EFFECT OF CHOICE-SET COMPOSITION ON ROUTE-CHOICE MODELS

By

AVINASH GEDA

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2014

To myself, my family, near, dear and all!

ACKNOWLEDGMENTS

First and foremost I would like to thank Dr. Siva Srinivasan for both the opportunity and for guidance throughout my master's. I am grateful for everything I have learned from him. Thank you for giving me such an interesting and challenging project.

Thanks also go to everyone who has helped with the technical aspects of my project, in particular to Drs. Tharanga, Toi and Yin. Your advice and guidance has been invaluable.

I would also like to thank Mr. William Sampson, Director of the McTrans Center, for supporting me with the funding for my master's.

Huge thanks go to my family and friends who have supported me throughout everything in life, not just my master's. In particular thank you to Lavanya, Sindhura, Sindhusha and Ramya for sticking with me and making it easier for me to work on the project. You are all amazing and thank you for always being there for me!

Special thanks go to my Mom, who always stood by me. She has taught me to keep my feet grounded while my head is held high, she has taught me to believe in myself and that I can achieve anything I set my mind to, so thank you Amma! Thank you for supporting me throughout my life.

Last but definitely not the least, I thank myself for sitting hours together on the chair of room number 511 with the hope that this project would one day help me earn my master's degree. It paid off!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

EFFECT OF CHOICE-SET COMPOSITION ON ROUTE-CHOICE MODELS

By

Avinash Geda

May 2014

Chair: Sivaramakrishnan Srinivasan
Major: Civil Engineering

Route choice models are the basis for the link level demand values. Unlike the traditional User Equilibrium and System Optimum models discrete level choice models provide a clear and better understanding of the demand forecasts for a future year. GPS based surveys have picked up lot of momentum in the recent past and the usage of the same data for the research helps lapse many limitations that are being caused due to the telephone based or other ways of route choice surveys. These GPS surveys give a clear picture of the actual chosen route. Although the data is available in a more accurate manner, it also takes in equally difficult amount of post processing and analysis of the GPS data to come up with proper understanding of the route choice. This study positions itself in the same context looking at the route choice as an alternative chosen from a set of alternatives that are potentially available for the traveler or decision maker.

CHAPTER 1
INTRODUCTION

The state of the practice approach to modeling route choice assumes that travelers choose the path of lowest generalized cost, with the generalized cost generally being a function of travel time and tolls. This substantive impact of travel times and costs on route choice decisions is also supported by surveys of stated traveler preferences (Wu, 2012) and the several empirical modeling exercises (Dhakar, 2012). At the same time, empirical models have also documented the strong impact of other route attributes such as the number of turns, facility type/speed, and number of intersections.

While many of the earlier empirical studies have been based on small samples of routes and travelers, the emergence of GPS-based surveys in the last decade and the availability of fine-grained GIS-based network data now provide us with substantial opportunity to develop multivariate route-choice models to study the marginal impacts of various attributes on route choice decisions. At the same time, these developments also introduce new challenges. One such challenge (which is also the focus of this study) is the identification of the appropriate choice-set for developing route choice models.

The GPS-based surveys only help us to identify the route chosen by the traveler; these surveys do not elicit information on other alternative routes considered for making the same trip. The availability of GIS-based roadway networks and the growing number of choice-set generation algorithms (Bekhor et al., 2006) allow us to construct several possible alternative routes. However, such processes do result in the generation of a very large number of route alternatives (subject to constraints on computational time and the granularity of the network). While a large number of alternatives may be advantageous in that they allow us to potentially consider the effects of a variety of attributes, they also impose computational constraints on model estimation and application and could reduce the overall predictive power of the model to the extent that irrelevant alternatives are included in the choice set. Therefore, empirical evidence is required to guide the selection of choice-sets for route choice modeling. Our review of the literature indicates that there is much variability in the size and composition of alternatives in the choice-sets used in past modeling exercises. Systematic comparisons of models estimated with varying choice-sets appear minimal.

In the light of the above discussions, the intent of this study is to examine the effects of choice-set size and composition on route choice models. Data from a large-scale GPS survey

along with a fine-grained road network are used to build several models. A choice set generation algorithm (the breadth-first search link elimination procedure) shown in the literature to be efficient for large-scale problems is used and the path-size logit is used as the econometric structure of the route choice models (again, shown to be an effective approach in the literature). Multiple choice-sets are constructed which vary in size (number of alternatives) and composition (the extent of similarity among the different routes in the choice set). Further, the analyses are undertaken separately for short- and long- routes recognizing that is theoretically possible to generate more alternative routes for a longer-distance trip.

The rest of this thesis is organized as follows. Chapter 2 presents a synthesis of the literature. Chapter 3 presents an extensive discussion of the data using several descriptive analysis. Chapter 4 presents the modeling methodology, estimations and the results associated with different compositions of choice sets for short and long trips. Chapter 5 looks into the conclusions of the study and discussion about the future directions.

CHAPTER 2
LITERATURE REVIEW

The process of developing route choice models from GPS data comprises three steps (Dhakar 2012): (1) map matching, (2) choice-set generation, and, (3) model estimation considering similarities among the different alternatives in the choice set. The focus of this study is on the second step of choice-set generation and as such the review of the literature presented in this Chapter focuses on this aspect. In particular, we opine that two characteristic features of the choice sets, namely, the size of the choice set (number of alternatives) and the variability (or similarity) across the different alternatives in the choice set are important aspects that could impact the computational requirements for model estimation/application and the predictive accuracy of the models. The reader is referred to Dhakar (2012) for an extensive discussion of map-matching procedures and to Smits et al. (2014) and Dhakar (2012) for discussions about alternate econometric methods for capturing correlations among the alternatives.

Table 2-1 presents a summary of various route-choice modeling studies from the stand point of choice-set composition. The studies are grouped in the order such that studies dealing with small scale and non GPS data in the beginning and with those GPS related trip data and large sample sizes are towards the end.

A universal choice set is a choice set which includes all the possible routes between a given OD pair. This universal choice set would contain a very large number of routes and cannot be defined practically. So, a subset of this universal choice set is considered in any of the studies related to choice set. It should also be noted that it is also not possible to produce that set of choices what a traveler exactly perceives because of the behavioral limitations and survey to that level is not supported in today's world. So, an analyst has to somehow come up with a set of those routes which are relevant to chosen route. An algorithm is needed to come up with all the relevant routes for a chosen route in a given network. A broad classification of choice set generation algorithms fall into shortest path, constrained enumeration and probabilistic methods. An exhaustive set of choice set generation algorithms available in the various broad classification are Biased Random Walk, Link Elimination, Link Penalty, Branch and Bound, Simulation, Labeling and Stochastic approaches. Many studies also use a combination of these choice set approaches to come up with unique choice set which contains route choices that are relevant to the chosen route (Bekhor et al., 2006 and Prato and Bekhor 2007)

Table 2-1.Overview of Choice Set Generation studies

Study	Data	Choice Set Sizes	Choice Set Generation Algorithms
Frejinger et al. (2009)	Synthetic dataset	Average choice set size 9.66	Biased random walk
Quattrone and Vieteatta, 2011	Road side survey (280 chosen routes) of trucks compared with on-board GPS (52 routes) trips	30	k-shortest path Combination of 5 criteria
Bierlaire and Frejinger (2008)	No real GPS data Reported trip dataset collected in Switzerland 780 observations	45	Stochastic Choice Set Generation
Bekhor et al. (2006)	No GPS data Questionnaire survey of faculty and staff at MIT, Boston Home-to-work 159 observations with choice sets consisting more than 1 routes	Max 51, median 30 and 25% have at least 40	Labeling and Simulation (48 draws)
Bovy et al. (2008)	Two datasets from different regions, Turin (182 OD pairs) and Boston (91 OD pairs)	Total of 188 routes for Boston and 236 for Turin.	Branch and Bound
Prato and Bekhor (2006)	No GPS data Web-based survey in Turin, Italy Home-to-work 236 chosen routes 339 possible alternatives 182 different ODs	Max 55, median 32 and 25% have at least 40	Branch and Bound
Prato and Bekhor (2007)	No GPS data Web-based survey in Turin, Italy Home-to-work 216 observations with at-least 5 alternatives except the chosen route	Branch & Bound: max 44, median 17,6.36% at least 40 Merged: max 55, median 32,31.36% at least 40	Link Elimination, Link Penalty, Simulation and Branch and Bound
Bekhor and Prato (2009)	No GPS data Turin and Boston datasets	Total of 188 routes for Boston and 236 for Turin.	Branch and Bound

Table 2-1.Continued

Study	Data	Choice Set Sizes	Choice Set Generation Algorithms
Spissu et al. 2011	GPS data from smart phones All purpose 393 observed routes Two leg survey of 1 week each	11 (one chosen route)	Simulation using Min-cost algorithm through existing Cagliari model with cost function of time and distance
Papinski and Scott, 2011b	GPS data Home-based Work Trips In all 237 trips	9 (but only 52 out of 237 could only produce)	k-shortest path algorithm
Frejinger and Bierlaire (2007)	Borlange GPS data 2978 observations 2244 unique observed routes 2179 OD pairs	Min 2, max 43, 93% less than15	Link Elimination
Schussler and Axhausen (2010)	GPS data collected in Zurich with an on-person GPS logger 1500 observations	20,60,100	Breadth First Search Link Elimination(BFSLE) and SCSG (Stochastic Choice Set Generation)
Dhakar (2012)	GPS data from Chicago for 2880 trips	5,10,15	Breadth First Search Link Elimination
Hood, Sall and Charlton (2011)	GPS data from San Francisco for 2777 bicycle trips	Doubly stochastic: avg. 51, Link Elimination: 96	Doubly stochastic: combination of stochastic and labeling, and Link Elimination

It should be understood that if the choice set generation approach is labeling, the size of choice set for all the OD pairs would be definite and constant, unless until there are any duplicate routes. It is the case with the other approaches like Link Elimination or Link Penalty etc. which introduces the irregularity in the size of choice set.

Once the choice set generation is done then comes the step with the sampling of the choices in the choice set. While the studies dealing with smaller scale data can actually consider the exhaustive level of choices generated, for data which has large data it becomes computationally intense and also there is no evidence why not to make the size of choice set same across all the OD pairs used for model estimation.

Frejinger 2009, Bekhor et al., 2006, Prato and Bekhor 2006, Prato and Bekhor 2007 used different approaches in comparison and combination to come up with choice sets for different OD pairs, they ended up building models with variable choice set size across the OD pairs considered for model estimation. Details about the choice set sizes and variations for the studies mentioned are available in Table 2-1.

While the large scale implementation of variable choice set size across the OD pairs used for estimation model is evidently available only with the study Frejinger and Bierlaire (2007) for the Borlange data set, most of the trips have choices generated are less than 15 (reported 93%), which can be a serious issue with respect to the limitations of the size of the network which might enable the traveler to have more than 15 alternatives.

The large scale implementation with constant choice set size across the OD pairs for the model estimation is done by Schussler and Axhausen 2010. They evaluated the models for different choice set sizes of 20, 60 and 100. While 60 and 100 might sound too unrealistic to say

that a traveler considers those many alternatives, they also failed to look at the similarities between the different choices in the choice set.

Another large scale implementation study which falls in line with Schussler and Axhausen 2010 is Dhakar 2012. While the choice set sizes remain constant across the OD pairs for model estimation and the sizes considered for different models being 5, 10 and 15, this study also does not really explain the sampling of these choices based on the variation among them.

One more study which comes into the category of the large scale implementation is the study by Hood et al., 2011, on the bicycle route choice. Although this study has its differences in terms of road networks, this study looks at the insights for bicycle network design. They looked at 2777 bike trips from 366 users. There were two types of choice set generation approaches employed for this of which the first one is the doubly stochastic method, which combines the stochastic and labelling approaches, and the reported average choice set size using this approach after filtering is 51, the second approaches is the link elimination using the Dijkstra's algorithm, and the reported choice set size is 96 for each OD pair. We believe is there is no evidence reported in the study which looks into account for the composition of the choice set.

Sampling is another impediment that occurs after the step of choice set generation. Ben Akiva 2009 addressed this using a sampling correction factor in the Path Size Logit model where they use a factor for correction of utilities which are considered for calculation of probabilities. While this accounts for the choice set generated to the choice set selection step, there is a need a for an approach to come up with the selection of alternative set for modeling from the choice set generated. As most of the studies listed in the Table deal with the exhaustive set of choices generated into the selection of the choices for model estimations, Schussler and Axhausen 2012 addressed this in a very detailed classified manner suggesting four different types of alternatives

selections for model estimations. They are Random reduction, Similarity distribution-based reduction, similarity-based reduction and rule-based reduction. While random reduction is a simplistic selection of choice set, similarity distribution-based and similarity-based look into the overlaps between the choices in the choice set with respect to the Path Size factor calculated as suggested from Ben Akiva and Bierlaire 1999, and the rule based reduction, removes the routes from the choice set which violate defined thresholds such as length, travel time etc. Implementation of all four reduction types are available in detailed in Schussler and Axhausen 2010. Dhakar 2012 employed BFSLE algorithm for choice set generation and employed a simple random reduction rule of selecting the first 'n' routes in the generated choice set, where 'n' is the size of the choice set.

A simple network level sampling is studied by Flotterod, et al., 2011, which essentially deals with the sampling of alternatives given a network, origin and destination. Although the algorithm presented has a clear approach for the sampling point, this study can be irrelevant in terms of concentration of the current study as it looks into the sampling after the choices being generated.

Once the selection is done, the most important step would be the preparation of the choice set for model estimation. It should be made sure at this step that the chosen route is included in the choice set. While those studies with no GPS data would add a chosen route estimated from survey based, GPS based data provide much realistic chosen route inclusion in the choice set.

Once the choice set is prepared, the models are to be estimated. It starts with a simple Multinomial Logit Model (MNL) which looks into the utility maximization of the choices involved. This looks at the deterministic and error in the utilities to be determined.

Enhancements to the simple MNL model can be like Cross Nested Logit (CNL) models etc. But there advanced sophistications to the simple MNL models with the introduction of correction factors in the deterministic component of the Utility term.

A detailed discussion on different model structures can be found in Dhakar 2012.

Prato et al., 2006, provides a detailed snapshot of the various "state of the practice" algorithms in choice set generation and also the various model structures for the model building. Although, there is no validation with respect to the real time data, the study critiques the various difficulties involved in different steps of the model building.

A discussion on the effect of different sizes of choice set is reported by Bliemer et al., 2008, for a single OD pair and synthetic data for choice set sizes of 6, 10 and 12. They researched on the effects of the different sizes across the different types of models and concluded that most of them do not have robustness of choice predictions at individual level.

Recent study by Vreeswijk, et al., 2014, look into the route choice perception of the travelers in the Dutch city of Enschede and found some interesting results of the perception in shortest path travel time and perceived travel time on the other choices. They reported that perception by the travelers is overestimated in both the cases.

The study by Prato et al., 2007, accounts for different choice set generation algorithms and different choice sets. They estimated models across the different model structures and observed that non-nested models perform better across all the various choice set compositions. The stress of this study is more on the model structure and did not really look into the aspects of homogenous size across different OD pairs and the overlap threshold with the actual route was imposed with overlap measure. As such, this is the closest study to our research but the objective was different in the first place.

Literature Review Summary: The crux of this study concentrates on the choice set composition, which can be deduced only after a choice set is generated. Ben-Akiva, Ramming, Schussler have tested all the above listed approaches for choice set generations. Schussler and Axhausen (2010) and Dhakar (2012) reported that link elimination is by far the best possible approach and hence the same is adopted for this study and a Breadth First Search Link Elimination (BFSLE) is employed to generate a choice sets for this study.

Moving forward, coming to the point of choice set composition, although there are variations with respect to the choice set size in various studies and similarity based reduction w.r.t path size factor are employed in Schusller and Axhausen (2010), as of our knowledge the uniqueness of the route in terms of distance overlap in the choice set is not exclusively available as a criterion of choice selection. This can be brought up with a commonly factor that relates to how one route is different from another. The commonly factor (CF) calculates the overlap between two routes in terms of distance and is calculated as $CF = \frac{L_{ij}}{\sqrt{L_i * L_j}}$, Where L_{ij} is the common distance of route and route ; L_i is the distance of route i and L_j is the distance of route j . Based on the restriction of this factor to various levels we can achieve more varied choices in a choice set. The size of the choice set which is depending on the number of choices in the choice set is also an important consideration before we decide on building the route choice models. These various sizes are researched by Schussler and Axhausen (2010), Dhakar (2012) and many others but the combination of this choice set size with respect to the variability between choices in the choice set lays base for this particular study.

The next step in the sequence after deciding on the choice set composition would be the model structure that should be adopted to build the discrete route choice model. A simplest approach for this purpose would be the traditional MNL model which depends on the utility

maximization. CNL and PSL are a bit advanced to the traditional MNL models. Due to the computational ease and flexibility for calculating the choice probabilities a Path Size Logit (PSL) model is considered for the study where we calculate the Path Size factor for every choice in the choice set which will be include in the Utility functions which determine the utility which in turn is used in the calculation of choice probabilities.

On a whole this study looks the choice set composition from the both the estimation and application perspective. By application, it means that the estimation of models is done on choice sets of different compositions and thus built models are applied onto different choice set compositions so as to observe how they compare in terms of various metrics and certain conclusions drawn from these comparisons are presented.

CHAPTER 3
DATA

The two major components of data used in this study are the GPS streams and the GIS-based roadway network characteristics. The GPS data come from the in-vehicle GPS-survey component of the Chicago Regional Household Travel Inventory (CRHTI) conducted, between January 2007 and February 2008. The raw data comprises of 6,089,852 GPS points from 9941 trips made by 408 HH vehicles (259 HHs).The roadway network for the study area was obtained from ArcGIS Data and Maps from ESRI.

Data Exploration: The analysis sample used in this study comprises 2742 trips between 2742 unique origin-destination pairs. These trips are at least 2 miles and 5 minutes long, and had a reasonable "chosen path" identified from the map-matching algorithm. The reader is referred to Dhakar (2012) for an extensive discussion of the map-matching and other data-generation procedures.

Figure 3-1 represents the percentage distribution of 2742 trips during different times of day. Early morning trips are those which have the mid-point time of the trip before 7AM, AM peak trips are those with have the mid-point time between 7AM-9AM, AM off peak trips are the remaining trips which have the mid-point time between 9AM and noon. PM off-peak trips are those with mid-point time between noon and 4PM, PM peak trips are those with mid-point time between 4PM-6PM, and evening trips are the remaining trips with mid-point time later than 6PM and before midnight. Figure 3-2 shows the day-of-the-week variation. Figure 3-3 identifies the trip purpose. Specifically, 22.6% of the trips are home-based originating at home, 28.12% of the trips are home-based with destination at home (i.e., "return home" trips), and 40% are non-home-based trips (neither end of the trip is home). It was not possible to classify the trip purpose for a few trips since the home location of the traveler was not definitely known and so the purpose of

these trips is labeled as “unknown”. These statistics show that the trips considered in our sample have a large variety and are unlike those considered in several other studies which are often focused on specific purpose (such as commute).

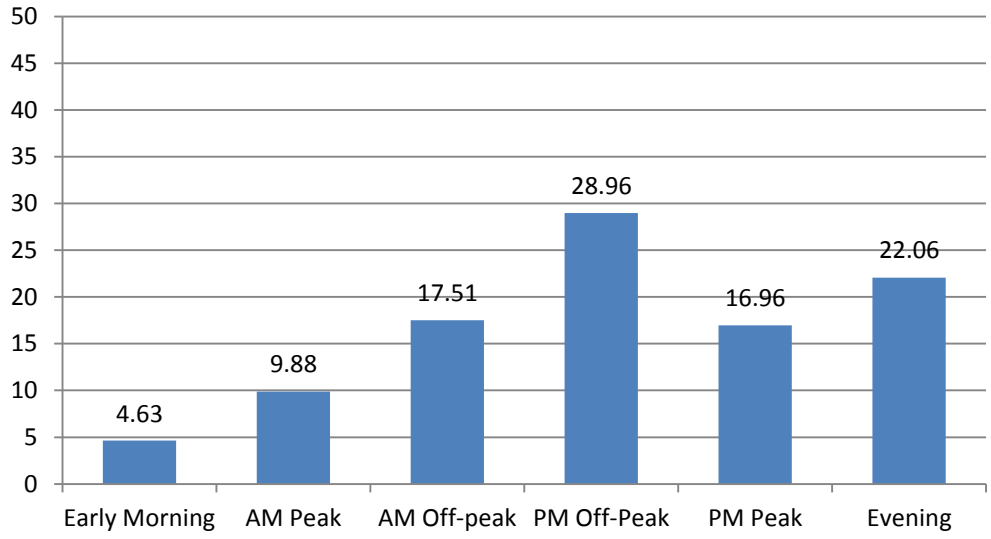


Figure 3-1.Percent Time of day trip distribution

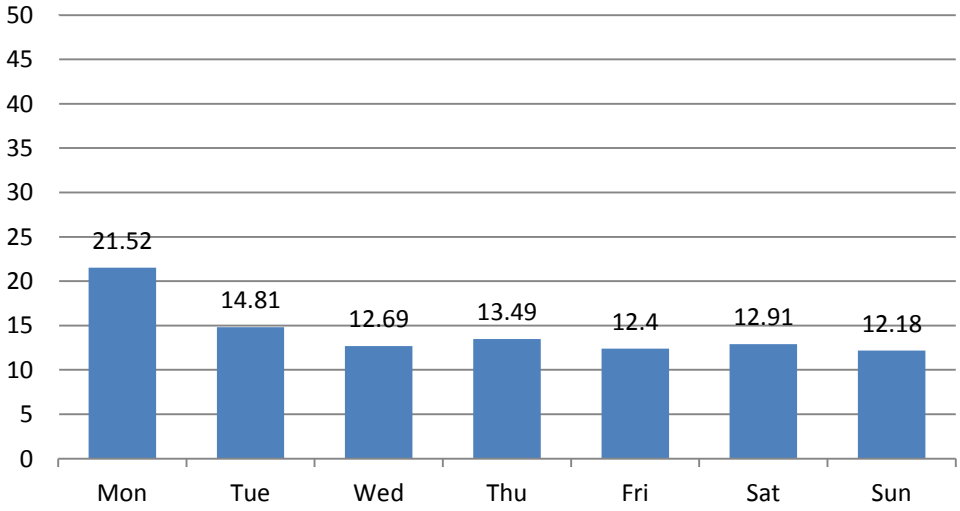


Figure 3-2.Percent Day of week-trip distribution

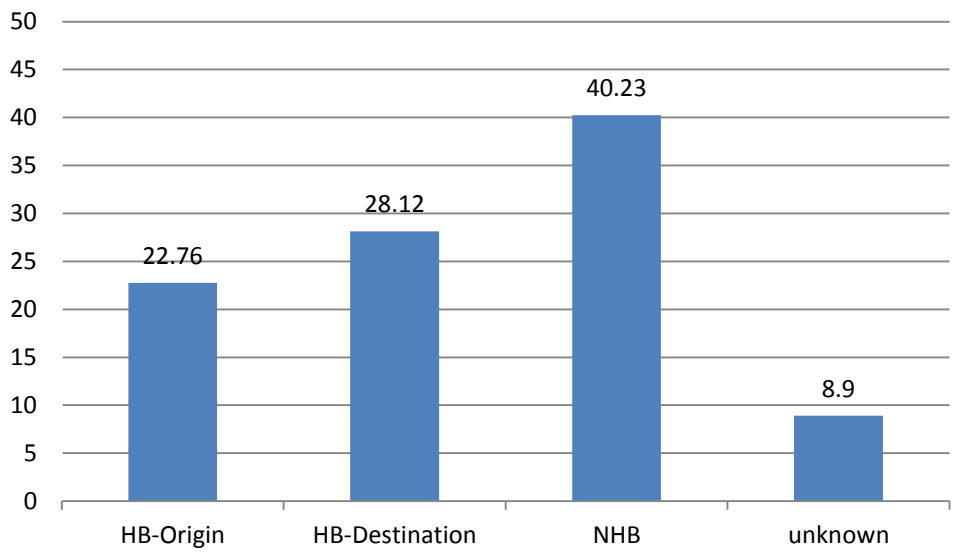


Figure 3-3.Trip Purpose distribution

The following Table summarizes several characteristics of the chosen route that were generated from by mapping the GPS traces to the GIS roadway network (the reader is referred to Dhakar 2012 for a discussion of the generation of these attributes).

Table 3-1.Route attributes' descriptive.

Attribute	Mean	Std. Dev.	5th percentile	Max
Total Distance (miles)	9.39	8.55	2.41	57.12
Total Time (minutes)	15.61	12.45	4.44	83.30
Intersection Count	20.73	16.29	4.00	136.00
Longest Leg Distance (miles)	3.83	6.22	0.44	48.52
Longest Leg Time (minutes)	5.64	7.66	0.81	58.60
Total Turns	3.16	2.36	0.00	22.00
Left Turns	1.48	1.33	0.00	10.00
Right Turns	1.54	1.40	0.00	12.00
Sharp Left Turns	0.09	0.30	0.00	3.00
Sharp Right Turns	0.05	0.23	0.00	2.00
Expressway Distance (miles)	1.10	4.19	0.00	40.43
Expressway Time (minutes)	1.23	4.70	0.00	45.99

Table 3-1.Continued

	Mean	Std. Dev.	5th percentile	Max
Longest Expressway Distance Leg (miles)	1.08	4.13	0.00	40.43
Longest Expressway Time Leg (minutes)	1.21	4.63	0.00	45.99
Arterial Distance (miles)	4.17	5.91	0.00	37.94
Arterial Time (minutes)	5.96	8.15	0.00	57.16
Longest Arterial Distance Leg (miles)	3.45	5.08	0.00	35.02
Longest Arterial Time Leg (minutes)	4.88	6.91	0.00	49.04
Local Road Distance (miles)	4.12	3.48	0.38	26.65
Local Road Time (minutes)	8.42	7.01	1.00	49.44
Longest Local Road Distance Leg (miles)	1.65	2.55	0.00	18.35
Longest Local Road Time Leg (minutes)	3.36	5.05	0.00	37.13
Max Speed (mph)	42.88	7.88	35.00	55.00
Mean Speed (mph)	34.84	6.14	25.96	53.67

Next, alternate routes are added to the choice set of each trip. Starting with the shortest free flow travel time path, alternatives are generated using the BFSLE algorithm (Dhakar 2012 for details). At first, all new routes generated were added into the choice set irrespective of the extent of overlap of the newly generated alternative with any of the previously generated alternatives (in other words, the maximum permissible commonly factor of the new route with any of the previous routes is 1). The algorithm was run for a maximum of 2 hours per trip or 100 choices limit.

The following Table 3-2 summarizes the number of alternatives generated. The reader will note that the results are presented separately for short (2-7 miles) and long (7+ miles) trips. The motivation for this is twofold. First, one might expect that shorter trips to have inherently fewer alternate route options than longer trips. Second, shorter trips are more likely to “closer” to the shortest-path trips than longer trips. To be sure, of the 1437 short trips in the sample, 366

trips (25.47%) have the shortest free flow travel time path overlapping with the chosen route for at least 95% while 160 of the 1305 (12.2%) long trips have the shortest free flow travel time path overlapping with the chosen route for at least 95%. As the choice-set generation algorithm starts with the shortest-time path and generates options, perhaps a greater number of alternatives are required for longer trips so that alternatives "close" to the chosen option are included. Finally, we acknowledge that the choice of cut-off distance for short and long trips is rather arbitrary, but with the 7 mile limit, we find that 52.4% of all trips are short and 47.6% of all trips are long giving us a reasonable sample size for modeling each case separately.

Table 3-2.Percentage distributions of the 2742 trips for various Choice Set size levels for short, long and total trips for Commonly Factor value '1'.

Number of Alternatives generated	Short distance trips (2-7 miles)		Long Distance trips (7+ miles)		Overall trips	
	Number of cases	Percent number of cases	Number of cases	Percent number of cases	Number of cases	Percent number of cases
At least 5	1431	99.6	1302	99.8	2733	99.7
At least 10	1401	97.5	1285	98.5	2686	98
At least 15	1338	93.1	1245	95.4	2583	94.2
At least 20	1165	81.1	954	73.1	2119	77.3
At least 25	1056	73.5	880	67.4	1936	70.6
At least 30	911	63.4	795	60.9	1706	62.2
At least 35	773	53.8	694	53.2	1467	53.5
At least 40	639	44.5	611	46.8	1250	45.6
At least 45	529	36.8	545	41.8	1074	39.2
At least 50	118	8.2	485	37.2	603	22

The reader will note that practically all trips (98%) had at least 10 routes generated and about 22% of all trips had 50 or more routes generated. Further, 37% of the longer trips had 50

or more routes compared to 8% of the shorter trips. Clearly, the generation of more alternatives for longer trips compared to shorter trips is quite reasonable.

As already indicated the first approach to generating alternatives did not consider the extent of overlap of a new alternative with any of the previous alternatives. Therefore, in the next step, we re-created the choice sets, but in this case, a new alternative was added to the choice set only if it did not overlap more than 95% with any of the previous alternatives already in the choice set (i.e., the maximum permissible commonly factor of the new route with any of the previous routes is 0.95). The following Table 3-3 summarizes the number of alternatives generated.

Table 3-3.Percentage distributions of the 2742 trips for various Choice Set size levels for short, long and total trips for Commonly Factor value '0.95'.

Number of Alternatives generated	Short distance trips (2-7 miles)		Long Distance trips (7+ miles)		Overall trips	
	Number of cases	Percent number of cases	Number of cases	Percent number of cases	Number of cases	Percent number of cases
At least 5	1415	98.5	1036	79.4	2451	89.2
At least 10	1319	91.8	442	33.9	1761	64.1
At least 15	1132	78.8	159	12.2	1291	47
At least 20	865	60.2	52	4	917	33.4
At least 25	673	46.8	17	1.3	690	25.1
At least 30	481	33.5	3	0.2	484	17.6
At least 35	343	23.9	0	0	343	12.5
At least 40	211	14.7	0	0	211	7.7
At least 45	108	7.5	0	0	108	3.9
At least 50	0	0	0	0	0	0

The reader will note the drastic reduction in the number of alternatives generated. Specifically, only 65% of the trips have 10 or more alternatives (it was 98% when no overlap

constraints were applied) and practically no trip has more than 50 alternatives. Further, the reduction in the number of alternatives is more dramatic for longer-distance trips. Clearly, while the BFSLE quickly generates lots of alternatives for the long-distance trips, many of them are fairly similar to some other routes in the choice set.

In a third choice-set generation step, the maximum permissible commonly factor of the new route with any of the previous routes was set at 0.90. The following Table 3-4 summarizes the number of alternatives generated. There is a further reduction in the number of options (especially in the case of long-distance trips).

Table 3-4.Percentage distributions of the 2742 trips for various Choice Set size levels for short, long and total trips for Commonly Factor value '0.90'.

Number of Alternatives generated	Short distance trips (2-7 miles)		Long Distance trips (7+ miles)		Overall trips	
	Number of cases	Percent number of cases	Number of cases	Percent number of cases	Number of cases	Percent number of cases
At least 5	1237	86.1	637	48.8	1874	68.3
At least 10	740	51.5	94	7.2	834	30.4
At least 15	341	23.7	8	0.6	349	12.7
At least 20	142	9.9	0	0	142	5.2
At least 25	60	4.2	0	0	60	2.2
At least 30	22	1.5	0	0	22	0.8
At least 35	9	0.6	0	0	9	0.3
At least 40	0	0	0	0	0	0
At least 45	0	0	0	0	0	0
At least 50	0	0	0	0	0	0

In consideration of the above results, this study will examine the following combinations with reasonable estimation sample size.

Table 3-5.Feasible choice set sizes with good amount of data availability

Max Allowed Commonly Factor among alternatives	Number of alternatives in the choice set	
	Short Distance	Long Distance
1	5,10,15	5,10
0.95	5,10	5
0.90	5	NA

Table 3-6 represents the number of trips available after the initial choice set for CF=1 is subjected to various samplings of choice set sizes and CF value levels.

Table 3-6.Number of trips available with different choice set sizes for various CF levels.

CF	Short			Long	
	CS5	CS10	CS15	CS5	CS10
1	1434	1412	1355	1304	1287
0.95	1423	1347	NA	1134	NA
0.9	1334	NA	NA	NA	NA

In the following Table 3-7, the maximum overlap of the choices in the choice set which will determine the inclusion of chosen route in the choice set is discussed about. The percentage presented in the Table indicates to what extent of that trips proportion for that particular case has a maximum overlap of at least 95% with the chosen route.

Table 3-7.Percent and number of trips with at least one choice having at least 95% overlap with chosen route

CF	Short			Long	
	CS5	CS10	CS15	CS5	CS10
1	521(36.3%)	583(41.3%)	626(46.2%)	229(17.6%)	261(20.3%)
0.95	528(37.1%)	590(43.8%)	NA	245(21.6%)	NA
0.9	527(39.5%)	NA	NA	NA	NA

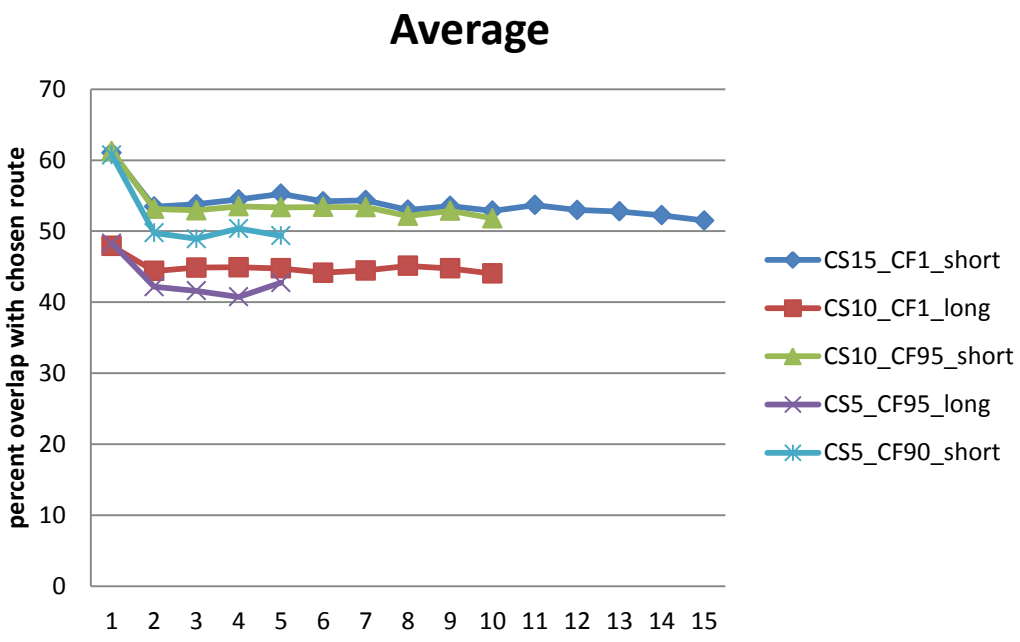


Figure 3-4.Average overlap of nth route with chosen route across the various CS and CF levels.

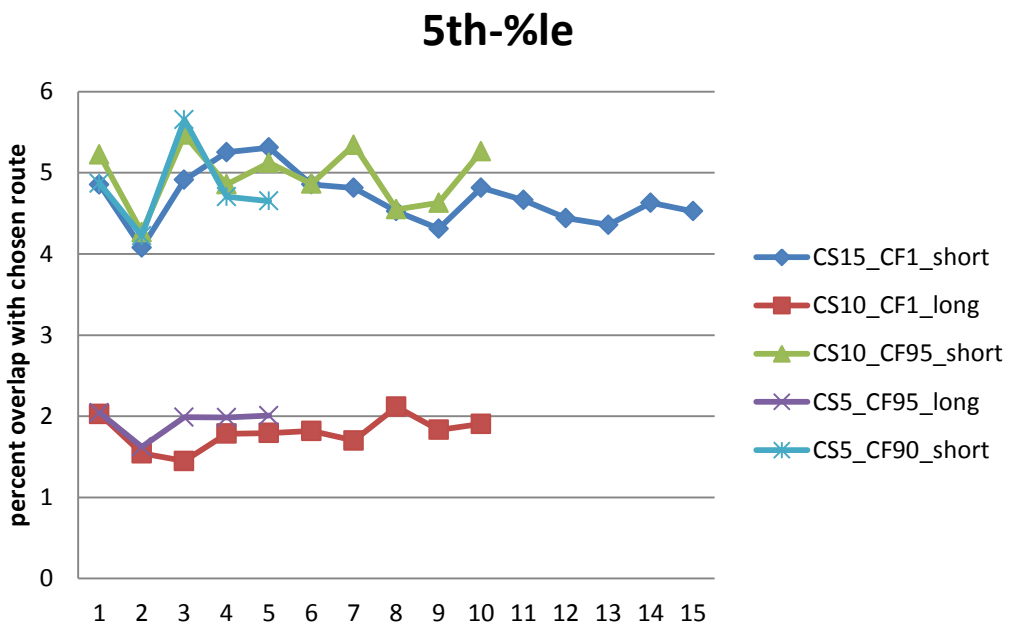


Figure 3-5.5^{th} percentile of overlap of nth route with chosen route across the various CS and CF levels.

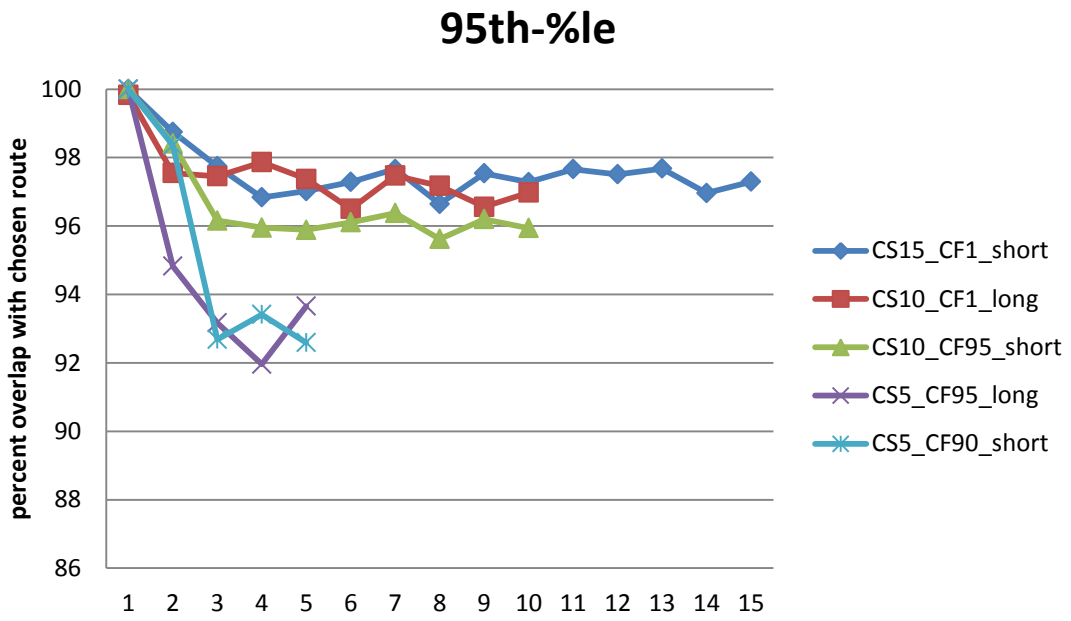


Figure 3-6.95th percentile of overlap of nth route with chosen route across the various CS and CF levels.

The above plots show the average, 5th percentile and 95th percentile, percent overlap of nth route of the choice set with the chosen route respectively. It is evaluated for the extreme cases of each of the long and short trip classifications. For example CS5_CF90_short from the ledger indicates that the plot is for the choice set of 5 routes obtained at the CF level of 0.9 for short trips.

CHAPTER 4
ROUTE CHOICE MODELS

This Chapter opens with the Path Size Logit Methodology adopted for this study and then talks about definitions of various explanatory variables involved in the model estimation and then looks into the cases of short and long trips. The rest of the Chapter is organized as model estimations and applications for short trips and model estimations for the long trips.

Path Size Logit Methodology

Path Size Logit (PSL) is the modeling methodology adopted for this study. Methodology first proposed by Ben Akiva and Bierlaire 1999, it has been proven to show good empirical performance. The model takes into account for the similarity between the choices in the form of a factor called Path Size (PS) factor within the deterministic component of the utility term of the model estimation. The PS factor for route 'i' is calculated based on the formula as reported by Ben-Akiva and Bierlaire 1999 as follows.

$$PS_i = \sum_{a \in \tau_i} \frac{L_a}{L_i} * \frac{1}{\sum_{l \in C} \delta_{al}} \tag{4-1}$$

Where $\tau_i$is the set of links in path 'i',

L_a is the length of link 'a' in path 'i',

L_i is the length of path 'i',

δ_{al} is the link-path incidence dummy, is equal to 1 if 'a' is on path 'i', and 0 otherwise.

It should be noticed that if a path does not share any links with any other routes, its PS factor value will be 1.

Once the PS factor is calculated for all the routes in the choice set, the probabilities are calculated from the probability expression as follows:

$$P(i) = \frac{\exp(U_i + \beta_{PS} * \ln PS_i)}{\sum_{j \epsilon C} \exp(U_j + \beta_{PS} * \ln PS_j)} \quad (4\text{-}2)$$

Where $U_i$and U_j are deterministic utilities for routes 'i' and 'j' in choice set 'C' and β_{PS} is the parameter for the Path Size to be estimated.

Estimation of the models looks in two different classes of the trip lengths as mentioned earlier, short trips which have the trip length between 2 and 7 miles and long trips for all the other trips with trip length greater than 7 miles. Various explanatory variables considered for the model estimations are defined in the Table 4-1.

Table 4-1.Definitions of Route Attributes

Attribute Name	Definition
Total Distance	Total length of the route
Total Time	Total free flow travel time for traversing the route
Left turns per minute	Number of left turns made traversing the route per unit free flow travel time
Right turns per minute	Number of right turns made traversing the route per unit free flow travel time
Intersection Count per minute	Number of intersections along the route per unit free flow travel time
Prop on expressways	Proportion of free flow travel time spent on expressways
Prop on arterials	Proportion of free flow travel time spent on arterials
Prop on local roads	Proportion of free flow travel time spent on local roads
Maximum speed	Maximum speed attained during the trip (speed limit)
Mean Speed	Average speed during the trip
Circuity	Deviation in terms of total length from the straight line distance between Origin and Destination.

A node is considered as an intersection if there are three or more segments meet on that node. Hence, the number of intersections is calculated by determining the number of nodes with three or more segments.

A leg is defined as the stretch of the route between two intersections. Therefore, the longest leg by distance and time is calculated as the maximum leg distance and leg time respectively for a route.

Number of turns in a route are determined by reading the directions output provided by the route solver in ArcGIS. The directions window explicitly specifies the types of turn, if required, along a route. The output also distinguishes the turns in terms of sharp and normal turns. The text in the output is read to determine the number of turns.

The roads in the network are classified into three categories: freeways, arterials, and local roads. The total distance and time on each road types is calculated and then the corresponding proportions are determined. The longest continuous travel (distance and time) made on each road type is also estimated.

Two measures of speed are calculated for a route: average speed, and maximum speed. The average speed is calculated by taking the time weighted average of the posted speeds on the segments of a route.

Circuity is used as a measure of the route distance deviation from the network-free straight line distance between the origin and destination. The straight line distance (SLD) is calculated using the Haversine formula of calculating distance between two points:

$$\text{SLD (miles)} = \text{ArcCos}[\text{Sin}(lat1)*\text{Sin}(lat2) + \text{Cos}(lat1)*\text{Cos}(lat2)*\text{Cos}(long2-long1)]*R$$

Where, lat1 and long1 are the latitude and longitude of a point ‘1’, and R is the earth radius (3949.99 miles).

The circuity is then calculated by taking the ratio of the route length with the straight line distance. The circuity is always greater than or equal to 1.

$$\text{Circuity} = \text{Route Length}/\text{SLD}$$

Short Trips

This is the first classification of the two types of trip lengths considered. The choice sets considered for this case with reasonable estimation sample sizes are 5, 10 and 15 for max commonly factor value of 1, 5 and 10 for max commonly factor value of 0.95 and 5 for max commonly factor value of 0.90. The comparison holds good only when the models are estimated using common set of OD pairs. So, the common set of OD pairs for all the choice set compositions considered came out to be 1249 unique OD pairs. A brief look at the descriptive of each of the explanatory variable associated with each of the route in the choice set is tabulated in Table 4-2. It should be noted that travel distance, free flow travel time, intersection count per minute of travel time, proportion time spent on arterials, proportion time spent on local roads, maximum speed, mean speed and circuity of the chosen route fall very close to that of the average of average of each of the attributes across the alternatives for the common OD pairs. But, the left turns and right turns per minute of the free flow travel time for the chosen route are on an average lesser than the average across the alternatives of the OD pairs considered.

Table 4-2.Comparison of the average of each of the explanatory variables with that of chosen route for short trips

	Chosen Route	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
Travel Distance	3.59	3.61	3.66	3.7	3.63	3.69	3.68
Travel time	6.88	6.62	6.69	6.78	6.66	6.79	6.8
Left turns per minute	0.22	0.4	0.46	0.48	0.41	0.46	0.4
Right turns per minute	0.24	0.42	0.48	0.5	0.42	0.47	0.41

Table 4-2.Continued

	Chosen Route	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
Intersection count per min	2.06	2.23	2.25	2.25	2.23	2.24	2.21
Prop time on expressways	0.003	0.01	0.01	0.01	0.01	0.01	0.01
Prop time on arterials	0.26	0.3	0.3	0.29	0.29	.2891	.2814
Prop on local roads	0.74	0.69	0.69	0.69	0.7	.7007	.7088
Max Speed	39.05	39.85	39.98	40.06	39.87	39.99	39.90
Mean Speed	32.43	33.95	33.96	33.94	33.86	33.82	33.65
Circuity	1.43	1.41	1.43	1.45	1.42	1.44	1.44
Ln(Path Size)	-0.82	-0.97	-1.42	-1.69	-0.93	-1.35	-0.81

The point of variance across different choice set sizes and commonly factor values is explained through the standard deviation of each of the explanatory variables across the different alternatives in the choice set for each of the compositions. The average and standard deviation statistics of the standard deviation of each of the explanatory variables across the alternatives in the choice set are tabulated in the Tables 4-3 and 4-4 respectively.

Although the behavior of variance is not standard across all the explanatory variable considered. The average std. deviation for the travel distance increases when more alternatives are added, keeping the commonly factor constant, indicating that different routes in terms of distance are being added. It also increases when size is kept constant and commonly factor is reduced, which reflects that the similarity of the routes in terms of distance is being varied. But for the case of travel time, when more alternatives are added for the same commonly factor the variance indicator of average standard deviation across the alternatives decrease, which essential tells that more similar routes in terms of travel time are being added, whereas the same increases when the choice set size is kept constant and the commonly factor is varied down, which is an indicator that routes which replace the existing routes are different from the existing similarities.

While there is not much of variability w.r.t the left turns and right turns per minute of free flow travel time across the different compositions, intersection count per minute of free flow travel time do not really tend to change with the number of alternatives being added for a constant commonly factor value, but there is an increase in the average standard deviation across the commonly factor value change. Also while the proportion time spent on different road types follow the trend of left turns and right turns per minute, max speed and mean speed follow the trends of travel distance and travel time respectively. Also, Circuity follows the similar trend of travel distance, while the natural logarithm of path size factor has a different trend, which has the variance indicator decreases with increase in choice set for a constant commonly factor value, but it increases in the other case of commonly factor value for constant choice set size.

Table 4-3.Average of std. deviation of different variables across the alternatives across the unique OD pairs considered for short trips

	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
Travel Distance	0.32	0.34	0.35	0.33	0.37	0.4
Travel time	0.75	0.73	0.73	0.77	0.77	0.86
Left turns per minute	0.19	0.18	0.18	0.19	0.18	0.19
Right turns per minute	0.19	0.18	0.18	0.19	0.18	0.19
Intersection count per min	0.35	0.35	0.35	0.37	0.37	0.42
Prop time on expressways	0.01	0.01	0.01	0.01	0.01	0.01
Prop time on arterials	0.11	0.1	0.1	0.11	0.11	0.13
Prop on local roads	0.11	0.1	0.1	0.11	0.11	0.13
Max speed	1.45	1.5	1.52	1.53	1.62	1.78
Mean speed	2.02	1.83	1.76	2.07	1.9	2.28
Circuity	0.14	0.15	0.15	0.15	0.16	0.17
Ln (Path Size)	0.32	0.45	0.5	0.31	0.43	0.3

Table 4-4.Std. deviation of std. deviation of different variables across the alternatives across the unique OD pairs considered for short trips

	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
Travel Distance	0.35	0.34	0.33	0.36	0.35	0.38
Travel time	0.81	0.74	0.69	0.82	0.77	0.8
Left turns per minute	0.09	0.07	0.07	0.09	0.07	0.09
Right turns per minute	0.09	0.07	0.07	0.09	0.07	0.09
Intersection count per min	0.29	0.23	0.21	0.29	0.23	0.29
Prop time on expressways	0.02	0.02	0.02	0.02	0.02	0.02
Prop time on arterials	0.1	0.08	0.08	0.1	0.08	0.1
Prop on local roads	0.1	0.08	0.08	0.1	0.08	0.1
Max Speed	2.54	2.28	2.23	2.61	2.36	2.8
Mean Speed	1.61	1.21	1.08	1.6	1.2	1.55
Circuity	0.2	0.18	0.18	0.21	0.19	0.21
Ln(Path Size)	0.12	0.12	0.11	0.12	0.11	0.1

The model estimations for the choice set sizes of 5,10 and 15 for commonly factor value of 1, and sizes 5, 10 for maximum commonly factor value of 0.95 and size 5 for max commonly factor value of 0.90 are reported in the Table 4-5.

Table 4-5.Model estimation results for short trips

Variable	CS Size 5 For CF = 1		CS Size10 For CF = 1		CS Size 15 For CF = 1		CS Size 5 For CF = 0.95		CS Size10 For CF = 0.95		**CS Size 5 For CF = 0.90**	
	Est	T stat	Est	T stat	Est	T stat	Est	T stat	Est	T stat	Est	T stat
Total Time	-0.14	-2.51	-0.14	-3.12	-0.16	-3.52	-0.09*	-1.69	-0.14	-3.09	0.02*	0.23
Left Turns/min	-6.15	-19.75	-7	-24.85	-7.11	-27.09	-6.3	-20.38	-7.15	-25.45	-6.2	-20.77
Right Turns/min	-6.39	-20.06	-7.01	-25.2	-7.14	-27.69	-6.29	-20.01	-6.97	-25.19	-6.14	-20.36
Intersection count/min	-0.65	-6.27	-0.54	-5.94	-0.47	-5.43	-0.61	-6.21	-0.55	-6.24	-0.56	-6.02
Prop time on Local road	3.13	9.02	2.62	8.84	2.53	9.05	2.73	8.19	2.57	8.76	2.2	7.23
Circuity	-0.42*	-1.86	-0.76	-3.76	-1.15	-5.36	-0.43*	-1.91	-0.72	-3.5	-0.56	-2.28
Ln (Path Size)	0.62	4.17	0.28	3.3	0.3	4.19	0.31	2.09	0.07*	0.79	-0.47	-2.93
Log Likelihood (at convergence)	-894.95		-1276.82		-1563.54		-911.174		-1293.29		-1004.49	
Log Likelihood (equal shares)	-2010.19		-2875.93		-3382.35		-2010.19		-2875.93		-2010.19	
R-squared	0.55		0.56		0.54		0.55		0.55		0.5	

Notes:

*indicates that the estimate is insignificant at 95% confidence level

R-squared = 1- {Log Likelihood (at Convergence)/Log Likelihood (equal shares)}

As expected, free flow travel time is found to be negatively associated with the probability of choosing a route, which indicates that the routes with higher travel times are not favored, keeping everything else considered for the model estimation constant. But, it is also the case that the travel time is insignificant at 95% confidence level for choice set sizes of 5 where the max commonly factor value allowed is 0.95 or 0.9. The left turns and right turns per minute of free flow travel time have a negative effect with the probability that a route from the set of alternatives can be chosen. While the trend shows that this effect gets stronger with increase in the choice set size for a constant max commonly factor value, it is not really much different across the different max commonly factor values without change in the choice set size. But, the number of intersections per minute of travel time has an intuitive effect in terms of probability of a certain would be chosen, the behavior across the compositions is peculiar, suggesting that the effect goes milder with increase in choice set size for constant max commonly factor threshold and the similar trend when max commonly factor is controlled down and choice set size is kept constant. Another intuitive result, which can be well supported by empirical evidence, is the explanatory variable “proportion time spent on local roads”. As the network is dense urban network of Chicago, and there is very small value for the average proportion time spent on expressways, when compared to arterials in general also, route with more time spent on local roads is favored more between two or more similar routes which have everything intact except the proportion time spent on local roads. And the effect of this variable is very much similar to that of intersection count per minute of free flow travel time in terms of variance in the magnitude across the different compositions of the choice set. Circuity is expected to have a negative effect on the probability of choosing a route, which is shown empirically, that more deviated a route is, from the straight line distance between the origin and destination, less chosen

that route is. Finally, the variable natural logarithm of path size is associated with the probability of choosing a route in both negative and positive ways at different composition levels. The positive sign indicates that a route which is significantly different from the other routes which share more similarity in common. But, while controlling for the commonly factor for a maximum of 0.90, the effect is reversed. While there are studies which reported positive effects for ln of path size in choosing a route (for ex. Prato and Bekhor, 2006; Prato and Bekhor, 2007; Bierlaire, and Frejinger, 2008), there can be chance for negative effects after controlling the similarities to an extent.

On a whole, it can be stated that all the models estimated are reasonably intuitive for all the effects of each of the attributes of the routes.

Cross Applications of Models and Comparisons: In order to better compare the models, or to determine which model performs better on which choice set compositions, brings the point of cross application of the models. While all the models estimated are deployed on to all the choice set compositions assembled, the results observed and discussed below paint a better picture to contrast between the models from the application standpoint also.

Tables 4-6, 4-7 paints a picture w.r.t to average expected overlap and standard deviation of the expected overlap respectively of the predicted route with the chosen route when the cross application of models estimated on the other choice set compositions is performed. Expected overlap is calculated by multiplying the predicted probabilities for each of the routes in the choice set with the overlap of each of the routes with the chosen routes and then adding them together. The expected overlap is given as:

$$E(overlap) = \sum_{i \in C} p_i * O_i \tag{4-3}$$

Where p_i= predicted probability of route 'i' in choice set 'C' when cross application is performed, O_i= overlap of route 'i' in choice set 'C' with the chosen route.

Table 4-6.Average expected overlap when cross application is performed for short trips

	Shortest Path	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
CS5 CF1	61.88	56.87	57.94	58.26	57.32	58.26	58.35*
CS10 CF1	61.88	54.41	56.21	56.69	55.31	56.82	57.27*
CS15 CF1	61.88	53.15	55.37	55.94	54.28	56.11	56.68*
CS5 CF95	61.88	55.94	57.14	57.53	56.42	57.5	57.56*
CS10 CF95	61.88	53.68	55.57	56.08	54.59	56.21	56.66*
CS5 CF90	61.88	54.59	55.95	56.48*	55.02	56.31	56.08

*indicates the highest value in the row excluding the shortest path.

Table 4-7.Std. deviation in average expected overlap when cross application is performed for short trips

	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
CS5 CF1	30.14	30.63	30.75	30.23	30.74	30.53
CS10 CF1	27.89	28.78	28.97	28.17	29.08	29.13
CS15 CF1	26.6	27.83	28.09	27.06	28.26	28.51
CS5 CF95	29.6	30.15	30.31	29.69	30.27	29.97
CS10 CF95	27.1	28.07	28.29	27.38	28.36	28.34
CS5 CF90	27.54	28.27	28.49	27.63	28.42	27.94

Rows in Table 4-6 indicate when the particular model estimated from the one in the column is cross applied on to the one in the row. Reader should notice that the cross application is performed on choice sets that are generated as is, and no chosen route is included in this step.

At an average level shortest path has a better overlap with chosen route than the expected overlap. But, as the purpose of this study determines the comparison across different models for choice set compositions, it can be observed that model estimated with 5 alternatives and maximum commonly factor threshold of 0.90 yields better results in terms of average expected overlap when applied almost all the cases except when the on the same choice set composition it is estimated from. And, the best performance on the choice set size of 5 for maximum commonly factor value of 0.90 is when the model estimated using 15 alternatives for max commonly factor of 1 is applied.

On a whole, the best of the best is when the model estimated using the choice set size 5 for maximum commonly factor value of 0.90 is applied onto the choice set with 5 alternatives for maximum commonly factor value of 1, which can result to a plausible meaning of putting in efforts for estimating the model in terms of controlling the commonly factor and applying it on to simple choice sets generated with no commonly factor limitations. That is essentially same as the application part gets easier and time saving with respect to generation of choice set for a model with efforts in estimation.

Probability of outperforming the shortest path in the choice set can be a good disaggregate level metric to look at the how better the model performs when compared to the shortest path. This metric is essentially the probability of path with an equal or better overlap than the shortest path. It is calculates as follows:

$$P_i^{outperform} = \sum_{j \in C} p_j * \delta_j \tag{4-4}$$

Where, p_j is the predicted probability of route 'j' in choice set 'C' corresponding to OD pair 'i', and δ_j is the overlap performance index, which is equal to 1 if overlap between route 'j' in choice set 'C' of OD pair 'i' and the chosen route for OD pair 'i' is equal to or greater than the overlap between the shortest time path and the chosen route, and 0 otherwise.

Tables 4-8 and 4-9 list the average probability of outperforming the shortest path and standard deviation of the same when cross application is performed.

Table 4-8.Average probability of outperforming the shortest path for short trips

	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
CS5 CF1	0.63	0.67	0.68*	0.65	0.68	0.68
CS10 CF1	0.52	0.57	0.58	0.54	0.59	0.6*
CS15 CF1	0.47	0.52	0.54	0.49	0.54	0.55*
CS5 CF95	0.62	0.67	0.68*	0.64	0.68	0.67
CS10 CF95	0.51	0.56	0.58	0.53	0.58	0.59*
CS5 CF90	0.58	0.62	0.64*	0.59	0.63	0.63

*indicates the highest value in the row in comparison

Table 4-9.Std. deviations in probabilities of outperforming the shortest path across the trips when cross application is performed for short trips

	CS5 CF1	CS10 CF1	CS15 CF1	CS5 CF95	CS10 CF95	CS5 CF90
CS5 CF1	0.33	0.32	0.31	0.32	0.31	0.3
CS10 CF1	0.34	0.33	0.33	0.33	0.33	0.31
CS15 CF1	0.34	0.33	0.33	0.33	0.33	0.31
CS5 CF95	0.33	0.32	0.31	0.32	0.31	0.3
CS10 CF95	0.34	0.33	0.33	0.33	0.33	0.31
CS5 CF90	0.34	0.33	0.33	0.33	0.33	0.32

The findings from the cross application of models in terms of outperforming the shortest path metric show that the model estimated using the highest number of alternatives for no control on commonly factor has the best performance when applied on the least choice set sizes for any

commonly factor threshold level. Where as in the rest of the three cases of choice set sizes 10 for max commonly factor values of 1,0.95 and 15 for max commonly factor value of 1, the model estimated using 5 alternatives and the best max commonly factor threshold case of 0.90 has the best performance.

Best of best performance is for the case when model estimated using choice set size of 15 for max commonly factor threshold value of 1 is applied on to the choice set size 5 for max commonly factor value of 1. It should be noted that values listed in the Tables are rounded to the second decimal place and the explanations are based on absolute values.

Long Trips

This is the second classification of the two types of trip lengths considered. The choice sets considered for this case with reasonable estimation sample sizes are 5 and 10 for max commonly factor value of 1 and 5 for max commonly factor value of 0.95. Just like in the case of short trips, the comparison holds good only when the models are estimated using common set of OD pairs. So, the common set of OD pairs for all the choice set compositions considered came out to be 1127 unique OD pairs. A brief look at the descriptive of each of the explanatory variable associated with each of the route in the choice set is tabulated in Table 4-10. It should be noted that travel distance, proportion time spent on expressways and maximum speed only, of the chosen route fall very close to that of the average of average of each of the attributes across the alternatives for the common OD pairs. But, the average free flow travel time and average circuity of the chosen route are significantly different from the average of the average of the alternatives across choice sets across the 1127 OD pairs considered. The left turns and right turns per minute of the free flow travel time for the chosen route are on an average lesser than the average across the alternatives of the OD pairs. While the proportion time spent on arterials for

chosen route is lesser than that of the average of average for each of the compositions, it is the other way round for the proportion time spent on local roads so as to balance the needle. The mean speed of the chosen route is also significantly lesser than that of the average of the average for each of the choice set compositions considered.

Table 4-10.Comparison of the average of each of the explanatory variables with that of chosen route for long trips

	Chosen Route	CS5 CF1	CS10 CF1	CS5 CF95
Travel Distance	15.59	14.94	15.03	14.88
Travel time	25.07	22.89	23.19	22.71
Left turns per minute	0.08	0.16	0.15	0.18
Right turns per minute	0.08	0.17	0.16	0.19
Intersection count per min	1.23	1.53	1.53	1.57
Prop time on expressways	0.07	0.07	0.06	0.06
Prop time on arterials	0.43	0.51	0.5	0.51
Prop on local roads	0.5	0.43	0.44	0.42
Max Speed	46.87	47.26	47.2	47.31
Mean Speed	37.5	40.23	39.94	40.47
Circuity	1.35	1.3	1.3	1.29
Ln (Path Size)	-0.63	-1.09	-0.9	-1.68

Table 4-11, 4-12 depict the average and standard deviation of the standard deviation of the various explanatory variables defined in the previous section across the alternatives in the choice set compositions considered for all the common unique 1127 OD pairs respectively.

The common trend with respect to variance across the alternatives of the choice set is depicted by the metric of standard deviation across the alternatives for each of the explanatory variables considered. The average of the same is tabulated in Table 4-11. Although, the variance indicator is not changing much for some explanatory variables like left, right turns and

intersection count per unit free flow travel time, proportion time spent on each of the road types, for everything else of the explanatory variables the trend in the variance indicator is pretty much decreasing with increase in number of alternatives or controlling for the max commonly factor threshold. This is very peculiar and contrasts to the short trips considered.

Table 4-11.Average of std. deviation of different variables across the alternatives across the unique OD pairs considered for long trips

	CS5 CF1	CS10 CF1	CS5 CF95
Travel Distance	0.71	0.83	0.61
Travel time	1.68	1.77	1.32
Left turns per minute	0.07	0.07	0.06
Right turns per minute	0.07	0.07	0.07
Intersection count per min	0.32	0.39	0.28
Prop time on expressways	0.03	0.04	0.03
Prop time on arterials	0.12	0.15	0.11
Prop on local roads	0.12	0.14	0.1
Max Speed	1.94	2.48	1.82
Mean Speed	2.26	2.57	1.88
Circuity	0.07	0.08	0.06
Ln (Path Size)	0.37	0.33	0.51

Table 4-12.Std. deviation of std. deviation of different variables across the alternatives across the unique OD pairs considered for long trips.

	CS5 CF1	CS10 CF1	CS5 CF95
Travel Distance	0.79	0.77	0.58
Travel time	1.84	1.78	1.27
Left turns per minute	0.04	0.04	0.03
Right turns per minute	0.04	0.04	0.03
Intersection count per min	0.34	0.35	0.25
Prop time on expressways	0.06	0.08	0.06
Prop time on arterials	0.09	0.09	0.07

Table 4-12.Continued

	CS5 CF1	CS10 CF1	CS5 CF95
Prop on local roads	0.09	0.09	0.07
Max Speed	2.72	2.94	2.27
Mean Speed	1.76	1.71	1.3
Circuity	0.07	0.07	0.06
Ln(Path Size)	0.15	0.12	0.15

Table 4-13 depicts the estimated coefficients for each of the reasonable explanatory variables considered for the model estimation, for each of the different choice set compositions mentioned.

Table 4-13.Model estimation results for long trips

Variable	CS Size 5 For CF = 1		CS Size10 For CF = 1		CS Size 5 For CF = 0.95	
	Est	T stat	Est	T stat	Est	T stat
Total Time	0.02*	0.54	0.12	2.9	0.17	4.13
Left Turns/min	-19.44	-13.37	-20.23	-16.62	-19	-15.16
Right Turns/min	-21.1	-14.95	-23.29	-18.42	-21.3	-16.57
Intersection count/min	-1.38	-6.68	-1.56	-8.83	-1.47	-9.15
Prop time on Local road	4.47	8.04	3.62	7.75	3.48	7.61
Circuity	1.51	2.06	-0.19*	-0.26	0.62*	0.84
Ln (Path Size)	1.91	8.17	0.98	7.55	0.7	2.98
Log Likelihood (at convergence)	-405.41		-598.64		-523.90	
Log Likelihood (equal shares)	-1813.84		-2595.01		-1813.84	
R-squared	0.78		0.77		0.71	

Notes:

*indicates that the estimate is insignificant at 95% confidence level
R-squared = 1- {Log Likelihood (at Convergence)/Log Likelihood (equal shares)}

The estimation results for this case of long trips are counter intuitive from some of the explanatory variables' standpoint. Although the signs of the coefficients left turns/min, right

turns/min, intersection count/min, proportion time spent on local roads are plausible, the signs on travel time and circuity make the case of counter intuition. The possible reasons for this can be explained from the Table 4-10, which says that the travel time of the chosen path is greater than the average on an average, which means that the choices generated starting from shortest path do not really converge to the point of chosen route for the alternatives in the choice set.

Table 4-14.Percentage of trips with at max one trip in choice set having the controversial variable for that trip greater than that of chosen route

Controversial variables	CS5 CF1	CS10 CF1	CS5 CF0.95
Travel Time	80.2%	71.2%	71.9%
Circuity	64.4%	51.8%	60.3%

Table 4-14 looks into detail for the counter-intuition of the signs on the coefficients of explanatory variables, travel time and circuity. This shows better understanding from the perspective of the choice set consideration from the choice set generation perspective. The numbers listed in the Table is the percentage number of trips which have at max only one trip in the choice set except the chosen route having either travel time or circuity to be greater than that of the chosen route. As almost all the numbers listed except the lowest for composition of choice set size 10 and max commonly factor of 1, every other number is above 60%, which means 60% of the data does not have more than route which has travel time or circuity greater than that of chosen route and hence the chosen route is prioritized as per maximizing log likelihood function and the model predicted a positive sign. The statistical insignificance of these variables is due to the strong correlations and the effects of these are picked up in other variables.

CHAPTER 5
CONCLUSIONS AND DISCUSSION

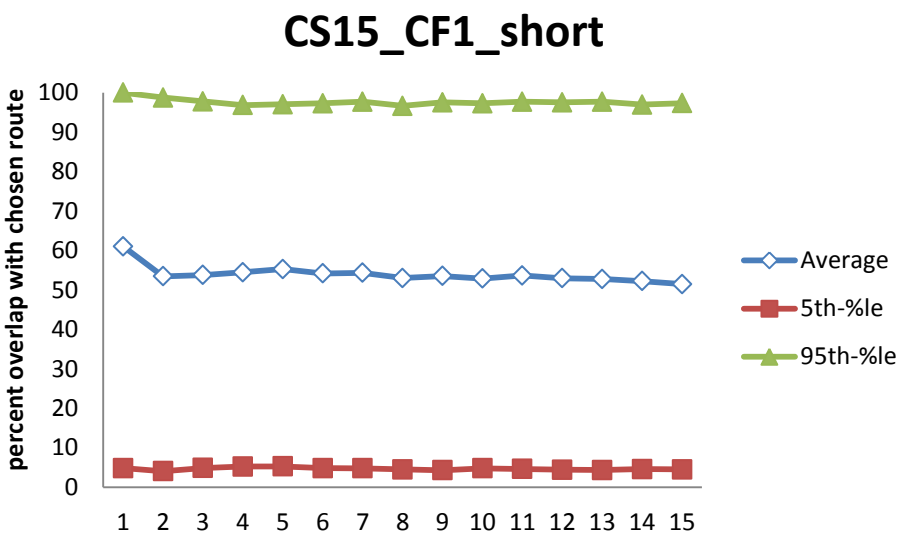
As the objective of the study determines, we looked into the point of effect of choice set composition on route choice models from the estimation and application perspective. Various choice set compositions with different choice set sizes and variability factor between the alternatives in the choice sets are considered for two trip length trips.

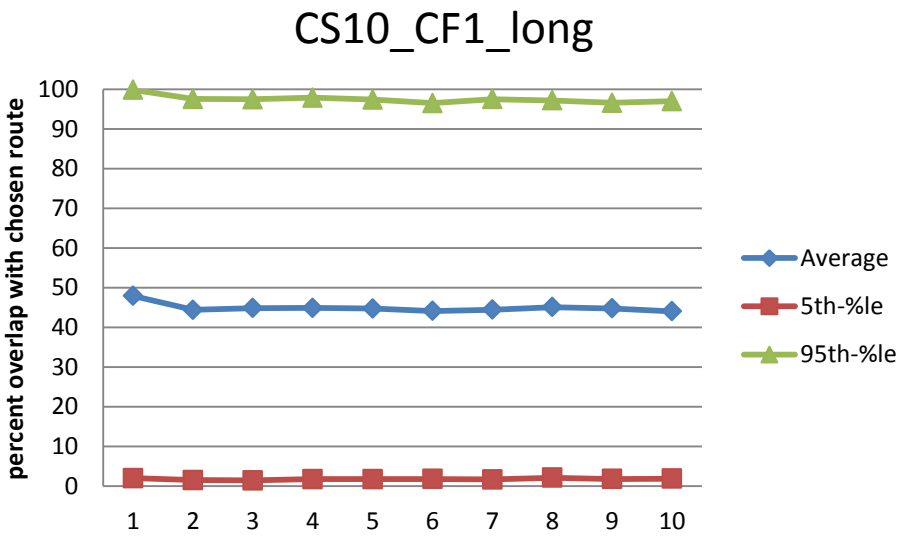
It is observed that the convergence of choice set generation to the chosen route will significantly affect the model structure. This is confirmed with respect to the classification we came up in terms of short and long trips. It is reported in Table 3-7 that the inclusion of chosen route in the choice set is very high in short distance trips than the long distance trips from the choice set generation perspective. Although, chosen route is forced into the estimation choice set for the purposes of estimation, and the choice set generation pattern starts from shortest path and moves forward with the link elimination algorithm, the longer trips had hard time converging to the chosen route with the limits for time and maximum number of alternatives. The choice set sizes considered in this study are not adequate to develop behaviorally reasonable models for long trips. Inclusion of more choice set sizes for long trips do not yield a good estimation sample for this particular study. So, the probable remedy can be increasing the run times or max number of alternatives threshold during the choice set generation step. On the other hand, as the short trips have the highest percentage for the inclusion of chosen path in the choice set, the model structures came out up to the intuition expected with good empirical evidence and quantitative measures.

Limitations: The sampling certainly has some limitations. Although there can be many more methods for sampling in reduction to a fixed size, as suggested by Schussler (2010), we did not employ those methods due to computational limitations. The selection of alternatives for

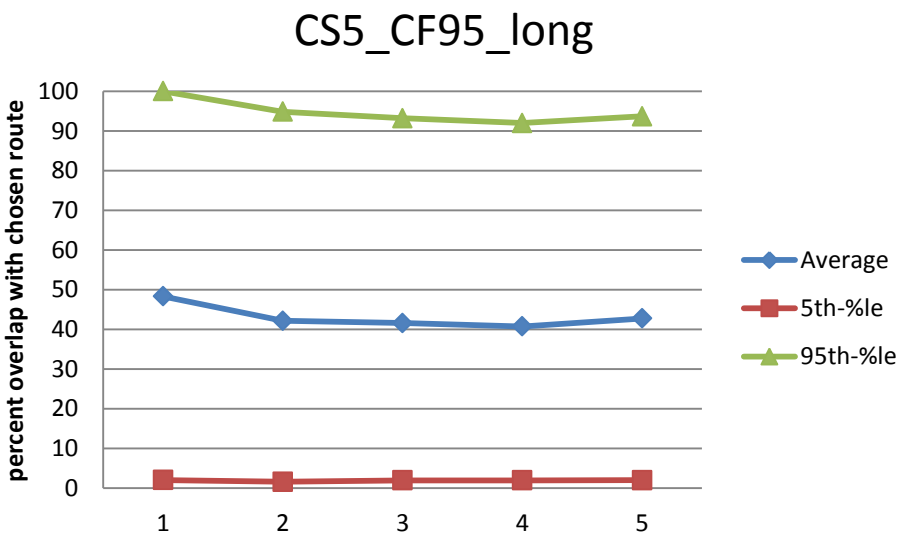
alternatives for this study only looks at the first ‘n’ alternatives of the different choice set generation patterns, despite there can be various combinations of the all the alternatives in the original choice set.

APPENDIX
PERCENT OVERLAP OF nth ROUTE WITH CHOSEN ROUTE






CS10_CF95_short
percent overlap with chosen route
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Average
5th-%le
95th-%le


CS5_CF95_long
percent overlap with chosen route
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
Average
5th-%le
95th-%le


CS5_CF90_short
percent overlap with chosen route
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
Average
5th-%le
95th-%le

LIST OF REFERENCES

1. Bekhor, S., M.E. Ben-Akiva, and S. Ramming. Evaluation of Choice Set Generation Algorithms for Route Choice Models. Annals of Operation Research, Vol. 144(1), 2006, pp. 235-247.

2. Bekhor, S., and C.G. Prato. Methodological Transferability in Route Choice modeling. Transportation Research Part B, Vol. 43(4), 2009, pp. 422-437.

3. Ben-Akiva, M.E., M.J. Bergman, A.J. Daly, and R. Ramaswamy. Modeling Inter-Urban Route Choice Behavior. In: Volmuller, J. Hamerslag, R. (Eds.), Proceedings of the 9th International Symposium on Transportation and Traffic Theory. VNU Science Press, Utrecht, The Netherlands, 1984, pp. 299-330.

4. Ben-Akiva, M.E., and M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. In Handbook of Transportation Science, pages 5-33. Springer US, 1999.

5. Bierlaire, M., and E. Frejinger. Route Choice Modeling with Network-Free Data. Transportation Research Part C, Vol. 16, 2008, pp. 187-198.

6. Bliemer, M.C.J., and P.H.L.Bovy. Impact of Route Choice Set on Route Choice Probabilities. Transportation Research Board: Journal of Transportation Research Board, no.2076, 2008, pp.10-19.

7. Bovy, P.H.L., S. Bekhor, and C.G. Prato. The Factor of Revisited Path Size: Alternative Derivation. Transportation Research Record: Journal of the Transportation Research Board, no. 2076, 2008, pp. 132-140.

8. Dhakar, N. Route Choice Modeling using GPS data. PhD Dissertation submitted to the University of Florida, 2012.

9. Flotterod, G. Bierlaire, M. Metropolis-Hastings sampling of paths, Transportation Research Part B, Vol. 48, pp: 53-66.

10. Frejinger, E., M. Bierlaire, and M. Ben-Akiva. Sampling of Alternatives for Route Choice Modeling. Transportation Research Part B, Vol. 43, 2009, pp. 984-994.

11. Frejinger, E., and M. Bierlaire. Capturing Correlation with Subnetworks in Route Choice Models. Transportation Research Part B, Vol. 41, 2007, pp. 363-378.

12. Hood, J. Sall, E. Charlton, B. A GPS based route choice model for San Fransisco, California, The International Journal of Transportation Research, 2011, Vol. 3, pp: 63-75.

13. Papinski, D., and D.M. Scott. Modeling Home-to-Work Route Choice Decisions Using GPS Data: A Comparison of Two Approaches for Generating Choice Sets. Presented at the Transport Research Board Annual Meeting 2011b, Washington D.C.

14. Prato, C.G., and S. Bekhor. Applying Branch-and-Bound Technique to Route Choice Set Generation. Transportation Research Record: Journal of the Transportation Research Board, no. 1985, 2006, pp. 19-28.

15. Prato, C.G., and S. Bekhor. Modeling Route Choice Behavior: How Relevant Is the Composition of Choice Set? Transportation Research Record: Journal of the Transportation Research Board, no. 2003, 2007, pp. 64-73.

16. Prato, C.G. Route Choice Modeling: Past, Present and Future Research Directions. Journal of Choice Modeling, Vol 2 (1), 2009, pp.65-100.

17. Quattrone, A., and A. Vietetta. Random and Fuzzy Utility models for Road Route Choice. Transportation Research Part E, Vol. 47, 2011, pp. 1126-1139.

18. Schussler, N. Accounting for Similarities in Destination Choice Modeling. PhD Dissertation submitted to ETH ZURICH, 2009.

19. Schussler, N., and K.W. Axhausen. Accounting for Route Overlap in Urban and Sub-urban Route Choice Decisions Derived from GPS Observations. Paper Presented at Proceedings of 12th International Conference on Travel Behavior Research, Jaipur, India, December 2009.

20. Smits, E.S., Bliemer, M.C.J., Pel, A., Arem van, B. On Route Choice Models with Closed Form Probability Expressions, Transportation Research Board Annual Meeting, 2014.

21. Spissu, E., I. Meloni, and B. Snjust. Behavioral Analysis of Choice of Daily Route with Data from Global Positioning System. Transportation Research Record: Journal of the Transportation Research Board, No. 2230, 2011, pp. 96-103.

22. Vreeswijk, J., Thomas, T., Berkum van, E., Arem van, B. Perception bias in route choice, Transportation Research Board Annual Meeting, 2014.

BIOGRAPHICAL SKETCH

Avinash Geda, originally from Andhra Pradesh, India, enrolled at the University of Florida in August of 2012. He joined the Transportation Graduate program at UF following completion of his B.Tech degree in Civil Engineering from the National Institute of Technology-Warangal, India. As a graduate assistant at the McTrans Center at the University of Florida, he worked on the analyzing and testing of the software level implementation of Highway Capacity Manual (HCM), Highway Capacity Software (HCS) under the supervision of Mr. William Sampson. He also served as teaching assistant for Mr. Sampson during his graduate studies at UF. He was working closely with Dr. Sivaramakrishnan Srinivasan in the area of route choice modeling using discrete choice Methodology.