How-To: Model Destination Choice

APRIL 2018

U.S. Department of Transportation
Federal Highway Administration

TMIP
Better Methods. Better Outcomes.

<table>
<tr><td>1. Report No.
FHWA-HEP-18-080</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient’s Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
How-To: Model Destination Choice</td><td colspan="2">5. Report Date
April 30, 2018</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Authors
Vince Bernardin, PhD; Clint Daniels; Jason Chen, PhD</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
RSG
55 Railroad Row
White River Junction, VT 05001</td><td colspan="2">10. Work Unit No. (TRAIS)</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTFH61-12-D-00013, T-12002</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
United States Department of Transportation
Federal Highway Administration
1200 New Jersey Ave. SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period Covered
September 2015 – April 2018</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
HEPP-30</td></tr>
<tr><td colspan="4">15. Supplementary Notes
The project was managed by Task Manager for Federal Highway Administration, Sarah Sun, who provided detailed technical directions.</td></tr>
<tr><td colspan="4">16. Abstract
This How-To guide walks the reader through the process of destination choice modeling. The guide describes the various data sources used in destination choice modeling, how the set of potential destinations is defined, how the model is specified and various factors are incorporated, how parameters are statistically estimated, how destination choice models are implemented in the context of larger travel modeling frameworks, and ultimately how they are evaluated and calibrated for use in travel forecasting.</td></tr>
<tr><td colspan="2">17. Key Words
Destination Choice, How-to, Spatial Choice, Trip Distribution, Location Choice, Spatial Interaction, Travel Model, Travel Forecasting</td><td colspan="2">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
79</td><td>22. Price
N/A</td></tr>
</table>

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

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

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

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

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

How-to: Model Destination Choice

April 2018

Federal Highway Administration

Table of Contents

List of Figures

List of Abbreviations

Abbreviations

OD	Origin-Destination
ODME	Origin-Destination Matrix Estimation
TAZ	Travel/Traffic Analysis Zone
TLFD	Trip Length Frequency Distribution
TMIP	Travel Model Improvement Program

1.0 Preface

1.1 Disclaimer

The views expressed in this document do not represent the opinions of FHWA and do not constitute an endorsement, recommendation or specification by FHWA. The document is based solely on the work of RSG and the TRB Travel Forecasting Resource Destination Choice Modeling Charrette.

1.2 Acknowledgments

Early draft materials for this document helped provide initial content which was then further developed by the participants of the Travel Forecasting Resource Destination Choice Modeling Charrette before being revised, edited, and developed into this document. The members of the charrette participants included:

Julie Dunbar (Dunbar)

Eric Miller, PhD (University of Toronto)

Rolf Moeckel, PhD (Technical University of Munich)

Jeff Newman, PhD (Cambridge Systematics)

Ram Pendyala, PhD (Arizona State University)

Rosella Picado (WSP)

Jennifer Weeks (TRB)

These participants generously donated some of their time for this study to develop the charrette materials which were also drawn upon in this document.

2.0 Introduction

2.1 The Problem of Trip Distribution in Space

This How-to guide addresses one of the most challenging problems in travel modeling: trip distribution. Both common experience in practice as well as academic research have found that trip distribution models are the largest source of error in travel modeling systems. (Zhou and Kockelman, 2002). The struggle to model the spatial distribution of trips is clearly understandable. It results from two basic facts. The solution space of origin-destination (OD) flows is very large. Most OD matrices in practice range from 500,000 to 25,000,000 cells, each representing an OD pair. This huge set of dependent variables is modeled using only a very small number of explanatory variables such as the travel time between zones and population and employees per zone without any information about the quality or cost of goods and services provided by these various destinations or the relationships between the individuals or communities living and working in these areas. It is little wonder therefore that distribution models struggle to reproduce the complex patterns observed in reality.

2.1.1 The Importance of the Problem

Reproducing and predicting where travelers are going to and from is of critical importance to travel modeling and transportation planning. Without an accurate representation or prediction of these spatial travel patterns it is impossible to accurately predict many important things including drivers' willingness-to-pay tolls (which depends on the length of their trip) or travelers' likelihood of changing modes to take advantage of a new transit service. Looking to the future, accurate spatial modeling will also be critical to answering new questions such as the likely effectiveness of various types of restrictions on deadheading/zero occupant vehicle trips by autonomous vehicles. Thus, while understanding the spatial patterns of travel is not always of paramount interest in itself, it underlies and is foundational to the ability to understand and predict many issues of interest to transportation.

2.2 Destination Choice Models

Destination choice models are a type of trip distribution or spatial interaction model which are formulated as discrete choice models, typically logit models. This flexible and extensible formulation allows destination choice models to provide a better behavioral basis for trip distribution than the traditional gravity-based trip distribution models, by allowing for a wider range of explanatory variables. Although technically gravity models can be considered a subset or special case of destination choice models, the term "destination choice models" typically is used to identify more general models that incorporate additional variables beyond size/attractions, impedance/friction factors and constants.

Destination choice models consistently reproduce observed travel patterns better than gravity models. Destination choice models perform better through the incorporation of additional variables and reflecting more complex statistical assumptions like spatial autocorrelation. (Bernardin et al., 2009) Logit-based destination choice models are therefore increasingly replacing gravity models for modeling the spatial distribution of trips in order to improve the overall travel model accuracy.

2.2.1 Increasing Use in Practice

As recently as the beginning of 2014, a TMIP survey of transportation agencies around the country confirmed that the majority of travel models still use gravity models for distributing trips in space. However, the same survey also confirmed that the portion of agencies using destination choice models in place of gravity models is increasing and had roughly tripled in less than nine years since a previous survey. As of 2018, destination choice models are now used by almost all of the top twenty-five largest metropolitan areas in the country and in just under half of statewide models (e.g., Arizona, California, Idaho, Iowa, Maryland, Michigan, New Hampshire, Ohio, Oregon, Tennessee, Wisconsin) as well as an increasing number of small and mid-sized metropolitan areas (e.g., South Bend, Evansville, and Columbus, Indiana; Ann Arbor, Michigan; Burlington, Vermont; Knoxville, Nashville, and Chattanooga, Tennessee; Charlottesville, Virginia; Charleston, South Carolina; and Jacksonville, Florida).

2.2.2 Improved Sensitivity to More Factors

The gravity model often exhibits incorrect demand elasticities; in particular, the model may respond illogically to changes in levels of service where improved accessibility to a given destination may cause a disproportionate increase in total trips, and/or an increase in trips using the mode(s) whose accessibility did not change. In both cases, the results are undesirable and may lead to erroneous assessments of the impact of transit or highway improvements.

Destination choice models overcome these gravity model limitations with more appropriate and sophisticated specifications of utility. Because the mathematical form of destination choice utility is very flexible, accounting for the uniqueness in the trip distribution pattern can be accomplished in intuitive ways. For example, modeling a natural barrier like a river in a gravity model usually requires K-factor (or explicit declaration), but, in a destination choice model, a term can be added to the utility equation, statistically estimated from observed data, and interpreted in terms of equivalent minutes of travel time. This latter approach is much more data-based and intuitive measure of the impact the river would have on a person's travel choice. In addition to psychological barriers like this, destination choice models frequently make use of traveler characteristics such as their income or their residence location as important explanatory variables. Walkability, availability and price of parking, accessibility and other variables can further improve the realism of destination choice models. Destination choice models can also incorporate effects such as spatial autocorrelation that simply cannot be incorporated in gravity models.

2.2.3 Improved Explanatory Power and Limitations

While a key advantage offered by destination choice models when compared to the more traditional gravity model is their ability to consider additional factors, at the same time it is also important to recognize destination choice models in practice today still struggle to explain the spatial distribution of personal travel. This is due in large measure to the importance of unobserved attributes such as the price and quality of goods and services provided at destinations.

It is surprisingly challenging to assess the state-of-the-practice in terms of how well (or poorly) gravity models perform and destination choice models outperform them for several reasons. First, from the early years of travel modeling it became frequent practice to only evaluate trip distribution models on the basis of how well they reproduce the observed trip length frequency distribution, rather than the actual observed OD patterns. While this is understandable at some level, given the limited observed OD data available at the time, it has led to the unfortunate situation that very few agencies or consultants developing distribution models even check or report a true goodness-of-fit statistic for spatial distribution models.

Moreover, there is a further technical difficulty that makes it difficult to compare spatial goodness-of-fit statistics across models. The most common spatial goodness-of-fit statistics such as pseudo-r-squared (or rho-squared) or squared error measures, all are dependent on the number/size of zones used. If the whole region is considered as a single zone, any model can perfectly predict the destination zone of internal trips. If the whole region is only represented by two zones, any model should get at least half of the destinations correct. The more zones used to represent the modeled region, the lower the goodness-of-fit measure should be expected. However, this is misleading as a model with more zones may actually be much better at capturing spatial patterns than one with less. This can be verified by comparing the goodness-of-fit of two models of the same region using two different zone structures in which the more detailed one nests within the less. If the goodness-of-fit is calculated for each model using its own zone system, the more aggregate model may well have the higher goodness-of-fit statistic, but if both models are compared using a goodness-of-fit statistic based on the more aggregate zone system, very likely the more disaggregate model will demonstrate the better goodness-of-fit.

For both these reasons it is somewhat difficult to make generalizations about the accuracy of spatial distribution models. However, some very rough generalizations can still be made based on the limited professional experience of the authors, with the caveat that they are dependent on the resolution of zone systems (so models with more zones should expect lower statistics than those with fewer). Assessing goodness-of-fit with rho-squared against the null model of observed household survey data, gravity models commonly explain only between ten and twenty percent of the observed variation in destinations; conventional destination choice models (without substantial fixed factors) often increase the explanatory power over gravity models of the same region by fifty to a hundred percent so they explain fifteen to forty percent of the observed variation in destinations. The struggle of destination choice models to explain observed OD patterns is still not surprising, given the importance of unobserved attributes such as the price and quality of goods and services provided at destinations. In many cases, a conventional destination choice model may have double the explanatory power of a gravity model, but, in the end, still explain less than half of the variation in the observed patterns.

2.2.4 Big Data and Destination Choice Modeling

Very recently, new data-driven modeling frameworks have allowed even more accurate representations of OD travel patterns. These approaches leverage new sources of passively collected, large sample location data (from mobile and in-vehicle devices). In one approach, conventional destination choice models have been incorporated in a larger pivot-point model framework. In another approach, this new information has been incorporated more directly within destination choice models using a constant rich, fixed factor utility specification. This new generation of destination choice models hold great promise for further improving the ability of models to represent and predict travel patterns.

2.3 Overview of this Guide

This guidebook is oriented to the travel modeling practitioner who wishes to develop a solid understanding of destination choice models. Following this introductory section, the remainder of the guide is organized in six sections:

- Data Sources for Destination Choice
- Destination Choice Set Formation
- Destination Choice Model Specification
- Destination Choice Model Estimation
- Destination Choice Model Implementation
- Destination Choice Model Calibration and Validation

Further theoretical discussions regarding destination choice modeling can be found in the appendices.

3.0 Data Sources

While it is possible to represent the selection of trip destinations more rigorously, destination choice models tend to require more data and higher fidelity data than traditional gravity models. Two types of data that are relevant for destination choice models: observed choice data and explanatory data. Observed choice data describe origin-destination flows that have been observed in a survey, by counting, or by passive data collection. Explanatory data, on the other hand, refer to input data that describe either destinations or characteristics of the traveler who chooses the destination.

3.1 Observed Choice Data

Observed choice data describe actual chosen origin-destination pairs. At minimum, these data provide a tally of observed trips at the level of zone-to-zone, origin-destination pairs. In some cases, to support more complex model specifications, surveys provide entire tours or trip-chains together with information on the traveler and specifics of the destination, such as an observation of a high-income worker going from home to work, from work to a restaurant and the restaurant back home. Often, such data are stratified by trip purpose, mode, time of day and various socio-demographic characteristics of the traveler.

For more information on observed choice data, the reader may want to refer to another recent TMIP publication: *Bridging Data Gaps: A TMIP Series on Understanding Origin-Destination Data*. This four-volume series provides a valuable resource with more in-depth information on the various different sources of observed choice or OD data and issues related to their collection, processing, and use for modeling and analysis.

3.1.1 Household Travel Surveys

Up until now the most common source for observed choice data have been household travel surveys. Origins and destinations are collected at the address or latitude-longitude level and translated into TAZ for data analysis and modeling. Long-distance data commonly are provided at a coarser geography like counties or metropolitan areas. Surveys have the benefit that they tend to provide rich information on the socio-demographic characteristics of the traveler as well as the purpose of the trip, mode of travel, and other contextual information. In addition to individual trips, surveys also commonly allow the analyst to identify entire tours. However, due to cost and respondent burden, rich survey data only covers a small sample of all OD pairs constituting OD space.

In larger regions, the sample size for a regional household travel survey is often between 0.4% to 0.6% of the region’s households (e.g., 4,000 to 6,000 households for a region with 1 million households). Smaller regions are more likely to have modestly higher sample sizes as a percentage of the region’s households. Participating households often report all travel for an assigned period, typically one weekday—usually in the spring or the fall to avoid the “atypical” summer and winter vacation/holiday periods. More recently, surveys conducted by smartphone have typically been

By way of example, a survey using traditional methods that obtains 5,000 households, with 2.5 people per household and 4 trips per person-day, will result in around 50,000 individual trip records. A typical MPO might have 2,000 traffic analysis zones (TAZs) in its travel demand model, producing 2,000 x 2,000, or 4 million possible origin-destination (OD) pairs. So, even if every survey trip was between a new OD pair, the survey would only cover 50,000/4,000,000, or 1.25% of all possible OD pairs. In reality, many MPOs include more than 2,000 TAZs and many survey trips are between the same OD pairs. As a result, the OD data from a household travel survey may often cover well below 1% of all possible OD pairs in a region.

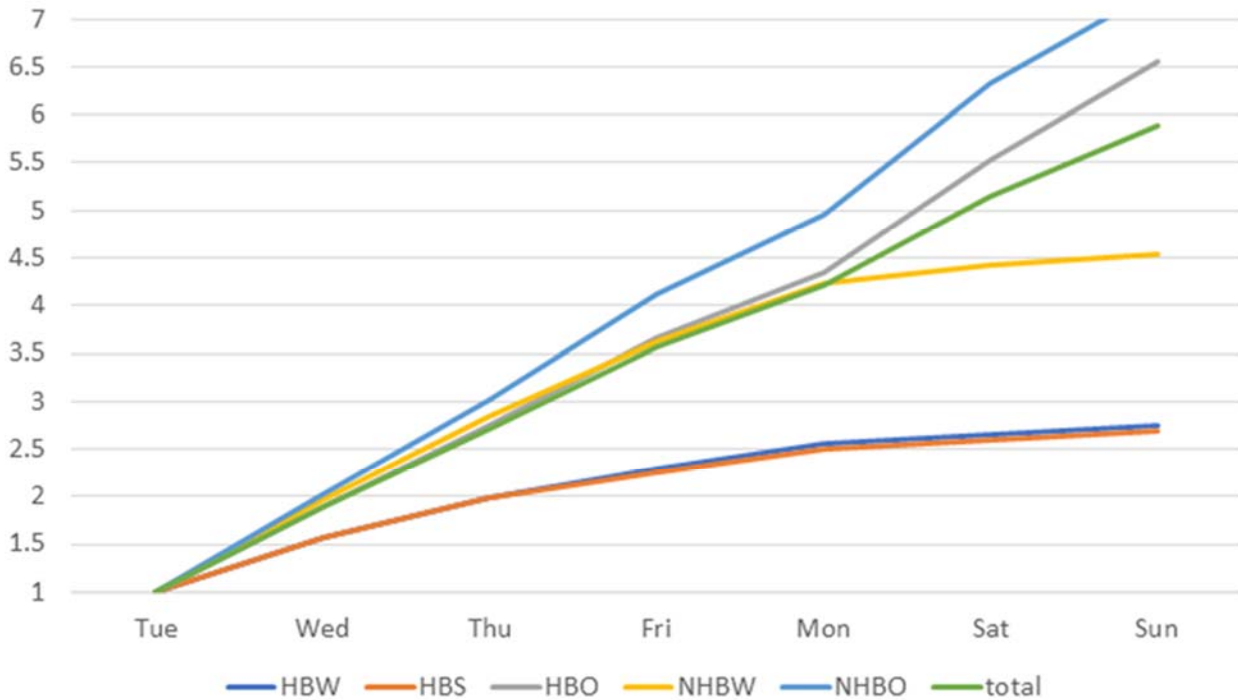


Figure 1. Cumulative number of unique trips as a multiple of Day 1 trips, based on SANDAG smartphone-based GPS travel survey data.

Source: FHWA

Additional days of data from smartphone surveys can help provide greater coverage to some extent, but it is still limited compared to fully passive data collection methods. Smartphone surveys, however, provide several other advantages as well. Use of smartphone apps for data collection improves accuracy of trip-end locations and time, provides data on routes used, increases willingness of younger travelers to participate, reduces respondent burden, and decreases trip under-reporting and other recall related problems. However, survey collection by smartphone must often be augmented by other response options such as online or phone since roughly two in ten adults (and a higher percentage of seniors) do not own smartphones. Smartphone apps can also be limited by different smartphone features, operating systems, and marketplaces.

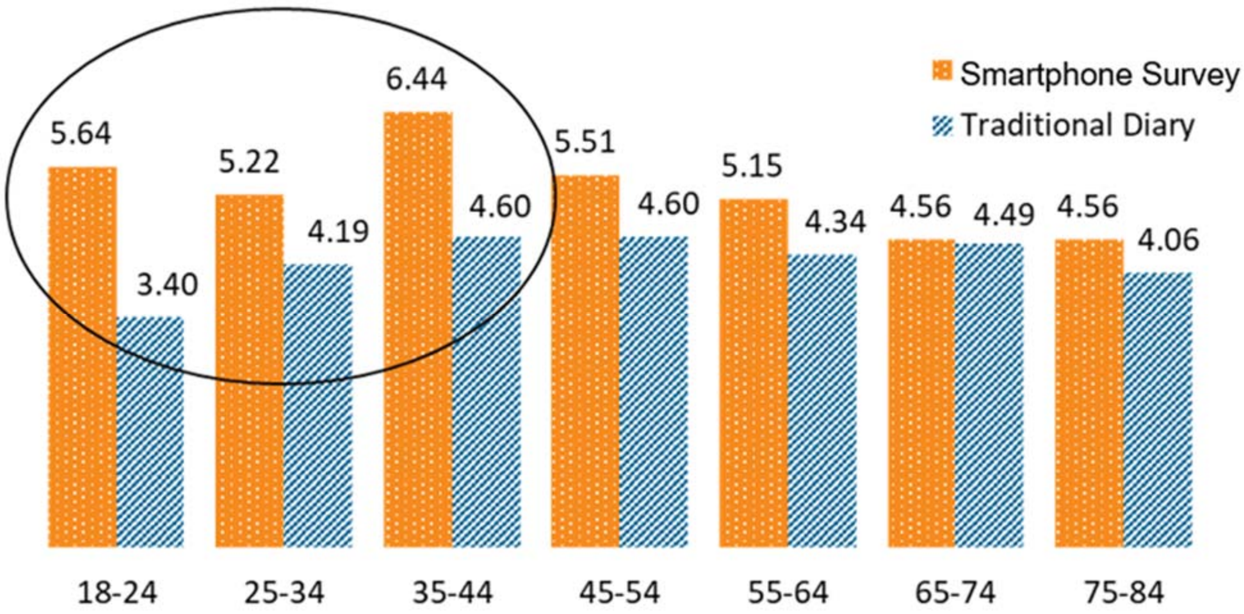


Figure 2. Trip rates by age from Raleigh-Durham, NC, showing under-reporting in traditional surveys.

Source: FHWA

3.1.2 U.S. Census Bureau Data

The U.S. Census bureau provides two data products with information on observed choice data, but only for work commute trips. These products are often used as supplementary or secondary data sources used for validating and sometimes calibrating, but not estimating, destination choice models.

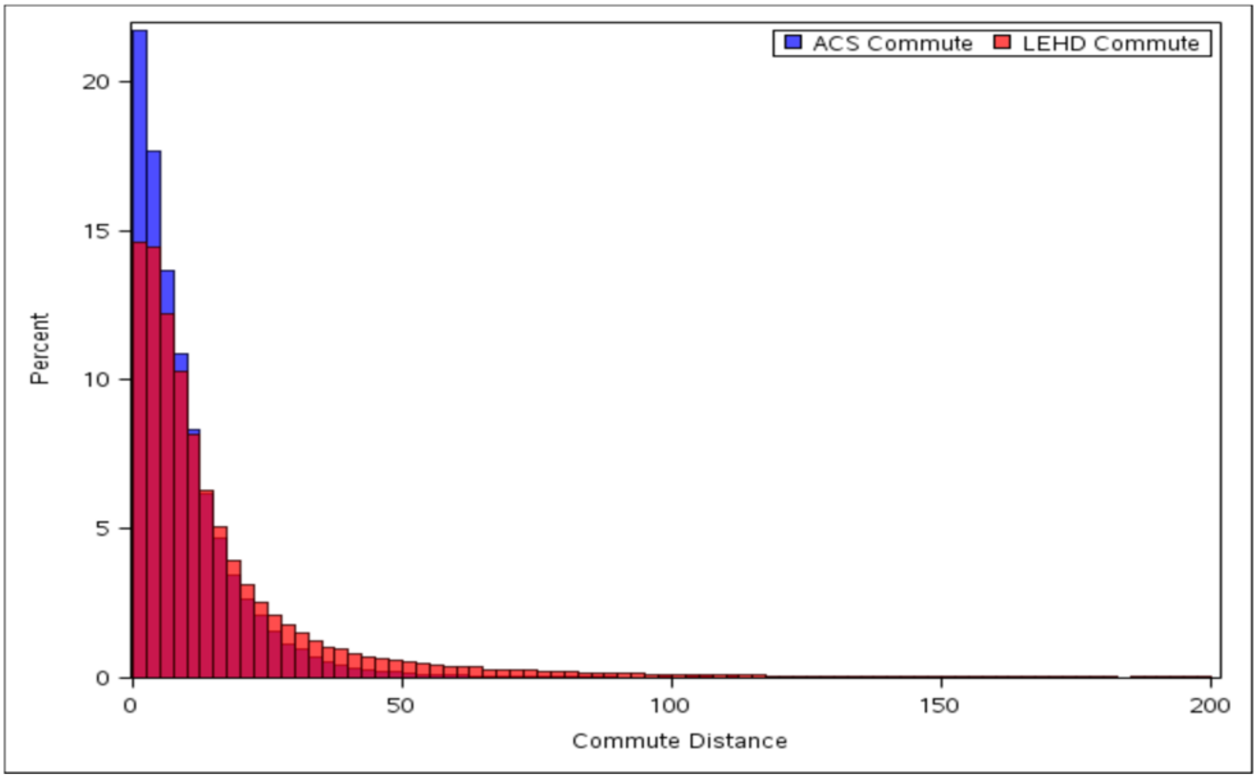


Figure 3. Comparison of work trip-length frequencies from LEHD and CTPP (source: Green et al., 2007).

The Census Transportation Planning Products Program (CTPP) collects data on work trips as part of the Census' American Communities Survey (ACS). CTPP is co-sponsored and hosted by the American Association of State Highway and Transportation Officials (AASHTO) in cooperation with the states. The dataset includes OD flows for various geographies from states and counties to block groups (or Census TAZs although these are being discontinued). The dataset also provides information on commuters like age, household structure, and income as well as the time of their commute and usual mode. Every year the ACS samples slightly over 1.5% of households, hence five-year ACS estimates are based on roughly a 7-8% sample. This number is much larger than any household survey, but still a modest sample compared the universe of administrative records used for the Census's LEHD data product.

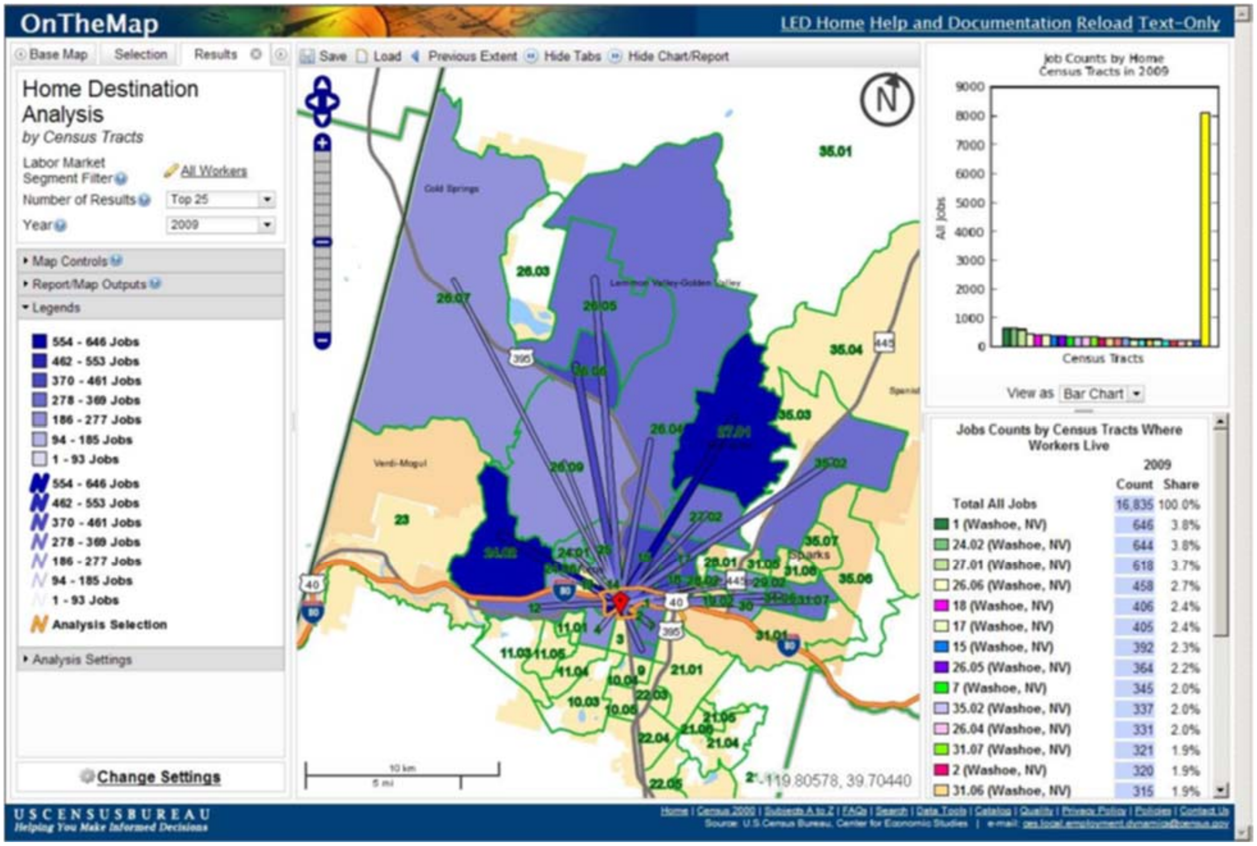


Figure 4. Screenshot of LEHD OnTheMap website visualization.

Source: US Census Bureau

The Longitudinal Employer-Household Dynamics (LEHD) is a joint project of the U.S. Census Bureau, the U.S. Bureau of Labor Statistics, and state employment security agencies, published and hosted by the Census Bureau. LEHD additionally offers the LEHD Origin-Destination Employment Statistics (LODES). These data provide commuter flows at the resolution of census blocks. Flow data are available segmented by three age groups, three income groups, and multiple industrial classifications.

The program uses administrative records from payroll taxes used for unemployment insurance and the quarterly census of employment and wages. Thus, the data is far more complete than any other dataset on commute flows. However, the data does not cover all workers; sole proprietors, railroad workers, and other special groups not covered by unemployment insurance are not included in the data. These exempt groups account from 5 to 20% of all workers in different regions and may be one reason for the under-representation of short commute trips in the LEHD data. The data also suffers from the “headquartering problem” where employers with multiple work sites may report all workers in one location. While the data is processed in attempt to address this, current methods have not been able to fully address this issue. This may also partially account for the under-estimation of short commute trips.

A process of disclosure proofing is also applied in which noise is intentionally introduced in the data at low levels of geography in order to protect the confidentiality of workers and firms. While the method is believed to protect data integrity at higher, more aggregate levels of geography, it introduces some error that can be difficult to correct for in some cases.

3.1.3 Passively Collected Location Data

In contrast to survey data, passively collected data do not ask people about their travel behavior explicitly but rather collect data passively through cellular phones, GPS devices, or other location-revealing technologies. While these data in themselves do not provide traveler characteristics or contextual information on trips (e.g., mode, purpose), these data have proven to be powerful because of their magnitude of coverage.

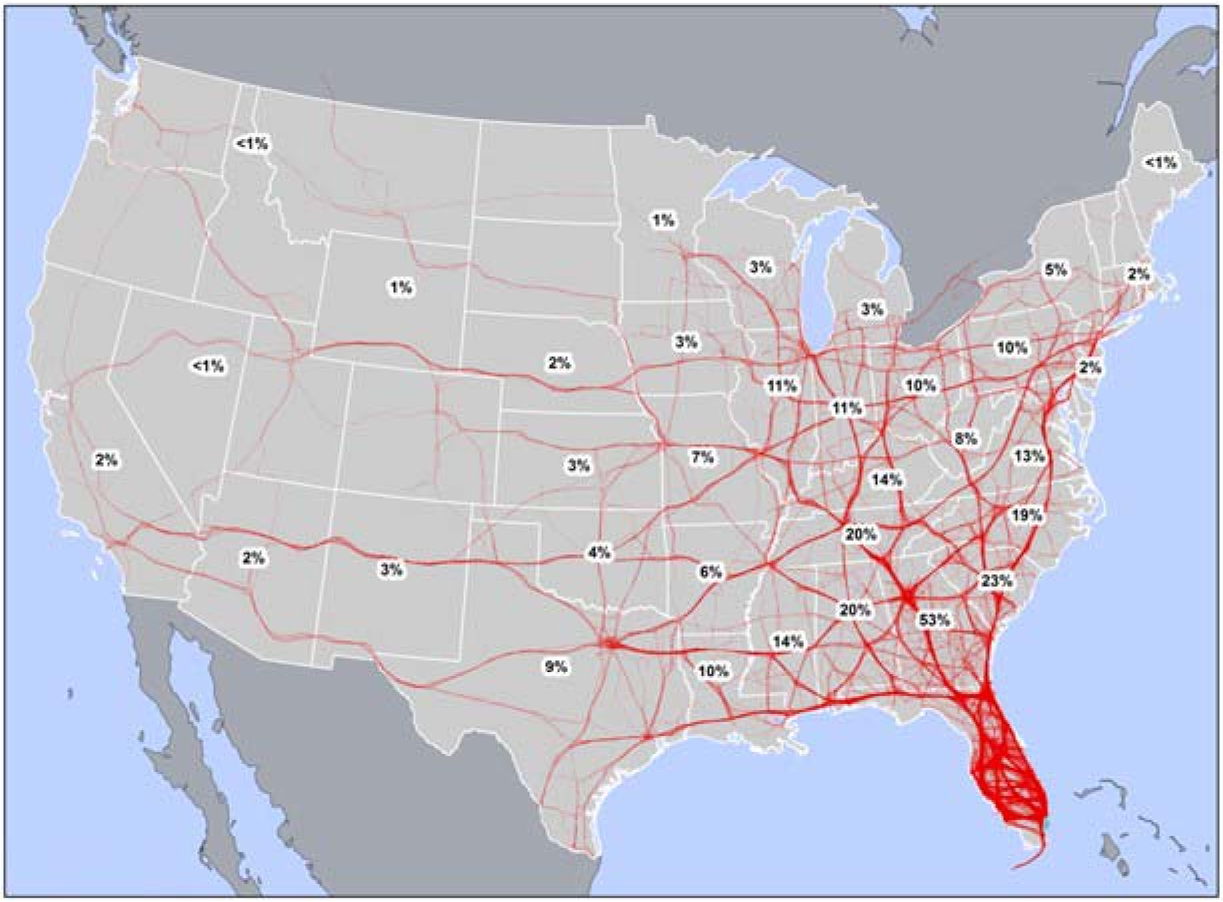


Figure 5. Passively collected truck GPS flows to/from Florida.

Source: ATRI

While traditional surveys often cover roughly one percent of the population for one to seven days, it is not uncommon for passively collected data to cover ten percent of the population for a month or more. These larger samples result in more comprehensive coverage of origin-destination pairs. Whereas surveys often provide observations on 1% or less of OD pairs, passive data often provides observations on a quarter or a third of possible OD pairs. In a Tennessee household travel survey, for example, the survey covered over 10,000 households generating approximately 40,000 origin-destination pairs in the statewide zone system, which is only 0.3 percent of all possible origin-destination zone pairs in Tennessee. Cell phone data, on the other hand, was able to capture 26 percent of all origin-destination zone pairs. Many origin-destination pairs particularly between rural areas have no actual travel to observe. Hence, it is believed that cell phone data was able to capture the majority of origin-destination pairs that are actually traveled. The almost complete coverage has important benefits for the estimation of destination choice models.

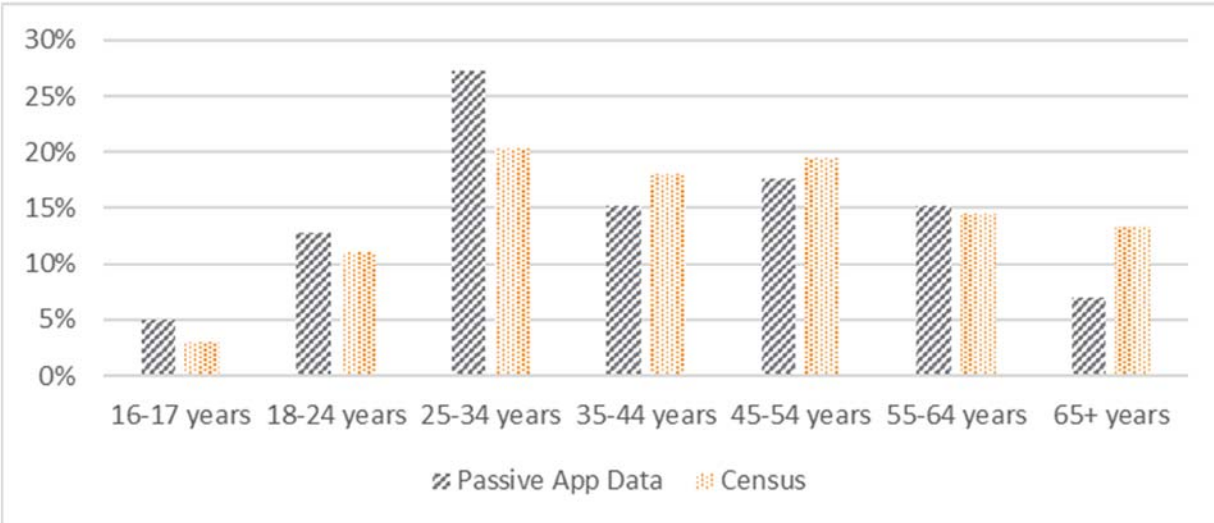


Figure 6. Systematic age bias in passive app data in Columbus, OH.

Source: FHWA

The two main disadvantages of passive OD data are the lack of choice-maker/traveler characteristics and other contextual information and systematic bias in the way the convenience sample represents the actual universe of all travel. The lack of mode and purpose information are particularly problematic for most modeling. Also, although passive data provides large sample data, often including millions of trips, it is still only a sample, and because it is not a controlled random sample, it is not representative of all travelers or trips. Commercially available datasets include only travelers with certain devices, carriers, and/or apps installed. Seniors and low-income populations are known to be under-represented in many datasets. Moreover, short-distance trips or short-duration activities are often under-represented in the data because they require more frequent observations of position which are not always available due to several factors including battery management, device and app usage. Passive OD data must be expanded to correct for these biases in order to properly and accurately represent OD patterns for destination choice modeling, and due to the lack of explanatory variables, the data must be used together with other datasets to support destination choice model estimation.

3.1.4 Traffic Counts

Traffic counts also provide valuable, albeit incomplete, information on origin-destination flows. Traffic counts commonly are provided by the local transportation engineer or Metropolitan Planning Organization. For instance, traffic counts along screenlines can provide information on aggregate district-to-district flows and are commonly used for this reason in destination choice model validation.

Traffic counts can also be used directly in model estimation to estimate model parameters simultaneously with survey data. Traffic counts may be used to impute origin-destination flows. In origin-destination matrix estimation (ODME), sometimes also called synthetic matrix estimation (SME), a trip matrix is synthesized that matches traffic count data.

ODME is a method to create a synthetic trip tables that resembles count data after assignment. (Willumsen, 1981) Such models often suffer from unexpectedly large differences in outcomes due to small changes in inputs (Aerde, 2003) as well as their inability to reconcile inconsistent or erroneous traffic counts. (Hazelton, 2003) As traffic counts do not distinguish trip purposes or user classes, ODME cannot provide trip tables by purpose or trip tables that distinguish travelers by socio-economic characteristics. Moreover, proper and responsible application of ODME reflects the importance of the initial seed distribution of OD patterns required by ODME and limits its distortion either through the formulation of the objective function for the optimization or the imposition of constraints. While in the past trip matrices generated with ODME flows were often only used if no other origin-destination data sources are available, the availability of good seed OD patterns from passive big data may now present a better foundation for ODME.

3.1.5 Other Sources of OD Data

The foregoing data sources are most commonly used for the estimation, calibration, and validation of destination choice models. However, OD data can occasionally also be provided by other data sources and used to support destination choice modeling in some contexts. Special surveys such as external cordon line surveys, visitor surveys, on-board transit ridership surveys, and roadside intercept surveys all can provide OD information that can be particularly valuable for special market segments in some areas. Establishment surveys can also provide valuable information although they typically only capture revealed choices of destinations without information on trips origins.

3.1.6 Visualization

While OD data does not necessarily have to be visualized to support destination choice model estimation, visualizations can be helpful in cleaning and validating the data itself as well as calibrating and validating models based on it. While matrices can convey OD patterns for the technically or numerically oriented, visualizations are often helpful.

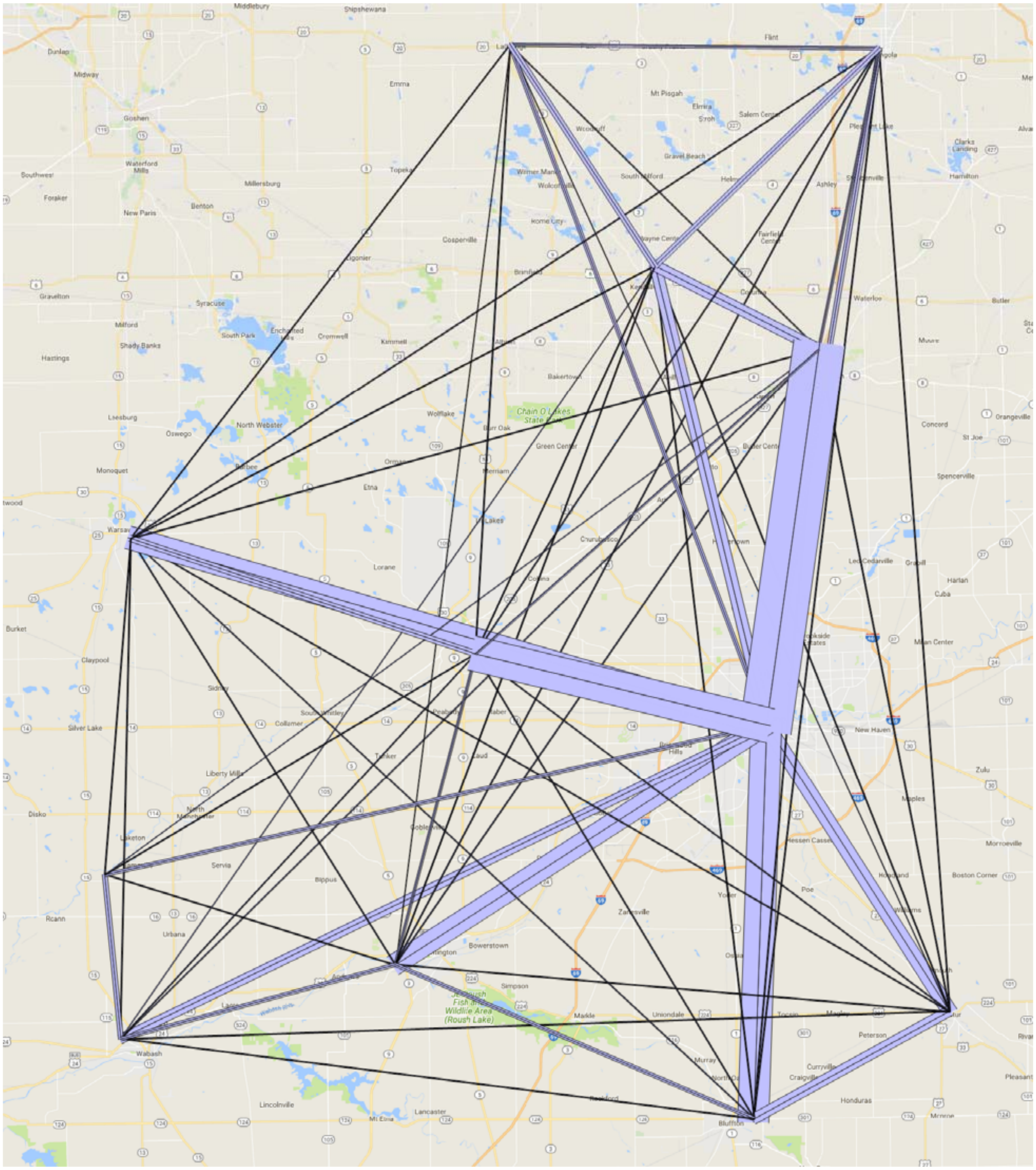


Figure 7. Desire lines for Northeast Indiana.
Source: FHWA

Desire lines are perhaps the most common visualizations of OD patterns, particularly at more aggregate levels; however, they are still often not easy to understand for those who are not familiar with them. Chord diagrams are another helpful way for visualizing OD patterns which have been becoming more popular recently.

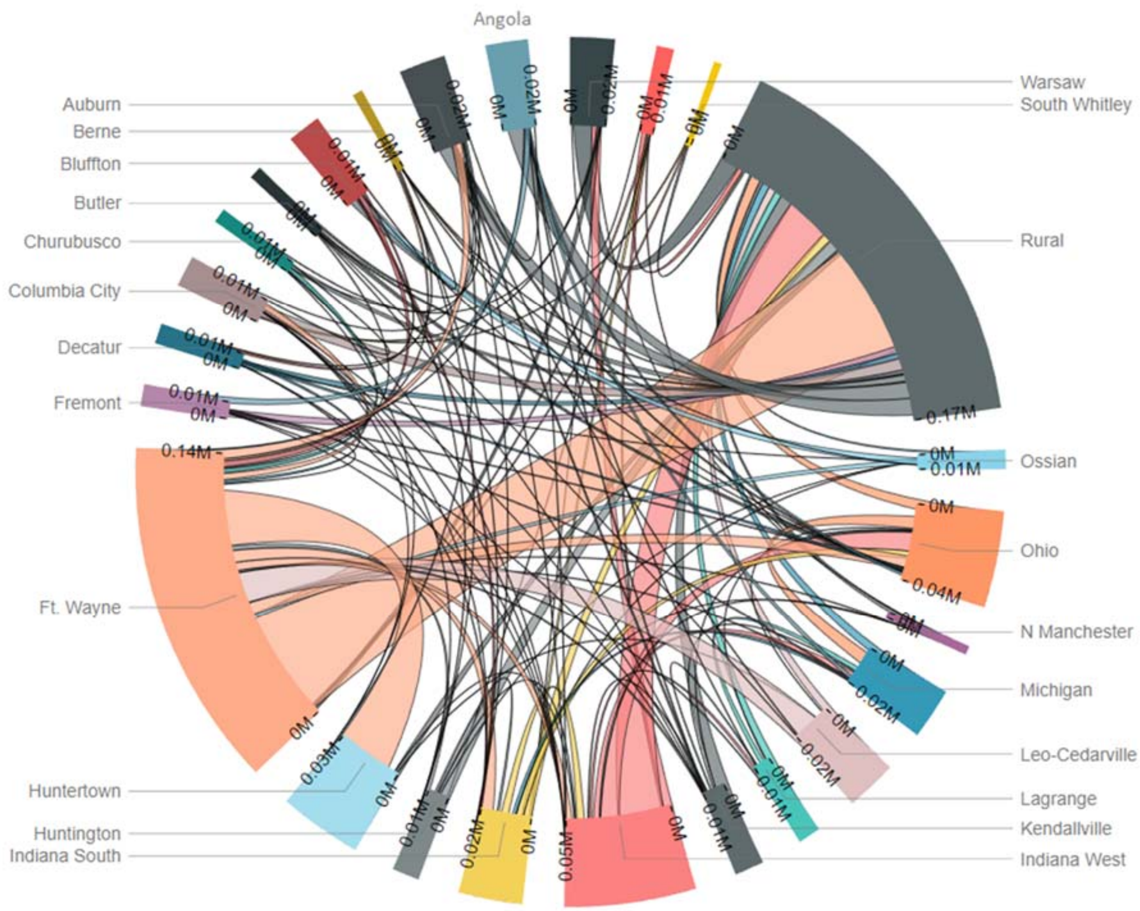


Figure 8. Chord diagram illustrating OD patterns in Northeast Indiana.

Source: FHWA

3.2 Explanatory Data

In addition to observed choice data, destination choice models also need information on potential destinations, such as retail facilities, parks or hotels, and the corresponding impedance to or difficulty of getting there. Similarly, information about the travelers, such as age, sex or income, are relevant when estimating destination choice models. These data often are called explanatory data.

3.2.1 Impedance Measures: Travel Time and Cost Data

Destination choice model requires information on the impedances or difficulty getting between zones. The impedance is commonly calculated as travel time, travel distance, or travel costs. A weighted combination of these three variables called generalized cost is also often used as impedance in models. The relative weight on time and cost imply a value of time (VOT). Matrices of shortest paths based on one of these impedance variables are a key input to destination choice models. In many cases, the variable used for finding shortest paths is the same as used in the destination choice model, but in some cases, a "skimmed" variable is used in the model, for instance, the travel time along the least generalized cost path.

Travel time, itself, is often comprised of several components. Although in-vehicle travel time is often dominant, terminal times for automobile trips and access, wait, and transfer times for transit are often important. Especially for transit impedances, these variables are often weighted up in relation to in-vehicle time.

Travel costs may include roadway tolls, parking, fuel costs, maintenance costs, and/or transit fare.

Trip distance is usually measured along the least generalized cost path (rather than point-to-point). In some cases, distance on higher and lower-class facilities is weighted differently to account for limited knowledge of lower class facilities or driver preference for design characteristics (e.g., wider lanes, faster speeds) of higher class facilities.

Mode Choice Logsums

The logsum or expected disutility of a mode choice model is sometimes also used as an impedance variable in destination choice models. This is done in order to make the models sensitive to impedances across several modes (as an alternative to weighted average impedances across modes or other "composite impedance" mechanisms). When this is done, under certain circumstances (which rarely actually obtain) the result is equivalent to a nested logit model of destination and mode choice. However, because mode choice is usually modeled after and conditional on destination choice when the data does not support this (Newman and Bernardin, 2010), the nesting parameter or coefficient on the mode choice logsum variable in the destination choice utility often must be asserted and constrained, and a second impedance variable such as distance used in order to produce a model that can replicate trip length frequencies. Given both the collinearity of these variables and since distance is often a component of the mode choice logsum, this is a potentially problematic utility specification that may result in unrealistic model sensitivity and responses to changes in travel time.

Transformations and Other Non-linear Techniques

Logarithmic and sometimes other transformations (polynomial expansions) are often used to transform travel time, distance or generalized cost as an impedance variable. Using the log puts more emphasis on differences between destinations that are close. For example, if one grocery store is 6 minutes away and another one is 10 minutes away, the difference of 4 minutes may be important for the trip maker. On the other hand, a grocery store that is 30 minutes away is not perceived as being as much further away than another one that is 26 minutes away. Even though, the difference is the same, a difference of 4 minutes is perceived to be more relevant for short-distance trips than for longer distance trips. The log-transformation nicely accounts for this perception.

Spline variables are also sometimes used in which different (decreasing) marginal impedance is applied to increasing ranges of an impedance variable. While this captures and can also reflect the same effect of decreasing sensitivity to impedance at greater impedances, the discontinuities it introduces in the utility and log-likelihood functions are reason to prefer log-transformations.

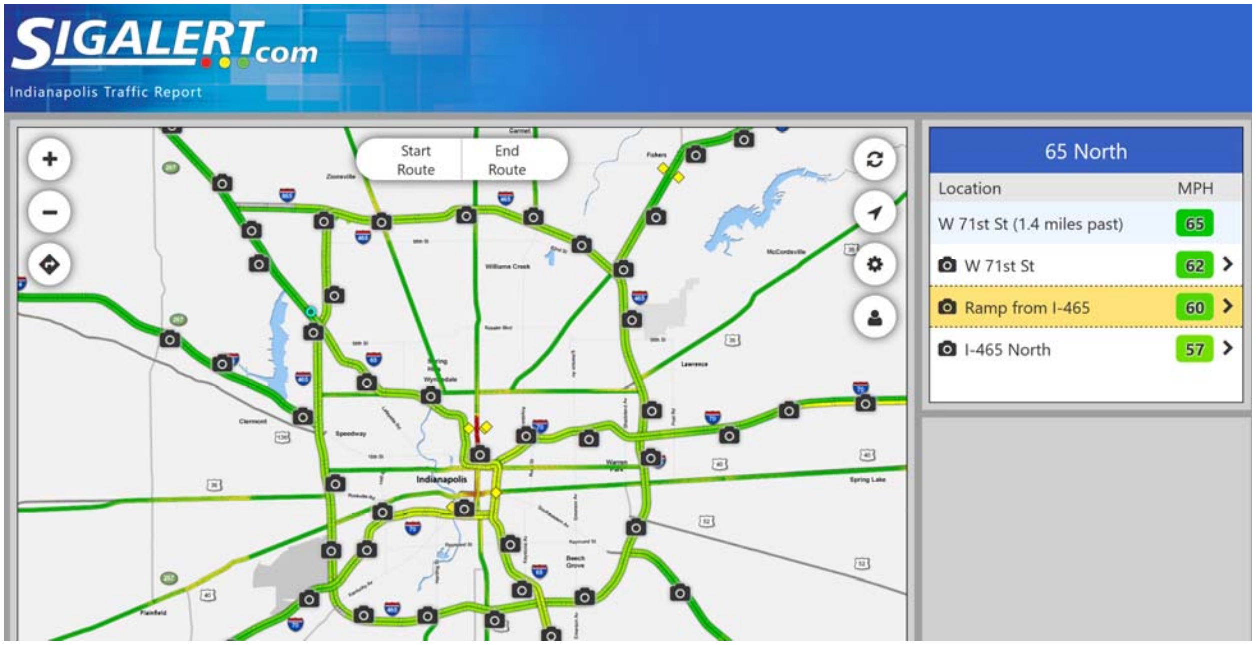


Figure 9. Real-time travel time data.

Source: Indianapolis Traffic Report

Online Sources for Impedances

Most commonly, impedances are skimmed from the network of a travel demand model. In some cases, however, such data are not available from the model, particularly when the model is still under development. Alternatively, these data may be acquired from data vendors and/or online (although users should be careful to observed terms of use for online resources). More commonly, online data sources are used for limited validation of travel times from model networks.

3.2.2 Land Use, Employment, and Demographic Data

Candidate destination zones are most commonly characterized using land use or socio-economic data such as the population and employment of the zones. Employment often is distinguished by industry sector like manufacturing, retail, and office. For example, shopping trips are mostly attracted by retail employment, while trips for visiting friends and family most frequently are attracted by population. An important limitation of these data is that categories tend to be relatively broad, part of necessity given the limited ability to forecast land use by detailed categories. Retail employment, for example, includes destinations as diverse as bakeries and car dealers, two very different retail facilities that in reality would attract very different trips. Further, is has been shown that larger facilities tend to attract more trips per employees than smaller facilities. Nevertheless, zonal land use data are the most common data source for modeling trip destinations.

Zonal population data usually are derived from Census Data that are provided at the block group level. Population forecasts for future years are either developed manually (e.g., scenario planning) or forecasted using land use models.

Zonal employment data commonly are developed from business registration data like the Quarterly Census of Employment and Wages (QCEW) or LODES. Clean-ups are necessary, because firms are commonly registered at their main site (or headquarters), and different branches are not coded explicitly. Alternatively, commercial databases like InfoUSA, Dun & Bradstreet, or Woods & Poole have been used, but suffer from many similar problems as public employment registries. Using both a public and a proprietary dataset together, however, has been shown to result in much more complete and accurate estimate of employment since their errors are not highly correlated.

3.2.3 Accessibility

Accessibilities describe the ease of travel between a zone to all other destination zones. Rather than an additional independent source of data, they are generally computed by combining travel time/impedance data and employment/land use or POI data. Accessibilities can be used in the utility functions of destination choice models to capture many important phenomena including spatial auto-correlation, trip-chaining efficiencies, and differential willingness-to-travel.

A special form of accessibilities is destination-mode choice logsums that combine travel times by various modes across destinations. Other multi-modal accessibilities can be produced by weighting mode-specific impedances by the share they are used by a particular user class, which allows to better represent the relevance of transit access for low-income households.

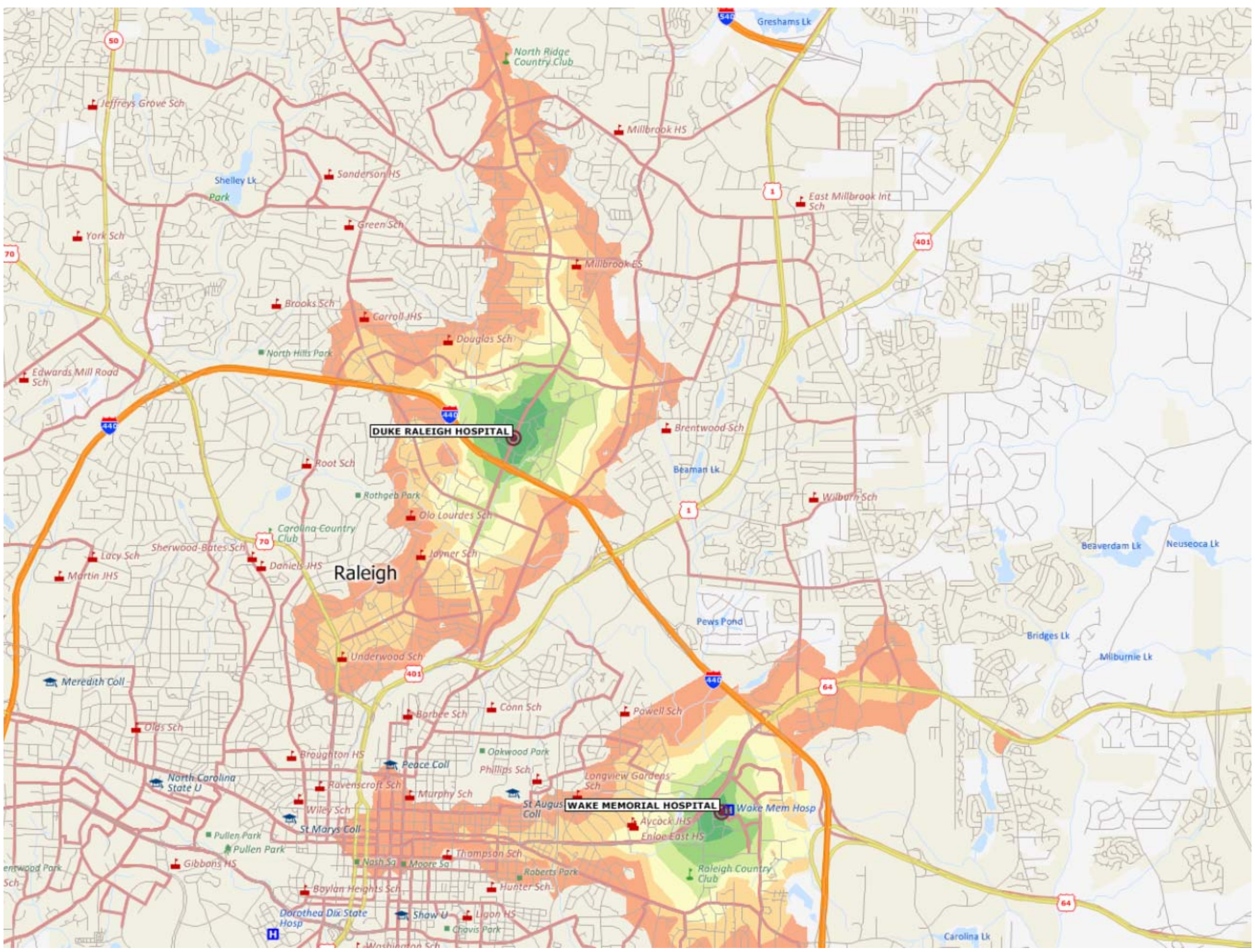


Figure 10. Transit accessibility to Hospitals in Durham, NC.

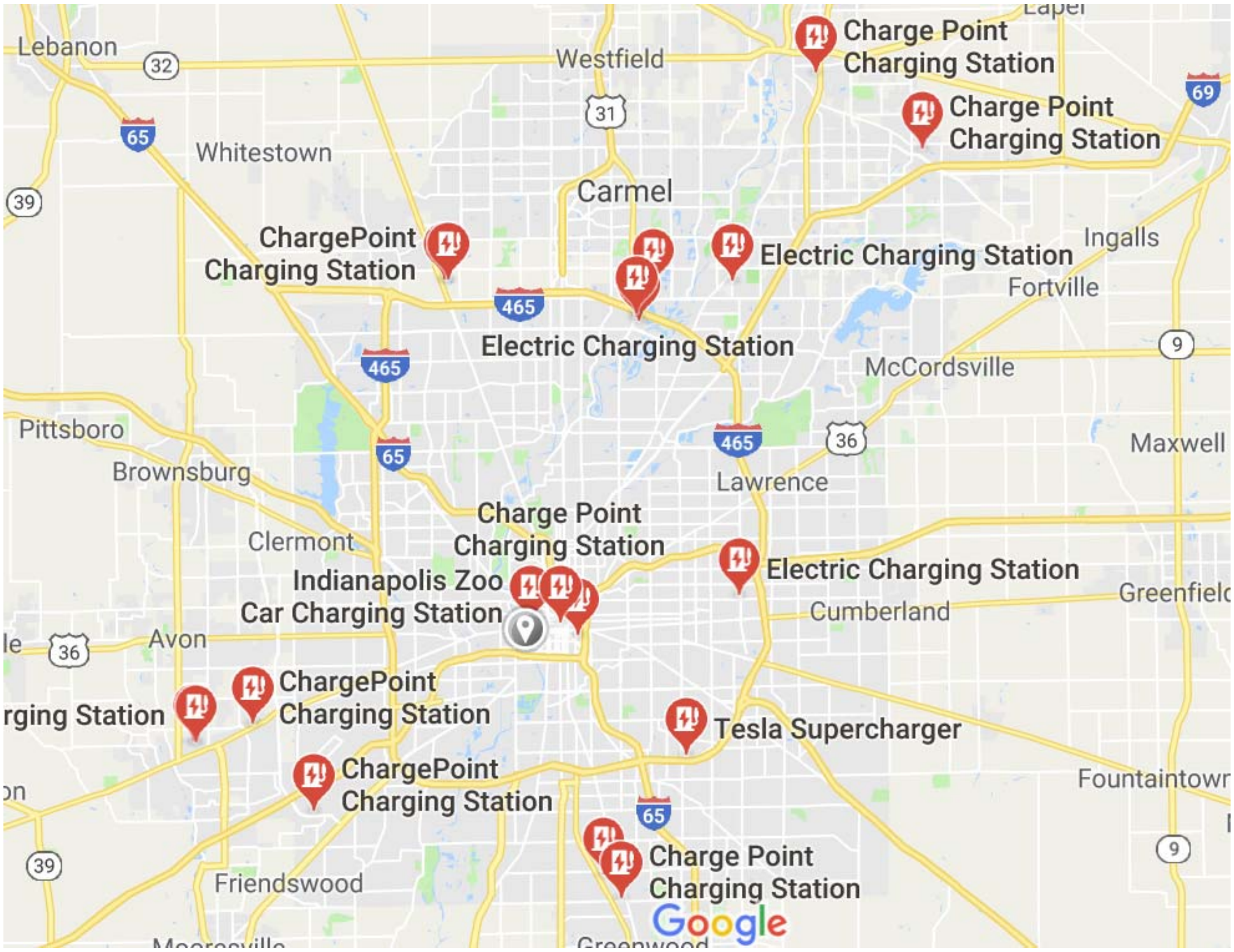
Source: FHWA

3.2.4 Passively-Collected Point-of-Interest (POI) Data

While land use data usually are based on census and business registration data, passively-collected data are gathered from online data sources, such as Cuebiq, Facebook, Foursquare, Google or Twitter. Often, these data are called Location-Based/Social Network (LBSN) data. These companies provide Application Programming Interfaces (API) to download the location, type and size of various trip attractors. Trip attractors include, for example:

- Restaurants and Bars
- Hotels
- Parks
- Ski Resorts
- Outdoor
- Medical facilities
- Grocery Stores

The availability of categories depends on the LBSN site. Most LBSN websites allow downloading a small sample for free, while larger samples require a fee.



Original Map: © 2018 Google®.

Figure 11. Car charging station POI data.

3.2.5 Choice-Maker Data

Choice makers often are stratified in different user groups. This classification may be done by income, household size, number of workers, number of cars, car sufficiency (usually defined as cars per worker), age of head of household, or any combination thereof. In essentially all cases, trip purposes further stratify destination choice modeling.

In traditional aggregate trip-based models, the stratification in destination choice is constrained by the stratification defined in trip generation. Activity-based models, on the other hand, commonly work with synthetic populations, and therefore, allow defining any stratification of user groups that works best in destination choice.

For estimation of destination choice models, choice-maker data generally is part of the observed choice data. Choice-maker data must be available from other sources for model applications but is generally available from Census data sources.

4.0 Destination Choice Set Formation

Destination choice models often take the form of multinomial logit (MNL) discrete choice models. Discrete choice models predict the probability that a choice-maker will choose a particular alternative from among a set of discrete alternatives. This list of alternatives is known as a choice set. Choice set formation or definition is a critical step in the specification, estimation, and application of all discrete choice models, including destination choice. The misspecification of choice sets can have adverse effects on parameter estimates and resultant computations of predicted choice probabilities. (Thill, 1992) The accurate definition of the destination choice set has been an issue of much interest to the profession and a variety of approaches have been developed and adopted in research and practice. With many travel demand model systems comprising thousands of zones, destination choice sets can prove to be extremely large, posing challenges for computational efficiency. On the one hand, methodological and computational advances now allow the use of the universe of locations (all zones) as the destination choice set. On the other hand, it has been speculated that the use of universal set of destinations as the choice set may compromise the behavioral representativeness of destination choice models such that the impedance measure captures not only willingness-to-travel but also the perception or consideration of alternatives, which could potentially bias the model's sensitivity. The analyst therefore needs to consider the pros and cons of alternative approaches when defining destination choice sets.

4.1 Using the Universe of Alternatives

Methodological and computational advances make it feasible to estimate discrete choice models with thousands of alternatives. In general, some consider it advisable to use the universe of destination choices when applying destination choice models in forecasting mode. For model estimation, sampling remains a popular approach for choice set formation, but the ability to use the universe of alternatives as the estimation choice set has proven increasingly appealing.

Using the entire universe of alternatives eliminates pitfalls associated with sampling approaches. Sampling approaches involve the use of Monte Carlo drawing procedures, thus rendering the formation of the choice set dependent on and susceptible to the choice set sampling process. In turn, model parameters and standard errors are also susceptible to the choice set sampling process. The gains from a statistical perspective should be weighed against the behavioral representativeness of such an approach when making decisions regarding the specification and estimation of destination choice models.

Theoretical behavioral basis for the use of the universe of alternatives together with psychological boundary terms and/or accessibility variables is grounded in both theory and research. These studies show how the availability/perception of alternatives in the choice set can be reflected implicitly by terms in the systematic utility function. (Cascetta and Papola, 2001) Psychological boundary terms and/or accessibility variables may play this role in destination choice models. (Fotheringham, 1991).

4.2 *Sampling Approaches*

In some cases, the multinomial logit (MNL) formulation for discrete models of destination choice makes it feasible to adopt sampling approaches without adversely affecting properties of parameter estimates. Generalized extreme value (GEV) based discrete choice models possess the desirable feature of accommodating sampling of alternatives without any deleterious effects due to their Independence of Irrelevant Alternatives (IIA) property. However, many destination choice models are not strictly GEV and do not observe the IIA property due to the use of attraction constraints and/or accessibility variables and in this case, sampling of alternatives can lead to biased parameters. Even so, considering that individuals are unlikely to consider thousands of alternatives when making location choices, it is appealing to adopt sampling approaches that may be more behaviorally realistic (from the standpoint that individuals can possibly gather and process information only for a subset of alternatives when making location decisions). In sampling approaches, samples of 30, 50, or 100 locations are commonly chosen from the universe of (feasible) choices – together with the chosen alternative. Multinomial logit models of destination choice are estimated on these sampled subsets of destination choices.

Two sampling approaches are commonly employed: random sampling and importance sampling. In random sampling approaches, the analyst selects a random sample of locations from the universe of (feasible) choices to constitute the consideration choice set. In this scheme, each alternative in the universe of (feasible) choices has an equal probability of being drawn into the consideration choice set.

In importance sampling approaches, the choice set composition method recognizes that some destinations are likely to be considered more highly (and thus considered more important or desirable) than others. An importance function is defined for each zone based on size and distance variables (its probability in a gravity model). Using Monte Carlo simulation procedures, a number of destinations are sampled with replacement from the importance probability distribution. Appropriate sampling correction factors then need to be applied in estimation to retain desirable properties of the maximum likelihood estimator.

4.3 Rule-Based Approaches

Rule-based approaches are largely based on assumptions that the analyst makes about criteria that define the inclusion or exclusion of an elemental alternative in a destination choice set. This approach to location choice set formation has been used in location choice model estimation for the Puget Sound Regional Council activity-based travel demand model. (Bowman et al., 2015) When setting rules for destination choice set formation, the following specific criteria should be considered.

4.3.1 Feasibility

Based on information contained in observed choice data, the analyst may establish feasibility criteria for inclusion of an element in a choice set. For example, based on a cumulative trip length distribution for shopping trips in a travel survey data set, the analyst may specify a distance threshold beyond which shopping locations would be considered infeasible and therefore excluded from the 'feasible' choice set. For example, Bowman et al. adopted a distance threshold equal to 125% of the longest trip distance (for a specific trip purpose) reported in the travel survey used for model estimation. While such feasibility criteria are often data-driven, they ignore heterogeneity in choice set formation and assume that a one-size-fits-all rule can be applied to the entire population. While a feasibility criterion may appear reasonable in the aggregate, it is unlikely to hold true in individual circumstances.

4.3.2 Awareness

As mentioned earlier, the universe of possible destinations can be very large. Individuals are not realistically able to consider and gather complete information on all possible destinations in the region. Given limited information gathering and processing capabilities of humans, and the possibility that individuals search until they are satisfied (satisficing rule), it is likely that individuals are aware of only a subset of alternatives as possible destinations and evaluate (in detail) only those alternatives that comprise the subset. Also, in many choice contexts, individuals may deliberately choose to narrow their search to a subset of alternatives, thus leading to the formation of an awareness set. It should be noted that awareness criteria can be combined with feasibility criteria to form a smaller subset of alternatives that constitute the intersection of these two sets of criteria. This smaller subset would then only include those alternatives that the individual considers feasible and is able to obtain full information to make an informed choice. In the absence of data about alternatives that individuals are aware of, it is difficult to establish robust awareness criteria for inclusion of alternatives in a choice set.

4.3.3 Trip Type – Land Use Compatibility

In most travel modeling contexts, it is possible to enhance feasibility criteria to consider the compatibility between trip type and land use characteristics. For example, one may choose to:

- exclude zones that have no jobs (employment) as possible work locations; exclude zones that have no student (enrollment) as possible school locations;
- exclude zones that have no retail employment as possible shopping locations;
- and exclude zones that have no housing stock as possible residential locations.

All these rules will reduce the size of the consideration choice set in location choice modeling.

The application of these compatibility rules is equivalent to using size variables in the utility equations to control the consideration/feasibility of zones. Thus, even if they are included in a choice set, zones that have a ‘zero’ size on a specific variable would be eliminated from consideration as a feasible destination for a specific trip type. This criterion may also consider the time-of-day at which a trip is being pursued; for example, if a shopping trip is being undertaken late at night, then only those destinations where establishments are open and operating at the time of the trip/activity (and known to the individual) would be included in the consideration choice set.

4.4 Time-Space Prism Approaches

Time-space prisms represent activity spaces within which individuals may travel and pursue activities. The notion of a time-space prism is derived from the field of time geography (Miller, 2009) and it has been used extensively in constructing activity-based travel demand model systems. The time-space prism constitutes a constrained action space that limits the range of destinations that an individual can visit.

The time-space prism is often defined by boundaries or anchors that describe places where an individual must be present by a certain time (or within a specific narrow time window). These spatial-temporal bounds define a prism, and the size of the prism is determined by the separation between the spatial-temporal boundaries and the speed of travel.

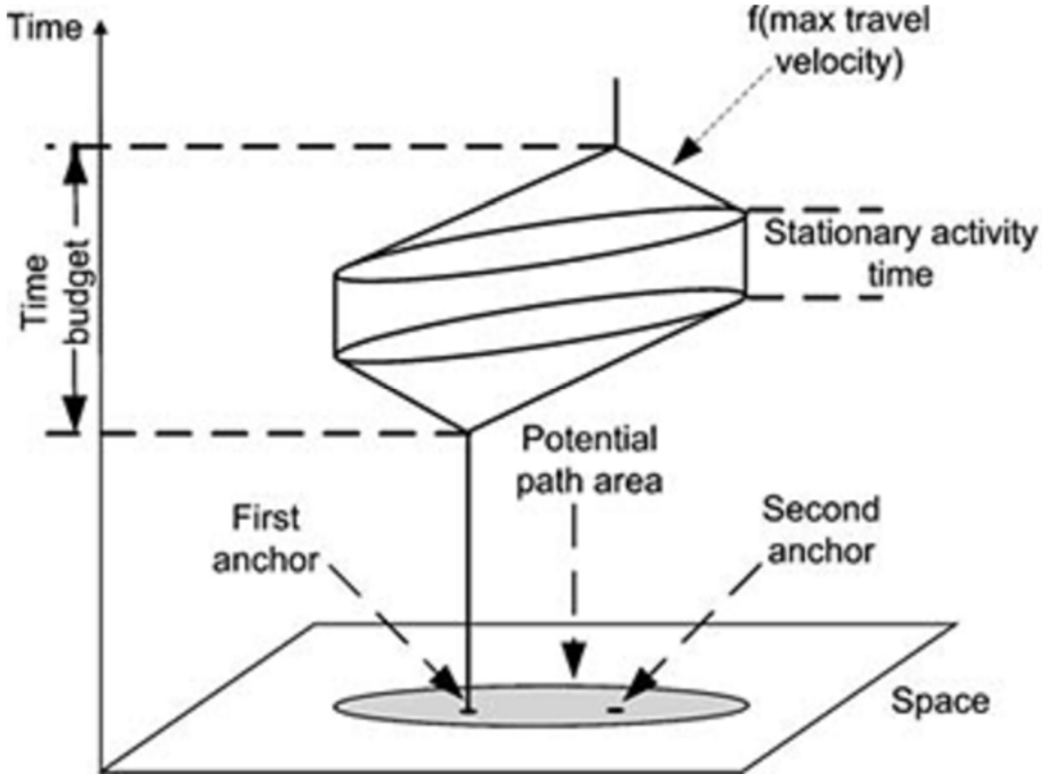


Figure 12. Space-time prism (Miller and Bridwell, 2009).

Source: Miller, 2009

Based on a knowledge of spatial-temporal constraints, the analyst can identify the feasible destination choice set as consisting of all possible locations that can be reached without violating a time-space prism constraint. The set of destinations that may be visited depends on the speed of travel; if the mode for a trip or tour is known a priori, then the speed of the mode can be used to determine prism size. If mode is determined after destination choice, then the set of reachable destinations may be identified based on the speed of the fastest available mode or an average speed of all feasible modes.

Care must be exercised in the use of this approach to define destination choice sets. Time-space prism constraints can be fuzzy and difficult to define, especially in the absence of specific data collected by survey samples about spatial-temporal constraints. For individuals who do not have anchor activities (such as work or school), the prism can be extremely large – resulting in a large set of possible destinations that may be reached without violating prism constraints. However, it is unlikely that individuals will consider such large choice sets when making location decisions.

Time-space prism constraints can be further combined with feasibility and/or awareness criteria as well as land use availability criteria to identify a final subset of destinations that may be considered for a trip.

4.5 The Modifiable Areal Unit Problem (MAUP)

The Modifiable Areal Unit Problem (MAUP) arises due to the geographic aggregation of location alternatives into zones or similar aggregate spatial units. Locations are often individual point or parcel locations (such as an individual home, building, office, or store), and yet choice alternatives in spatial location models are represented as zonal aggregations of such locations. The definition and delineation of traffic analysis zones (TAZ) affects the magnitudes and properties of model parameter estimates.

MAUP is thus a source of statistical bias that can significantly affect inference regarding the influence of various factors on destination choices (Fotheringham, 1991). The MAUP is an issue in destination choice modeling even though size terms used as explanatory variables in destination choice utility equations enter in log form, thus ensuring that choice probabilities change in direct proportion to the change in intensity of activities.

Because of the MAUP, caution should be exercised in the application of destination choice models in forecasting mode. In particular, the zonal configuration used in model estimation should be retained in model application as well. Substantial changes to the zone system could invalidate a destination choice model and require re-calibration beginning with re-estimation. A destination choice model estimated on one zonal configuration should not be applied to a vastly different zonal configuration (destination choice models should not be transferred between metropolitan areas). While very minor or modest deviations in zonal configuration (e.g., a few zone splits) may be acceptable in application mode, such deviations should be kept to a minimum. The choice set should therefore consist of TAZ are defined with care, essentially minimizing spatial autocorrelation problems through the definition of disaggregate homogeneous zones.

Within the mathematical formulation of the destination choice model, the implications of the MAUP on the size terms can be addressed from two perspectives: either the boundaries of the zones can be considered as artificial constructs of the model or the boundaries of the zones can be considered as meaningful (if limited) expression of spatial auto-correlation.

In the first case, zone boundaries are meaningless, then the individual activity opportunities within a zone should be considered no more like each other than the opportunities in other zones; this is achieved mathematically by fixing the coefficient on the size log term in the utility function exactly equal to 1.0. If, on the other hand, the zonal boundaries are meaningful, that implies that there is greater similarity among the individual activity opportunities within a zone than there is between those activities and others outside the zone. In this case, the coefficient on the size log term in the utility function should be constrained to take on a value between 0 and 1, analogous to a nested logit model for activities.

4.6 Disaggregate Representation and Allocation of Activity Locations

With the adoption of increasingly disaggregate representation of space, the travel demand modeling profession is increasingly utilizing fine-grained units of geography to identify locations and destinations. For example, activity locations can be represented as points (on network links) or individual parcels. More recently, activity-based models have been adopting the use of microanalysis zones (MAZ) as a unit of geography to represent destination and location choices at a fine-grained spatial resolution and better capture transit access and egress legs.

More sophisticated procedures in which synthetic populations generated at an aggregate geographic scale are further spatially sub-allocated to the parcel level have also been developed (Zhu, 2014). Such procedures employ iterative processes at multiple geographic scales to allocate households to individual parcels while controlling for known parcel capacity constraints and estimated conditional distributions of household attributes by building type. However, much of the detailed information upon which these procedures are based may not be capable of being forecast with accuracy.

5.0 Destination Choice Model Specification

The most common specification of destination choice is as a multinomial logit (MNL) model. Singly constrained gravity models – which are commonly used in aggregate, trip-based models – can be shown to be a special case of a GEV multinomial choice model. More complex formulations are universal or mother logit models and do not observe the IIA property, nor are they any form of GEV model.

Typically, zone-based destination choice models will incorporate a utility function that includes two categories of explanatory factors: qualitative factors (how good are the choices in a given destination zone), and quantitative factors (how many individual choices are in a zone). The usage of qualitative explanatory factors is common in virtually all choice models. For destination choice models, these commonly include impedance, accessibility, psychological boundaries, and other destination qualities, as well as traveler attributes. The quantitative factors, typically labeled as size terms or attractions, are an unusual feature of destination choice models, which arise because the "alternatives" represented in the model, often TAZs, are not actually the choices, but they represent a pool of choices. The actual choice is instead one particular activity point (job, store, theatre seat, etc.) within the zone. Due to this distinction, factors that represent the quantity (instead of quality) of choices in a zone need to be treated differently in the mathematical formulation, being included in log form.

This section begins by specifying a simple gravity model as a destination choice model and then adds various terms to the utility function to develop a rich utility specification. While not all of these terms or factors may be relevant in all destination choice contexts, many are widely applicable.

The following discussion assumes familiarity with the general formulation of MNL models shown below for reference, describing the probability (P) of a traveler or type of traveler (h) choosing a destination (j) given an origin (i).

$$P_{j'|ih} = \frac{e^{V_{ij'}}}{\sum_j e^{V_{ij}}}$$

Figure 13. Equation. General formulation of MNL models.

(For those interested in a good introductory text on topic with relevance to travel modeling, see Koppelman and Bhat's *Self Instructing Course in Mode Choice Modeling*.) The destination choice problem is generally presented with reference to an individual decision-maker. However, the model is equally applicable to aggregate, zone-based formulations. This section describes the specification of the destination choice utility function in general terms.

5.1 Size Terms/Attractions

Destination choice models are usually represented with some level of aggregation of the alternatives. That is, the "alternatives" or destinations represented in the model, often TAZs, are not actually the choices, but they represent a pool of choices. For example, the destination choice model may express the choice of a work trip destination as TAZ 123, but in actuality the destination is one particular job among however many jobs there are within that TAZ; if there are more jobs in the TAZ, there are more actual sub-alternatives to choose within the modeled alternative of TAZ 123. The aggregate choices in many ways are similar to a nested logit model, with the aggregations (zones) corresponding to the nests, except we only observe the choice at the nest level, not at the elemental alternative level. To incorporate this detail into the utility function for the destination choice model, we must provide a representation of the number of individual unique alternatives available within the zone.

In general, when thinking about which variables and parameters are part of the size terms in the utility function, the questions to consider is whether the data represents how many opportunities there are (if so, it's a size term) or how good (or bad) the opportunities are (in which case, not a size term). The exact nature of the quantitative term will generally vary based on the trip purpose being modeled. For work trips, it is typical to include measures of employment, either in total or by industry type (the latter being preferred if disaggregate employment information for travelers is also available by industry type). For non-work purposes, it is typical to include only particular relevant industry categories (e.g. retail employment for shopping purposes, restaurant employment for meal purposes, etc.) and other socio-economic features of the zones as well (e.g. households or population for social purposes). In these cases, the size term is typically a linear combination of different types of employment or other variables, for example:

$$Size_j = \alpha_1 \times RetailEmp + \alpha_2 \times ServiceEmp + \alpha_3 \times IndustrialEmp + \cdots$$

Figure 14. Equation. Linear combination of different types of employment or other variables.

In gravity models, the relative weights or α parameters are estimated independent of the calibration of an impedance function or friction factors. However, this sequential estimation can result in biased parameters and sensitivities. Therefore, the α parameters should either be asserted on a fundamental theoretical basis or estimated simultaneously with other destination choice model parameters.

The aggregation of alternative destinations in zones has been shown to result to result in aggregation bias in model parameters (Ye et al., 2012), but is generally necessary both for computational tractability and due to fundamental data limitations.

The size term always enters the utility function in log form. (See Size Terms in Aggregate Choice Models to understand the theory by which this can be derived.) The log formulation is necessary so that the choice probability of a destination is directly proportional to the number of opportunities at the destination. In other words, if the number of jobs at a destination doubles, all else equal, then the choice probability of this destination approximately doubles. If size or attractions were the only factor affecting destination choice (i.e., travelers did not care about distance, time, etc.), the systematic utility (V) of each destination (j) could be written as follows:

$$V_j = \ln(Size_j)$$

Figure 15. Equation. Systematic utility of each destination.

A corollary of the size term log specification is that the choice probabilities are invariant with respect to the scale of the size term. That is, the choice probabilities remain the same when the entire size term is multiplied by an arbitrary factor. For this reason, by convention one of the variables in the size term is given a coefficient value of 1. (Daly, 2918) Doing so is optional in model application, but necessary when estimating the model, since otherwise the estimation problem is undetermined.

5.2 Distance/Impedance Terms

Perhaps the most fundamental terms in the utility function for destination choice models are measures of distance, travel time, or more generally, impedance. These terms represent the effort required to get to various alternative destinations from a known origin. Impedances commonly combine information on travel time, travel cost, and distance by various modes. (See 3.2.1 Impedance Measures: Travel Time and Cost Data.)

Incorporating the impedance (*t*) between origin (*i*) and destination (*j*) along with the size term, the systematic utility (*V*) of can be written:

$$V_{j|i} = \ln(Size_j) + \beta \times t_{ij}$$

Figure 16. Equation. Systematic utility including impedance and size term.

where the parameter β is the relative importance of travel impedance (*t*).

In this formulation, equivalent to the singly constrained gravity model with an exponential impedance function, the utility (and therefore probability) of a destination depends on the impedance or spatial separation between the trip origin and the destination, and the size or attractions at the destination. This is the simplest representation of destination choice utility.

Other types of impedance functions can be used, as well. Log transformations of the impedance variable are equivalent to the use of a power function impedance function in a gravity model, while the use of both the linear and log forms is equivalent to the use of a gamma function. (Daly, 1982)

$$V_{j|i} = \ln(Size_j) + \beta_1 \times t_{ij} + \beta_2 \times ln\left(t_{ij}\right)$$

Figure 17. Equation. Systematic utility with more explanatory variables.

The coefficient of the impedance variables (β) can be generic (i.e., the same for all decision-makers), or it can vary for certain types of travelers (*h*). For example, it is often found that higher income workers tend to be willing to choose work locations that are farther from home than other workers, all else equal. This is represented in the utility function by a less negative coefficient on distance impedance for high income workers than the distance coefficient used for other workers.

$$V_{j|ih} = \ln(Size_j) + \beta_{1h} \times t_{ij} + \beta_{2h} \times ln\left(t_{ij}\right)$$

Figure 18. Equation. Systematic utility with more explanatory variables.

Since trip-maker characteristics are the same for all destinations, the way to represent the effect of these variables on destination choice is to interact them with one of the impedance variables or by partially or fully segmenting the model. With full segmentation, there are two entirely separate utility functions; whereas, with partial segmentation, some terms of the utility are common to all choice-makers while others are specific to particular market segments. An example of a partially segmented model is when the worker industry is known, and the size variable becomes a function of worker industry. An example of a fully segmented model is to specify different models (utility functions), one per household car sufficiency segment, in a trip-based model.

5.3 Alternative Specific Constants

Unlike for many other choice models, it is not common to incorporate alternative-specific constants for every destination zone. The use of constants results in some convenient mathematical properties. Most notably, for MNL models, the other parameters of the model will have unbiased estimators even in the presence of non-uniform sampling. Without constants, in contrast, parameter estimates may be biased. However, including a complete set of alternative-specific constants can result in other complications: if the number of zonal alternatives approaches or exceeds the number of sampled destination observations, the model parameters will be over-determined and model estimation will simply fail.

To update the original simple gravity specification, constants (*c*) for each zone or for certain groups of zones (or zone pairs) can be added:

$$V_{j|ih} = c_j + \ln(Size_j) + \beta_{1h} t_{ij} + \beta_{2h} ln(t_{ij})$$

Figure 19. Equation. Systematic utility with alternative specific constants.

The household survey data traditionally used to estimate destination choice models did not provide enough observations to support the inclusion of a full set of constants. Instead of employing a complete set of constants for every alternative, constants have commonly been used for just a partial set of alternatives. For example, the model could include a constant for destinations located in Central Business District areas, or for destinations that include a regional shopping mall.

Now, however, the availability of large scale passive OD data sets can support the estimation of not only a full set of destination constants but also some OD constants (aggregate district-to-district interaction terms). The use of constant rich utility specifications is therefore a new approach but has the dual advantage of substantially improving destination choice models' goodness-of-fit and resulting in less biased parameter estimates and model sensitivities.

5.4 Psychological Boundaries

The utility function may also include disutility factors associated with crossing boundary lines, which may represent real geographic features (e.g. rivers, railroads, freeways, ridge lines, intervening rural areas) or socio-political boundaries (e.g. state lines, county lines, neighborhood boundaries, etc.). These psychological boundaries or barriers terms in the utility functions of destination choice models have important theoretical basis in cognitive or behavioral geography and are closely and importantly related to Lynch's concepts of Edges and Districts as elements of mental maps. The cognitive psychology literature has documented the effect of such boundaries in distorting the perception of spatial relationships such as distance. (For a good review of cognitive geography, see Mark et al., 1999.)

Denoting a psychological boundary between origin (*i*) and destination (*j*) as *b*, the systematic utility becomes:

$$V_{j|ih} = c_j + \ln(Size_j) + \beta_{1h} t_{ij} + \beta_{2h} ln(t_{ij}) + \beta_{3h} b_{ij}$$

Figure 20. Equation. Systematic utility with psychological boundaries.

These terms may be understood as playing an important and implicit role in the availability/perception of alternatives in destination choice sets (see Cascetta and Papola, 2001, for discussion of how the availability/perception of alternatives in the choice set can be reflected implicitly by terms in the systematic utility function). In this sense, these psychological boundary terms can be important and serve as a representation of the fuzzy boundaries of travelers' mental/cognitive maps which define their destination choice sets. The use of these terms may, therefore, complement the use of the universe of destinations as the choice set and perhaps complicate other schemes of choice set formation.

5.5 Agglomeration effects and competing destinations

In the utility functions described thus far, two destinations that are equidistant from the origin and have the same size (number of jobs, for example), will exhibit the same choice probability, all else equal. However, one of these destinations may be in a Central Business District, while the other may be in a suburb. The CBD destination may be more attractive because it more conveniently affords opportunities for conducting other activities, such as going out for lunch, shopping, recreation. In essence, part of the attractiveness of a destination lies in the accessibility that it provides to other activities. This effect can be introduced in a destination choice model by adding accessibility variables. Note that the accessibilities (*A*) are calculated from each destination zone (*j*) to attractions (*S*) all other destinations (*k*):

$$A_j = \ln\left[\sum_k S_k \times e^{\beta' \times t_{jk}}\right]$$

Figure 21. Equation. Accessibility measurement.

The impedance term embedded in the accessibility may or may not be the same as the impedance term in the upper level of the utility function. It is not unusual, for instance, for β' to be more negative than the upper level decay parameter β.

Conversely, accessibility variables can help to differentiate between destinations that compete. An example of competing destinations is retail locations for incidental shopping. The use of accessibility variables to capture differential spatial competition among alternatives lead Fotheringham to formulate the Competing Destinations model, which was later shown to be a special case of a destination choice model and adapted to more general use in this context. (Bhat et al., 1998) Both of these effects can be incorporated in destination choice models through the careful specification of multiple accessibility variables such as accessibility to complements (A^C_j) and accessibility to substitutes (A^S_j). (Bernardin et al., 2009)

$$V_{j|ih} = c_j + \ln(Size_j) + \beta_{1h} t_{ij} + \beta_{2h} ln(t_{ij}) + \beta_{3h} b_{ij} + \beta_{4h} A_j^C + \beta_{5h} A_j^S$$

Figure 22. Equation. Systematic utility with accessibility measurement.

Figure 23. Agglomeration (trip-chaining) and spatial competition (autocorrelation) effects captured by accessibility.

Source: FHWA

5.6 Other Destination Qualities

Other attributes of the various destinations can also be included in a destination choice model. Some examples of such variables sometimes included are

- Walkability measures
- Measures of the diversity of land uses
- Parking costs (or a simple indicator for paid parking)
- Miles of coastline/lakeshore

These or other attributes of destinations can also be incorporated in destination choice models in a straightforward manner. Where N destination quality variables (x) are incorporated, the systematic utility can then be specified:

$$V_{j|ih} = c_j + \ln(Size_j) + \beta_{1h} t_{ij} + \beta_{2h} ln(t_{ij}) + \beta_{3h} b_{ij} + \beta_{4h} A_j^C + \beta_{5h} A_j^S + \sum_{n=6toN} \beta_{nh} x_n$$

Figure 24. Equation. Systematic utility with additional travel attributes.

5.7 Equilibrium Constraints and Shadow Pricing

It is sometimes desirable to apply a doubly-constrained model, as doing so reflects equilibrium conditions between supply and demand. Doubly-constrained means that the sum of predicted trips to each destination matches, or is at least proportional to, agreed-upon destination control totals. For example, it is desirable that a model that predicts usual school location for university students results in as many students selecting a school TAZ as the reported university enrollment at the TAZ. Similarly, it is desirable that the prediction of workers at their workplace is proportional to the number of jobs at each TAZ. A destination choice model is not guaranteed to be doubly-constrained; the model may allocate more students to some schools than are enrolled in it, while allocating fewer than actual students to the other schools.

$$V_{j|ih} = c_{ij} + p_j + \ln(Size_j) + \beta_{1h} t_{ij} + \beta_{2h} ln(t_{ij}) + \beta_{3h} b_{ij} + \beta_{4h} A_j^C + \beta_{5h} A_j^S + \sum_{n=6toN} \beta_{nh} x_n$$

Figure 25. Equation. Systematic utility with shadow pricing constraints.

In order to enforce a double or attraction constraint, a special constant may be added to each alternative. (Daly, 1982) This constant is, in fact, the lagrangian multiplier corresponding to the relaxation of the equilibrium constraint. More commonly, these constants are called shadow prices, borrowing the terminology from economics, in which the term denotes as here an unobserved (non-monetary) component of cost beyond the observed cost that can be inferred by the market's observed equilibrium. Ideally, these constants should be estimated simultaneously with other model parameters. However, this is often computationally challenging and not supported by statistical software, therefore, they are typically calculated in an iterative fashion:

1. Apply the unconstrained model (i.e., model without shadow prices)
2. Sum the model predictions for each destination, across all origins (i.e., the column sums of the origin/destination matrix)
3. Compare the column sums to the destination control total (school enrollment, jobs, etc.)
4. Calculate the natural log of the ratio of control destinations over predicted destinations
5. Add this term to the utility function, apply the model again, and repeat the process until the shadow prices converge

While modifying a singly-constrained model in this fashion is common practice, it is known that ignoring constraints when estimating destination choice models can lead to biased estimates. There is some danger, therefore, that incorporating shadow prices via model calibration, as described above, may not be sufficient to correct for a fundamental model parameter bias. The difficulty however lies in that standard logit estimation software cannot estimate models with constraints.

5.8 Intrazonal Dummy Variables

It is common practice to introduce a special binary indicator (dummy) variable to denote the diagonal of the impedance matrix where the origin and destination zones are the same. As intrazonal impedances are commonly estimated in a different manner than other, interzonal impedances, these terms are often necessary and are frequently adjusted in calibration.

5.9 Traveler Characteristics

The observant reader will have noticed that any of the qualitative parameters (β) have been subscripted with *h*, to denote that they may take on different values for different market segments. Automobile availability and income are the most commonly used traveler attributes which are used to segment destination choice models. However, other traveler characteristics can be used as well, and they can be used as interaction terms rather than for segmentation. The best example of this is residential accessibility. Residential accessibility is sometimes interacted with impedance on the theoretical basis that when people choose where to live they also choose how much they are willing to travel. This theory has been born out in robust parameter estimates for these interaction terms. Given both the strong theory and empirical results, there is some reason to prefer this to segmentation by income.

$$V_{j|ih} = c_{ij} + p_j + \ln(Size_j) + \beta_0 A_i t_{ij} + \beta_{1h} t_{ij} + \beta_{2h} ln(t_{ij}) + \cdots$$

Figure 26. Equation. Systematic utility with income segmentation.

Source: FHWA

6.0 Destination Choice Model Estimation

Once observed choices and explanatory variables from data are related by formulating a utility function, the challenge becomes estimating the parameters that quantify these relationships or how explanatory variables contribute to destination choice probabilities. Rather than a one-time effort, this is commonly an iterative process in which alternative specifications of the utility function are tested. The parameter estimation process is based in statistical/econometric theory and generally relies on maximum likelihood estimation (MLE) techniques. Specialized software or custom programming is generally required. Algorithmic approaches to MLE for destination choice models generally fall into two general families: gradient-based and metaheuristics.

6.1 Gradient-Based Approaches

The traditional approach for choice model parameter estimation is maximum likelihood estimation. This method is based on a probabilistic evaluation of the model given the observed explanatory/exogenous data, the proposed model structure, and any given set of model parameters. Model estimation is the process for finding the set of parameters that maximizes the model's likelihood of the observed choices. In practice it is preferable to maximize the logarithm of the likelihood function instead; this transformation does not change the location of the maximum parameters, and the resulting calculations are simpler and more numerically stable.

For regular MNL models that have a strictly linear-in-parameters utility function (i.e., for destination choice models with no parameters to estimate embedded inside size or accessibility terms or shadow prices), it is well known that the log of the likelihood function is both smooth and globally convex. This means that, for any initial guess at the model parameters, a gradient descent maximization algorithm will eventually converge to the global optimum. Put more simply, the analysis always gets the same result from parameter estimation and it is guaranteed to be correct. Moreover, because this kind of model is relatively common and simple to use, there are a variety of off-the-shelf computer programs that will be able to handle the parameter estimation efficiently.

However, destination choice models typically are not simple MNL models with a strictly linear-in-parameters utility function. Many times, for instance, the size term is parameterized inside its log function, which can potentially introduce non-convexity in the overall log likelihood function. This non-convexity can be addressed through a variety of techniques (some discussed below) although the computational effort typically increases greatly for achieving a reliable globally optimal solution for non-convex problems.

6.2 Metaheuristics and Machine Learning

While gradient-based methods are efficient for parameter estimation problems where the maximum likelihood function can be shown to be globally convex, many destination choice models do not meet this criterion. Two common reasons that destination choice models have non-convex likelihood functions is because they use accessibility variables in the utility function to reflect the jointness or inter-dependence of destination choices due to trip-chaining and/or because they include shadow prices or other constants to enforce a market equilibrium or attraction constraint. An example of a likelihood function with multiple optima for a destination choice model with an accessibility variable is shown in Figure 27.

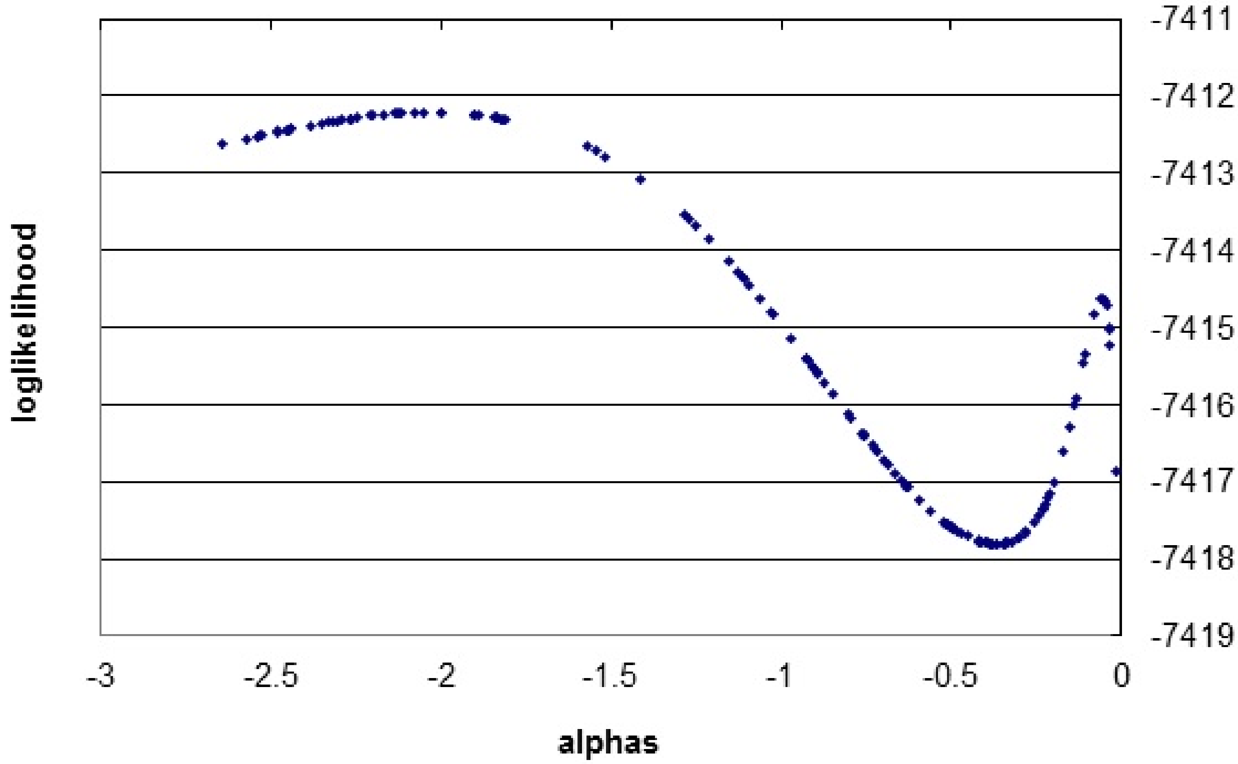


Figure 27. Nonconvex likelihood function of a destination choice model with embedded accessibility parameters.

Source: FHWA

The use of gradient-based methods to estimate parameters for models with non-convex likelihood functions can result in sub-optimal solutions. For example, if gradient-based methods were used to estimate the accessibility parameter for the model with the likelihood function illustrated above, assuming the starting parameter value was zero, the algorithm would find the local maxima close to zero, not the global maximum close to -2. One alternative that is sometimes proposed, is to use gradient-based algorithms with multiple parameter sets for starting points, but this approach is both labor and computationally intensive and quickly becomes increasingly complex as the number of parameters to be estimated increases. Alternatively, various meta-heuristics for non-linear optimization and/or machine learning methods can be used for robust parameter estimation for models which are liable to multiple optima. Genetic algorithms have been applied for this task and more recently other machine learning methods such as boosted decision trees have also been reported.

6.2.1 Genetic Algorithms

A genetic algorithm begins by generating an initial population or set of solutions, each solution being a complete set of parameters. The algorithm then applies each candidate solution, often using the ultimate model application code. Another module evaluates the fitness of each solution, calculating its log likelihood by comparing the predicted probabilities from the model application to the observed data. Once the fitness of all the candidate solutions in the population has been evaluated, the least fit solutions die or are removed from the population, the best solution is cloned to start the next generation's population and the remaining members of the new population are created by either mating or randomly recombining two solutions from the parent population (whose probability of reproducing is a function of their fitness) or by mutating a single solution from the parent population (again, whose probability of being selected for mutation is a function of its fitness). The process iterates until no further progress can be made and a maximum likelihood solution is produced.

The genetic programming approach has significant advantages and disadvantages compared to traditional analytic gradient-based methods. The key disadvantage is its computational intensity. It is computational expensive and take significantly long time to estimate model parameters. This is clearly a significant impediment to its widespread use. Better parallel processing may be able to significantly reduce the run time.

Despite the run time, the genetic programming approach has several attractive advantages over traditional estimation methods. The method is robust to multiple optima which are possible for constrained or joint/conditional models. It obviates the need for sampling of alternatives common in destination choice model estimation, simplifying the estimation and improving the statistical efficiency of the estimator, making better use of limited survey data or full use of large passive datasets. It allows benefits from the use of inequality constraints on parameters which can be used to enforce that a parameter has the logical sign. It allows for otherwise non-linear-in-parameters utility specifications, such as the estimation of embedded distance decay parameters in accessibility terms important for controlling for spatial autocorrelation and trip-chaining effects. It also greatly reduces the opportunity for inconsistencies between estimation and ultimate application of the model for forecasting when the application code is used in estimation.

7.0 Destination Choice Model Implementation

Destination choice models can be implemented in various ways in different travel modeling frameworks. They can be applied disaggregately in activity-based models using Monte Carlo simulation or in aggregate trip-based models using matrices. In both contexts, there are important issues related to how destination choice models are integrated with the larger model system. Key issues include how various destination choices are related to each other, how choices of destination and mode are related to each other, and how the larger model system achieves an equilibrium between travel demand and supply, commonly through iterative feedback loops. Destination choice models can also be implemented in an incremental or pivoting framework.

7.1 Aggregate vs Disaggregate

Destination choice models can be applied disaggregately in activity-based models using Monte Carlo simulation or aggregately in trip-based models using matrices. In an aggregate model the destination choice selection probabilities are applied to all trips produced in a given TAZ at once, while in a disaggregate framework, a destination is predicted for individual trips, one at a time.

7.1.1 Aggregate Applications

The application of a destination choice model in an aggregate framework, such as in a trip-based model, follows the same practice as commonly used for mode choice models. Destination selection probabilities are calculated for each trip market segment, and then they are multiplied by the total trip productions predicted for the market segment. The selection probabilities are treated as “shares”, to be used in allocating the trips produced in an origin zone to all destination zones in proportion to the selection probabilities.

In an aggregate framework, care must be taken to use a consistent trip segmentation from trip generation through mode choice, particularly when mode choice logsums are used in the destination choice utility. A good practice is to use the same trip segmentation in the destination and mode choice models, and to use a trip generation segmentation that can be easily collapsed into the destination choice segments.

Given that most regional models have thousands of TAZs, the selection probabilities can be very small for the majority of TAZs. It is important to use adequate floating-point precision when applying these models to avoid introducing rounding and/or truncation errors. Aggregate applications of destination choice models can sometimes take advantage of optimized routines in travel modeling software packages for applying gravity models by reformulating the destination choice model as a gravity model with k factors where the k factor is calculated as the residual of the utility function beyond the impedance and size terms.

Doubly-constraining a destination choice model in an aggregate framework is typically accomplished by the application of iterative-proportional fitting (IPF), with target destination controls that are proportional to the model attractions

7.1.2 Disaggregate Applications

The application of destination choice models in a disaggregate framework, such as an activity-based model, requires applying Monte Carlo simulation to select a single TAZ as the trip or tour destination, given selection probabilities. The use of Monte Carlo simulation is common in the implementation of other probabilistic models, such as those used to predict mode choice, auto ownership or tour frequency in activity-based models.

In a disaggregate model the mode choice logsums used in the destination choice models are oftentimes referred to as "representative" logsums, because they omit certain decision-maker characteristics, in the interest of maintaining model run times within practical limits. For example, while the model system may use distributed values of time, the mode choice logsums may be computed only for a small number of value of time classes. Similarly, representative logsums may be used when the mode choice model includes explanatory variables such as age or household composition.

It is also common in a disaggregate framework to use a sample of alternatives, rather than the entire destination choice set, when applying the model. This strategy is also used to control model run times, since resolving the destination choice probabilities over thousands of alternatives can be computationally onerous, although it is becoming less and less so.

Doubly-constraining a destination choice model in a disaggregate framework is typically accomplished with shadow prices. Shadow prices need to be calculated iteratively, since there is no exact a priori formula that will result in a doubly-constrained model.

7.2 System Integration Issues

Key issues include how various destination choices are related to each other, how choices of destination and mode are related to each other, and how the larger model system achieves an equilibrium between travel demand and supply, commonly through iterative feedback loops.

7.2.1 Inter-related Destination Choices

In the context of daily travel, the choice of many destinations are made jointly or conditioned on the choice of other destinations. For example, the choice of where to get dinner on the way to home from university classes in the evening is strongly conditioned on not only the home location of the traveler but also their school location choice. Other destination choices are conditioned on other individual's destination choices. For instance, when two friends decide to meet for lunch in the middle of their work days, the lunch destination is conditioned on both travelers' work locations. Various frameworks can and have been used to reflect the conditioning or jointness of destination choices. These include the use of sequences or hierarchies of destination choices, the use of "rubber-banding" to calculate travel impedance in terms of diversion or marginal travel cost, and the use of accessibility variables to capture agglomeration or convenience effects of destinations.

In traditional four-step travel models, all destination choices are assumed to be independent. This is reflected in the fact that gravity or destination choice models in this context are run in parallel and independently of each other. Both activity-based, hybrid, and advanced trip-based frameworks have developed different approaches to relaxing this assumption of independence. Each of these approaches in presented in summary below.

The Activity-Based Simulation Approach

Although there is variation in the details of various activity-based simulation approaches, they generally all share some common characteristics, owing to their common descent from the approach developed for early activity-based models in Portland and San Francisco. The first is that long-term choices of mandatory activity locations such as work and school are modeled first and subsequent daily destination choices are conditioned on these. The second is that each tour is assumed to have a primary destination (which may be work or school or a location such as a retail store chosen that day), and intermediate stops locations are iteratively chosen by sequentially adding stops on the outbound or inbound half-tour on the way to or from the primary destination based on the marginal cost of adding that stop to the tour. This approach of using marginal travel costs for intermediate stop location choices is sometimes referred to as rubber-banding and is illustrated in Figure 28.

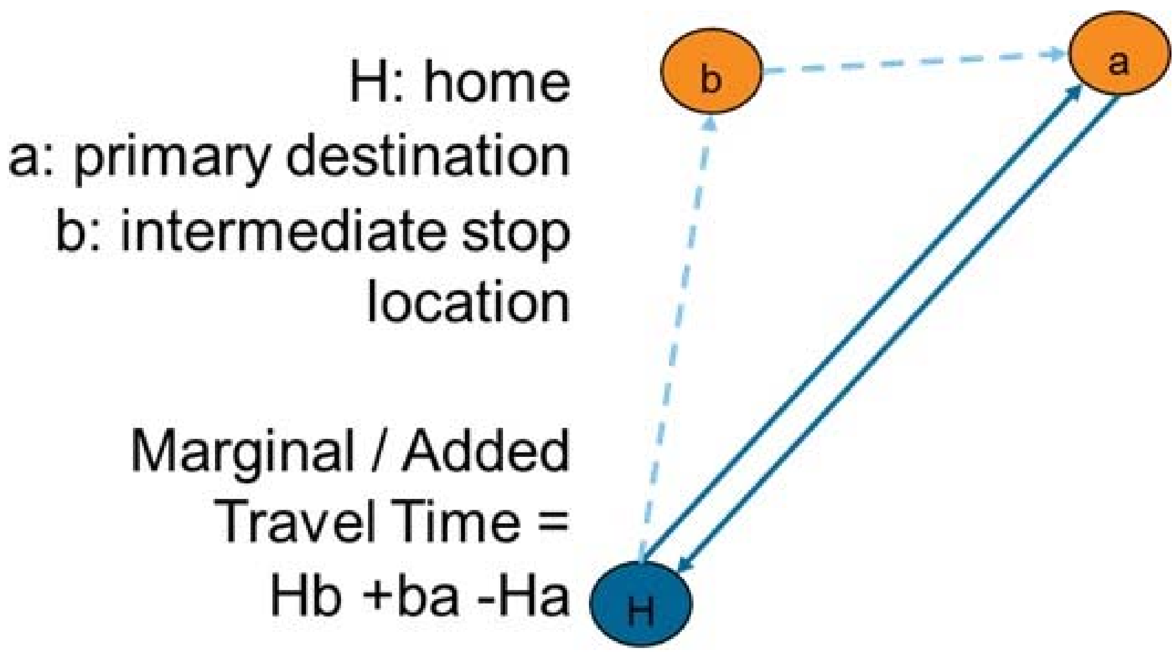


Figure 28. Sequential choice of intermediate stop locations with rubber-banding.

Source: FHWA

Two key ways that various activity-based model formulations differ is in their approach to choice set formation and the use of accessibility variables in the utility function (whether logsums of other destination choices or proxies thereof). The disaggregate simulation framework of activity-based models conceptually allows the explicit representation of space time constraints to define the feasible choice set. However, these constraints can be binding or fuzzy to various degrees and computationally challenging to implement. Therefore, they are sometimes represented through the use of simpler, more easily computed proxy variables, for instance, giving some indication of "time pressure" based on general information about the daily activity pattern and/or tour. Some destination choice models (generally non-mandatory tour primary destination choices) include accessibility variables to indicate the convenience of the primary destination to potential intermediate or secondary stop locations, but precise utility formulations vary considerably, for instance whether accessibility variables are calculated for specific individuals, etc.

Through these various techniques, the activity-based approach can incorporate the jointness of destination choices, particularly for destinations on the same tour, including space-time constraints to with some degree of explicitness. Moreover, the approach has some behavioral plausibility in terms of schedule formation, particularly for mandatory tours. However, the assumption that all tours have a “primary” destination which anchors the tours and on which the other location choices are conditional (but not necessarily the converse) is a strong assumption that is not necessarily supported for many non-mandatory tours. For this reason, there is some variation in the heuristics used to identify a tour’s primary destination, some of which place more importance on duration of activity, others on distance from the home/work anchor, and others on an asserted hierarchy of activity purposes. The approach of building tours sequentially, adding one stop location at a time, also basically requires a simulation framework and fundamentally limits the applicability of the approach to aggregate modeling frameworks because the size of the matrices requires grows exponentially with the number of stops.

The Hybrid Approach

An alternative approach for aggregate tour-based or hybrid models was developed for a model for the Knoxville region (Bernardin, 2008; Bernardin and Conger, 2010) and has subsequently been applied in roughly half a dozen other locations. In order to facilitate aggregate modeling and avoid the requirement of simulation and the variation it introduces, the approach is designed to evade the dimensional explosion of the solution space by partitioning the problem into two location choices, regardless of the number of stops on a tour. In the first stage, all stop locations (of a particular purpose, for a particular market segment) are chosen jointly based on their distance from the anchor location (home) and their accessibility to each other. In the second stage, each chosen destination for a trip is assigned an origin from which it is visited, with the constraint that the number of trips to each location must equal the number of trips from that location over the course of the day.

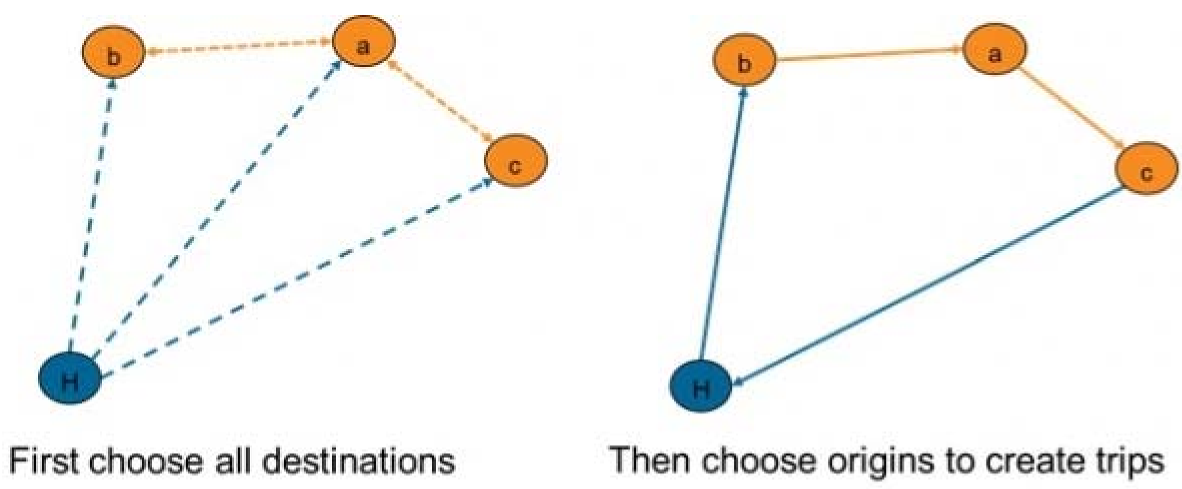


Figure 29. Stop location and sequence choices.

Source: FHWA

This approach can represent the jointness of destination choices including agglomeration effects in stop location choice (Bernardin et al., 2009), and avoids the requirement of identifying a primary destination for all tours but can only capture space-time constraints implicitly. Moreover, the second stage choice of origin or stop sequence can be challenging to interpret and is difficult to implement efficiently. For that reason, the following approach was developed.

The Advanced Trip-Based Approach

A third alternative framework for representing the conditional nature of destination choices has also been applied in aggregate trip-based models (Bernardin and Chen, 2016) in more than a half dozen states. The approach does not require a tour-based framework, but rather uses the traditional trip-based framework, simply shifting the choice of destinations for non-home-based trips to after and conditional on the choice of home-based trips.

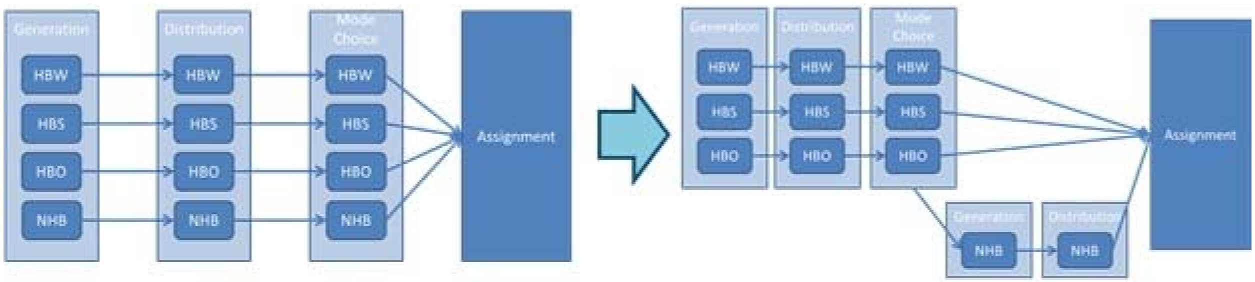


Figure 30. Conditional non-home-based destination choice models.

Source: FHWA

This approach does not require any modification to existing home-based trip component models, and all choices remain easily interpretable. Like the hybrid approach, it can capture agglomeration effects in destination choices, but cannot explicitly represent space-time constraints. It has not been mathematically proven to result in trips consistent with tours with the same rigor as the hybrid approach, but it has been demonstrated to result in more consistent destination (and mode) choices in practical applications. The figure below illustrates how the advanced trip-based approach produces non-home-based-work trip locations consistent with home-based work locations for a new residential development in the far south of the Salt Lake City region in contrast with the traditional four-step model which predicts new non-home-based work trips in the far north of the region even though it predicts no new home-based work trips there.

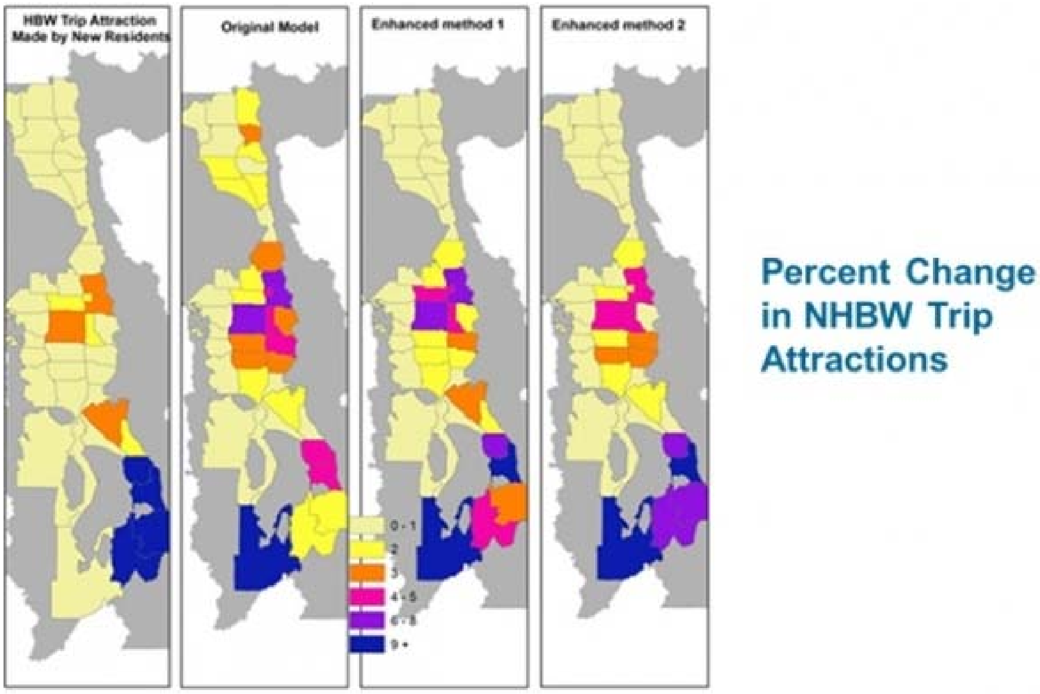


Figure 31. Salt Lake City example from TMIP How-to: Model Non-Home-Based Trips.

Source: FHWA

7.2.2 Destination and Mode Choice

Travelers' choices of destination and mode are importantly related. In the four-step model, mode choice was originally formulated as conditional on destination choice, being applied as the third step after the second step of trip distribution. This approach has the advantage of allowing mode choice models to be formulated using actual travel times between origins and destinations as level-of-service variables. Moreover, this formulation of mode choice models conditional on destination choice has been institutionalized in various guidelines for model development and application (such as for FTA grant applications).

Building on this approach, efforts to link mode and destination choice have most commonly involved the use of mode choice logsums as a multimodal impedance variable in destination choice models. In theory, if properly specified, this approach results in the destination-mode choice model system taking the mathematical form of a nested logit model. Such models can be estimated simultaneously as a single model, but the explosion of alternatives in this framework typically makes simultaneous estimation computationally intensive if not intractable.

More commonly, the models are estimated sequentially, with the results of the mode choice model being used in the destination choice model estimation. This approach has been shown to be a form of limited information maximum likelihood and while unbiased is statistically inefficient compared to full information maximum likelihood estimation.

However, this widespread approach commonly leads to problems such as the inability to predict realistic trip length frequency distributions without the use of other, highly correlated explanatory variables such as distance and the inability to estimate valid parameters for the mode choice logsum variable in the destination choice model such that this parameter is often asserted in practice. Some have interpreted these problems as an indication that the traditionally assumed choice hierarchy is incorrect and formulated alternative joint mode and destination choice models with the reverse hierarchy. In the reverse hierarchy destination choice is conditional on mode choice and destination choice logsums are used as explanatory variables in mode choice.

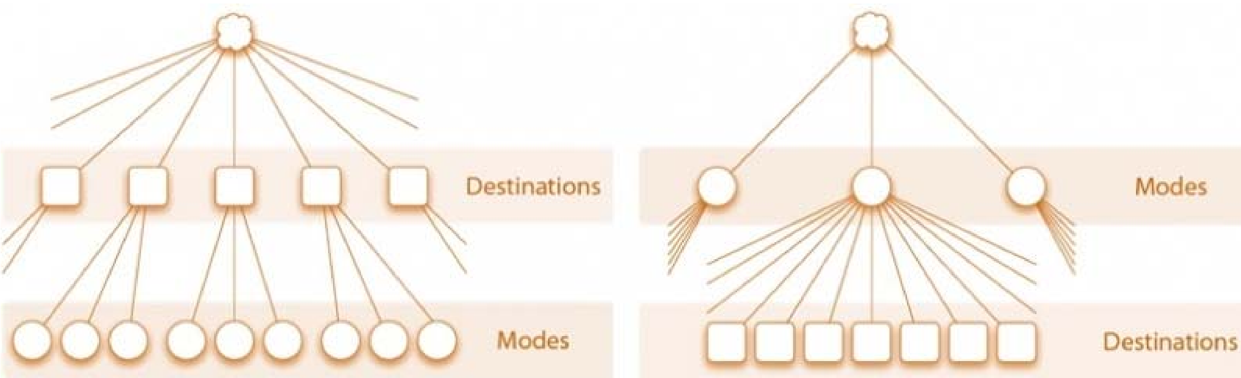


Figure 32. Traditional and reverse mode and destination choice hierarchy.

Source: FHWA

The traditional hierarchy implies that mode choices are more elastic (or more likely to change in response to shocks such as changes in travel times or fares) than destination choices. While this assumption is reasonable in certain travel markets or for certain market segments such as choice riders with personal vehicles available as an alternative, in many circumstances particularly in the United States, it may be more reasonable to assume that destinations are more elastic than modes. Transit captive destination choice set may be determined by local bus service, for instance, in many cases. The assertion of the traditional choice hierarchy, by enforcing this hierarchy of elasticities, may be a cause of optimism bias in transit forecasts.

7.2.3 System Equilibrium and Feedback

Originally, trip distribution models were applied using impedance measures based on free-flow or uncongested travel times. In this case the travel times on which the spatial distribution of trips is based are generally inconsistent with the travel times predicted by assignment in the larger model system. Combined distribution models were formulated and the practice of “feedback” developed to address this problem, especially to support emissions modeling which requires realistic travel times or speeds. In the context of travel forecasting, feedback generally refers to iterating the entire or several steps of the travel demand-network modeling system. At minimum, it generally means the feedback of travel times from assignment to distribution or destination choice. It was eventually recognized and proved that feedback models are equivalent to and can be formulated as combined distribution-assignment models.

The feedback process is common and can be required under certain air quality conformity conditions. Although incorporating some form of system feedback is now quite common, there is little consistency in the details of how this feedback is implemented. Feedback can and has been implemented by feeding back trip tables or flow matrices, travel time or skim matrices, link flows or link travel times. Averaging can be done using the method of successive averages (MSA) or fixed factor methods. There are even more different criteria in use to measure the convergence of feedback loops or combined model systems. A key development in recent years regarding feedback methods was the recognition that “naïve” feedback without any averaging of flows or travel times across iterations may not converge.

7.3 Data Driven Applications

There are generally two methods for using travel demand models together with passive origin-destination (OD) data or incorporating this data in travel demand models. The first approach uses travel demand models (usually of more traditional, aggregate designs) to pivot from OD matrices developed from passive data and traffic counts. The second approach instead uses these OD matrices to develop fixed factors or constants which are incorporated into the travel model; this approach is more attractive for activity-based demand simulation models although it can also be applied with aggregate trip-based travel models. The following sections describe and discuss these two similar and related but alternative approaches.

7.3.1 Pivot-Point Methods

The most common approach to using travel demand models together with an independently data-derived trip matrix is to apply the change in OD travel patterns predicted by a model to the data-driven OD matrix.

This approach typically uses rules or a weighting scheme to combine additive pivoting and multiplicative pivoting. Additive pivoting works by subtracting the modeled OD matrix for the base case from the modeled OD matrix for the alternative and adding this difference to the data-derived OD matrix. Multiplicative pivoting works by dividing the modeled OD matrix for the alternative by the modeled OD matrix for the base case and multiplying the data-derived OD matrix by this growth factor. Multiplicative pivoting is generally preferred for normal, moderate growth or changes, but can produce poor forecasts in some cases, particularly when there are very few or no trips for an OD pair in one or more of the matrices. Rules or weighting based on the absolute number of trips are therefore commonly used to select or combine the two basic pivoting methods.

Pivot-point methods have the clear advantage of requiring relatively little or no modification to an existing travel demand model and hence relatively little effort to apply. However, when incorporated in a model rather than used for an individual forecast, they can require careful management and updating of an input file for the base-case modeled OD matrix. This has little impact on the application of the model for routine forecasting, but it can complicate updates to the model, including zone splits.

Pivot-point methods also are attractive because they are straightforward and easy to understand in concept. Many professionals are already familiar with pivot-point methods from their use to pivot individual traffic counts to produce facility-specific forecasts.

Pivoting on ODs rather than highway network link volumes is less familiar to many in the United States, but it has long been common in Europe and Australia and is quickly growing in use in the United States in response to the advent of big OD data. Pivot-point modeling can substantially improve forecasts by removing the error in a travel demand model’s base-case OD matrix. This error is known to be the largest source of error in traffic modeling. Thus, pivot-point methods promise substantially improved accuracy in forecasting. However, pivot-point methods have no effect on the sensitivity of the travel model or resulting forecast to changes in travel time, tolls, land use, or other factors.

This can be viewed in either a positive or negative light. On the one hand, the independence of the model's sensitivity to the approach can alleviate any concerns related to overfitting or over-specification. On the other hand, this same independence of the model's sensitivity to the approach also means that the information in the big OD data does not necessarily improve the sensitivity of the travel model or resulting forecast to changes in travel time, tolls, land use, or other factors. The large amount of error in base-case models suggests the strong possibility of under-specification errors in existing or traditional models which may translate into over-sensitivity of models to travel times, tolls, land-use variables, and other factors, and pivot-point methods do not help to address this issue.

While the inability of pivot-point methods to address underspecification errors affecting model sensitivities is an important theoretical concern, one of the main drawbacks of pivot-point approaches in practice is the inability of applying the approach at the level of disaggregate demand in demand simulation models such as activity-based models or supply chain simulation models. The fixed-factor approach presented in the following section offers an alternative method that can be applied to disaggregate demand simulation models as well as traditional aggregate models. In summary, pivot-point approaches may not be theoretically ideal or practical for use with activity-based or supply chain simulation models, but they are easy to apply with many travel models and can substantially reduce error.

Fixed-factor or constant rich approaches involve a deeper integration of big OD data into a travel model. As such, they generally require more effort, but they can also potentially yield greater benefits than pivot-point methods and are applicable to activity-based or supply chain simulation models as well as more traditional aggregate trip-based models.

7.3.2 Fixed-Factor/Constant Rich Methods

The fixed-factor approach works by incorporating a set of constants into the spatial (gravity, destination, or activity location choice) model components of a travel demand modeling system. These factors are estimated in a statistically rigorous way to allow the model to reproduce expanded big OD data with minimal error. Fixed factors or constants can be specific to individual or groups of origins or destinations or OD pairings.

These constants are importantly different than traditional k factors sometimes used in gravity models in that they can be systematically statistically estimated from a sound support of big OD data; whereas, k factors were developed in an ad hoc fashion based on survey or traffic count data that often could not actually support them. Despite this important distinction, some historical abuses of k factors still make some professionals hesitant or fearful of constant rich approaches. Individuals with a classical statistical background may also have a hesitancy due to fears of over-specification errors. However, while errors and abuse are possible in any statistical modeling, and some level of caution is always an important component in good judgment, in the new context of the availability of passive data, conscientious professionals should reconsider constant rich approaches in an open and unbiased way. The emergence of a new generation of constant rich approaches is driven by a real change in the context of the data and analysis methods available, the evaluation of which is should not be overly burdened by data-poor and poorly structured use of k factors that bear little resemblance to contemporary methods. In addition to passive data, machine learning analysis methods are another new factor driving contemporary constant rich approaches that are also worthy of further consideration. Machine learning also provides a perspective that is more concerned with under-specification errors than over-specification errors, which may be helpful in balancing certain schools of classical statistical thought.

ORIGIN SUPER-DISTRICT	DESTINATION SUPERDISTRICT												GRAND TOTAL
	1	2	3	4	5	6	7	8	9	10	11	12	
1	0.5%	0.1%	-0.2%	0.0%	0.0%	-0.1%	-0.2%	-0.2%	-0.1%	0.0%	-0.1%	-0.3%	-0.6%
2	0.2%	0.2%	0.1%	0.0%	0.1%	-0.1%	0.0%	0.1%	0.1%	0.0%	0.0%	-0.1%	0.5%
3	-0.2%	0.0%	0.3%	-0.2%	-0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-0.1%	-0.3%
4	0.0%	0.2%	-0.2%	0.1%	0.0%	-0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.3%
5	0.1%	0.1%	-0.1%	0.0%	0.4%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%
6	-0.2%	-0.1%	0.0%	-0.1%	0.0%	0.2%	0.2%	-0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
7	-0.1%	0.0%	0.1%	0.0%	0.1%	0.0%	0.3%	0.1%	0.1%	0.0%	-0.1%	0.0%	0.5%
8	-0.1%	0.0%	0.0%	0.0%	0.0%	-0.1%	0.1%	0.3%	-0.1%	0.0%	0.0%	0.1%	0.2%
9	-0.1%	0.0%	0.0%	0.0%	0.0%	-0.1%	0.1%	0.0%	0.6%	0.0%	0.0%	0.0%	0.5%
10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.1%	0.0%	0.3%
11	-0.1%	-0.1%	0.0%	-0.1%	0.0%	0.0%	-0.1%	0.0%	0.0%	0.2%	0.1%	-0.2%	-0.2%
12	-0.3%	-0.4%	-0.2%	-0.2%	0.0%	-0.1%	-0.3%	0.0%	-0.1%	0.0%	-0.2%	0.0%	-1.8%
Grand Total	-0.1%	0.0%	-0.1%	-0.4%	0.5%	-0.3%	0.2%	0.2%	0.5%	0.4%	-0.3%	-0.7%	0.0%

Figure 33. Destination choice models with fixed factors vs. passive cell phone-based ODs in Chattanooga.

Source: FHWA

Fixed-factor methods can be developed in two importantly different ways. First, a sequential estimation approach in which the factors are estimated after and independently of other model parameters is like pivot-point methods in that it does not affect model sensitivities for good or ill, and it is easier to apply. This method has been successfully applied in practice (for instance, in Chattanooga, see Figure 33). Second, simultaneous estimation of fixed factors together with other model parameters requires more effort, but it also offers the potential for better results by addressing likely under-specification errors and potential model over-sensitivities. Over-specification errors are still possible, though this is less of an issue with passive data.

8.0 Destination Choice Model Calibration and Validation

In practice, destination choice models can rarely be applied for forecasting exactly as they are estimated. Calibration adjustments are commonly required for several reasons. Sometimes application of the model to application data sets produce results that differ in some important ways from the results when the model is applied to the estimation data sets. In some cases, such differences can be caused or exacerbated by inconsistencies between the model estimation and application (such as different sources for explanatory variables like income or travel time or the omission of constraints in estimation). Careful and thoughtful adjustments in keeping with good professional judgment can be required to ensure the applied model demonstrates both reasonable ability to replicate observed travel patterns (from both estimation data and in some cases, other independent data sources for validation) and reasonable response properties or elasticities to key variables.

8.1 Validation

Validation refers to the comparison of model results to independent observed data not used in parameter estimation to assess if a calibrated model reasonably represents the actual system. Initial comparison of the results from the application of the estimated model with independent validation data rarely confirms the validity of the model as estimated. The resulting process of adjusting the model specification and/or model parameters to achieve validation is referred to as calibration. Good professional judgment is required to balance the desire to achieve good fit between model results and validation data with the danger of distorting model sensitivities to change.

8.1.1 Comparison of Predicted and Observed OD Matrices

The results of a destination choice model can be validated against OD matrices that are estimated or observed based on independent real-world data not used in model estimation. Since destination choice models are typically estimated from household survey data, the most commonly used sources of validation data are traffic counts, Census journey-to-work data, and more recently, passive OD data.

Traffic counts can be used to validate destination choice models in several ways. Traffic counts along screenlines can be compared to the results of assigning the modeled OD matrix to a network, or in some cases, directly to aggregate district-to-district flows. Traffic counts can also be used to estimate OD matrices using origin-destination matrix estimation (ODME) routines. A comparison of model predicted OD matrices and ODME generated OD matrices can in some cases provide key insights on the validity and performance of the destination choice model, but only if the ODME process is used judiciously with good seed OD data and reasonable constraints on the perturbation of the seed.

Passively collected data from mobile or in-vehicle devices are also now being used to infer OD matrices which can be used to validate destination choice models. In practice, traffic counts and passive OD data can and are often combined by using passive OD data as the seed for ODME or using traffic counts in some other way to expand passive OD data. Combining data from multiple sources (e.g., network sensors, count stations, smartphone traces, and GPS-enabled/connected devices or vehicles) may offer considerable promise in further enhancing the profession's ability to derive accurate OD matrices that reflect ground truth on spatial-temporal patterns of mobility. Comparisons between model implied OD matrices and real-world OD matrices by time of day period can also be valuable in ensuring that the temporal distribution of travel is also being captured accurately in the model system. There are many ways to measure the distance (extent of similarity or dissimilarity) between two matrices (see the discussion of calibration measures below). Although it is desirable for model predicted OD matrices to replicate ground truth conditions closely, care must be exercised in the extent to which model parameters and constants are adjusted to match observed conditions. Analysts should exercise caution to avoid over-fitting to ground truth conditions lest the adjustments in model parameters or constants result in a model that replicates current conditions accurately at the expense of offering robust forecasts in future years under alternative scenarios.

8.1.2 Sensitivity Analysis

As travel demand models are largely used for forecasting and predicting travel demand under alternative built environment and socio-economic/demographic scenarios, it is important to check the reasonableness of a model with respect to its predictive ability. While the ability to replicate ground truth conditions in the base year may be a necessary condition for validation, it is not necessarily a sufficient condition. To be certain of validity of the model under a wide variety of application scenarios, the model should ideally be dynamically validated using sensitivity tests. For a series of scenarios defined by changes in system conditions, the destination choice model should be applied and resulting changes in spatial patterns of travel demand should be examined for their reasonableness, degree of change/sensitivity, and predictive accuracy. In some instances, there may be real-world data (such as that available before-and-after an infrastructure improvement or a major land use change) and the predictive accuracy of the model can be compared against such longitudinal data. The possibility of such before-and-after comparisons is increasingly feasible given the new availability of cost effective passive OD data. Alternatively, the changes in spatial patterns of travel predicted by the model can be checked for reasonableness based on local knowledge of travel demand and its elasticity with respect to changes in system conditions. In some cases, there may also be statistics (measures of elasticity or change) published in the literature; such published sources may offer a basis to confirm the general validity of predictions output by a destination choice model. An example of such a publication is the TCRP 95 Report series that documents Traveler Response to System Changes.

8.2 Calibration and Validation Measures

A variety of measures can be used to evaluate the validity of destination choice models. Comparisons to trip length frequency distributions remain the most common approaches although it has been demonstrated that models can easily be over-calibrated to reproduce trip length frequency distributions at the expense of their ability to accurately reproduce actual spatial interaction patterns. (Ye et al., 2012) Evaluation of the validity of destination choice models should therefore be based on actual comparisons of the predicted versus actual observed OD flows whenever possible. These comparisons frequently take the form of comparisons of district-to-district summaries of more detailed zonal OD flows, such as comparison of modeled county level commute flows to CTPP estimates. However, it is also possible to calculate a correlation coefficient or other goodness-of-fit measures depending on the type of data available. It is important to be aware, however, that spatial goodness-of-fit statistics are dependent on spatial aggregation or the zone system used, such that, for instance, district level statistics will appear better than zone level statistics and zone level statistics appear better than parcel level statistics comparing the same model to the same observed data. This is consistent with a general understanding of forecasting: given some observations about any given person, it is relatively easy to predict what city they work in, harder to predict the neighborhood, and quite difficult to accurately predict the exact building where they work.

8.2.1 Trip Length Frequency Distribution

As noted above, comparisons of trip length frequency distributions (TLFD) are the most common validation checks performed for destination choice models and are generally considered a minimum requirement for documenting model validity in practice. Poor fit of the modeled TLFD to an observed TLFD clearly indicates that the model is not reproducing observed OD patterns. However, the converse is not true - good agreement between modeled and observed TLFDs does not guarantee that the model is reproducing observed OD patterns. There are many OD patterns with the same TLFD some of which can bear little resemblance to each other at all. Moreover, as is also noted above, over-fitting modeled TLFD to observed TLFD can actually degrade their replication of the observed OD patterns as measured by more robust goodness-of-fit measures. (Ye et al., 2012) Therefore, while TLFD checks are generally recommended as a minimum standard of validation checking for destination choice models, whenever possible they should also be accompanied by evaluation of other calibration measures.

The trip length frequency distribution (TLFD) is calculated as the percentage of trips falling into certain ranges of distance (or duration), say 0-1 mile, 1-2 mile, 2-3 miles, etc. Figure 34 shows an example comparison of modeled and observed TLFDs.

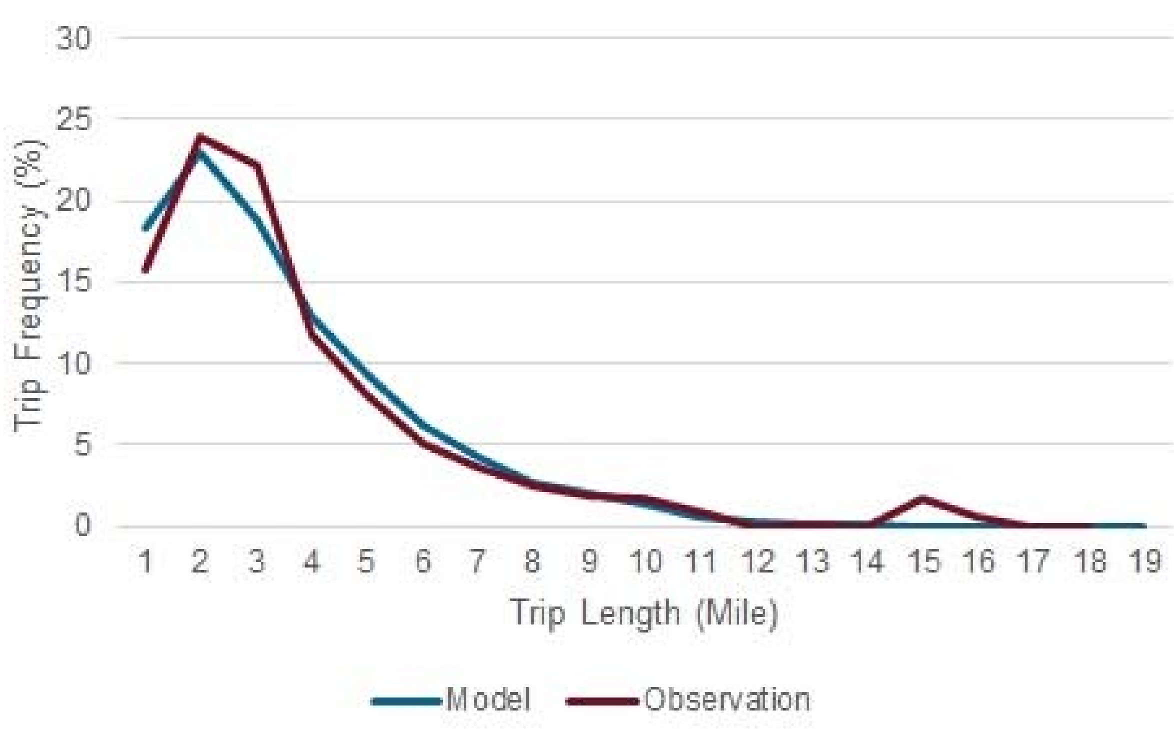


Figure 34. Example comparison of modeled and observed trip length frequency distributions.

Source: FHWA

TLFD checks are usually performed separately for each trip purpose and/or market segment. For work trips, Census journey-to-work data (CTPP and/or LODES) are often used as a validation data source for observed trip lengths, in addition to the data from household travel survey. For non-work trips, model TLFD is usually compared with household travel survey data. It is important to keep in mind that TLFDs from travel survey data can be lumpy due to limited sample size, and in such cases, the model TLFD should be smoother than the survey TLFD. Caution should be used in comparing modeled TLFD to TLFD observed in passive data sources since these data are susceptible to systematic bias with respect to trip lengths which must be carefully corrected using appropriate expansion techniques in order to support valid comparisons.

In addition to visual comparison of TLFDs through graphs, these comparisons are sometimes quantified using metrics such as average trip lengths and/or coincidence ratios. The coincidence ratio measures the area under both TLFDs, as defined below:

$$CR = \left\{ \frac{\sum_{\tau}[\min(PM_t PO_t)]}{\sum_{\tau}[\max(PM_t PO_t)]} \right\}$$

Figure 35. Equation. Coincidence ratio.

CR = Coincidence Ratio

PM_t = Proportion of modeled distribution in interval T

Po_t = Proportion of observed distribution in interval T

T = Histogram interval for time, distance, or other impedance measure

Based on this definition, the coincidence ratio is defined on the range from 0 to 1 with 1 indicating identical distributions and values closer to 1 generally being more desirable within reason (although values very close to 1 may indicate over-fitting).

8.2.2 Intrazonal Trip Percentages

Intrazonal trips are trips with both origin and destination in the same zone and constitute the diagonal of the OD matrix. Comparison of the modeled and observed percentage of intrazonal trips are an important calibration check for destination choice models. Attention to intrazonal trips is important because in model applications, intrazonal impedances or travel times are generally calculated by a different method than interzonal impedance or travel times and both this fact and sometimes differences in the nature of very short trips can lead to unreasonably high or low shares of intrazonal trips. Some agencies have used criteria such as that regionwide modeled intrazonal percentages should be within three percent of observed intrazonal percentages. These criteria may not be reasonable for all applications but large deviations between modeled and observed intrazonal percentages generally indicate the need for a calibration adjustment.

8.2.3 Goodness of Fit Statistics

Formal goodness-of-fit statistics such as log likelihoods and rho-squared (or pseudo-r-squared) statistics are commonly used in the estimation of destination choice models to compare the model's predictions to observed patterns in the estimation data. These statistics are particularly valuable because they can summarize in a single measure the agreement of modeled and observed overall OD patterns - and distinguish, for instance, between different patterns with the same TLFD. Good goodness-of-fit statistics can therefore provide a much greater level of confidence in the validity of a destination choice model than traditional TLFD comparisons.

However, as noted above, it is important to keep in mind that spatial goodness-of-fit statistics are dependent on spatial aggregation or the zone system used, such that, for instance, district level statistics will appear better than zone level statistics and zone level statistics appear better than parcel level statistics comparing the same model to the same observed data. Bearing this is mind, it is difficult to formulate general criteria for goodness-of-fit statistics what constitutes a good value differs with the resolution of the zone system. Moreover, even though the scale does vary with the geographic resolution of the zones used, some generalizations can still be made. For example, aggregate destination choice models at the level of metropolitan traffic analysis zones commonly have rho-squared statistics in the range of 0.15 to 0.45.

While the variation of these statistics with zone sizes does limit the usefulness of these statistics for the establishment of validation guidelines or comparisons between different models, goodness-of-fit statistics can provide very valuable information about how much of the observed variation in OD patterns is explained or reproduced by the model. In particular, the rho-squared statistic can be understood in this way as the portion of the observed variation in destination choices at the level of the zones used explained by the model.

As valuable as they are, in practice these goodness-of-fit statistics are rarely re-calculated after estimation in the process of calibration. Wider use of formal goodness-of-fit statistics as a calibration measure could be an important advance to the state of the practice.

The most common formal goodness-of-fit statistic for destination choice models is the adjusted rho-squared statistic. The adjusted rho-squared is derived by comparing the log-likelihood of the model versus the observed data to the log-likelihood of a uniform discrete distribution versus the observed data. For this reason, rho-squared is sometimes also referred to as pseudo-r-squared because it can be thought of as analogous to or an extension of the common r^2 used as a standardized goodness-of-fit measure for regression models derived from the sum of squared errors. Both statistics are measured on the range of 0 to 1 with higher values indicating greater model explanatory power and goodness-of-fit to observed data. The formula for the adjusted rho-squared is given below:

$$\rho_0^2 = 1 - \frac{LL(model\ distribution) - degress\ of\ freedom}{LL(uniform\ distribution) - degress\ of\ freedom}$$

Figure 36. Equation. Adjusted rho-squared.

In turn, the log-likelihood of a probability distribution versus a set of discrete observations (weighted trips by mode in this case) is defined as follows:

$$LL = \sum_{o,d,m} w_{odm} \ln(P_{odm})$$

Figure 37. Equation. Log-likelihood of a distribution.

Although formal goodness-of-fit statistics are often calculated using disaggregate data in the context of parameter estimation and are thought of in this context, they can be calculated in aggregate modeling contexts as well. Although it is not common to think of trip matrices as probability distributions, they can be easily understood and interpreted this way. Model trip tables can be easily converted into probability distributions by dividing each cell by the grand sum of the matrix (or of all matrices). A uniform probability distribution is simply a matrix (or matrices) of constants equal to the inverse of the dimensions of the distribution (number of zones x number of zones x modes).

8.2.4 Aggregate District-to-District Flows

Interpreting a direct comparison of OD flow matrices is not possible (without the use of goodness-of-fit statistics) at the level of traffic analysis zones when these number in the hundreds or thousands, and even if it were, great care would need to be taken in considering the precision of the observed data at this level of resolution. However, by grouping zones into more aggregate districts, meaningful and valuable comparisons can be made between modeled OD flows between a dozen or so districts. Such checks are a best practice in destination choice model validation. They are most commonly used to compare county-level journey-to-work flows from the model to Census data but can also be used to compare modeled flows for other trip types to household survey or passive data. While less comprehensive and rigorous than formal goodness-of-fit statistics, aggregate OD flow comparisons can provide valuable insights into what aspects of observed OD patterns a destination choice model does and does not capture and sometimes thereby suggest improvements to the model specification or specific calibration adjustments.

The same information can also be presented visually using scatterplots. The example below compares a home-based work destination choice model to Census journey-to-work data from two different time periods.

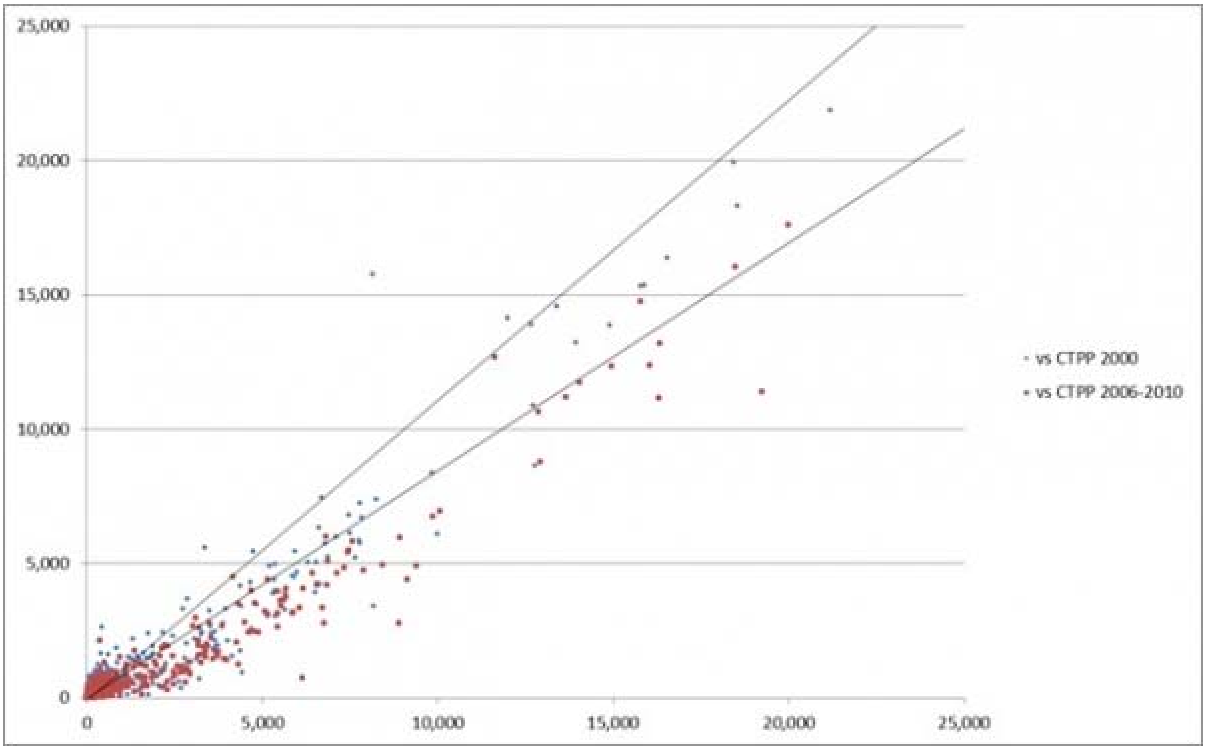


Figure 38. Scatterplot comparing county-to-county journey-to-work flows from CTPP and a statewide model.

Source: FHWA

In some cases, especially such as statewide models, the use of a log-log scale can be helpful for examining the fit of the model at different scales or orders of magnitude.

8.2.5 Screenline Counts

Traffic counts along a screenline or cutline can be used to validate destination choice models in several ways. When a screenline partitions the model zone system into two distinct districts, a direct comparison can be made to aggregate district-to-district flows in the OD matrix. In other cases, the modeled OD matrix can be assigned to the model network and modeled network flows can be compared against screenline counts. In this context, it is important to keep in mind that if the model results do not reproduce screenline counts that this can be the result of problems either with the destination choice model or with the network assignment procedures.

8.3 Calibration Strategies

The objective of model calibration is to ensure that the estimated model, when applied to the entire region, reproduces the aggregate regional travel patterns, as observed in the OD data that was the source for model estimation, as well as in other, independent OD data sources. There are many reasons why the estimated model, when first applied, fails to reproduce the observed travel patterns:

- There may be errors upstream of the distribution model, for example in the level of service matrices, mode choice logsums, trip productions, home locations, employment estimates, and other model inputs.
- There may be an error in the model application software.
- The model may have been estimated with a very small sample of OD records, such that many OD flows may be missing or under/over-represented.
- Some key characteristic of the travel market may have been missed in the model estimation.
- The model estimation may have resulted in insignificant parameters for some variables that are nonetheless important to retain in the final model.
- Accessibility variables and other techniques used to represent agglomeration and destination constraints may not fully capture the effect on the constraints on the estimated parameters.

It is incumbent upon the model developer to identify the most appropriate corrective action – adequate software testing, verification of all model inputs, re-estimation with additional explanatory variables and/or exploring market segmentation. However even after such steps have been taken, it is oftentimes necessary to adjust the estimated model parameters to improve the fit of the model to the observed aggregate data. Possible calibration actions include adjusting the coefficients of the impedance terms, adjusting the coefficient for intrazonal trips and other possible indicator variables, and adjusting the size term coefficients. The calculation of shadow prices for usual work and school location models is generally considered a model calibration step. The flexibility of the model structure allows for accounting for systematic differences that may emerge when the model is applied, and that may lead to additional utility function terms. Finally, when sampling is used to construct the destination choice set, sometimes adjusting the sampling utility function can lead to a better model fit.

Appendix A Theoretical Foundations of Destination Choice

Three primary theoretical starting points for developing destination choice models dominate current practice:

- Gravity models
- Entropy maximization (also known as information minimization) models
- Random utility models

These three modeling approaches are, under appropriate assumptions, mathematically equivalent, and so are special cases of what can be generally called spatial interaction models. All these models attempt to address the same problem, as illustrated in Figure 39, in which spatial interactions (usually trips) between locations in space (typically traffic zones) are to be predicted, given limited, more macro, information concerning these interactions, such as the number of trips originating in each zone and/or the number of trips destined to each zone. For more of the mathematical development of these theories and demonstration of their equivalence, the reader is referred to the TFResource.org page on this topic.

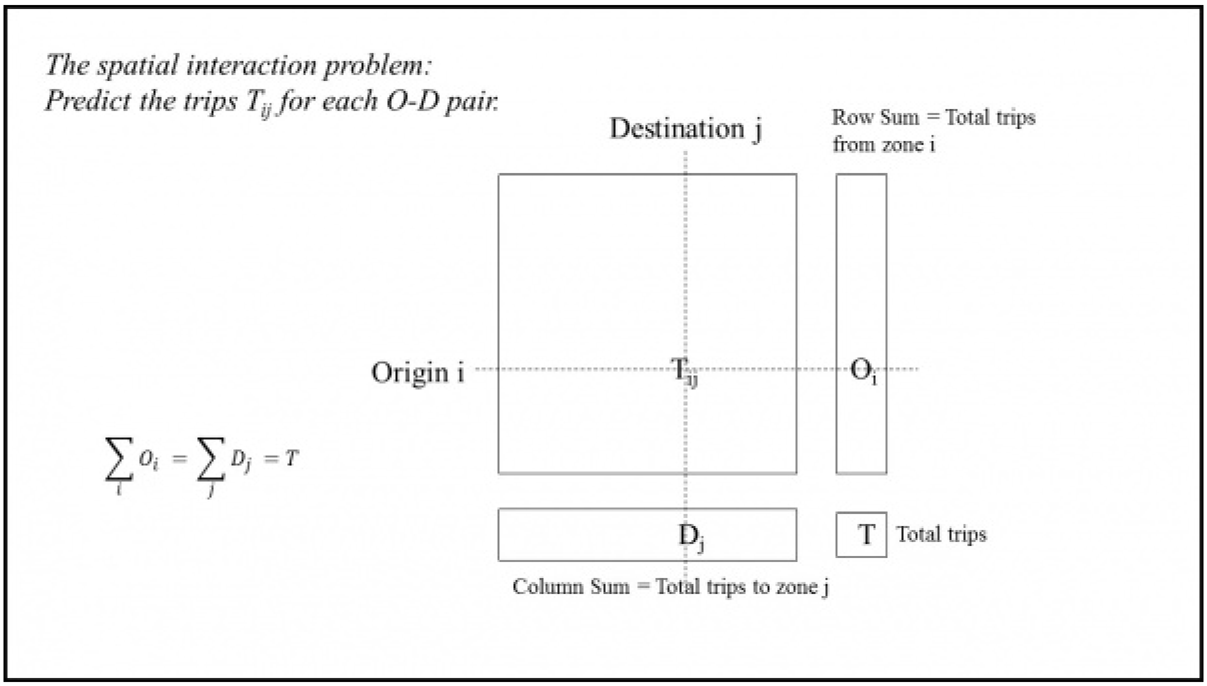


Figure 39. The spatial interaction problem.

Source: FHWA

8.4 Gravity Models

Gravity models have been in use by geographers, market researchers, transportation modelers and many others for well over a hundred years. The starting point for these models, as the name implies, is Newton's Law of Gravity which states that the gravitational force (or interaction) between the bodies is proportional to their masses (size) and inversely related to the distance between them: bigger bodies closer together have a greater interaction.

Gravity models of human spatial interaction adopt the same assumption: the amount of interaction between two locations (usually represented by trips, but could also be flows of money, information, etc.) is proportional to the “size” (“attractiveness”) of the two locations and the extent of their physical separation (measured in distance or travel time).

8.5 Entropy Maximization Models

While it is intuitively plausible that should trips go to “bigger” (more attractive) destinations as well as to destinations that are closer to, rather than farther from, the trip origin, gravity models have always been criticized for their apparently ad hoc derivation: why should human interactions necessarily follow the same “law” as gravitational bodies? Beginning with Alan Wilson’s seminal paper in 1967, a sound statistical theory underlying gravity models was developed. Wilson showed that the statistically most likely trip matrix is given by maximizing the entropy function which results in precisely the doubly constrained gravity model.

In other words, the “ad hoc” gravity model, “properly specified” is the statistically most likely model of a trip OD matrix, given known constraints. This provides very strong theoretical support for “gravity-like” spatial interaction models. Other important points to note include: A specific entropy model specification is determined by the choice of constraints imposed on the model. Arbitrarily complex specifications can be generated, providing that an appropriate constraint set can be specified. In particular, the impedance functional form derives from the constraint(s) written concerning transportation level-of-service variables. In the classic example, imposing the constraint that the predicted system-wide average travel should equal the observed average time in the base data yields a negative exponential impedance function. If instead, one wrote a constraint in which the predicted average of log of travel time equals the observed average value, then the resulting impedance function would take the form of a negative power function.

8.6 Random Utility Models

By far the most common type of destination choice model used in practice is some form of random utility model, usually a multinomial logit model or a nested logit model (e.g., a nested destination-mode choice model). Random utility (discrete choice) models are used throughout travel demand modeling given their strong theoretical foundations in microeconomic theory and their practical and efficient analytical function forms. Logit destination choice models are widely used for a variety of reasons including:

- Flexibility in specifying the utility function (any relevant variable can be readily included).
- Readily available parameter estimation software.
- Familiarity with the method.
- Computational efficiency.
- Support for both disaggregate (person-level) and aggregate (trip flows) formulations.

8.7 Mathematical Equivalence of Gravity, Entropy and Logit Models

It is commonplace in the literature to state that “destination choice” (i.e., disaggregate logit) models are superior in performance to “gravity models”. This, however, is a somewhat misleading statement in that it reflects the common practice in terms of how “gravity” and “logit” models are typically implemented, rather than fundamental differences in the mathematics of the two approaches. In practice, “gravity” models are often aggregate (based on O-D flows instead of individual trips) and very simply specified in terms of both attraction/size variables and impendence functions (including sometimes the use of distance rather than travel times). “Logit” models, on the other hand, are usually disaggregate (based on individual trips) and can have an extensive set of explanatory attraction variables in the utility function. Given this typically more extensive set of explanatory variables, it is not surprising that such “logit” models outperform the more simply specified “gravity” models.

But, as Daly (1982), first observed, gravity models can be shown to be a special case of nested logit models where the nests are degenerate, aggregate alternatives. Similarly, Anas (1983) observed, “gravity” models as derived through entropy maximization can be formulated at the disaggregate (individual trip) level as well and can incorporate any number of explanatory variables. In particular, any linear-in-the-parameters utility function typically used in logit destination choice models can be replicated in an entropy model. Further, if consistently defined at the same level of aggregation, the same set of explanatory variables and the same base data are used for parameter estimation, then it can be shown that the estimated parameters for the two models will be identical. Thus, logit and entropy (gravity) models are, in fact, not different models but are mathematically the same model.

This mathematical equivalency with entropy models only holds for multinomial logit models, not for random utility models in general. The ability to theoretically derive logit models from two very different starting points, one behavioral (people choose alternatives so as to maximize their personal utility) and one statistical (deriving most likely choice probabilities given known constraints on these probabilities), however, is striking and arguably reinforces the case for use of logit models in applications where the underlying assumptions of the model (e.g., statistical independence of the alternatives) holds.

8.8 Other Destination Choice Model Formulations

Historically, other approaches to destination choice models have been developed, including intervening opportunities models and competing destinations models. In general, these approaches tend to be computationally more intensive without generating improved fits to observed data than more conventional methods and so are rarely used in current practice. Brief descriptions of these methods are provided here for historical documentation.

8.8.1 Intervening Opportunities Models

The intervening opportunities model was first proposed by Stouffer (1940) and extended by Schneider (1960) and Golding and Davidson (1970). Most recently, a version of the model has been resurrected by McArdle et al. (2012) who motivates the model from the theory of radiation in physics. Stouffer’s original model hypothesized that the number of OD trips is proportional to the number of opportunities at destination zone and inversely proportional to the number of intervening opportunities between the origin and destination. For a detailed discussion of intervening opportunities models, see Hutchinson (1974).

8.8.2 Competing Destinations Models

Competing destinations models (Fotheringham, 1983) have received a fair amount of attention in the geography literature. They have been used in practice by at least one transportation planning agency. Fotheringham's technique, which introduces an accessibility measure, has now become a common practice in destination choice modeling (following Bhat et al., 1998).

8.8.3 Machine Learning

Although they are not known to have been applied in practice, both decision trees (Thill and Wheeler, 2000) and neural networks (Tillema et al., 2005; Fischer and Reismann, 2010) have been proposed and explored as possibly applicable to destination choice.

Appendix B Size Terms in Aggregate Choice Models

Sometimes, a discrete choice is made from a very large pool of possible choices. In these circumstances, it may be useful to aggregate choices together, and represent a set of choices as a single meta-choice. This is particularly common in destination choice models, where the individual possible destinations are aggregated together as traffic analysis zones.

The aggregate choice in many ways represents a nested logit model, with the aggregations corresponding to the nests, except we only observe the choice at the nest level, not at the elemental alternative level.

8.9 Basic Aggregate Models

To start with, we can make some assumptions:

1. The individual elemental alternatives within each zone or aggregate are homogeneous. That is, each such alternative has the same systematic utility, $V_i=\beta X_i$
2. The particular locations of the zonal or aggregation boundaries are arbitrary and have no systematic meaning themselves.
3. The number of individual elemental alternatives within each zone or aggregate is directly observable.

Using these assumptions, we can derive a reasonably simple aggregate/zonal choice model.

The usual form of the nested logit model calculates the probability of an alternative as $P_{nest}\, P_{alt|nest}$.

In the case of aggregate choices, we do not observe the choice, but only the nest, so we only care about P_{nest}. The nested formula for that term is

$$P_{nest} = \frac{exp(V_{nest})}{\sum_{j \in nests} exp(V_j)}$$

Figure 40. Equation. Probability of nest.

with

$$V_{nest} = \mu_{nest} ln\left[\sum_{i \in nests} exp\left(\frac{V_i}{\mu_{nest}}\right)\right].$$

Figure 41. Equation. Utility of nest.

Using assumption 2, we know that μ_{nest} must be 1, as we want the aggregation nesting structure to collapse to a multinomial logit model. Further, our first assumption is that all the V_i are equal, so the terms inside the summation can collapse together, leaving

$$V_{nest} = ln[N_{nest}\, exp(V_I)] = V_I + ln[N_{nest}]$$

Figure 42. Equation. Utility of nest.

with N_{nest} as the number of discrete elemental alternatives inside the nest or size variable.

Under the assumptions we laid out above, estimating an aggregate model is actually quite simple. We can simply define a variable for each aggregate alternative that has a value of ln(N_{nest}), and including it in an MNL model, with a beta coefficient constrained to be equal to 1.

One thing to be careful of in these models: the log likelihood at "zeros" model should include the parameter on log(N_{nest}) equal to 1, not 0. This is because this is not a parameter we are estimating in the model, it is a direct function of the structure of aggregation, which we have imposed externally.

In application, however, sometimes we want to relax some of the assumptions we outlined above, which can introduce some complications.

8.1 Relaxation of Arbitrary Boundaries Assumption

Relaxing the assumption of arbitrary boundaries puts μ_{nest} back into the equation for V_{nest}:

$$V_{nest} = \mu_{nest}\, ln[N_{nest}\, exp(V_I/\mu_{nest})] = V_I + \mu_{nest}\, ln[N_{nest}]$$

Figure 43. Equation. Utility of nest.

The logsum parameter thus appears as a coefficient on log(N_{nest}). This may or may not be a good idea for transportation models. In an intra-urban model, if the boundaries of zones are at the TAZ level, which are small sectors drawn only for modelling purposes, relaxing this assumption probably doesn't make sense. If the boundaries are aligned with political boundaries (counties, towns) that have differing taxing, administration, or other policies, it might be OK to relax this assumption. In a long-distance travel model, if the boundaries are aligned with metropolitan areas, then it is certainly reasonable to relax the arbitrary bounds assumption.

Relaxing this constraint doesn't require any special methods beyond the standard MNL tools. All that is necessary is to relax the constraint on the parameter attached to ln(N_{nest}), so that it no longer must exactly equal 1.0. Of course, we still need to ensure that the estimated parameter is in the interval (0,1]. Also, for the log likelihood at "zeros" model we should still consider the default value of the parameter on ln(N_{nest}) equal to 1, not 0.

9.0 References

Anas, A., “Discrete choice theory, information theory, and the multinomial logit and gravity models”, Transportation Research B 17, 1983, 13-23.

Aerde, Michel, Hesham Rakha, and Harinarayan Paramahamsan (2003) Estimation of Origin-Destination Matrices: Relationship Between Practical and Theoretical Considerations. In: *Transportation Research Record: Journal of the Transportation Research Board* 1831 (-1):122-130. doi: 10.3141/1831-14.

Bart Kosko (1986). "Fuzzy Cognitive Maps" (PDF). International Journal of Man-Machine Studies. 24: 65–75.

Bernardin, V. (2008) A Trip-Based Travel Demand Framework Consistent with Tours and Stop Interaction. Dissertation. Northwestern University, Evanston, IL.

Bernardin, V., F. Koppelman, and D. Boyce. (2009) “Enhanced Destination Choice Models Incorporating Agglomeration Related to Trip Chaining While Controlling for Spatial Competition.” Transportation Research Record: Journal of the Transportation Research Board, No. 2132: p. 143-151.

Bernardin, V. and M. Conger (2010) “From Academia to Application: Results from the Calibration and Validation of the First Hybrid Accessibility-based Model.” Transportation Research Record: Journal of the Transportation Research Board, No. 2176: p. 50-58.

Bernardin, V., S. Trevino, G. Slater & J. Gliebe. (2015) “Simultaneous Travel Model Estimation from Survey Data and Traffic Counts.” Transportation Research Record: Journal of the Transportation Research Board, No. 2496: p. 69-76.

Bernardin, V. and J. Chen. (2016) New Methods for Improving Non-Home-Based Trips in Trip-Based Models. Presented at the 95th Annual Meeting of the Transportation Research Board, Washington, DC.

Bhat, C., A. Govindarajan, and V. Pulugurta. (1998) Disaggregate Attraction–End Choice Modeling: Formulation and Empirical Analysis. In Transportation Research Record 1645, TRB, National Research Council, Washington, D.C., 1998, pp. 60–68.

Borgers, A., and H. Timmermans. Choice Model Specification, Substitution and Spatial Structure Effects: A Simulation Experiment. Regional Science and Urban Economics, Vol. 17, 1987, pp. 29-47

Bowman, J., M. Bradley, Y. Shiftan, T. K. Lawton, M. Ben-Akiva. (1998) Demonstration of an Activity-Based Model System for Portland. 8th World Conference on Transport Research, July 12-17, 1998, Antwerp, Belgium.

Bowman, J., M. Bradley, and S. Childress (2015) Location Choice Model Estimation. SoundCast: Activity-Based Travel Forecasting Model. Puget Sound Regional Council, Seattle, WA.

Cascetta, E. and A. Papola. (2001) Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transportation Research Part C: Emerging Technologies, Vol. 9, No. 4, pp. 249–263. https://doi.org/10.1016/S0968-090X(00)00036-X

Chow, L.-F., F. Zhao, M.-T. Li, and S.-C. Li. (2005) Development and Evaluation of Aggregate Destination Choice Models for Trip Distribution in Florida. In Transportation Research Record: Journal of the Transportation Research Board of the National Academies, Washington, D.C., pp. 18-27.

Daly, A. (1982) Estimating Choice Models Containing Attraction Variables. Transportation Research B, Vol. 16B, No. 1, pp. 5–15.

Daly, A., J. Fox and J. Tuinenga. (2005) Pivot-Point Procedures in Practical Travel Demand Modeling. Presented at the 45th Congress of the European Regional Science Association, Amsterdam, The Netherlands, August 2005.

Fischer, M. and M. Reismann. (2010) A Methodology for Neural Spatial Interaction Modeling. *Geographical Analysis*, 34(3), pp. 207-228.

Fotheringham, A. S. (1983) Some Theoretical Aspects of Destination Choice and Their Relevance to Production-Constrained Gravity Models. Environment and Planning, Vol. 15A, pp. 1121–1132.

Fotheringham, A. S. (1986) Modeling Hierarchical Destination Choice. Environment and Planning, Vol. 18A, pp. 401–418.

Fotheringham, A.S. and D.W.S. Wong (1991) The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. Environment and Planning A, 23(7), pp. 1025-1044.

Fox, J., A. Daly and B. Patruni. (2012) Enhancement of the Pivot Point Process used in the Sydney Strategic Model. Bureau of Transport Statistics, Transport for New South Wales.

Golding, S. and K.B. Davidson (1970) A Residential Land Use Prediction Model for Transportation Planning, Proceedings, Australian Road Research Board, Melbourne, 5-25

Hansen, Walter G. (1959) How accessibility shapes land use. Journal of the American Institute of Planners 25: 73-76.

Hazelton, M. L. (2003) Some comments on origin-destination matrix estimation. In: *Transportation Research Part A - Policy and Practice* 37 (10):811-822. doi: Doi 10.1016/S0965-8564(03)00044-2.

Hirtle, S. and J. Jonides. 1985. Evidence of Hierarchies in Cognitive Maps. Memory and Cognition, Vol. 13, pp. 208–217.

Hutchinson, B.G. (1974) Principles of Urban Transport Systems Planning, New York: McGraw-Hill

Jonnalagadda, N., J. Freedman, W. A. Davidson, and J. D. Hunt. (2001) Development of Microsimulation Activity-Based Model for San Francisco: Destination and Mode Choice Models. In Transportation Research Record: Journal of the Transportation Research Board of the National Academies, Washington, D.C., pp. 25-35.

Lee, Y., V. Bernardin and D. Kall, (2016) Big Data and Advanced Models on a Mid-Sized City's Budget: The Chattanooga Experience. Presented at the 15th National Tools of the Trade Conference, Charleston, SC, September 2016.

Lynch, Kevin (1960). The Image of the City. Cambridge MA: MIT Press.

Mark, D., C. Fresksa, S. Hirtle, R. Lloyd, and B. Tversky. (1999) Cognitive models of geographical space. International Journal of Geographical Information Science, Vol. 13, No. 8, pp. 747-774. http://dx.doi.org/10.1080/136588199241003

McArdle, G., A. Lawlor, E. Furey, and A. Pozdnoukhov. (2012) City-scale Traffic Simulation from Digital Footprints. Proceedings of the Fifth ACM International Workshop on Urban Computing, Beijing, China.

McFadden, D. (1978) Modelling the choice of residential location. Spatial Interaction Theory and Residential Location (Karlquist A. Ed., pp. 75-96). North Holland, Amsterdam.

Miller, H. and S.A. Bridwell (2009) A Field-Based Theory for Time Geography. Annals of the Association of American Geographers 99(1), pp. 49-75.

MTC and PB, Inc. (2013) Travel Model Two: Initial Representation of Space. Metropolitan Transportation Commission with Parsons Brinckerhoff, Inc., San Francisco, CA.

Newman, J., and V. Bernardin. (2010) “Hierarchical Ordering of Nests in a Joint Mode & Destination Choice Model.” Transportation, Vol. 37, No. 4: p. 677-688.

Satomura, T., J. Kim, G. Allenby. (2011) "Multiple-Constraint Choice Models with Corner and Interior Solutions". Marketing Science, Vol. 30, No. 3, pp. 481-490

Schneider, M. (1960) Panel Discussion on Inter-Area Travel Formulas, Bulletin No. 253, Highway Research Board, Washington, D.C.

Shiftan, Y. (1998) Practical Approach to Model Trip Chaining. Transportation Research Record, No. 1645, pp. 17-23.

Stevens, A. and P. Coupe. (1978) Distortions in Judged Spatial Relationships. Cognitive Psychology, Vol. 10, pp. 422–437.

Stouffer, S.A. (1940) “Intervening Opportunities: A Theory Relating Mobility and Distance”, American Sociological Review, 5(6), 845-867

Thill, J-C. (1992) Choice Set Formation for Destination Choice Modeling. Progress in Human Geography 16(3), pp. 361-382.

Thill, J-C. and A. Wheeler. (2000) Knowledge Discovery and Induction of Decision Trees in Spatial Decision Problems. In *Spatial Economic Science*. Springer, Berlin.

Tillema, F., K. Van Zuilekom and M. Van Maarseveen. (2006) Comparison of Neural Networks and Gravity Models in Trip Distribution. *Computer-Aided Civil and Infrastructure Engineering*, 21(2), pp. 104-119.

Train, K. (2003). Discrete choice methods with simulation. Cambridge university press.

Willumsen, L. G. (1981) Simplified transport models based on traffic counts. In: *Transportation* 10 (3):257-278. doi: 10.1007/BF00148462.

Wilson, A.G. (1970) “A Statistical Theory of Spatial Distribution Models”, Chapter 3 in R. Quandt (ed) The Demand for Travel: Theory and Measurement, Lexington, Mass: Lexington Books, 55-82.

Winfield, M., R. Castellano, and Y. Kimmel. (2016) Predicting a New User's First Travel Destination on AirBnB.

Ye, X., W. Cheng and X. Jia (2012) "Synthetic Environment to Evaluate Alternative Trip Distribution Models." Transportation Research Record: Journal of the Transportation Research Board, No. 2302: p. 111-120.

Zhao, Y. and K. Kockelman (2002) 'The Propagation of Uncertainty through Travel Demand Models', *Annals of Regional Science* 36 (1), pp.145-163

Zheng, J. and J.Y. Guo (2008) Destination Choice Model Incorporating Choice Set Formation. Compendium of Papers of the 87th Annual Meeting of the Transportation Research Board, Washington, D.C.

Zhu, Y. and J. Ferreira Jr (2014) Synthetic Population Generation at Disaggregated Spatial Scales for Land Use and Transportation Microsimulation. Transportation Research Record: Journal of the Transportation Research Board 2429, pp. 168-177.

U.S. Department of Transportation
Federal Highway Administration
Office of Planning, Environment, and Realty
1200 New Jersey Avenue, SE
Washington, DC 20590

April 2018

FHWA-HEP-16-078

U.S. Department of Transportation
Federal Highway Administration