual travelers. The lower two tabulations show the distribution of parties and individuals across the joint distribution of purpose and party size. It can be seen that the final tabulation of travelers matches the input joint distribution of purpose and party size, demonstrating that this step of the model is producing the synthetic population as intended.

Table 10. Synthetic population of individual travelers and parties summarized by trip purpose and party size.

Parties	1	2	3	4	5 or more	Total
Business	68,855	7,050	1,227	433	413	77,978
NonBusiness	47,274	23,079	4,647	3,554	2,051	80,605
Total	116,129	30,129	5,874	3,987	2,464	158,583
Travelers	1	2	3	4	5 or more	Total
Business	68,855	14,100	3,681	1,732	2,065	90,433
NonBusiness	47,274	46,158	13,941	14,216	10,255	131,844
Total	116,129	60,258	17,622	15,948	12,320	222,277
Percent Parties	1	2	3	4	5 or more	Total
Business	43.4%	4.4%	0.8%	0.3%	0.3%	49.2%
NonBusiness	29.8%	14.6%	2.9%	2.2%	1.3%	50.8%
Total	73.2%	19.0%	3.7%	2.5%	1.6%	100.0%
Percent Travelers	1	2	3	4	5 or more	Total
Business	31.0%	6.3%	1.7%	0.8%	0.9%	40.7%
NonBusiness	21.3%	20.8%	6.3%	6.4%	4.6%	59.3%
Total	52.2%	27.1%	7.9%	7.2%	5.5%	100.0%

Allocation to Census Tracts

Overview

The simulation sample is comprised of a set of travel parties with round trip itineraries. The origin county at the start of the travel parties' outbound trips is known, along with the destination county of their outbound trip. This model step then allocates each travel party in the simulation sample to specific origin and destination Census tracts within their origin counties and destination counties.

Data Sources

The data used by this model step describes the distribution of population and employment within each of the counties in the study area and are available from public sources:

- U.S. Census geographical relationships between the county and sub-county summary levels of Block, Block Group, and Census Tract.
- 2010 U.S. Census population at the Census Tract level.
- 2008 employment by industry by Census Tract from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics data (http://lehd.ces.census.gov/).
- 2008 employment by ZIP Code from the U.S. Census County Business Patterns data (http://www.census.gov/econ/cbp/)

The population data were obtained from the 2010 U.S. Census. The study areas comprise 13,273 tracts in the East Coast study area and 7,909 tracts in the West Coast study area. Table 11 summarizes the population in the parts of the study area in each state.

Total employment and hospitality sector employment (NAICS 72) of each Census Tract for 2008 were obtained from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics program for use in applying this model. The dataset used was LEHD Origin-Destination Employment Statistics (LODES). Table 11 summarizes the total employment and hospitality employment in the parts of the study area in each state. At the time of processing these data, the Census Bureau had not made available the data for Massachusetts, which is the final state to be added to this relatively new data source, and instead ZIP code level employment from the U.S. Census County Business Patterns data were used.

Table 11. Population, employment, and hospitality employment by state.

State/Region*	Population	Total Employment	Hospitality Employment
Connecticut	3,574,097	1,575,309	106,126
Delaware	897,934	392,294	31,489
District of Columbia	601,723	621,524	54,762
Maine	478,805	224,881	20,454
Maryland	5,668,368	2,383,588	184,822
Massachusetts	6,547,629	2,924,913*	232,603*
New Hampshire	819,087	372,312	28,024
New Jersey	8,791,894	3,732,237	268,508
New York	13,874,816	6,025,444	398,364
Pennsylvania	7,396,902	3,196,508	233,356
Rhode Island	1,052,567	435,352	39,850
Virginia	5,939,131	2,595,203	216,880
West Virginia	175,208	46,465	5,041
East Coast Total	**55,818,161**	**21,601,117**	**1,587,676**
Los Angeles Region	17,877,006	7,037,049	613,704
Sacramento Region	2,414,783	908,149	74,520
San Diego Region	3,095,313	1,230,279	129,459
San Francisco Region	7,468,390	3,263,252	278,910
Remainder of California	5,834,108	1,846,391	155,352
West Coast Total	**36,689,600**	**14,285,120**	**1,251,945**

*LODES data has not been released for Massachusetts as of July 2013, and instead ZIP code County Business Patterns data were used

Approach

As described above, the input demand is represented as county-to-county flows. To add additional spatial detail to the model to capture variation in the accessibility of stations and airports across counties (which are sometimes very large, as in the case of the West Coast), it uses Census Tracts as the spatial unit for airport and rail station accessibility measures. The increased spatial resolution that can be used in the model as a result of converting to a disaggregate approach means that the airport and rail station accessibility measures for a particular travel party will be more realistic. For example, an auto access time specific to travel between a Census Tract and an airport or rail station will be more realistic than using an average value for travel from the entire county.

Using this more detailed spatial resolution means allocating each travel party from a county to a Census Tract at each of the origin and destination ends in order to select the correct accessibility measures for the travel party. At the origin (home end) of the round trip, the proposed approach to allocation is using Census Tract population since most trips begin at home.

At the destination end of the outbound trips, the destinations vary between other homes (typically for trips with a purpose of visiting friends and family), hotels, and places of work. To model the allocation at the destination, the model uses a combination of Census Tract population and Census Tract employment in the hospitality sector, following an approach similar to one used during the U.S. DOT Travel Analysis Framework project to develop county-to-county origin-destination trip tables for air and rail. The allocation calculation is as follows:

1. For each Census Tract, input variables include "Pop," population in units of number of people and "Emp_Hosp," hospitality employment in units of number of jobs.
2. A variable that is a function of the population and hospitality employment is calculated for each Census Tract:
   FPopHospEmp = Pop * 0.144 + Emp_Hosp * 0.172
   Where 0.144 is the county population coefficient from the visitor trip allocation model
   And 0.172 is the county employment in hospitality sector coefficient from the visitor trip allocation model
3. For each county, the total FPopHospEmp is calculated by summing over the Census Tracts within the county.
4. For each Census Tract, its proportion of county visitor trips is calculated as Census Tract FPopHospEmp/County Total FPopHospEmp.
5. This set of Census Tract proportions is used as a set of sampling probabilities for allocating each visiting traveler group's trip end to that county to a specific destination Census Tract in the simulation.

The Research Team recognizes that this simple allocation approach omits potentially important explanatory variables that are known to affect the propensity of people to make long distance trips, such as household income. More in depth model estimation work with suitable data would be helpful in improving the allocation approach and models.

Simulate income categories and vehicle availability

Overview

The simulation sample that is an input to the mode choice models is comprised of a set of travel parties with round trip itineraries where the counties and Census Tract at the start and end of their trips, and the trip purpose are known. This step in the model simulates the household income category and vehicle availability for the travel party given their trip purpose to allow the subsequent choice models to be applied correctly to each travel party.

Data Sources

The additional data used by this model step describes the distributions of household income in 2008 and vehicle availability among air and rail travelers as well as the household income and vehicle ownership of households in the air and rail travelers home Census Tracts. The data sources are as follows:

- Air passenger surveys for the New York region, New England, the Baltimore-Washington area, Los Angeles, the Bay Area, and San Diego.
- Travel survey data collected in California by California High Speed Rail Authority.

- American Community Survey data for 2006–2010 at the Census Tract level describing household income and vehicle ownership, available from the U.S. Census Bureau (http://www.census.gov/acs/www/).

Additional data that are used to more completely describe the characteristics of the travel parties were developed from the air passenger surveys and rail survey data from California described herein. Table 12 shows the distribution derived from the survey data. In general, travelers have vehicles available at the home end of the trip. The income distribution is skewed higher among business travelers.

Table 12. Joint distribution of trip purpose, household income in 2008, and vehicle availability at the home end of a round trip.

Purpose	Household Income in 2008	Vehicle availability	Percentage
Business	$125,000 or more	No	0.4%
Business	$125,000 or more	Yes	21.7%
Business	$75,000 to $125,000	No	0.2%
Business	$75,000 to $125,000	Yes	10.6%
Business	$25,000 to $75,000	No	0.5%
Business	$25,000 to $75,000	Yes	9.3%
Business	<$25,000	No	0.1%
Business	<$25,000	Yes	0.5%
Non Business	$125,000 or more	No	0.3%
Non Business	$125,000 or more	Yes	16.7%
Non Business	$75,000 to $125,000	No	0.3%
Non Business	$75,000 to $125,000	Yes	13.1%
Non Business	$25,000 to $75,000	No	0.8%
Non Business	$25,000 to $75,000	Yes	21.3%
Non Business	<$25,000	No	0.7%
Non Business	<$25,000	Yes	3.4%

Approach

The disaggregate nature of the simulation model means that supporting market segmentation is possible. Several types of segmentation of the demand are incorporated, which each require the simulation of corresponding characteristics for each of the simulated travel parties; this must be done ahead of the subsequent steps that apply choice models for each travel party conditional on the market segment that the travel party falls in. At this stage in the model, the party size and trip purpose is already known. In this step, the household income category and vehicle availability for the party are developed conditional on the trip purpose of the party. The general approach is to draw from the joint distributions of those variables observed from the air passenger and rail travel data.

Airport Choice and Station Choice

Overview

This step in the model applies airport and station choice models to each travel party to select likely airports and rail stations for their origins and destinations. The simulation sample that is input into this step is comprised of a set of travel parties with roundtrip itineraries where their origin county and Census Tract at the start of their outbound trip (and their final destination at the end of their return trip) are known, along with the destination county and Census Tract of their outbound trip. In addition, several characteristics of the party are known: party size; trip purpose (business versus non-business); income category; and vehicle availability at the origin end of their trip.

Data Sources

The additional data required by the airport and station choice models are as follows:

- Aviation data—flight availability and fares by airport pair and airport enplanements. These data were developed using publicly sourced aviation data, including T100, on-time performance data, and DB1B.
- Rail data—data items parallel to the aviation data using Amtrak O-D data and Amtrak schedules.
- Transit accessibility measures for each Census Tract—derived from CTPP journey to work data.
- Highway accessibility to airports and stations—developed by applying shortest path algorithm from each Census Tract within 150 miles of each airport and station to a highway network built from a national street centerline geodatabase.

Airfare distributions by decile for each airport pair in the study area with observed travel in the DB1B data were developed for 2008 for use in the airport choice model. Figure 5 shows the distributions for the top 40 airport pairs (omitting the maximum values, or 100% value, from the chart for clarity, as in several cases these values reach several thousand dollars). In general, the distributions are very similar, clustering in a band that reaches between $250 and $350 at the 90th percentile. However, three airport pairs have markedly higher fares: BOS-DCA, LGA-DCA, and LGA-BOS. These airports' fares are between approximately $600 and $800 at the 90th percentile.

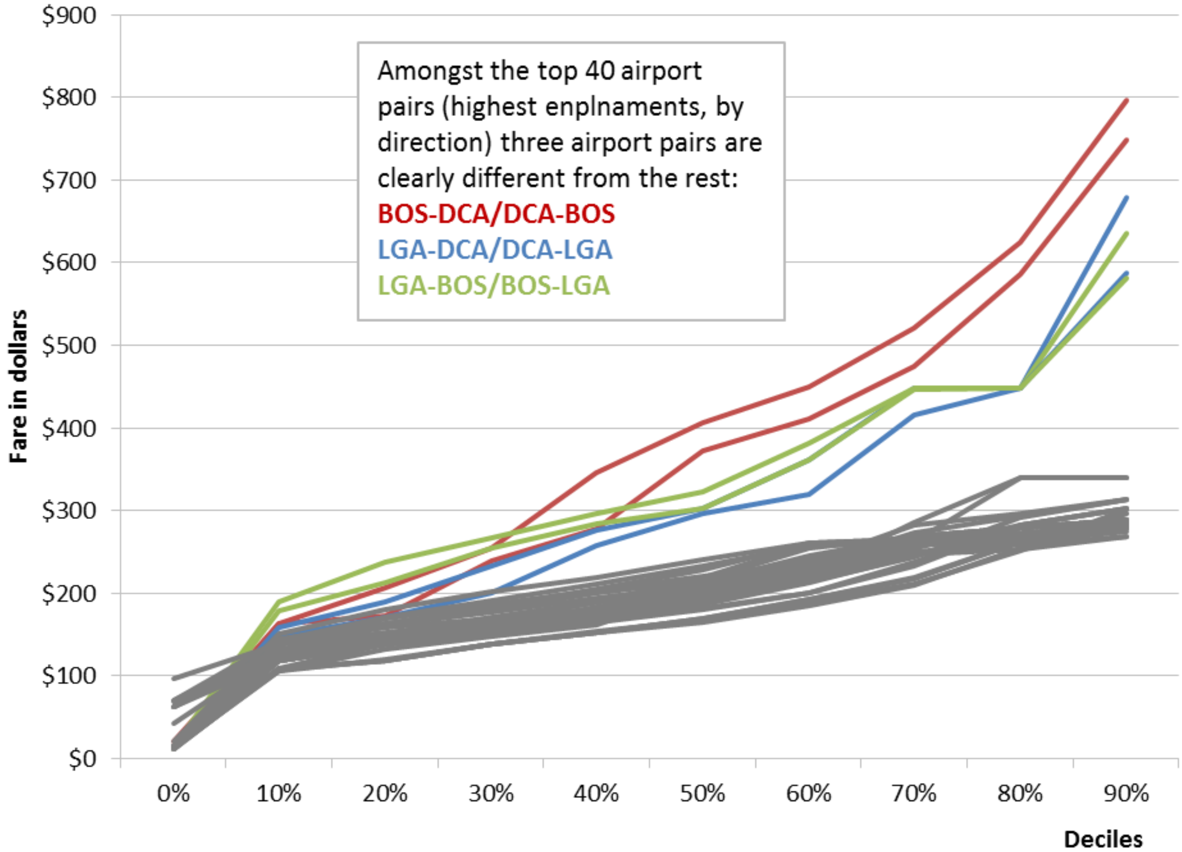

Figure 5. Air fare distribution in 2008 by decile for airport pairs in study area.

Table 13 shows the airport pairs in each study area with the highest levels of air service. The source for these data is the BTS Airline on-time performance data (http://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=120&Link=0), which shows the service scheduled and actually operated, in this case for 2008. Unlike the Official Airline Guide, this is a publicly available source of airline level-of-service data; therefore, this is preferable as the data can be included in the model for distribution without licensing issues.

Table 13. Airport pairs with highest levels of air service (direct flights per day in 2008)*.

West Coast Study Area

ORIGIN	DEST	Flights/Day
SFO	LAX	38
LAX	SFO	37
LAX	SAN	31
SAN	LAX	31
SJC	LAX	24
LAX	SJC	24
LAX	OAK	21
OAK	LAX	21
SAN	SFO	18
SFO	SAN	17
OAK	SAN	16
SAN	OAK	16
SMF	LAX	16
LAX	SMF	16
BUR	OAK	15
OAK	BUR	15

East Coast Study Area

ORIGIN	DEST	Flights/Day
LGA	BOS	33
BOS	LGA	33
DCA	LGA	30
LGA	DCA	30
DCA	BOS	24
BOS	DCA	24
JFK	BOS	21
BOS	JFK	20
PHL	BOS	17
BOS	PHL	17
BOS	BWI	13
BWI	BOS	13
BOS	IAD	11
BWI	PVD	11
PVD	BWI	11
BOS	EWR	11

* Direct flights per day are the average of all flights in 2008

Rail schedule data were developed using the published Amtrak schedule; these data incorporate Amtrak's connecting buses in California (this was also incorporated on the demand side into the development of the rail county-to-county trip tables).

The Research Team developed rail fare assumptions by estimating the relationship between rail fares and trip distance, segmented by East and West Coast markets (which have notably different fare levels). Average fares between an O-D pair are calculated as the sum of revenue on all non-express services (including both coach and business seats) divided by the sum of ridership. Figure 6 and Figure 7 show the observed average fare and rail mileage data (in blue); they also show a fitted function that describes the relationship (in red). There is a notable difference in fare-mileage patterns between the two market segments—the East Coast fare appears to have a polynomial relationship with the trip mileage while the West Coast fare is more likely to have a linear relationship with the trip mileage. A restriction was imposed during fitting the polynomial function; this restriction dictated that the fare will only increase as the mileage increases.

Over the past few years, Amtrak has announced several fare increases. Fares were first assumed to have a 4% annual increase from 2004 to 2008 in both markets to capture this increasing trend. Current (coach seats) fares in 2012 for station pairs with the highest demand was obtained from Amtrak's online reservation system as a calibration reference. Table 14 and Table 15 show the fare differences in 2004 and 2012 for 10 O-D pairs with the highest ridership in 2004 in each market. The data show that 4% is not sufficient to capture the increase in the West Coast market fare between 2004 and 2012, and the annual increase rate was adjusted to 10% for the West Coast market.

The final rail fare functions used to develop 2008 station-to-station fares are:

East coast: $fare = -0.000186358 \cdot mileage^2 + 0.295565 \cdot mileage + 14.062870$
West coast: $fare = 0.167421 \cdot mileage + 8.301447$

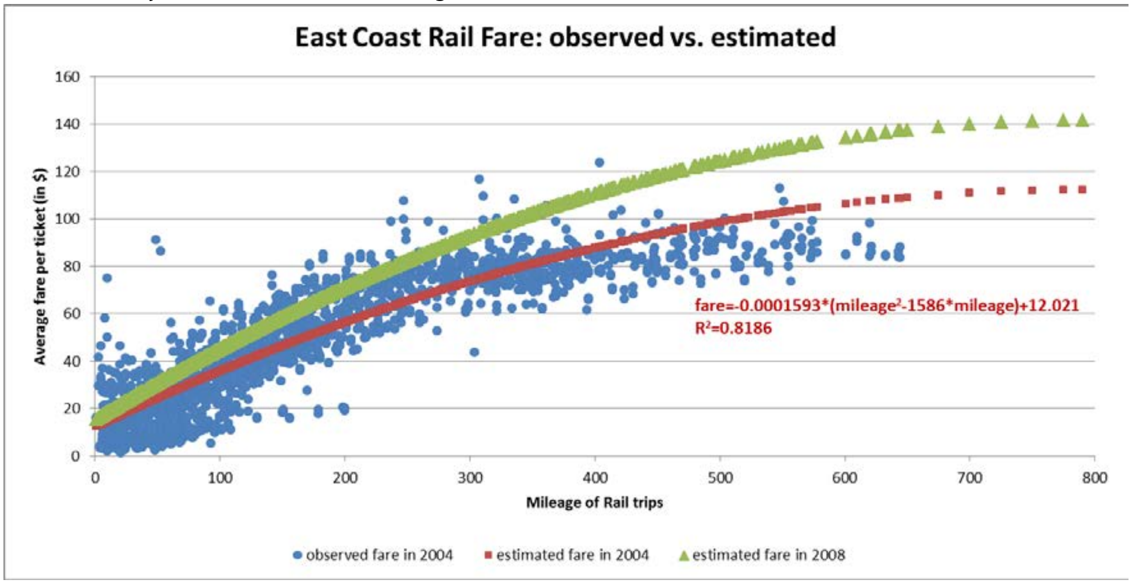

Figure 6. Relationships between East Coast rail fare and trip mileage.

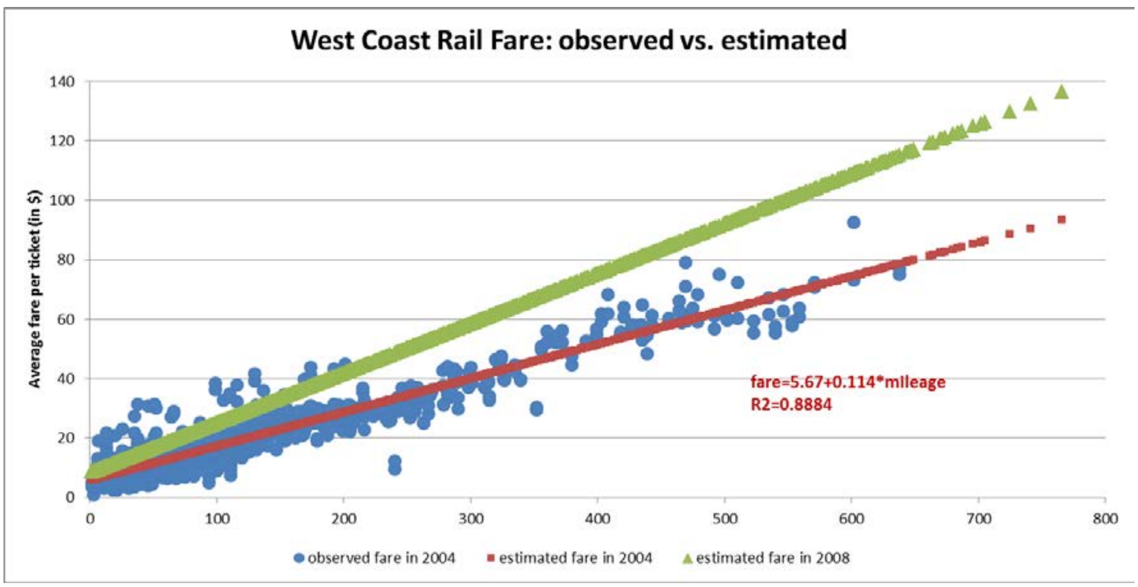

Figure 7. Relationships between West Coast rail fare and trip mileage.

Table 14. East Coast rail fare at station O-D pairs with max demand, observed versus estimated.

Origin/Dest.		Avg. Time (in Minutes)	Coach ridership	Mileage	Avg. fare in 2004	Estimated fare in 2008	Coach fare, 2012
PHL	NYP	75.7	1,159,037	91	37.67	39.42	36.00
WAS	NYP	174.3	849,100	226	72.48	71.34	82.00
NYP	TRE	54.0	544,378	58	14.84	30.58	41.00
NYP	PJC	50.8	533,470	49	7.26	28.10	34.00
NYP	ALB	157.8	516,715	142	41.60	52.28	41.00
WAS	PHL	96.7	398,338	135	43.86	50.57	50.00
NWK	PJC	32.4	325,826	39	6.82	25.31	34.00
NYP	NWK	16.1	239,400	11	7.16	17.29	28.00
NYP	BAL	145.9	212,864	185	69.61	62.36	73.00
NYP	BOS	247.7	210,484	231	62.93	72.39	71.00

Rail station codes are defined in the list of abbreviations on page 4

Table 15. West Coast rail fare at station O-D pairs with max demand, observed versus estimated.

Origin/ Dest.		Avg. Time (in Minutes)	Coach ridership	Mileage	Avg. fare in 2004	Estimated fare in 2008	Coach fare, 2012
SAN	LAX	160.23	274,520	128	23.39	29.82	37.00
EMY	SAC	111.67	156,601	85	11.55	22.59	28.00
SOL	LAX	126.36	108,143	103	20.95	25.62	30.00
OSD	LAX	110.56	93,660	87	16.44	22.93	28.00
SAC	RIC	85.44	93,595	77	10.84	21.25	26.00
SAC	MTZ	60.13	96,656	58	8.09	18.06	28.00
SAC	OKJ	115.50	87,207	90	11.83	23.43	28.00
FUL	SAN	134.91	75,130	102	16.83	25.45	29.00
SAN	SNC	81.02	70,457	70	12.08	20.07	21.00
DAV	EMY	80.54	70,271	71	9.64	20.24	26.00

Rail station codes are defined in the list of abbreviations on page 4

For the transit accessibility measure, 2000 CTPP data describing journey to work by mode was processed for all Census Tracts and counties in the study area to calculate transit mode shares for the journey to work. These data are used in the model as a proxy for transit accessibility from that Census Tract and county rather than attempting to develop transit network data, which is beyond the scope of the type of sketch planning model developed as part of this research.

The transit accessibility measure (TAM) was computed as a weighted sum of relevant transit shares:

TAM (ij) = 0.25*TS_org_county(i) + 0.5*TS(ij) + 0.25* TS_des_county (j)

Where,

Symbol	Description
TAM (ij)	Transit accessibility measure between census tracts i and j
TS_org_county(i)	Transit share of all journey to work (JTW) trips originating in the county containing census tract i
TS_des_county(j)	Transit share of all journey to work (JTW) trips terminating in the county containing census tract j
TS(ij)	Transit share of all journey to work (JTW) trips between census tracts i and j

Highway accessibility data for each airport and station were developed by applying a shortest-path algorithm from each Census Tract within 150 miles of each airport and station to a highway network built from a national street centerline geodatabase. Maps describing the data are shown in Figure 8, Figure 9, Figure 10, and Figure 11.

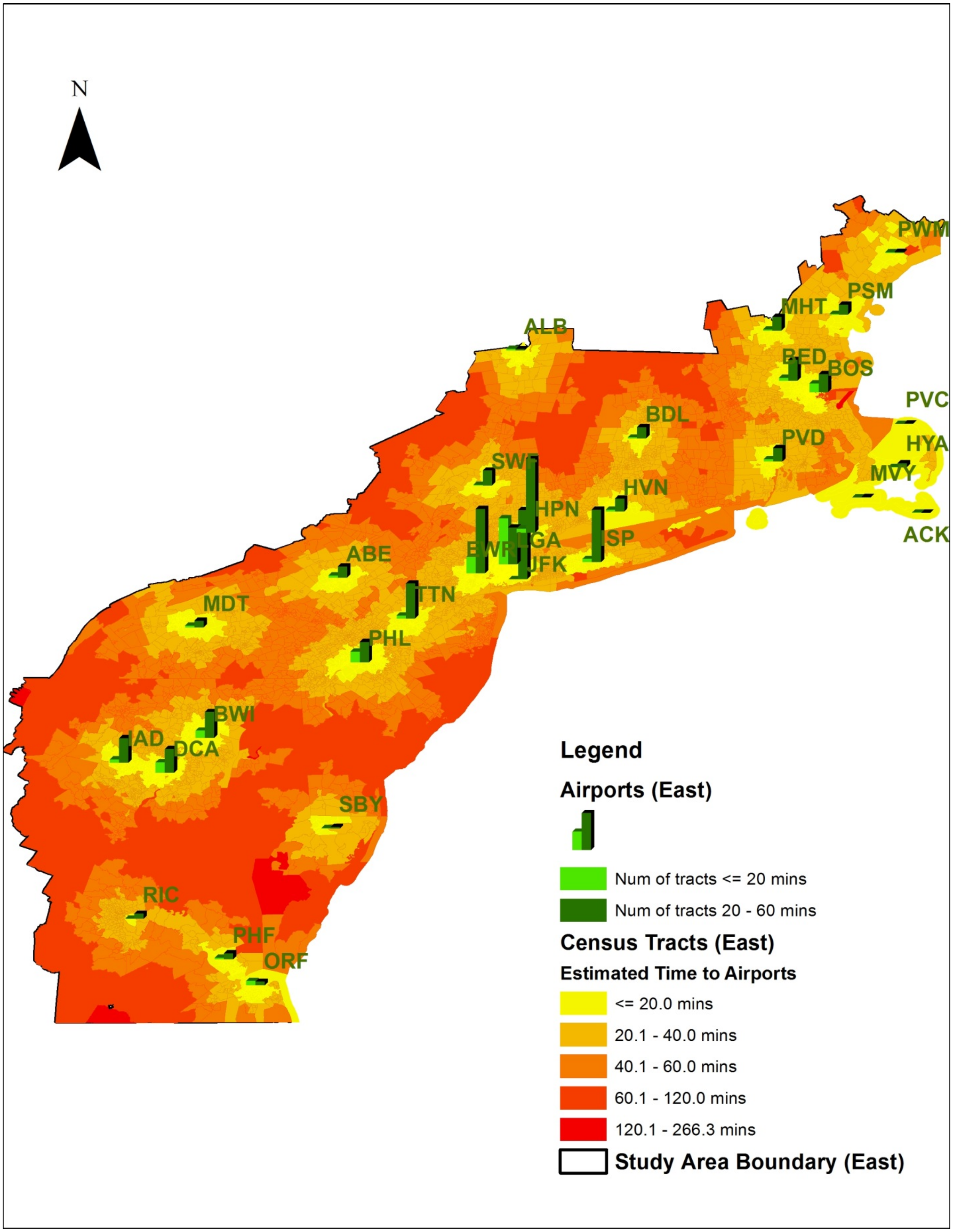

Figure 8. Highway access times to East Coast airports.

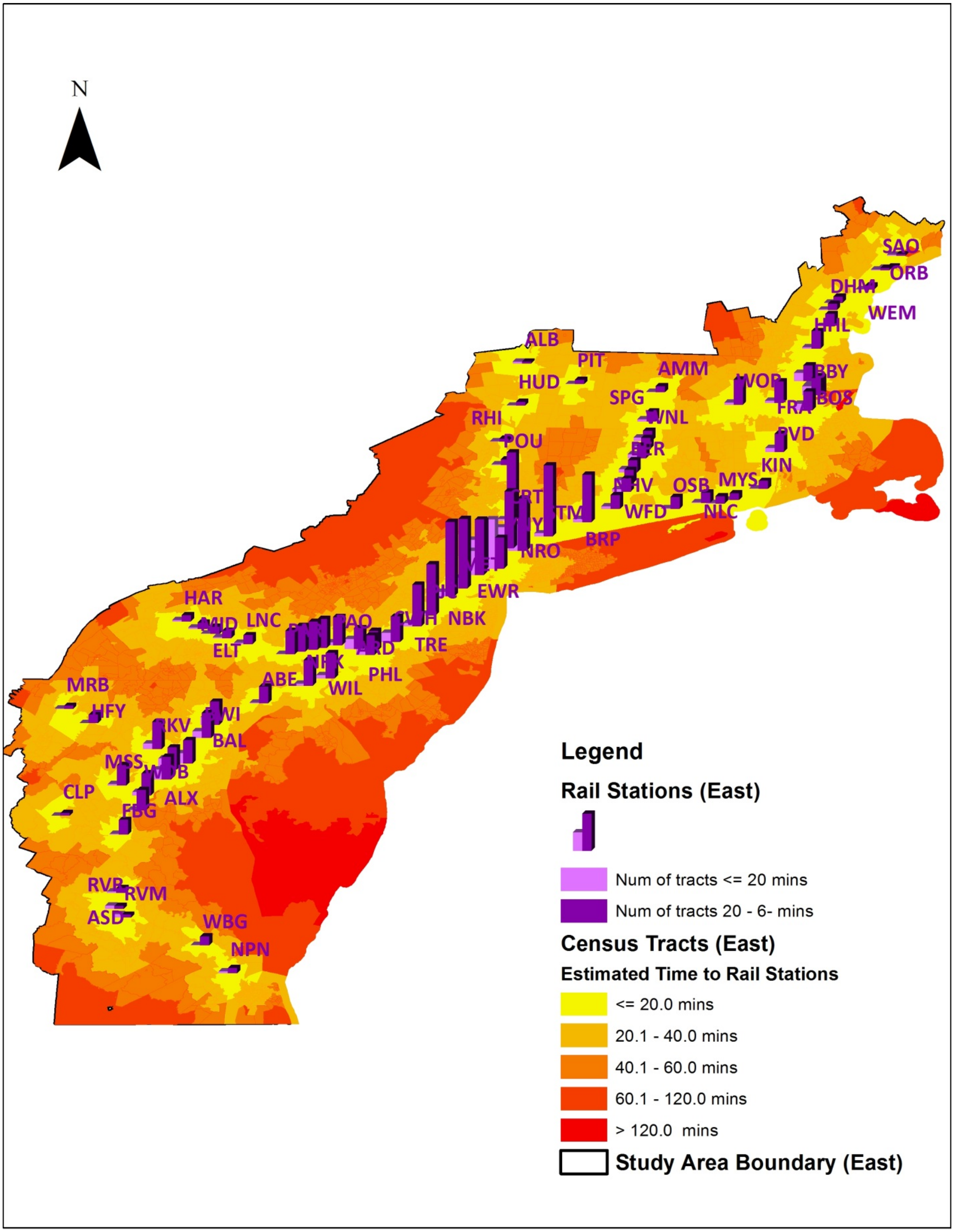

Figure 9. Highway access times to East Coast rail stations.

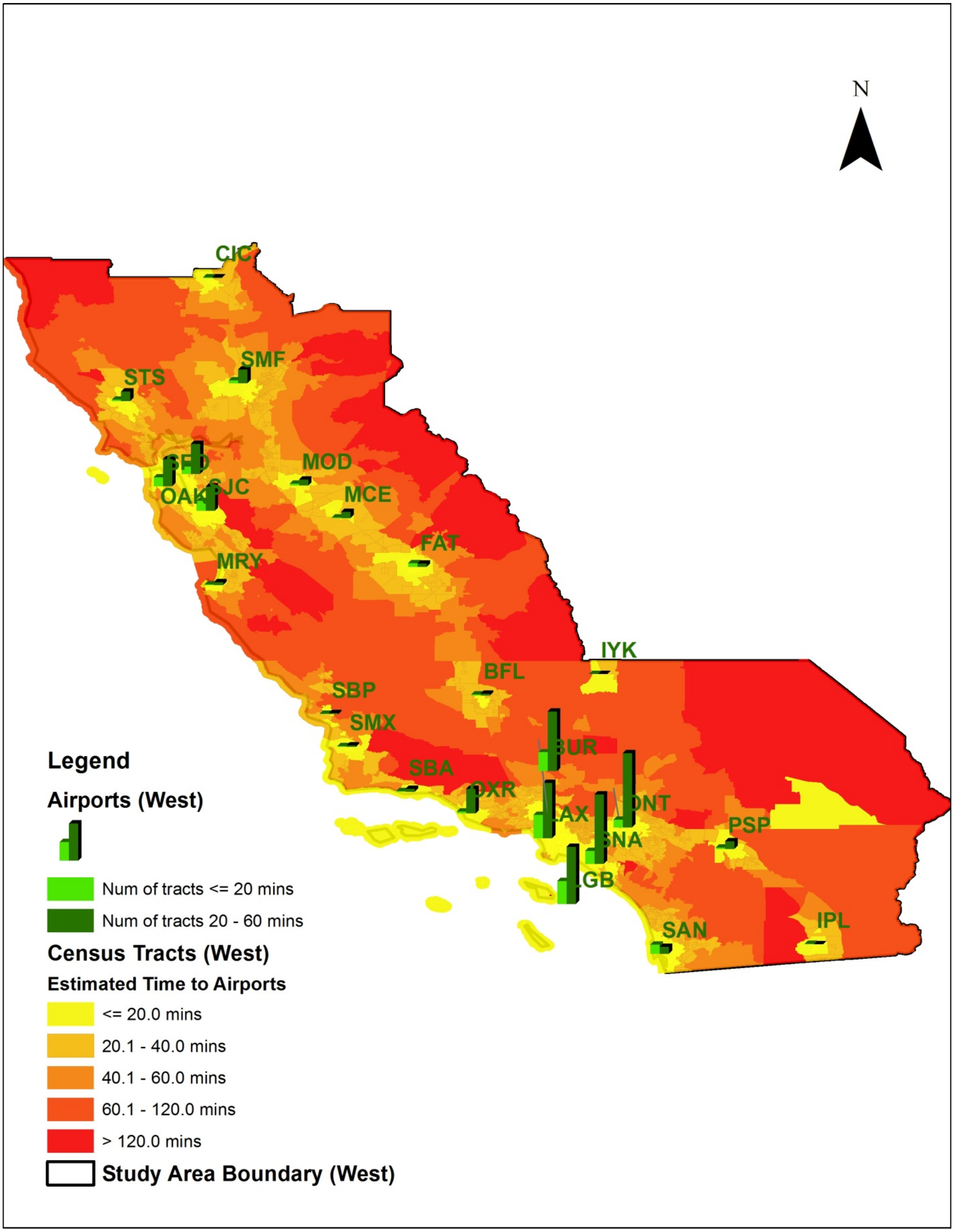

Figure 10. Highway access times to West Coast airports.

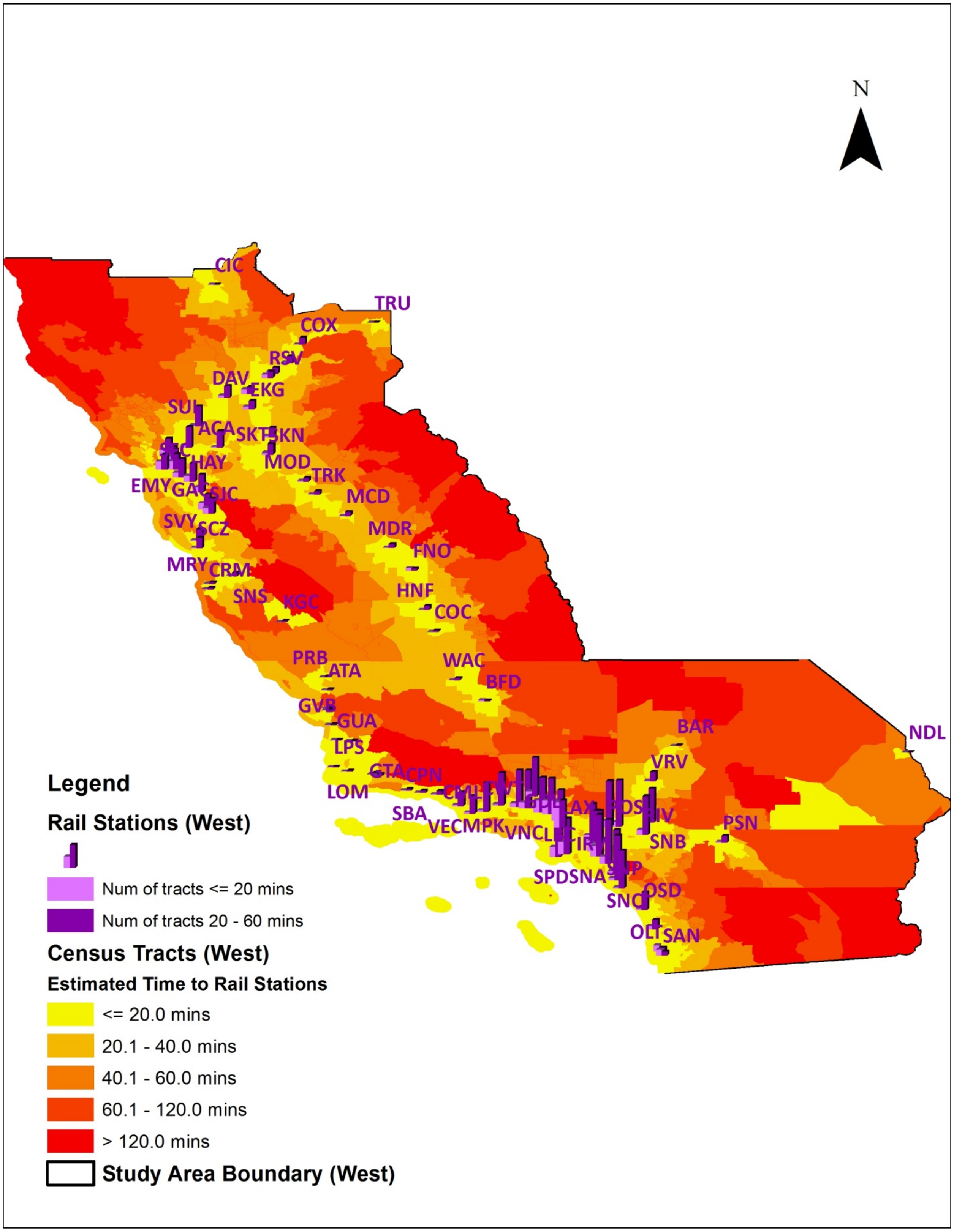

Figure 11. Highway access times to West Coast rail stations.

Approach

While the primary choice model in the framework is the main mode choice between air and rail, the model incorporates an airport choice and a station choice model to define the specific air and rail service options traded off in the main mode choice model by each travel party. The selection of particular airports and rail stations could be achieved using a simple algorithm, such as selecting the closest airport and rail stations, but instead the model incorporates an airport choice and a station choice model into the framework to allow consideration of additional variables such as the relative level of air service at different airports. As a result, travel parties first choose between the available airports and then between the available rail stations at each end of their trip to select a preferred pair of airports and pair of rail stations for use in the main mode choice model.

The model that is being transferred is a joint multinomial logit airport choice and ground access mode choice model with alternatives specified for each combination of airport and mode choice. It was developed by members of the Research Team for the West of Hudson Regional Transit Access Study (WHRTAS). This included airport choice between the airports in the New York region. The variables in the model were grouped into two categories: airport choice parameters, which affect choice between airports, and mode choice parameters, which affect choice between modes, but also (collectively) influence airport choices. The model was applied for four segments, and only considered the end of the air trip within the New York region.

Several of the model coefficients are specific to the New York region to help better fit to the idiosyncrasies of travel across the Hudson River and to and from Manhattan. Given that the model was estimated specifically for the New York Region, it is noted by the Research Team that applying the model in other regions is likely to explain airport choice less well than a model estimated using locally collected airport choice data would. It is recommended that analysts attempting to transfer the model to other regions review whether parameters in the model such as the distance coefficient (which might partially account for the difficulty of traveling longer distances to airports in a region like New York with geographical barriers such as major rivers) are relevant to other regions.

For application in this model framework, the joint airport and access mode choice model has been simplified, with elements of the access mode choice model and the region specific dummy variables removed, and adjustments made to incorporate the transit accessibility measure. In addition, the context of the choice being modeled in this model is somewhat different than the region-specific context of the WHRTAS application. In this model, the airport choice model builds a choice set of possible air routes that each air traveler might use. At the point of application of the airport choice model, each travel group's trip is defined with a specific origin tract and destination tract. A set of accessible airports are selected at either end of the trip, starting with those within 75 miles of the origin and 75 miles of the destination, or until the closest three airports are included at each of the origin and destination. Air service data is then used to select the airport pairs from airports close to the origin and destination that have air service (either direct or with a reasonable transfer, based on review of the on-time performance data and the DB1B data to understand routes that are operated and actually used). This set of airport pairs becomes the choice set for the airport choice model, which chooses the airport pair for the travel party.

The change from modeling just one of the air trips to modeling the complete trip, from origin to destination, means there are now just two segments—one for each trip purpose, business and non-business. This change also means that the concept of resident and visitor is used to differentiate between the origin end of the trip, where the air traveler is a resident, and the destination end of the trip, where the air traveler is a visitor.

Table 16 shows the airport choice model parameters, which were extracted from the complete model specification. The specific meaning of each parameter and the input data used with each parameter are described below. One simplification made to the model is that while access travel time is considered, access cost is no longer included in the model specification. Air fare is still included in the model specification, and consistent with the approach used to estimate the models, air fare is in terms of total party cost while travel time elements are in terms of the amount of time each person's travel takes and not in terms of total party travel times.

Table 16. Airport choice model parameters.

Parameter Description	Parameter Name	Non-Business		Business	
		Coeff	T-Stat	Coeff	T-Stat
Millions of Annual Enplanements, Origin Airport	EnplanementsO	0.17	27.8	0.0278	3.2
Millions of Annual Enplanements, Destination Airport	EnplanementsD	0.094	16.1	0.0948	10.6
Direct Service Constant	DirectConst	1.73	11.9	2.72	10.1
Direct Service Flights (flight per day)	DirectFlights	0.0204	23.1	0.0293	18.2
Distance (highway miles), Origin Airport (see text for definition)	DistanceO	0.052	7.6	0.067	5.9
Distance (highway miles), Destination Airport (see text for definition)	DistanceD	0.0704	9.9	0.0907	9.0
Air Fare ($, for all party)	AirFare	-0.0134	-11.2	-0.0135	-4.2
Access In Vehicle Travel Time (minutes), Origin	AccessIVTTO	-0.0191	-33.8	-0.0202	-23.5
Access In Vehicle Travel Time (minutes), Destination	AccessIVTTD	-0.0103	-19.7	-0.0138	-16.4
Transit Accessibility, Origin	TransitO	1	-	1	-
Transit Accessibility, Destination	TransitD	1	-	1	-
Model Statistics for complete models, segmented by resident/non-resident					
Observations		9,048/6,932		4,710/3,722	
Final Log Likelihood		-16,929/-12,451		-7,836/-6,684	
Rho-Squared (0)		0.325/0.338		0.349/0.337	

As explained above, the airport choice model evaluates a set of possible alternative air trips from an origin Census Tract, OTract, to a destination Census Tract, DTract, via several accessible origin airports, OAirport(s), and several accessible destination airports, d airport(s).

The utility for each OTract-OAirport, OAirport-DAirport, DAirport-DTract route is created from three terms in a utility equation as follows:

Utility(Air Trip, OTract-DTract) = U(OAirport-DAirport) + OTract-OAirport + DTract-DAirport

Where

U(OAirport-DAirport) = EnplanementsO * Origin Airport Enplanements +
EnplanementsD * Destination Airport Enplanements +
DirectConst * Dummy (Direct Flights) +
DirectFlights * Number of Direct Flights
AirFare * Party Air Fare

- Origin Airport Enplanements is the number of enplanements at the origin airport in millions of annual enplanements. This uses the resident coefficient for the enplanements parameter from the WHRTAS model.
- Destination Airport Enplanements is the number of enplanements at the destination airport in millions of annual enplanements. This uses the visitor coefficient for the enplanements parameter from the WHRTAS model.
- Dummy (Direct Flights) is a dummy variable that is applied if there is at least one daily direct flight between the origin and destination airports. This uses the coefficient for the Destination Constant parameter from the WHRTAS model.
- Number of Direct Flights is the number of daily direct flights from the origin to the destination airport. This uses the coefficient for the Destination Flights parameter from the WHRTAS model.
- Party Air Fare is the air fare for the party in dollars. The air fare for the party is the per person fare drawn from the airfare distributions by decile for each airport pair in the study multiplied by the party size.

U(OTract-OAirport) = DistanceO * max(20-Distance to Airport,0) +
AccessIVTTO * Access in Vehicle Travel Time +
TransitO * Transit Accessibility Measure

- Distance (highway), origin airport is a positive utility for airport proximity. For distances over 20 miles it has a value of zero. For distance of less than 20 miles, the coefficient is multiplied by 20 – highway distance to the airport (which therefore peaks at a value of 20 with zero distance from the airport). This uses the resident coefficient for the distance parameter from the WHRTAS model.
- Access in Vehicle Travel Time, origin, is the negative utility associated with travel time from the origin tract to the origin airport and is multiplied by the number of minutes of auto travel time for that trip. This uses the resident coefficient for the Access In Vehicle Travel Time—Not Commuter Rail parameter from the WHRTAS model.
- Transit Accessibility, origin, is the asserted coefficient that is multiplied by the transit accessibility measure for the origin tract – origin airport combination. The development of this measure is described above. Since this parameter has been added to the model, an asserted value for the coefficient of one is used so that value of utility equals the value of the transit accessibility measure. Future research may be able to include this variable in a model specification and estimate a coefficient.

U(DTract-DAirport) = DistanceD * max(20-Distance to Airport,0) +
AccessIVTTD * Access in Vehicle Travel Time +
TransitD * Transit Accessibility Measure

- Distance (highway), destination airport is a positive utility for airport proximity. For distances over 20 miles, it has a value of zero. For distance of less than 20 miles, the coefficient is multiplied by 20 – highway distance to the airport (which therefore peaks at a value of 20 with zero distance from the airport). This uses the visitor coefficient for the distance parameter from the WHRTAS model.
- Access in Vehicle Travel Time, destination, is the negative utility associated with travel time from the destination airport to the destination tract and is multiplied by the number of minutes of auto travel time for that trip. This uses the visitor coefficient for the Access In Vehicle Travel Time – Not Commuter Rail parameter from the WHRTAS model.

- Transit Accessibility, destination airport, is the asserted coefficient that is multiplied by the transit accessibility measure for the destination tract – destination airport combination. The development of this measure is described above.

Once the utilities for the three sections of each OTract-OAirport, OAirport-DAirport, DAirport-DTract route in the choice set is calculated, they are summed together to create a total utility. As this is a multinomial logit model, each total utility is then exponentiated, and the choice probability for each alternative is calculated as the exponentiated utility for that alternative divided by the sum of all of the exponentiated utilities.

The choice probabilities are converted to a cumulative probability distribution and used to simulate a choice for each travel party. In this way, the alternative with the highest utility is the mostly likely to be chosen, but some travel parties will choose less attractive alternatives.

A similar approach is used to apply the station choice model, with some additional simplification in comparison to the airport choice context. In this case, all stations within 25 miles or a minimum of the 2 closest stations are included for the origin and destination ends of the trip. The station pairs with rail service between them (including those with connecting rail service) are selected. A smaller set of parameters are used to define the utility of each pair of stations: distance from the origin to the access station and from the egress station to the destination (using a similar positive proximity value as in the airport choice model); access and egress in vehicle travel time; and the number of trains per day between the two stations. The utility for each OTract-OStation, OStation-DStation, DStation-DTract route is created as follows and then the logit model is simulated to select station choices for each travel party:

Utility(Rail Trip, OTract-DTract) = U(OStation-DStation) + OTract-OStation + DTract-DStation

Where

U(OStation-DStation) = DirectConst * Dummy (Direct Trains) +
DirectFlights * Number of Direct Trains

- The Direct Service Constant is a dummy variable that is applied if there is at least one daily direct train between the origin and destination station. This uses the coefficient for the Destination Constant parameter from the WHRTAS model.
- The Direct Service Flights coefficient is multiplied by the number of daily direct trains from the origin to the destination station. This uses the coefficient for the Destination Flights parameter from the WHRTAS model.

U(OTract-OStation) = DistanceO * max(20-Distance to Station,0) +
AccessIVTTO * Access in Vehicle Travel Time +
TransitO * Transit Accessibility Measure

- Distance (highway), origin airport is used as a positive utility for station proximity. For distances over 20 miles it has a value of zero. For distance of less than 20 miles, the coefficient is multiplied by 20 – highway distance to the station (which therefore peaks at a value of 20 with zero distance from the station). This uses the resident coefficient for the distance parameter from the WHRTAS model.
- Access in Vehicle Travel Time, origin, is the negative utility associated with travel time from the origin tract to the origin station and is multiplied by the number of minutes of auto travel time for that trip. This uses the resident coefficient for the Access In Vehicle Travel Time—Not Commuter Rail parameter from the WHRTAS model.
- Transit Accessibility, origin, is the asserted coefficient that is multiplied by the transit accessibility measure for the origin tract – origin station combination.

U(DTract-DStation) = DistanceD * max(20-Distance to Station,0) +
AccessIVTTD * Access in Vehicle Travel Time +
TransitD * Transit Accessibility Measure

- Distance (highway), destination airport is a positive utility for station proximity. For distances over 20 miles it has a value of zero. For distance of less than 20 miles, the coefficient is multiplied by 20 – highway distance to the station (which therefore peaks at a value of 20 with zero distance from the station). This uses the visitor coefficient for the distance parameter from the WHRTAS model.
- Access in Vehicle Travel Time, destination, is the negative utility associated with travel time from the destination station to the destination tract and is multiplied by the number of minutes of auto travel time for that trip. This uses the visitor coefficient for the Access In Vehicle Travel Time – Not Commuter Rail parameter from the WHRTAS model.
- Transit Accessibility, destination airport, is the asserted coefficient that is multiplied by the transit accessibility measure for the destination tract – destination station combination.

Main Mode Choice

Overview

This step in the model applies the main mode choice model to choose between traveling by air or rail, based on the characteristics of the travel party, the accessibility of the airports and rail stations at origin and destination end of the trip, and the level of air and rail service between them. The input to this model is a set of travel parties with round trip itineraries that include the origin and destination Census Tracts and an assumed airport pair and rail station pair for their trip.

Data Sources

The data used in the airport and station choice models, describing the competing air and rail levels of service and fares, are used again in the main mode choice model.

Approach

The main mode choice model is a modified version of the California High Speed Rail Authority long trip main mode choice model. While the California High Speed Rail Authority model evaluates choice between auto, air, conventional rail, and HSR, in this application the auto mode (for the main mode) is not considered. Furthermore, the base application has been calibrated with just air and conventional rail available, with the option to add alternative improved passenger rail service in scenario testing.

The extracted coefficients for just the air, rail, and HSR choice for long distance only travel are shown in Table 17. The California High Speed Rail Authority access and egress mode choice models were not used as part of this model, but the logsum values have been retained. Instead of applying those values to the intended logsum values, the travel times and transit accessibility values derived from the airport, and station choice models for access and egress from the chosen airports and stations have been utilized to bring some measure of influence of both highway and transit accessibility at the access and egress ends of the trip into the main mode choice model. The access and egress travel times have been converted to utilities using the coefficients from the California High Speed Rail Authority access and egress mode choice models (i.e., coefficient values of -0.06 for business travel and -0.03 for non-business travel.)

Table 17. Coefficients extracted from the California High Speed Rail Authority long trip main mode choice model.

Parameter	Non Business		Business	
	Coeff.	T-Stat	Coeff.	T-Stat
Air Constant	0.6898	2.8	-1.645	-4.7
Conventional Rail Constant	0.6149	2.6	-0.387	-0.9
High Speed Rail Constant	1.434	7	-0.3503	-1.1
Cost ($)	-0.035	-18.5	-0.01626	-12.8
In vehicle travel time (min)	-0.011	-14.2	-0.016	-11.1
Service headway (min)	-0.003	-3.5	-0.003	-3.7
Air Travel in a Group	-0.5061	-3.7	-0.3375	-2.7
High Income - Air	0	-	1.018	4.5
High Income - Conventional Rail	0	-	0.5237	1.2
High Income - High Speed Rail	0	-	0.9807	4.8
Access Mode Choice Logsum	0.2134	3.8	0.115	3.1
Egress Mode Choice Logsum	0.3974	7.1	0.1561	3.8
Model Statistics (for complete model including auto alternative				
Observations	8,075		2,918	
Final Log Likelihood	-3,933		-1,969	
Rho-Squared (0)	0.31		0.389	
Rho-Squared (cons)	0.155		0.163	

The utility equations for the alternatives included in the main mode choice are as follows:

U(Air) =
Air Constant +
Cost * Party Air Fare +
In vehicle travel time * In vehicle travel time by Air +
Service headway * Air Service Headway +
Air Travel in a Group * Dummy (Party size greater than one) +
High Income Air * Dummy (High Income) +
Access Mode Choice Logsum * (Access Egress IVTT * Access In Vehicle Travel Time + Transit Accessibility Measure) +
Egress Mode Choice Logsum * (Access Egress IVTT * Egress In Vehicle Travel Time + Transit Accessibility Measure)

U(Conventional Rail) =
Conventional Rail Constant +
Cost * Party Conventional Rail Fare +
In vehicle travel time * In vehicle travel time by Conventional Rail +
Service headway * Conventional Rail Service Headway +
High Income - Conventional Rail * Dummy (High Income) +
Access Mode Choice Logsum * (Access Egress IVTT * Access In Vehicle Travel Time + Transit Accessibility Measure) +

Egress Mode Choice Logsum * (Access Egress IVTT * Egress In Vehicle Travel Time + Transit Accessibility Measure)

U(High Speed Rail) = High Speed Rail Constant +
Cost * Party High Speed Rail Fare +
In vehicle travel time * In vehicle travel time by High Speed Rail +
Service headway * High Speed Rail Service Headway +
High Income - High Speed Rail * Dummy (High Income) +
Access Mode Choice Logsum * (Access Egress IVTT * Access In Vehicle Travel Time + Transit Accessibility Measure) +
Egress Mode Choice Logsum * (Access Egress IVTT * Egress In Vehicle Travel Time + Transit Accessibility Measure)

Where:

- Party Air Fare, Party Conventional Rail Fare, and Party High Speed Rail Fare are the input party fares in dollars for the air and rail modes respectively.
- Air Service Headway, Conventional Rail Service Headway, and High Speed Rail Service Headway are the inputs headways in minutes between air and rail services for the respective airport to airport and station to station routes.
- Air Travel in a group is a dummy variable with a value of one when the size of the travel party is more than one. Its negative value indicated that the rail modes are more attractive to those traveling in a group.
- Dummy (High Income) is a dummy variable with a value of one when the household income assigned to the travel party is more than $75,000 per year.
- Access In Vehicle Travel Time and Egress In Vehicle Travel Time are the times taken in minutes to travel between the origin tract and the airport/station and the destination tract and the airport/station, as also used in the airport and station choice models.
- Transit Accessibility Measure is the measure for transit accessibility estimated for travel between the origin tract and the airport/station and the destination tract and the airport/station, as also used in the airport and station choice models.

The model has been lightly calibrated to the observed county to county air vs. rail shares by adjusting the alternative specific constants, adjusting model’s scale, and by adjusting the sensitivity to travel time. This has not been done separately for the two coastal regions (recognizing the low observed shares in the West Coast due to the relatively small amount of existing rail service). The following specific calibration adjustments were made:

- The scale of the model has been doubled (i.e., all coefficients multiplied by two).
- The constants in the model were normalized on the conventional rail alternative.
- The air alternative specific constant in the East Coast model has been increased by 1 and in the West Coast model has been decreased by 2.
- The in vehicle travel time coefficient in the East Coast model has been increased by a factor of 1.2.

No O-D pair specific adjustments were made to attempt to force the model to match specific county to county modal volumes. The estimated and calibrated coefficients for each segment in the East and West Coast models are shown in Table 18.

Table 18. Estimated and calibrated coefficients in the main mode choice model.

	Non Business			Business		
Parameter	Est.	East Coast	West Coast	Est.	East Coast	West Coast
Air Constant	0.6898	1.1498	-1.8502	-1.645	-1.516	-4.516
Conventional Rail Constant	0.6149	0	0	-0.387	0	0
High Speed Rail Constant	1.434	1.6382	1.6382	-0.3503	0.0734	0.0734
Cost ($)	-0.035	-0.07	-0.07	-0.01626	-0.03252	-0.03252
In vehicle travel time (min)	-0.011	-0.0264	-0.022	-0.016	-0.0384	-0.032
Service headway (min)	-0.003	-0.006	-0.006	-0.003	-0.006	-0.006
Air Travel in a Group	-0.5061	-1.0122	-1.0122	-0.3375	-0.675	-0.675
High Income - Air	0	0	0	1.018	0.9886	0.9886
High Income – Con. Rail	0	0	0	0.5237	0	0
High Income - High Speed Rail	0	0	0	0.9807	0.914	0.914
Access Mode Choice Logsum	0.2134	0.4268	0.4268	0.115	0.23	0.23
Egress Mode Choice Logsum	0.3974	0.7948	0.7948	0.1561	0.3122	0.3122

Table 19 and Table 20 show a summary of the top 10 metropolitan areas on the East and West Coasts respectively and compare the modeled air and rail volumes with the reference trips table flows (which are denoted as observed data, although it is important to note that they are also the result of some modeling work to combine passenger data for the main mode movements with access survey data). Since the model is run using a sampling approach, when the outputs are scaled up to annual trips, there is a small discrepancy of a few thousand trips between the observed and modeled totals for each CBSA.

The following figures, Figure 12 to Figure 15, show additional comparisons between the modeled air and rail volumes and the reference trip table flows. The first two figures show similar information to the two tables, with modeled and observed air and rail volumes for the largest metropolitan areas on the East and West Coast respectively. The second two figures compare origin to destination volumes for the largest origin to destination pair metropolitan areas in terms of total travel demand, again for the East and West Coast respectively. In general, the modeled results compare well with the reference trip tables, although some discrepancies are apparent when inspecting specific origin to destination pairs as would be expected with a planning model of this type that has not been tightly calibrated for each origin to destination pair.

Table 19. Base year summary for East Coast model, top 10 metropolitan areas.

CBSA Title	Mod Air	Mod Rail	Obs Air	Obs Rail	Diff Air	Diff Rail
New York-Northern New Jersey-Long Island, NY-NJ-PA	852,350	1,920,450	796,876	1,965,163	(55,474)	44,713
Washington-Arlington-Alexandria, DC-VA-MD-WV	809,600	1,062,400	797,317	1,068,252	(12,283)	5,852
Boston-Cambridge-Quincy, MA-NH	867,250	704,450	885,984	675,212	18,734	(29,238)
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	252,950	1,205,700	225,452	1,227,225	(27,498)	21,525
Baltimore-Towson, MD	257,100	554,450	310,538	494,977	53,438	(59,473)
Providence-New Bedford-Fall River, RI-MA	198,850	221,300	195,810	224,549	(3,040)	3,249
Albany-Schenectady-Troy, NY	36,550	179,150	50,134	160,459	13,584	(18,691)
Richmond, VA	141,950	62,000	129,546	74,093	(12,404)	12,093
New Haven-Milford, CT	59,050	135,200	36,898	155,029	(22,152)	19,829

Table 20. Base year summary for West Coast model, top 10 metropolitan areas.

CBSA Title	Mod Air	Mod Rail	Obs Air	Obs Rail	Diff Air	Diff Rail
Los Angeles-Long Beach-Santa Ana, CA	2,372,350	800,500	2,312,793	866,934	(59,557)	66,434
San Francisco-Oakland-Fremont, CA	1,725,250	317,550	1,759,708	290,084	34,458	(27,466)
Sacramento--Arden-Arcade--Roseville, CA	712,700	277,100	704,775	285,393	(7,925)	8,293
San Diego-Carlsbad-San Marcos, CA	656,050	332,900	604,342	384,344	(51,708)	51,444
San Jose-Sunnyvale-Santa Clara, CA	694,300	75,950	715,679	59,755	21,379	(16,195)
Riverside-San Bernardino-Ontario, CA	337,100	40,550	324,343	54,095	(12,757)	13,545
Oxnard-Thousand Oaks-Ventura, CA	122,650	62,500	84,428	98,807	(38,222)	36,307
Santa Barbara-Santa Maria-Goleta, CA	80,100	54,000	20,362	110,135	(59,738)	56,135
Vallejo-Fairfield, CA	83,300	45,400	79,623	47,858	(3,677)	2,458

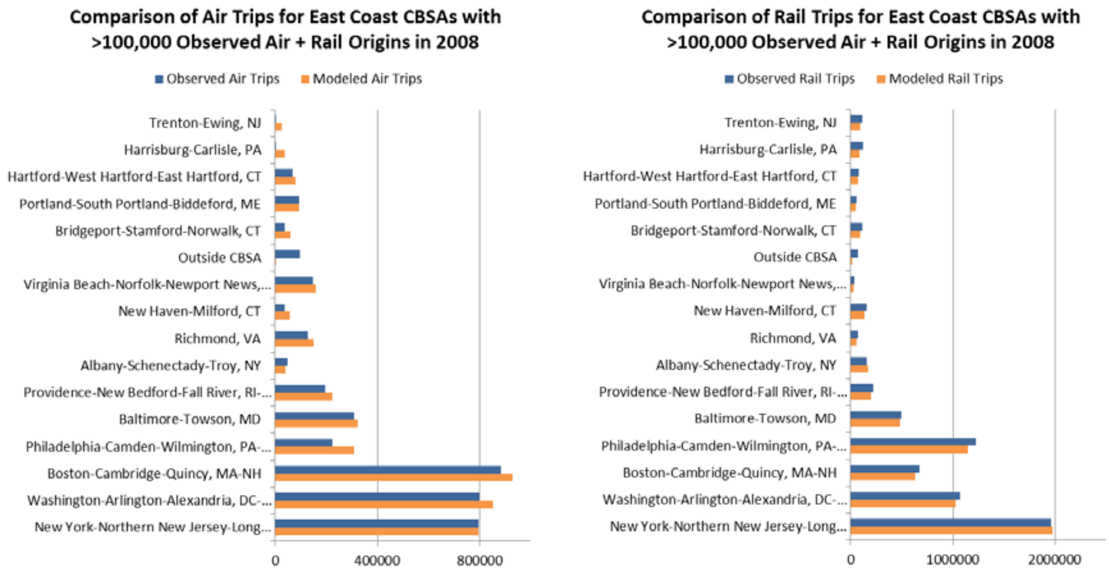

Figure 12. Comparison of East Coast modeled origins with reference trip tables.

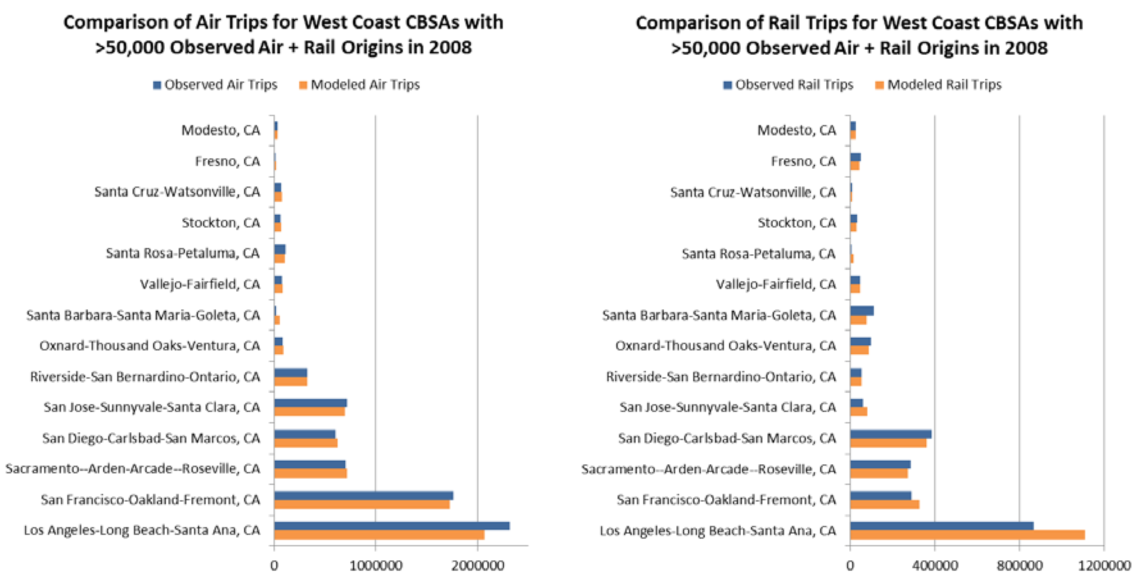

Figure 13. Comparison of West Coast modeled origins with reference trip tables.

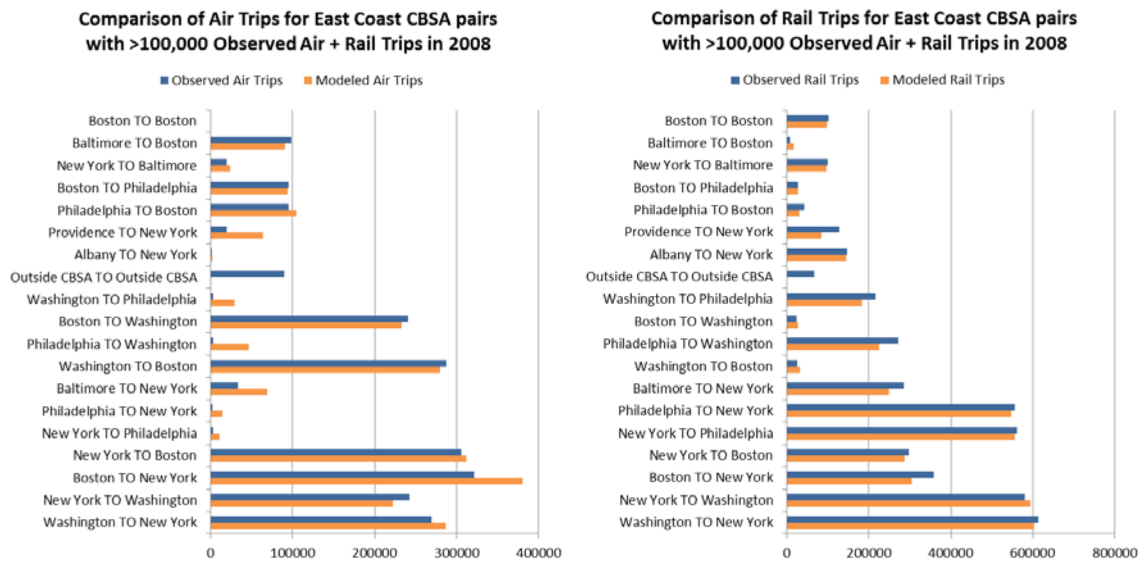

Figure 14. Comparison of East Coast modeled origin to destination trips with reference trip tables.

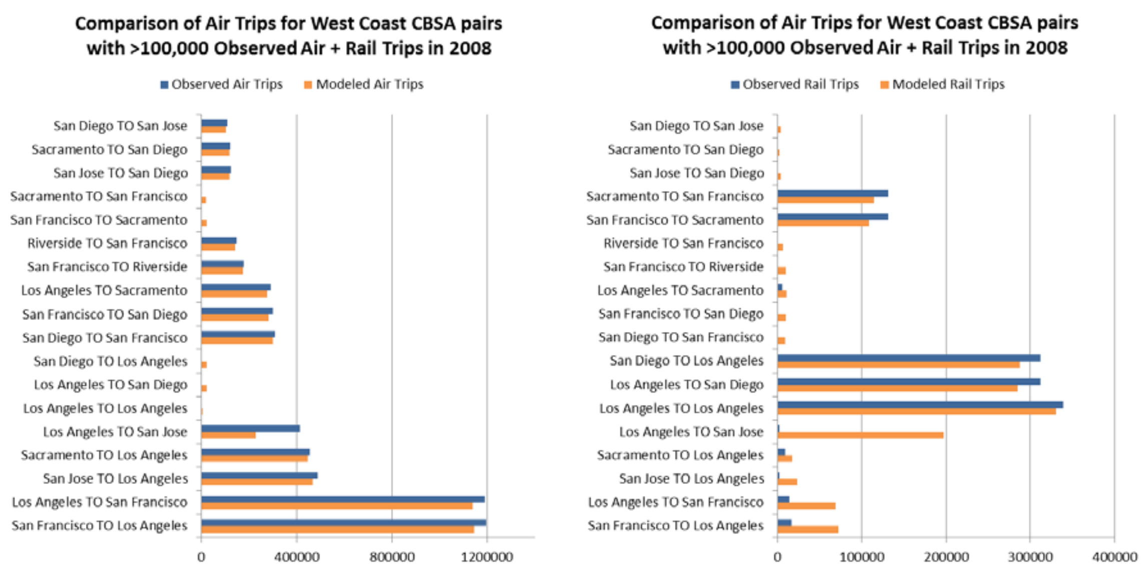

Figure 15: Comparison of West Coast modeled origin to destination trips with reference trip tables.

Airline Response

Overview

The final step in the model sequence evaluates the response of airlines to the amount of rail service in place in the corridor in terms of a change in airline level of service. This model forms part of a feedback loop in the model, where, once the adjusted airline level of service has been estimated, the main mode choice model is applied again using the new airline level of service.

Data Sources

The input data used for estimation of this model are derived from the air and rail level-of-service data described earlier in this chapter. The rail stations and airports in the East Coast and West Coast study areas were mapped to the common geographic regions of Core Based Statistical Areas (CBSAs).

Approach

The airline response model predicts how airlines adjust their service in response to improvement in rail service. Conceptually, that response includes several elements. First, to the extent that improved rail service captures traffic that previously traveled by air, airlines may reduce seat capacity on affected segments. Second, either as a result of offering reduced seat capacity, or as a direct response to competition from rail, the distribution of capacity among different types of airlines—mainline network carriers, regional affiliates, and low-cost carriers—may change. Finally, as either as a direct response to rail competition, or as a secondary response to diminished traffic, carriers may change their fleet mixes in corridors with improved rail service.

To model these responses, the Research Team has performed econometric analyses on cross-sectional airline service data in order to isolate the direct and indirect effects of competition from existing rail services. This approach assumes that the impact of introducing rail service improvements can be captured by the change in “rail competitiveness” that results; thus, the impact can be extrapolated by observing the effect of existing variation in that factor. Ways of measuring this factor are discussed further below. An alternative would be to expand the geographic scope of the analysis to Europe and Japan, where rail service improvements such as HSR are already present. Limits on data availability and questions about the applicability of the results to the United States meant that the Research Team did not pursue this approach.

The model development started with an investigation of the difference between non-stop air markets with and without rail service. To begin, the non-stop market data were segmented into markets (origin-destination pairs of CBSAs) with no rail service, markets with through rail service, and markets with connecting rail service that requires one transfer; also, the cumulative distribution of the ratio of flights in a market to the enplaned passengers summed over all markets served by the origin and destination (with annual enplaned passengers in thousands) were reviewed. The enplaned passengers summed over all markets served by the origin and destination were used to recognize the network aspect of airline service, where the level of air service between two airports is a function of the level of air travel demand for more than just that particular origin and destination; the following analysis using the population in the origin and destination markets was intended to understand the particular origin to destination market in isolation.

This shows that, for example, 60% of the markets with no rail service and 80% of the markets with through or connecting rail service are in markets where the ratio of flights to enplaned passengers is 10% or less. In fact, over the full range of values on the x-axis, the Cumulative Distribution Function (CDF) representing markets with rail service is greater than the CDF for markets without rail service. In short, when the frequency in a market and total demand at both the origin and destination of the market is controlled for, markets with rail service have less air service than markets without rail service.

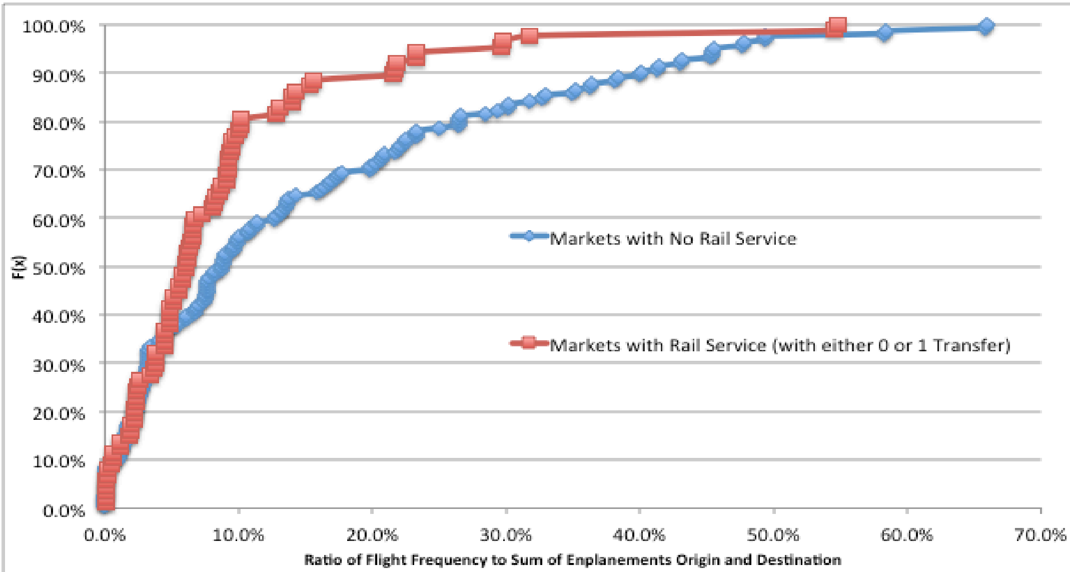

Figure 16. Cumulative distribution function for the ratio of flight frequency to the sum of enplanements at market origins and destinations for markets without rail service and with through or connecting rail service.

The same analysis was performed to limit markets with rail service to markets with only through rail service. The cumulative distribution of the ratio of flights in a market to the annual enplaned passengers summed over all markets served by the origin and destination is shown in Figure 17. The finding is consistent with that of Figure 16.

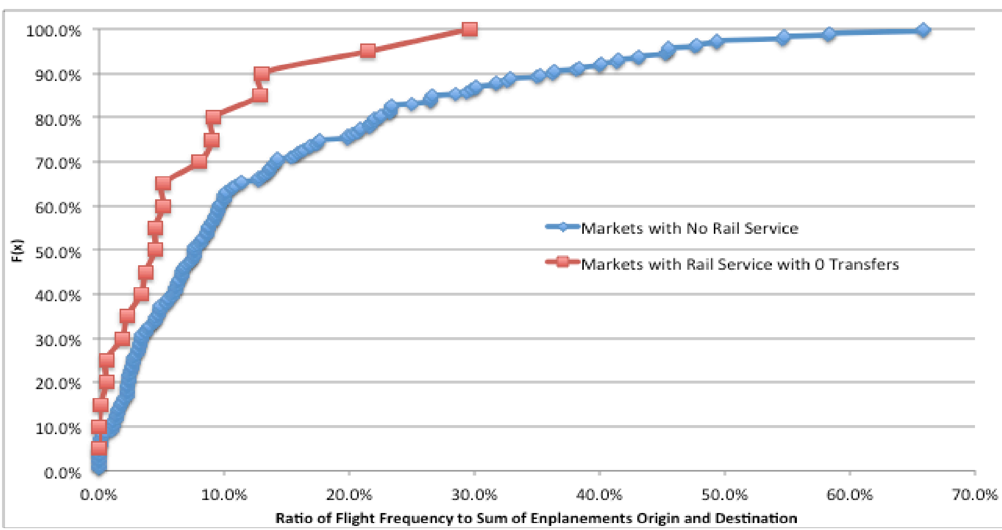

Figure 17. Cumulative distribution function for the ratio of annual flight frequency in 2008 to the sum of enplanements at market origins and destinations for markets without rail service and with through or connecting rail service.

The Research Team next reviewed the cumulative distribution of the ratio of market flight frequency to the sum of the population at the origin and destination of that market (in thousands). Figure 18 considers the CDF for markets with through or connecting rail service and for markets without rail service. The differences between the two CDFs are less significant than when annual enplanements are considered, but the conclusion is the same: when frequency is controlled in a market along with total population at both the origin and destination of the market, the markets with rail service have less air service than markets without rail service. The same conclusion is reached when rail service is limited to through service (Figure 19).

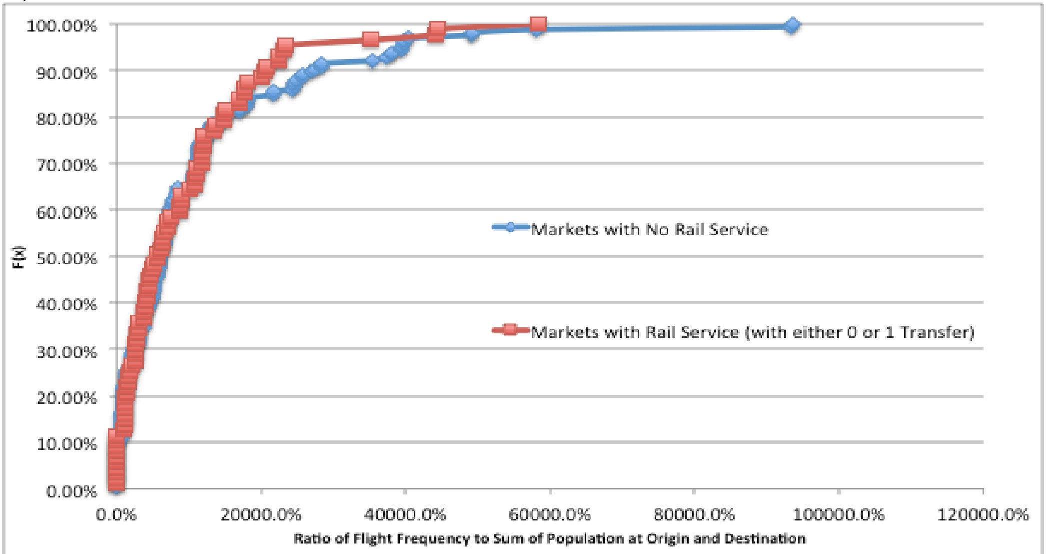

Figure 18. Cumulative distribution function for the ratio of annual flight frequency in 2008 to the sum of population at market origins and destinations for markets without rail service and with through or connecting rail service.

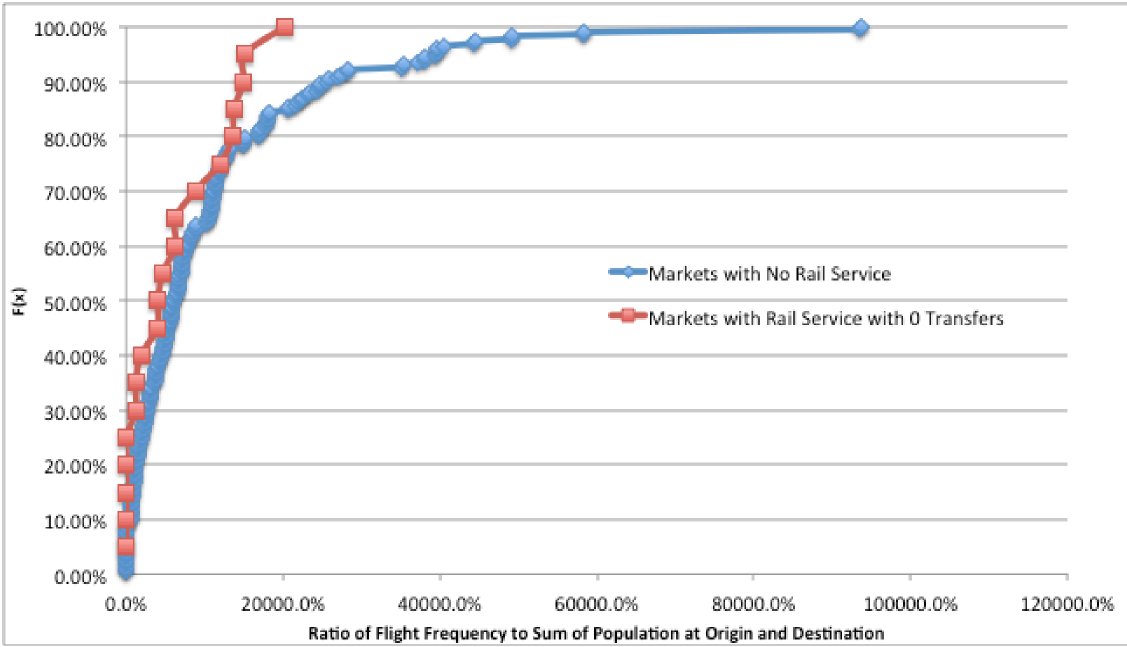

Figure 19. Cumulative distribution function for the ratio of annual flight frequency in 2008 to the sum of population at market origins and destinations for markets without rail service and with through or connecting rail service.

To further investigate how rail service affects air service, the Research Team performed a multiple regression analysis. Using a set of short- and medium-haul markets for which air and rail service data have been developed, a model was estimated in which the number of annual non-stop one-way air flights for a given origin to destination pair of CBSAs is the dependent variable. To avoid endogeneity bias, the model was estimated using only demand-side variables and qualitative rail service variables (those not directly related to the amount of air or rail travel). The unit of observation for the model is the CBSA pair.

The model specification is:

$$\ln(flts_{ij}) = \alpha_0 + \alpha_1 \cdot \ln(pop_i \cdot pop_j) + \alpha_2 \cdot \ln(hhinc_i \cdot hhinc_j) + \alpha_3 \cdot d500_{ij} + \beta_1 \cdot railserv_{ij} + \beta_2 \cdot \ln(railtime_{ij}) + \varepsilon$$

where:

$flts_{ij}$	Is the total flights between CBSA i and CBSA j in 2008
pop_i	Is the population of CBSA i in 2008
$hhinc_i$	Is the population weighted average of the median household incomes of the counties in CBSA i in 2008
$d500_{ij}$	Is a dummy variable indicating whether the distance between CBSA i and CBSA j is less than 500 miles
$railserv_{ij}$	Is a dummy variable indicating whether there is rail service between CBSA i and CBSA j
$railtime_{ij}$	Is geometric average of the rail travel time between CBSA i and CBSA j, or 0 if no rail service available
ε	Is a stochastic error term
$\alpha_0 - \alpha_3, \beta_1 - \beta_2$	Are coefficients to be estimated

The model was estimated on a set of 127 directional CBSA markets that had at least 150 flights in 2008. The results appear in Table 21. For comparison, estimation results for the same model without the rail service variables is also included. In both models, all coefficients are significant and have the expected signs. The R^2 is somewhat low, but this is to be expected given the simplicity of the model, in particular the exclusion of network effects. The number of flights increases with population and is strongly sensitive to household income. Segments under 500 miles also tend to have more flights. The presence of rail service has a strong effect on the number of flights, but this effect attenuates with rail travel time.

Table 21. Airline response model coefficients.

	With Rail Service		Without Rail Service	
Variable	Coefficient Estimate	p-value	Coefficient Estimate	p-value
intercept	-45.79706	<.0001	-46.36265	<.0001
$Pop_i \cdot Pop_j$	0.54979	<.0001	0.49510	<.0001
$Hhinc_i \cdot Hhinc_j$	1.61779	0.0003	1.71358	0.0001
$d500_{ij}$	1.60822	0.0267	1.45873	0.0362
$railserv_{ij}$	-2.36714	0.0305	--	--
$railtime_{ij}$	0.35683	0.0404	--	--
Model Statistics				
R^2	0.43		0.41	
N	127		127	

The units of the variables in the airline response model are as follows:

- Railtime is rail travel time in minutes.
- Hhinc is median household income in $ for 2008.
- Pop is 2008 population from the U.S. Census Bureau estimates.
- Flts is flights per year between CBSA i and CBSA j.

Figure 20 summarizes the implications of the model concerning the impact of rail service on air service. If rail travel times are around one hour, then this eliminates about 60% of the flights that would be expected without rail service. This effect attenuates so that a 7-hour rail service would result in about 80% of the flights expected under a non-rail scenario. The effect disappears entirely at a rail travel time of around 11 hours.

It should be noted that although the coefficient estimates are significant, given the overall R^2 value of the model indicates that the coefficients only explain a portion of the variation in air service, some uncertainty remains about the relationship in Figure 20. Further analysis, which is beyond the scope of this research project, is required to reduce this uncertainty. As estimated, the responsiveness of air service to changes in rail service is relatively high. However, given that the model is based on a cross sectional analysis as opposed to being based on changes over time, use of the model to forecast changes will ideally be validated in future research against before and after survey information that observed the airline response in a particular market when rail level of service was changed.

In addition to the remaining uncertainty, several recognized aspects of competition between air and rail were not included in the model, including frequency of rail service and rail fare levels. It is expected that both factors influence the market share that is attracted to rail, and hence the size of the residual air market, which in turn influences the number of flights needed to serve this market. Other considerations which were deemed beyond the scope of the current effort but should be considered in future research are the impacts of the number of airlines serving a market and the type of equipment used, and the number of airports serving a given metropolitan area.

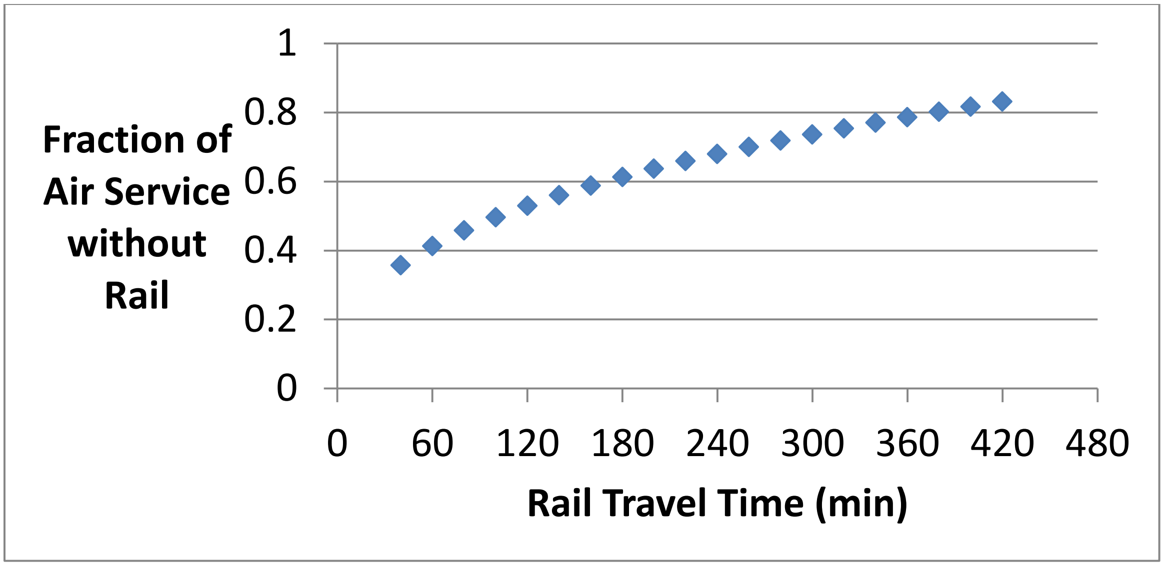

Figure 20. Impact of rail service and rail travel time on air service.

The airline response model follows the main mode choice in the model framework. In application it is used to estimate a change in air service if the modeled scenario includes a change in rail service. The model is applied as follows:

- For the "Base" scenario, which is the business as usual case for the model and is run first before alternative model scenarios are tested, the airline response model is applied using the base inputs for rail service. This produces a base estimate of air service according to the airline response model
- When an alternative scenario is run, the airline response model is run again with the new rail service as inputs. This produces an alternative estimate of air service according to the airline response model.
- The difference in air service estimates between the base and alternative scenario is then used to adjust the air service inputs used in the main mode choice model.
- The final step is to the reapply the main mode choice model using the adjusted air service inputs. The output from this second application of the main mode choice model is then output by the model.

Model Transferability and Updates

The Air/Rail Diversion Model is currently implemented in two study areas, on the East and West Coast. This section provides an overview of how the model could be both transferred to other regions outside of the two study areas where it has been implemented and how the model be updated, with some work by a competent analyst.

Model transfer would require the development of inputs for a new region and possibly some recalibration of the model but would not require code changes to either the R code used to simulate the model or to the Excel GUI's visual basic code. The input files are built from a combination of currently readily available public data, while the observed origin destination trip tables are expected to be published soon by the US DOT. The model user's guide explains more about the format of the input files and the text above describes data sources, but a comprehensive set of data processing tools to prepare the inputs for any region have not been developed as part of this research project.

Model updates, such as changes to the model specification of a choice model, would require edits to the open source R code used to run the model simulation. Since the code is not compiled, it is readily accessible to any model user and an analyst with competency in R programming could update the model. It is very unlikely that changes to the excel GUI would be required as the model simulation code and the excel GUI are separate applications with relatively loose connections that are not dependent on the details of the model itself. The model's user's guide and commentary within the R code itself provide an overview of the R code structure and guidance on how to adjust and update the model.