Making advanced travel forecasting models affordable through model transferability

A Research Project Sponsored by FHWA
under the Broad Agency Announcement DTFH61-10-R-00013

Final Report

September 7, 2013

Prepared by

John L. Bowman, Ph. D.
Bowman Research and Consulting
28 Beals Street, Brookline, MA 02446 USA
+1-617-232-8189 John_L_Bowman@alum.mit.edu http://JBowman.net

Mark Bradley
Bradley Research & Consulting
524 Arroyo Ave., Santa Barbara, CA 93109, USA.
+1-805-564-3908 mark_bradley@cox.net

Joe Castiglione
Resource Systems Group
600 B Street Suite 2202, San Diego CA 92101, USA
+1-617-251-5111 jcastiglione@rsginc.com

Table of Contents

1. Executive Overview

Research Problem Statement

The objective of this research is to empirically test and demonstrate the transferability of advanced (activity-based, or "AB") models between regions. If these methods are shown to be transferable, regions can borrow and adapt activity-based models from other regions, thereby avoiding most of the cost and expertise needed to field a large household survey and estimate entirely new models. Such a result could greatly accelerate the adoption of this advanced modeling method throughout the United States, especially among small and mid-sized agencies.

Using current data obtained through the 2008-2009 National Household Travel Survey (NHTS) "add-on" program, the principal investigators estimate advanced activity-based models simultaneously for six regions, four in California and two in Florida. Statistical tests are applied to identify regional differences in the model coefficients. The absence of statistically different estimates for key model coefficients across regions would provide strong empirical evidence of transferability.

Regions Included in the Study

This study examines six regions, including four in California and two in Florida, as listed in Table 1.1, with summary statistics, including the size of the NHTS sample, in Table 1.2.

Table 1.1: Regions included in the transferability study

REGION	AGENCIES INVOLVED
California Regions	
Fresno	Council of Fresno County Governments (FresnoCOG)
Northern San Joaquin Valley (NSJV. Or '3county')	Merced County Association of Governments (MCAG) San Joaquin Council of Governments (SJCOG) Stanislaus Council of Governments (StanCOG)
Sacramento	Sacramento Area Council of Governments (SACOG)
San Diego	San Diego Association of Governments (SANDAG)
Florida Regions	
Jacksonville	Florida DOT District 2, North Florida Transportation Planning Organization (NFTPO)
Tampa	Florida DOT District 7

Table 1.2: Basic statistics of the six regions in the transferability study

	Fresno	NSJV	Sacramento	San Diego	Jacksonville	Tampa
NHTS Households	380	660	1,310	6,000	1,050	2,500
REGION Households	288,857	458,731	805,292	1,081,082	551,353	1,361,724
Sq Miles	6,017	4,913	6,197	4,262	2,532	3,275
Number of TAZ	1,967	3,758	1,502	4,682	1,309	2,251
Number of Census Blocks	27,891	39,312	49,282	33,084	30,899	54,310
Avg Census Block Size (acres)	137.9	78.0	80.3	82.4	56.0	36.3
Median Census Block Size (acres)	5.6	4.5	5.4	6.7	4.8	5.1
HH / Sq Mi	48	93	130	254	218	416
Avg Households/Census Block	10.4	11.7	16.6	34.2	17.8	26.4
Avg Employment/Census Block	12.7	11.7	19.7	45.9	20.7	26.9
Employment (Total)	353,216	459,910	969,838	1,519,582	638,195	1,462,137
Education	35,768	52,028	73,688	149,540	35,787	106,766
Food	22,862	10,836	58,102	90,013	46,619	129,529
Government	36,837	16,290	67,103	205,780	72,801	70,147
Industrial	68,443	140,303	119,305	202,605	111,102	206,713
Medical	36,304	48,081	108,036	101,251	70,395	198,171
Office	44,521	54,789	204,288	170,027	112,511	358,637
Retail	35,946	55,881	131,781	151,504	87,031	192,973
Service	5,393	24,609	202,884	424,454	100,445	98,815
Agricultual/Resource/ Construction	67,142	57,093	4,652	24,408	1,504	100,386
Avg commute time (2006-2008 ACS)	23.9	30.0	28.5	27.9	28.6	28.4
Modeled modes (in region's original model)	5	5	10	25	8	11
No. auto skim periods	4	4	5	6	4	2
No. transit skim periods	2	2	5	6	2	2
Network software	Cube	Cube	Cube	TransCAD	Cube	Cube

Transferability

Any travel demand model, whether it be a standard "4 step" model or an advanced activity-based model, consists of two key components: the input data, and the parameterized relationships used to "translate" the input data into output predictions of travel demand. The critical question addressed in this study is whether or not the parameterized relationships in advanced travel demand models are transferable. In discrete choice travel demand models, these relationships are typically in the form of utility functions: $U_{ia} = a_a + bx_{ia} + e_{ia}$. The question is how transferable the values of the coefficients represented by *a* and *b* are from one regional context to another. The two main approaches available for answering this question are:

(a) **Application-based**: Estimate utility coefficients based on observed choices in one region, apply that model to data from a different region, and see how well the model predicts observed choices in the other region.

(b) **Estimation-based**: Estimate utility coefficients based on observed choices from two regions, using the same exact model specification for both regions, and test whether the resulting estimated utility coefficients are statistically different from one another.

The estimation-based approach, which is adopted for this study, has important advantages over the application-based approach. The most obvious advantage is that the estimation-based approach allows for explicit statistical tests of the differences in coefficients, and thus the ability to address a wide variety of hypotheses regarding transferability. A related advantage is that the estimation-based approach can be used to test the transferability of the coefficients for specific types of variables, while the application-based approach can only test the transferability of the model as a whole.

A potential drawback of the estimation-based approach is that it is such a strong statistical test that any hypothesis of transferability is likely to be rejected unless the input data from the different regions are highly consistent. In other words, the statistical test will confound the effects of data inconsistency with the effects of different underlying relationships. For this reason, this research project places a strong emphasis on data quality and consistency. As discussed in later chapters, the study is designed to meet the necessary data consistency condition, using data from the same NHTS survey for all regions, and taking great care to make the spatial land use data and network skim data from the various regions as consistent as possible.

Another important data issue affecting the study of transferability relates to the quantity of the survey data used to estimate the model coefficients. If there is not enough data when a model is estimated, some important coefficients may be statistically insignificant and others may be completely inestimable. As with the data consistency issue, problems associated with inadequate quantity of data tend to confound the tests of transferability. However, since the quantity of data cannot be increased in this study to eliminate the issue, efforts are made to minimize the confounding of results arising from inadequate quantity of data. In particular, the summarization of results accounts for inestimability of coefficients arising from small sample sizes. Also, in evaluating the results, efforts are made to point out situations where the effects of data quantity might be misinterpreted as transferability effects.

The summary of findings for this study identifies the hypotheses tested, the conclusions drawn, and the limits imposed by unavoidable data issues.

Activity-Based Model Framework

The models estimated in this project are components of the activity-based (AB) model framework commonly known as DaySim, which has been used in Sacramento since 2007 and Seattle since 2009, and has been enhanced and implemented for use in all of the regions of this project except San Diego. The DaySim framework is used for this study as a matter of convenience: it is the framework that has been developed by the principals in this study. The reader is referred to Chapter 4 for more details about DaySim. This transferability project includes 14 of the 21 different DaySim models, including all of the most important models, so it is uniquely comprehensive among existing transferability tests for AB model systems. The DaySim models are listed below, with the models shown in bold font that are included in this study:

Regular Workplace Location
Regular School Location
Auto Ownership
Daily Activity Pattern
Exact Number of Tours
Work Tour Primary Destination Choice
Other Tour Primary Destination Choice
Work-Based Subtour Generation
Work Tour Main Mode Choice (two versions tested, as explained in Chapter 4)
School Tour Main Mode Choice
Escort Tour Main Mode Choice
Work-based subtour Main Mode Choice
Other Tour Main Mode Choice
Work Tour Time Period Choice
School Tour Time Period Choice
Work-based subtour Time Period Choice
Other Tour Time Period Choice
Intermediate Stop Generation
Intermediate Stop Location
Trip Mode Choice
Trip Departure Time

The wide variety of variables in these models includes person and household characteristics, Census Block-level land use and accessibility, zone-to-zone accessibility, endogenous variables related to the predicted activity pattern and related schedule pressure, endogenous variables related to people who work from home, or who use various modes to get to work. For purposes of summarizing results of the transferability tests, each model variable is classified into one of the following types (two types if it is an interaction variable):

1. A-constant
2. P-person
3. H-household
4. D-day pattern

5. T-tour/trip
6. I-impedance
7. U-land use
8. W-time window
9. C-logsum
10. G-size variable
11. L-log size multiplier

Model Estimation and Transferability Testing Approach

Base Estimation Approach

The models listed above are estimated using a common, consistent data set spanning all six regions. In particular:

- The observed choices are taken from a household travel survey, which has been processed in such a way as to reflect the assumptions and conditional relationships among choice components of the DaySim model system. There are records representing each household, person, household-day, person-day, tour and trip.

- For each observed choice, relevant information is drawn from a file of Census Block attributes and files of zone-to-zone impedance information (road and transit skims).

- Estimation of upper level models uses composite variables (logsums) calculated from the models of lower level choices.

The Transferability Testing Process

The transferability tests are set up and executed using the following sequence of steps:

(1) For each choice model to be tested, a base model specification is developed, including all explanatory variables to be tested. For this study the SACOG version is used as the base specification because it is the most rigorously developed specification within the DaySim family and serves as the basis of the other existing DaySim implementations. Also, the current transferability study lacks budget for the development of a new specification to serve as the basis of comparison. Each variable in each model is given a one- or two-letter code denoting what type of variable it is.

(2) For each of the 15 models, the base model estimation data set is created for each of the 6 regions, giving 90 separate estimation data sets. This is done using the DaySim software, which uses the same model code for both model estimation and application. The models had already been coded into DaySim and tested extensively in application for other projects, which helped to avoid errors in setting up the estimation data, and to ensure consistency in the data processing across the six regions. DaySim automatically creates consistent data and control files for use by the ALOGIT model estimation software package. For each of the 90 base models (15 model types times 6 regions), the base model is estimated using ALOGIT.

Any variables that cannot be estimated in the base version are dealt with by leaving the variable in the model but constraining the coefficient to a specific value (usually 0).

(3) A program created for this project reads the ALOGIT estimation results (extension .f12) files for the 90 different base runs, and automatically creates new data files and control files to estimate 36 different model specifications for each of the 15 models. For a specified one of the 15 model types, the program merges the separate estimation data files for the 6 regions into a single combined estimation data file, creates ALOGIT estimation control files for all 36 models used in the transferability tests, and creates an ALOGIT "batch run" file instructing the ALOGIT software to estimate all 36 models in succession.

(4) After the 36 model runs for all 15 model types are completed successfully, a second program created for this project compiles and tabulates all estimation results for further viewing and analysis. It reads in the estimation results (.F12) files, and writes out two types of files:

   a. A single, space-delimited metafile, containing a record for each coefficient in each estimated model. The content of this file is documented in Appendix 2.

   b. Comma-delimited (.csv) files showing all estimation results in tabular form. This is available as a workbook (xlsx format), separate from this report.

(5) The metadata file from the previous step is analyzed using an SPSS syntax file created for this project that creates the tables and charts presented in Chapter 6.

The model specifications for transferability tests

As noted above, up to 36 different models are estimated for each model type. These fall into two main types: the "base models", and the "difference models".

Twelve of the models are base models. Six of them are the region-specific models. The other six fall into four different types:

- **2-state**: Using the data from all 6 regions in both states
- **1-state**: (a) Using only the data from the 4 California regions, and (b) using only the data from the 2 Florida regions
- **2-state+ASC**: Using the data from all 6 regions in both states, but using separate alternative-specific constants for each region.
- **1-state+ASC**: (a) Using only the data from the 4 California regions, and (b) using only the data from the 2 Florida regions, but using separate alternative-specific constants for each region in that state.

These six models are used as a basis for comparison for the remaining 24 "difference models". The concept behind the difference models is as follows:

- Estimate a model across data from multiple regions (6 regions for the 2-state models, and either 4 or 2 regions for the 1-state models).

- For a single selected region, add a second set of all the coefficients that are in the base model. These are referred to as “difference variables”, because if the coefficient value that is estimated across all included regions is the same as the coefficient value that represents the single selected region, then the estimated value for the second, region-specific coefficient will be 0. The second coefficient essentially measures the difference between the correct coefficient value for the selected region and the correct value for the sum of all regions **except** the selected region.

- Estimating a full model with difference variables gives transferability evidence in two ways:

    - The significance of the difference coefficient on each variable provides evidence of the transferability of that particular variable between the selected region and the other regions included in the model.

    - A chi-squared model fit test between the full difference model and the corresponding base model provides evidence on the transferability of the model as a whole. (Note that this is quite a strict test to use for data from different regions. It is more commonly used for data from a single region, to test the significance of adding or subtracting variables from a model.)

The 24 difference models include:

- 1-6: Difference variables for each of the 6 regions relative to the 2-state base

- 7-12: Difference variables for each of the 6 regions relative to the 1-state base

- 13-18: Difference variables for each of the 6 regions relative to the 2-state+ASC base

- 19-24: Difference variables for each of the 6 regions relative to the 1-state+ASC base

Each of the types of specifications, model types, and variable types can provide different types of evidence related to our research hypotheses presented in previous chapters. The following chapter presents a careful analysis of the resulting evidence.

Summary of Findings

Though it is not possible to make a definitive statement about the transferability of activity-based models based on a single study, this study has been able to provide some new and unique evidence. Overall, although the strictest statistical test (chi-squared test) usually rejects the hypothesis that models based on data from different regions are statistically indistinguishable, it is also true that most of the individual coefficients are not significantly different from one region to the next. In addition, this study shows the substantial improvement of estimability that occurs with large survey samples. Based on these findings, **the most important conclusion of this study is that, although estimation of models using a large local sample is best, it is better to transfer models that are based on a large sample from a comparable region than it is to estimate new models using a much smaller local sample.**

However, this conclusion does **not** mean that metropolitan regions can relegate survey data collection and model development to the past, and simply borrow a model from others who have gone before. Even if a comparable region and its model can be found, survey data should nevertheless be collected for purposes of calibrating components of the model, such as activity and tour generation, that cannot be calibrated using traffic count data. And this study has only scratched the surface in identifying the factors that make two regions comparable, as described in the next paragraphs.

Although small sample sizes limit the ability to draw strong conclusions about comparability among the four California regions and two Florida regions included in this study, there is some substantial evidence of comparability among them. This is shown in Figure 1.1, where it can be seen that, for all regions, the differences from the 2-state model are insignificant for over 80% of the coefficients. However, Tampa stands out as less comparable than the others, and this study did not identify the reason. The California regions are more comparable within state than across states, perhaps because of the presence of Tampa in the two-state comparison. The issue with Tampa draws attention to the likelihood that there may be factors that would cause two regions, even two regions within the same state, to be bad candidates for a model transfer.

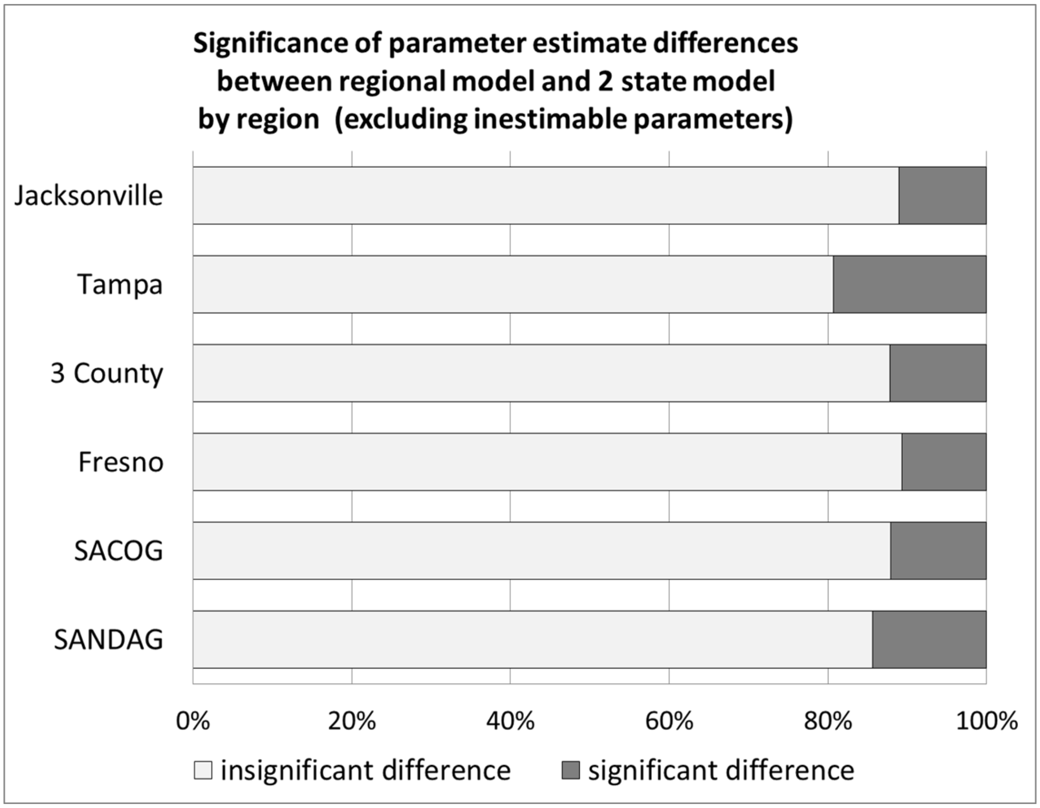


Figure 1.1

This study did not explore comparability for regions in states other than California and Florida, estimability and comparability for a full spectrum of sample sizes—especially samples with more than 2,500 households, or comparability in categories other than state boundaries, such as urban density, size, or demographic make-up. For example, university towns or cities with large seasonal retirement population may be distinctly different in ways that make transferring from other regions inadvisable, and this study lacks evidence to draw conclusions one way or the other. These remain as important avenues for further research.

On the other hand, there may be good reasons for transferring a model from a region that is NOT currently comparable if there is reason to believe that it will be comparable in the future. For example, it may be that a region that is growing rapidly and/or adding new travel options would lack the data to develop a model that would serve it well even if it could conduct a very large household survey. The diversity of conditions needed to estimate the coefficients of the model simply may not exist within the region. In a case like that, perhaps a model transfer should be considered.

This study is also limited in its ability to determine what sample size is large enough for local estimation, because the largest sample in this study includes only 6,000 households and the rest are 2,500 or less. However, as shown in Figure 1.2, where estimation results for each region are compared to the 2-state combined models, the results are clear that **a sample of 6,000**

households provides much better information for estimating AB model coefficients than samples of size 2,500 or less. It is also likely that samples considerably larger than 6,000 would substantially improve estimation results, enabling significant coefficient estimates for important small population segments.

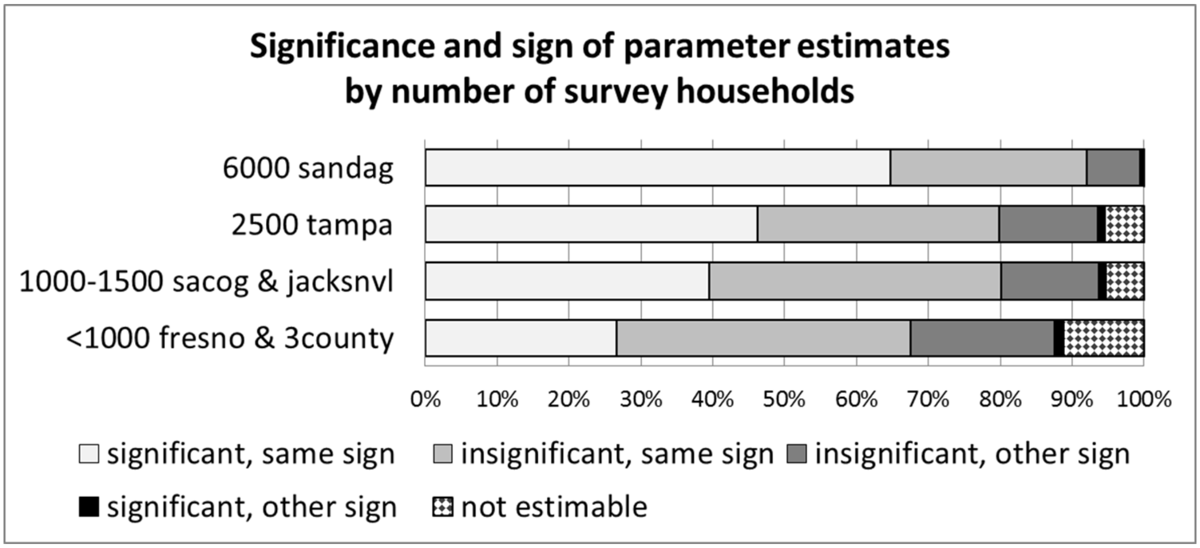


Figure 1.2

The following sections provide more context to these overall conclusions, as well as additional findings about the NHTS, sample size issues, and transferability hypotheses tested in this study.

Adequacy of the 2009 National Household Travel Survey (NHTS) data for estimating activity-based models

A unique aspect of this project is that identical models are estimated using data from six different regions, but all using data from the same household travel survey. Because the opportunity to purchase an add-on sample to the NHTS is available to all state and regional agencies, it may be useful to assess how well the data support the estimation of the component models of the activity-based model system (the DaySim v.1.8 model system in this case). The findings are:

- In terms of survey design, the NHTS 2009 survey supports the estimation of almost all of the component models of advanced AB model systems. Three notable exceptions are student's usual school location, transit pass ownership, and available of free parking at the usual workplace, which should be added to future NHTS surveys.

- In terms of sample design, there are some more serious issues:

    - The sample sizes for most of the individual regions used in this study are not adequate to estimate statistically significant parameters for many of the variables tested. This is particularly true for some of the rarer types of households and persons, such as low-income households, zero-vehicle households, and persons who use transit and bicycle.

- o There are many households in the sample for which one or more household members did not complete the travel diary. The result is that households with incomplete data must be excluded from estimation for some models, which exacerbates any sample size problems, particularly among larger households.

Addressing the above issues would increase the usability of future NHTS data for AB model development. The above findings related to the NHTS data should also be considered in the development of other household surveys for use with AB models.

Issues regarding sample size and composition

The adequacy of the survey sample size can be judged mainly from (a) how many of the coefficients in the models can be estimated with statistically significant precision (i.e. t-statistics of 1.9 or higher), and (b) how many of the coefficients can be estimated at all. The latter issue of non-estimable parameters usually arises in the case of variables that apply to only small segments of the population and for which no variation in choice behavior can be observed. An example is a case in which no households in the lowest income category choose the walk mode, so a low-income variable for the walk mode cannot be estimated.

Overall sample size has a large impact on coefficient estimability, with the inability to estimate a coefficient at all decreasing with sample size, and the ability to estimate significant coefficents of the correct sign increasing with sample size. Because the largest sample in this study included 6,000 households, conclusions about the benefits of samples larger than 6,000 cannot be drawn. Nevertheless, what can be concluded is that, among the sample sizes included in this study, much is to be gained from larger sample size.

The variation in results in response to sample size appears to affect all types of models, including mode choice, destination choice, time of day choice, tour and trip generation, and auto ownership. The problems associated with small samples arise in all models in the estimation of coefficients related to important but small or hard-to-sample segments of the population, such as very low or very high income households, households that do not own cars, and young adult households. The effects are most apparent in cases where specific alternatives have a small number of observed cases.

Tests of specific hypotheses regarding model transferability

In this study, model transferability between regions is tested through the use of region-specific “difference models”. Starting with a “base model”, estimated on all regions within both states (or else all regions in a single state), additive “difference coefficients” are added for all variables for observations from a single region. The difference coefficients thus measure the difference between the coefficient estimated only for that specific region and the coefficient estimated for all regions except that specific region. If the difference coefficient is not significantly different from zero, it is an indication that the model is transferable between the specific region and the other regions included in the base model—at least for that particular variable. This approach tends to work best with adequate sample sizes. With small sample sizes, although it is more difficult to estimate a local version of the model, it is also more difficult to prove whether or not the local version is statistically different from models estimated on other regions, because there is less data to use in estimating the difference coefficients and thus they will tend to be less significant, regardless of the true transferability. Therefore, this test is less useful in judging variations in sample sizes, and the comparisons tend to be most meaningful along model dimensions not defined by region (where sample sizes differ), such as comparisons between types of models or between types of variables (including data from all regions without differentiating results by region).

Another way that transferability is tested in this study, rather than looking at each variable separately, is to look at the change in model fit as a whole when the region-specific difference variables are included. This is done with the standard chi-squared likelihood ratio test. Such a test is typically used to test different model specifications on the same estimation data set, whereas in this study it is used to test the same model specification on different estimation data sets. The likelihood ratio test is a rather strict test to use, because it tests whether the two models being compared are the same in every respect. Indeed the hypothesis that the different data sets yield "the same model" (in statistical terms) is rejected with fairly high significance in most of the cases. Nevertheless, the results of the likelihood ratio test can be used in a comparative manner, finding out which models show the most and the least improvement in model fit due to allowing region-specific coefficients.

Each of the transferability hypotheses raised in Chapter 3 and tested in Chapter 7 is discussed below, with attention to the practical implications of the research findings:

(Hypothesis 1) ***Variables that apply to population segments defined by characteristics of individuals and/or their situational context (i.e., segment-specific variables) will tend to be more transferable than variables that are more generic and apply to all individuals*.**

Examples of generic variables, as the term is used here, include alternative-specific constants, cost and time variables, if they apply equally to all members of the population. Segment-specific variables include the same types of variables (e.g., alternative-specific constants, cost and time variables) that apply only to a specific segment of the population, such as females with children (i.e., the variable equals zero for all other members of the population). As the term is used here, segment-specific variables also include variables, such as income or age, that differ among the members of the population, causing the effect of the coefficient in the utility function to differ according to, for example, income or age.

This hypothesis is supported by the data, with 90% of the estimable coefficients that are population-segment-specific showing no significant difference from the base model, compared to 82% for coefficients that apply to the entire population, as illustrated in Figure 1.3. Of course, estimating coefficients for segment-specific variables requires adequate sample size for each segment, which requires larger sample size overall. The smaller effective sample sizes applying to these variables in estimation may be part of the reason that fewer significant differences result. Nevertheless, the results support the theory that variables with more socio-demographic specificity should show more stability across regions. This finding also supports the idea that it may be better to transfer activity-based models from a region with a large enough survey sample to include segmentation detail, rather than estimating on a much smaller local survey sample that does not allow such detail.

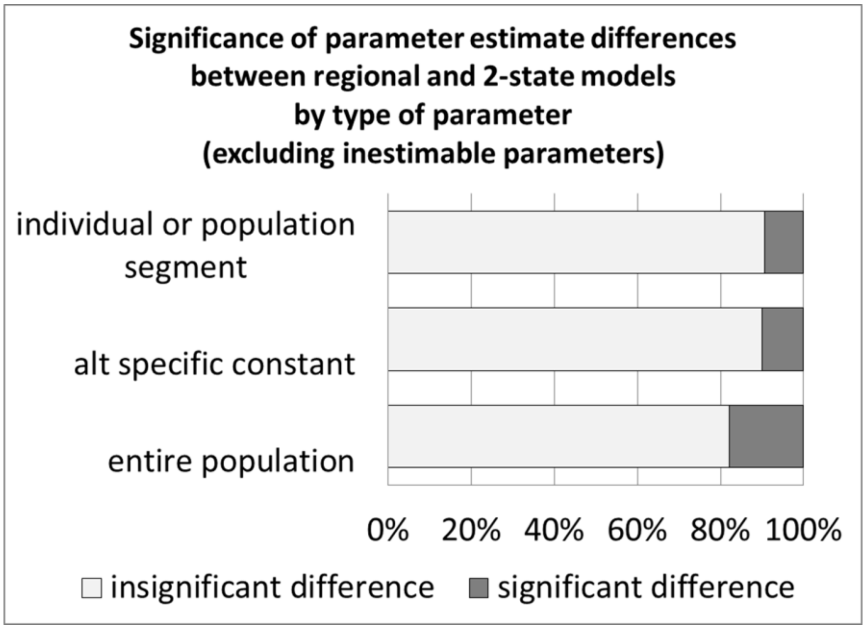


Figure 1.3

(Hypothesis 2) ***Variables that are segment-specific will tend to be more transferable than alternative-specific constants.***

This hypothesis is not supported by the tests. Compared to the segment-specific variables, the alternative-specific constants (ASCs) show a slightly lower percent of estimable coefficients across all models with significant differences from the base model (9.6% vs. 9.9%, as shown in Figure 1.3). The ASCs are also more estimable, with only 5% of cases not estimable, compared to 9% of segment-specific variables. This finding implies that alternative-specific constants may be fairly transferable between regions as well, which would be reassuring in a situation when there is very little observed local data at all to use in calibrating the model to local conditions. When such data do exist, however, it is still good practice to use local data to re-calibrate the alternative-specific constants of a transferred model whenever necessary. Although data sets from surveys such as NHTS may not be adequate for complete re-estimation of an activity-based model system, they will typically be adequate for such a simple calibration exercise.

(Hypothesis 3) ***Models that deal with social organization (activity generation and scheduling) will tend to be more transferable than models that deal mainly with spatial organization (mode choice and location choice)*.**

The data strongly support this hypothesis. As shown in Figure 1.4, the models emphasizing “spatial organization” result in 16% of coefficients with significant differences across regions and 16% non-estimable on single-region data, compared to only 10% and 5% respectively for models emphasizing “social organization”. This conforms to the expectation that different US regions will tend to show stability in general socio-cultural patterns, as those are less influenced by the spatial dispersion of opportunities or the availability and quality of particular travel modes. In practical terms, this finding suggests that when transferring an activity-based model system across regions, it would be warranted to spend more effort in calibrating and adjusting the mode choice and destination choice models than the other models.

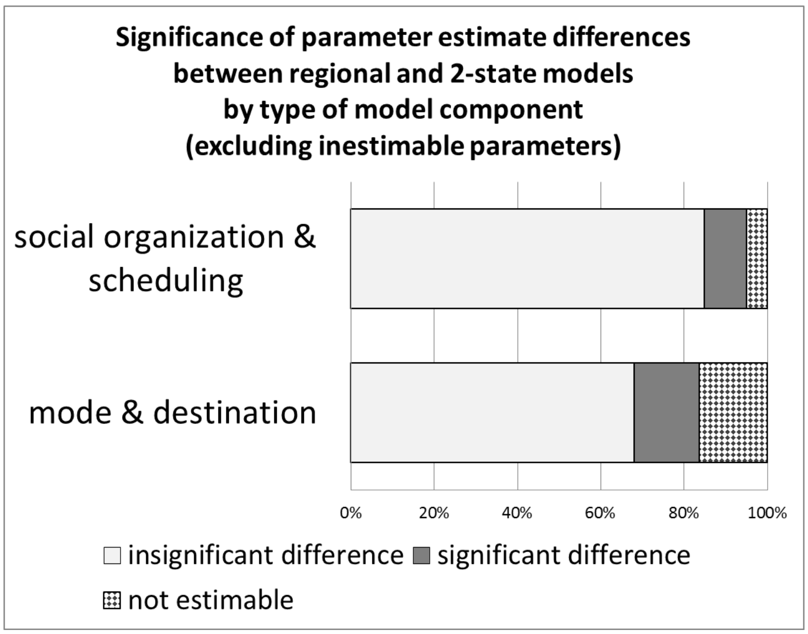


Figure 1.4

(Hypothesis 4) ***Models for different regions within the same state will tend to be more transferable than models for regions in different parts of the country*****:**

In most cases, the data support this finding (see Figure 1.5), with the California regions showing higher transferability for the one-state California base model than for the two-state base model, and with Tampa showing higher transferability for the one-state Florida model than for the two-state model. The exception is that the Jacksonville region shows somewhat less transferability with the one-state Florida model (i.e. with Tampa) than with the two-state model. This suggests that in at least some aspects of travel behavior, Jacksonville residents behave more like residents of one or more of the California regions than like Tampa residents.

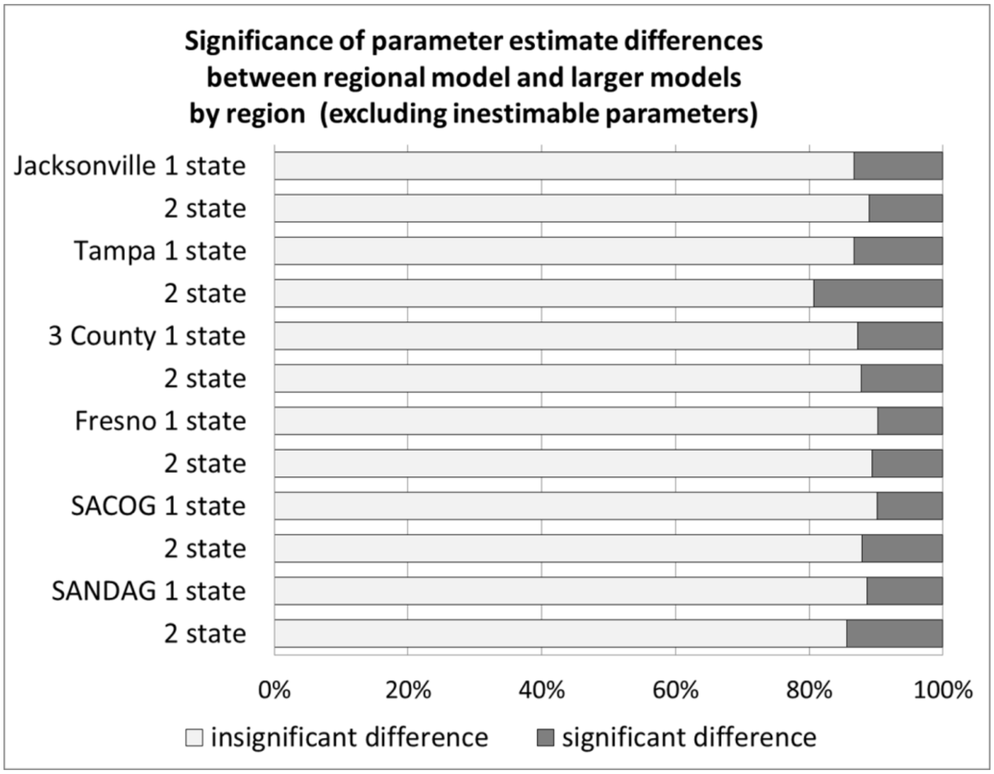

Significance of parameter estimate differences between regional model and larger models by region (excluding inestimable parameters)
Jacksonville 1 state
2 state
Tampa 1 state
2 state
3 County 1 state
2 state
Fresno 1 state
2 state
SACOG 1 state
2 state
SANDAG 1 state
2 state
0%
20%
40%
60%
80%
100%
insignificant difference
significant difference

Figure 1.5

2. Introduction

Research problem statement

The objective of this research is to empirically test and demonstrate the transferability of advanced (activity-based, or "AB") models between regions. If these methods are shown to be transferable, regions can borrow and adapt activity-based models from other regions, thereby avoiding most of the cost and expertise needed to field a large household survey and estimate entirely new models. Such a result could greatly accelerate the adoption of this advanced modeling method throughout the United States, especially among small and mid-sized agencies.

Using current data obtained through the 2008-2009 National Household Travel Survey (NHTS) "add-on" program, the principal investigators estimate advanced activity-based models simultaneously for six regions, four in California and two in Florida. Statistical tests are applied to identify regional differences in the model coefficients. The absence of statistically different estimates for key model coefficients across regions would provide strong empirical evidence of transferability.

Regions included in the study

This study examines six regions, including four in California and two in Florida, as listed in Table 2.1 and shown in Figure 2.1.

Table 2.1: Regions included in the transferability study

REGION	AGENCIES INVOLVED
California Regions	
Fresno	Council of Fresno County Governments (FresnoCOG)
Northern San Joaquin Valley (NSJV. Or '3county')	Merced County Association of Governments (MCAG) San Joaquin Council of Governments (SJCOG) Stanislaus Council of Governments (StanCOG)
Sacramento	Sacramento Area Council of Governments (SACOG)
San Diego	San Diego Association of Governments (SANDAG)
Florida Regions	
Jacksonville	Florida DOT District 2, North Florida Transportation Planning Organization (NFTPO)
Tampa	Florida DOT District 7

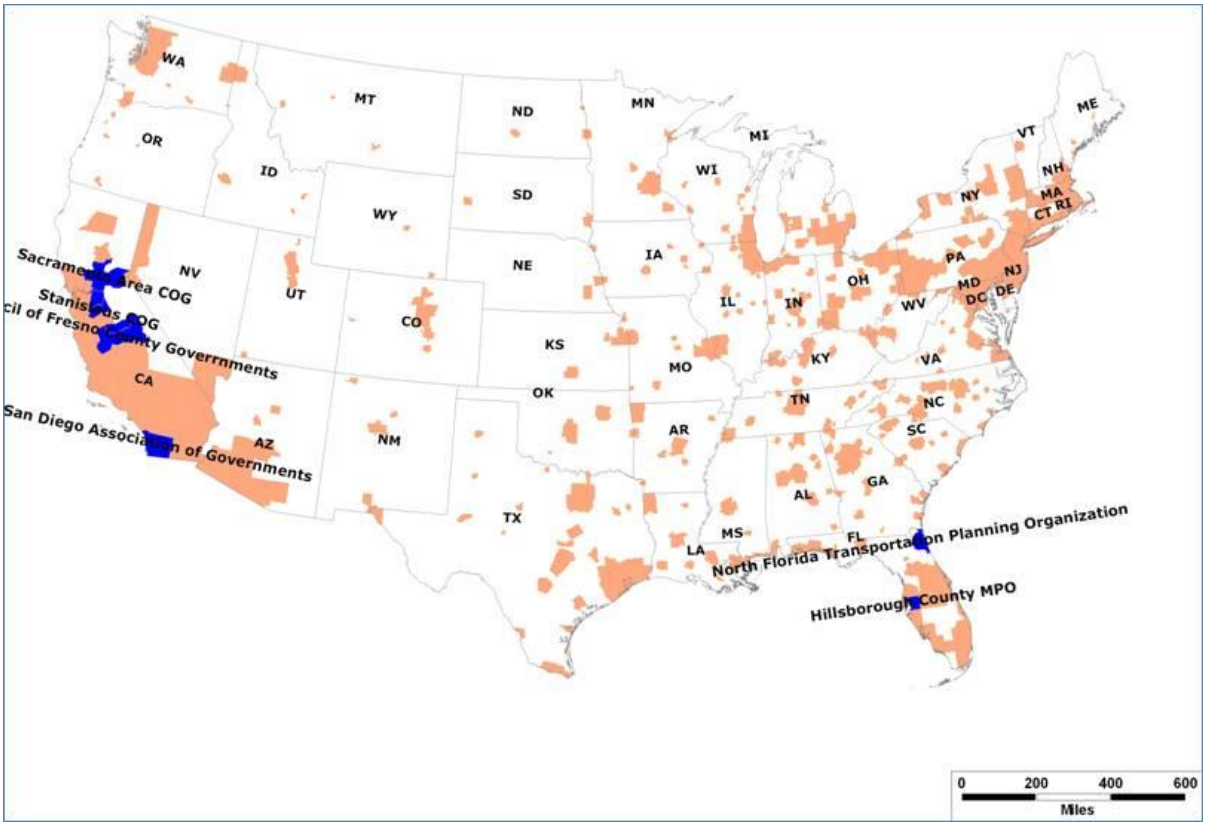


Figure 2.1 --Regions included in the transferability study

Organization of this document

This document discusses the idea of transferability, describes the regions, data, models and methods used in the study, and presents the results, as well as supporting documentation.

Chapter 3 provides a general discussion of issues related to travel demand model transferability. It includes specific research hypotheses and questions that are tested and answered using the research approach.

Following the discussion of transferability, the next two chapters provide further background information for the study. Chapter 4 provides background information about the specific activity-based travel demand models for which transferability is tested. Chapter 5 describes the regions included in the study and explains the data that is required by the models and how it is prepared.

Chapter 6 describes in detail the research approach used to test the hypotheses set out in Chapter 3. It includes a conceptual explanation of the methods used to formulate the models and conduct the tests, as well as a description of the mechanics employed to manage the estimation and summarization of a very large number of models.

Chapter 7 presents the research results, directly addressing the hypotheses and questions laid out in Chapter 3. Chapter 8 draws conclusions, compares this study to two other recetn studies and recommends avenues of further research.

Appendix 1 describes the NHTS survey data. Appendix 2 is a data dictionary of the data file produced in the estimation and testing process containing the information used for the analysis of results. Appendix 3 provides summary statistics for all base and difference models. Appendix 4 summarizes estimation results by detailed variable type. A separate workbook in .xlsx format provides detailed estimation results and summary statistics for each model.

3. Model Transferability

Key Components of Travel Demand Models: Data and Parameterized Relationships

Any travel demand model, whether it be a standard "4 step" model or an advanced activity-based model, consists of two key components: the input data, and the parameterized relationships used to "translate" the input data into output predictions of travel demand. All travel demand models tend to use the same types of input data, including:

1. Spatial land use data on households, employment, and school enrolment
2. Road and transit networks, and corresponding zone-to-zone measures of travel times and costs (typically referred to as network "skims")
3. Data on observed travel choices—typically in the form of travel/activity diary data from a sample of households.
4. Various data on "auxiliary" types of travel demand, such as freight trips, external trips to/from areas outside the region, and any special generators such as airports.
5. Other data used for model validation, such as traffic counts and transit passenger counts.

The types of parameterized relationships included in travel demand models include:

1. Models of mode choice (and related network route choice functions)
2. Models of time of day choice
3. Models of location choice
4. Models of tour and/or trip generation
5. Models of auto ownership and availability

All applied model systems are built from these same essential elements, with key distinctions arising in the level of detail at which they are treated. For example, most 4-step model systems use simple gravity models of trip end location ("distribution") at the zonal level, while advanced activity-based model systems have various types of location models such as usual work and school locations, tour primary destinations, and intermediate stop destinations. In addition, some advanced models treat location choice at the levels of individual parcels or Census blocks rather than more aggregate traffic analysis zones (TAZ). As another example, standard 4-step models typically use fixed time of day factors, while more advanced approaches include explicit models of activity scheduling and trip departure time choice.

Using Model Estimation to Test Transferability of Parameterized Relationships

The critical question addressed in this study is whether or not the parameterized relationships in advanced travel demand models are transferable. In discrete choice travel demand models, these relationships are typically in the form of utility functions:

$$U_{ia} = a_a + bx_{ia} + e_{ia} \quad , \tag{1}$$

where:

- U_{ia} is the utility of individual i for alternative a
- a_a is an alternative-specific constant for alternative a
- x_{ia} is a vector of attributes of the alternative a and/or the individual i
- b is a vector of utility coefficients corresponding to the variables in vector x
- e_{ia} is the "residual error term", capturing all non-systematic effects that are not captured by the other utility components.

The question is how transferable the values of the coefficients represented by a and b are from one regional context to another. Another way of stating the question is: If individuals in two different regions share the same person and household characteristics and face the same sets of alternatives (meaning that they have the same vector of explanatory variables in x), will they have the same probability of choosing each alternative?

In reality, one can never hope to find exactly the same choice alternatives in different regions, so it is necessary to address the transferability question in a less direct manner. In practice, the two main approaches available are:

(c) **Application-based**: Estimate utility coefficients based on observed choices in one region, apply that model to data from a different region, and see how well the model predicts observed choices in the other region. Although not typically called a "transferability test", this approach is carried out quite often in practice, when a model from one region is taken and re-calibrated to choice data in another region.

(d) **Estimation-based**: Estimate utility coefficients based on observed choices from two regions, using the same exact model specification for both regions, and test whether the resulting estimated utility coefficients are statistically different from one another.

The estimation-based approach, which is adopted for this study, has important advantages over the application-based approach. The most obvious advantage is that the estimation-based approach allows for explicit statistical tests of the differences in coefficients, and thus the ability to address a wide variety of hypotheses regarding transferability. A related advantage is that the estimation-based approach can be used to test the transferability of the coefficients for specific types of variables, while the application-based approach can only test the transferability of the model as a whole. This feature makes the estimation-based approach a much stronger test of transferability, because in the application-based approach, just by random chance, a model might happen to predict choices quite well even though the individual coefficients in the models are not appropriate for the region.

The Impact of Data on the Transferability Analysis

A potential drawback of the estimation-based approach, and perhaps the reason why it has not been used often in practice, is that it is such a strong statistical test that any hypothesis of transferability is likely to be rejected unless the input data from the different regions are highly consistent. In other words, the statistical test will confound the effects of data inconsistency with the effects of different underlying relationships. Data inconsistency can arise from methods of collecting the data, the level of aggregation used, the units of measurement used, the categorization and coding of survey questions, and so forth. A necessary condition for model transferability is that the data specification be as compatible as possible between regions. This issue applies not only to survey data, but also to other input data related to land use and travel supply. For this reason, this research project places a strong emphasis on data quality and consistency. As discussed in later chapters, the study is designed to meet the necessary data consistency condition, using data from the same NHTS survey for all regions, and taking great care to make the spatial land use data and network skim data from the various regions as consistent as possible.

Another important data issue affecting the study of transferability relates to the quantity of the survey data used to estimate the model coefficients. If there is not enough data when a model is estimated, some important coefficients may be statistically insignificant and others may be completely inestimable.

As with the data consistency issue, problems associated with inadequate quantity of data tend to confound the tests of transferability. However, since the quantity of data cannot be increased in this study to eliminate the issue, efforts are made to minimize the confounding of results arising from inadequate quantity of data. In particular, the summarization of results accounts for inestimability of coefficients arising from small sample sizes. Also, in evaluating the results, efforts are made to point out situations where the effects of data quantity might be misinterpreted as transferability effects.

On the other hand, this research has the opportunity to address questions about how much sample is required for estimating the models required by advanced travel demand model systems because the six regions in the study have widely varying survey sample sizes. These questions are important because the high cost of data is one of the big reasons that model transferability is attractive. In this project, the following three questions related to data and estimability are raised, and subsequently addressed:

Question E1: What sample size is adequate for local estimation?

Question E2: How does combining data samples improve estimability?

Question E3: Which models are more estimable at the regional level?

Transferability Hypotheses Addressed in this Research

In forming hypotheses about transferability, it is useful to relate back to the generic utility equation (1). It is commonly accepted that the more precisely an independent variable in vector *x* is defined, the more likely it is that the corresponding utility coefficient in vector *b* will be transferable across regions. Therefore, coefficients associated with variables that apply to population segments defined by the characteristics of the specific person, tour or trip are probably more transferable than coefficients associated with variables that apply uniformly to all segments of the population. As an example, consider a coefficient for travel cost. It is well-established in travel demand modeling (and micro-economic theory) that individuals with lower incomes tend to be more cost-sensitive than individuals with higher incomes. That means that a cost coefficient estimated specifically for the lowest income quartile will tend to be more negative than cost coefficients that are estimated specifically for other income quartiles. Correspondingly, one would expect the coefficient for a variable defined as “travel cost, for persons with household income below $20,000” to be more transferable than one for a variable that is defined as “travel cost, for all persons”. In the latter case, an estimated coefficient would depend somewhat on the income distribution in the region, while in the former case the income distribution is already controlled for in the variable definition[1]. As another example, one would expect zones with high levels of employment to be, by definition, more likely to be chosen as usual workplace locations. Adding more detail, one expects zones with high numbers of retail and service jobs to be more attractive as work locations for lower income workers, while zones with more office and medical employment tend to be more attractive as work locations for higher income workers.

Following from the discussion above, one might expect that coefficients from an advanced activity-based model will be more transferable than coefficients from a more aggregate trip-based model. Activity-based models tend to include a greater variety of explanatory variables regarding person and household characteristics, whereas standard trip-based models typically only include one or two segmentation variables such as household size, income or car ownership. Furthermore, trip-based models only include a small number of trip purposes, with little or no information about the tour context of non-home-based trips. In contrast, most activity-based models deal explicitly with both the tour and trip context of each trip, as well as incorporating

[1] This a good example of how transferability and data quantity issues interact. Specifying multiple income categories should make the model more transferable, but because each coefficient is estimated from only the portion of sample in its income category, a small sample will tend to make the estimated coefficients less significant. If the estimated values across two regions happen to be similar, their differences may register as insignificant simply because the sample is small. The models may appear to be transferable when in reality the sample is too small to identify significant difference that are actually present.

information on the full day's activity schedule of the individual and other household members. As an example, consider a mode choice model for a work-based tour (a chain of two or more trips that begin at a person's workplace and return back to that same place during the same workday). In activity-based models, one can capture the effect that the mode used for a work-based tour is highly related to the mode that the person uses to commute between home and work on that same day. Such variables cannot be incorporated in an aggregate trip-based model, and as a result, that model will include fewer systematic effects that can be explained by the model.

This project tests transferability only for disaggregate, activity-based models; the hypothesis that they are more transferable across regions than aggregate trip-based models is not addressed. However, at the outset of the project the expectation was that the activity-based models would have a reasonable chance of being transferable across regions, thus making it worthwhile to carry out formal tests, while the expectation was not so high for standard trip-based models. Furthermore, the research question is much more timely for activity-based models, as many large, medium and even small public agencies consider whether or not to adopt these methods and, if so, to what extent an existing activity-based model can be transferred from another region.

The transferability question cannot be answered definitively on the basis of this single study, although it does provide valuable evidence. Also, it is unlikely that there will ever be a single "yes or no" answer in any case. It may be that certain types of variables and coefficients, or entire component models, are good candidates for transferring from one region to another, while other types of variables and/or models may be better re-estimated or re-calibrated based on local data. It may also be possible to say something about the geographic extent of transferability (i.e. that transferability is greater between regions in the same part of the country or between regions of similar size) , although to deal in a comprehensive way with that issue would clearly require additional studies using more data sets from more regions.

Based on the discussion above, four hypotheses are identified and tested in the project. For each hypothesis, one or more specific questions is asked, which the subsequent analysis attempts to answer.

(Hypothesis 1) *Variables that apply to population segments defined by characteristics of individuals and/or their situational context (i.e., segment-specific variables) will tend to be more transferable than variables that are more generic and apply to all individuals*. The reasons behind this hypothesis are discussed at some length above.

> Question H1: Are coefficients defined by individual characteristics or population segments more transferable then those defined for the entire population?

(Hypothesis 2) *Variables that are segment-specific will tend to be more transferable than alternative-specific constants (the 'a' coefficient in equation 1).* This is because alternative-specific constants (ASCs) tend to be "catch-all" coefficients for all effects not captured by the rest of the systematic utility (the *x* and *b* vectors in equation 1), and thus are less predictable from one context to another. This is an extension of the logic behind (H1).

Question H2: Are coefficients defined by population segments more transferable than alternative specific constants?

(Hypothesis 3) *Models that deal with social organization (activity generation and scheduling) will tend to be more transferable than models that deal mainly with spatial organization (mode choice and location choice)*. The spatial attributes that determine mode and location choices are more likely to be influenced by local variations in the types of travel modes available, the historical development of settlement patterns, and possible differences in how the spatial data has been prepared. Models of activity generation and scheduling, on the other hand, deal with behavior that appears similar across regions—people have to sleep, eat, go to work, go to school, shop, do errands, care for children, etc., and they have 24 hours per day to accomplish this.

Question H3.1: Which models are more transferable across states?

Question H3.2: Are models that deal with social organization more transferable across states than those that deal mainly with spatial organization?

Question H3.3: Which models are more transferable within California?

Question H3.4: Within California, are models that deal with social organization more transferable than those that deal mainly with spatial organization?

Question H3.5: Which models are more transferable within Florida?

Question H3.6: Within Florida, are models that deal with social organization more transferable than those that deal mainly with spatial organization?

(Hypothesis 4) *Models for different regions within the same state will tend to be more transferable than models for regions in different parts of the country*: Differences in climate, historical development, economic prosperity, lifestyle, and so forth may give rise to somewhat different activity patterns in different parts of the country, even after controlling for all other segmentation and situational variables. (Note: It would be possible to pose similar hypotheses related to the size and urbanization level of different regions, but a test of such hypotheses may require a larger number of regions.)

Question H4.1: Can a region use models developed from a state or multi-state sample?

Question H4.2: Are California models more transferable within California than they are across California and Florida?

Question H4.3: Are Florida models more transferable within Florida than they are across California and Florida?

Question H4.4: Is a region's model essentially the same as the combined within-state or two-state model?

Finally, this research project does not deal with temporal transferability of models, as all of the survey data is from a single year. It is likely that model would be more transferable to a different year in the same region than to a different region in the same year. This is because there may be "self selection" effects that draw people with different travel preferences to live in different regions. For example, someone who does not like to drive but is not averse to using transit may be more likely to settle in a large urban region, while someone with the opposite preferences may be more likely to live in a more rural region. Unless a region changes drastically, one would not expect those same "self selection" differences to arise within the same region over time.

4. Activity-Based Model Framework

The models estimated in this project are components of the activity-based (AB) model framework commonly known as DaySim, which has been used in Sacramento since 2007 and Seattle since 2009, and has been enhanced and implemented for use in all of the regions of this project except San Diego. The DaySim framework is used for this study as a matter of convenience: it is the framework that has been developed by the principals in this study.

The following description, diagram and table of the existing Sacramento model system provide a good sense of the framework[2]

> Figure [4.1] is a flow diagram showing the relationships among DaySim's component models, which are also listed in Table [4.1]. The models themselves are numbered hierarchically in the table… The hierarchy embodies assumptions about the relationships among simultaneous real world outcomes. In particular, outcomes from models higher in the hierarchy are treated as known in lower level models. It places at a higher level those outcomes that are thought to be higher priority to the decision maker. The model structure also embodies priority assumptions that are hidden in the hierarchy, namely the relative priority of outcomes on a given level of the hierarchy. The most notable of these are the relative priority of tours in a pattern, and the relative priority of stops on a tour. The formal hierarchical structure provides what has been referred to by Vovsha, Bradley and Bowman (2004)[3] as downward vertical integrity.
>
> Just as important as downward integrity is the upward vertical integrity that is achieved by the use of composite accessibility variables to explain upper level outcomes. Done properly, this makes the upper level models sensitive to important attributes that are known only at the lower levels of the model, most notably travel times and costs. It also captures non-uniform cross-elasticities caused by shared unobserved attributes among groups of lower level alternatives sharing the same upper level outcome.

This transferability project includes the model types shown in bold in Table 4.1.These are 14 of the 21 different models in DaySim, including all of the most important models.

[2] From SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution,by Mark Bradley, John L. Bowman and Brice Griesenbeck, 2010, Journal of Choice Modeling. Available at www.jbowman.net.

[3] Vovsha, Peter, Mark A. Bradley and John L. Bowman (2004) Activity-based travel forecasting models in the United States: Progress since 1995 and Prospects for the Future , presentation at the EIRASS Conference on Progress in Activity-Based Analysis, May 28-31, 2004, Vaeshartelt Castle, Maastricht, The Netherlands

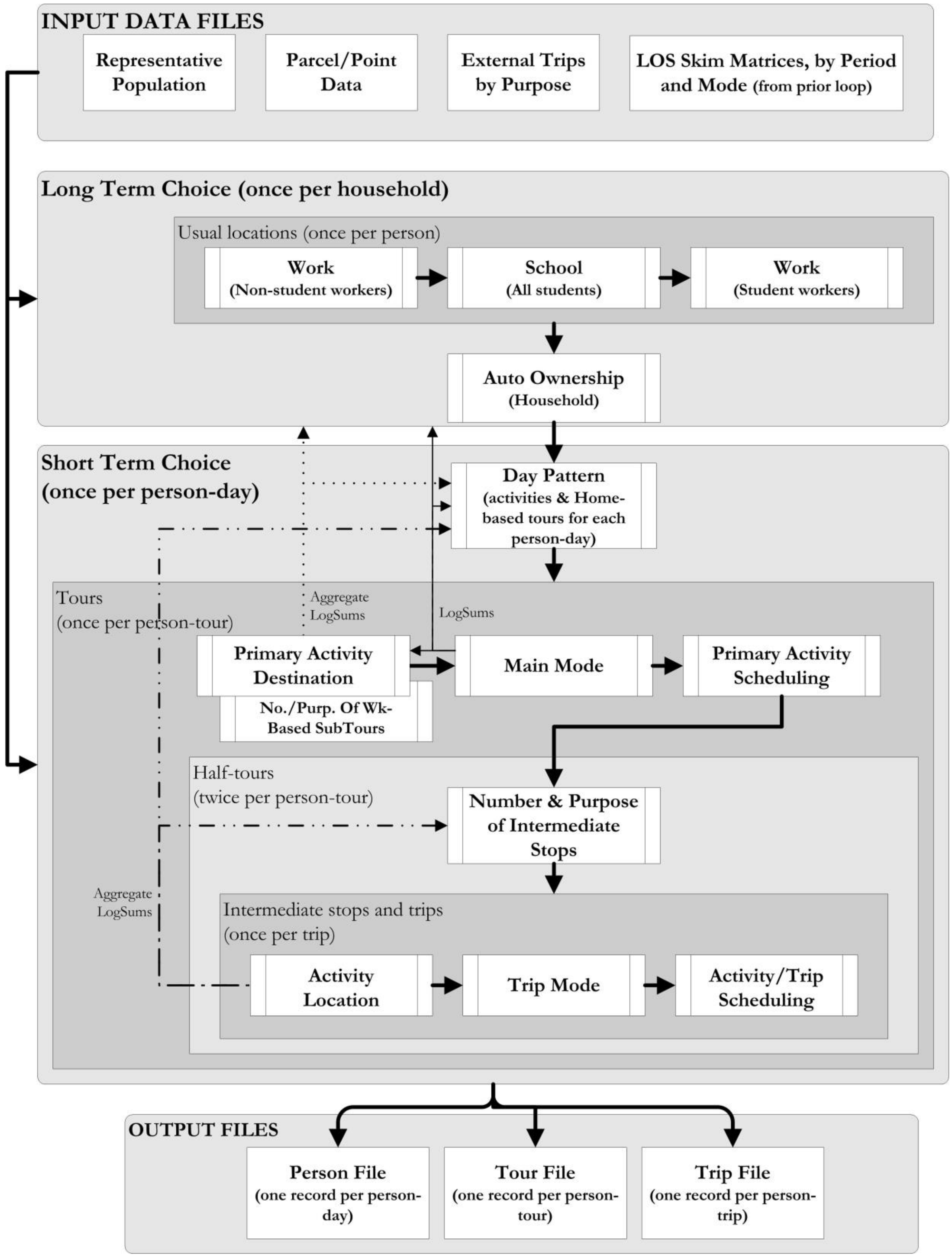


Figure 4.1: DaySim Flow Diagram

Table 4.1. Component Models of DaySim

Model Name	Level	What is predicted	Model Type**
Regular Workplace Location	Worker	Workplace location zone and parcel	Spatial
Regular School Location	Student	School location zone and parcel	Spatial
Auto Ownership	Household	Auto ownership	Social
Daily Activity Pattern	Person-day	0 or 1+ tours for 7 activity purposes. 0 or 1+ stops for 7 activity purposes	Social
Exact Number of Tours	Person-day	For purposes with 1+ tours, 1, 2 or 3 tours.	Social
Work Tour Primary Destination Choice **Other Tour Primary Destination Choice**	(Sub)Tour	Primary destination zone and parcel	Spatial
Work-Based Subtour Generation	Work Tour	Number and purpose of any subtours made during a work tour	Social
Work Tour Main Mode Choice * **School Tour Main Mode Choice** Escort Tour Main Mode Choice Work-based subtour Main Mode Choice **Other Tour Main Mode Choice**	(Sub)Tour	Main tour mode	Spatial
Work Tour Time Period Choice School Tour Time Period Choice Work-based subtour Time Period Choice **Other Tour Time Period Choice**	(Sub)Tour	The time period arriving and the time period leaving primary destination	Social
Intermediate Stop Generation	Half Tour	Number and activity purpose of any intermediate stops made on the half tour, conditional on day pattern	Social
Intermediate Stop Location	Trip	Destination zone and parcel of each intermediate stop, conditional on tour origin, destination, and location of any previous stops	Spatial
Trip Mode Choice	Trip	Trip mode, conditional on main tour mode	Spatial
Trip Departure Time	Trip	Departure time within 30 min. periods, conditional on time windows remaining from previous choices	Social

* Two different versions of this model were tested, as explained below.
**As used in Hypothesis 3 (see Chapter 3)

Table 4.2, split across two pages, provides a more detailed picture of what types of explanatory variables are included in each of the model types that are part of the DaySim system. It is clear from the table that the models include several types of variables that are not typically included in standard 4-step model systems. These include:

- A wide variety of person and household characteristics.
- A variety of parcel level land use and accessibility variables.
- Various zone-level accessibility variables.

- Endogenous variables related to the predicted activity pattern and related schedule pressure
- Endogenous variables related to people who work from home, or who use various modes to get to work

Table 4.2 - Part 1: Variables included in Sacramento DaySim models (P = predicted, X = explanatory)

	Residential location	Usual work location	Usual school location	Auto ownership	Day activity pattern	Work-based tour generation	Tour destination choice	Tour mode choice	Tour time of day choice	Stop frequency and purpose	Intermediate stop location	Trip mode choice	Trip time of day choice
Household characteristics													
Household size	X		X	X	X			X		X		X	
Household number of workers	X			X	X			X					
Household income	X	X	X	X	X		X	X	X	X	X	X	
Household includes children				X	X		X	X		X	X	X	
Household includes people age 65+	X			X	X		X	X	X	X			
Household is non-family household				X	X								
Household number of driving age people				X	X	X	X	X				X	
Household has no cars				P		X	X	X				X	
Household has fewer cars than workers				P				X					
Household has fewer cars than adults				P	X	X	X	X				X	
Housing unit type	X												
Person characteristics													
Full time worker		X		X	X		X	X	X	X			
Part time worker		X		X	X		X	X	X	X			
Non-working adult					X		X	X	X	X		X	
University student		X	X	X	X		X		X				X
Driving age child		X	X	X	X		X	X	X	X	X	X	X
Child age 5-15			X		X		X	X	X	X		X	X
Child age under 5			X	X	X		X	X	X	X			X
Age is 65 or older				X	X		X	X	X	X			
Age is 51-65					X			X					
Age is 26-35					X							X	
Age is 18-25					X							X	
Gender		X			X			X		X	X	X	
Usual workplace is home		P			X								
Parcel-level land use variables													
Service employment (density)		X	X	X		X	X				X		
Educational employment (density)		X	X				X				X		
Government employment (density)		X	X				X				X		
Office employment (density)		X	X				X				X		
Retail employment (density)		X		X		X	X				X		
Restaurant employment (density)		X		X		X	X				X		
Medical employment (density)		X		X		X	X				X		
Industrial employment (density)		X					X				X		
Total employment density		X					X				X		
Household density		X	X				X				X		
University student enrollment (density)		X	X				X				X		
K-12 student enrollment (density)		X	X			X	X				X		
Mixed use balance		X			X		X	X			X	X	

Table 4.2 Part 2: Variables included in Sacramento DaySim models (P = predicted, X = explanatory)

	Residential location	Usual work location	Usual school location	Auto ownership	Day activity pattern	Work-based tour generation	Tour destination choice	Tour mode choice	Tour time of day choice	Stop frequency and purpose	Intermediate stop location	Trip mode choice	Trip time of day choice
Parcel-level accessibility variables													
Parking density		X					X				X		
Average parking price				X				X			X	X	
Street intersection density		X			X		X	X		X	X	X	
Distance to nearest transit stop				X	X			X			X	X	
Zone-level accessibiliy variables													
Auto and transit costs								X	X		X	X	X
Auto, transit and non-motorized times								X	X		X	X	X
Transit connectivity/availability								X	X		X	X	X
Auto time on very congested links									X				X
Driving distance		X	X				X	X			X		
Mode choice accessibility logsum		X	X	X	X		X						
Mode/destination accessibility logsums		X			X		X						
Intermediate stop accessibility logsums					X					X			
Endogenous activity pattern variables													
Number of home-based tours in pattern					P	X		X	X	X			X
Pattern has multiple tours for the purpose					P	X	X		X				
Pattern has stop(s) for the purpose					P		X		X				
Pattern includes work or school tour					P		X			X			
Purpose of tour					P		X	X	X	X	X	X	X
Tour is work-based subtour						P	X		X	X	X	X	X
Intermediate stop purpose								X		P	X	X	X
Number of intermediate stops on half tour										P	X	X	X
Outbound or return tour direction										X	X	X	X
Endogenous location, mode, TOD variables													
Work tour is not to usual workplace						X	P						
Tour mode is auto, transit, etc.								P		X	X	X	
Mode used to get to work								P				X	
Tour time periods of the day									P	X	X	X	X
Unscheduled time remaining in the day							X		P	X	X		X
Trip mode is auto, transit, etc.												P	X
Trip time period of the day													P

For purposes of summarizing results of the transferability tests, each model variable is classified into one of the following types (two types if it is an interaction variable):

1. A-constant
2. P-person
3. H-household
4. D-day pattern
5. T-tour/trip
6. I-impedance
7. U-land use
8. W-time window
9. C-logsum
10. G-size variable
11. L-log size multiplier

Table 4.3 shows a cross tabulation of the model coefficients by model type and variable type, accounting for interaction variables that are two types.

Table 4.3. Model Coefficients by Model Type and Variable Type

	Variable Type											
Model Type (mtype)	A-constant	P-person	H-household	D-day pattern	T-tour/trip	I-impedance	U-land use	W-time window	C-logsum	G-size variable	L-log size mult	Total
Usual work location	14	17	6	0	0	9	18	0	6	14	1	48
Auto ownership	4	0	14	0	0	4	4	0	2	0	0	24
Person-day tour generation	28	76	46	0	0	0	1	0	2	0	0	126
Exact number of tours	13	34	22	23	0	0	0	0	8	0	0	86
Work tour time of day	27	9	6	13	9	4	0	10	0	0	0	69
Work tour mode (detailed LOS)	5	9	12	0	0	14	20	0	0	0	0	58
Work tour mode (combined LOS)	5	3	9	4	0	3	7	0	0	0	0	31
WB subtour generation	8	0	2	2	0	0	1	0	1	0	0	14
School tour mode	5	7	12	0	0	3	6	0	0	0	0	32
Other tour destination	12	2	2	0	44	29	14	1	5	12	1	62
Other HB tour time of day	25	12	0	17	29	4	0	10	0	0	0	86
Other HB tour mode	5	3	11	4	7	3	8	0	0	0	0	41
Intermediate stop generation	7	5	4	11	55	0	1	20	0	0	0	100
Intermediate stop location	9	2	3	0	45	35	19	1	0	9	1	66
Trip time of day	25	5	0	2	10	2	0	3	0	0	0	45
Total	192	184	149	76	199	110	99	45	24	35	3	
As Primary Variable	157	159	112	74	113	67	99	45	24	35	3	888
As Interaction Variable	35	25	37	2	86	43	0	0	0	0	0	

In general, the 14 selected models are transferred "as is" from the latest version of DaySim that was recently updated for SACOG. The SACOG version is used as the basis because it is the most rigorously developed specification within the DaySim family and serves as the basis of the other existing DaySim implementations. Also, the current transferability study lacks budget for the development of a new specification to serve as the basis of comparison. A few minor changes are made, as follows, in most cases these are simplifications to make the models somewhat easier to estimate and test:

- The auto ownership model specification is simplified somewhat
- The size variable functions in the location choice models are generally simplified a bit to improve estimability of the models on the smaller data sets.
- The DailyActivityPattern model is simplified so that it still predicts 0 or 1+ tours for the 7 different purposes, but does not predict 0 or 1+ stops. (This reduces the number of alternatives in the model from 2,080 to 64, and reduces the number of coefficients considerably, making the model much more practical for making many testing runs.
- The IntermediateStopGeneration model alternative availability is no longer constrained by the results of the DailyActivityPattern, to be consistent with that adjusted model which no longer predicted if stops were made or not.
- The drive-to-transit (park and ride) alternative was eliminated from the Work Tour Mode Choice model, because there were few, if any, observed park and ride tours for most regions, and because all regions did not have park and ride lot files ready in time for the project.
- None of the base model specifications included nesting coefficients for nested logit significantly different from 1.0, even the mode choice models. So, to make the test run process more tractable, those models were estimated as MNL rather than NL.
- The logsums calculated from mode and destination choice models and used as explanatory variables in other models are calculated from the original SACOG version of the mode and destination choice models, rather than from the new version being estimated.

One model is tested in a more detailed version. The mode choice models in DaySim use generalized cost functions with pre-specified weights on the various time and cost coefficients, based on the recent SHRP 2 C04 study. A more detailed version of the Work Tour Main Mode Choice model is also tested that allows separate coefficients to be estimated for each of the travel time and cost coefficients. This extra test brings the total number of models tested to 15.

5. The Regions and Their Data

Table 5.1 shows basic statistics of the six regions in the study. This table indicates that there is a fair amount of variation in the number of NHTS households available in each region. At one extreme, San Diego has a fairly robust sample of 6,000 households, while at the other end the Fresno dataset includes only 380 households.

Broadly, the six regions can be grouped into two primary classes by household and employment size. San Diego, Tampa and Sacramento are larger regions, each with approximately 1 million households, and between 1 million and 1.5 million jobs. Jacksonville, Northern San Joaquin County (NSJV) and Fresno are smaller, typically with 0.5 million households and employment. Household density varies widely, ranging from 48 households per square mile in Fresno to over 400 in Tampa. In Fresno and Northern San Joaquin Valley, agriculture and industry are major sources of employment, whereas in other regions, government, office and service employment are more important. In concept, the AB models can control for these differences, although in practice, large differences might affect transferability.

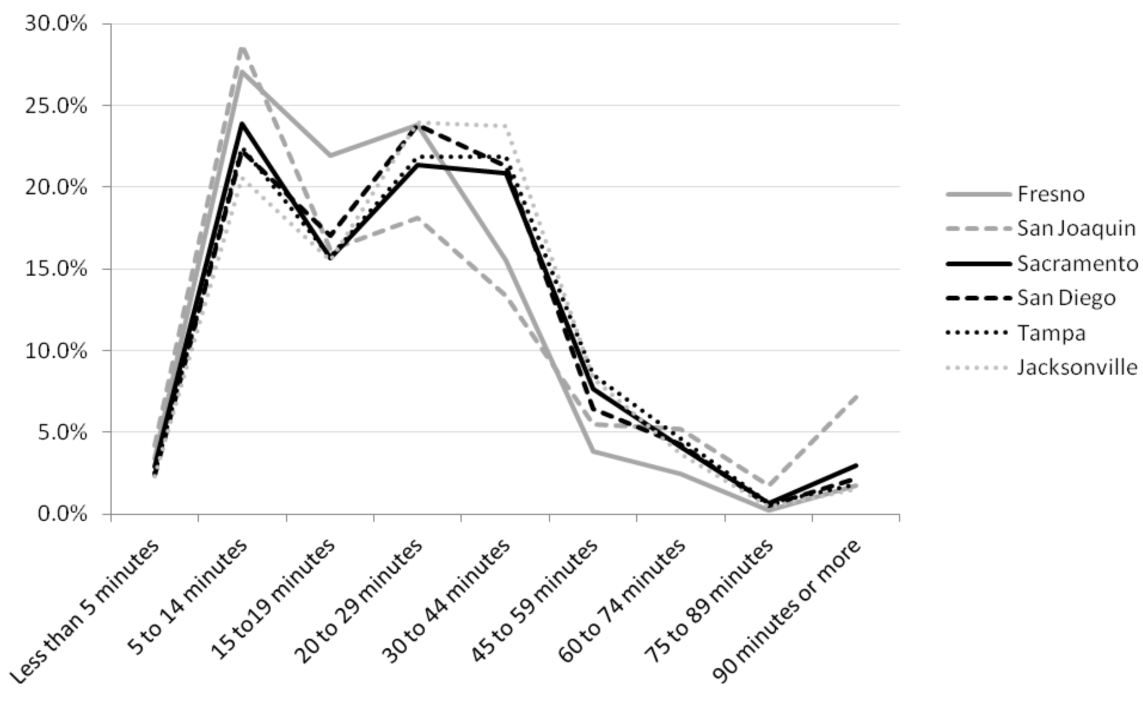
In terms of their total geographic extent, the six regions are more comparably sized, although the Jacksonville and Tampa regions are somewhat smaller than the California regions. Despite this relative comparability of size, the degree of spatial resolution in the original regional models varies by region, with San Diego incorporating the most detailed zone system with 4,700 TAZs covering 4,300 square miles, or less than one square mile per TAZ to Fresno, which 2,000 TAZs covering 6,000 square miles. A potential cause of non-transferability of models estimated and applied at the TAZ level is variation across regions in the size of the TAZ.

In order to control for this variation, the project team developed all data and performed all model estimation using the "microzone" geographic resolution. Microzones are essentially Census Blocks, with minor modifications to ensure that all microzones "nest" within the existing regional TAZ structure. This nesting is important because the network impedances used in model estimation are at the TAZ level due to issues associated with building network skims at the microzone level. Throughout this report, the terms "Census Block" and "microzone" are used interchangeably because in most cases they are the same. Preparation of model estimation datafiles at the Census Block level was facilitated by the fact that the original DaySim implementations for Sacramento, Fresno, Tampa and Jacksonville all used parcel-level household, employment, and enrollment data that could be easily aggregated to Census Blocks. However, availability of parcel-level information is not required in order to develop Census Block-based models. The project team developed a tool that disaggregates TAZ level information to Census Blocks using Census and Longitudinal Employer-Household Dynamics (LEHD) data. This tool was used to develop the datafiles for the Northern San Joaquin Valley region, In San Diego, the household, employment, and enrollment data was provided at the level of Master Geographic Reference Areas (MGRAs), which have a resolution comparable to Census blocks. The Census Block data prepared for each of the regions includes essential information on housing, employment by sector, enrollment by school type, as well as detailed information on urban form, transit access, and proximity buffers.

To the greatest extent possible, this effort seeks to build upon existing model development efforts and practices, and therefore must rely upon the existing TAZ structure for network skim methods and resolution. In addition to geographic resolution, skims are also distinguished by the amount of temporal detail or time periods they represent, as well as the amount of transportation sub-modal detail they represent. Most of the regions use at least 4 time periods for representing the changes in auto impedances by time of day, but the majority only use two time periods (peak and off peak) to represent changes in transit services. Most of the regions employ between 5 and 10 submodes in mode choice and network assignment, although San Diego incorporates significantly more submodal detail. In order to have a set of skims by mode and time period that was defined consistently across all the regions, the project team developed a simplified set of skims for each region. These included skims of walk and bike, single occupancy and high occupancy vehicles, and a generalized transit mode.

Observed average commute travel times derived from the 2006-2008 American Community Survey are reported in Table 5.1, while Figure 5.1 illustrates the observed distributions of these travel times for the six regions. Interestingly, Table 5.1 shows that the Northern San Joaquin counties have the highest average travel times, although Figure 5.1 illustrates that it also has the highest share of very short commutes and the highest share of very long commutes. These long commutes are likely workers travelling the long distance to the San Francisco Bay Area for employment. It is also striking that the average commute travel times and the distribution of commute travel times is remarkably similar across the four largest regions (San Diego, Tampa, Sacrament, and Jacksonville). The AB model accounts for residents who work outside the region, so the otherwise similar commute travel time distribution might enhance transferability.

Figure 5.1: Commute Travel Time Distribution by Region



Source: 2006-2008 American Community Survey

Table 5.1: Basic statistics of the six regions in the transferability study

	Fresno	NSJV	Sacramento	San Diego	Jacksonville	Tampa
NHTS Households	380	660	1,310	6,000	1,050	2,500
REGION Households	288,857	458,731	805,292	1,081,082	551,353	1,361,724
Sq Miles	6,017	4,913	6,197	4,262	2,532	3,275
Number of TAZ	1,967	3,758	1,502	4,682	1,309	2,251
Number of Census Blocks	27,891	39,312	49,282	33,084	30,899	54,310
Avg Census Block Size (acres)	137.9	78.0	80.3	82.4	56.0	36.3
Median Census Block Size (acres)	5.6	4.5	5.4	6.7	4.8	5.1
HH / Sq Mi	48	93	130	254	218	416
Avg Households/Census Block	10.4	11.7	16.6	34.2	17.8	26.4
Avg Employment/Census Block	12.7	11.7	19.7	45.9	20.7	26.9
Employment (Total)	353,216	459,910	969,838	1,519,582	638,195	1,462,137
Education	35,768	52,028	73,688	149,540	35,787	106,766
Food	22,862	10,836	58,102	90,013	46,619	129,529
Government	36,837	16,290	67,103	205,780	72,801	70,147
Industrial	68,443	140,303	119,305	202,605	111,102	206,713
Medical	36,304	48,081	108,036	101,251	70,395	198,171
Office	44,521	54,789	204,288	170,027	112,511	358,637
Retail	35,946	55,881	131,781	151,504	87,031	192,973
Service	5,393	24,609	202,884	424,454	100,445	98,815
Agricultual/Resource/ Construction	67,142	57,093	4,652	24,408	1,504	100,386
Avg commute time (2006-2008 ACS)	23.9	30.0	28.5	27.9	28.6	28.4
Modeled modes (in region's original model)	5	5	10	25	8	11
No. auto skim periods	4	4	5	6	4	2
No. transit skim periods	2	2	5	6	2	2
Network software	Cube	Cube	Cube	TransCAD	Cube	Cube

6. Model Estimation and Transferability Testing Approach

Base Estimation Approach

The models described in Chapter 4 are estimated using a common, consistent data set spanning all of the regions described in Chapter 5. In particular:

- The observed choices are taken from a household travel survey, which has been processed in such a way as to reflect the assumptions and conditional relationships among choice components that are depicted in Figure 4.1. There are records representing each household, person, household-day, person-day, tour and trip.

- For each observed choice, relevant information is drawn from a file of ‘microzone’ attributes and files of zone-to-zone impedance information (road and transit skims). As described previously, the microzones are equivalent to Census blocks, except for cases where the region’s zone (TAZ) system are inconsistent with Census block boundaries, in which case some blocks are sub-divided along the TAZ boundaries.

- Estimation of upper level models use composite variables (logsums) calculated from the models of lower level choices.

There are two important differences between the common estimation data used for this study and the data sets used to estimate models for the various agencies:

- The observed choices all come from the 2008-2009 National Household Travel Survey (NHTS). This provides a uniformity that allows the data sets to be combined for estimation and improves the uniformity of definitions across regions. In estimation, it is still possible to identify which region each observation is from, enabling the estimation of region-specific and state-specific coefficients, the primary method for identifying regional and state-level differences in the model estimation results (as described in the following section).

- Since each region has its own geography, network, and network model software (used to generate the skim matrices), the spatial land use attributes and network impedance information for the regional subsets are drawn from those separate data sets before combining into a single database for model estimation. As discussed earlier, this could potentially introduce differences across regions that could cause differences in model estimation results across regions that would be impossible to distinguish from behavioral differences across the populations. This is problematic because if the models prove to be non-transferable, it is not possible to tell conclusively whether it is caused by differences in the way people behave or differences in the impedance and zonal data. While this problem is impossible to avoid completely, it is mitigated by two factors. First, the consistent use of Census blocks for microzones, and the use of consistent buffering methods to generate "neighborhood" and density variables means that the land use data is not much affected by differences in the regions' zone systems. Second, only a subset of variables and models depend on the travel time and distance measures from the network skims, and the mode choice models use pre-specified weights on those variables (taken from the SHRP 2 C04 study) rather than estimating new coefficient on those variables. Thus, any inconsistencies in the network assignment and skimming methods should not greatly affect the transferability results. (A more detailed version of the Work Tour Mode Choice model was also tested, specifically to look at the transferability of travel time and cost coefficients.)

Transferability Testing Approach

The process

The transferability tests were set up and executed using the following sequence of steps:

(1) For each choice model to be tested, a base model specification was developed, including all explanatory variables to be tested. As described earlier, in most cases this specification was kept the same as the recently revised DaySim-SACOG specification, except for a few specific instances where the model structure or specification was simplified somewhat to improve estimability.

(2) As described above, each variable in each model was given a one- or two-letter code denoting what type of variable it is. This coding aided in automatically classifying variables during later steps in the process.

(3) For each of the 15 models, the base model estimation data set was created for each of the 6 regions, giving 90 separate estimation data sets. This step was done using the DaySim software framework, which uses identical model code for both model estimation and application. Because the models (excluding any modifications) had already been coded into DaySim and tested extensively in application for other projects, this process was important for both (a) avoiding errors in setting up the estimation data, and (b) ensuring consistency in the data processing across the six regions.

(4) DaySim automatically creates consistent data and control files for use by the ALOGIT model estimation software package. For each of the 90 base models (15 model types times 6 regions), the base model was estimated using ALOGIT. In some cases, the models were not estimable at this stage. The most common reasons for that were:

   a. There was no variation in an independent variable for a specific choice alternative, so the coefficient would go to positive or negative infinity. This was most common in cases where there were very few observed choices for an alternative (e.g. the bike and transit modes in the mode choice models, or extremely late or early periods in the time of day models), or in cases where there were very few observations with a modeled characteristic (e.g. 0-car households or very low income households in the smallest (Fresno) data set).

b. There were other estimation issues due to very high co-linearity in the variables. This was most common for the land use variables, and tended to have most affect in the size variable functions for the location choice models.

Any variables that could not be estimated in the base version were dealt with by leaving the variable in the model but constraining the coefficient to a specific value (usually 0).

(5) A program was created to read in the ALOGIT estimation results (extension .f12) files for the 90 different base runs, and to automatically create new data files and control files to estimate 36 different model specifications for each of the 15 models. For a specified one of the 15 model types, the program performs the following main tasks:

a. Merges the separate estimation data files for the 6 regions into a single combined estimation data file, ensuring that the data record structure is exactly the same for every observation.

b. Creates ALOGIT estimation control files for all 36 models used in the transferability tests. These are listed under variable 'Mspec' in the Appendix 2, and described in more detail in the following paragraphs.

c. Creates an ALOGIT "batch run" file instructing the ALOGIT software to estimate all 36 models in succession.

This automated procedure proved very efficient in performing the large number of needed estimation runs, and has the further advantage that it can be re-used if it is ever decided to add further data from additional regions into the study.

(6) After the 36 model runs for all 15 model types were completed successfully, a second program was written to compile and tabulate all estimation results for further viewing and analysis. This program reads in the estimation results (.F12) files, and writes out two types of files:

a. Comma-delimited (.csv) files showing all estimation results in tabular form, including model summary statistics, plus the labels, coefficients, and t-statistics for all coefficients. These .csv files were then imported into Excel, where they were pasted into a standard tabular format for viewing, and combined into a single workbook. That workbook is included separately with this report.

b. A single, space-delimited metafile, containing the results for every model and coefficient estimated in the project in a single file. The content of this file is documented in Appendix 2.

(7) The metadata file from the previous step was analyzed in SPSS to create the tables and charts presented in the following chapter.

The model specifications for transferability tests

As noted above, up to 36 different models were estimated for each model type. These fall into two main types: the "base models", and the "difference models".

There were twelve base models estimated. Six of them are the region-specific based models described in Step 4 above. These were also re-estimated as part of the sequence of 36 runs, but were not used in the formal transferability tests. They were used mainly for the analysis of parameter estimability reported in the next chapter, particularly as it relates to sample size.

Six additional base models were estimated as well, falling into four different types:

- **2-state**: Using the data from all 6 regions in both states
- **1-state**: (a) Using only the data from the 4 California regions, and (b) using only the data from the 2 Florida regions
- **2-state+ASC**: Using the data from all 6 regions in both states, but using separate alternative-specific constants for each region.
- **1-state+ASC**: (a) Using only the data from the 4 California regions, and (b) using only the data from the 2 Florida regions, but using separate alternative-specific constants for each region in that state.

These six models were used as a basis for comparison for the remaining 24 "difference models". The concept behind the difference models is as follows:

- Estimate a model across data from multiple regions (6 regions for the 2-state models, and either 4 or 2 regions for the 1-state models).
- For a single selected region, add a second set of all the coefficients that are in the base model. These are referred to as "difference variables", because if the coefficient value that is estimated across all included regions is the same as the coefficient value that represents the single selected region, then the estimated value for the second, region-specific coefficient will be 0. The second coefficient essentially measures the difference between the correct coefficient value for the selected region and the correct value for the sum of all regions **<u>except</u>** the selected region.
- Estimating a full model with difference variables gives transferability evidence in two ways:
    - The significance of the difference coefficient on each variable provides evidence of the transferability of that particular variable between the selected region and the other regions included in the model.

- A chi-squared model fit test between the full difference model and the corresponding base model provides evidence on the transferability of the model as a whole. (Note that this is quite a strict test to use for data from different regions. It is more commonly used for data from a single region, to test the significance of adding or subtracting variables from a model.)

For a simple example of a difference model, imagine a logit model that only has two alternatives and two independent variables. The base model has the utility functions:

$$V(1) = 0$$

$$V(2) = a + b1 . x1 + b2.x2$$

The difference model would then have the utility functions:

$$V(1) = 0$$

$$V(2) = a + b1.x1 + b2.x2 \quad + R.(a' + b1'.x1 + b2'.x2) .$$

where R is a dummy variable that is 1 for the selected “difference region” and 0 for all other regions

The second half of the V(2) equation is a copy of the first half, but the coefficients a’, b1’ and b2’ are only estimated on the observations for the selected region where R=1. So, for that selected region, the resulting coefficients for x1 and x2 are (b1+b1’) and (b2+b2’) and the resulting estimate for the alternative specific constant is (a+a’). The primed coefficients thus measure the differences between the original coefficients and the resulting, combined coefficients for the selected region.

Consider the California data with 4 regions. The base model for the “1-state+ASC” case would be written as:

$$V(1) = 0$$

$$V(2) = a + b1 . x1 + b2.x2 + (a'^2.R^2 + a'^3.R^3 + a'^4.R^4)$$

This specification adds a difference variable on the alternative-specific constant for each of the 4 regions except one, which effectively allows each of the four regions to have its own constant.

The difference model for region 4, relative to this base model is then:

$$V(1) = 0$$

$$V(2) = a + b1.x1 + b2.x2 + (a'^2.R^2 + a'^3.R^3 + a'^4.R^4) + R^4.(b1'.x1 + b2'.x2) .$$

Because there was already a difference coefficient on the constant a for region 4, only difference variables for x1 and x2 need to be added. Also, because the region-specific constants are allowed, the base estimates of b1 and b2 could change as well, which could in turn affect the values of the difference coefficients b1' and b2'.

To allow for the testing of these different types of effects, 24 difference models were estimated:

- 1-6: Difference variables for each of the 6 regions relative to the 2-state base
- 7-12: Difference variables for each of the 6 regions relative to the 1-state base
- 13-18: Difference variables for each of the 6 regions relative to the 2-state+ASC base
- 19-24: Difference variables for each of the 6 regions relative to the 1-state+ASC base

Each of the types of specifications, model types, and variable types can provide different types of evidence related to our research hypotheses presented in Chapter 3. The following Chapter presents a careful analysis of the resulting evidence.

7. Results

This Chapter reports the results of the model estimation procedures described in Chapter 6, addressing the questions developed in Chapter 3, first related to data and estimability, and then related to transferability.

Comparisons of Coefficient Significance and Estimability

The questions of coefficient estimability and significance are addressed by examining the estimability and significance of the coefficients of each separately estimated regional model, as well as the one-state and two-state models. Also, each estimable coefficient is identified as having the same or different sign as the corresponding coefficient estimated from the entire two-state sample. Thus, each model coefficient falls into one of five categories:

Significant, with the same sign as in the two-state model
Insignificant, with the same sign as in the two-state model
Insignificant, with different sign than in the two-state model
Significant, with different sign than in the two-state model
Not estimable (but its counterpart in the two-state model is estimable)

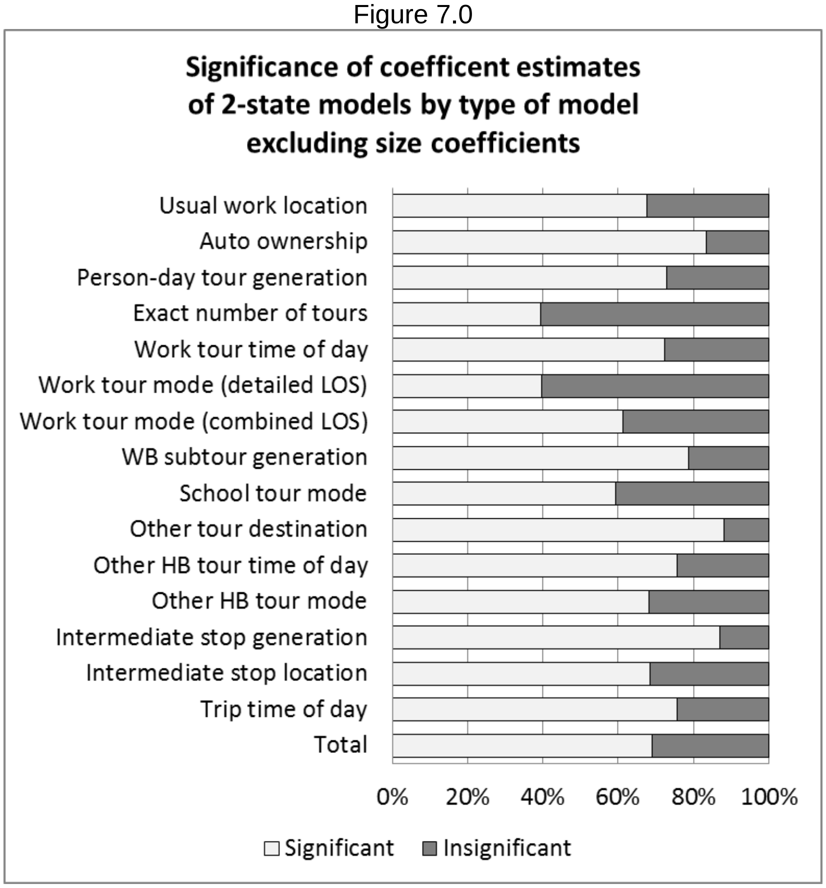
From the standpoint of estimability, the best category is on top, and it gets worse moving down the list, assuming that the two-state sample is the most likely to have the correct sign because of its large size. Having the same sign is better than having a different sign. If the sign is the same, it is better for the estimate to be significant, and if the sign is different, it is better for the estimate to be insignificant. The worst case is if a coefficient that is estimable in the two-state sample is not estimable in the smaller sample.

Table 7.0 and Figure 7.0 show the significance of the coefficient estimates for all fifteen model types in the two-state base models. The size variables of the location choice models are excluded from Table 7.0 and Figure 7.0 because their nonlinear form in the model make their significance measures incomparable to the other coefficients.

Table 7.0 Coefficients by model type and significance in the base two-state models (excluding size coefficients)

Model Type	Significant	Insignificant	Total Number
Usual work location	68%	32%	34
Auto ownership	83%	17%	24
Person-day tour generation	73%	27%	126
Exact number of tours	40%	60%	86
Work tour time of day	72%	28%	69
Work tour mode (detailed LOS)	40%	60%	58
Work tour mode (combined LOS)	61%	39%	31
WB subtour generation	79%	21%	14
School tour mode	59%	41%	32
Other tour destination	88%	12%	50
Other HB tour time of day	76%	24%	86
Other HB tour mode	68%	32%	41
Intermediate stop generation	87%	13%	100
Intermediate stop location	68%	32%	57
Trip time of day	76%	24%	45
Total	69%	31%	853

Figure 7.0



For the subsequent estimability results, the coefficients are evaluated in comparison to this two-state base. Also, data from only fourteen of the fifteen model types are included; the work tour mode choice model with detailed impedance variables is excluded because in the two-state model, which serves as the basis for evaluation of the smaller models' signs, it includes several coefficients with the wrong sign.

Note that the results for the detailed work mode choice model support the idea that very few RP data sets will support the accurate estimation of detailed travel time and cost coefficients. Since there have been numerous careful SP and RP studies related to this, including the SHRP 2 C04 synthesis study, it appears that it will nearly always be better to transfer travel time and cost coefficients (and related co-variant effects) from such prior research rather than attempting to estimate completely new coefficients that are subject to a great deal of error and uncertainty.

Tables 1-a through 1-n in Appendix 3 show a detailed summary of the estimation results, with a table for each of the fourteen relevant model types. Each table shows the percentage of coefficients that fall in each of the five categories listed above, for each of the nine "base" models—the two-state base mode, each of the two states separately, and the six regions separately. Appendix 4 shows summaries of estimation results by coefficient type.

In the remainder of this Chapter, the results are summarized in different ways to address the different research questions. The results described in the next few sections are derived from the same information used to create the more detailed tables in Appendix 3, but aggregated along different categories to answer particular research hypotheses.

Significance and estimability by number of survey households

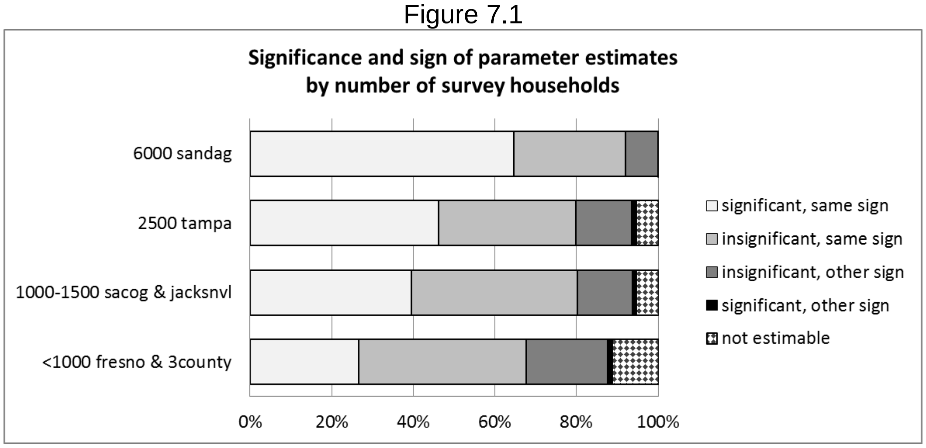
Question E1: What sample size is adequate for local estimation?

For this question, Table 7.1 and Figure 7.1 summarize the results by combining all model types for each region, and combining regions into four survey size categories. As can be seen in the table and figure, the ability to estimate statistically significant coefficients increases substantially as sample size increases from under a thousand to as many as 6000. Although these results don't definitively say what sample size is adequate, the large improvement between Tampa results with 2500 households and the SANDAG results with 6000 suggests that samples much smaller than 6000 should be discouraged. Furthermore, although it is impossible from these results to know how much improvement would come from regional samples greater than 6000, the large difference between Tampa and SANDAG also gives a hint that increasing above 6000 could yield even more improvement.

Table 7.1

	<1000 fresno & 3county	1000-1500 sacog & jacksnvl	2500 tampa	6000 sandag
Significance and sign				
significant, same sign	27%	39%	46%	65%
insignificant, same sign	41%	41%	34%	27%
insignificant, other sign	20%	14%	14%	7%
significant, other sign	1%	1%	1%	0%
not estimable	11%	5%	5%	0%

Figure 7.1



Significance and estimability by geography

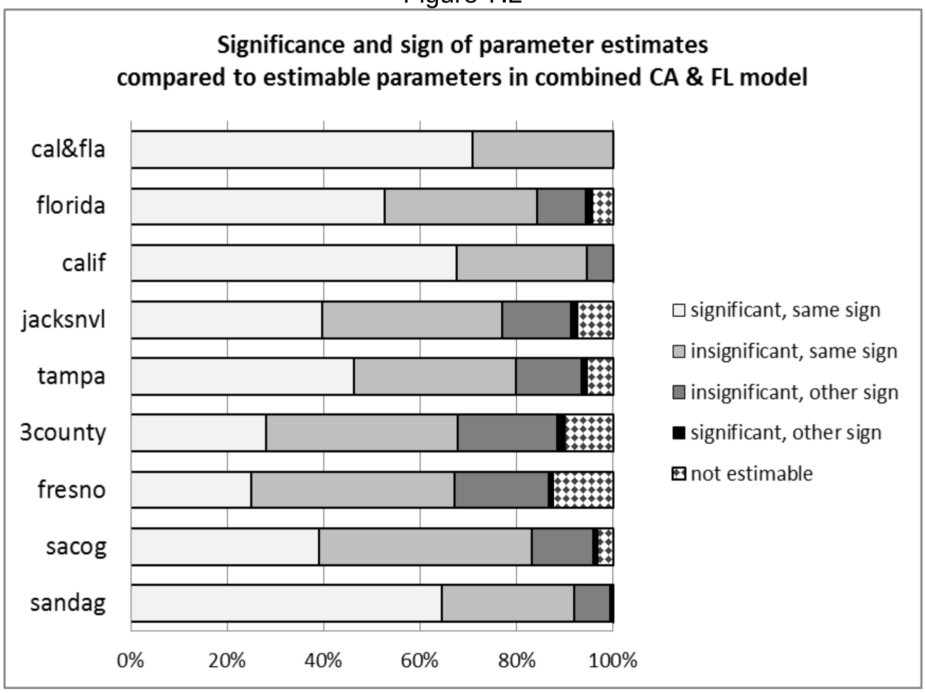
Question E2 : How does combining data samples improve estimability?

For this question, Table 7.2 and Figure 7.2 summarize by again combining results for all model types, but keeping results for each region separate and also showing results for the one-state and two-state models. Table 7.2 also shows the number of survey households for each model. In most cases, combining samples across regions within California and Florida substantially improves the ability to estimate coefficients, to achieve the same sign as the two-state sample, and to get statistically significant coefficient estimates. However, the comparison of results for SANDAG and California shows little improvement when adding data from other California regions to the SANDAG sample, suggesting that if a region has a sample as large as SANDAG's, adding data from other regions may not be beneficial. This stands in contrast to Fresno, with the smallest survey sample, where estimability, sign and significance are all substantially worse than in the California and two-state models.

Table 7.2

	sandag	sacog	fresno	3county	tampa	jacksnvl	calif	florida	cal&fla
Number of survey households	6002	1311	381	657	2517	1335	8351	3852	12203
Significance and sign									
significant, same sign	65%	39%	25%	28%	46%	40%	68%	53%	71%
insignificant, same sign	27%	44%	42%	40%	34%	37%	27%	31%	29%
insignificant, other sign	7%	13%	20%	20%	14%	14%	5%	10%	0%
significant, other sign	0%	1%	1%	2%	1%	1%	0%	1%	0%
not estimable	0%	3%	13%	10%	5%	7%	0%	4%	0%

Figure 7.2



Significance and estimability by type of model component

Question E3: Which models are more estimable at the regional level?

For this question the results are summarized by combining results across states, but looking at three different categorizations of model type. Table 7.3 and Figure 7.3 look at the model's level in the model system hierarchy. Table 7.4 and Figure 7.4 look at models by the type of choice represented by the model. Here, the models classified as 'social organization' include auto ownership and those that define the day in terms of the number and purposes of tours and intermediate stops. Table 7.5 and Figure 7.5 also look at type of choice but in more broadly defined categories of timing and social organization vs spatial organization, as used to address transferability hypothesis 3.

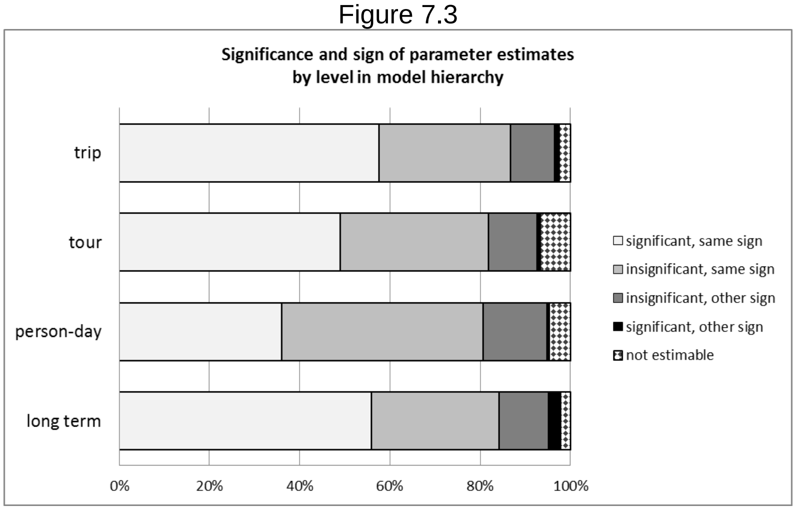
From the perspective of model hierarchy, the trip and long term models are more estimable at the regional level, followed by the tour models and lastly the person-day models. From the perspective of model type, the mode choice models are least estimable. Using the categorization of hypothesis three, which contrasts models of social organization and timing from those of spatial organization (mode and destination), the models of spatial organization are less estimable. (Note however, in Table 7.0, that the average for the person-day models is brought down quite significantly by the value for the Exact Number of Tours model, but that the Person Day Tour Pattern model, which is far more important model in determining forecasts, has a much higher percentage of significant variables (73% vs. 40%). The Exact Number of Tours model only explains cases where people make multiple tours for the same purpose during a single day, which is relatively rare, so such a model has less information to inform the coefficient values.)

By level in model hierarchy

Table 7.3

	long term	person-day	tour	trip
Number of estimated parameters	513	1908	2898	1809
Significance and sign				
significant, same sign	56%	36%	49%	58%
insignificant, same sign	28%	45%	33%	29%
insignificant, other sign	11%	14%	11%	10%
significant, other sign	3%	1%	1%	1%
not estimable	2%	5%	7%	2%

Figure 7.3

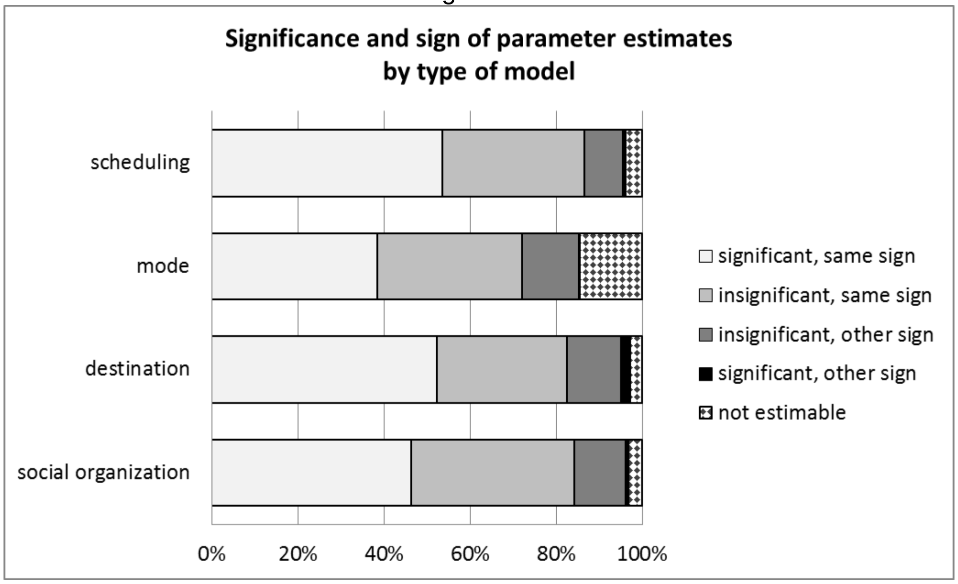


By type of model

Table 7.4

	social organization	destination	mode	scheduling
Number of estimated parameters	3150	1242	936	1800
Significance and sign				
significant, same sign	46%	52%	39%	54%
insignificant, same sign	38%	30%	33%	33%
insignificant, other sign	12%	13%	13%	9%
significant, other sign	1%	2%	0%	1%
not estimable	3%	3%	14%	4%

Figure 7.4

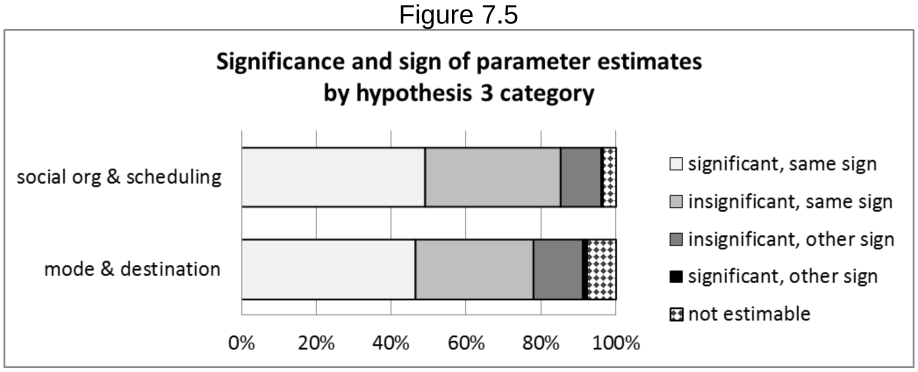


By hypothesis 3 category

Table 7.5

	mode & destination	social org & scheduling
Number of estimated parameters	1011	2429
Significance and sign		
significant, same sign	46%	49%
insignificant, same sign	32%	36%
insignificant, other sign	13%	11%
significant, other sign	1%	1%
not estimable	8%	3%

Figure 7.5



Comparison of Estimated Coefficients

This section compares regional models in several different ways using the four following difference model specifications:

1. Two-state base with difference coefficients for the named region
2. One-state base with difference coefficients for the named region
3. Two-state base with ASCs for all regions and difference coefficients for the named region
4. One-state base with ASCs for all regions and difference coefficients for the named region

Some comparisons use only one base for comparison, while others use all four.

The comparisons include estimation result data from all fifteen model types, including the work tour mode choice model with detailed impedance variables, which was excluded from the above analysis of significance and estimability.

Most of the comparisons summarize the results by identifying the percentage of coefficients in each of three categories:

1. Regional model coefficient insignificantly different from the larger model's coefficient
2. Regional model coefficient significantly different from the larger model's coefficient
3. Model coefficient can be estimated in the larger model but not in the regional model

In some cases it is useful to look at all three categories. In other cases it is more useful to focus only on coefficients that can be estimated at both the regional and larger geography, comparing the percentages of insignificant and significant differences; this is one way to remove confounding effects caused by differences in degree of estimability across regions.

Tables 2-a through 2-n in Appendix 3 show a detailed summary of the difference model results, with a table for each of the fourteen relevant model types. Each table shows the percentage of coefficients that fall in each of the three categories listed above, for each of the 24 base model type/region combinations (the four base model types above times six regions). Appendix 4 provides a summary of difference model results by detailed variable type.

The results described in the next few sections are derived from the same information used to create the tables in Appendix 3, but aggregated along different categories to answer particular research hypotheses.

Differences between regional and two-state models by type of coefficient

This section compares the regional models to the two-state base, summarizing the results for three major types of coefficients. It addresses the two following questions arising from hypotheses formed in the early stages of the project:

Question H1: Are coefficients defined by individual characteristics or population segments more transferable then those defined for the entire population?

Yes, Figure 7.7b shows that, after controlling for differences of estimability across the coefficient types, substantially more of the coefficients defined by individual characteristics or population segments were insignificantly different between the regional and 2-state models, compared to those defined on the entire population.

Question H2: Are coefficients defined by population segments more transferable than alternative specific constants?

No, Figure 7.7b also indicates that ASCs are just as transferable as coefficients defined by population segments.

Table 7.7

Parameter type	insignificant difference	significant difference	not estimable	Number of parameters
entire population	73%	16%	10%	2309
alt specific constant	85%	9%	5%	942
individual or population segment	82%	9%	9%	1838

Figure 7.7a

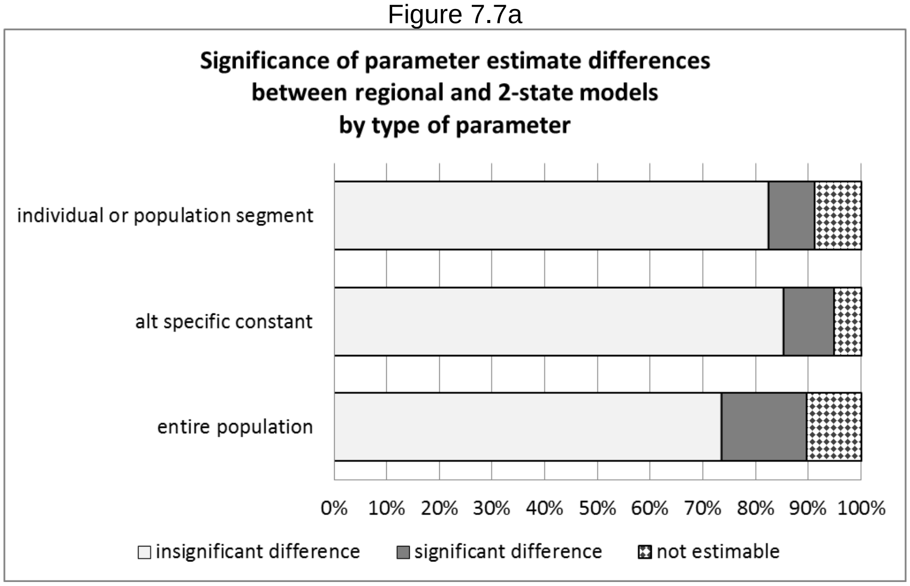
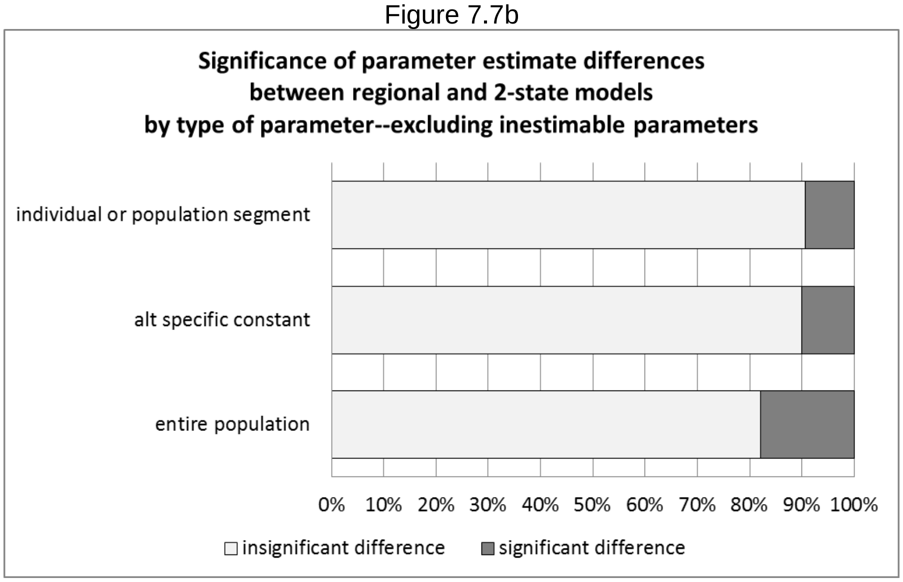


Figure 7.7b



Differences between regional and two-state models, and between regional and one-state models by type of model component

This section compares the regional models to the two-state base, summarizing the results for various types of models, from the perspectives of the model's level in the model hierarchy, the type of choice represented by the model, and a categorization scheme aimed at answering the third hypothesis raised early in the project. As it turns out, the corresponding questions for the comparisons of regions to their corresponding one-state models have the same answers, so they are addressed together here, and separate Tables and Figures are provided which show that the answers are the same.

Three questions are addressed:

Question H3.1: Which models are more transferable across states?

Question H3.3: Which models are more transferable within California?

Question H3.5: Which models are more transferable within Florida?

As shown in Figures 7.8b, 7.9b, and 7.10b, from the perspective of model hierarchy, the person day models are more transferable, followed by the tour, trip and long-term models. From the perspective of model type, the destination choice models are least transferable and the other types are similar.

Question H3.2: Are models that deal with social organization more transferable than those that deal mainly with spatial organization?

Question H3.4: Within California, are models that deal with social organization more transferable than those that deal mainly with spatial organization?

Question H3.6: Within Florida, are models that deal with social organization more transferable than those that deal mainly with spatial organization?

In this analysis, the models classified as 'social organization' include auto ownership and those that define the day in terms of the number and purposes of tours and intermediate stops. They exclude models that determine destination, mode and timing of tours and stops. Figures 7.8b, 7.9b and 7.10b show that social organization is similar in transferability to mode and timing choice, but more transferable than destination choice. When the social organization category is combined with timing choice and separated from the combined mode and destination choice category, social organization and timing models are more transferable.

Differences between regional and two-state models

Table 7.8

		insignificant difference	significant difference	not estimable	Number of parameters
Level	long term	74%	23%	3%	342
	person-day	86%	8%	6%	1262
	tour	76%	10%	14%	2279
	trip	78%	18%	4%	1206
Type	social organization	85%	10%	5%	2090
	destination	71%	25%	4%	828
	mode	65%	7%	27%	971
	scheduling	84%	10%	6%	1200
Hypothesis 3	mode & destination	68%	16%	16%	1799
	social org & scheduling	85%	10%	5%	3290

Figure 7.8a

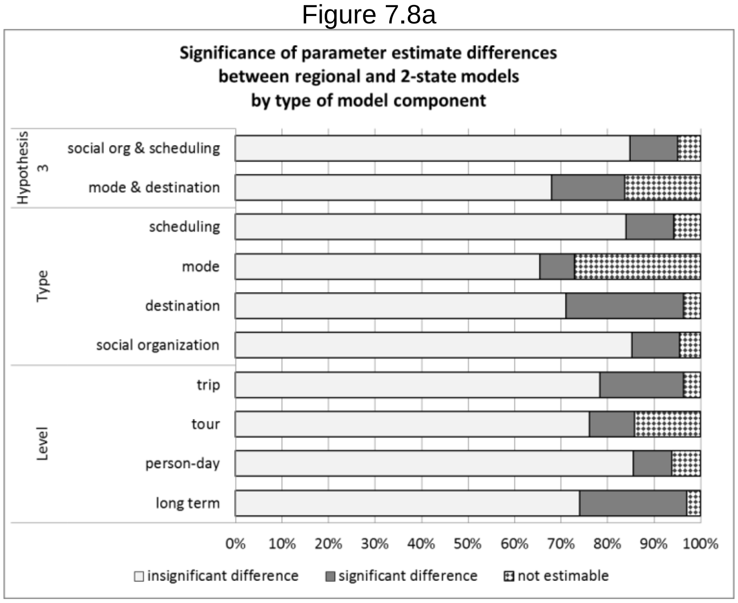
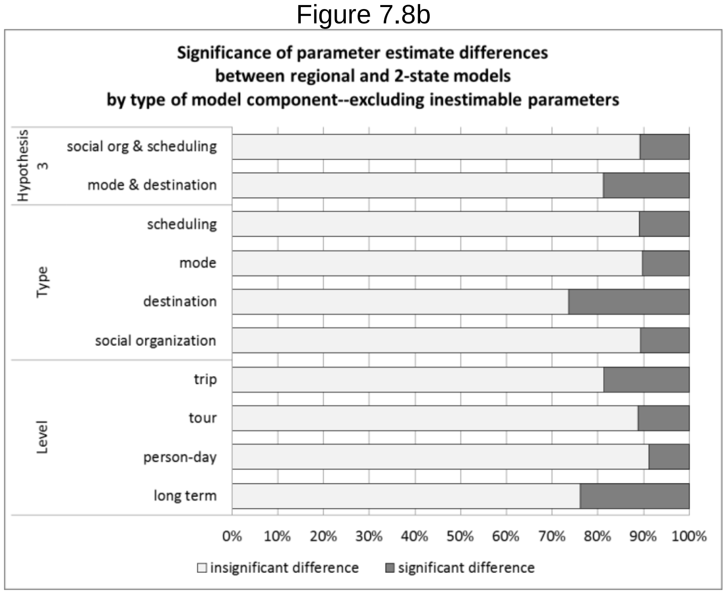


Figure 7.8b



Differences between regional and California models

Table 7.9

		insignificant difference	significant difference	not estimable	Number of parameters
Level	long term	79%	18%	3%	228
	person-day	85%	8%	8%	839
	tour	80%	7%	13%	1519
	trip	80%	16%	4%	804
Type	social organization	85%	9%	6%	1391
	destination	76%	21%	3%	552
	mode	71%	6%	23%	647
	scheduling	86%	7%	7%	800
Hypothesis 3	mode & destination	73%	13%	14%	1199
	social org & scheduling	86%	8%	6%	2191

Figure 7.9a

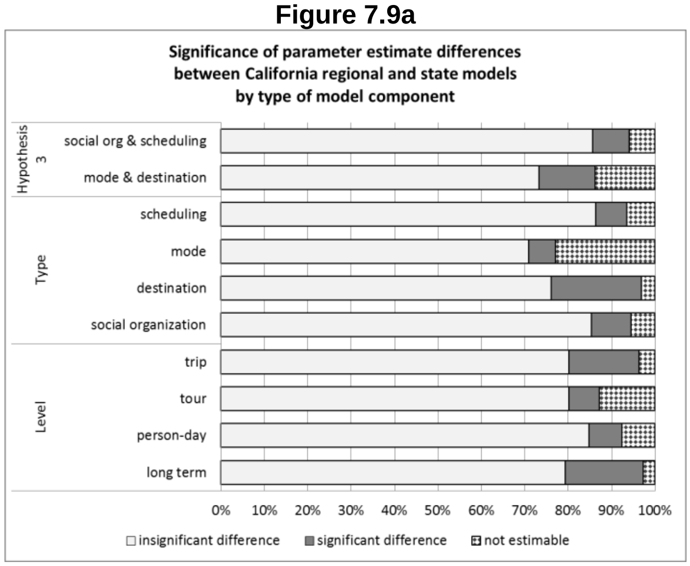
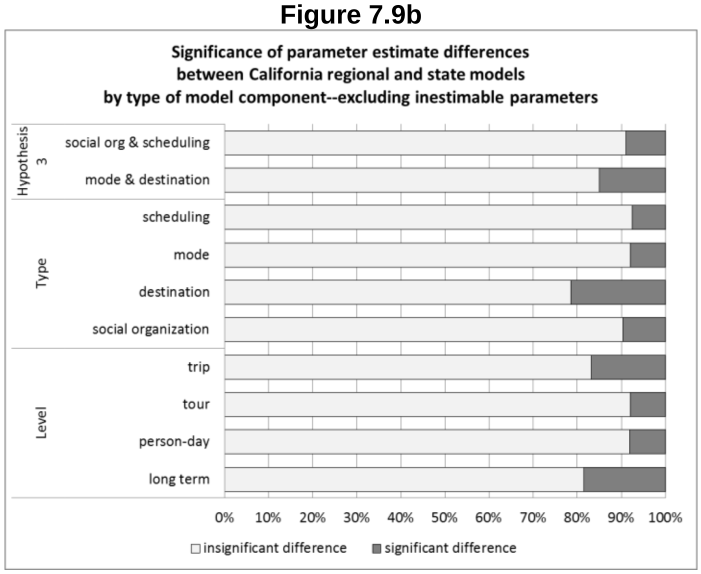


Figure 7.9b



Differences between regional and Florida models

Table 7.10

		insignificant difference	significant difference	not estimable	Number of parameters
Level	long term	77%	18%	5%	114
	person-day	85%	9%	6%	422
	tour	71%	9%	20%	760
	trip	76%	18%	6%	402
Type	social organization	85%	11%	5%	698
	destination	75%	19%	7%	276
	mode	52%	7%	40%	324
	scheduling	82%	12%	6%	400
Hypothesis 3	mode & destination	63%	13%	25%	600
	social org & scheduling	84%	11%	5%	1098

Figure 7.10a

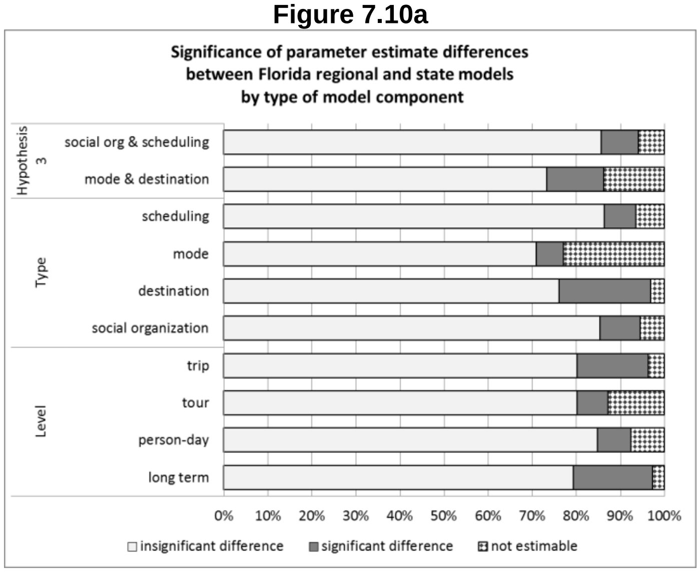
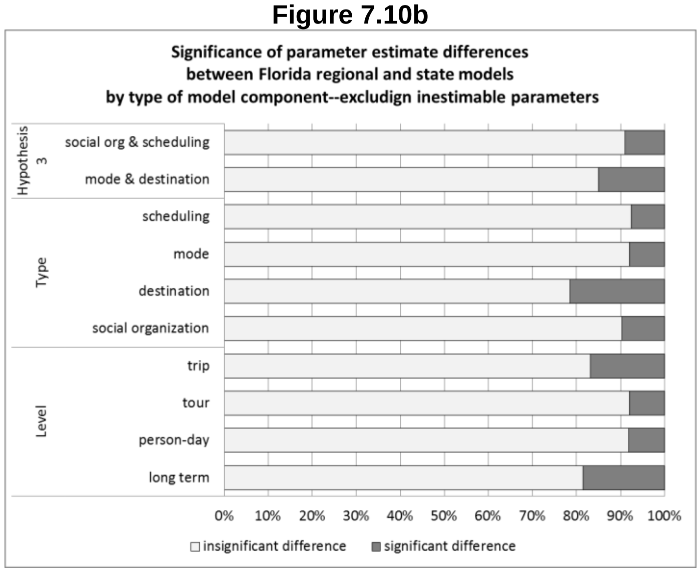


Figure 7.10b



Differences between regional estimates and those from combined one-state and two-state samples by region

This section compares each region's models to all four larger base models, focusing on separate results for each region. The model specifications that include ASCs for all regions in the base model are included because they might more closely resemble how a transferred model might be used in practice, where some of the ASCs would be calibrated to meet base year aggregate control values for the region. However, most of the conclusions in this section could be drawn regardless of whether ASCs for each region are included in the base model. Three questions are addressed:

Question H4.1: Can a region use models developed from a state or multi-state sample?

Figure 7.11b shows that, among coefficients that can be estimated at both levels, the percentage of significantly different coefficients ranges between 8% and 19%. Whether any or all of these levels is acceptable or not is a matter of judgment. However, as shown in Figure 7.11a, for most of the regions with smaller samples, such as Fresno, (survey sizes are noted in Table 7.11), a similar or greater percentage of coefficients was not estimable using only the region's sample. This highlights a trade-off that occurs when a region has the choice between using a small regional sample or a much larger multi-region sample: the quality of the model developed from a small regional sample is severely compromised, resulting in a lot of inestimable or insignificant coefficients, but the regional fit of a model estimated from a multi-region sample is imperfect. It is probably better to transfer a good model based on a sample of 6000 or more households than to develop a local model based on a sample as small as 2000 or less households.

Question H4.2: Are California models more transferable within California than they are across California and Florida?

Yes, as shown in Table 7.11 and Figure 7.11b, the percentage of insignificant differences between the regional model and the California model is the same or greater, and the percentage of significant differences is the same or smaller, than when the regional model is compared to the 2-state model. However, the differences are fairly small, and for the three-county region they are barely measurable.

Question H4.3: Are Florida models more transferable within Florida than they are across California and Florida?

The answer differs for Tampa and Jacksonville. As with the California regions, and again looking at Table 7.11 and Figure 7.11b, Tampa's differences from the 2-state model are greater than they are from the Florida model. However, Jacksonville's results are more similar to the 2-state model than to the Florida model, suggesting that Jacksonville's model is more similar to the California models than to the Tampa model. One possible reason for this is that Tampa has a larger fraction of retired and seasonal residents compared to Jacksonville. (See Appendix 1 comparing the data from the various regions.)

Table 7.11

	Base model type	insignificant difference	significant difference	not estimable	Number of survey households
sandag	2 state	85%	14%	1%	6002
	1 state	87%	11%	2%	
	2 state & ASC	89%	10%	1%	
	1 state & ASC	90%	8%	2%	
sacog	2 state	85%	12%	3%	1311
	1 state	87%	10%	3%	
	2 state & ASC	87%	9%	4%	
	1 state & ASC	88%	8%	4%	
fresno	2 state	74%	9%	17%	381
	1 state	75%	8%	17%	
	2 state & ASC	72%	7%	21%	
	1 state & ASC	72%	7%	21%	
3county	2 state	76%	11%	13%	657
	1 state	76%	11%	13%	
	2 state & ASC	75%	9%	17%	
	1 state & ASC	76%	8%	17%	
tampa	2 state	74%	18%	9%	2517
	1 state	76%	12%	12%	
	2 state & ASC	75%	14%	11%	
	1 state & ASC	74%	11%	15%	
jacksnvl	2 state	79%	10%	11%	1335
	1 state	76%	12%	12%	
	2 state & ASC	79%	7%	14%	
	1 state & ASC	74%	11%	15%	

Figure 7.11a

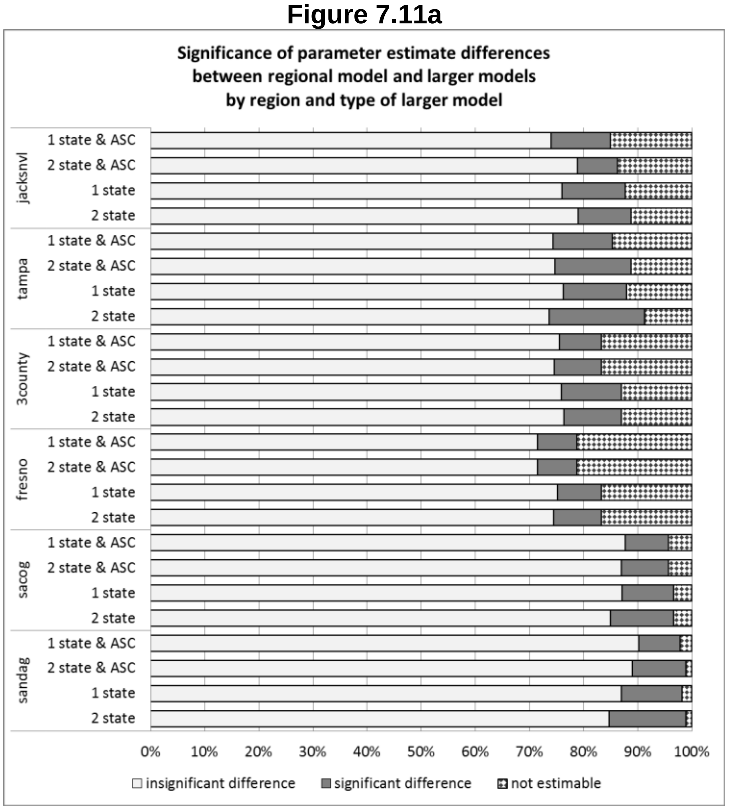
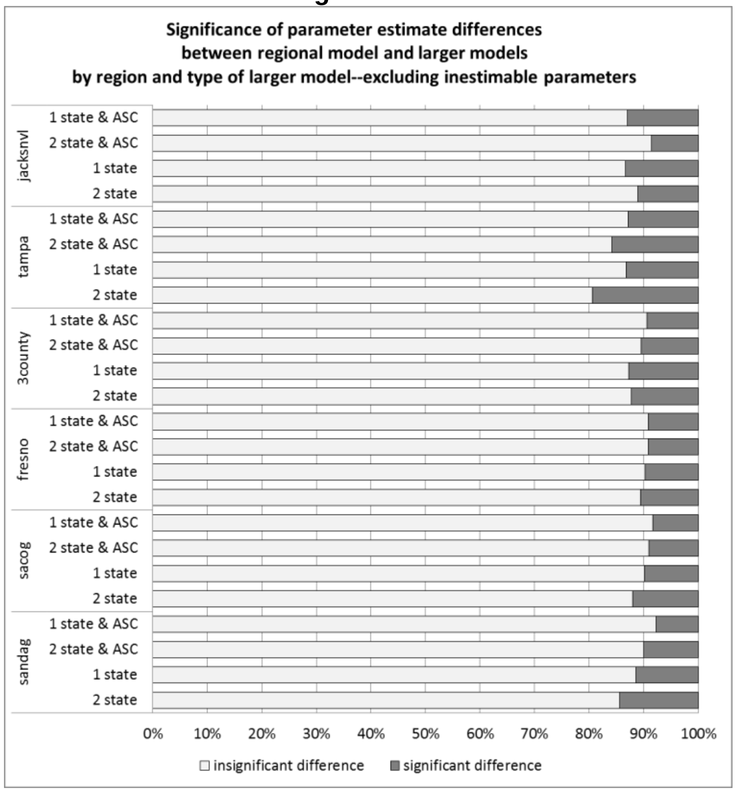


Figure 7.11b



Summaries of Likelihood Ratio Tests of Model Equivalence

In contrast to the preceding sections, which give summaries based on comparisons of individual coefficients, this section summarizes likelihood ratio tests, each of which evaluates whether a particular regional model is identical to its associated larger model.

For a likelihood ratio test, two models, one a restricted version of the other, are estimated by the method of maximum likelihood. If the two models are essentially the same, then the additional estimated coefficients in the unrestricted version will be statistically insignificant and there will be little difference between the final likelihood values of the two models. Under the null hypothesis that the models are the same, the likelihood ratio statistic will have a chi-squared distribution with the number of degrees of freedom equal to the difference in the number of estimated coefficients between the unrestricted and restricted models; that is,

$-2(L(\hat{\theta}_R) - L(\hat{\theta}_U)) \dot{\sim} \chi^2_q$, where $L(\hat{\theta}\ \ )$ is the loglikelihood measure from maximum likelihood estimation, R is the restricted model, U is the unrestricted model, and q is the number of degrees of freedom. This statistic can be used to find the probability, under the null hypothesis, that the difference between estimation results of U and R would be as great as those observed, or, in other words, the degree of confidence that the two models are the same. Conversely, it provides the probability that there is at least some difference between the two models.

For the summary reported in this section, each test consists of a pair of model runs for one of the 15 model types (auto ownership, work tour mode choice, etc). The unrestricted model is one of four base models with a full set of difference coefficients for a named region:

1. Two-state base with difference coefficients for the named region
2. One-state base with difference coefficients for the named region
3. Two-state base with ASCs for all regions and difference coefficients for the named region
4. One-state base with ASCs for all regions and difference coefficients for the named region

The restricted model is the same model specification but without the difference coefficients for the named region.

Table 7.12 summarizes the results of these tests. For each region and base model type, it tabulates the percentage of test results falling into each of three categories, representing low, medium and high significance of the differences. It also gives the number of estimated model pairs contributing to the summary tabulation for each base model type (spanning all regions), and the number of survey households in each region. Figures 7.12a through 7.12d provide the significance of differences in graphical form, one figure for each base model type.

The primary purpose of generating this summary is to address the following question for the various regions:

Question H4.4: Is the region's model essentially the same as the combined within-state or two-state model?

It is clear from the length of the dark bands in the figures that for all states and base model types, the confidence exceeds 95% for a substantial portion of the models that the region’s model is <u>not</u> identical to the base model estimated using combined data across the larger geography. On the other hand, for a substantial portion of the models, represented by the medium grey bands, the confidences ranges only between 75 and 95%, and for another substantial group, represented by the lightest bands, the confidence that the region’s model is not identical to the larger geography’s model is under 75%. It is notable that the confidence that the region’s model is different is positively correlated with the region’s survey sample size. For a region with a small sample size, such as Fresno with only 381 households, the sample lacks the information needed to confidently distinguish it from the other regions, regardless of whether or not it really is different. This corroborates an earlier conclusion that it is probably better for a region that could only afford a small sample to transfer a model than to develop one from a small sample. In contrast, for a region with a large sample size, such as SANDAG, the sample contains the information needed to confidently distinguish it from the other regions. This is true even though its large size gives it the most influence over the estimation results of the restricted model.

Comparing results for the four different base model specifications sheds some light on preferable approaches for transferring models. First, comparing the 2-state results in Figure 7.12a with the 1-state results in Figure 7.12b reveals that, for all regions, the region’s model is as similar or more similar to the 1-state model than it is to the 2-state model. The same holds true when comparing Figures 7.12c and 7.12d, with the exception of Jacksonville, corroborating previous evidence that the Jacksonville model may be more similar to the California models than to the Tampa model. Nevertheless, most of the evidence here is that transferability is greater within state than across states.

Comparing the 2-state results in Figure 7.12a with the 2-state results where the base includes region-specific ASCs shows a greater similarity to the base in all regions when the base accounts for regional differences in the ASCs. This suggests that if a shared-data approach is used for model estimation, it may be good practice to estimate region-specific ASCs. It also suggests that, if a region is adopting a model that has been estimated on another region's data, it would be good practice to calibrate the ASCs, to the extent that aggregate data in the borrowing region is available for the calibration.

Table 7.12

Base model type	Significance of differences	sandag	sacog	fresno	3county	tampa	jacksnvl	Number of estimated model pairs
2 state	Low (under 75%)	7%	21%	29%	0%	13%	20%	13
	Medium (75-95%)	27%	29%	43%	29%	7%	27%	23
	High (over 95%)	67%	50%	29%	71%	80%	53%	51
1 state	Low (under 75%)	20%	21%	43%	14%	33%	27%	23
	Medium (75-95%)	20%	29%	36%	29%	13%	20%	21
	High (over 95%)	60%	50%	21%	57%	53%	53%	43
2 state & ASC	Low (under 75%)	33%	33%	42%	25%	17%	50%	24
	Medium (75-95%)	8%	17%	33%	25%	17%	8%	13
	High (over 95%)	58%	50%	25%	50%	67%	42%	35
1 state & ASC	Low (under 75%)	33%	33%	42%	33%	36%	33%	25
	Medium (75-95%)	8%	25%	33%	17%	9%	17%	13
	High (over 95%)	58%	42%	25%	50%	55%	50%	33
Number of survey households		6002	1311	381	657	2517	1335	

Figure 7.12a

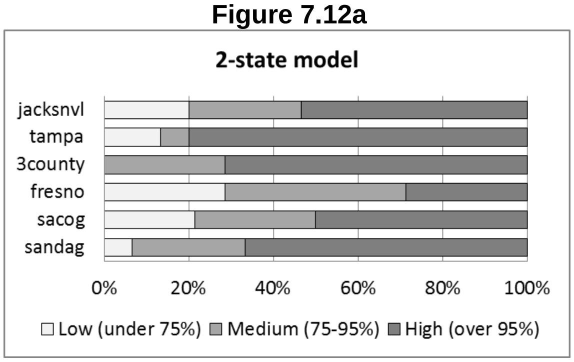


Figure 7.12b

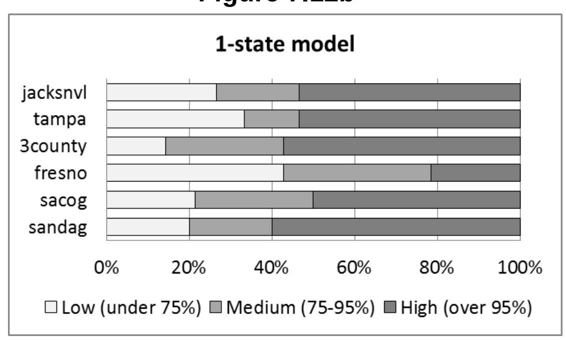


Figure 7.12c

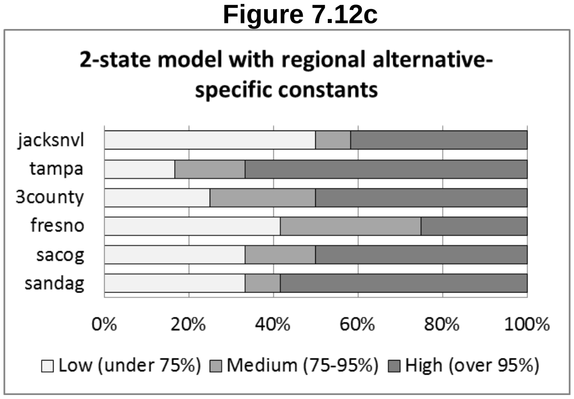
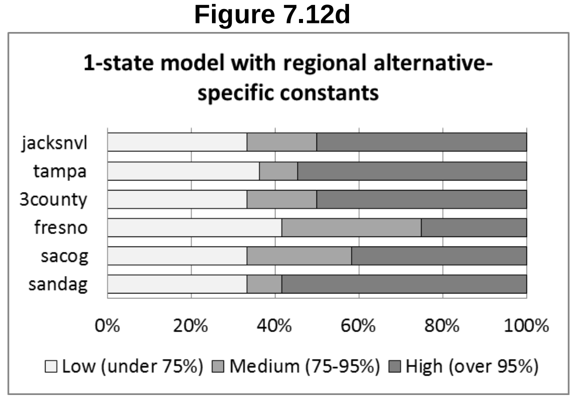


Figure 7.12d



It is possible to summarize the results of the likelihood ratio tests in other ways to cast light on previously addressed questions. Table 7.13 and Figures 7.13a through 7.13d summarize the results by four model types, to provide another look at the following question:

Question H3.2: Are models that deal with social organization more transferable than those that deal mainly with spatial organization?

As can be seen in both the table and Figures 7.13c and 7.13d, the destination choice models are excluded from the summaries in the two cases where the base model includes region-specific alternative-specific constants because destination choice models do not have alternative-specific constants.

In both Figures 7.13a and 7.13b, it can be seen that the destination choice models have the greatest statistical difference between the region and the larger one-state and two-state models, followed fairly closely by the mode choice models. The social organization and scheduling models are much more similar across geographic levels. This result is different than the earlier result based on summarization of individual coefficients (see Table 7.8 and Figure 7.8b), which suggested that the mode choice models are nearly as transferable as the scheduling and social organization models.

Not surprisingly, when comparisons are made across the four base model types, summarizing by model type shows a similar pattern as summarizing by region: the regional models are more similar to the one-state model base than to the two-state base, and they are more similar to the base models with region-specific ASCs than to those without region-specific ASCs.

Table 7.13

Base model type	Significance of differences	social organization	destination	mode	scheduling	Number of estimated model pairs
2 state	Low (under 75%)	20%	7%	8%	22%	13
	Medium (75-95%)	33%	13%	21%	33%	23
	High (over 95%)	47%	80%	71%	44%	51
1 state	Low (under 75%)	27%	13%	17%	50%	23
	Medium (75-95%)	33%	7%	29%	17%	21
	High (over 95%)	40%	80%	54%	33%	43
2 state & ASC	Low (under 75%)	40%	0%	21%	39%	24
	Medium (75-95%)	10%	0%	21%	28%	13
	High (over 95%)	50%	0%	58%	33%	35
1 state & ASC	Low (under 75%)	40%	0%	17%	53%	25
	Medium (75-95%)	10%	0%	33%	12%	13
	High (over 95%)	50%	0%	50%	35%	33

Figure 7.13a

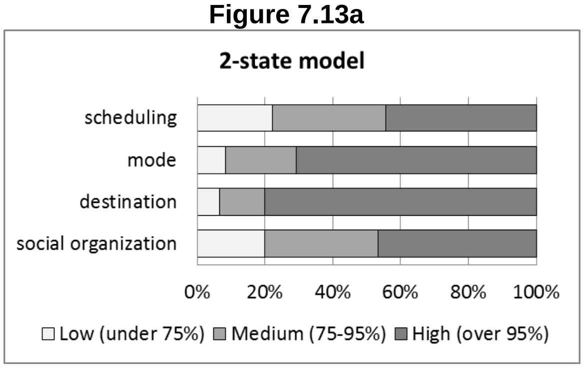
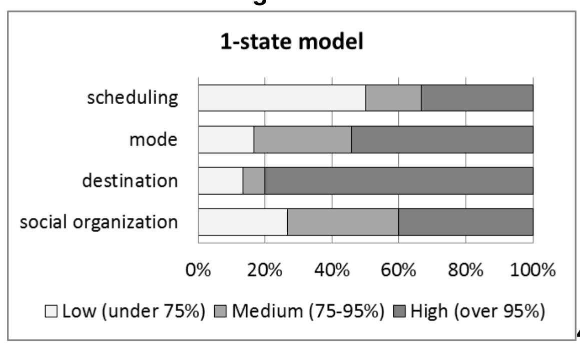


Figure 7.13b



4

Figure 7.13c

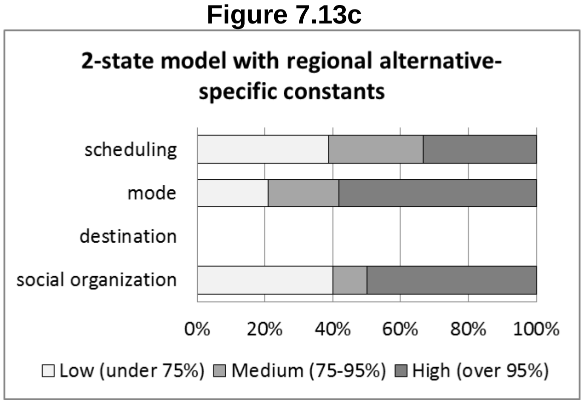
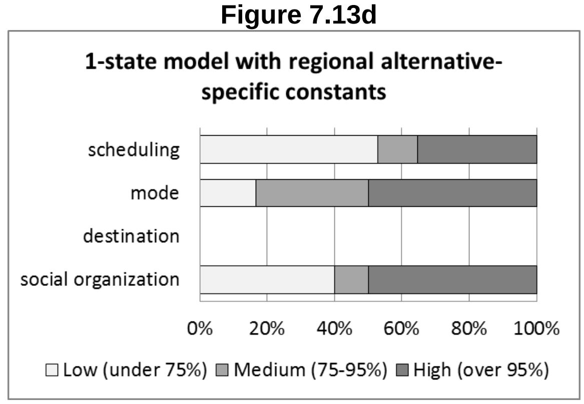


Figure 7.13d



8. Conclusions

Summary of Findings

Though it is not possible to make a definitive statement about the transferability of activity-based models based on a single study, this study has been able to provide some new and unique evidence. Overall, although the strictest statistical test (chi-squared test) usually rejects the hypothesis that models based on data from different regions are statistically indistinguishable, it is also true that most of the individual coefficients are not significantly different from one region to the next. In addition, this study shows the substantial improvement of estimability that occurs with large survey samples. Based on these findings, **the most important conclusion of this study is that, although estimation of models using a large local sample is best, it is better to transfer models that are based on a large sample from a comparable region than it is to estimate new models using a much smaller local sample.**

However, this conclusion does **not** mean that metropolitan regions can relegate survey data collection and model development to the past, and simply borrow a model from others who have gone before. Even if a comparable region and its model can be found, survey data should nevertheless be collected for purposes of calibrating components of the model, such as activity and tour generation, that cannot be calibrated using traffic count data. And this study has only scratched the surface in identifying the factors that make two regions comparable, as described in the next paragraphs.

Although small sample sizes limit the ability to draw strong conclusions about comparability among the four California regions and two Florida regions included in this study, there is some substantial evidence of comparability among them. However, Tampa stands out as less comparable than the others, and this study did not identify the reason. The California regions are more comparable within state than across states, perhaps because of the presence of Tampa in the two-state comparison. The issue with Tampa draws attention to the likelihood that there may be factors would cause two regions, even two regions within the same state, to be bad candidates for a model transfer.

This study did not explore comparability for regions in states other than California and Florida, estimability and comparability for a full spectrum of sample sizes—especially samples with more than 2,500 households, or comparability in categories other than state boundaries, such as urban density, size, or demographic make-up. For example, university towns or cities with large seasonal retirement population may be distinctly different in ways that make transferring from other regions inadvisable, and this study lacks evidence to draw conclusions one way or the other. These remain as important avenues for further research.

On the other hand, there may be good reasons for transferring a model from a region that is NOT currently comparable if there is reason to believe that it will be comparable in the future. For example, it may be that a region that is growing rapidly and/or adding new travel options would lack the data to develop a model that would serve it well even it could collect a very large household survey. The diversity of conditions needed to estimate the coefficients of the model simply may not exist within the region. In a case like that, perhaps a model transfer should be considered.

This study is also limited in its ability to determine what sample size is large enough for local estimation, because the largest sample in this study includes only 6,000 households and the rest are 2,500 or less. However, the results suggest that **sample sizes of less than 6,000 households should be discouraged. It is also very likely that samples considerably larger than 6,000 would substantially improve estimation results**, enabling significant coefficient estimates for important small population segments.

The following paragraphs provide more context to these overall conclusions, as well as additional findings about the NHTS, sample size issues, and transferability hypotheses tested in this study.

Adequacy of the 2009 National Household Travel Survey (NHTS) data for estimating activity-based models

A unique aspect of this project is that identical models are estimated using data from six different regions, but all using data from the same household travel survey. Because the opportunity to purchase an add-on sample to the NHTS is available to all state and regional agencies, it may be useful to assess how well the data support the estimation of the component models of the activity-based model system (the DaySim v.1.8 model system in this case). The following findings can be helpful in the implementation of future NHTS surveys, as well as other household surveys designed for use with AB models:

- In terms of survey design, the NHTS 2009 survey supports the estimation of almost all of the component models of DaySim. The exceptions are:
    - Most AB model systems have a longer term model of students' usual school locations, and NHTS 2009 did not ask this question. Although it is possible to impute this information for students who actually went to school during the travel day, this may be a biased sample of all students.
    - DaySim 1.8 has models of transit pass ownership and of the availability of free parking at the usual workplace, both of which are useful in predicting mode choice. Neither of these variables was collected in NHTS 2009, but could be added in future surveys by adding fairly simple questions.
- In terms of sample design, there are some more serious issues:
    - The sample sizes for most of the individual regions used in this study are not adequate to estimate statistically significant parameters for many of the variables tested. This is particularly true for some of the rarer types of households and persons, such as low-income households, zero-vehicle households, and persons who use transit

and bicycle. These issues arise in many household travel surveys, not just NHTS. For NHTS, the use of equal subsamples for every day of the week means that 28% of travel days are on weekends, and are not useful for the typical regional models of weekday travel. It should be considered to devote a smaller percentage of the sample to weekends, unless models of weekend travel are important for the region.

- There are many households in the sample for which one or more household members did not complete the travel diary. In the case of children under age 5, this was by design, and is not a major problem for modeling, since the large majority of travel and activities of those children can be imputed from the joint trips they make with the adults in the household—assuming that all adults complete the travel diaries. In the 2009 NHTS data, however, there are many cases in which one or more household adults did not complete a travel diary. As a result, it is not possible to explicitly model all joint travel and activities in the household. This is not a major problem for the DaySim 1.8 model system, which does not contain models that explicitly link the joint tours and trips across household members, but it would be a problem with later versions of DaySim that do contain such models. Vovsha et al.(2011) also report such issues when using the NHTS data for estimating AB models for Phoenix and Tucson. The result is that households with incomplete data have to be excluded from estimation for some models, which further exacerbates any sample size problems (particularly among larger households). For the next NHTS, a more strict definition of data for a "complete" household should be considered.

Issues regarding sample size and composition

The adequacy of the survey sample size can be judged mainly from (a) how many of the coefficients in the models can be estimated with statistically significant precision (i.e. t-statistics of 1.9 or higher), and (b) how many of the coefficients can be estimated at all. The latter issue of non-estimable parameters usually arises in the case of variables that apply to only small segments of the population and for which no variation in choice behavior can be observed. An example is a case in which no households in the lowest income category choose the walk mode, so a low-income variable for the walk mode cannot be estimated.

In regard to overall sample size, only the San Diego region, with roughly 6,000 households seems to be adequate to estimate statistically significant coefficient values for the majority of variables in most models. The next largest sample is for the Tampa region, with only about 2,500 households. The percent of coefficients in all (base) models with significant estimates of the expected sign fell from 65% for San Diego to 46% for Tampa, and then to 39% for Jacksonville and Sacramento (sample sizes of 1,000-2,000 HH) , and to 27% for Fresno and Northern San Joaquin Valley (less than 1,000 HH). The percent of parameters that are not estimable at all goes up as sample size goes down, with 0% for San Diego, rising to 11% for Fresno, the region with the smallest sample. From this analysis it cannot be concluded that samples larger than 6,000 households are of little or no value, for two reasons. First, the only variables considered in this study are those included in the DaySim v1.8 models, which were originally specified using data from a local Sacramento region survey of roughly 5,000 households. A larger sample would have almost certainly yielded models with additional significant coefficients. The second reason is that this study does not include any samples of size larger than 6,000, so it provides no empirical evidence about larger sample sizes. In this study, no regions are included with a sample size in between 2,500 and 6,000 households, which is a fairly large gap. Thus, it is also difficult to say how much improvement a sample of, say, 4,000 would provide compared to a sample of 2,500. One way in which this research can usefully be extended is to include a number of additional regions with varying sample sizes that cover the range from 2,500 to 6,000 more completely (while simultaneously covering states in more parts of the US). It would also be desirable to test estimability and transferability with sample sizes in excess of 6,000. Nevertheless, what can be concluded is that, among the sample sizes included in this study, much is to be gained from larger sample size.

As an alternative to transferring models from other regions, an option is to combine survey data from similar regions. For example, the Sacramento, Fresno and Northern Joaquin regions are all in fairly close geographic proximity to each other, and combining data from those three areas would yield a sample of 2,500 households, similar to the sample size for Tampa. Note that this does not require combining the regions' zone systems, networks, etc. into a "megaregion", but it does require at least merging the data sets after they have been prepared for estimation, as was done for this project. However, the current study provides only limited information about what constitutes "similar" for purposes of combining data. By explicitly evaluating differences of individual regions' results from combined one-state and two-state samples it only provides information about defining 'within state' as a basis of similarity, and the results are mixed, as described below. A practical issue in combining data across regions is that the process can be time-intensive and complex, which might place it beyond the scope of many regional modeling efforts.

The variation in results in response to sample size appears to affect all types of models, including mode choice, destination choice, time of day choice, tour and trip generation, and auto ownership. The problems associated with small samples arise in all models in the estimation of coefficients related to important but small or hard-to-sample segments of the population, such as very low or very high income households, households that do not own cars, and young adult households. The effects are most apparent in cases where specific alternatives have a small number of observed cases. In the tour mode choice models, for example, there are very few cases in which people chose the transit and bicycle modes, particularly in the Florida regions and the more rural California regions with smaller sample sizes. The actual mode shares for transit and bicycle in those regions tend to be quite small, and, even though some oversampling was done in NHTS to include more households from neighborhoods likely to have higher transit and non-motorized mode usage, the resulting shares for those modes in the NHTS data for the selected regions is still very low. (In future NHTS efforts, it may be worth the effort to oversample even more heavily in such neighborhood areas, and/or to reconsider the areas that should be oversampled. Areas around universities and other areas with a high proportion of young, single households may be advantageous, both to get a wider variety of mode choice behavior, and to counteract the typical survey sampling bias towards older households.)

Tests of specific hypotheses regarding model transferability

In this study, model transferability between regions is tested through the use of region-specific "difference models". Starting with a "base model", estimated on all regions within both states (or else all regions in a single state), additive "difference coefficients" are added for all variables for observations from a single region. The difference coefficients thus measure the difference between the coefficient estimated only for that specific region and the coefficient estimated for all regions except that specific region. If the difference coefficient is not significantly different from zero, it is an indication that the model is transferable between the specific region and the other regions included in the base model—at least for that particular variable. This approach tends to work best with adequate sample sizes. With small sample sizes, although it is more difficult to estimate a local version of the model, it is also more difficult to prove whether or not the local version is statistically different from models estimated on other regions, because there is less data to use in estimating the difference coefficients and thus they will tend to be less significant, regardless of the true transferability. Therefore, this test is less useful in judging variations in sample sizes, and the comparisons tend to be most meaningful along model dimensions not defined by region (where sample sizes differ), such as comparisons between types of models or between types of variables (including data from all regions without differentiating results by region).

Another way that transferability is tested in this study, rather than looking at each variable separately, is to look at the change in model fit as a whole when the region-specific difference variables are included. This is done with the standard chi-squared likelihood ratio test. Such a test is typically used to test different model specifications on the same estimation data set, whereas in this study it is used to test the same model specification on different estimation data sets. The likelihood ratio test is a rather strict test to use, because it tests whether the two models being compared are the same in every respect. Indeed the hypothesis that the different data sets yield "the same model" (in statistical terms) is rejected with fairly high significance in most of the cases. Nevertheless, the results of the likelihood ratio test can be used in a comparative manner, finding out which models show the most and the least improvement in model fit due to allowing region-specific coefficients.

Each of the transferability hypotheses raised in Chapter 2 and tested in Chapter 6 are discussed below, with attention to the practical implications of the research findings:

(Hypothesis 1) *Variables that apply to population segments defined by characteristics of individuals and/or their situational context (i.e., segment-specific variables) will tend to be more transferable than variables that are more generic and apply to all individuals*.

This hypothesis is supported by the data, with 82% of the variables that are population-segment-specific showing no significant difference from the base model, compared to 73% for variables that apply to the entire population. Looking at it the other way around, 27% of full-population variables show significant cross-region differences or are not estimable, compared to only 18% for segment-specific variables, a relative difference of 50%.

Of course, estimating segment-specific variables requires adequate sample size for each segment, which requires larger sample size overall. The smaller effective sample sizes applying to these variables in estimation may be part of the reason that fewer significant differences result. Nevertheless, the results support the theory that variables with more socio-demographic specificity should show more stability across regions. This finding also supports the idea that it may be better to transfer activity-based models from a region with a large enough survey sample to include segmentation detail, rather than estimating on a much smaller local survey sample that does not allow such detail.

(Hypothesis 2) ***Variables that are segment-specific will tend to be more transferable than alternative-specific constants.***

This hypothesis is <u>not</u> supported by the tests. Compared to the segment-specific variables, the alternative-specific constants (ASCs) show the same percent of variables across all models with significant differences from the base model (9%). The ASCs are also more estimable, with only 5% of cases not estimable, compared to 9% of segment-specific variables.

This finding implies that alternative-specific constants may be fairly transferable between regions as well, which would be reassuring in a situation when there is very little observed local data at all to use in calibrating the model to local conditions. When such data do exist, however, it is still good practice to use local data to re-calibrate the alternative-specific constants of a transferred model whenever necessary. Although data sets from surveys such as NHTS may not be adequate for complete re-estimation of an activity-based model system, they will typically be adequate for such a simple calibration exercise (which can usually be done by manually adjusting the constants and comparing the observed versus predicted choice shares, rather than using formal model estimation methods).

(Hypothesis 3) ***Models that deal with social organization (activity generation and scheduling) will tend to be more transferable than models that deal mainly with spatial organization (mode choice and location choice).***

The data strongly support this hypothesis. The models emphasizing "spatial organization" result in 16% of variables with significant differences across regions and 16% non-estimable on single-region data, compared to only 10% and 5% respectively for models emphasizing "social organization". This conforms to the expectation that different US regions will tend to show stability in general socio-cultural patterns, as those are less influenced by the spatial dispersion of opportunities or the availability and quality of particular travel modes.

In practical terms, this finding suggests that when transferring an activity-based model system across regions, it would be warranted to spend more effort in calibrating and adjusting the mode choice and destination choice models than the other models. Vovsha et al (2011) also report that destination/location choice models can show different results across regions, particularly with the sensitivity to travel distance, so calibration to local data on trip distance distributions is recommended, unless the model is transferred from a region with very similar geography and dispersion of employment and population.

(Hypothesis 4) ***Models for different regions within the same state will tend to be more transferable than models for regions in different parts of the country*:**

In most cases, the data support this finding, with the California regions showing higher transferability for the one-state California base model than for the two-state base model, and with Tampa showing higher transferability for the one-state Florida model than for the two-state model. The exception is the Jacksonville region, which shows somewhat less transferability with the one-state Florida model than with the two-state model. This suggests that in at least some aspects of travel behavior, Jacksonville residents behave more like residents of one or more of the California regions than like Tampa residents.

While this study is ambitious in using data from six different regions, findings such as this one make it evident that six regions in two states is still a quite limited number to work with if one wishes to generalize about transferability within versus between states or different types of regions. For this reason, it would be advisable to extend the current study to additional regions, as mentioned in the overall summary above and described in the suggestions for further research below.

It is important to recognize that all of the transferability findings above are inter-related in the sense that the level of detail in the specification of the models being tested can substantially influence the findings. For example, the finding for Hypothesis 1 suggests that if one were to test the transferability of a "stripped down" version of the DaySim models that included far fewer variables that are segmented by activity purpose and household and person characteristics, then it is likely that those models would be assessed as somewhat less transferable than the ones tested here. The inclusion of segment-specific detail tends to influence the "social organization" models the most, so the finding for Hypothesis 3 that such models appear more transferable than "spatial organization" models is due in part to the inclusion of those variables that capture the similarities in how certain types of households and persons tend to engage in and schedule activities. Similarly, if such variables were not included in the models, then differences across regions in the prevalence of different household, person, and tour characteristics would tend to be captured in the alternative-specific-constants instead of the omitted variables, so the finding for Hypothesis 2 might change as well. In general, the AB model systems in use in the US include more of this type of behavioral detail than most 4-step models do, so they will also tend to be more suitable for transferring to other regions.

Comparison To Two Recent Transferability Studies

In this section, the research conducted in this project is compared to the research of Vovsha, et al (2011)[4] and Sikder and Pinjari (2012)[5], both of which use an estimation-based approach to test

[4] Vovsha, P, S Gupta, J Freedman, W Sun and V Livshits (2011) Workplace Choice Model: Comparison of Spatial Patterns of Commuting in Four Metropolitan Regions, Presented at the 91st Annual Meeting of the Transportation Research Board, January 2012, Washington, D.C.

the transferability of components of activity-based (AB) model systems. The comparison touches on research method and scope, as well as on the research findings.

Vovsha, et al

Vovsha, et al compare estimation results of one closely related pair of models for four metropolitan regions in the United States. The models constitute the choice of work location via (1) the binary choice between working at home and working away from home, and (2) the traffic analysis zone of the work location, given that it is not at home. The four regions are San Diego, Tucson, Phoenix and Chicago. Except for the estimation of the home-vs-away binary choice model using a joint Phoenix-Tucson data set, in which the only region-specific coefficient tested is the alternative-specific constant, the authors rely on informal comparison of model coefficients estimated separately for the four regions.

Vovsha, et al, acknowledge the need for more rigorous statistical testing using pooled data sets, and this is one of the big differences between their work and this project. In this project, models are estimated using combined data sets across multiple regions in two states, testing for differences of a region's models from the other regions within its state, and from a combined two-state data set. In addition, this project examines fourteen different model components, rather than one. It also conducts formal tests for differences for all model coefficients, not just alternative-specific constants, using tests for differences in individual coefficients as well as difference tests for each model as a whole.

Vovsha, et al, report substantial differences in the model estimation results across regions and conclude that, for the work location choice, estimation using local data is important. Despite the differences in scope and rigor of testing, their conclusion is consistent with this project's finding that the location choice models are the least transferable of the various types of models in an AB model system.

Sikder and Pinjari

Sikder and Pinjari examine spatial transferability of an activity generation and duration model among regions within Florida, and between Florida and California. The model predicts jointly the amount of time spent out-of-home for each of eight activity purposes, and corresponds most closely to the person-day tour generation model in this study. It uses the same NHTS survey data that is used in this project. It relies primarily on demographic characteristics and a few simple spatial attributes to explain behavior, without considering the effect of travel conditions for the out-of-home activities, which is a major difference from the models tested in this project. Their study deals only with adults who are not employed rather than the entire population. Since their study deals only with one model, and for a limited subset of the population, it is unable to draw broad conclusions about transferability of entire model systems.

---

[5] Sikder, S and A R Pinjari (2012) Spatial Transferability of Person-Level Daily Activity Generation and Time-Use Models: An Empirical Assessment, Presented at the 92nd Annual Meeting of the Transportation Research Board, January 2013, Washington, D.C.

The authors estimate the model for each of three regions in Florida, for a combined Florida data set, and for a combined California data set. They transfer each of the three regional Florida models to the other two regions and two additional Florida regions in two different ways: (1) without making any changes, and (2) after re-estimating the models on the new region's data allowing only the alternative-specific constants to change. They transfer models between Florida and California in the same way, using the state-wide combined data samples. This method of conducting pairwise comparisons contrasts with the current project's approach of comparing a region's model to that of combined one-state and two-state models. Their approach provides more information about transferability between two specific regions. For example, their pairwise comparisons with five Florida regions enable them to say that the model transfers better among urban regions than it transfers from urban to rural regions. On the other hand, the pairwise approach provides less information about the suitability of using pooled data to supply the model for a region, and also less information about the similarity of a particular region's model to another state. In particular, although their study determines that there are substantial differences between the Florida and California models, and provides evidence that the Tampa model is the most different from the other Florida regional models, it is unable to detect that Jacksonville's model is more similar to the California model than it is to the Tampa model, as is found in the current project.

Sikder and Pinjari use several statistical measures to evaluate transferability. These include one measure based on the likelihood function and two measures based on estimated shares, all of which compare the measure for the borrowing region to that for the donor region when the model is applied to the estimation data set. They also include two measures of the magnitude of change in model results when the values of a major input variable in the estimation data set are changed. Here also, the measures compare the magnitude of the change for the borrowing region to that for the donor region. These last two sensitivity measures enable the authors to draw conclusions that cannot be drawn in the current project, which does not include sensitivity tests in model application. In particular, they find that, although re-estimating the constants in the transferred model improves its fit with the borrowing region's estimation data, it does not help, and sometimes even hurts, the transferred model's predictive performance when the input data changes. This leads them to emphasize the importance of testing model sensitivity in the investigation of model transferability.

Suggestions for Further Research

Several opportunities exist to extend the research of this project in order to provide additional valuable information about the transferability of advanced travel demand models. The first natural extension would be to increase the number of states and regions included, selecting two or more regions with substantial NHTS samples from each of several additional states. This would lead to more robust answers to the hypotheses and questions already addressed. It would provide stronger statistical support for the conclusions, helping to clarify which answers are particular to California and Florida and which are more general. It would provide more evidence about the benefit of increasing sample sizes in the range between 2,500 and 6,000 households (the gap between the two largest samples in this study—Tampa and San Diego). It also would provide stronger evidence about the advisability of estimating models using multi-state samples and transferring model specifications across state lines versus doing so only within a state.

Based on identification of the largest add-on samples from the 2009 NHTS, it is advisable to add data for at least two regions in at least three of the states/areas listed below, with the most likely candidate regions in parenthesis:

- New York state (Albany, Rochester, Syracuse)
- Texas (Dallas, Austin, San Antonio)
- Midwest (Omaha NE; Cedar Rapids IA; Indianapolis IN)
- South (Greensboro NC; Charlotte NC; Nashville TN)

This would bring the total number of regions used to at least 12, with at least 2 in each of 5 states/areas of the US.

A second natural extension would be to pool regions for combined estimation and transferability testing according to categories other than state lines. This would help determine the best criteria for matching donor and borrower region, or for pooling data. Many categorization schemes might be good candidates, and careful thought would need to be given in the selection of candidates to test. One attractive categorization to test is based on urban density or size or a combination of density and size. One or more categorizations could be based on demographic distribution, using a statistic such as median household income or proportion of population over or under a certain age. Another could be based on the degree of availability or use of public transit in the region.

A third extension would be to enhance technical aspects of the estimation and testing approach used in the study. Two desired technical enhancements are known at this time:

<u>Differences in model scale</u>: Model scale relates to the absolute size of the coefficients. Because utility functions have no natural units, all estimated coefficients are normalized relative to the residual error term (*e* in Chapter 3, equation 1). If there is more unexplained, residual "error" in the choices from one region relative to the others, then <u>all</u> estimated coefficients for that region will tend to be smaller (in absolute value) than for the other regions. If almost all of the difference coefficients for a region are in the same direction (for example, all of the coefficients are significant with the opposite sign from the base coefficient), then that is a strong indication of scale difference. Although it would be very difficult to estimate scale difference coefficients simultaneously with all of the other coefficients, it would be possible to "pre-scale" the data by some proportion prior to estimation to test whether this change makes the model appear much more transferable.

<u>Logsum coefficients</u>: It would be very difficult to apply the additive difference coefficient approach to structural model coefficients such as the logsum coefficients in nested logit (NL) models. NL is most typically used for mode choice models. In that case, different modes may

be available in different regions, which further complicates matters. Nesting is also used in other types of models to isolate “special” alternatives, such as the “stay at home” alternative in tour generation models, the “work at home” alternative in work location models, and the “home school” alternative in school location models. For nested models, the most feasible approach to test the transferability of logsum coefficients may be to estimate a completely separate model on each data set and simply compare the estimated coefficients. Another approach may be to estimate the nested logit models sequentially as two MNL models and test transferability separately for them.

Given that the evidence is mounting that location choice models are the least transferable, a fourth extension would be to develop and test the transferability of alternate model specifications for the location choice models. This could help identify the causes of the inferior transferability as well as techniques to improve it while preserving the quality of the specification. A breakthrough in this area could increase the potential for transferring entire model systems on a widespread basis and provide some guidelines for dealing with the location choice models in that context.

A fifth research extension would be to conduct model application tests for various scenarios of changed input data. For each scenario and each region, the tests would involve applying the entire AB model system several times, once using the region’s own estimation results, and once each for estimation results from several pooled data sets (one-state, two-state, and any other pooling categorizations that had been implemented, such as urban density as described above). The scenario sensitivity of the various specifications would be compared to evaluate the forecast quality of the various transfer categories. In doing this, the sensitivity of standard trip outputs would be measured, as well as tour-level and day-level outcomes relevant to the scenario being examined.

All of the above avenues of research are especially promising as extensions of the existing research because they can naturally build on data, software and techniques that are already in place.

Acknowledgments

The authors wish to recognize the significant contributions of others toward the completion of this research project. California Department of Transportation, Florida Department of Transportation, Sacramento Area Council of Governments, San Diego Association of Governments, Fresno County Council of Governments, San Joaquin County Council of Governments, Stanislaus County Council of Governments, Merced County Association of Governments, North Florida Transportation Planning Organization and Fehr & Peers provided the necessary data. Resource Systems Group helped condition the data for use in the project. David Ory (Metropolitan Transportation Commission), Joel Freedman (Parsons Brinckerhoff), Brian Gregor (Oregon Department of Transportation), Abdul Pinjari (University of South Florida) and Wu Sun (San Diego Association of Governments) served on the peer review panel and provided excellent feedback. Scott Smith (USDOT Volpe Center) provided technical support for the peer review meeting. Brian Gardner (FHWA System Planning and Analysis Team) oversaw the project, and Supin Yoder (FHWA System Planning and Analysis Team) served as project manager, working regularly with the authors to assure the project's success.

Appendix 1: Detailed NHTS data summaries

This appendix provides an initial analysis of the 2008-09 NHTS data from the California and Florida add-on samples, with particular focus on the six regions that will be compared in the model transferability tests. (This analysis combines results for Fresno with those for the northern three-county region of the San Joaquin Valley).

The data was first prepared by processing the household, person and trip records to form additional records on travel tours and person travel days. This appendix looks at the data at all of those different levels. A series of tables and graphs are provided, with description and comments.

Household variables

Tables 1-3: Observations by Day of Week and Region: These tables, like many other ones in this appendix, divide the columns by region:

Sacramento: The SACOG region, including Sacramento, Yolo, Yuba, Sutter, Placer and El Dorado counties.

San Joaquin: The Northern San Joaquin Valley, including Fresno, Stanislaus, San Joaquin and Madera counties. (All four are separate single-county MPO's.)

San Diego: The SANDAG region, including only San Diego county.

Jacksonville: The North Florida TPO region, including Duval, St. Johns, Nassau and Clay counties. (A sub-region of Florida DOT Region 2.)

Tampa: The Florida DOT Region 7, including Hillsborough, Piniellas, Pasco, Citrus and Hernando counties.

For comparison purposes, we also include two other "regions", the rest of California and the rest of Florida, respectively.

The "total" row of Table 1 shows that San Diego is the largest region with 6,000 surveyed households, while Tampa is the second largest with 2,500 households. The Jacksonville, Sacramento and San Joaquin regions all have in the range 1,000-1,300 households, while the rest of the state includes 12,000-13,000 households for both CA and FL. These relative totals are in line with the relative sizes of the actual populations in the regions, with the exception of SANDAG, which paid for a higher sampling rate in the CA add-on sample.

A comparison of the columns shows that roughly the same number of households had survey travel days on each of the seven days of the week. When the NHTS household weights are applied in Table 3, the total percent in each day becomes identical at 14.3%, indicating that the weights were calculated in part to even the sample across days of the week, although the fractions do not change very much relative to Table 2.

For the "longer term" models of auto ownership and usual workplace location, we can use all of the survey households in estimation. For models at the level of person-days, tours or trips, on the other hand, we will need to screen out the households with weekend travel days to focus only on weekdays. For the next series of tables, however, we include all households, because there is no systematic relationship between household types and days of the week (and there is no separate NHTS household weight for weekdays only).

Table 1: Day of Week by Region (unweighted)

Count

		Region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
dow	Sunday	188	164	860	155	373	1888	1811	5439
	Monday	186	124	897	189	369	1858	1695	5318
	Tuesday	182	154	813	177	346	1881	1767	5320
	Wednesday	184	144	856	154	341	1825	1706	5210
	Thursday	190	145	879	157	363	1870	1771	5375
	Friday	188	158	835	157	355	1779	1735	5207
	Saturday	193	149	862	151	367	1773	1745	5240
Total		1311	1038	6002	1140	2514	12874	12230	37109

Table 2: Day of Week by Region (unweighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
dow	Sunday	14.3%	15.8%	14.3%	13.6%	14.8%	14.7%	14.8%	14.7%
	Monday	14.2%	11.9%	14.9%	16.6%	14.7%	14.4%	13.9%	14.3%
	Tuesday	13.9%	14.8%	13.5%	15.5%	13.8%	14.6%	14.4%	14.3%
	Wednesday	14.0%	13.9%	14.3%	13.5%	13.6%	14.2%	13.9%	14.0%
	Thursday	14.5%	14.0%	14.6%	13.8%	14.4%	14.5%	14.5%	14.5%
	Friday	14.3%	15.2%	13.9%	13.8%	14.1%	13.8%	14.2%	14.0%
	Saturday	14.7%	14.4%	14.4%	13.2%	14.6%	13.8%	14.3%	14.1%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 3: Day of Week by Region (weighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
dow	Sunday	14.2%	13.8%	14.0%	14.5%	14.8%	14.4%	14.2%	14.3%
	Monday	14.5%	13.3%	14.4%	16.7%	14.1%	14.3%	14.1%	14.3%
	Tuesday	12.0%	14.3%	13.3%	16.1%	12.1%	14.6%	14.6%	14.3%
	Wednesday	15.5%	14.9%	13.4%	14.7%	14.2%	14.2%	14.3%	14.3%
	Thursday	14.3%	14.7%	14.9%	12.6%	14.9%	14.2%	14.3%	14.3%
	Friday	14.2%	14.8%	14.2%	13.9%	14.6%	14.3%	14.2%	14.3%
	Saturday	15.2%	14.1%	15.7%	11.6%	15.3%	14.1%	14.3%	14.3%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 4 and 5: Observations by Household Size and Region: Looking first at the unweighted numbers in Table 4, all regions look similar in terms of household size distribution, except for Tampa and the Rest of Florida, which contain more 1-person households and fewer large households. Jacksonville also contains fewer large households. In general, we can expect Florida to have more senior households, which tend to be smaller. The fractions change considerably with weighting in Table 5, indicating that there were different sampling and/or non-response rates by household size. The gap between states remains, however, with the Florida regions having more 1-person households and fewer with 4 or more people. San Joaquin has the fewest single person households.

Table 4: Household Size by Region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hhtot	1 person	23.8%	21.2%	23.8%	22.1%	28.2%	23.1%	25.7%	24.4%
	2 people	42.9%	42.1%	40.6%	48.3%	50.5%	40.6%	49.6%	44.6%
	3 people	14.3%	13.8%	14.6%	14.0%	10.7%	15.0%	11.7%	13.5%
	4 people	12.4%	11.3%	12.8%	10.3%	6.9%	13.3%	8.6%	11.1%
	5+ people	6.6%	11.7%	8.1%	5.3%	3.6%	8.0%	4.4%	6.5%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 5: Household Size by Region (weighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hhtot	1 person	29.0%	22.6%	25.9%	29.4%	31.0%	25.3%	28.8%	26.9%
	2 people	31.5%	29.7%	31.6%	34.2%	37.6%	29.7%	36.8%	32.5%
	3 people	15.6%	14.7%	15.9%	16.6%	14.6%	16.8%	14.8%	15.9%
	4 people	15.2%	14.9%	14.9%	12.1%	10.8%	16.0%	12.6%	14.5%
	5+ people	8.6%	18.2%	11.6%	7.7%	6.0%	12.1%	7.0%	10.3%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 6 and 7: Observations by Workers in Household and Region: Tampa and the Rest of FL have more 0-worker HH than the other regions, although the weighting in Table 7 reduces the percent of 0-workers households considerably. (Survey response rates tend to be highest among senior and non-working households, and the weighting adjusts for that, among other things.) San Joaquin has more 0-worker households than the other CA regions.

Table 6: Workers in Household by Region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
workers	No workers	38.9%	39.1%	36.6%	40.2%	51.7%	35.8%	46.9%	41.0%
	1 worker	32.6%	35.1%	36.3%	36.3%	29.6%	36.7%	31.9%	34.4%
	2 workers	24.5%	22.2%	22.7%	21.0%	16.8%	23.2%	18.5%	21.1%
	3+ workers	4.0%	3.7%	4.3%	2.5%	1.9%	4.3%	2.7%	3.5%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 7: Workers in Household by Region (weighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
workers	No workers	27.8%	30.0%	25.3%	28.4%	32.6%	25.1%	31.8%	27.8%
	1 worker	41.2%	40.7%	44.5%	45.8%	41.4%	44.6%	42.6%	43.6%
	2 workers	26.0%	23.6%	24.9%	22.5%	22.9%	25.0%	21.7%	23.8%
	3+ workers	5.0%	5.7%	5.3%	3.3%	3.2%	5.4%	3.9%	4.8%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 8 and 9: Observations by Household Type and Region: Here we created a 10-category indicator of household type, based on size, worker status, and presence of children of different ages. One striking number is that the percent of single worker households is almost twice as high in the weighted table as in the unweighted table, while the number of two person senior households is less than half as high. The weighting also increases the percent of households with children. There are very few single student households, even after weighting. (Presumably the weighting does not reflect group quarters such as dormitories.) Tampa and the Rest of FL region have more senior households than the other regions—particularly with 2 persons. Jacksonville, however, has fewer senior HH than the other FL regions. In Table 9, San Joaquin has the highest percentage of "traditional" family households with children under 16. In fact, compared to Tampa, San Joaquin has over twice as high a percent of households with children under age 5.

Table 8: Household Type by Region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hhtype	single worker	7.6%	6.4%	7.7%	7.2%	6.2%	8.0%	7.1%	7.5%
	single senior	11.7%	11.6%	12.5%	11.7%	17.5%	11.3%	14.8%	13.1%
	single non-worker	4.3%	3.3%	3.3%	3.1%	4.5%	3.6%	3.6%	3.6%
	single student	.2%		.2%	.2%	.0%	.2%	.1%	.1%
	two workers	11.7%	9.0%	9.7%	9.9%	8.9%	9.6%	8.7%	9.4%
	two seniors	11.1%	11.3%	10.0%	11.5%	17.0%	10.1%	16.6%	12.8%
	other two person/no kids	18.8%	20.4%	19.4%	25.2%	23.3%	19.0%	22.6%	20.8%
	three+ person/no kids	6.1%	8.2%	7.0%	6.3%	5.3%	7.6%	6.2%	6.8%
	single adult with kids	2.1%	2.2%	1.9%	1.9%	1.4%	2.2%	1.7%	1.9%
	adults with kids under 5	8.2%	10.6%	9.8%	7.4%	4.1%	8.7%	5.2%	7.4%
	adults with kids under 16	13.0%	13.4%	13.8%	11.6%	8.4%	14.4%	9.8%	12.2%
	adults with kids all 16+	5.3%	3.8%	4.7%	4.1%	3.3%	5.3%	3.4%	4.4%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 9: Household Type by Region (weighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hhtype	single worker	15.4%	10.3%	14.1%	15.9%	15.8%	13.9%	14.3%	14.1%
	single senior	9.2%	8.6%	8.8%	9.3%	11.3%	7.9%	11.3%	9.2%
	single non-worker	4.1%	3.7%	2.8%	4.0%	3.8%	3.3%	3.0%	3.2%
	single student	.4%		.2%	.2%	.1%	.2%	.1%	.2%
	two workers	11.1%	7.5%	9.6%	10.0%	11.1%	8.8%	8.7%	9.0%
	two seniors	5.4%	5.9%	5.2%	4.4%	8.4%	4.6%	8.5%	6.1%
	other two person/no kids	13.5%	15.1%	14.8%	16.9%	16.8%	14.2%	17.3%	15.3%
	three+ person/no kids	6.3%	9.6%	7.6%	7.3%	6.7%	8.6%	7.7%	8.1%
	single adult with kids	3.4%	2.9%	2.8%	3.6%	2.3%	3.3%	3.0%	3.1%
	adults with kids under 5	11.5%	15.7%	13.8%	11.6%	7.6%	12.4%	8.6%	11.2%
	adults with kids under 16	14.3%	17.0%	15.0%	12.8%	12.1%	16.8%	13.5%	15.3%
	adults with kids all 16+	5.6%	3.8%	5.2%	4.0%	4.1%	6.0%	4.0%	5.1%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 10 and 11: Observations by Own vs. Rent and Region: The most striking thing here is not the difference between regions, but that the percent of renting households is more than twice as high after weighting is applied in Table 11. Renting is noticeably less common in FL than in CA, although the percentages for Sacramento and Jacksonville are similar, in the middle range.

Table 10: Own/Rent by Region (unweighted)

% within region

		Region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
hownrent	own	84.3%	81.0%	79.4%	88.8%	90.4%	79.8%	90.3%	84.4%
	rent	15.7%	19.0%	20.6%	11.2%	9.6%	20.2%	9.7%	15.6%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 11: Own/Rent by Region (weighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
hownrent	own	64.2%	59.7%	56.4%	65.4%	70.0%	56.3%	70.0%	61.7%
	rent	35.8%	40.3%	43.6%	34.6%	30.0%	43.7%	30.0%	38.3%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 12 and 13: Observations by Residence Type and Region: The fraction of households living in single family detached housing is highest in Sacramento, San Joaquin and lowest in San Diego. Weighting decreases this fraction for all regions. Overall, this data item looks suspect, and will not be tested in the choice models in any case, as it is rarely available for travel demand forecasting models.

Table 12: Residence Type by Region (unweighted)

% within region

		Region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hrestype	single family detached	82.0%	82.6%	69.9%	78.1%	69.8%	73.0%	70.0%	72.0%
	duplex/rowhouse	14.4%	13.1%	25.8%	13.4%	19.0%	22.7%	20.9%	21.5%
	apartment/condo	3.4%	4.2%	4.0%	8.2%	10.9%	3.9%	8.6%	6.1%
	mobile home	.1%		.0%	.1%	.0%	.1%	.1%	.1%
	dorm/rented room						.0%		.0%
	Other			.1%			.1%	.1%	.1%
	missing	.1%	.1%	.1%	.2%	.2%	.2%	.2%	.2%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 13: Residence Type by Region (weighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hrestype	single family detached	71.3%	72.8%	56.2%	67.6%	63.0%	59.1%	61.3%	61.1%
	duplex/rowhouse	25.5%	23.2%	39.8%	26.0%	29.0%	37.4%	32.1%	34.1%
	apartment/condo	3.0%	3.9%	3.5%	6.4%	7.8%	3.2%	6.3%	4.5%
	mobile home	.0%		.1%	.1%	.0%	.0%	.1%	.0%
	dorm/rented room						.0%		.0%
	other			.3%			.2%	.1%	.1%
	missing	.1%	.0%	.1%	.1%	.1%	.2%	.2%	.2%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 14 and 15: Observations by Car Ownership and Region: This is the dependent variable for the auto ownership models. Weighting increases the number of 0-vehicle and 1-vehicle households substantially. The Florida regions have the most 1-car households—probably many of them senior households. Sacramento has the fewest 0-vehicle HH, and San Joaquin the most, although the difference is not large.

Table 14: Car Ownership by Region (unweighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
cars	no cars	3.4%	5.2%	4.9%	4.0%	5.3%	4.9%	4.5%	4.7%
	1 car	25.7%	27.7%	26.8%	27.3%	39.3%	26.4%	34.9%	30.2%
	2 cars	40.3%	39.2%	39.9%	43.3%	38.4%	40.6%	40.4%	40.3%
	3 cars	19.6%	18.3%	17.5%	17.8%	12.7%	17.9%	14.3%	16.4%
	4+ cars	11.0%	9.5%	10.9%	7.5%	4.3%	10.2%	6.0%	8.4%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 15: Car Ownership by Region (weighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
cars	no cars	5.4%	7.9%	6.3%	7.2%	6.3%	8.1%	6.7%	7.4%
	1 car	32.5%	32.6%	32.4%	36.9%	42.1%	31.5%	40.1%	34.9%
	2 cars	38.0%	35.4%	39.3%	39.5%	38.3%	37.5%	37.9%	37.7%
	3 cars	15.0%	15.8%	13.6%	10.2%	9.7%	14.4%	10.9%	13.1%
	4+ cars	9.1%	8.2%	8.5%	6.1%	3.5%	8.5%	4.4%	7.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 16 and 17: Observations by Income Group and Region: Weighting shifts the incomes lower somewhat. Only 8 percent of household have missing income data, which is a fairly low percentage. San Joaquin has the highest percentage of low-income households (almost 30% under $25,000 in Table 17), while Sacramento and San Diego have the highest percentages of high-income HH (about 20% over $100,000, similar to the Rest of CA). The lower income in San Joaquin may help to explain its higher percentage of 0-vehicle HH, mentioned above.

Table 16: Income Group by Region (unweighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
inccat	$0 - 24,999	17.1%	22.4%	16.6%	18.2%	24.1%	17.6%	22.1%	19.5%
	$25,000 - 49,999	20.8%	26.7%	22.4%	23.3%	28.5%	21.4%	26.3%	23.8%
	$50,000 - 99,999	32.3%	27.3%	28.3%	29.9%	25.2%	27.1%	27.0%	27.4%
	$100,000 and over	23.6%	16.8%	25.3%	20.3%	13.0%	26.4%	15.1%	21.1%
	missing	6.2%	6.9%	7.4%	8.3%	9.1%	7.4%	9.5%	8.2%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 17: Income Group by Region (weighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
inccat	$0 - 24,999	23.0%	29.2%	22.6%	22.5%	24.6%	24.3%	25.4%	24.6%
	$25,000 - 49,999	22.7%	26.6%	24.6%	27.2%	29.7%	22.6%	26.5%	24.5%
	$50,000 - 99,999	30.3%	24.1%	26.9%	28.4%	25.3%	25.6%	26.2%	26.0%
	$100,000 and over	19.5%	13.6%	20.3%	15.3%	13.3%	21.5%	13.3%	18.1%
	missing	4.5%	6.4%	5.6%	6.7%	7.1%	6.0%	8.6%	6.8%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 18 and 19: Regional Sample Means on Household Variables: The relative mean values indicate several of the same things found by looking at the relative distributions:

Sacramento and San Diego have the highest incomes, and San Joaquin the lowest.

Sacramento has the highest car ownership, and Florida is lower than California, particularly Tampa.

Tampa and the rest of FL (but not Jacksonville) have the fewest workers per HH and the most seniors per HH.

San Joaquin has the most children per HH in all age categories, and Tampa the least.

Table 18: Regional Sample Means (unweighted)

StatisticsReportMean

	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
Car ownership	2.16	2.05	2.10	2.00	1.74	2.08	1.85	1.99
Income (non-missing)	66,360	56,527	66,690	62,7666	52,380	67,013	55,723	61,859
Household size	2.38	2.59	2.45	2.30	2.10	2.47	2.19	2.34
Full time workers	.71	.70	.72	.69	.53	.74	.60	.67
Part time workers	.22	.20	.21	.16	.15	.21	.16	.19
University students	.04	.05	.04	.03	.02	.04	.02	.03
Other adults age 65+	.47	.49	.48	.51	.69	.46	.66	.55
Other adults age <65	.47	.54	.48	.50	.42	.50	.42	.46
Children age 16+	.07	.07	.07	.06	.05	.08	.05	.06
Children age 5-15	.30	.40	.32	.25	.18	.32	.21	.27
Children age 0-4	.11	.15	.13	.10	.06	.12	.07	.10

Table 19: Regional Sample Means (weighted)

Statistics=Mean

	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
Car ownership	1.95	1.89	1.92	1.73	1.63	1.89	1.68	1.82
Income (non-missing)	59,610	50,560	59,586	55,653	52,007	59,731	52,209	56,720
Household size	2.48	2.94	2.62	2.39	2.28	2.67	2.36	2.55
Full time workers	.81	.79	.83	.80	.75	.85	.75	.81
Part time workers	.27	.25	.25	.18	.20	.24	.21	.23
University students	.04	.05	.05	.04	.02	.06	.03	.05
Other adults age 65+	.31	.33	.31	.30	.42	.30	.44	.35
Other adults age <65	.45	.60	.51	.50	.41	.55	.43	.50
Children age 16+	.08	.10	.08	.08	.07	.10	.06	.08
Children age 5-15	.36	.58	.40	.31	.29	.42	.31	.38
Children age 0-4	.17	.22	.18	.17	.11	.17	.12	.15

Person Variables

Tables 20-22: Observations by Person Type and Region: Now, we look at observations at the person level rather than the household level. Person type is an 8-category variable, created based on age, student status, and employment status. It is one of the key variables used for segmentation in the travel models.

Tables 20 and 21 show the same results we would expect from the preceding discussion, with Tampa having the most seniors and seniors being weighted down in all of the regions, and San Joaquin having the most children and Tampa the fewest. NHTS also provides a separate person weight, which corrects for that fact that there are missing person records in some households. Because children under age 5 did not fill in travel diaries, these children were removed from the NHTS weighting process (and later data sets), so have person weights of 0. Comparing the weights between Tables 21 and 22, all person types were weighted upward somewhat in the person-level corrections except for adults age 65+ (seniors), who are weighted downward even further (because non-response rates were lowest in that group).

NOTE: NHTS provided us with age information on all of the persons who are missing from the person (and trip) data files because they did not provide travel-day data. That additional information was used to correct the household composition and type variables reported above. However, they did not provide employment status or student status data on those missing persons, so it is not possible to always have a correct count of the number of workers or students in each household.

Table 20: Person Type by Region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
pptyp	full time worker	32.2%	29.8%	32.4%	32.7%	27.2%	32.8%	29.4%	31.2%
	part time worker	9.8%	8.3%	9.3%	7.5%	7.7%	9.4%	8.0%	8.8%
	other adult age 65+	20.4%	19.9%	20.2%	22.9%	34.0%	19.6%	31.0%	24.3%
	other adult age<65	16.3%	16.6%	15.7%	18.4%	17.5%	15.9%	15.8%	16.1%
	college student	1.8%	1.9%	1.8%	1.3%	.9%	2.0%	1.2%	1.6%
	child age 16+	2.7%	2.2%	2.5%	2.0%	2.1%	2.8%	1.9%	2.4%
	child age 5-15	11.8%	14.7%	12.3%	10.3%	7.8%	12.2%	9.3%	11.1%
	child age 0-4	5.1%	6.5%	5.8%	5.0%	2.9%	5.2%	3.4%	4.6%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 21: Person Type by Region (weighted by Household Weight)

% within region

		Region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
pptyp	full time worker	35.4%	30.8%	35.4%	36.7%	35.7%	35.7%	35.1%	35.4%
	part time worker	11.7%	9.8%	10.8%	8.5%	9.6%	10.3%	9.7%	10.1%
	other adult age 65+	13.2%	12.0%	12.5%	12.9%	19.3%	11.7%	19.2%	14.3%
	other adult age<65	14.4%	15.7%	14.4%	17.0%	14.5%	15.3%	13.9%	14.9%
	college student	2.0%	2.0%	2.3%	1.7%	1.2%	2.4%	1.5%	2.0%
	child age 16+	2.7%	2.7%	2.6%	3.0%	2.6%	3.1%	2.1%	2.8%
	child age 5-15	13.1%	18.4%	14.2%	12.4%	11.8%	14.3%	13.2%	13.9%
	child age 0-4	7.4%	8.5%	7.8%	7.8%	5.3%	7.1%	5.4%	6.7%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 22: Person Type by Region (weighted by Person Weight)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
pptyp	full time worker	39.2%	34.7%	39.1%	36.8%	37.5%	39.1%	37.5%	38.4%
	part time worker	13.5%	11.3%	12.2%	9.2%	9.5%	11.2%	9.8%	10.8%
	other adult age 65+	10.7%	9.5%	9.7%	9.0%	14.8%	9.4%	14.4%	11.1%
	other adult age<65	15.1%	16.8%	15.5%	16.8%	14.9%	16.4%	14.4%	15.7%
	college student	3.5%	3.5%	4.1%	2.9%	2.3%	3.7%	2.6%	3.3%
	child age 16+	3.3%	3.2%	3.3%	3.6%	3.4%	3.8%	2.6%	3.4%
	child age 5-15	14.7%	20.9%	16.0%	13.7%	13.3%	16.3%	13.8%	15.5%
	child age 0-4								
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 23-24: Observations by Gender and Region: There are more males in the sample than females in all regions, although most of that difference disappears with person weighting in Table 24. Children under age 5 in the CA data are missing gender information, but they also have a person weight of 0, so are not included in the weighted table.

Table 23: Gender by Region (unweighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
pgend	missing	5.1%*	6.5%*	5.8%*			5.2%*		3.2%*
	male	44.4%	44.2%	44.2%	45.6%	46.2%	44.8%	46.3%	45.2%
	female	50.5%	49.2%	50.0%	54.4%	53.8%	50.0%	53.7%	51.5%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

* all cases of missing data on gender are for children under age 5, for whom no travel diary data was collected

Table 24: Gender by Region (weighted by Person Weight)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
pgend	male	49.7%	50.4%	50.2%	49.1%	49.1%	49.9%	49.2%	49.7%
	female	50.3%	49.6%	49.8%	50.9%	50.9%	50.1%	50.8%	50.3%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Person-Day Travel Variables

We now move on to look at summary variables of the respondents' travel on the survey diary day. Some models in Activity-Based model systems operate at the person-day level, so it is useful to look at the data at this level. Because we are looking at this data mainly in terms of estimating the models and less to find differences between the regions, all of the tables in this section are unweighted.

Table 25: Mean number of tours/person by purpose and day of week: This table shows why it is not a good idea to group data from weekends and weekdays together in a model of travel patterns. Saturday and Sunday are clearly different from the other days, with only about 20% as many work or school or medical tours, and also fewer serve passenger (pick up or drop off) tours, but more tours for all other purposes (particularly on Sunday for "personal business", which includes going to church). In contrast, there are no large differences between the tour rates for weekdays Monday through Friday.

Table 25: Mean number of tours/person by purpose and day of week (unweighted)

Statistics=Mean

	dow							
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Total
Home-based work	.06	.30	.32	.32	.31	.30	.09	.24
Home-based school	.02	.11	.13	.13	.12	.10	.01	.09
Home-based serve passenger	.08	.14	.16	.15	.14	.15	.09	.13
Home-based personal-business	.25	.09	.11	.11	.10	.09	.12	.12
Home-based shopping	.29	.27	.26	.25	.24	.28	.40	.28
Home-based meal	.11	.06	.07	.07	.08	.10	.13	.09
Home-based social visit	.13	.07	.07	.08	.10	.09	.15	.10
Home-based recreation	.17	.17	.16	.16	.15	.17	.21	.17
Home-based medical	.01	.07	.07	.07	.06	.05	.01	.05
Total home-based tours	1.13	1.28	1.34	1.33	1.30	1.34	1.24	1.28
Work-based subtours	.00	.05	.05	.06	.05	.05	.01	.04

For the remaining tables in this appendix, we have excluded all weekend travel days, and used only the weekday travel information that will be used for model estimation.

Tables 26 and 27: Weekday travel day begins and ends at home by Region and Day of Week: Most applied models of day activity patterns assume that a person begins and ends their day at home. Yet, in reality that is not always the case. In the NHTS data, about 93% of the travel days begin and end at home. There is very little difference by region, with the lowest fraction is in San Joaquin (91.7%) and the highest in Jacksonville (93.4%). The differences are somewhat larger by day of week, with Friday having a larger percentage of travel days ending away from home (e.g. people going away for the weekend).

Table 26: Weekday travel day begins and ends at home by Region (unweighted)

% within region

		Region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
behome	does not begin or end at home	2.1%	2.4%	1.7%	2.1%	1.6%	2.1%	2.2%	2.1%
	begins away from home	1.8%	2.7%	1.8%	1.3%	1.4%	2.1%	1.5%	1.8%
	ends away from home	3.4%	3.2%	3.4%	3.2%	3.9%	3.1%	2.9%	3.2%
	begins and ends at home	92.6%	91.7%	93.0%	93.4%	93.1%	92.6%	93.3%	92.9%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 27: Weekday travel day begins and ends at home by Day of Week (unweighted)

% within dow

		Dow					Total
		Monday	Tuesday	Wednesday	Thursday	Friday	
behome	does not begin or end at home	1.9%	2.1%	2.1%	1.8%	2.4%	2.1%
	begins away from home	2.1%	1.6%	1.5%	1.7%	2.3%	1.8%
	ends away from home	2.5%	2.5%	2.6%	3.0%	5.3%	3.2%
	begins and ends at home	93.5%	93.8%	93.8%	93.5%	90.1%	92.9%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 28: Person-days with weekday travel days, begin and end at home: For the remaining tables in this section, we include only travel days that begin and end at home, as these will be the sample for the person-day-level models that we test. This table shows the unweighted sample size for those models by person type and region. As is typical, the samples for college students and high school students are the "thinnest", so we may not be able to tell much about the transferability of models for those groups.

Table 28: Person-days with full weekday travel diaries, begin and end at home (unweighted)

Count

		Region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
pptyp	full time worker	621	447	2848	531	844	6199	4714	16204
	part time worker	179	139	827	117	244	1813	1314	4633
	other adult age 65+	390	310	1854	399	1132	3838	5204	13127
	other adult age<65	304	252	1417	294	575	3056	2615	8513
	college student	29	30	153	20	20	377	186	815
	child age 16+	50	34	225	34	71	520	312	1246
	child age 5-15	215	229	1100	178	235	2337	1464	5758
Total		1788	1441	8424	1573	3121	18140	15809	50296

Tables 29 and 30: Weekday travel days by response type and region and person type: Not all travel diary information is provided from a diary by the person who made the trips. Sometimes travel is reported by proxy by another household member, and sometimes the data is reported only by recall when the person did not use travel diary. Both of these differences can affect the completeness and quality of the travel data. The tables show that there are 54% of cases where the person used the diary and reported their own travel to the interviewer. There are 18% of cases where the person used the diary, and then another person reported that data by proxy, and 19% of cases where a person reported their own travel by recall (and not from the diary). The most suspect cases are the 9% where data is reported by recall and by proxy—one person reporting from memory what another person did on the travel day. Table 30 shows the percentages by person type. Children under 16 generally did not fill out travel diaries, and high-school children used the diary in about one half of cases. For both groups, in about one third of cases their travel was reported by proxy. There is less difference among the adult types. Seniors used proxy more often, perhaps because they spent the travel day together so that they knew what the other person did. The differences by region in Table 29 are less pronounced. San Joaquin has slightly higher percentages on all the proxy and recall categories—probably because this region has the most children.

Table 29: Response type by Region (unweighted)

% within region

		Region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
rtype	used diary, no proxy	53.7%	49.1%	55.0%	55.4%	58.4%	52.1%	55.5%	54.1%
	used diary, by proxy	17.7%	18.9%	16.6%	17.5%	17.4%	18.1%	18.2%	17.8%
	did not use diary, no proxy	19.3%	21.2%	19.3%	19.1%	16.2%	20.1%	17.5%	18.9%
	did not use diary, by proxy	9.3%	10.8%	9.2%	7.9%	8.0%	9.6%	8.8%	9.2%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 30: Response type by Person Type (unweighted)

% within pptyp

		pptyp							Total
		full time worker	part time worker	other adult age 65+	other adult age<65	college student	child age 16+	child age 5-15	
rtype	used diary, no proxy	65.3%	69.1%	57.4%	60.0%	48.7%	30.5%	.3%	54.1%
	used diary, by proxy	15.7%	16.4%	24.3%	24.2%	20.2%	19.0%	.2%	17.8%
	did not use diary, no proxy	15.1%	11.1%	9.4%	9.5%	24.5%	35.0%	67.2%	18.9%
	did not use diary, by proxy	4.0%	3.4%	8.9%	6.3%	6.5%	15.5%	32.3%	9.2%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 31: Mean reported tours and stops per person-day by response type (full time workers only): One interesting question is whether the use of proxy and recall influences the data quality. In Table 31, we control by person-type by only including the person-days of full time workers. The results show a clear difference in the number of tours and stops reported for each response type, with fewer trips when reported by proxy, particularly in combination with not using the diary. The causality of this finding is not obvious—it could be that people with simpler travel patterns (or who stayed home all day) had less need to use the diary, or that it was simpler to report them by proxy. However, there were very few full time workers who stayed home all day on the diary day, so it is likely that there is also some effect of the response type on the completeness of the data. We will control for these differences when estimating the models.

Table 31: Mean reported tours and stops per person-day by response type (full time workers only) (unweighted)

Statistics=Mean,pptyp=full time worker

	rtype				
	used diary, no proxy	used diary, by proxy	did not use diary, no proxy	did not use diary, by proxy	Total
Home-based work	.85	.71	.86	.61	.82
Home-based school	.01	.01	.01	.00	.01
Home-based serve passenger	.13	.12	.09	.06	.12
Home-based personal-business	.08	.06	.05	.04	.07
Home-based shopping	.18	.14	.13	.09	.17
Home-based meal	.07	.05	.05	.03	.06
Home-based social visit	.05	.04	.05	.03	.05
Home-based recreation	.14	.10	.11	.05	.13
Home-based medical	.03	.02	.02	.02	.03
Total home-based tours	1.55	1.26	1.38	.93	1.45
Work-based subtours	.18	.10	.14	.04	.16
Intermediate stops on tours	1.49	1.08	1.06	.48	1.32

Table 32 and 33: Number of home-based tours during travel day by person type and region: These tables show the distribution of the number of home-based trip chains, each with two or more trips, are made during the day. People with 0 tours did not leave home during the diary day. By person type, the number of "stay at home" days is highest for seniors (25%) and other non-working adults (22%) and lowest for full-time workers (5%). Multiple tour making (2+ tours in the day) is highest for part time workers and college students. The differences by region in Table 33 are much smaller. In general, the percentage of "stay at home" days is somewhat higher for FL than for CA.

Table 32: Home-based tours during travel day by person type (unweighted)

% within pptyp

		pptyp							
		full time worker	part time worker	other adult age 65+	other adult age<65	college student	child age 16+	child age 5-15	Total
Hbtours	0	5.4%	8.3%	25.1%	21.8%	9.1%	10.1%	10.9%	14.4%
	1	57.2%	42.6%	43.4%	38.1%	47.0%	53.0%	57.2%	48.7%
	2	27.5%	30.4%	21.7%	23.2%	30.9%	28.3%	26.2%	25.5%
	3	7.2%	12.8%	7.2%	11.2%	9.9%	7.6%	5.0%	8.2%
	4	2.0%	4.6%	2.0%	3.7%	2.7%	.6%	.7%	2.4%
	5	.5%	1.0%	.4%	1.4%	.4%	.3%	.0%	.6%
	6	.1%	.2%	.1%	.4%				.2%
	7	.0%	.1%	.0%	.1%				.0%
	8	.0%	.0%		.0%				.0%
	10			.0%					.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 33: Home-based tours during travel day by region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
hbtours	0	13.1%	14.3%	13.3%	15.5%	16.0%	13.7%	15.5%	14.4%
	1	48.2%	50.0%	47.9%	48.0%	47.8%	49.0%	49.1%	48.7%
	2	26.2%	24.1%	27.3%	24.9%	24.4%	25.3%	24.9%	25.5%
	3	8.6%	7.9%	8.4%	9.0%	8.5%	8.5%	7.6%	8.2%
	4	3.2%	2.8%	2.3%	2.0%	2.4%	2.5%	2.1%	2.4%
	5	.6%	.5%	.7%	.5%	.6%	.6%	.5%	.6%
	6	.1%	.2%	.1%	.1%	.1%	.2%	.2%	.2%
	7		.1%	.0%		.2%	.0%	.1%	.0%
	8			.0%			.0%	.0%	.0%
	10							.0%	.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 34 and Figure 1: Mean reported tours and stops per person-day by person type: The differences in tours by purpose by the person types are logical, with students mainly making school tours, workers mainly making work tours, and other adults mainly making shopping and other discretionary tours. Serve passenger tours are highest for part time workers and younger non-working adults. Workers tend to make the most "extra" intermediate stops on tours (1.32 per day), while children make the fewest stops.

Table 34: Mean reported tours and stops per person-day by person type (unweighted)

Statistics=Mean

	pptyp							
	full time worker	part time worker	other adult age 65+	other adult age<65	college student	child age 16+	child age 5-15	Total
Home-based work	.82	.55	.01	.03	.11	.08	.00	.33
Home-based school	.01	.04	.01	.03	.64	.72	.69	.12
Home-based serve passenger	.12	.25	.09	.33	.17	.05	.08	.15
Home-based personal-business	.07	.12	.14	.14	.08	.06	.06	.10
Home-based shopping	.17	.31	.40	.40	.20	.12	.10	.27
Home-based meal	.06	.08	.11	.09	.07	.05	.05	.08
Home-based social visit	.05	.08	.11	.11	.09	.10	.10	.09
Home-based recreation	.13	.18	.19	.20	.13	.17	.17	.17
Home-based medical	.03	.07	.13	.10	.03	.03	.02	.07
Total home-based tours	1.45	1.68	1.20	1.43	1.51	1.37	1.27	1.38
Work-based subtours	.16	.04	.00	.00	.00	.00	.00	.05
Intermediate stops on tours	1.32	1.32	.92	1.07	.91	.72	.64	1.07

Figure 1: Mean reported tours and stops per person-day by person type (unweighted)

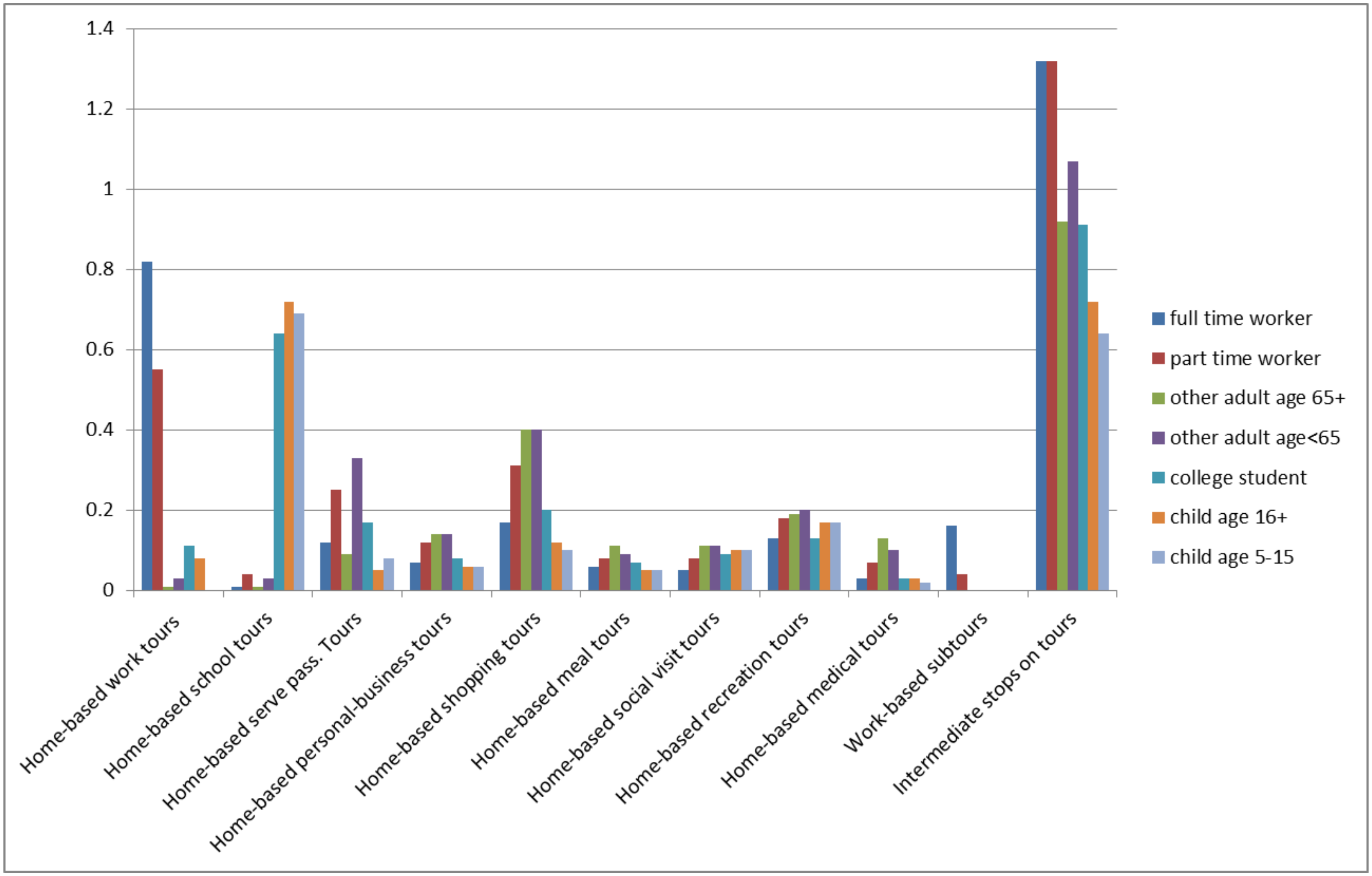


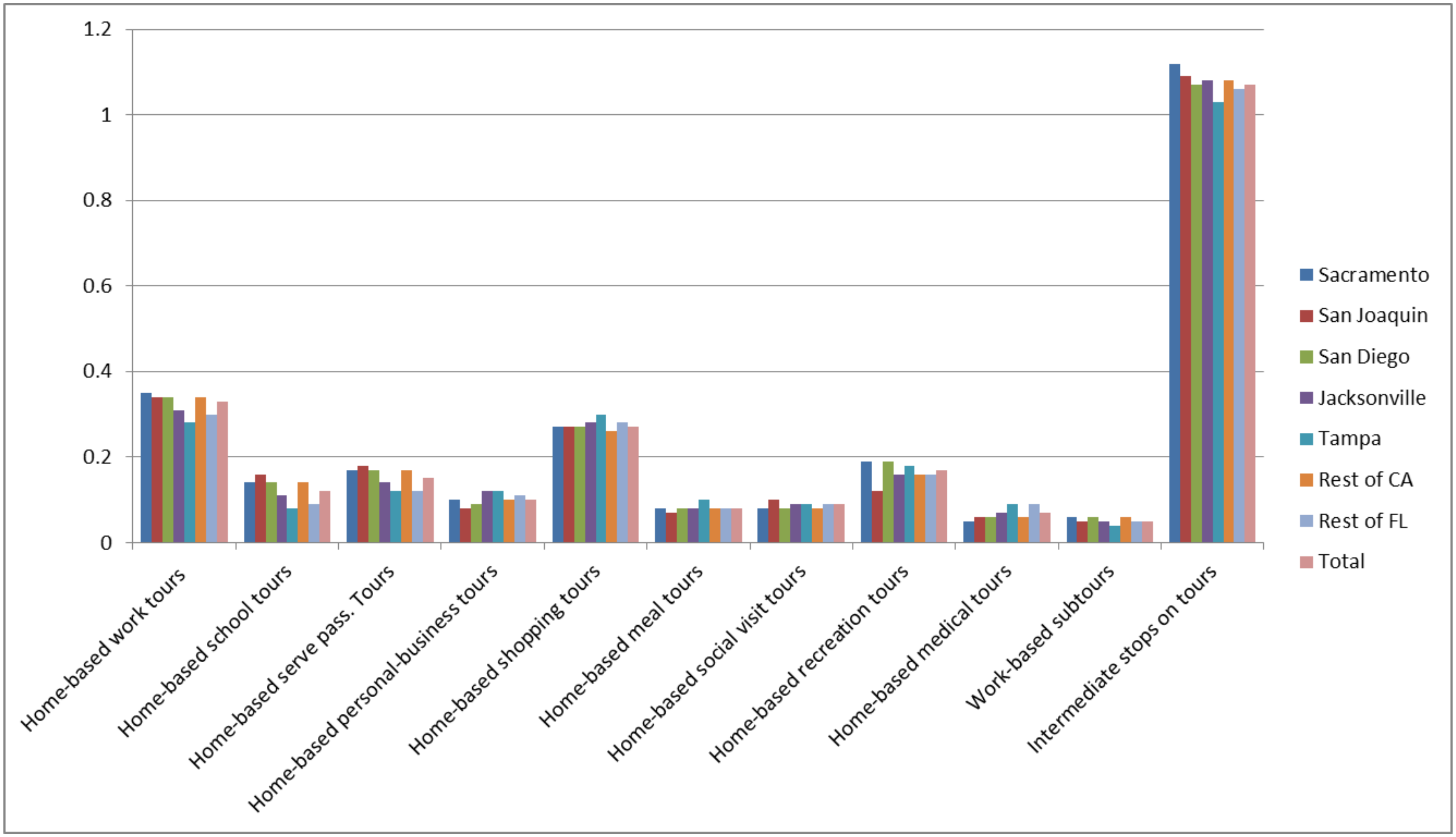
Table 35 and Figure 2: Mean reported tours and stops per person-day by region: The differences across regions are much smaller than across person types, and are likely due to differences in person-type distribution across the regions—as indicated by more school tours in San Joaquin (by children) and more shopping tours in Tampa (by seniors).

Table 35: Mean reported tours and stops per person-day by region (unweighted)

Statistics=Mean

	pptyp							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
Home-based work	.35	.34	.34	.31	.28	.34	.30	.33
Home-based school	.14	.16	.14	.11	.08	.14	.09	.12
Home-based serve passenger	.17	.18	.17	.14	.12	.17	.12	.15
Home-based personal-business	.10	.08	.09	.12	.12	.10	.11	.10
Home-based shopping	.27	.27	.27	.28	.30	.26	.28	.27
Home-based meal	.08	.07	.08	.08	.10	.08	.08	.08
Home-based social visit	.08	.10	.08	.09	.09	.08	.09	.09
Home-based recreation	.19	.12	.19	.16	.18	.16	.16	.17
Home-based medical	.05	.06	.06	.07	.09	.06	.09	.07
Total home-based tours	1.43	1.38	1.41	1.36	1.36	1.40	1.34	1.38
Work-based subtours	.06	.05	.06	.05	.04	.06	.05	.05
Intermediate stops on tours	1.12	1.09	1.07	1.08	1.03	1.08	1.06	1.07

Figure 2: Mean reported tours and stops per person-day by region (unweighted)



Tour variables

Tables 36 and 37: Weekday tour main purpose by region: Table 36 shows the absolute number of weekday tour records by main purpose in the different regions. These will be used as the basis for tour-level model estimation. The lowest number of tours is for medical visits in the smaller regions, while the largest numbers of home based tours are for work and shopping. The distribution of tours by purpose shown in Table 37 shows very little difference across the regions. One noticeable difference is a low percentage of recreation tours and a high percentage of school tours in San Joaquin, both perhaps due to a higher percentage of children.

Table 36: Weekday Tour Purpose by Region (unweighted)

Count

		Region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
pdpurp	Work	656	511	2993	519	918	6608	5051	17256
	School	248	233	1210	177	255	2639	1463	6225
	serve passenger	313	287	1499	227	398	3285	2043	8052
	personal business	186	120	819	198	389	1953	1856	5521
	shopping	525	420	2381	460	993	5037	4760	14576
	Meal	190	135	864	174	350	1896	1643	5252
	social visit	154	151	727	143	303	1458	1544	4480
	recreation	350	180	1658	264	575	3135	2666	8828
	Medical	85	88	506	118	292	1152	1427	3668
Total		2707	2125	12657	2280	4473	27163	22453	73858

Table 37: Weekday Tour Purpose by Region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
pdpurp	Work	24.2%	24.0%	23.6%	22.8%	20.5%	24.3%	22.5%	23.4%
	School	9.2%	11.0%	9.6%	7.8%	5.7%	9.7%	6.5%	8.4%
	serve passenger	11.6%	13.5%	11.8%	10.0%	8.9%	12.1%	9.1%	10.9%
	personal business	6.9%	5.6%	6.5%	8.7%	8.7%	7.2%	8.3%	7.5%
	shopping	19.4%	19.8%	18.8%	20.2%	22.2%	18.5%	21.2%	19.7%
	Meal	7.0%	6.4%	6.8%	7.6%	7.8%	7.0%	7.3%	7.1%
	social visit	5.7%	7.1%	5.7%	6.3%	6.8%	5.4%	6.9%	6.1%
	recreation	12.9%	8.5%	13.1%	11.6%	12.9%	11.5%	11.9%	12.0%
	Medical	3.1%	4.1%	4.0%	5.2%	6.5%	4.2%	6.4%	5.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 38-40: Weekday tour main mode by region: Here we look at the main mode used on a tour. The main mode is specified according to a hierarchy, with walk the lowest and "other" the highest, with the order of hierarchy as shown in the tables, from bottom to top. So, a walk trip can be part of any tour, but an "other mode" trip can only be part of an "other mode" tour. There are very few transit tours, particularly for drive-access to transit, and particularly in the regions other than San Diego. There are very few bicycle or school bus tours in those regions as well. Overall, almost 85% of tours are by car, roughly half of those SOV and half HOV. There is a somewhat higher percentage of walk trips in CA and a higher percentage of school bus trips in FL, but overall the differences are quite small.

Table 40 shows that most transit tours use local bus, with some by commuter bus, commuter rail and light rail. The accuracy of the NHTS submode coding is questionable, as there should be many more light rail tours relative to commuter rail in the Sacramento region.

It is clear that we will not be able to test transferability of models across regions with respect to transit mode choice, and probably not for bike and school bus mode choice. We will be able to say more about walk mode choice and car occupancy choice.

Table 38: Weekday Tour Main Mode by Region (unweighted)

Count

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
tmodetp	Walk	300	228	1418	213	441	3259	2008	7867
	Bike	56	25	137	39	53	378	307	995
	drive alone	1075	806	5126	1001	1943	10703	9729	30383
	shared ride 2	631	536	3011	545	1227	6549	6065	18564
	shared ride 3+	541	455	2535	398	662	5103	3581	13275
	walk to transit	25	23	190	7	17	569	144	975
	drive to transit	14	3	13	1	1	70	21	123
	school bus	30	22	106	60	76	241	392	927
	Others	35	27	121	16	53	291	206	749
Total		2707	2125	12657	2280	4473	27163	22453	73858

Table 39: Weekday Tour Main Mode by Region (unweighted)

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
tmodetp	Walk	11.1%	10.7%	11.2%	9.3%	9.9%	12.0%	8.9%	10.7%
	Bike	2.1%	1.2%	1.1%	1.7%	1.2%	1.4%	1.4%	1.3%
	drive alone	39.7%	37.9%	40.5%	43.9%	43.4%	39.4%	43.3%	41.1%
	shared ride 2	23.3%	25.2%	23.8%	23.9%	27.4%	24.1%	27.0%	25.1%
	shared ride 3+	20.0%	21.4%	20.0%	17.5%	14.8%	18.8%	15.9%	18.0%
	walk to transit	.9%	1.1%	1.5%	.3%	.4%	2.1%	.6%	1.3%
	drive to transit	.5%	.1%	.1%	.0%	.0%	.3%	.1%	.2%
	school bus	1.1%	1.0%	.8%	2.6%	1.7%	.9%	1.7%	1.3%
	Others	1.3%	1.3%	1.0%	.7%	1.2%	1.1%	.9%	1.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 40: Weekday Tour Transit Sub-mode by Region (unweighted)

Count

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
tpathtp	n/a	2668	2099	12454	2272	4455	26524	22288	72760
	local bus	19	18	146	7	17	451	115	773
	light rail	2	0	24	0	1	71	13	111
	commuter bus	7	7	22	0	0	70	23	129
	commuter rail	10	1	10	1	0	46	14	82
	Ferry	1	0	1	0	0	1	0	3
Total		2707	2125	12657	2280	4473	27163	22453	73858

Table 41 and Figure 3: Mean number of trips per weekday tour by main purpose and region: Each tour has at least two trips, and may have more if there intermediate stops made along the way. The mean number of stops varies mainly by purpose, with work and serve passenger tours having the most (about 3.0 trips per tour), and meal and recreation tours having the least (about 2.3 trips per tour). There is no clear pattern by region, with all regions having an average very close to 2.7 trips per tour.

Table 41: Mean number of trips per weekday tour by Tour Main Purpose and Region (unweighted)

Pdpurp	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
Work	2.9726	2.9569	3.0057	3.0270	2.9303	2.9660	2.9547	2.9695
School	2.5645	2.4335	2.5678	2.6893	2.3961	2.5873	2.6152	2.5785
serve passenger	2.9872	2.9756	2.9226	2.9427	3.1181	2.9431	3.1669	3.0076
personal business	2.5323	2.9333	2.4921	2.3788	2.4833	2.5765	2.5178	2.5369
Shopping	2.8667	2.6476	2.7564	2.9065	2.7986	2.7272	2.8256	2.7774
Meal	2.3632	2.6222	2.3785	2.2874	2.5143	2.3766	2.4108	2.3997
social visit	2.5649	2.5762	2.4209	2.3846	2.5710	2.5316	2.4333	2.4804
Recreation	2.2229	2.2056	2.2720	2.2235	2.2139	2.2769	2.2772	2.2668
Medical	2.7294	2.8409	2.7925	2.9746	2.9897	2.8229	2.9720	2.8931
Total	2.7156	2.7228	2.6928	2.7193	2.7033	2.7041	2.7334	2.7124

Figure 3: Mean number of trips per weekday tour by Tour Main Purpose and Region (unweighted)

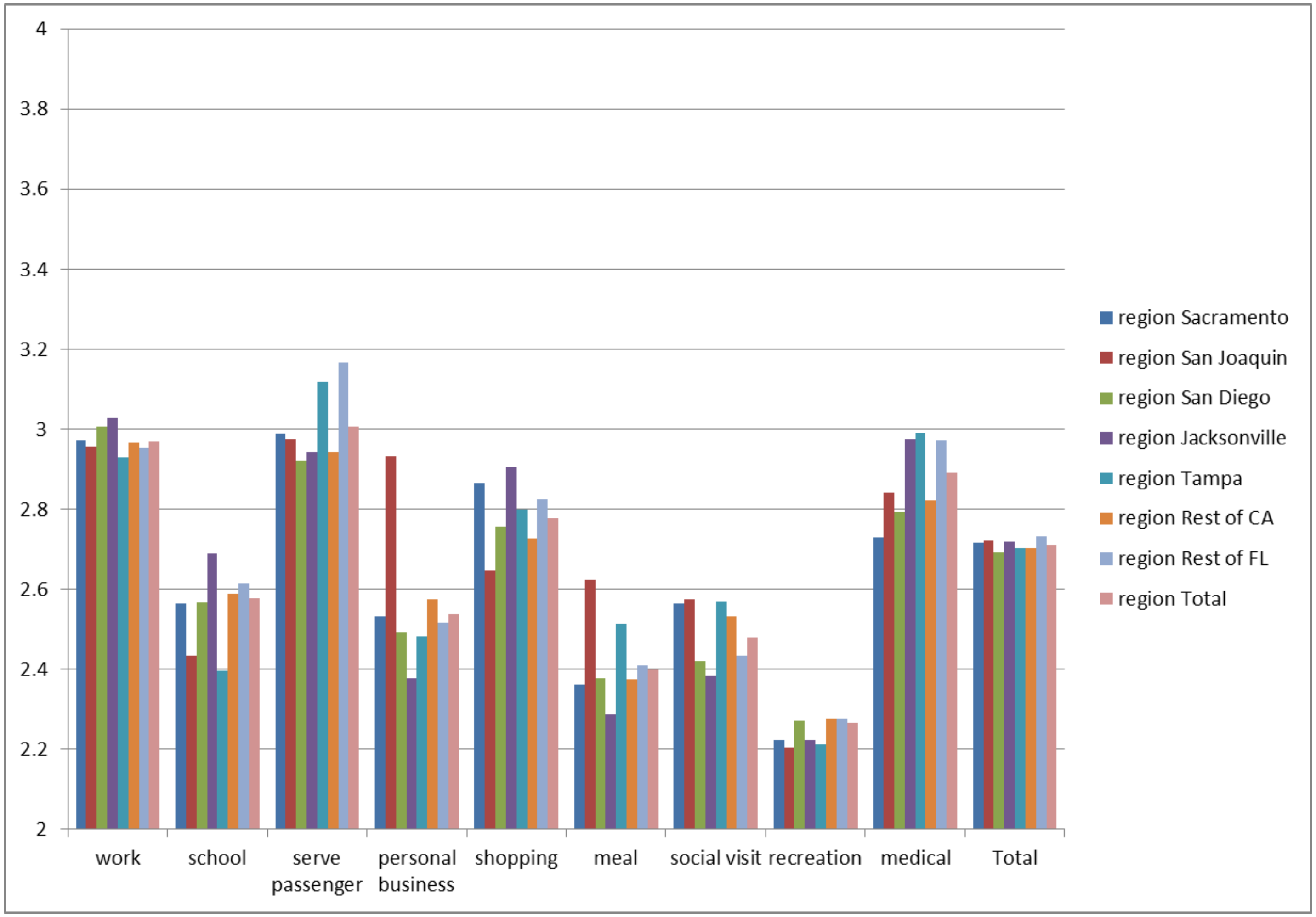
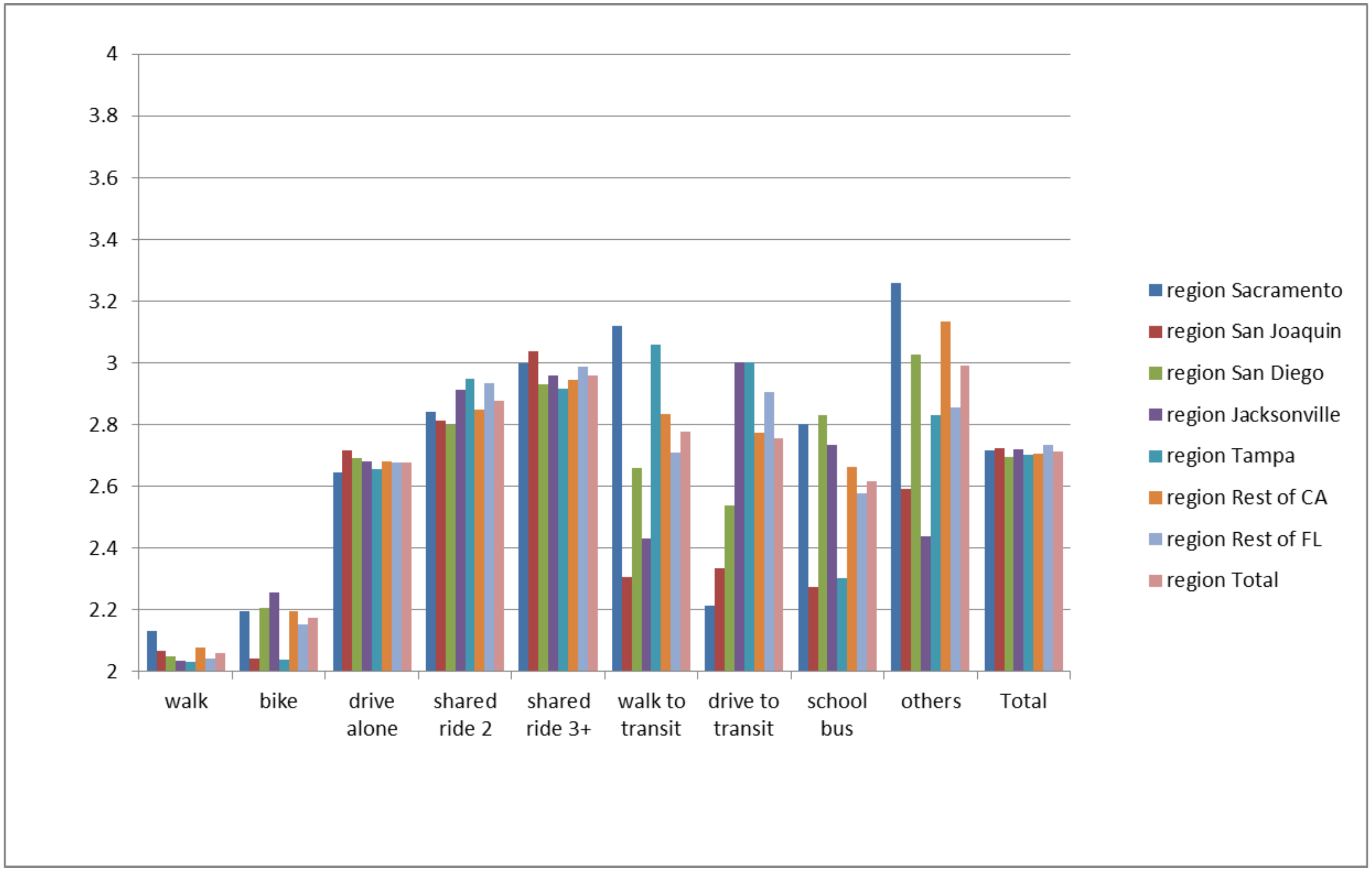


Table 42 and Figure 4: Mean number of trips per weekday tour by main mode and region: There are significant differences in trips per tour by mode, with walk and bike tours having less than 2.2 trips, on average, and shared ride tours having almost 3.0 trips per tour. There are larger variations in the transit, school bus and other modes by region, perhaps due to the small sample sizes. It is a common finding that auto tours are most conducive to making extra stops, particularly for picking up and dropping off passengers.

Table 42: Mean number of trips per weekday tour by Tour Main Mode and Region (unweighted)

Tmodetp	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
Walk	2.1300	2.0658	2.0472	2.0329	2.0295	2.0770	2.0403	2.0601
Bike	2.1964	2.0400	2.2044	2.2564	2.0377	2.1931	2.1531	2.1729
drive alone	2.6428	2.7159	2.6908	2.6813	2.6547	2.6789	2.6764	2.6783
shared ride 2	2.8399	2.8116	2.7977	2.9138	2.9470	2.8484	2.9334	2.8750
shared ride 3+	2.9963	3.0374	2.9314	2.9598	2.9154	2.9445	2.9863	2.9576
walk to transit	3.1200	2.3043	2.6579	2.4286	3.0588	2.8348	2.7083	2.7774
drive to transit	2.2143	2.3333	2.5385	3.0000	3.0000	2.7714	2.9048	2.7561
school bus	2.8000	2.2727	2.8302	2.7333	2.3026	2.6639	2.5765	2.6160
Others	3.2571	2.5926	3.0248	2.4375	2.8302	3.1340	2.8544	2.9893
Total	2.7156	2.7228	2.6928	2.7193	2.7033	2.7041	2.7334	2.7124

Figure 4: Mean number of trips per weekday tour by Tour Main Mode and Region (unweighted)



Trip variables

Tables 43 and 44 and Figures 5 and 6: Observed weekday trip distribution by hour of day and region: The time of day models at the tour and trip levels are based on the departure times of trips during the day. These tables and graphs show the departure time distribution across the day, with both weighted and unweighted tables. (NHTS trip-level weights are used for the weighted tables.) San Joaquin and the Rest of CA show the most pronounced peaking in the AM and PM peaks, with Tampa showing the least peaking. These differences may be due to a different mix of trip and tour purposes, arising from the different person types in the regions. Jacksonville and Sacramento show the most pronounced midday peak at around noon. The use of weighting does not change the shape or ordering of the curves noticeably between Figures 5 and 6.

Table 43: **Observed weekday trip distribution by hour of day and region (unweighted)**

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
xhour	4 am	.1%	.1%	.0%	.3%	.1%	.1%	.0%	.1%
	5 am	.6%	.6%	.6%	.5%	.4%	.5%	.4%	.5%
	6 am	2.1%	1.8%	1.6%	2.2%	1.3%	1.5%	1.5%	1.6%
	7 am	4.3%	4.7%	5.3%	4.9%	5.0%	4.5%	4.6%	4.7%
	8 am	7.3%	8.8%	7.5%	6.8%	5.7%	8.4%	6.6%	7.5%
	9 am	5.9%	5.1%	5.8%	5.6%	5.9%	5.5%	6.0%	5.8%
	10 am	5.5%	5.5%	5.7%	5.8%	7.2%	5.5%	6.3%	5.9%
	11 am	6.8%	5.9%	6.2%	5.9%	7.6%	6.1%	7.1%	6.5%
	12 pm	7.8%	7.5%	7.5%	8.6%	8.1%	7.5%	7.9%	7.7%
	1 pm	6.9%	6.7%	7.5%	7.3%	7.9%	7.5%	7.6%	7.5%
	2 pm	7.6%	7.3%	7.1%	6.9%	7.4%	7.3%	7.8%	7.4%
	3 pm	9.2%	10.2%	9.0%	8.5%	8.0%	8.9%	7.9%	8.6%
	4 pm	8.0%	8.6%	8.5%	8.7%	7.4%	8.0%	8.0%	8.1%

	5 pm	8.3%	8.0%	7.9%	7.3%	7.9%	7.8%	8.0%	7.9%
	6 pm	6.6%	6.8%	6.9%	7.2%	6.8%	7.2%	7.0%	7.0%
	7 pm	5.1%	4.4%	5.1%	5.4%	5.1%	5.3%	5.1%	5.2%
	8 pm	3.5%	3.1%	3.1%	4.1%	3.4%	3.3%	3.5%	3.4%
	9 pm	2.8%	2.7%	2.3%	2.2%	2.5%	2.4%	2.3%	2.4%
	10 pm	1.0%	1.2%	1.4%	1.2%	1.2%	1.5%	1.4%	1.4%
	11 pm	.6%	.5%	.6%	.6%	.6%	.7%	.6%	.6%
	12 am	.1%	.2%	.1%	.1%	.1%	.1%	.1%	.1%
	1 am	.1%	.0%	.1%	.0%	.1%	.0%	.0%	.1%
	2 am	.1%	.1%	.1%	.1%	.0%	.1%	.1%	.1%
	3 am	.0%	.1%	.0%	.0%	.0%	.0%	.0%	.0%
	4 am		.1%	.0%	.0%	.0%	.0%	.0%	.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 44: Observed weekday trip distribution by hour of day and region (weighted)

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
xhour	4 am	.1%	.1%	.0%	.2%	.1%	.1%	.1%	.1%
	5 am	.8%	.7%	.7%	.4%	.6%	.6%	.5%	.6%
	6 am	2.3%	2.4%	1.9%	2.8%	1.5%	1.8%	1.8%	1.9%
	7 am	5.3%	5.4%	6.2%	6.0%	6.2%	5.2%	5.9%	5.5%
	8 am	7.9%	9.7%	8.1%	7.5%	7.0%	9.6%	8.1%	8.9%
	9 am	5.6%	4.9%	5.3%	5.2%	5.9%	5.1%	5.9%	5.3%

10 am	4.4%	4.4%	4.7%	4.4%	5.4%	4.6%	5.0%	4.7%
11 am	5.6%	4.9%	5.0%	4.9%	5.6%	5.1%	5.3%	5.2%
12 pm	7.2%	6.3%	6.7%	7.4%	6.4%	6.3%	6.6%	6.5%
1 pm	6.1%	5.8%	6.7%	6.1%	6.9%	6.8%	6.5%	6.6%
2 pm	6.8%	7.1%	6.4%	6.3%	6.5%	6.7%	6.7%	6.7%
3 pm	9.2%	10.0%	9.1%	9.0%	7.3%	8.8%	7.8%	8.6%
4 pm	8.4%	9.0%	8.8%	8.6%	7.3%	8.1%	7.6%	8.0%
5 pm	8.6%	8.0%	8.1%	7.7%	9.6%	8.0%	8.7%	8.3%
6 pm	6.4%	7.7%	7.0%	7.8%	8.1%	7.7%	7.6%	7.6%
7 pm	5.7%	4.7%	5.7%	6.1%	5.6%	6.0%	5.9%	5.9%
8 pm	3.7%	3.5%	3.4%	3.9%	4.3%	3.9%	4.1%	3.9%
9 pm	3.1%	2.9%	2.7%	2.9%	3.1%	2.7%	2.6%	2.8%
10 pm	1.4%	1.5%	1.8%	1.6%	1.5%	1.8%	1.8%	1.7%
11 pm	.8%	.7%	.9%	1.1%	.9%	.8%	.9%	.9%
12 am	.2%	.2%	.2%	.1%	.1%	.1%	.3%	.2%
1 am	.1%	.0%	.2%	.0%	.1%	.1%	.1%	.1%
2 am	.2%	.1%	.1%	.1%	.1%	.1%	.1%	.1%
3 am	.1%	.1%	.1%	.0%	.0%	.1%	.1%	.1%
4 am		.0%	.0%	.0%	.0%	.0%	.1%	.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 5: Observed weekday trip distribution by hour of day and region (unweighted)

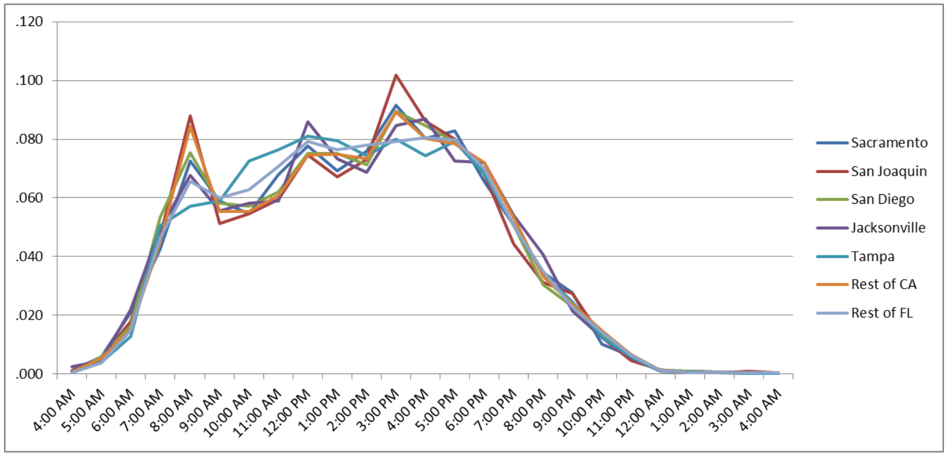
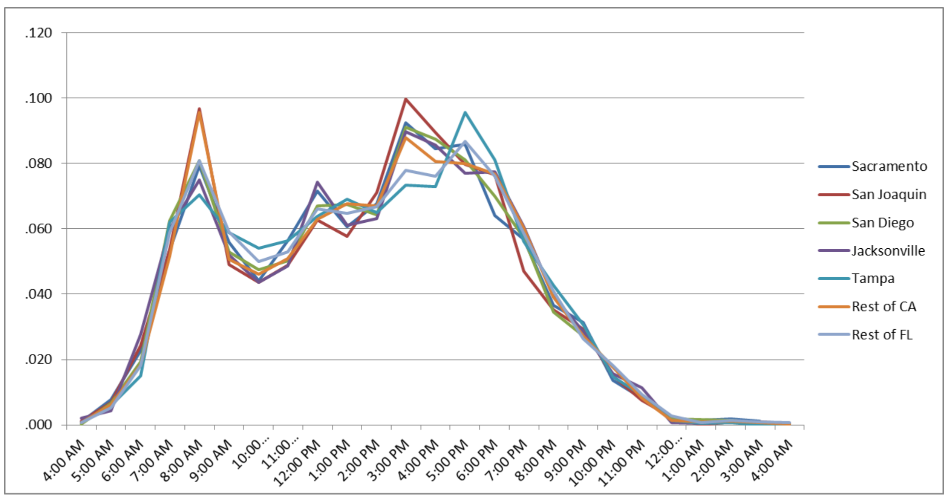


Figure 6: Observed weekday trip distribution by hour of day and region (weighted)



Tables 45 and 46 and Figures 7 and 8: Observed weekday trip distribution by hour of day and trip destination purpose: The most pronounced differences in time of day are related to trip purpose, as one would expect. Trips to school have a very pronounced peak at around 8 AM, while trips to work and for serve passenger have a somewhat smaller peak at that same hour. Trips to meals show a peak at the lunch hour and a smaller one at the dinner hour. Trips to medical and shopping destinations are highest in the midday, while trips for the other purposes (recreation, social visits, personal business, and returning home, are most common in the early evening hours. Again, the use of trip weights does not change the picture substantially.

Table 45: **Observed weekday trip distribution by hour of day and destination purpose (unweighted)**

		home	work	school	serve passenger	personal business	shopping	meal	social visit	recreation	medical	Total
xhour	4 am	.0%	.3%		.0%	.1%	.0%	.0%	.0%	.1%		.1%
	5 am	.1%	2.3%	.1%	.4%	.3%	.1%	.2%	.2%	1.2%	.2%	.5%
	6 am	.6%	6.5%	1.1%	1.4%	1.7%	.4%	.9%	.6%	2.8%	.7%	1.6%
	7 am	1.3%	15.1%	23.7%	8.8%	4.2%	1.0%	2.6%	1.6%	4.9%	2.4%	4.7%
	8 am	2.6%	17.4%	42.5%	17.5%	6.6%	2.4%	3.3%	3.5%	6.6%	8.4%	7.5%
	9 am	2.8%	10.2%	10.7%	6.6%	7.9%	5.5%	3.8%	6.1%	7.7%	12.8%	5.8%
	10 am	3.4%	6.2%	2.6%	3.5%	9.8%	9.6%	3.5%	7.7%	6.5%	15.0%	5.9%
	11 am	4.8%	4.6%	2.0%	3.7%	8.1%	11.4%	7.2%	7.9%	6.0%	12.2%	6.5%
	12 pm	6.4%	7.1%	2.1%	4.7%	7.1%	10.8%	17.6%	7.6%	5.1%	7.4%	7.7%
	1 pm	5.8%	11.0%	2.0%	4.2%	7.2%	10.2%	10.8%	7.4%	5.1%	9.2%	7.5%
	2 pm	7.0%	7.1%	2.0%	8.2%	6.2%	10.5%	5.9%	6.5%	5.2%	11.2%	7.4%
	3 pm	10.6%	4.6%	2.8%	11.3%	7.9%	10.1%	4.0%	7.3%	6.0%	9.2%	8.6%
	4 pm	10.8%	3.2%	2.0%	8.6%	6.7%	8.8%	4.5%	8.1%	8.3%	5.9%	8.1%
	5 pm	11.2%	2.1%	2.0%	8.0%	6.3%	6.9%	8.1%	9.5%	9.0%	3.1%	7.9%
	6 pm	9.6%	1.1%	2.6%	5.3%	7.4%	5.0%	12.1%	11.2%	9.9%	1.1%	7.0%

	7 pm	7.1%	.6%	1.5%	2.9%	7.9%	3.4%	8.2%	8.0%	8.7%	.5%	5.2%
	8 pm	5.9%	.3%	.1%	1.9%	2.2%	2.0%	4.5%	3.8%	3.8%	.4%	3.4%
	9 pm	4.8%	.2%	.0%	1.4%	1.1%	1.0%	1.8%	1.8%	1.8%	.1%	2.4%
	10 pm	3.1%	.1%	.0%	.9%	.8%	.4%	.6%	.8%	.9%	.1%	1.4%
	11 pm	1.4%	.0%	.0%	.5%	.4%	.1%	.3%	.3%	.4%	.0%	.6%
	12 am	.3%	.0%	.0%	.1%	.1%	.0%	.1%	.0%	.1%		.1%
	1 am	.1%			.0%	.0%	.0%	.0%		.0%	.0%	.1%
	2 am	.2%	.0%		.0%	.0%	.0%	.1%	.0%	.0%	.0%	.1%
	3 am	.1%			.0%		.0%		.0%			.0%
	4 am	.1%	.0%		.0%	.0%	.0%		.0%	.0%		.0%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 46: Observed weekday trip distribution by hour of day and destination purpose (weighted)

		home	work	school	serve passenger	personal business	shopping	meal	social visit	recreation	medical	Total
xhour	4 am	.0%	.3%		.0%	.0%	.1%	.1%	.0%	.2%		.1%
	5 am	.1%	2.6%	.1%	.5%	.4%	.2%	.3%	.3%	1.2%	.2%	.6%
	6 am	.7%	7.3%	1.1%	1.6%	2.0%	.6%	.9%	.4%	2.7%	.7%	1.9%
	7 am	1.3%	15.0%	23.5%	10.1%	4.6%	1.1%	2.7%	1.3%	3.8%	2.4%	5.5%
	8 am	2.8%	17.5%	43.2%	20.0%	6.4%	3.0%	3.5%	3.6%	4.6%	8.8%	8.9%
	9 am	2.4%	10.3%	9.7%	6.0%	7.4%	4.8%	3.1%	5.1%	5.7%	13.0%	5.3%
	10 am	2.5%	6.1%	2.9%	2.8%	8.1%	7.9%	3.2%	7.1%	5.4%	13.4%	4.7%
	11 am	3.7%	4.3%	1.9%	3.2%	6.9%	9.5%	7.0%	6.5%	4.6%	11.2%	5.2%
	12 pm	5.2%	6.4%	2.2%	4.1%	6.3%	9.4%	16.9%	7.1%	4.1%	7.7%	6.5%

1 pm	5.1%	10.0%	2.2%	3.4%	6.3%	9.4%	10.1%	6.0%	4.9%	9.2%	6.6%
2 pm	6.4%	6.9%	1.9%	7.6%	5.8%	9.2%	6.0%	5.9%	4.7%	9.6%	6.7%
3 pm	11.1%	4.9%	2.9%	11.6%	7.9%	9.0%	4.2%	7.6%	5.5%	9.6%	8.6%
4 pm	10.8%	3.1%	1.5%	8.2%	7.2%	9.3%	4.0%	8.3%	8.7%	7.7%	8.0%
5 pm	11.4%	2.5%	2.3%	8.1%	6.7%	8.3%	7.7%	10.6%	10.6%	3.4%	8.3%
6 pm	10.3%	1.3%	3.0%	5.2%	8.0%	7.1%	11.3%	11.0%	12.2%	1.6%	7.6%
7 pm	8.0%	.7%	1.5%	2.8%	9.6%	5.3%	8.4%	9.3%	10.9%	.6%	5.9%
8 pm	6.3%	.4%	.1%	1.8%	3.0%	3.2%	6.6%	5.2%	5.3%	.5%	3.9%
9 pm	5.4%	.2%	.0%	1.2%	1.4%	1.5%	2.4%	2.7%	2.8%	.2%	2.8%
10 pm	3.6%	.2%	.0%	.8%	1.2%	.8%	.9%	1.2%	1.5%	.1%	1.7%
11 pm	1.8%	.0%	.0%	.8%	.4%	.1%	.6%	.5%	.6%	.1%	.9%
12 am	.4%	.0%	.0%	.1%	.3%	.1%	.2%	.0%	.0%		.2%
1 am	.2%			.0%	.0%	.0%	.0%		.0%	.0%	.1%
2 am	.3%	.0%		.0%	.0%	.0%	.1%	.0%	.0%	.0%	.1%
3 am	.2%			.0%		.0%		.0%			.1%
4 am	.1%	.0%		.0%	.0%	.0%		.1%	.1%		.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 7:Observed weekday trip distribution by hour of day and destination purpose (unweighted)

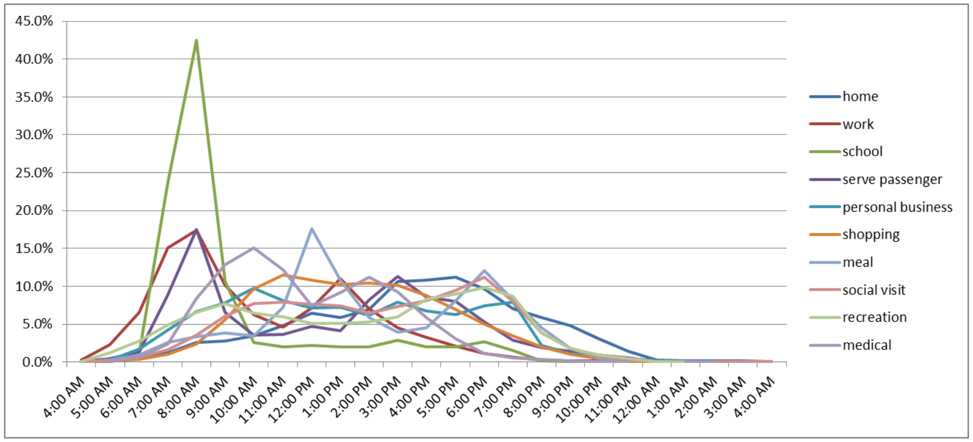
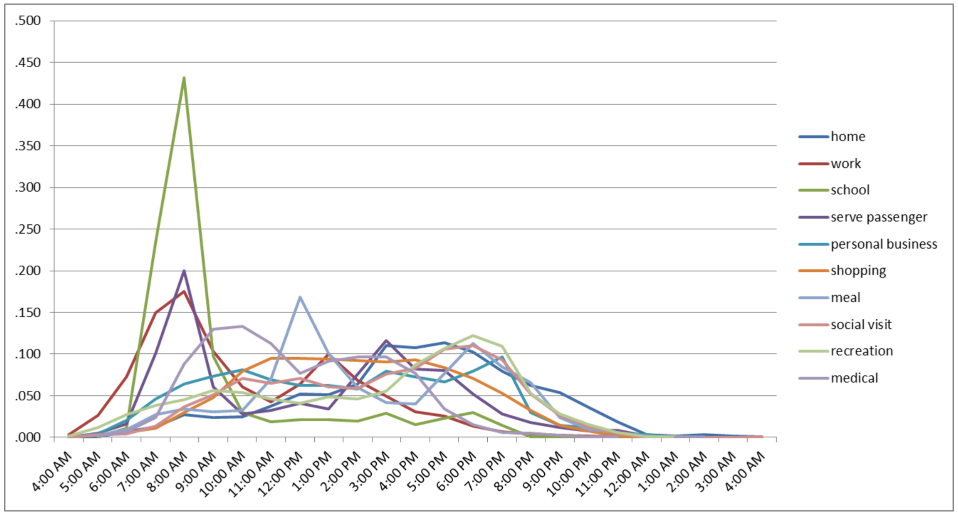


Figure 8:Observed weekday trip distribution by hour of day and destination purpose (weighted)



Tables 47 and 48 and Figures 9 and 10: Observed weekday trip distribution by reported travel duration and region: Respondents' self-reported trip durations were grouped in 5-minute intervals and analyzed to look at the distribution of journey times. The graphs show some "kinking" at 30, 45 and 60 minutes, as is often found with self-reported journey times. There is not a huge difference between the regions, but there is a consistent pattern, with San Joaquin having the highest percentage of short-duration trips, and Jacksonville showing her percentages in the higher durations.

Table 47: **Observed weekday trip distribution by reported travel duration and region (unweighted)**

% within region

		region							
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
tripdur	.00	3.6%	4.1%	3.2%	3.7%	3.3%	3.4%	3.6%	3.5%
	5.00	24.2%	26.2%	21.2%	19.4%	22.2%	23.3%	20.9%	22.1%
	10.00	22.7%	22.1%	22.3%	19.9%	21.3%	22.1%	21.2%	21.8%
	15.00	18.1%	19.4%	19.8%	20.2%	19.5%	18.5%	19.1%	19.0%
	20.00	9.0%	8.8%	10.8%	9.9%	10.5%	9.4%	10.3%	10.0%
	25.00	4.5%	4.3%	5.2%	5.1%	5.1%	4.5%	5.2%	4.9%
	30.00	8.1%	6.6%	8.0%	9.0%	7.8%	7.6%	8.6%	8.0%
	35.00	1.9%	1.4%	2.1%	3.3%	2.2%	2.0%	2.2%	2.1%
	40.00	1.7%	1.1%	1.6%	2.2%	1.6%	1.6%	1.9%	1.7%
	45.00	2.0%	1.7%	2.1%	2.6%	2.4%	2.2%	2.3%	2.2%
	50.00	.7%	.6%	.6%	.9%	.7%	.8%	.7%	.7%
	55.00	.4%	.3%	.3%	.5%	.4%	.5%	.5%	.5%
	60.00	1.1%	1.1%	.9%	1.3%	1.1%	1.5%	1.3%	1.3%
	65.00	.2%	.2%	.2%	.2%	.2%	.2%	.3%	.2%
	70.00	.3%	.3%	.2%	.2%	.2%	.2%	.3%	.2%

	75.00	.2%	.3%	.3%	.4%	.3%	.5%	.4%	.4%
	80.00	.1%	.1%	.1%	.1%	.2%	.2%	.1%	.1%
	85.00	.0%	.2%	.0%	.1%	.1%	.1%	.1%	.1%
	90.00	.2%	.3%	.3%	.3%	.3%	.5%	.3%	.4%
	95.00	.1%	.1%	.0%	.0%	.0%	.1%	.0%	.1%
	100.00	.1%	.0%	.0%	.1%	.0%	.1%	.0%	.1%
	105.00	.0%	.1%	.1%	.1%	.0%	.1%	.1%	.1%
	110.00	.0%	.1%	.0%	.0%	.1%	.0%	.0%	.0%
	115.00	.0%		.0%		.0%	.0%	.0%	.0%
	120.00+	.7%	.6%	.5%	.4%	.4%	.6%	.5%	.5%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 48: **Observed weekday trip distribution by reported travel duration and region (weighted)**

% within region

		region							Total
		Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	
tripdur	.00	3.2%	4.1%	2.9%	3.4%	3.0%	3.0%	3.6%	3.2%
	5.00	24.3%	24.5%	19.9%	19.0%	23.6%	21.9%	19.9%	21.5%
	10.00	22.7%	22.8%	21.5%	18.2%	20.4%	21.3%	20.2%	21.0%
	15.00	17.2%	19.0%	19.9%	19.7%	18.0%	18.0%	18.7%	18.3%
	20.00	9.1%	8.9%	11.5%	9.5%	10.7%	9.6%	10.3%	9.9%
	25.00	5.2%	3.9%	5.2%	5.5%	5.2%	4.6%	5.4%	4.9%
	30.00	8.3%	7.8%	8.7%	10.7%	8.0%	8.6%	9.2%	8.7%
	35.00	1.8%	1.5%	2.2%	3.7%	2.7%	2.3%	2.2%	2.3%

40.00	1.8%	1.4%	1.8%	2.2%	1.8%	1.8%	2.2%	1.9%
45.00	2.2%	1.3%	2.2%	3.2%	2.5%	2.4%	2.7%	2.4%
50.00	.7%	.7%	.7%	.8%	.7%	.9%	.8%	.9%
55.00	.4%	.3%	.4%	.3%	.5%	.5%	.6%	.5%
60.00	1.1%	1.6%	1.2%	1.6%	1.0%	1.9%	1.5%	1.6%
65.00	.2%	.1%	.2%	.1%	.1%	.3%	.4%	.3%
70.00	.3%	.4%	.2%	.2%	.2%	.3%	.3%	.3%
75.00	.4%	.2%	.3%	.5%	.3%	.5%	.4%	.5%
80.00	.2%	.1%	.1%	.1%	.1%	.2%	.2%	.2%
85.00	.0%	.2%	.1%	.1%	.1%	.1%	.1%	.1%
90.00	.2%	.4%	.3%	.4%	.4%	.7%	.5%	.6%
95.00	.1%	.1%	.1%	.0%	.0%	.1%	.1%	.1%
100.00	.1%	.0%	.0%	.0%	.0%	.1%	.0%	.1%
105.00	.0%	.1%	.1%	.0%	.0%	.1%	.1%	.1%
110.00	.0%	.1%	.0%	.1%	.1%	.0%	.0%	.0%
115.00	.0%		.0%		.0%	.1%	.0%	.0%
120.00	.6%	.6%	.6%	.4%	.5%	.8%	.7%	.7%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 9: Observed weekday trip distribution by reported travel duration and region (unweighted)

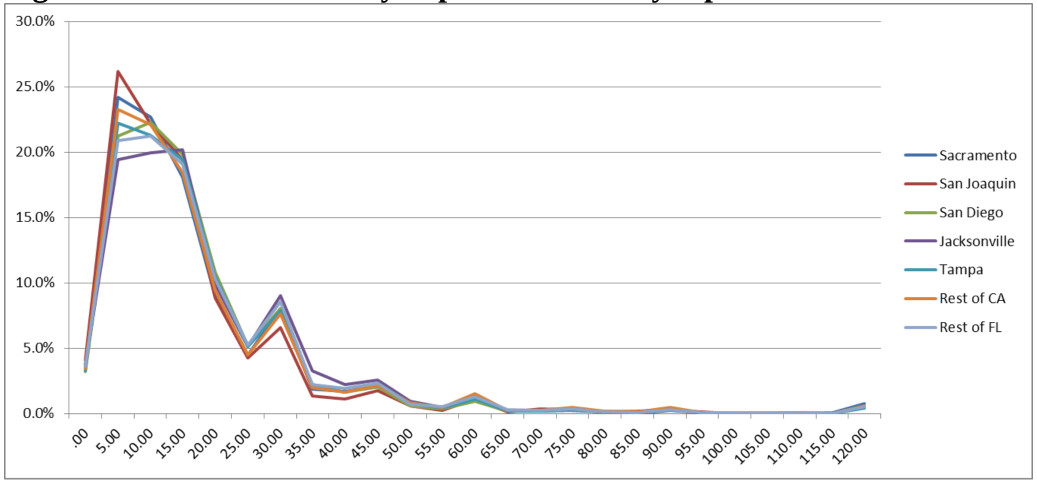
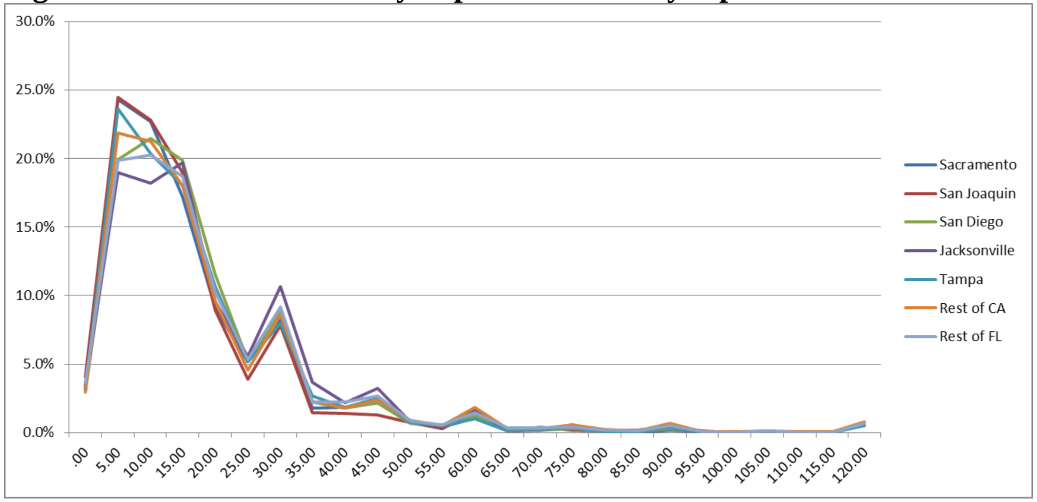


Figure 10: Observed weekday trip distribution by reported travel duration and region (weighted)



Tables 49 and 50 and Figures 11 and 12: Observed weekday trip distribution by reported travel duration and trip destination purpose: The differences in trip duration are more pronounced across trip purposes than across regions. Meal, shopping, and serve passenger trips are consistently the shortest, while work trips and medical trips are the longest.

Table 49: Observed weekday trip distribution by reported travel duration and destination purpose (unweighted)

		home	work	school	serve passenger	personal business	shopping	meal	social visit	recreation	medical	Total
tripdur	.00	2.7%	2.1%	2.4%	3.8%	3.5%	5.6%	4.7%	5.8%	2.5%	1.3%	3.5%
	5.00	19.9%	16.7%	23.2%	25.7%	23.2%	29.2%	29.0%	19.7%	18.8%	12.2%	22.1%
	10.00	21.3%	17.5%	23.4%	24.3%	22.6%	25.0%	24.6%	19.3%	20.3%	17.6%	21.8%
	15.00	19.6%	17.3%	17.8%	19.3%	19.9%	17.6%	18.8%	18.7%	22.4%	20.7%	19.0%
	20.00	10.2%	11.2%	10.4%	9.3%	9.7%	8.6%	8.3%	11.2%	10.5%	13.1%	10.0%
	25.00	5.0%	6.8%	5.0%	4.6%	4.8%	3.5%	3.4%	4.7%	4.6%	7.9%	4.9%
	30.00	9.0%	10.4%	7.1%	5.6%	7.6%	5.0%	5.5%	8.7%	10.4%	11.5%	8.0%
	35.00	2.3%	3.5%	2.5%	1.7%	1.5%	1.2%	1.1%	2.2%	1.8%	3.3%	2.1%
	40.00	1.9%	3.1%	2.0%	1.3%	1.2%	1.0%	.8%	1.5%	1.5%	2.2%	1.7%
	45.00	2.5%	3.6%	1.8%	1.3%	1.7%	1.2%	1.1%	2.4%	2.5%	3.3%	2.2%
	50.00	.8%	1.4%	.9%	.6%	.7%	.4%	.3%	.6%	.7%	.9%	.7%
	55.00	.5%	.8%	.4%	.3%	.4%	.3%	.2%	.3%	.3%	.7%	.5%
	60.00	1.5%	1.9%	1.1%	.7%	1.1%	.6%	.8%	1.8%	1.5%	2.0%	1.3%
	65.00	.3%	.5%	.3%	.1%	.2%	.1%	.1%	.2%	.2%	.4%	.2%
	70.00	.3%	.4%	.4%	.2%	.2%	.1%	.1%	.2%	.2%	.3%	.2%
	75.00	.5%	.7%	.3%	.3%	.3%	.1%	.2%	.4%	.4%	.6%	.4%

	80.00	.2%	.3%	.1%	.1%	.2%	.1%	.0%	.2%	.1%	.1%	.1%
	85.00	.1%	.2%	.0%	.0%	.1%	.0%	.1%	.1%	.1%	.1%	.1%
	90.00	.4%	.6%	.3%	.2%	.3%	.1%	.2%	.5%	.4%	.6%	.4%
	95.00	.1%	.1%	.1%	.0%	.1%	.0%	.0%	.1%	.1%	.0%	.1%
	100.00	.1%	.1%	.0%	.1%	.0%	.0%	.0%	.1%	.0%	.1%	.1%
	105.00	.1%	.1%	.0%	.0%	.1%	.0%	.0%	.2%	.1%	.1%	.1%
	110.00	.1%	.0%	.0%	.0%	.0%	.0%	.0%	.1%	.0%	.0%	.0%
	115.00	.0%	.0%		.0%	.1%	.0%	.0%	.0%	.0%	.0%	.0%
	120.00	.8%	.6%	.4%	.4%	.5%	.2%	.3%	.8%	.6%	.9%	.5%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 50: Observed weekday trip distribution by reported travel duration and destination purpose (weighted)

		home	work	school	serve passenger	personal business	shopping	meal	social visit	recreation	medical	Total
tripdur	.00	2.6%	1.8%	2.2%	3.7%	3.1%	5.3%	5.3%	5.4%	2.6%	1.2%	3.2%
	5.00	19.4%	15.9%	21.6%	27.1%	22.1%	28.5%	31.7%	19.4%	18.3%	10.7%	21.5%
	10.00	20.2%	16.6%	22.2%	25.1%	22.3%	24.3%	24.9%	19.3%	20.4%	16.4%	21.0%
	15.00	18.8%	17.1%	16.9%	18.7%	19.1%	17.7%	17.0%	17.8%	22.4%	18.6%	18.3%
	20.00	10.1%	11.1%	11.0%	8.8%	9.8%	8.5%	7.3%	11.0%	9.7%	13.8%	9.9%
	25.00	4.8%	6.7%	5.0%	4.4%	5.4%	3.5%	3.2%	4.5%	4.5%	8.2%	4.9%
	30.00	9.8%	11.0%	7.8%	5.3%	8.2%	5.7%	5.3%	9.1%	11.4%	12.9%	8.7%
	35.00	2.5%	3.4%	2.7%	1.4%	1.4%	1.4%	1.4%	2.4%	1.7%	2.9%	2.3%
	40.00	2.1%	3.1%	2.5%	1.0%	1.4%	1.2%	.8%	1.4%	1.3%	2.2%	1.9%
	45.00	2.6%	4.0%	2.2%	1.3%	1.8%	1.4%	.8%	3.3%	2.8%	3.2%	2.4%
	50.00	.9%	1.5%	1.0%	.7%	1.0%	.4%	.4%	.5%	.7%	1.1%	.9%

55.00	.6%	.9%	.5%	.3%	.3%	.2%	.3%	.4%	.4%	1.0%	.5%
60.00	1.9%	2.2%	1.6%	.8%	1.8%	.9%	.7%	2.4%	1.7%	2.7%	1.6%
65.00	.2%	.6%	.3%	.1%	.2%	.1%	.1%	.3%	.1%	.4%	.3%
70.00	.4%	.5%	.4%	.2%	.3%	.1%	.1%	.2%	.2%	.4%	.3%
75.00	.6%	.8%	.4%	.3%	.3%	.1%	.1%	.4%	.3%	.8%	.5%
80.00	.2%	.4%	.2%	.2%	.2%	.0%	.1%	.2%	.1%	.1%	.2%
85.00	.1%	.2%	.1%	.1%	.1%	.0%	.1%	.0%	.2%	.1%	.1%
90.00	.7%	.9%	.5%	.1%	.4%	.2%	.1%	.7%	.5%	1.4%	.6%
95.00	.1%	.1%	.1%	.0%	.1%	.0%	.0%	.1%	.0%	.1%	.1%
100.00	.0%	.1%	.1%	.1%	.0%	.0%	.0%	.2%	.0%	.3%	.1%
105.00	.1%	.2%	.0%	.1%	.1%	.1%	.0%	.1%	.1%	.0%	.1%
110.00	.1%	.0%	.0%	.0%	.0%	.0%	.0%	.1%	.0%	.1%	.0%
115.00	.1%	.0%		.0%	.1%	.0%	.0%	.0%	.0%	.0%	.0%
120.00	1.0%	.8%	.6%	.3%	.6%	.3%	.3%	1.0%	.5%	1.2%	.7%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 11: Observed weekday trip distribution by reported travel duration and destination purpose (unweighted)

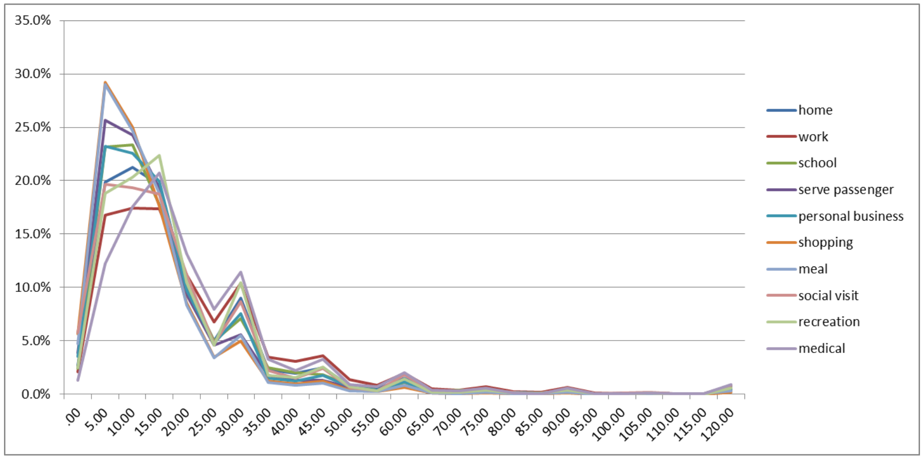
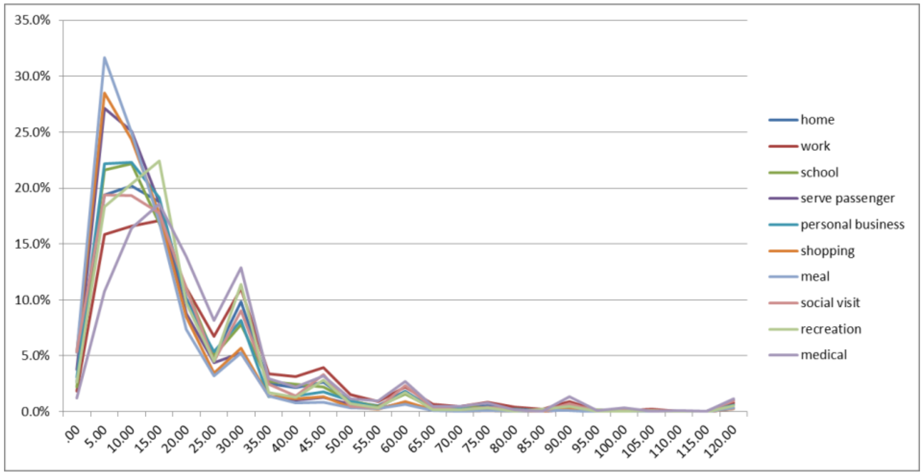


Figure 12: Observed weekday trip distribution by reported travel duration and destination purpose (weighted)



Tables 51 and 52 and Figures 13 and 14: Mean observed weekday trip duration by mode and region: The average durations for walk, bike, and car trips are around 15 minutes for all of the regions, while the average durations for transit and school bus trips are 30 minutes or more for all regions. There is not a great deal of congestion in any of the regions that would cause longer trip durations.

Table 51: Mean Observed Trip Duration by Mode and Region (unweighted)

mode	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
walk	14.1546	13.1302	14.9809	13.3976	13.8635	15.2544	13.0572	14.4535
bike	18.5781	21.0400	22.8505	17.4524	17.4375	19.8156	17.3023	19.2572
drive alone	16.9929	15.6390	16.8168	18.2703	16.7383	17.5209	17.4769	17.2921
shared ride 2	15.8601	14.6857	15.7774	17.5089	16.6827	16.0393	17.3036	16.4665
shared ride 3+	14.7675	15.6335	15.9901	18.2509	18.5580	15.6744	17.9007	16.5454
transit	41.7500	48.9556	42.5489	63.0000	41.0732	45.3209	44.9970	44.7265
school bus	35.6047	34.7692	32.4379	32.7830	31.0073	31.2928	32.3171	32.1337
others	21.7031	19.3220	24.0544	25.5429	26.4952	25.0036	27.9150	25.3929
Total	16.5210	15.6483	16.7575	18.0775	17.0334	17.2782	17.5067	17.1947

Table 52: Mean Observed Trip Duration by Mode and Region (weighted)

Mode	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
walk	13.9475	13.5692	15.2575	13.2778	13.1000	15.5163	12.5692	14.7071
bike	15.0019	24.1605	21.3516	15.5294	13.5095	19.5894	17.4774	18.6927
drive alone	18.1365	16.6922	17.5932	18.9945	17.4922	18.8191	19.3360	18.7091
shared ride 2	15.5662	14.1684	15.6568	18.0951	16.7621	16.1509	17.2518	16.4019
shared ride 3+	13.8389	15.4981	15.9947	17.5392	16.3014	16.0983	17.1281	16.2449
transit	42.6240	48.2119	42.5478	71.2692	33.4196	45.8446	42.5253	45.1835
school bus	30.3333	36.6431	33.2609	32.6033	31.4328	32.3626	33.0166	32.6900
others	21.1685	17.2526	23.9322	31.8616	29.7617	25.7781	29.2842	26.4064
Total	16.7034	16.1822	17.4291	18.9329	17.1412	18.4142	18.5308	18.1673

Figure 13: Mean Observed Trip Duration by Mode and Region (unweighted)

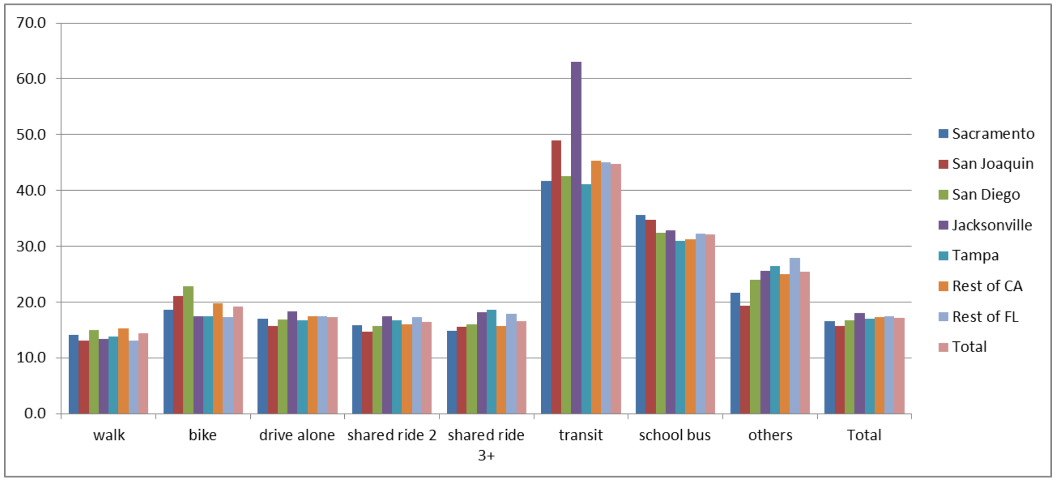
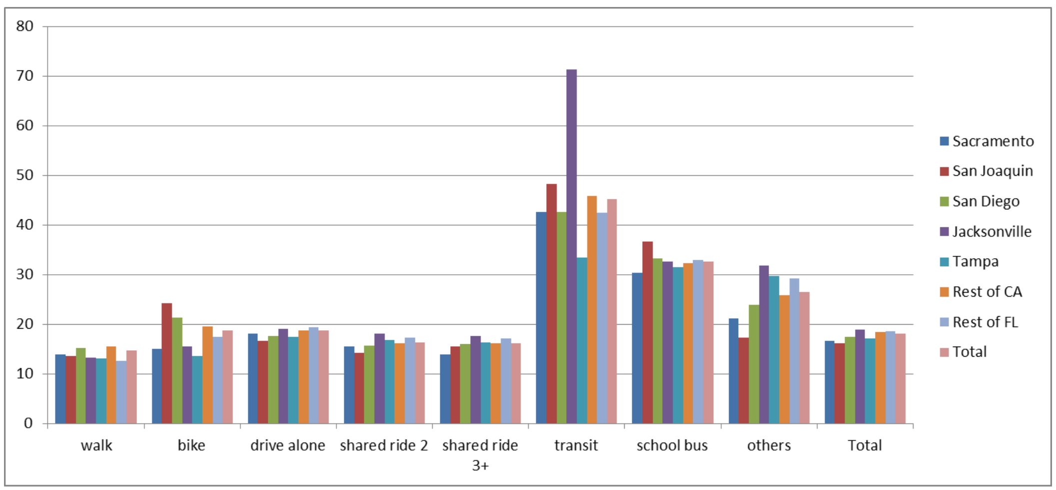


Figure 14: Mean Observed Trip Duration by Mode and Region (weighted)



Tables 53 and 54 and Figures 15 and 16: Mean observed weekday trip duration by mode and region: Again, the differences in means are more pronounced by purpose than by region, although some consistent trends are seen by region as well. The San Joaquin region has the shortest mean duration for all of the purposes except personal business and medical, while the Jacksonville region has the longest mean duration for almost all of those same purposes. These differences may be related to land use patterns and the physical size of the various urban areas, a question that the model estimation will address.

Table 53: **Mean Observed Trip Duration by Destination Purpose and Region (unweighted)**

dpurp	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
home	17.5004	16.7574	17.6866	19.5703	17.8079	18.3802	18.6368	18.2625
work	20.8227	19.3251	20.6618	21.6586	20.7187	21.9285	21.4185	21.3707
school	15.4280	13.7649	15.4049	21.8148	18.9818	16.2200	18.9074	16.8452
serve passenger	14.7313	12.6976	14.5978	16.4378	17.3915	13.9545	16.5022	14.9462
personal business	17.4290	18.4842	15.3053	16.8525	14.8750	17.1034	15.8284	16.2983
shopping	12.4483	11.5955	13.1473	13.8642	13.3805	13.3503	13.6691	13.3591
meal	13.0356	12.7434	13.3768	13.3073	14.5807	13.1336	14.6728	13.7448
social visit	18.1600	15.9033	18.0911	18.7431	18.7081	18.1615	17.9614	18.0636
recreation	16.8344	17.5894	18.6430	18.2514	17.8516	18.7338	17.5845	18.2132
medical	19.4211	23.0735	20.3868	21.7560	20.9817	21.4070	21.8673	21.3995
Total	16.5210	15.6483	16.7575	18.0775	17.0334	17.2782	17.5067	17.1947

Table 54: Mean Observed Trip Duration by Destination Purpose and Region (weighted)

Dpurp	region							
	Sacramento	San Joaquin	San Diego	Jacksonville	Tampa	Rest of CA	Rest of FL	Total
home	17.5307	17.4922	18.3435	20.5337	17.8465	19.6715	19.6816	19.3404
work	20.4597	20.3148	21.0896	22.2959	19.9921	23.0587	23.1900	22.5846
school	16.3454	15.3148	16.6771	21.6017	18.7399	17.8217	20.1603	18.2022
serve passenger	15.3932	11.9446	15.0106	16.6157	16.6794	14.0279	15.7135	14.5649
personal business	16.9562	18.5313	16.4313	16.5568	13.7761	17.9944	17.2258	17.3802
shopping	13.2422	11.7702	13.6378	13.8228	13.6365	14.3988	13.8659	14.0013
meal	11.4122	12.4502	13.1433	12.3261	13.3311	12.9673	13.7376	13.0896
social visit	16.3901	17.4522	19.0131	20.5531	17.2000	19.0904	19.0815	18.8110
recreation	16.7354	17.5900	18.5608	19.7574	18.2068	19.0545	17.3035	18.4152
medical	18.9371	22.9883	21.8946	21.3596	22.5133	24.4020	23.2352	23.4889
Total	16.7034	16.1822	17.4291	18.9329	17.1412	18.4142	18.5308	18.1673

Figure 15: Mean Observed Trip Duration by Destination Purpose and Region (unweighted)

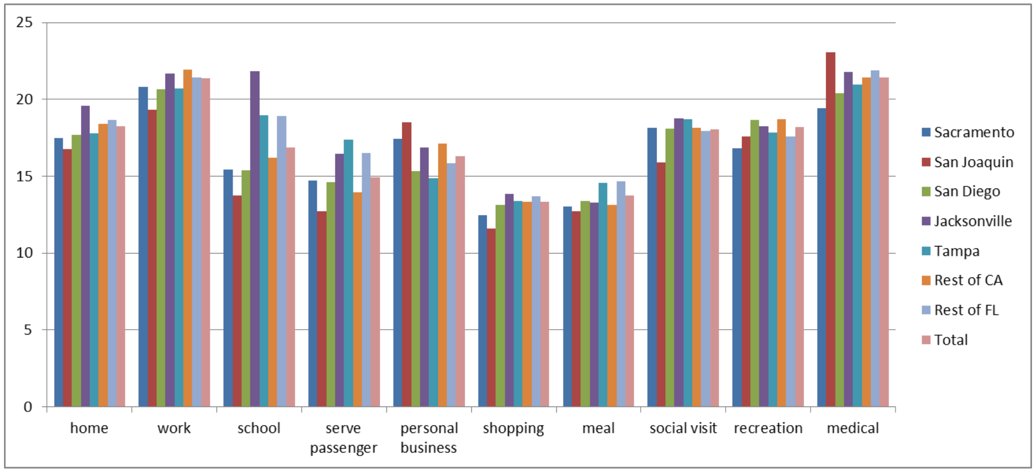
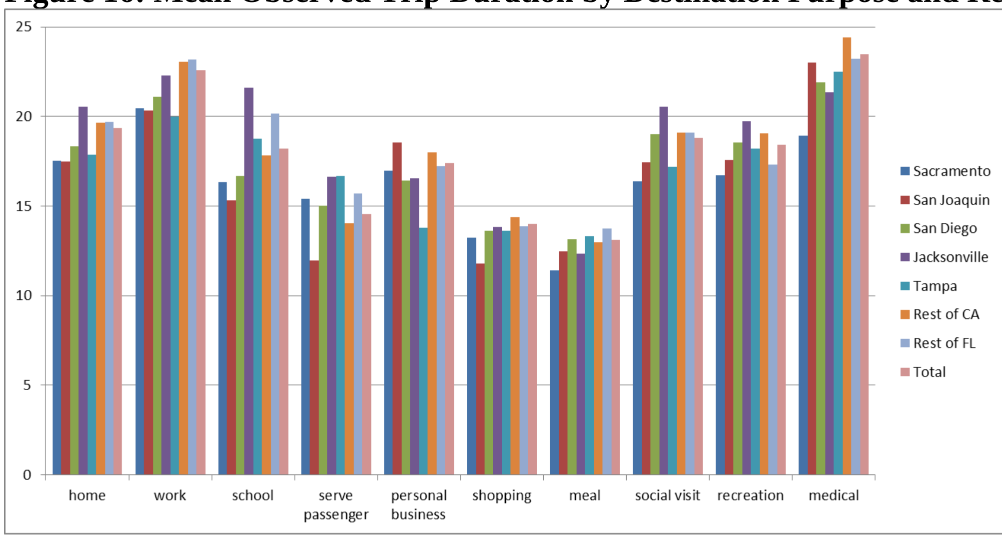


Figure 16: Mean Observed Trip Duration by Destination Purpose and Region (weighted)



Preliminary Summary Observations

Impacts of small sample size. Except for San Diego, all the regions have small sample sizes relative to typical regional household survey sizes used for estimating activity schedule models (Table 1), and this will, in general, limit our ability to find statistical differences across regions. Beyond that, there will be some cases where small sample will almost surely prevent meaningful comparisons across regions. These include children under the age of 5, for whom no travel diaries were collected (Table 28), and children age 16+ and college students, for whom the sample sizes are especially small (also Table 28), as well as trips involving bicycle and transit modes (Tables 38-40.)

Differences across regions. There are some differences in the population across the regions, notably with regard to household type (Table 9), income group (Table 17) and person type (Table 22). The model specifications can, to a great extent, control for these differences, so that the differences might not limit model transferability. Of the three major differences noted here, the DaySim model specifications used as the starting point for the transferability testing may be weakest in using household type to condition the specifications. Improving the model specifications to control for different household types might enhance model transferability.

In addition to differences in population, there are some differences across regions in trip distribution by hour of the day (Figure 6), and in trip duration (Figure 10). It remains to be seen whether or not factors controlled for in the models, such as purpose (Figure 8), and travel impedance, will account for these differences.

Appendix 2: Data dictionary of the estimation results data file

This appendix is the data dictionary for the estimation results data file, from which all summary tables and figures were produced. The file includes a record for each coefficient in each estimated model. A model was estimated for each combination of Model Type (mtype) and Model Spec (mspec). However, the location choice models (mtypes 1, 10 and 14) do not have models for Model Specs 25-36, those with a base that has alternative-specific constants (ASCs) for each region. This is because the location choice models do not have ASCs.

Name	Type	Wid.	Dec.	Label	Values	Description
Mname	String	7	0	Model Name (mname)	workloc autownx idpatt1 exactt1 wttime1 wtmode1 wtmode2 subtour stmode1 otdest1 ottime1 otmode1 instop1 stoploc trptim1	1 'Usual work location' 2 'Auto ownership' 3 'Person-day tour generation' 4 'Exact number of tours' 5 'Work tour time of day' 6 'Work tour mode (detailed LOS)' 7 'Work tour mode (combined LOS)' 8 ' WB subtour generation' 9 'School tour mode' 10 'Other tour destination' 11 'Other HB tour time of day' 12 'Other HB tour mode' 13 'Intermediate stop generation' 14 'Intermediate stop location' 15 'Trip time of day'
Mtype	Num.	1	0	Model Type (mtype)	1 'Usual work location' 2 'Auto ownership' 3 'Person-day tour generation' 4 'Exact number of tours' 5 'Work tour time of day' 6 'Work tour mode (detailed LOS)' 7 'Work tour mode (combined LOS)' 8 ' WB subtour generation' 9 'School tour mode' 10 'Other tour destination' 11 'Other HB tour time of day' 12 'Other HB tour mode' 13 'Intermediate stop generation' 14 'Intermediate stop location' 15 'Trip time of day'	

Name	Type	Wid.	Dec.	Label	Values	Description
Mspec	Num.	1	0	Model Spec (mspec)	1 '2State_base' 2 'Calif_base' 3 'Florida_base' 4 '2State_base_ASC' 5 'Calif_base_ASC' 6 'Florida_base_ASC' 7 'SANDAG_base' 8 'SACOG_base' 9 'Fresno_base' 10 '3County_base' 11 'Tampa_base' 12 'Jacksnv_base' 13 'SANDAG_Dif_2S' 14 'SACOG_Dif_2S' 15 'Fresno_Dif_2S' 16 '3County_Dif_2S' 17 'Tampa_Dif_2S' 18 'Jacksnv_Dif_2S' 19 'SANDAG_Dif_1S' 20 'SACOG_Dif_1S' 21 'Fresno_Dif_1S' 22 '3County_Dif_1S' 23 'Tampa_Dif_1S' 24 'Jacksnv_Dif_1S' 25 'SANDAG_Dif_2S_ASC' 26 'SACOG_Dif_2S_ASC' 27 'Fresno_Dif_2S_ASC' 28 '3County_Dif_2S_ASC' 29 'Tampa_Dif_2S_ASC' 30 'Jacksnv_Dif_2S_ASC' 31 'SANDAG_Dif_1S_ASC' 32 'SACOG_Dif_1S_ASC' 33 'Fresno_Dif_1S_ASC' 34 '3County_Dif_1S_ASC' 35 'Tampa_Dif_1S_ASC' 36 'Jacksnv_Dif_1S_ASC'	1-3: Base model without ASCs for all regions 4-6: Base model with ASCs for all regions 7-12: Region-speciifc models 13-18: 2-state base with coefficient differences for named region 19-24: 1-state base with coefficient differences for named region 25-30: 2-state base with ASCs for each coefficient and coefficient differences for named region 31-36: 1-state base with ASCs for each coefficient and coefficient differences for named region
Dataused	Num.	1	0	Data Used (dataused)	1 'sandag' 2 'sacog' 3 'fresno' 4 '3county' 5 'tampa' 6 'jacksnvl' 7 'calif' 8 'florida' 9 'cal&fla'	Data set used from model estimation. 7 'calif' includes all 4 CA regions 8 'florida' includes both FL regions 9 'cal&fla' includes all 6 regions
Diftype	Num.	1	0	Type of Difference Model (diftype)	0 'none' 1 'ASCs free' 2 'one regoin' 3 'one region & ASCs free'	0 'none': for mspec 1-3, 7-12 '1 ASCs free': for mspec 4-6 '2 one region' for mspec 13-24 '3 one region & ASCs free' for mspec 25-36

Name	Type	Wid.	Dec.	Label	Values	Description
Difregion	Num.	1	0	Difference Region (difregion)	0 'none' 1 'sandag' 2 'sacog' 3 'fresno' 4 '3county' 5 'tampa' 6 'jacksnvl'	region for which difference coefficients are estimated
Basespec	Num.	1	0	Base Model (basespce)	0 'n/a' 1 '2State_base' 2 'Calif_base' 3 'Florida_base' 4 '2State_base_ASC' 5 'Calif_base_ASC' 6 'Florida_base_ASC'	the base estimated model. '_ASC' means ASCs were estimated for all regions
Totpar	Num.	2	0	Number of Coefficients Total (totpar)		
Estpar	Num.	2	0	Number of Coefficients Estimated (estpar)		
totdif	Num.	2	0	Number of Difference Coefficients Total (totdif)		
estdif	Num.	2	0	Number of Difference Coefficients Estimated (estdif)		
nobs	Num.	5	0	Number of Observations (nobs)		
niter	Num.	1	0	Number of Estimation Iterations (niter)		
converge	Num.	1	0	Model Converged? (converge)	0 'no' 1 'yes'	
like0	Num.	9	2	Log Likelihood (0) (like0)		Log likelihood with no estimated coefficients
likec	Num.	9	2	Log Likelihood (constants) (likec)		Log likelihood with a full set of alternative specific constants
likef	Num.	9	2	Likelihood (final) (likef)		Log likelihood with final estimated values
chidf	Num.	2	0	Degrees of Freedom (U - R) (chidf)		Difference in number of estimated coefficients between unrestricted model (U) and restricted model (R). --unrestricted model includes region-specific difference coefficients --restricted model excludes them
chival	Num.	6	2	Chi Squared Statistic (chival)		-2(Likef(R) - Likef(U))

Name	Type	Wid.	Dec.	Label	Values	Description
chiprob	Num.	6	2	Chi Squared Probability (chiprob)		Under the null hypothesis that U and R are the same model (ie that the region's model is no different than the base model), this is the probability that chival would be less than what was actually observed. That is, chiprob is the probability that the two models are different.
pnum	Num.	2	0	Coefficient Number (pnum)		
plabel	String	10	0	Coefficient Label (plabel)		
ptype1	Num.	1	0	Coefficient Type 1 (ptype1)	0 'none' 1 'A-constant' 2 'P-person' 3 'H-household' 4 'D-day pattern' 5 'T-tour/trip' 6 'I-impedance' 7 'U-land use' 8 'W-time window' 9 'C-logsum' 10 'G-size variable' 11 'L-log size mult'	
ptype2	Num.	1	0	Coefficient Type 2 (ptype)	same as ptype1	This identifies a second type in the case of variables specified as interactions, such as a size variable (11) for a particular tour purpose (5)
Difvar	Num.	1	0		0 'no' 1 'yes'	Is this a coefficient for a difference variable?
Basevar	Num.	2	0			
Expsign	Num.	2	0		-1 'negative' 0 'unknown' 1 'positive'	Expected sign, a priori (Note: this was not used in final reporting)
Constr	Num.	1	0		0 'no' 1 'yes'	Was this coefficient constrained to a particular value in estimation?
Coeff	Num.	11	8			Final coefficient value, either constrained or estimated
tstat	Num.	7	3		-9999 'not estimated'	asymptotic t statistic value if coefficient was estimated; -9999 otherwise
constrbase	Num.	1	0		0 'no' 1 'yes'	Was this coefficient also constrained to a particular value in the associated two-state base model (mspec=1)?
Coeffbase	Num.	11	8			Final coefficient value, either constrained or estimated, in the associated two-state base model
Tstatbase	Num.	7	3			asymptotic t statistic value from the associated two-state base model if the coefficient was estimated; -9999 otherwise

Appendix 3: Summary Statistics for all Models

Summary Statistics for all Base Models

Table 1a

Usual work location

Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	7526	54	48	0.128	67%	33%	0%	0%	0%
Calif_base	5584	54	48	0.126	67%	24%	9%	0%	0%
Florida_base	1942	54	47	0.150	55%	27%	6%	9%	3%
SANDAG_base	4195	54	48	0.125	61%	27%	9%	3%	0%
SACOG_base	807	54	39	0.118	39%	39%	15%	6%	0%
Fresno_base	247	54	45	0.108	21%	55%	15%	0%	9%
3County_base	335	54	37	0.194	15%	45%	30%	0%	9%
Tampa_base	1265	54	47	0.160	42%	30%	21%	3%	3%
Jacksnv_base	677	54	47	0.142	42%	27%	18%	9%	3%

Size variables excluded from the significance and estimability columns (the five columns on the right

Table 1b

Auto ownership

Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	12203	24	24	0.310	83%	17%	0%	0%	0%
Calif_base	8351	24	24	0.303	83%	13%	0%	4%	0%
Florida_base	3852	24	24	0.334	79%	13%	8%	0%	0%
SANDAG_base	6002	24	24	0.302	83%	8%	4%	4%	0%
SACOG_base	1311	24	24	0.297	46%	50%	4%	0%	0%
Fresno_base	381	24	24	0.323	54%	38%	4%	4%	0%
3County_base	657	24	24	0.297	58%	21%	21%	0%	0%
Tampa_base	2517	24	22	0.357	75%	8%	8%	0%	8%
Jacksnv_base	1335	24	24	0.314	71%	13%	13%	4%	0%

Table 1c

Person-day tour generation

Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	18770	127	126	0.442	73%	27%	0%	0%	0%
Calif_base	13268	127	126	0.444	68%	28%	4%	0%	0%
Florida_base	5502	127	126	0.436	44%	44%	11%	0%	0%
SANDAG_base	9567	127	126	0.443	66%	28%	6%	1%	0%
SACOG_base	2023	127	126	0.438	40%	52%	8%	0%	0%
Fresno_base	656	127	122	0.450	20%	44%	33%	0%	3%
3County_base	1022	127	122	0.430	20%	49%	24%	4%	3%
Tampa_base	3449	127	125	0.432	37%	50%	12%	1%	1%
Jacksnv_base	2053	127	126	0.439	37%	49%	13%	1%	0%

Table 1d

Exact number of tours

Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	20795	98	86	0.655	40%	60%	0%	0%	0%
Calif_base	14781	98	86	0.659	36%	55%	9%	0%	0%
Florida_base	6014	98	83	0.640	23%	52%	21%	0%	3%
SANDAG_base	10715	98	86	0.661	33%	55%	13%	0%	0%
SACOG_base	2283	98	75	0.644	22%	45%	20%	0%	13%
Fresno_base	673	98	49	0.701	5%	40%	13%	0%	43%
3County_base	1110	98	73	0.603	14%	44%	24%	2%	15%
Tampa_base	3820	98	80	0.629	19%	48%	27%	0%	7%
Jacksnv_base	2194	98	76	0.645	16%	47%	26%	0%	12%

Table 1e

Work tour time of day

Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	5347	74	69	0.223	72%	28%	0%	0%	0%
Calif_base	3924	74	69	0.219	71%	25%	4%	0%	0%
Florida_base	1423	74	67	0.234	67%	26%	3%	1%	3%
SANDAG_base	2832	74	69	0.216	68%	29%	3%	0%	0%
SACOG_base	617	74	64	0.207	36%	45%	12%	0%	7%
Fresno_base	173	74	61	0.219	29%	35%	25%	0%	12%
3County_base	302	74	57	0.220	36%	30%	14%	1%	17%
Tampa_base	883	74	67	0.232	64%	28%	4%	1%	3%
Jacksnv_base	540	74	63	0.230	45%	32%	14%	0%	9%

Table 1f

Work tour mode (combined LOS)

Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	5282	31	31	0.611	61%	39%	0%	0%	0%
Calif_base	3866	31	31	0.594	48%	45%	6%	0%	0%
Florida_base	1416	31	21	0.662	29%	39%	0%	0%	32%
SANDAG_base	2794	31	31	0.606	48%	39%	13%	0%	0%
SACOG_base	601	31	30	0.578	23%	52%	23%	0%	3%
Fresno_base	173	31	19	0.404	10%	39%	13%	0%	39%
3County_base	298	31	18	0.503	6%	39%	13%	0%	42%
Tampa_base	880	31	21	0.608	19%	42%	6%	0%	32%
Jacksnv_base	536	31	17	0.539	10%	32%	13%	0%	45%

Table 1g

WB subtour generation									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	3447	14	14	0.393	79%	21%	0%	0%	0%
Calif_base	2553	14	14	0.385	64%	21%	14%	0%	0%
Florida_base	894	14	14	0.405	50%	43%	7%	0%	0%
SANDAG_base	1852	14	14	0.383	64%	21%	14%	0%	0%
SACOG_base	410	14	13	0.355	36%	36%	21%	0%	7%
Fresno_base	96	14	8	0.528	21%	21%	0%	14%	43%
3County_base	195	14	12	0.364	36%	29%	21%	0%	14%
Tampa_base	537	14	13	0.407	43%	36%	14%	0%	7%
Jacksnv_base	357	14	13	0.380	7%	57%	29%	0%	7%

Table 1h

School tour mode									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	1910	38	33	0.251	59%	41%	0%	0%	0%
Calif_base	1527	38	32	0.264	56%	34%	6%	0%	3%
Florida_base	383	38	22	0.267	19%	28%	19%	0%	34%
SANDAG_base	1092	38	32	0.274	59%	28%	9%	0%	3%
SACOG_base	226	38	29	0.224	13%	38%	34%	3%	13%
Fresno_base	77	38	22	0.052	9%	28%	28%	0%	34%
3County_base	132	38	17	0.255	16%	19%	13%	3%	50%
Tampa_base	211	38	22	0.285	13%	19%	34%	0%	34%
Jacksnv_base	172	38	21	0.240	22%	28%	13%	0%	38%

Table 1i

Other tour destination									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	14210	72	62	0.257	88%	12%	0%	0%	0%
Calif_base	10220	72	62	0.276	84%	8%	6%	2%	0%
Florida_base	3990	72	62	0.219	69%	22%	2%	6%	0%
SANDAG_base	7512	72	62	0.273	80%	10%	10%	0%	0%
SACOG_base	1520	72	61	0.293	57%	33%	8%	2%	0%
Fresno_base	434	72	62	0.272	35%	41%	24%	0%	0%
3County_base	754	72	62	0.288	39%	47%	14%	0%	0%
Tampa_base	2674	72	62	0.223	71%	16%	10%	2%	0%
Jacksnv_base	1316	72	62	0.215	51%	35%	12%	2%	0%

Size variables excluded from the significance and estimability columns (the five columns on the right

Table 1j

Other HB tour time of day									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	15377	95	86	0.252	76%	24%	0%	0%	0%
Calif_base	10628	95	86	0.252	71%	26%	3%	0%	0%
Florida_base	4749	95	84	0.251	60%	26%	10%	1%	2%
SANDAG_base	7726	95	86	0.252	66%	30%	3%	0%	0%
SACOG_base	1617	95	84	0.252	33%	57%	8%	0%	2%
Fresno_base	460	95	84	0.231	22%	55%	20%	1%	2%
3County_base	825	95	82	0.243	30%	50%	15%	0%	5%
Tampa_base	3083	95	84	0.248	52%	31%	14%	0%	2%
Jacksnv_base	1666	95	83	0.254	40%	44%	12%	1%	3%

Table 1k

Other HB tour mode									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	12583	41	41	0.358	68%	32%	0%	0%	0%
Calif_base	8546	41	41	0.365	63%	34%	2%	0%	0%
Florida_base	4037	41	40	0.347	61%	24%	12%	0%	2%
SANDAG_base	6226	41	41	0.370	61%	27%	12%	0%	0%
SACOG_base	1320	41	41	0.352	46%	34%	20%	0%	0%
Fresno_base	386	41	38	0.331	37%	37%	20%	0%	7%
3County_base	614	41	40	0.344	41%	29%	27%	0%	2%
Tampa_base	2623	41	40	0.360	54%	32%	12%	0%	2%
Jacksnv_base	1414	41	39	0.322	49%	34%	12%	0%	5%

Table 1l

Intermediate stop generation									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	63630	100	100	0.536	87%	13%	0%	0%	0%
Calif_base	45144	100	100	0.539	81%	14%	5%	0%	0%
Florida_base	18486	100	100	0.528	70%	24%	6%	0%	0%
SANDAG_base	32605	100	100	0.540	77%	15%	8%	0%	0%
SACOG_base	6963	100	100	0.532	54%	36%	9%	1%	0%
Fresno_base	2014	100	99	0.530	40%	48%	11%	0%	1%
3County_base	3562	100	100	0.542	45%	38%	15%	2%	0%
Tampa_base	11727	100	100	0.534	58%	33%	9%	0%	0%
Jacksnv_base	6759	100	100	0.517	55%	37%	7%	1%	0%

Table 1m

Intermediate stop location									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	12140	77	66	0.228	68%	32%	0%	0%	0%
Calif_base	8711	77	66	0.236	77%	23%	0%	0%	0%
Florida_base	3429	77	62	0.213	43%	29%	21%	2%	5%
SANDAG_base	6705	77	66	0.234	70%	27%	4%	0%	0%
SACOG_base	1300	77	64	0.239	41%	43%	13%	4%	0%
Fresno_base	427	77	52	0.222	29%	32%	27%	4%	9%
3County_base	279	77	54	0.224	13%	45%	32%	0%	11%
Tampa_base	2278	77	61	0.201	41%	30%	14%	7%	7%
Jacksnv_base	1151	77	58	0.232	38%	29%	21%	0%	13%
Size variables excluded from the significance and estimability columns (the five columns on the right									

Table 1n

Trip time of day									
Model Spec (mspec)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Coefficients Significant Same Sign	Coefficients Insignificant Same Sign	Coefficients Insignificant Other Sign	Coefficients Significant Other Sign	Coefficients Not Estimable
2State_base	15554	48	45	0.523	76%	24%	0%	0%	0%
Calif_base	10946	48	45	0.518	73%	22%	4%	0%	0%
Florida_base	4608	48	44	0.536	71%	18%	7%	2%	2%
SANDAG_base	7862	48	45	0.517	76%	18%	7%	0%	0%
SACOG_base	1696	48	44	0.515	51%	38%	9%	0%	2%
Fresno_base	507	48	38	0.517	31%	47%	7%	0%	16%
3County_base	881	48	40	0.521	36%	24%	24%	4%	11%
Tampa_base	2887	48	43	0.550	64%	20%	11%	0%	4%
Jacksnv_base	1721	48	42	0.508	60%	22%	9%	2%	7%

Summary Statistics for all Difference Models

Table 2a

Usual work location

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	7526	106	95	0.133	47	36%	64%	0%	100%
2State_base	sacog	7526	106	86	0.129	38	24%	76%	0%	100%
2State_base	fresno	7526	106	92	0.128	44	3%	88%	9%	91%
2State_base	3county	7526	106	84	0.128	36	15%	76%	9%	100%
2State_base	tampa	7526	106	94	0.131	46	42%	55%	3%	100%
2State_base	jacksnvl	7526	106	94	0.131	46	33%	64%	3%	100%
Calif_base	sandag	5584	106	95	0.128	47	27%	73%	0%	100%
Calif_base	sacog	5584	106	86	0.127	38	21%	79%	0%	100%
Calif_base	fresno	5584	106	92	0.126	44	6%	85%	9%	89%
Calif_base	3county	5584	106	84	0.125	36	21%	70%	9%	0%
Florida_base	tampa	1942	106	92	0.154	45	21%	76%	3%	100%
Florida_base	jacksnvl	1942	106	93	0.153	46	21%	76%	3%	100%

Size variables excluded from the significance and estimability columns

Table 2b

Auto ownership

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	12203	48	48	0.311	24	29%	71%	0%	100%
2State_base	sacog	12203	48	48	0.310	24	21%	79%	0%	99%
2State_base	fresno	12203	48	48	0.310	24	21%	79%	0%	98%
2State_base	3county	12203	48	48	0.310	24	8%	92%	0%	99%
2State_base	tampa	12203	48	46	0.315	22	13%	79%	8%	100%
2State_base	jacksnvl	12203	48	48	0.311	24	25%	75%	0%	100%
Calif_base	sandag	8351	48	48	0.302	24	17%	83%	0%	90%
Calif_base	sacog	8351	48	48	0.302	24	21%	79%	0%	91%
Calif_base	fresno	8351	48	48	0.302	24	13%	88%	0%	93%
Calif_base	3county	8351	48	48	0.303	24	17%	83%	0%	99%
Florida_base	tampa	3852	48	46	0.342	22	13%	79%	8%	100%
Florida_base	jacksnvl	3852	48	46	0.342	22	13%	79%	8%	100%
2State_base_ASC	sandag	12203	64	64	0.314	20	25%	75%	0%	100%
2State_base_ASC	sacog	12203	64	64	0.313	20	15%	85%	0%	89%
2State_base_ASC	fresno	12203	64	64	0.313	20	15%	85%	0%	99%
2State_base_ASC	3county	12203	64	64	0.313	20	5%	95%	0%	99%
2State_base_ASC	tampa	12203	64	62	0.315	18	15%	75%	10%	100%
2State_base_ASC	jacksnvl	12203	64	64	0.314	20	20%	80%	0%	100%
Calif_base_ASC	sandag	8351	56	56	0.302	20	15%	85%	0%	93%
Calif_base_ASC	sacog	8351	56	56	0.302	20	20%	80%	0%	95%
Calif_base_ASC	fresno	8351	56	56	0.302	20	15%	85%	0%	97%
Calif_base_ASC	3county	8351	56	56	0.302	20	10%	90%	0%	99%
Florida_base_ASC	tampa	3852	48	46	0.342	18	15%	75%	10%	100%
Florida_base_ASC	jacksnvl	3852	48	46	0.342	18	15%	75%	10%	100%

No ASC differences for cases with ASC model base

Table 2c

Person-day tour generation

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	18770	254	252	0.441	126	7%	93%	0%	93%
2State_base	sacog	18770	254	252	0.441	126	10%	90%	0%	80%
2State_base	fresno	18770	254	248	0.441	122	6%	90%	3%	82%
2State_base	3county	18770	254	248	0.441	122	14%	83%	3%	96%
2State_base	tampa	18770	254	251	0.442	125	10%	89%	1%	94%
2State_base	jacksnvl	18770	254	252	0.441	126	8%	92%	0%	93%
Calif_base	sandag	13268	254	252	0.443	126	7%	93%	0%	88%
Calif_base	sacog	13268	254	252	0.443	126	6%	94%	0%	86%
Calif_base	fresno	13268	254	248	0.443	122	6%	90%	3%	62%
Calif_base	3county	13268	254	248	0.444	122	13%	84%	3%	96%
Florida_base	tampa	5502	254	251	0.434	125	11%	88%	1%	94%
Florida_base	jacksnvl	5502	254	251	0.434	125	11%	88%	1%	94%
2State_base_ASC	sandag	18770	366	364	0.441	98	4%	96%	0%	55%
2State_base_ASC	sacog	18770	366	364	0.441	98	7%	93%	0%	46%
2State_base_ASC	fresno	18770	366	360	0.441	94	6%	90%	4%	71%
2State_base_ASC	3county	18770	366	360	0.441	94	10%	86%	4%	98%
2State_base_ASC	tampa	18770	366	363	0.441	97	12%	87%	1%	100%
2State_base_ASC	jacksnvl	18770	366	364	0.441	98	8%	92%	0%	98%
Calif_base_ASC	sandag	13268	310	308	0.443	98	2%	98%	0%	49%
Calif_base_ASC	sacog	13268	310	308	0.443	98	4%	96%	0%	58%
Calif_base_ASC	fresno	13268	310	304	0.443	94	8%	88%	4%	66%
Calif_base_ASC	3county	13268	310	304	0.443	94	8%	88%	4%	95%
Florida_base_ASC	tampa	5502	254	251	0.434	97	11%	88%	1%	100%
Florida_base_ASC	jacksnvl	5502	254	251	0.434	97	11%	88%	1%	100%

No ASC differences for cases with ASC model base

Table 2d

Exact number of tours

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	20795	196	170	0.654	84	5%	93%	2%	94%
2State_base	sacog	20795	196	161	0.653	75	7%	82%	11%	82%
2State_base	fresno	20795	196	135	0.654	49	0%	60%	40%	36%
2State_base	3county	20795	196	159	0.655	73	12%	75%	13%	100%
2State_base	tampa	20795	196	166	0.653	80	12%	82%	6%	75%
2State_base	jacksnvl	20795	196	162	0.653	76	4%	86%	11%	80%
Calif_base	sandag	14781	196	168	0.657	82	7%	89%	4%	98%
Calif_base	sacog	14781	196	161	0.657	75	5%	85%	11%	78%
Calif_base	fresno	14781	196	135	0.657	49	0%	60%	40%	25%
Calif_base	3county	14781	196	159	0.659	73	14%	73%	13%	100%
Florida_base	tampa	6014	196	156	0.635	73	6%	80%	14%	70%
Florida_base	jacksnvl	6014	196	156	0.635	73	6%	80%	14%	70%
2State_base_ASC	sandag	20795	252	214	0.653	72	7%	92%	1%	94%
2State_base_ASC	sacog	20795	252	205	0.653	63	6%	83%	11%	82%
2State_base_ASC	fresno	20795	252	182	0.653	40	0%	58%	42%	6%
2State_base_ASC	3county	20795	252	204	0.654	62	11%	76%	13%	100%
2State_base_ASC	tampa	20795	252	210	0.653	68	13%	82%	6%	62%
2State_base_ASC	jacksnvl	20795	252	206	0.653	64	6%	83%	11%	74%
Calif_base_ASC	sandag	14781	224	188	0.657	70	7%	90%	3%	99%
Calif_base_ASC	sacog	14781	224	181	0.657	63	6%	83%	11%	80%
Calif_base_ASC	fresno	14781	224	158	0.657	40	0%	58%	42%	6%
Calif_base_ASC	3county	14781	224	180	0.658	62	13%	75%	13%	100%
Florida_base_ASC	tampa	6014	196	156	0.635	61	7%	78%	15%	66%
Florida_base_ASC	jacksnvl	6014	196	156	0.635	61	7%	78%	15%	66%

No ASC differences for cases with ASC model base

Table 2e

Work tour time of day

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	5347	148	136	0.222	67	4%	93%	3%	91%
2State_base	sacog	5347	148	133	0.222	64	0%	93%	7%	1%
2State_base	fresno	5347	148	130	0.222	61	12%	77%	12%	75%
2State_base	3county	5347	148	126	0.222	57	4%	78%	17%	83%
2State_base	tampa	5347	148	136	0.223	67	19%	78%	3%	100%
2State_base	jacksnvl	5347	148	132	0.222	63	1%	90%	9%	42%
Calif_base	sandag	3924	148	135	0.217	66	1%	94%	4%	15%
Calif_base	sacog	3924	148	133	0.217	64	0%	93%	7%	1%
Calif_base	fresno	3924	148	130	0.218	61	7%	81%	12%	63%
Calif_base	3county	3924	148	126	0.218	57	4%	78%	17%	90%
Florida_base	tampa	1423	148	132	0.230	65	9%	86%	6%	28%
Florida_base	jacksnvl	1423	148	130	0.231	63	9%	83%	9%	77%
2State_base_ASC	sandag	5347	272	236	0.221	41	0%	98%	2%	35%
2State_base_ASC	sacog	5347	272	234	0.221	39	0%	93%	7%	1%
2State_base_ASC	fresno	5347	272	231	0.221	36	10%	76%	14%	80%
2State_base_ASC	3county	5347	272	227	0.221	32	2%	74%	24%	63%
2State_base_ASC	tampa	5347	272	236	0.221	41	19%	79%	2%	88%
2State_base_ASC	jacksnvl	5347	272	233	0.221	38	2%	88%	10%	58%
Calif_base_ASC	sandag	3924	210	184	0.216	40	0%	95%	5%	15%
Calif_base_ASC	sacog	3924	210	183	0.216	39	0%	93%	7%	3%
Calif_base_ASC	fresno	3924	210	180	0.217	36	5%	81%	14%	70%
Calif_base_ASC	3county	3924	210	176	0.217	32	2%	74%	24%	67%
Florida_base_ASC	tampa	1423	148	132	0.222	40	10%	86%	5%	0%
Florida_base_ASC	jacksnvl	1423	148	130	0.231	38	10%	81%	10%	85%

No ASC differences for cases with ASC model base

Table 2f

Work tour mode (detailed LOS)

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	5282	120	116	0.529	58	14%	86%	0%	100%
2State_base	sacog	5282	120	111	0.529	53	7%	84%	9%	100%
2State_base	fresno	5282	120	73	0.529	15	0%	26%	74%	47%
2State_base	3county	5282	120	82	0.529	24	3%	38%	59%	85%
2State_base	tampa	5282	120	85	0.531	27	9%	38%	53%	100%
2State_base	jacksnvl	5282	120	79	0.529	21	2%	34%	64%	48%
Calif_base	sandag	3866	120	113	0.513	55	9%	86%	5%	98%
Calif_base	sacog	3866	120	111	0.514	53	7%	84%	9%	100%
Calif_base	fresno	3866	120	73	0.514	15	0%	26%	74%	40%
Calif_base	3county	3866	120	82	0.514	24	0%	41%	59%	90%
Florida_base	tampa	1416	120	48	0.551	21	5%	31%	64%	94%
Florida_base	jacksnvl	1416	120	48	0.551	21	5%	31%	64%	94%
2State_base_ASC	sandag	5282	140	133	0.530	53	9%	91%	0%	100%
2State_base_ASC	sacog	5282	140	128	0.528	48	8%	83%	9%	97%
2State_base_ASC	fresno	5282	140	90	0.529	10	0%	19%	81%	72%
2State_base_ASC	3county	5282	140	99	0.529	19	4%	32%	64%	90%
2State_base_ASC	tampa	5282	140	103	0.531	23	6%	38%	57%	100%
2State_base_ASC	jacksnvl	5282	140	98	0.529	18	0%	34%	66%	61%
Calif_base_ASC	sandag	3866	130	123	0.513	50	6%	89%	6%	96%
Calif_base_ASC	sacog	3866	130	121	0.513	48	6%	85%	9%	92%
Calif_base_ASC	fresno	3866	130	83	0.515	10	0%	19%	81%	64%
Calif_base_ASC	3county	3866	130	92	0.515	19	0%	36%	64%	90%
Florida_base_ASC	tampa	1416	120	48	0.551	18	6%	28%	66%	96%
Florida_base_ASC	jacksnvl	1416	120	48	0.551	18	6%	28%	66%	96%

No ASC differences for cases with ASC model base

Table 2g

Work tour mode (combined LOS)

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	5282	62	62	0.613	31	16%	84%	0%	100%
2State_base	sacog	5282	62	61	0.612	30	10%	87%	3%	100%
2State_base	fresno	5282	62	50	0.610	19	16%	45%	39%	76%
2State_base	3county	5282	62	49	0.611	18	6%	52%	42%	100%
2State_base	tampa	5282	62	52	0.612	21	0%	68%	32%	100%
2State_base	jacksnvl	5282	62	48	0.611	17	3%	52%	45%	97%
Calif_base	sandag	3866	62	62	0.595	31	10%	90%	0%	100%
Calif_base	sacog	3866	62	61	0.595	30	13%	84%	3%	100%
Calif_base	fresno	3866	62	50	0.593	19	10%	52%	39%	72%
Calif_base	3county	3866	62	49	0.594	18	3%	55%	42%	100%
Florida_base	tampa	1416	62	38	0.658	17	0%	55%	45%	55%
Florida_base	jacksnvl	1416	62	38	0.658	17	0%	55%	45%	55%
2State_base_ASC	sandag	5282	82	79	0.614	26	15%	85%	0%	100%
2State_base_ASC	sacog	5282	82	78	0.612	25	8%	88%	4%	100%
2State_base_ASC	fresno	5282	82	67	0.611	14	12%	42%	46%	79%
2State_base_ASC	3county	5282	82	66	0.612	13	8%	42%	50%	100%
2State_base_ASC	tampa	5282	82	70	0.611	17	0%	65%	35%	91%
2State_base_ASC	jacksnvl	5282	82	67	0.612	14	4%	50%	46%	97%
Calif_base_ASC	sandag	3866	72	72	0.595	26	8%	92%	0%	100%
Calif_base_ASC	sacog	3866	72	71	0.594	25	12%	85%	4%	100%
Calif_base_ASC	fresno	3866	72	60	0.593	14	8%	46%	46%	79%
Calif_base_ASC	3county	3866	72	59	0.594	13	4%	46%	50%	100%
Florida_base_ASC	tampa	1416	62	38	0.658	14	0%	54%	46%	69%
Florida_base_ASC	jacksnvl	1416	62	38	0.658	14	0%	54%	46%	69%

No ASC differences for cases with ASC model base

Table 2h

WB subtour generation

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	3447	28	28	0.391	14	0%	100%	0%	47%
2State_base	sacog	3447	28	27	0.392	13	0%	93%	7%	84%
2State_base	fresno	3447	28	22	0.392	8	0%	57%	43%	35%
2State_base	3county	3447	28	26	0.392	12	0%	86%	14%	76%
2State_base	tampa	3447	28	27	0.390	13	0%	93%	7%	16%
2State_base	jacksnvl	3447	28	27	0.391	13	0%	93%	7%	63%
Calif_base	sandag	2553	28	28	0.382	14	0%	100%	0%	55%
Calif_base	sacog	2553	28	27	0.384	13	0%	93%	7%	84%
Calif_base	fresno	2553	28	22	0.384	8	7%	50%	43%	46%
Calif_base	3county	2553	28	26	0.384	12	0%	86%	14%	85%
Florida_base	tampa	894	28	26	0.396	12	0%	86%	14%	30%
Florida_base	jacksnvl	894	28	26	0.396	12	0%	86%	14%	30%
2State_base_ASC	sandag	3447	60	54	0.388	6	0%	100%	0%	46%
2State_base_ASC	sacog	3447	60	53	0.389	5	0%	83%	17%	69%
2State_base_ASC	fresno	3447	60	52	0.388	4	0%	67%	33%	44%
2State_base_ASC	3county	3447	60	53	0.388	5	0%	83%	17%	58%
2State_base_ASC	tampa	3447	60	53	0.388	5	0%	83%	17%	4%
2State_base_ASC	jacksnvl	3447	60	54	0.388	6	0%	100%	0%	46%
Calif_base_ASC	sandag	2553	44	39	0.382	6	0%	100%	0%	68%
Calif_base_ASC	sacog	2553	44	38	0.382	5	0%	83%	17%	58%
Calif_base_ASC	fresno	2553	44	37	0.382	4	0%	67%	33%	44%
Calif_base_ASC	3county	2553	44	38	0.382	5	0%	83%	17%	69%
Florida_base_ASC	tampa	894	28	26	0.396	5	0%	83%	17%	30%
Florida_base_ASC	jacksnvl	894	28	26	0.396	5	0%	83%	17%	30%

No ASC differences for cases with ASC model base

Table 2i

School tour mode

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	1910	76	64	0.258	31	9%	84%	6%	100%
2State_base	sacog	1910	76	62	0.249	29	9%	78%	13%	90%
2State_base	fresno	1910	76	55	0.249	22	3%	63%	34%	88%
2State_base	3county	1910	76	50	0.251	17	3%	48%	48%	98%
2State_base	tampa	1910	76	54	0.259	21	6%	56%	38%	100%
2State_base	jacksnvl	1910	76	54	0.255	21	6%	56%	38%	100%
Calif_base	sandag	1527	76	62	0.261	30	6%	84%	9%	90%
Calif_base	sacog	1527	76	61	0.262	29	9%	78%	13%	96%
Calif_base	fresno	1527	76	54	0.262	22	0%	66%	34%	94%
Calif_base	3county	1527	76	49	0.262	17	3%	48%	48%	82%
Florida_base	tampa	383	76	43	0.264	21	16%	47%	38%	99%
Florida_base	jacksnvl	383	76	43	0.264	21	16%	47%	38%	99%
2State_base_ASC	sandag	1910	96	81	0.267	26	11%	81%	7%	100%
2State_base_ASC	sacog	1910	96	79	0.265	24	11%	74%	15%	97%
2State_base_ASC	fresno	1910	96	72	0.265	17	4%	56%	41%	82%
2State_base_ASC	3county	1910	96	68	0.265	13	0%	46%	54%	55%
2State_base_ASC	tampa	1910	96	72	0.267	17	7%	52%	41%	100%
2State_base_ASC	jacksnvl	1910	96	72	0.264	17	4%	56%	41%	55%
Calif_base_ASC	sandag	1527	86	71	0.262	25	7%	81%	11%	97%
Calif_base_ASC	sacog	1527	86	70	0.262	24	11%	74%	15%	97%
Calif_base_ASC	fresno	1527	86	63	0.261	17	0%	59%	41%	82%
Calif_base_ASC	3county	1527	86	59	0.261	13	0%	46%	54%	55%
Florida_base_ASC	tampa	383	76	43	0.264	17	19%	41%	41%	100%
Florida_base_ASC	jacksnvl	383	76	43	0.264	17	19%	41%	41%	100%

No ASC differences for cases with ASC model base

Table 2j

Other tour destination

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	14210	142	123	0.259	61	43%	57%	0%	100%
2State_base	sacog	14210	142	122	0.257	60	18%	82%	0%	63%
2State_base	fresno	14210	142	123	0.257	61	22%	78%	0%	100%
2State_base	3county	14210	142	123	0.257	61	20%	80%	0%	100%
2State_base	tampa	14210	142	123	0.259	61	37%	63%	0%	100%
2State_base	jacksnvl	14210	142	123	0.258	61	20%	80%	0%	100%
Calif_base	sandag	10220	142	123	0.276	61	27%	73%	0%	100%
Calif_base	sacog	10220	142	122	0.275	60	8%	92%	0%	2%
Calif_base	fresno	10220	142	123	0.276	61	22%	78%	0%	100%
Calif_base	3county	10220	142	123	0.276	61	22%	78%	0%	100%
Florida_base	tampa	3990	142	123	0.221	61	18%	82%	0%	100%
Florida_base	jacksnvl	3990	142	123	0.221	61	18%	82%	0%	100%

Size variables excluded from the significance and estimability columns

Table 2k

Other HB tour time of day

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	15377	190	170	0.252	84	9%	88%	2%	81%
2State_base	sacog	15377	190	170	0.252	84	7%	91%	2%	69%
2State_base	fresno	15377	190	170	0.252	84	5%	93%	2%	81%
2State_base	3county	15377	190	168	0.252	82	5%	91%	5%	92%
2State_base	tampa	15377	190	170	0.252	84	16%	81%	2%	100%
2State_base	jacksnvl	15377	190	169	0.252	83	5%	92%	3%	81%
Calif_base	sandag	10628	190	170	0.252	84	5%	93%	2%	74%
Calif_base	sacog	10628	190	170	0.252	84	3%	94%	2%	37%
Calif_base	fresno	10628	190	170	0.252	84	6%	92%	2%	83%
Calif_base	3county	10628	190	168	0.252	82	5%	91%	5%	67%
Florida_base	tampa	4749	190	167	0.250	83	8%	88%	3%	67%
Florida_base	jacksnvl	4749	190	167	0.250	83	8%	88%	3%	67%
2State_base_ASC	sandag	15377	314	270	0.251	59	8%	89%	3%	70%
2State_base_ASC	sacog	15377	314	270	0.251	59	7%	90%	3%	56%
2State_base_ASC	fresno	15377	314	270	0.251	59	7%	90%	3%	86%
2State_base_ASC	3county	15377	314	268	0.251	57	8%	85%	7%	90%
2State_base_ASC	tampa	15377	314	270	0.251	59	15%	82%	3%	100%
2State_base_ASC	jacksnvl	15377	314	269	0.251	58	5%	90%	5%	66%
Calif_base_ASC	sandag	10628	252	220	0.251	59	7%	90%	3%	72%
Calif_base_ASC	sacog	10628	252	220	0.251	59	5%	92%	3%	21%
Calif_base_ASC	fresno	10628	252	220	0.251	59	8%	89%	3%	91%
Calif_base_ASC	3county	10628	252	218	0.251	57	5%	89%	7%	67%
Florida_base_ASC	tampa	4749	190	167	0.250	58	7%	89%	5%	45%
Florida_base_ASC	jacksnvl	4749	190	167	0.250	58	7%	89%	5%	45%

No ASC differences for cases with ASC model base

Table 2l

Other HB tour mode

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	12583	82	82	0.359	41	7%	93%	0%	100%
2State_base	sacog	12583	82	82	0.358	41	12%	88%	0%	100%
2State_base	fresno	12583	82	79	0.358	38	7%	85%	7%	87%
2State_base	3county	12583	82	81	0.358	40	7%	90%	2%	97%
2State_base	tampa	12583	82	81	0.359	40	15%	83%	2%	100%
2State_base	jacksnvl	12583	82	80	0.358	39	10%	85%	5%	100%
Calif_base	sandag	8546	82	82	0.365	41	5%	95%	0%	100%
Calif_base	sacog	8546	82	82	0.365	41	15%	85%	0%	100%
Calif_base	fresno	8546	82	79	0.364	38	10%	83%	7%	92%
Calif_base	3county	8546	82	81	0.364	40	5%	93%	2%	97%
Florida_base	tampa	4037	82	79	0.347	39	10%	85%	5%	100%
Florida_base	jacksnvl	4037	82	79	0.347	39	10%	85%	5%	100%
2State_base_ASC	sandag	12583	102	102	0.359	36	8%	92%	0%	100%
2State_base_ASC	sacog	12583	102	102	0.360	36	11%	89%	0%	100%
2State_base_ASC	fresno	12583	102	99	0.359	33	8%	83%	8%	75%
2State_base_ASC	3county	12583	102	101	0.359	35	6%	92%	3%	91%
2State_base_ASC	tampa	12583	102	101	0.360	35	14%	83%	3%	100%
2State_base_ASC	jacksnvl	12583	102	100	0.359	34	11%	83%	6%	98%
Calif_base_ASC	sandag	8546	92	92	0.365	36	6%	94%	0%	98%
Calif_base_ASC	sacog	8546	92	92	0.366	36	14%	86%	0%	100%
Calif_base_ASC	fresno	8546	92	89	0.365	33	11%	81%	8%	81%
Calif_base_ASC	3county	8546	92	91	0.365	35	6%	92%	3%	86%
Florida_base_ASC	tampa	4037	82	79	0.347	34	8%	86%	6%	92%
Florida_base_ASC	jacksnvl	4037	82	79	0.347	34	8%	86%	6%	92%

No ASC differences for cases with ASC model base

Table 2m

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
Intermediate stop generation										
2State_base	sandag	63630	200	200	0.536	100	13%	87%	0%	100%
2State_base	sacog	63630	200	200	0.536	100	17%	83%	0%	100%
2State_base	fresno	63630	200	196	0.536	96	9%	87%	4%	100%
2State_base	3county	63630	200	200	0.536	100	10%	90%	0%	100%
2State_base	tampa	63630	200	200	0.536	100	19%	81%	0%	100%
2State_base	jacksnvl	63630	200	200	0.536	100	12%	88%	0%	100%
Calif_base	sandag	45144	200	200	0.539	100	14%	86%	0%	98%
Calif_base	sacog	45144	200	200	0.539	100	15%	85%	0%	100%
Calif_base	fresno	45144	200	196	0.539	96	9%	87%	4%	100%
Calif_base	3county	45144	200	200	0.539	100	9%	91%	0%	100%
Florida_base	tampa	18486	200	200	0.528	100	15%	85%	0%	100%
Florida_base	jacksnvl	18486	200	200	0.528	100	15%	85%	0%	100%
2State_base_ASC	sandag	63630	228	228	0.536	93	14%	86%	0%	100%
2State_base_ASC	sacog	63630	228	228	0.536	93	16%	84%	0%	100%
2State_base_ASC	fresno	63630	228	224	0.536	89	9%	87%	4%	100%
2State_base_ASC	3county	63630	228	228	0.536	93	9%	91%	0%	100%
2State_base_ASC	tampa	63630	228	228	0.536	93	19%	81%	0%	100%
2State_base_ASC	jacksnvl	63630	228	228	0.536	93	12%	88%	0%	100%
Calif_base_ASC	sandag	45144	214	214	0.539	93	15%	85%	0%	97%
Calif_base_ASC	sacog	45144	214	214	0.539	93	14%	86%	0%	100%
Calif_base_ASC	fresno	45144	214	210	0.539	89	9%	87%	4%	100%
Calif_base_ASC	3county	45144	214	214	0.539	93	9%	91%	0%	100%
Florida_base_ASC	tampa	18486	200	200	0.528	93	15%	85%	0%	100%
Florida_base_ASC	jacksnvl	18486	200	200	0.528	93	15%	85%	0%	100%

No ASC differences for cases with ASC model base

Table 2n

Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
Intermediate stop location										
2State_base	sandag	12140	152	131	0.229	65	27%	73%	0%	100%
2State_base	sacog	12140	152	129	0.225	63	30%	70%	0%	0%
2State_base	fresno	12140	152	117	0.198	51	20%	71%	9%	0%
2State_base	3county	12140	152	119	0.225	53	11%	79%	11%	0%
2State_base	tampa	12140	152	126	0.228	60	38%	55%	7%	100%
2State_base	jacksnvl	12140	152	123	0.228	57	18%	70%	13%	94%
Calif_base	sandag	8711	152	130	0.236	64	27%	73%	0%	100%
Calif_base	sacog	8711	152	129	0.231	63	29%	71%	0%	0%
Calif_base	fresno	8711	152	117	0.205	51	18%	73%	9%	0%
Calif_base	3county	8711	152	119	0.226	53	18%	71%	11%	0%
Florida_base	tampa	3429	152	118	0.212	56	18%	68%	14%	100%
Florida_base	jacksnvl	3429	152	118	0.213	56	18%	68%	14%	100%

Size variables excluded from the significance and estimability columns

Table 2o

Trip time of day										
Base Model (basespce)	Difference Region (difregion)	Number of Observations (nobs)	Number of Coefficients Total (totpar)	Number of Coefficients Estimated (estpar)	Adjusted Rho Squared (rhosqadj)	Degrees of Freedom (U - R) (chidf)	Significant Difference Coefficients	Insignificant Difference Coefficients	Difference Coefficients Not Estimable	Chi Squared Probabilty (chiprob)
2State_base	sandag	15554	96	89	0.523	44	22%	76%	2%	99%
2State_base	sacog	15554	96	89	0.523	44	7%	91%	2%	100%
2State_base	fresno	15554	96	83	0.523	38	20%	64%	16%	99%
2State_base	3county	15554	96	85	0.523	40	31%	58%	11%	97%
2State_base	tampa	15554	96	88	0.523	43	27%	69%	4%	99%
2State_base	jacksnvl	15554	96	87	0.523	42	18%	76%	7%	96%
Calif_base	sandag	10946	96	89	0.518	44	18%	80%	2%	100%
Calif_base	sacog	10946	96	89	0.518	44	4%	93%	2%	99%
Calif_base	fresno	10946	96	83	0.518	38	18%	67%	16%	98%
Calif_base	3county	10946	96	85	0.518	40	31%	58%	11%	96%
Florida_base	tampa	4608	96	85	0.535	41	24%	67%	9%	98%
Florida_base	jacksnvl	4608	96	85	0.535	41	24%	67%	9%	98%
2State_base_ASC	sandag	15554	208	176	0.522	20	40%	60%	0%	99%
2State_base_ASC	sacog	15554	208	176	0.523	20	10%	90%	0%	100%
2State_base_ASC	fresno	15554	208	175	0.522	19	40%	55%	5%	99%
2State_base_ASC	3county	15554	208	174	0.522	18	45%	45%	10%	99%
2State_base_ASC	tampa	15554	208	176	0.522	20	45%	55%	0%	100%
2State_base_ASC	jacksnvl	15554	208	175	0.522	19	20%	75%	5%	95%
Calif_base_ASC	sandag	10946	152	130	0.517	20	25%	75%	0%	100%
Calif_base_ASC	sacog	10946	152	130	0.517	20	10%	90%	0%	100%
Calif_base_ASC	fresno	10946	152	129	0.517	19	40%	55%	5%	99%
Calif_base_ASC	3county	10946	152	128	0.517	18	45%	45%	10%	99%
Florida_base_ASC	tampa	4608	96	86	0.535	19	45%	50%	5%	99%
Florida_base_ASC	jacksnvl	4608	96	85	0.535	18	45%	50%	5%	100%
No ASC differences for cases with ASC model base										

Appendix 4: Summary for all Coefficient Types

This appendix provides figures showing summaries of estimability and differences from 2-state base models by type of coefficient at the finest type categorization used in the analysis. Size variables and log size multipliers are excluded from the summaries. In cases where a coefficient is identified by two types, the following order takes precedence in the summaries:

C-logsum
I-impedance
U-land use
W-time window
T-tour/trip
D-day pattern
P-person
H-household
A-constant

Table 1: Number of coefficients by type

Coefficient Type	2-state base models	all base models	difference models
A-constant	152	1368	1884
P-person	154	1386	3878
H-household	83	747	2272
D-day pattern	62	558	1488
T-tour/trip	102	918	2436
I-impedance	94	846	1716
U-land use	76	684	1640
W-time window	45	405	1056
C-logsum	24	216	444
Total	792	7128	16814

Figure 1

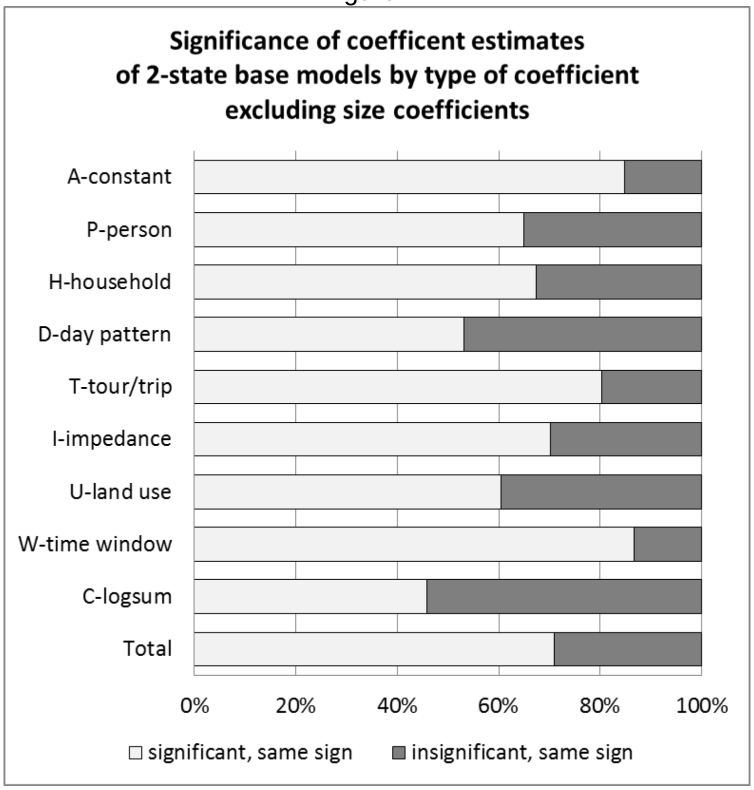


Figure 2

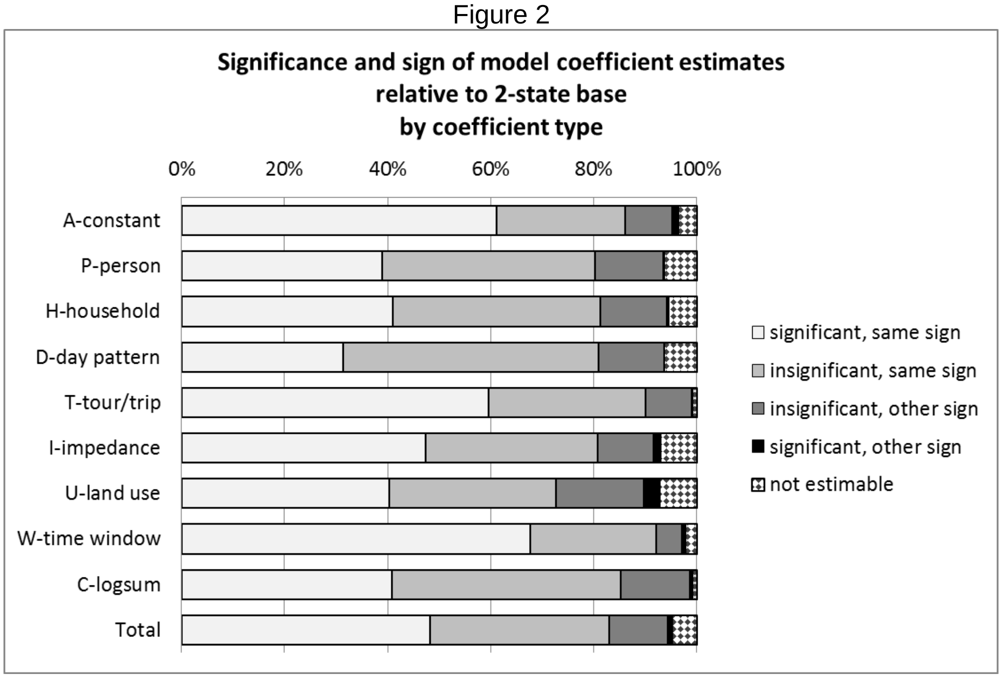
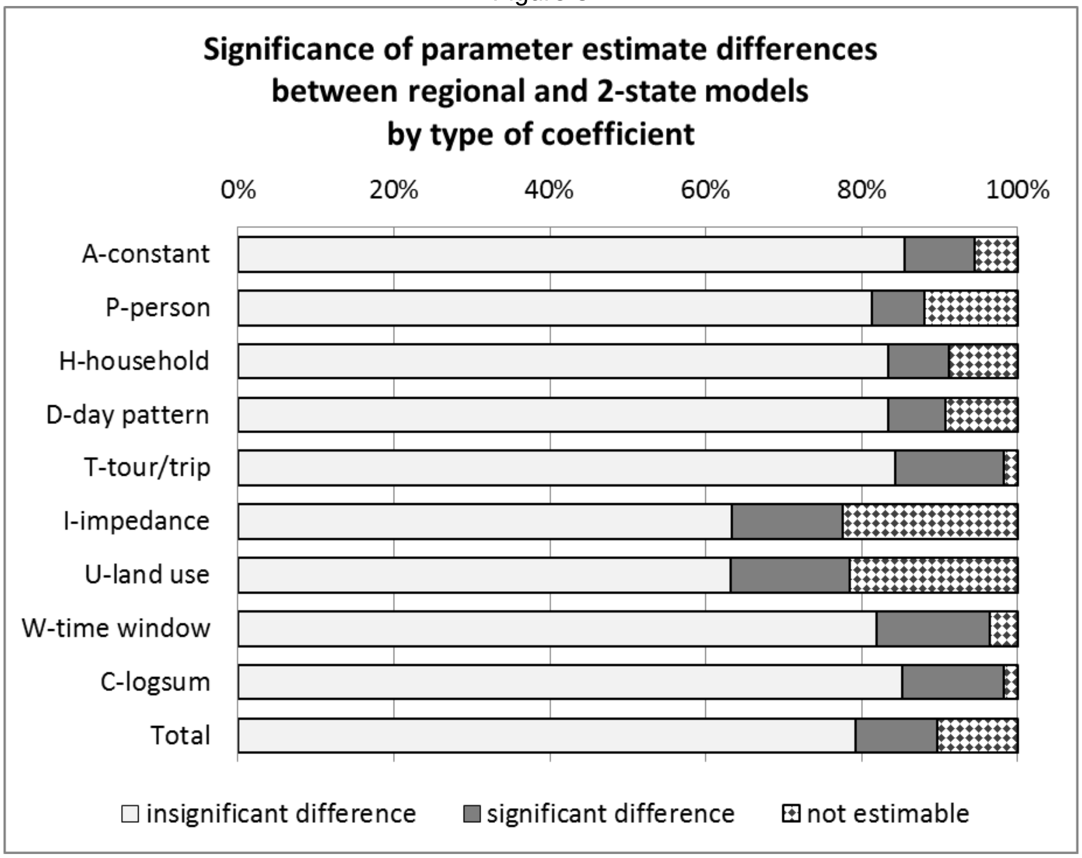


Figure 3


Significance of parameter estimate differences between regional and 2-state models by type of coefficient
0%
20%
40%
60%
80%
100%
A-constant
P-person
H-household
D-day pattern
T-tour/trip
I-impedance
U-land use
W-time window
C-logsum
Total
insignificant difference
significant difference
not estimable

Appendix 5: Feedback from Peer Review Panel

This appendix contains a memo summarizing feedback offered to the project team by the peer review panel at the project peer review meeting. It was drafted by Scott Smith of the U.S. DOT Volpe Center and accepted by the members of the panel. The final report includes edits in response to the panel's recommended changes to the final report. All seven recommendations were addressed except for the literature review, which was excluded due to budget limitations.

Memorandum on the Findings of the Peer Review Panel for Advanced Model Transferability

Peer Review Panelists

Name	**Organization**
David Ory	Metropolitan Transportation Commission (chairperson)
Joel Freedman	Parsons Brinckerhoff
Brian Gregor	Oregon Department of Transportation
Abdul Pinjari	University of South Florida
Wu Sun	San Diego Association of Governments
Scott Smith	Volpe Center (technical support)

This memorandum summarizes the peer review meeting for the report entitled **Making Advanced Travel Forecasting Models Affordable through Model Transferability,** co-authored by John Bowman, Mark Bradley and Joe Castiglione. Both the report and the peer review were sponsored by the Federal Highway Administration (FHWA) Office of Planning. The peer review meeting was held May 14-15, 2013 at the offices of Resource Systems Group in San Diego, and was based on a draft version of the report, dated April 26, 2013. The remainder of this memorandum is organized as follows:

- Questions for the panelists
- Panel findings including strengths and limitations of the report, as well as suggested modifications
- Areas for further research
- An appendix which contains the meeting agenda and list of attendees

<u>Questions for the panelists</u>

Before the meeting, the following questions were presented to panel members:

1**. Critique and make suggestions about method**, so that methodological issues can be documented in the final report and methods can be improved in future transferability studies. Some questions to trigger discussion: How can we better deal with the effect of sample size on the statistical results? What other statistical tests might be useful?

2. **Provide insight about conclusions** that can (or cannot) be drawn from the completed analysis, so that the project team can enhance the Executive Overview and the Conclusions chapter. Questions: How shall we interpret the disappointing results of the Chi square tests? What important conclusions might be apparent from the detailed analysis in chapter 7 that failed to make it into the conclusions and executive overview?

3. **Evaluate understandability** for different audiences and suggest minor editorial changes to the Final Report, to correct errors or improve readability.

4. **Suggest avenues of further research** that would be of most value in the ongoing study of model transferability, with a focus on ways to leverage and extend the work that has been done in this project.

Panel findings

The research that was presented is timely, relevant and useful. Activity-based models can be data-hungry, and mid-size Metropolitan Planning Organizations (MPOs) may find it difficult to obtain enough household survey data. Given limited budgets for survey work and model development, the option of transferring an existing model is often attractive.

The automated, systematic estimation of multiple activity-based models for six regions was a large undertaking, carried out in an efficient manner. It was a comprehensive effort, in that it considered 15 separate models within an activity based model system.

A good set of hypotheses was presented, and a consistent set of models and tools were used for evaluation, considering geographies that are smaller than the Transportation Analysis Zone (TAZ). Finally, the report was well-organized, well-written and clear. It would be easily readable by the intended MPO audience, although they might struggle with the question of how to apply it.

A limitation of the report is that the comparisons as presented would provide little guidance for an MPO who is considering transferring a model. One measure used was the numbers of coefficients of varying types whose parameter estimates were significantly different between a regional model and a 1-state or 2-state model. With more than 800 coefficients in total, such an analysis is concise, but did not provide a deep understanding of the coefficients that were different. It implicitly assumes that all coefficients are equally important, while in reality, some are far more important than others. Questions that might be asked include: What coefficients are most important? How do the sensitivities compare? Which tend to be transferable, and which are not? Simply counting the numbers of different coefficients also makes it difficult to compare this work to previous papers on transferability.

The Chi-squared test for assessing the similarity of two models is not particularly useful, because it almost always shows that two models are different, even though the differences may be practically insignificant. It may be better to use a transferability index, which would provide an indication of the degree to which two models are different.[6]

All of the models used a single specification, a modified version of the Sacramento (SACSIM) specification. Although use of a single specification is important from a comparison standpoint,

[6] For example, see the paper by Sikder, Pinjari, Srinivasan, Nowrouzian, "Spatial Transferability of Travel Forecasting Models," presented at the 4th Innovations in Travel Modeling Conference, Tampa, Florida, and downloaded from http://www.trb.org/conferences/InnovationsinTravelModeling2012.aspx

it leaves open the question of whether some of the transferability results are simply due to the chosen specification. Could transferability be improved by changing the model specification? Also, how well would the existing Sacramento model transfer to the model for Sacramento that was estimated for this project?

The report did not anticipate what readers might criticize, in particular the choice of the modified SACSIM model as a base model specification

Finally, the conclusion, that "although estimation of models using a large local sample is best, it is better to transfer models that are based on a large sample from a comparable region than it is to estimate new models using a much smaller local sample," requires more nuance and caveats. For some special markets, this might not be true. Also, data collection will still be needed, for calibration if nothing else.

The panel recommends the following changes to the final report:

1. Discuss transferability more broadly. Model transferability is a larger topic than similarity of coefficients, and the distinction needs to be clearer. Transferability also has value beyond inexpensive model development. For a region that is rapidly growing or adding new travel options, it might be better to transfer a model from a larger region than to attempt to estimate a model based on the region as it exists now. Some discussion of the role of transferability and data collection (surveys) should be included.
2. Consider adding a literature review. The academic community is not likely to take this work seriously unless a literature review is included. A brief literature review may help to provide context for the two studies that were mentioned near the end of the report, in the section "Comparison To Two Recent Transferability Studies." On the other hand, the panel recognized that an exhaustive literature review is most likely out of scope for this research.
3. Clarify how the logsums are computed.
4. Motivate the selection of base model (modified SACSIM) specification.
5. Summarize results by categories of variables. Table 4.3 presents 11 types of variables, but little use of them is made in the report.
6. Create a data dictionary for excel spreadsheet. The spreadsheet of model results is well organized and should be made publicly available. It will be more useful if a data dictionary explaining the variables is added.
7. Add more nuance to the conclusions, by noting that data collection will still be necessary for calibration, and that the transferability findings may not apply where the regions are significantly different. For example, a university town may have significantly different characteristics from any of the regions studied here.

<u>Areas for further research</u>

The report suggested that the research could be extended by adding other regions where add-on samples are available from the National Household Travel Survey (NHTS). Although there would be some value in creating a multi-state model specification, panel members felt that it should be a higher priority to perform more in-depth analysis on the existing data. Simply adding more regions with small samples would add little value.

It may be helpful to replace the SACSIM model specification with a new specification that is based on the combined 2-state data.

Some areas to explore in a more in-depth analysis using the current data include different measures of transferability, exploration of individual parameters, scale differences, variables that address differences among regions, and weighted estimation.

As noted earlier, the Chi-squared test is of limited use. Other options, such as the transfer index mentioned earlier, would provide more information than a yes-no answer, and should be explored. The transfer index measures the goodness of fit of a transferred model versus the locally specified model. It thus provides an indication of the extent of transferability.

The combined models include well over 800 parameters. Some are, obviously, more important than others in their influence on model predictions. It would be valuable to identify which parameters are the most important, and do some sensitivity testing on them, to determine elasticities. Are the same parameters important across the various models, and do these parameters transfer easily from model to model? What is the sensitivity to policy changes?

Since the utility functions used in these models have no units, coefficients are scaled relative to the residual error term. Differing scales on the regional models could confound the transferability results, and it would be helpful to determine whether there are significant scale differences among the models that were tested.

Region-wide variables could be used to address differences among regions. Examples include the history of growth in the region (very different for, say, Phoenix and Detroit), an overall congestion index, or an index of income that accounts for cost of living (A particular household income might be less than the median household income in San Diego, but more than the median in Fresno). They could help to address the question of the evolution of a region over time: what if one region becomes more like another region?

A weighted estimation approach that controls for differing sample sizes could be explored.

It may be helpful to explore the largest data set (the 6000 households from San Diego) to examine sample size effects, to see what happens to estimability and significance for sample sizes that are between 2500 and 6000. Response curves could be developed for the estimability and significance of coefficients for different models as a function of sample size.

Although the estimation-based approach used in this report allows for explicit statistical tests of the differences in coefficients, it would be helpful to explore the application based approach, and perform sensitivity tests to see if the models differ significantly in their predictions of observed choices. Such an application-based approach is more common in studies of transferability, and more directly relates to the decisions that these models may be supporting.

Moving beyond transferability and the current report, a number of areas were identified as research priorities:

Destination location choice models appear to be the weakest and least transferable in the model chain. Improving the estimation of the usual workplace location would add significant value. Research questions include

- How does sample size affect location choice models?
- How does the choice of geographic subunits, ranging from TAZ to parcel, affect location choice models?
- What can be done with improved land use data, such as assessments or parcel level data from Google?
- How can GPS-assisted travel surveys help?
- What is the dependence between destination choice and mode choice?

A second research interest is to better understand the sources of variation and uncertainty, so that more robust models can be built with respect to the variety of futures that may occur. Questions include

- How can sampling and simulation variation be characterized?
- How can temporal stability be characterized? What happens as an area grows, or its culture changes?

A third research interest is the modeling of special markets (such as university towns), and how insights from them can be applied elsewhere. Under what circumstances is transit oriented development effective (the so-called myth of the transit village)? Universities often have restrictive parking policies; what can be learned from them?

Appendix: Meeting Agenda and Attendees

PEER REVIEW AGENDA		
Day One—Tuesday, May 14, 2013		
8:30	Welcome and Introductions	Brian Gardner, FHWA System Planning & Analysis Team Leader Supin Yoder, FHWA Project Manager
9:00	Project Introduction (report chapters 2-4)	John L Bowman, Principal Investigator
9:30	The Regions and Their Data (report chapter 5 & Appendix 1)	Joe Castiglione, Co-investigator
10:00	Model Estimation and Transferability Testing Approach (report chapter 6 & Appendices 2,3)	Mark Bradley, Principal Investigator
11:30	Results (report chapters 7-8)	John L Bowman and Mark Bradley
12:00	Lunch (working if behind schedule)	
1:00	Results (continued)	
2:30	Current ideas for further research	John L Bowman and Mark Bradley
3:00	Separate closed session for panel	Panel led by panel chairperson
4:30	End of first day	
Day Two—Wednesday, May 15, 2013		
8:30	Panel closed session, final prep for feedback	Panel chairperson and other members
9:00	Presentation of panel feedback	Panel chairperson and other members
10:00	Open discussions	Brian Gardner & Supin Yoder
11:30	Concluding comments	John Bowman, Mark Bradley & Joe Castiglione
12:00	Dismissal	

Attendees:

Name	**Organization**
John Bowman	Bowman Research and Consulting
Mark Bradley	Resource Systems Group
Joe Castiglione	Resource Systems Group
Joel Freedman	Parsons Brinckerhoff
Brian Gardner	FHWA (May 14 only)
Brian Gregor	Oregon DOT
David Ory	Metropolitan Transportation Commission (chairperson)
Abdul Pinjari	University of South Florida
Scott Smith	Volpe Center, US DOT
Wu Sun	San Diego Association of Governments
Supin Yoder	FHWA