

5-1-2015

## Mode Choice Modeling Using Personalized Travel Time and Cost Data

CUTR

Follow this and additional works at: [https://digitalcommons.usf.edu/cutr\\_nctr](https://digitalcommons.usf.edu/cutr_nctr)

---

### Recommended Citation

"Mode Choice Modeling Using Personalized Travel Time and Cost Data," National Center for Transit Research (NCTR) Report No. CUTR-NCTR-RR-2015-07, Center for Urban Transportation Research, University of South Florida, 2015.

DOI: <https://doi.org/10.5038/CUTR-NCTR-RR-2015-07>

Available at: [https://scholarcommons.usf.edu/cutr\\_nctr/28](https://scholarcommons.usf.edu/cutr_nctr/28)

This Technical Report is brought to you for free and open access by the National Center for Transit Research (NCTR) Archive (2000-2020) at Digital Commons @ University of South Florida. It has been accepted for inclusion in Research Reports by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact [digitalcommons@usf.edu](mailto:digitalcommons@usf.edu).



# **Mode Choice Modeling Using Personalized Travel Time and Cost Data**

**Final Report**

**May 2015**

**PROJECT NO.**  
2117-9060-02-C

**PREPARED FOR**  
National Center for Transit Research (NCTR)

## **Disclaimer**

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program and the Florida Department of Transportation, in the interest of information exchange. The U.S. Government and the Florida Department of Transportation assume no liability for the contents or use thereof.

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

# Metric Conversion

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
<b>LENGTH</b>				
<b>in</b>	inches	25.4	millimeters	mm
<b>ft.</b>	feet	0.305	meters	m
<b>yd.</b>	yards	0.914	meters	m
<b>mi</b>	miles	1.61	kilometers	km
<b>VOLUME</b>				
<b>fl. oz.</b>	fluid ounces	29.57	milliliters	mL
<b>gal</b>	gallons	3.785	liters	L
<b>ft<sup>3</sup></b>	cubic feet	0.028	cubic meters	m <sup>3</sup>
<b>yd<sup>3</sup></b>	cubic yards	0.765	cubic meters	m <sup>3</sup>
NOTE: volumes greater than 1000 L shall be shown in m <sup>3</sup>				
<b>MASS</b>				
<b>oz.</b>	ounces	28.35	grams	g
<b>lb.</b>	pounds	0.454	kilograms	kg
<b>T</b>	Short tons (2000 lb.)	0.907	megagrams (or "metric ton")	Mg (or "t")
<b>TEMPERATURE (exact degrees)</b>				
<b>°F</b>	Fahrenheit	$\frac{5}{9}(F-32)$ or $(F-32)/1.8$	Celsius	°C

# Technical Report Documentation

1. Report No. 2117-9060-02-C		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Mode Choice Modeling Using Personalized Travel Time and Cost Data			5. Report Date May 2015		
6. Performing Organization Code					
7. Author(s) Mahmoud Javanmardi Ph.D. Candidate Department of Civil and Materials Engineering University of Illinois at Chicago 842 W. Taylor Street Chicago, IL 60607-7023			8. Performing Organization Report No.		
9. Performing Organization Name and Address National Center for Transit Research Center for Urban Transportation Research (CUTR) University of South Florida 4202 East Fowler Avenue, CUT100 Tampa, FL 33620-5375			10. Work Unit No. (TRAIS)		
11. Contract or Grant No.					
12. Sponsoring Agency Name and Address  Research and Innovative Technology Administration U.S. Department of Transportation Mail Code RDT-30, 1200 New Jersey Ave SE, Room E33 Washington, DC 20590-0001			13. Type of Report and Period Covered		
14. Sponsoring Agency Code					
15. Supplementary Notes					
16. Abstract <p>Microsimulation based frameworks have become very popular in many research areas including travel demand modeling where activity-based models have been in the center of attention for the past decade. Advanced activity-based models synthesize the entire population of the study region and simulate their activities in a way that they can keep track of agents' resources as well as their spatial location. However, the models that are built for these frameworks do not take into account this information sufficiently. This paper describes the importance of accurate information by analyzing a travel survey and proposes a methodology to generate the actual alternatives that individuals had when making their trips using Google Maps API and RTA's Goroo TripPlanner. The study reveals how some alternative modes' travel information are limited/unavailable in the traditional travel time skim matrix which must be taken in to account in our mode choice models. Also, statistics regarding available alternatives and the constraints people encounter when making a choice are presented and considered accordingly. In order to examine the importance of accurate alternative information for mode choice modeling, two mode choice models are then developed for comparison. One of the models is estimated from the accurate alternative information described in this paper and the other is estimated from the traditional travel time skim matrix obtained from the regional travel demand model. Once the superiority of the proposed method is approved through the comparison, a nested mode choice model encompassing all modes is estimated based on the developed choice set and model significance and goodness-of-fit measures are presented accordingly.</p>					
17. Key Words			18. Distribution Statement No restrictions		
19. Security Classification (of this report)  Unclassified		20. Security Classification (of this page)  Unclassified		21. No. of Pages 31	
22. Price					



**Urban Transportation Center at the University of Illinois at Chicago**

# **Mode Choice Modeling Using Personalized Travel Time and Cost Data**

**MAY 2015**

*This report was produced with funding from the National Center for Transit Research (NCTR), a US DOT-OST National University Transportation Center*

# **Mode Choice Modeling Using Personalized Travel Time and Cost Data**

## **Mahmoud Javanmardi**

Ph.D. Candidate  
Department of Civil and Materials Engineering  
University of Illinois at Chicago  
842 W. Taylor Street  
Chicago, IL 60607-7023  
Tel: (312) 996-3441  
Email: [mjavan2@uic.edu](mailto:mjavan2@uic.edu)

## **Mehran Fasihzaman Langerudi**

Ph.D. Candidate  
Department of Civil and Materials Engineering  
University of Illinois at Chicago  
842 W. Taylor Street  
Chicago, IL 60607-7023  
Tel: (312) 996-3441  
Email: [mfasih2@uic.edu](mailto:mfasih2@uic.edu)

## **Ramin Shabanpour Anbarani**

Ph.D. Student  
Department of Civil and Materials Engineering  
University of Illinois at Chicago  
842 W. Taylor Street  
Chicago, IL 60607-7023  
Tel: (312) 996-3441  
Email: [rshaba4@uic.edu](mailto:rshaba4@uic.edu)

## **Abolfazl (Kouros) Mohammadian, Ph.D.**

Professor  
Department of Civil and Materials Engineering  
University of Illinois at Chicago  
842 W. Taylor Street  
Chicago, IL 60607-7023  
Tel: (312) 996-9840  
Email: [kouros@uic.edu](mailto:kouros@uic.edu)

## Contents

Abstract.....	4
Introduction.....	5
Literature Review.....	6
Study Area and Data .....	7
Descriptive Study.....	9
Choice Set Formation Methodology.....	13
Personalized Travel Time and Cost .....	13
Resource Constraints .....	14
Model Analysis.....	14
MNL Models Comparison .....	14
Nested Logit Model.....	20
Conclusion .....	24
References.....	25



## **Abstract**

Microsimulation based frameworks have become very popular in many research areas including travel demand modeling where activity-based models have been in the center of attention for the past decade. Advanced activity-based models synthesize the entire population of the study region and simulate their activities in a way that they can keep track of agents' resources as well as their spatial location. However, the models that are built for these frameworks do not take into account this information sufficiently. This paper describes the importance of accurate information by analyzing a travel survey and proposes a methodology to generate the actual alternatives that individuals had when making their trips using Google Maps API and RTA's Goroo TripPlanner. The study reveals how some alternative modes' travel information are limited/unavailable in the traditional travel time skim matrix which must be taken in to account in our mode choice models. Also, statistics regarding available alternatives and the constraints people encounter when making a choice are presented and considered accordingly. In order to examine the importance of accurate alternative information for mode choice modeling, two mode choice models are then developed for comparison. One of the models is estimated from the accurate alternative information described in this paper and the other is estimated from the traditional travel time skim matrix obtained from the regional travel demand model. Once the superiority of the proposed method is approved through the comparison, a nested mode choice model encompassing all modes is estimated based on the developed choice set and model significance and goodness-of-fit measures are presented accordingly.

## Introduction

Traditionally transportation related models were designed to deal with aggregate information or specific market segments; however, with advancement in technology, much more information has become available to researchers as the demand for more policy sensitive models has increased. These advances have made microsimulation frameworks very popular in the past two decades (1-6).

Although Microsimulation frameworks could potentially be sensitive to many important factors, in which modelers and planners have always been interested, they also introduce many challenges and issues that if not taken into account, they could highly bias their outcomes. In the area of regional travel demand modeling, activity-based models have been the center of attention for the past decade and many researchers have proposed and developed different frameworks (1,7,8). These microsimulation frameworks try to replace trip-based four-step approaches, which were highly aggregate, by simulating individuals daily/weekly/monthly activities while conserving many institutional and resource constraints (1).

In this type of frameworks, census public data is utilized to synthesize the entire population of the study region along with their socio-demographic/economic attributes (8). For most of the currently developed activity-based models, this information is used during the activity generation and scheduling part and is not carried to the traffic assignment step of the program. However, more advanced frameworks take advantage of this information even during the traffic assignment. The state of the art frameworks have an integrated traffic assignment module that is run concurrently with the activity generation module. The agents have full access to the network information and also availability of resources at each time step so that they could update their decisions accordingly. This level of details is available to all of the sub models within the larger framework.

The traditional frameworks, however, were built mostly for a limited market segment and were not necessarily based on disaggregate and accurate information. They focus on the Traffic Analysis Zones (TAZ) to monitor the aggregate behavior of a city such as land use, commuting patterns, and location of citizens of different categories (9, 10). Therefore, enforcing the integrity of resources/constraints was not feasible in such simulation environments. This aggregate treatment constrains the mode choice process for each individual. Availability of cars for each household member (i.e. other members have not taken the car already), relation between modes of travel in tour segments, and availability of transit modes are among the factors which are not investigated comprehensively in the literature. On the other hand,

activity-based models are designed to investigate activities and travels of individuals for a longer duration (usually more than 24 hours). All individuals could be tracked during this period and institutional and resource constraints could be enforced. Therefore, for instance, a mode choice model in an activity-based framework could take advantage of vehicle availability knowledge at a time, while a mode choice model used in a four-step framework, could just rely on the number of vehicles a household owns. Moreover, advanced technology has enabled researchers to collect accurate information in terms of household information, their travel behavior and network conditions that could affect individuals' travel and its attributes. The technology, powerful processors, multithreading and distributed computing techniques have also made complicated microsimulation tasks much faster and more feasible. They could take advantage of accurate network and public transit information and take them into account during the simulation. For example, in such environments it is feasible to use the location of individuals, the transit system availability and proximity at the time of interest to make an alert decision when choosing a mode for travel.

With this introduction, the purpose of this study is to describe the advantages and importance of accurate data and resource constraints in the models within mode choice of activity-based frameworks. In the following sections, first, a literature review is conducted. Then, the study area and the data are described. Following that a descriptive analysis of mode choice is presented. Next, the choice set formation methodology including the personalized travel time/cost and resource constraints are described. The study is then followed by the result section that provides models comparison estimated by the new and the traditional choice set formation approaches. Finally, a nested mode choice model for the Chicago metropolitan is estimated using the new choice set formation approach and the goodness-of-fit measures for the mode choice problem is explored.

## **Literature Review**

Travel mode choice behavior in the context of disaggregate models has been extensively investigated over the last decades. Disaggregate mode choice modeling, which is based on individual analysis, entails various unobservable traveler's attitude and perceptions, which affect his/her utilities of activity and mode, and in turn, his/her activity and mode decisions. The majority of studies in this stream have been focused on econometric theory of random utility maximization. It assumes that an individual's choice is determined by the indirect utilities of each alternative and the individual can choose the one that maximizes his/her utility level (11). For example, Koppelman (12) used a multinomial logit model to predict mode share changes in response to a range of transit service improvements. Later on, a great deal of advances has been done following this line of research. Advanced modelling frameworks including the mixed multinomial logit model and the generalized extreme value (GEV) models (13) have been adopted

to investigate travel model choice behavior. Mixed logit models have been applied to model mode choice and to incorporate both observed and unobserved heterogeneities (14). The assumption of independence from irrelevant alternatives (IIA) has been addressed by a series of studies on nested logit and generalized nested logit models (e.g. 15, 16).

It can be inferred clearly from the literature that individual and household socio-demographics play an important role in travel mode choice decisions. Specifically, gender, income, car ownership, employment status effect travel mode decisions (17, 18). In addition, (19-21) showed the prominent influence of residential location, neighborhood type, and urban form in determining the favored travel mode for commute.

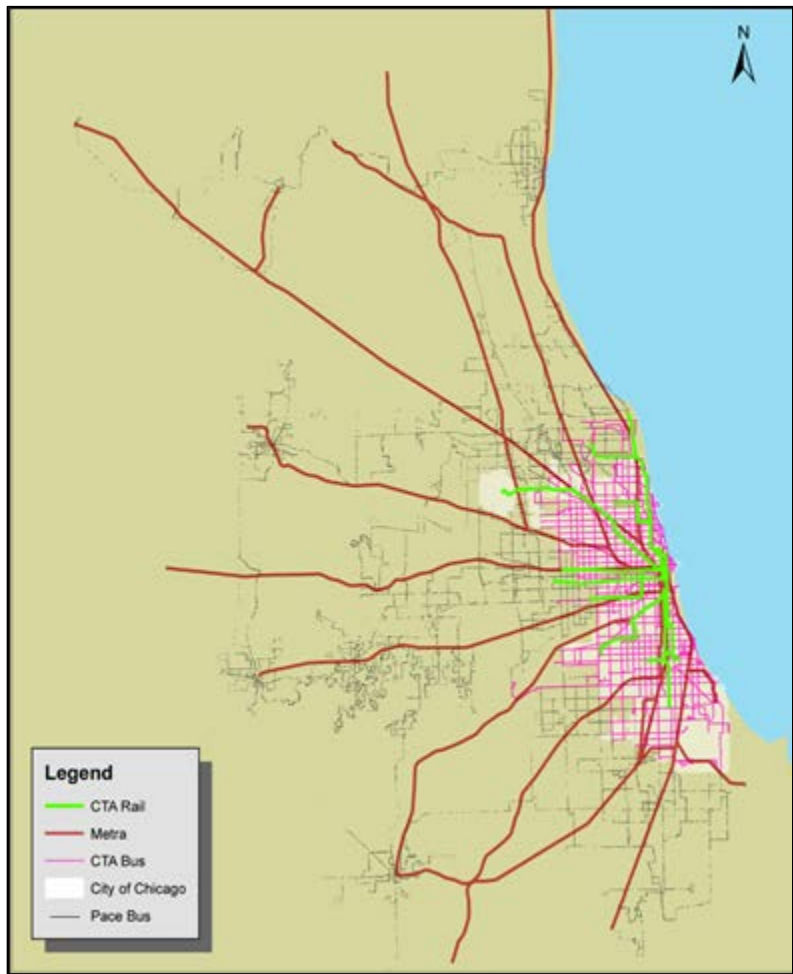
Willingness to pay, which is the amount of money that travelers are inclined to pay to reduce their travel time by unit time, is another factor which has been extensively investigated (17, 18). In more recent research studies (22, 23), reliability of travel time is also incorporated within the framework to compute the value of travel time.

Another important dimension of mode choice process is definition of choice set for each individual. Some of the recent studies tried to consider choice set heterogeneity by considering varying choice sets. For example, (24) defined a theoretical framework using a two-stage process to account for variable choice sets. The first stage involves the choice set generation and the second stage is a conditional choice model, given the choice set. However, few studies (25, 26) have applied this framework due to the large number of choice sets and difficulty in estimating probabilities of considering such choice sets empirically (27).

## **Study Area and Data**

Public transportation in the Metropolitan of Chicago is made available by three major agencies including CTA (with both Bus and Subway fleet) which serves the city of Chicago residents, Pace system which mainly serves intra and inter Suburban trips and finally Metra that provides service to intra and inter suburban trips. Figure 1 depicts the area where each agency serves.

The data used in this study comes from the CMAP Travel Tracker Survey, which is a comprehensive travel, and activity survey conducted by Chicago Metropolitan Agency for Planning (CMAP). The survey was designed for use in a regional travel demand model and was conducted over a 14 months period beginning in January 2007. More than 14,000 households participated in the survey and recorded their activities and travels for one or two days. At the end, more than 218,005 activities were recorded. The survey dataset was analyzed, invalid records were eliminated, and trips and home-based tours were formed and linked to socio-demographic information of individuals. Also, mode choices were defined as



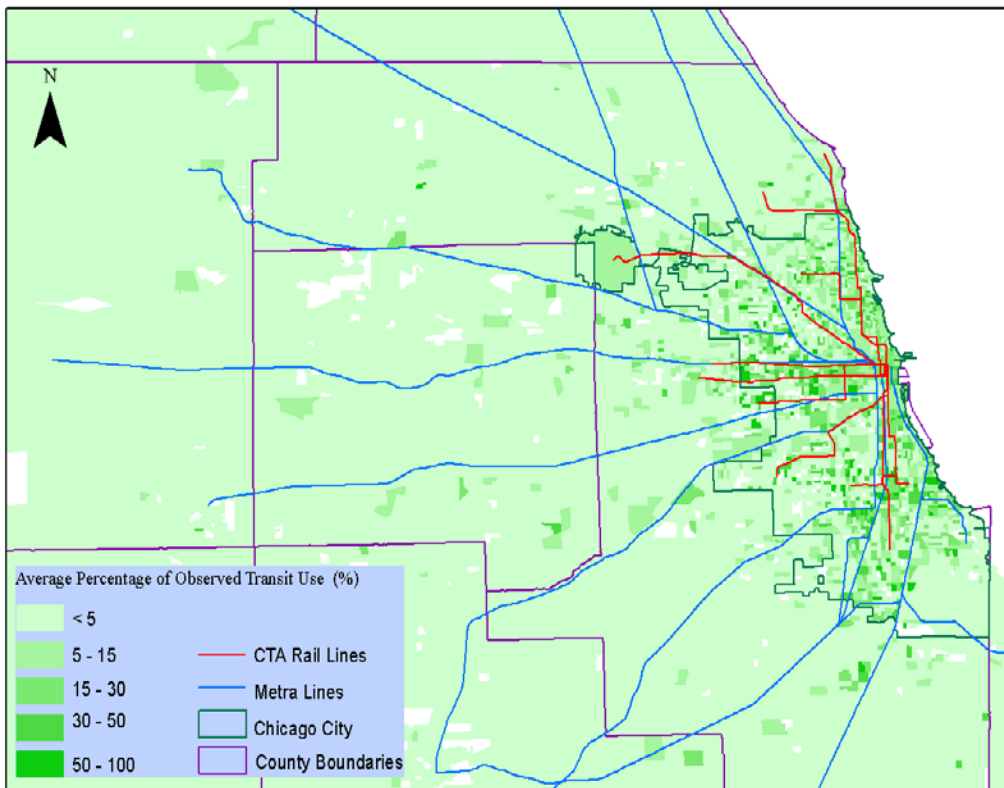
**Figure 1.** Chicago Public Transportation Network

the dominant mode of the trip. If the trip had multiples legs, then the dominant mode was selected from the leg with the longest travel time.

It is noteworthy that home-based trips with their origin at-home were selected for these mode choice models since trips that are not home-based are highly associated with their respective home-based origin trip within a tour (A tour is a combination of trips, which starts and ends at home). This means when an individual starts a tour which is comprised of several trips, the mode of the first trip that originates from

home highly affects the mode of the later trips. Therefore, the models developed are mainly focused on the trips that start from home.

Figure 2 represents the ridership percentage in the greater Chicago area block groups based on the CMAP data. As expected, most of the suburban areas have ridership as low as 5%; however, the city area has ridership as much as 50% or higher in certain block groups. One of the resource constraint parameters in mode choice modeling could be the unavailability of transit which could be inferred from the low ridership in certain areas. In case, transit is not offered for a number of trips in the data, the alternative



**Figure 2.** Transit Ridership in the Greater Chicago Area

modes must be restrained accordingly. This issue, which affects the choice set of an individual, is investigated in the following section.

## Descriptive Study

Before describing the choice set formation methodology, a comprehensive descriptive analysis was conducted to investigate the important factors in choice set formation. One of the analyses that were conducted was the distribution of modes, when previous mode of a trip was auto-drive. In Table 1, trips are categorized into two groups, “return-home” and “others”. The table shows that almost all return-home trips and more than 96% of other trips use the same mode as their previous mode (i.e. auto-drive). It is

highly possible that if the sub-tours are formed, the rate of auto-drive mode will be higher than 96.5%. This represents a constraint that individuals encounter when choosing a mode and confirms the expectation that when a transit or auto-drive tour starts, the mode is typically fixed in the tour. This conclusion highly affects the structure of a mode choice model in terms of adopting tour-based approach; a tour-based approach that relies on the mode of the initial home-based trip. However, in other cases like work-based tours, the mode of the non-home-based trips could be different from the initial trip mode. Since these trips are less frequent in the data, the focus of this study is on home-based mode choice analysis.

**Table 1**  
Distribution of mode when previous trip mode is auto drive

Mode	return home		others	
	count	Percentage	count	Percentage
Auto / Van / Truck Driver	33,885	99.36%	30,061	96.54%
Walk	42	0.12%	558	1.79%
Auto / Van / Truck Passenger	144	0.42%	344	1.10%
Metra Train	4	0.01%	61	0.20%
CTA Train	0	0.00%	33	0.11%
School Bus	0	0.00%	23	0.07%
Taxi	2	0.01%	18	0.06%
OTHER	0	0.00%	9	0.03%
More than one transit provider	3	0.01%	9	0.03%
Bike	22	0.06%	7	0.02%
Private shuttle bus	0	0.00%	6	0.02%
CTA Bus	1	0.00%	4	0.01%
Pace Bus	0	0.00%	4	0.01%
Dial a ride/Paratransit	1	0.00%	1	0.00%

Another type of constraint that should be taken into account is vehicle availability. In traditional mode choice models, it has been represented by household number of vehicles. Although this question is typically asked in surveys, including CMAP survey, data mining and tracking households' members' activities revealed many cases that the only household vehicle was being used by another household member when the trip started. Therefore, auto-drive should be eliminated from the other household member's choice set. Table 2 represents the mode distribution for individuals with one household vehicle ownership, which was being used by another household member at the time of the corresponding trip. This information should be incorporated in the model estimation procedure.

**Table 2**  
Selected mode when another household member is using the only vehicle (HHVEH=1)

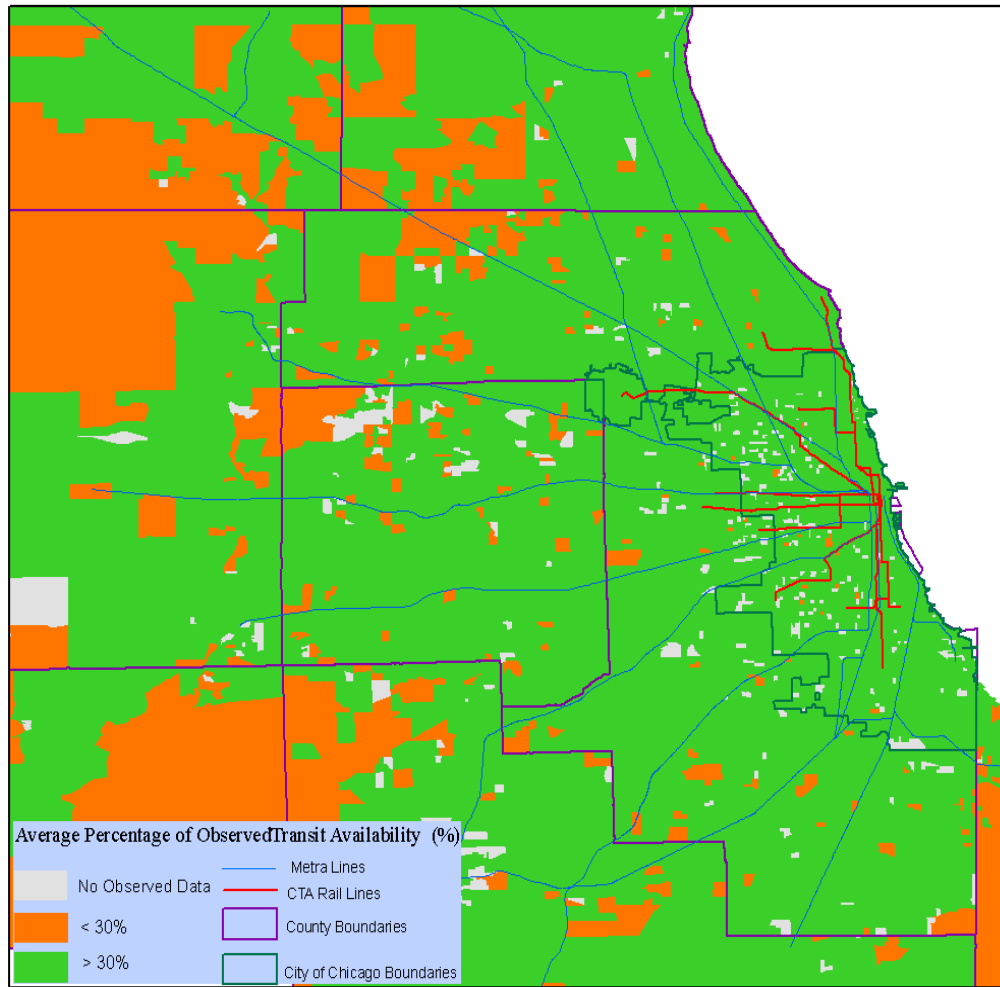
<b>Mode</b>	<b>Count</b>	<b>Total</b>	<b>Percentage</b>
Walk	746	4,323	17%
Bike	64	407	16%
Auto / Van / Truck Driver	444	24,798	2%
Auto / Van / Truck Passenger	5,840	7251	81%
CTA Bus	127	688	18%
CTA Train	77	640	12%
Pace Bus	12	71	17%
Metra Train	72	529	14%
Private shuttle bus	7	63	11%
Dial a ride/Paratransit	12	19	63%
School Bus	163	381	43%
Taxi	25	150	17%
Local Transit (NIRPC region)	3	16	19%
More than one transit provider	31	270	11%
OTHER (SPECIFY)	3	29	10%

The other relevant choice set constraint was transit availability. Figure 3 shows the block groups where transit was barely available for the observed trips. The red block groups show that less than 30% of the observed trips had at least one transit alternative at the time when the trips occurred. In other words, the red block groups represent the transit deserts in the Chicago area and could be the target of future transit projects. It was observed that for many observations there are no transit stops or stations close to the origin or destination. Moreover, some transit agencies do not operate around the time and the day when the trip started. The conclusion for both cases is that the transit alternative was not available and it should be eliminated from the estimation choice set. Table 3 shows that almost 80% of the trips did not have any CTA alternatives, approximately 75% of the trips did not have any Pace alternatives and 75% did not have any Metra alternatives, while 47% of trips did not have any transit alternatives.

**Table 3**  
Transit Alternative Availability for Observed Trips

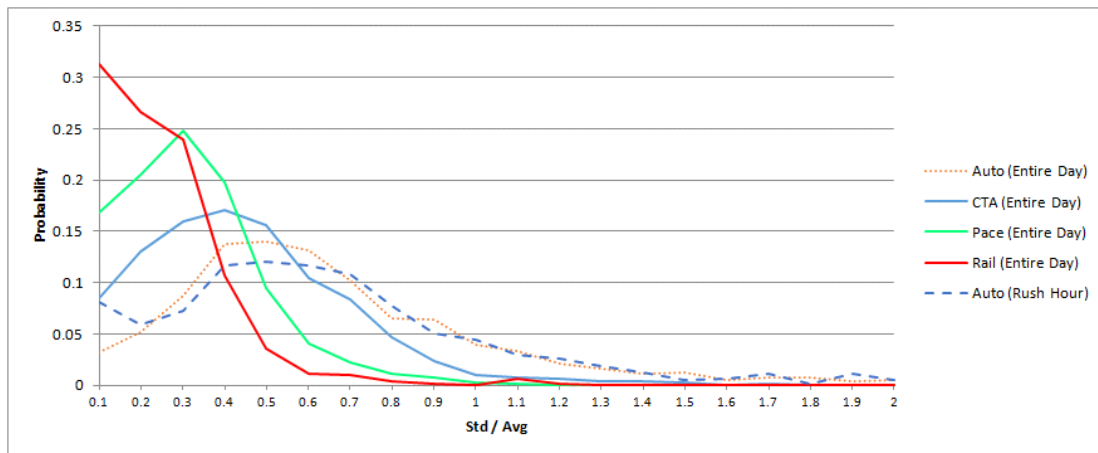
	<b>CTA (Bus and Subway)</b>	<b>Pace</b>	<b>Heavy Rail</b>	<b>Any Transit</b>
Not-Available	124,548 (80%)	115,880 (75%)	116,128 (75%)	73,724 (47%)
Available	30,975 (20%)	39,643 (25%)	39,395 (25%)	81,799 (53%)
Sum:	155,523	155,523	155,523	155,523





**Figure 3.** Transit Availability for the Observed Trips

Figure 4 displays the travel time reliability of different modes represented as the distribution of  $\frac{STD}{Average}$  for zone-to-zone travel times (for zones with more than 5 observations) for different modes. As it can be seen, the travel times between zones highly varies and follows a distribution. For example, for CTA, the



**Figure 4.** Transit Distribution of  $\left(\frac{STD}{Average}\right)$  of zone to zone travel times for different modes of travel

average error is around 40% while for auto (both for entire day and rush hour) the error is higher. The distributions represented in this graph describe why average travel time skim assumption from a regional travel demand model might not be a reliable approach for alternative mode travel times. The rush-hour distribution shows that travel time reliability decreases for the observed rush hour trips.

## Choice Set Formation Methodology

Based on the descriptive analysis presented earlier, this paper proposes a choice set formation methodology that considers two important factors of personalized travel time/cost as well as resource constraints in mode choice modeling.

## Personalized Travel Time and Cost

Traditionally when developing a mode choice model, TAZ level travel time skim information for the alternative modes is generated from a regional travel demand model, and average travel times are used to represent zonal travel times. Some modelers have tried to improve this method by using more disaggregate zones, for instances Micro Analysis Zones (MAZs) (28). However, the use of personalized travel time and cost data can better monitor the individual's travel behavior for the observed trip. For this purpose, a software application was developed to record the observed trips' detail information from *Google Maps API* and also *RTA's Goroo TripPlanner* website. The information included exact point-to-point travel times, available alternative modes, disaggregate access and egress distances, the nearest available transit stations/stops, transfer information, among others.

In order to find transit travel times, Google relies on the information that each transit agency provides in GTFS format. Each transit agency uploads its fleet schedule, route, and fare structure while Google takes advantage of this data to find the best route. The website also offers travelers to choose one of the three transit alternatives: Bus, Subway, and Rail. For this study, all three options were selected and the suggested route information was recorded for the alternative modes of the observed trips. To be consistent with observed mode definitions, the mode of the alternatives was selected based on the longest trip leg.

Furthermore, since all trips legs were known (when more than one transit mode is used), the fare associated with each trip could be exactly calculated, even though each agency has a complicated fare system in terms of transfers, origin and destination station, and age of the travelers. The software was also designed to query the travel times of trips at the same day of week and exact time of the day that the original trips were made. For driving mode, this process was repeated 10 times, and the average was recorded. It is worth mentioning that although travel times in 2014 are not necessarily the same as they

were in 2007-2008, when the original trips were made, this approach seems to be an ideal travel time estimation method.

## **Resource Constraints**

As emphasized in the descriptive study section, resource availability in terms of either vehicle or transit was strictly imposed on the choice set formation. A data mining approach was employed to set the vehicle availability for the observed trip. The choice of auto-drive was eliminated for the observed trips for which no vehicle was found available. Furthermore, availability of all transit alternatives was experimented through the Google Map API software developed. All the unavailable transit alternatives were eliminated from the choice set of individuals. These resource constraints can eventually be considered in the activity-based travel demand models due to the disaggregate information that is available in these models.

## **Model Analysis**

In this section, two multinomial logit models are estimated based on the new and the traditional choice set formation approaches and are compared in terms of their significance and goodness-of-fit measures. Then a nested mode choice model for the Chicago metropolitan, which is estimated using the new choice set formation approach, is elaborated.

## **MNL Models Comparison**

As mentioned earlier, the hypothesis was that the new choice set formation would provide a more reliable mode choice model for application compared to the mode choice models developed from traditional choice set formation approach that uses TAZ level travel time skim information without taking into account the resource constraints.

For testing the hypothesis, this section discusses the comparison of two mode choice models estimated from both approaches. In order to compare these two models in terms of both estimation and prediction potential, they are developed based on the same modeling structure. The training sample is an 80% filtered random sample from the CMAP Travel Tracker Survey data, and the remaining 20% is used for the testing sample. Since the TAZ-level travel time skim information provided by CMAP does not specify the type of transit modes, the models are developed using a transit general mode. Even though the modeling structure could affect the prediction performance of the models, being aware of the drawbacks of a Multinomial Logit structure, most importantly the IIA assumption, both models are estimated with Multinomial Logit. However, this assumption does not degenerate the results considering that it is imposed on both models.

The mode choice models developed for this study are estimated on home-based trip observations with home as the trip origin. This eliminates the problem of associated/correlated trips in tours. Explanatory variables that are used to model the choice situation are defined in Table 4 and result of the estimated MNL models are presented in Table 5. Five different modes are considered for the models including walk, bike, auto-drive, auto-passenger, and transit. A number of key variables were examined to specify the corresponding utility functions. Variables used to specify the utility function of walking mode include age and travel time. For biking, due to the infrequency of data, travel time is the only factor to form the utility function. Auto mode including drive and passenger could be dependent on a number of socio-demographic and economic attributes including gender, income, number of household members, number of vehicles as well as trip characteristics such as origin, destination, activity type, time of day, gas price and day of week. The model shows that female individuals drive more frequently than male ones. Furthermore, income is a significant factor in increasing the drive mode. Driving to the CBD area during the rush hours is a less frequent drive. Activity type is another factor that affects the mode preference; shopping activities are more executed with the auto mode. On the other hand, trips for work related activities are less made with the passenger mode. Individuals are more inclined to drive on the weekends; less to use passenger mode. Passenger mode is more probable with higher number of household members and number of vehicles which is intuitive based on vehicle availability for driving as well as the passenger mode. For Transit, Number of transfers, transit access distance, travel time and fare are the negative factors for the ridership while having no vehicles is an implicit factor in increasing the likelihood for transit due to the elimination of auto-drive mode for those individuals.

Even though a wider range of variables could be used in the models, the ones used are selected equivalently in both of the models to keep the inference unbiased. For the goodness of fit of the models, the likelihood ratio index is calculated based on the statistic suggested by McFadden (1974). This index although different in concept is analogous to the R-square in linear regression models. The result shows that the likelihood ratio for the model associated with the new choice set formation is higher. This represents the significance of the choice set formation approach which is constructed by considering accurate alternative mode information.

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)} \quad (1)$$

In addition to the likelihood ratio, which is an unintuitive method to compare likelihood based models, in this study, the models are also tested for their prediction potential. Table 6 shows the predictability measures of the simulated models over the 20% test data and the observed data. Mode share is calculated

for the observed data and the associated 95% confidence intervals are obtained based on approximate binomial sampling variance estimation through the formulation below:

$$\hat{p} \mp t_{\alpha/2} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\hat{p}(1 - \hat{p})}{n - 1}} \approx \hat{p} \mp t_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n - 1}} \quad (2)$$

In which n represents the number of observations for the test sample and N is the total number of trips where the survey is extracted from. N by far exceeds the number n; that is why the term is typically omitted in practical applications. As the predictability result shows, none of these models can predict the mode shares with 95% significance level; however, the model estimated from the new choice set formation results in mode shares closer to the range of the observed mode shares. Most importantly, it shows a better performance in correctly predicting the transit choices as is observed in the test data. Also, in terms of auto and passenger predictions, it seems superior to its counterpart.

**Table 4**

Description of explanatory variables used in the models

<b>Variable</b>	<b>Description</b>	<b>Average</b>	<b>Std. Dev.</b>
TravelTime	Travel time in hours	0.39	0.43
TravelCost	Travel Cost in Dollars	1.09	1.29
nVeh	Number of vehicles	2.02	0.97
AgeOver20 (Y N)	If age is over 20	0.83	0.39
AgeOver65 (Y N)	If age is over 65	0.16	0.37
TOTTR1	Total number of tours in the day	1.66	1.02
nBusStopOrigin	Number of CTA bus stops in the origin zone	1.80	6.17
nBusStopDestination	Number of CTA bus stops in the destination zone	9.08	33.87
nMetraStopOrigin	Number of Metra stations in the origin zone	0.05	0.24
nMetraStopDestination	Number of Metra stations in the destination zone	0.11	0.46
DistanceWalk	If the walking distance to destination is over 0.4 km (0.25 miles)	0.995	0.07
TourPurposeWork (Y N)	If the trip purpose is work	0.29	0.45
TourPurposeShop (Y N)	If the trip purpose is shopping	0.13	0.34
Female (Y N)	If the individual is female	0.53	0.50
PaceAccesskm	Access distance from origin when Pace is main mode (km)	0.49	1.50
PaceEgresskm	Egress distance to destination when Pace is main mode(km)	0.28	0.97
ctaAccesskm	Access distance from origin when CTA is main mode (km)	0.14	0.34
ctaEgresskm	Egress distance to destination when CTA is main mode(km)	0.10	0.28
railAccesskm	Access distance from origin when Metra is main mode (km)	0.97	2.88
railEgresskm	Egress distance to destination when Metra is main mode(km)	0.67	2.17
Income	Household Income	79828	35124
nCTATransfer	Number of transfers when CTA is main mode	0.09	0.36
nPaceTransfer	Number of transfers when pace is main mode	0.27	0.69
nFastRailTransfer	Number of transfers when Metra fast access is main mode	0.39	0.82
nSlowRailTransfer	Number of transfers when Metra slow access is main mode	0.40	0.83
DestCBD_RushHr (Y N)	If the destination is CBD during rush hour	0.04	0.19
Weekend (Y N)	If the trip occurred in the weekend	0.12	0.32

**Table 5**  
**MNL Models Estimated on google Map and Travel time Skim Data**

	Home-based Mode Choice from Google Travel Time (Trip Origin is home)									
	<i>Walk</i>	<i>Bike</i>	<i>Drive</i>	<i>Passenger</i>	<i>Transit</i>	<i>Walk</i>	<i>Bike</i>	<i>Drive</i>	<i>Passenger</i>	<i>Transit</i>
Constant	2.77 (52.8)		2.19 (36.7)	2.19 (36.7)	2.93 (38.8)	3.67 (66.7)		1.97 (33.6)	1.97 (33.6)	2.48 (33.6)
TravelTime	-1.04 (-37.2)	-1.04 (-37.2)	-1.04 (-37.2)	-1.04 (-37.2)	-1.04 (-37.2)	-3.28 (-51.6)	-3.28 (-51.6)	-3.28 (-51.6)	-3.28 (-51.6)	-3.28 (-51.6)
TravelCost			-0.13 (-13.1)	-0.13 (-13.1)	-0.13 (-13.1)			-0.11 (-12)	-0.11 (-12)	-0.11 (-12)
nVeh			1.40 (59.85)	0.66 (27.3)				1.46 (60.1)	0.71 (28.5)	
Female (Y N)			0.33 (10.42)					0.33 (10.1)		
Income (10 <sup>-5</sup> )					-0.77 (-11.2)					-0.95 (-13.45)
DestCBD_RushHr (Y N)			-1.72 (-14.9)	-0.73 (-7.6)				-1.59 (-13.5)	-0.85 (-8.9)	
TourPurposeShop (Y N)			0.82 (14.25)	0.2 (2.35)				0.94 (15.7)	0.31 (4.98)	
TourPurposeWork (Y N)				-1.8 (-48.65)	0.45 (9.1)				-1.8 (-48.7)	0.51 (10.1)
Weekend (Y N)			0.09 (1.65)	-0.18 (-3.5)				0.15 (2.75)	-0.13 (-2.3)	
AccessDistance					-0.39 (-8.5)					-0.11 (-3.68)
EgressDistance					-0.23 (-7)					-0.08 (-2.5)
AgeOver65 (Y N)	-1 (-16)					-1.12 (-17.11)				
nTransfers					-0.17 (-6.62)					
<b>Log-likelihood at zero</b>	-56536.45					-56536.45				
<b>Log-likelihood at convergence</b>	-21314.24					-24876.4				
<b>Likelihood Ratio Index</b>	62.3%					56%				
<b>Sample size</b>	40497					40497				

**Table 6**  
**Comparison of Google-Based and Travel time Skim MNL Model Prediction Results**

Mode	Observed Data			Proposed-Model			Travel Time Skim-Model		
	Mode Share	Variance	CI (95%) For Real Mode Share	Simulated Mode Share	Correct Predictability (Matches the Observed model)	Accept	Simulated Mode Share	Correct Predictability (Matches the Observed model)	Accept
<i>Walk</i>	3.880%	0.0715%	(3.809%, 3.951%)	3.913%	28.50%	Y	3.892%	21.55%	Y
<i>Bike</i>	0.660%	0.0300%	(0.630%, 0.690%)	0.795%	1.92%	N	0.647%	3.30%	Y
<i>Auto</i>	68.621%	0.1718%	(68.450%, 68.793%)	67.250%	81.08%	N	66.921%	76.55%	N
<i>Passenger</i>	22.400%	0.1543%	(22.246%, 22.554%)	24.090%	59.11%	N	24.86%	58.8 %	N
<i>Transit</i>	4.438%	0.0762%	(4.362%, 4.515%)	3.956%	27.82%	N	3.680%	22.54.%	N

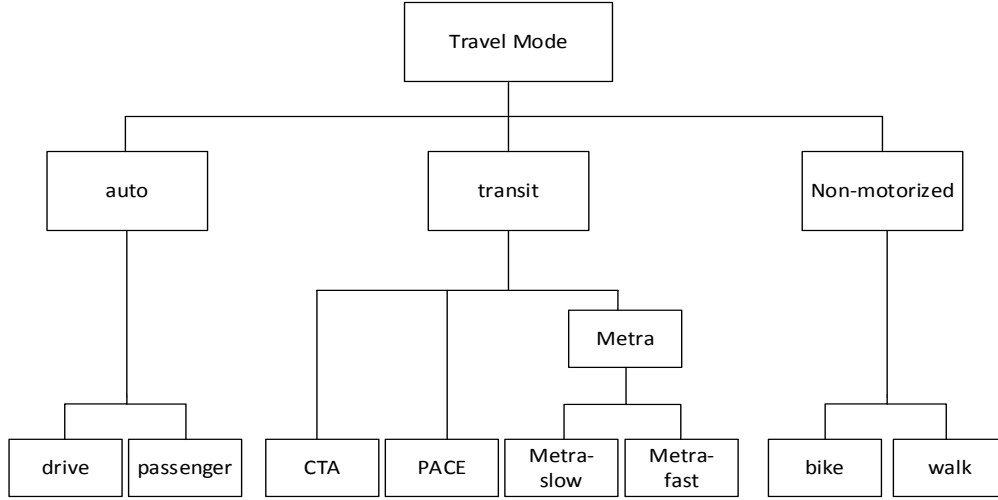


## Nested Logit Model

Following the comparison, this section presents a nested logit model developed from the new choice set formation. The Nested logit model, which is an extension of MNL, was introduced by (29) to capture the potential correlation among alternatives. Although IIA is maintained within each nest, it is relaxed to Independence of irrelevant nests (IIN) for alternatives in different nests. This allows variance of errors to differ across the nests, and yet the choice probabilities have a closed-form formula.

The transit modes used for this model cover all the transit agency providers in the greater Chicago region, which makes the use of a nested logit model more interpretable. To develop a mode choice model, first, it was required to select the alternatives that were going to be used in the model. As it was mentioned before, each transit agency targets specific type of trips and spatial location in the region, their fare structure, quality of ride, distance between stations/stops, and many other factors are consistent within each agency. Therefore, it was decided to make each agency as a separate alternative to travelers in addition to walk, bike, auto-drive, and auto-passenger modes. Since the access distance to Metra stations significantly differs from place to place in the suburbs and many people tend to drive to transit stations, while some others simply walk, it was decided to separate these two options as metra-fast and metra-slow, respectively.

Therefore, a three level NL model was developed in this study to explain choice of travel modes. As shown in figure 5, the tree of the NL model is structured in a way that the upper level has three limbs, which are *auto*, *transit*, and *non-motorized*. The *auto* limb divides into *drive* and *passenger* choices. The *transit* limb includes *Metra* branch (which is divided into *Metra-fast* and *Metra-slow* choices), *CTA*, and *Pace* alternatives. *Walk* and *bike* are choices in *Non-motorized* limb.



**Figure 5.** Tree Structure for the Nested Logit model

The probability of choosing each mode is given by equation 3, assuming a linear-in-parameter function of explanatory variables for the systematic utility, and maximizing the likelihood of occurrence for the sampled observations (30). In this equation, subscript (n) which denotes person n, is omitted for ease of representation, ( $P_k$ ) is the probability of choosing an upper-level nest (limb), ( $P_{j|k}$ ) is the conditional probability of choosing a lower-level nest (branch) given an upper-level nest, and ( $P_{i|j}$ ) is the conditional probability of choosing an alternative given a lower-level nest, if exists.

$$\begin{aligned}
 P_i &= P_k * P_{j|k} * P_{i|j} \\
 &= \frac{\exp\left(\frac{1}{\mu_k} \tau_k\right)}{\sum_{k' \in K} \exp\left(\frac{1}{\mu_{k'}} \tau_{k'}\right)} * \frac{\exp\left(\frac{\mu_k}{\mu_j} \tau_j\right)}{\sum_{j' \in J} \exp\left(\frac{\mu_k}{\mu_{j'}} \tau_{j'}\right)} * \frac{\exp(\mu_j V_i)}{\sum_{i' \in I} \exp(\mu_j V_{i'})} \quad (3)
 \end{aligned}$$

In this equation,  $\tau_j$  and  $\tau_k$  are, respectively, given by equations 4 and 5, and  $\mu$  is the coefficient of inclusive value (IV), also known as inverse logsum parameter.

$$\tau_j = \ln\left(\sum_{i' \in N_j} \exp(\mu_j V_{i'})\right) \quad (4)$$

$$\tau_k = \ln\left(\sum_{j' \in N_k} \left(\frac{\mu_k}{\mu_{j'}} \tau_{j'}\right)\right) \quad (5)$$

Table 7 presents results of NL model including coefficients of all significant variables and their t-statistics. All the variables have meaningful signs and are significant at a 99 percent confidence interval. Likelihood ratio test is also conducted to measure the overall goodness of fit of the model. The NL model has a likelihood ratio index of 66.7 percent, which is higher than 62.3 percent likelihood ratio index of the MNL model. The significant IV parameters shown in Table 7 convey a misspecification bias in the MNL model. The IV parameters have to meet certain conditions in order for a nested logit model to be consistent with the global utility maximization theory. They should be positive and less than 1. If the IV approaches 1, this represents a lower correlation within nest alternatives. When IV is equal to 1 in a nest, probabilities of alternatives in that nest would be estimated with a simple logit. In model results, IV for transit nest is 0.92 representing that there is low correlation among *CTA*, *Pace*, and *Metra*. As mentioned earlier, specifications of each transit agency such as spatial accessibility and distance between stops are specific. This fact leads to high value of IV parameter in the transit nest. Similarly, distance to *Metra* stations for people who have this option in their neighborhood, determines whether they drive or walk to the stations.

Adult individuals (over than 20 years old) tend to drive rather than walk and seniors (age>65) prefer the passenger mode. Sign of these coefficients are compatible with these facts. A person who travels more than 1 tour in a day is not willing to take a passenger mode. Obviously, higher density of transit stops encourages people to use transit modes. That is why their signs in both origin and destination zones are positive. In the same manner, increase in access and egress distances of transit modes dispirits people to ride transit modes. People prefer to use transit options for work tours. In contrast, they prefer to drive for shopping tours since these tours usually consist of several legs and they need more flexibility for these trips. Travelers tend to use transit modes rather than driving for trips that are destined to the CBD area during the rush hour. Finally, spending time in traffic jams, scarcity of available parking, and higher fuel consumption are among the factors that discourage people to drive.

**Table 7**  
**NL Model Estimated on google Map Data**

Home-based Mode Choice from Google Travel Time (Trip Origin is home)								
Nests (level 1)	Non-motorized		auto		transit			
Constant					Metra			
Alternatives	Walk	Bike	Drive	Passenger	CTA	Pace	Metra slow	Metra fast
<b>Nests (level 2)</b>		-1.87 (-5.2)				1.82 (9.16)		
TravelTime	-0.39 (-16.18)	-0.39 (-16.18)	-0.39 (-16.18)	-0.39 (-16.18)	-0.39 (-16.18)	-0.39 (-16.18)	-0.39 (-16.18)	-0.39 (-16.18)
TravelCost	-0.059 (-9.94)	-0.059 (-9.94)	-0.059 (-9.94)	-0.059 (-9.94)	-0.059 (-9.94)	-0.059 (-9.94)	-0.059 (-9.94)	-0.059 (-9.94)
AgeOver20 (Y N)	-0.17 (-2.73)		0.83 (20.86)					
AgeOver65 (Y N)				0.42 (9.55)				
TOTTR1				-4.89 (-11.01)				
nStopOrigin					0.022 (6.53)		0.77 (1.66)	
nStopDestination					0.0089 (13.96)			
DistanceWalk	-0.95 (-2.83)							
TourPurposeWork (Y N)					1.78 (22.83)		1.036 (3.27)	
TourPurposeShop (Y N)			0.14 (6.85)					
Female (Y N)			0.45 (16.86)					
AccessDistance						-0.80 (-5.05)	-0.45 (-3.13)	
EgressDistance						-0.33 (-2.40)		-0.49 (-1.47)
Income(10 <sup>-5</sup> )			0.26 (15.08)		-0.46 (-5.93)	-2.01 (-8.11)		
nTransfers					-0.33 (-4.91)		-0.27 (-1.77)	
DestCBD_RushHr (Y N)			-0.55 (-6.48)		0.77 (6.17)	1.44 (4.26)		
Inclusive value parameters:	0.58 (6.36)		0.34 (19.82)			0.89 (18.66)		0.95 (23.06)
<b>Log-likelihood at zero</b>	-56536.45							
<b>Log-likelihood at convergence</b>	-18719.01							
<b>Likelihood Ratio Index</b>	66.7%							
<b>Sample size</b>	40497							

## Conclusion

This paper demonstrates the importance of utilizing personalized travel time/cost and resource constraints in microsimulation-based mode choice models. A descriptive analysis was conducted on CMAP travel tracker survey to reveal the spatiotemporal mode choice constraints that individuals encountered when choosing the selected mode. These constraints should be taken into account when developing a model or simulating travel. Based on the useful information revealed in the descriptive analysis section, a methodology for improving the choice set formation of model development is proposed. The proposed methodology queries personalized travel time and cost information from Google Map and Goroo trip planner and links the attributes of the observed trips to the estimation data. Relevant resource constraints such as vehicle or transit availability at the time of the observed trip are also investigated to find out the available alternatives for the individual.

To examine the effectiveness of the proposed choice set formation approach, a mode choice model developed from this approach was compared to another model. The other model was estimated based on the traditional TAZ level travel time skim data for the alternative modes with no resource constraint considerations. The comparison approves the efficacy of the proposed choice set formation and suggests the use of this approach for future mode choice models. Finally, a nested mode choice model estimated from the new choice set approach is presented with detail mode alternatives for the Chicago metropolitan. For future work, the authors plan to propose a mode choice framework that incorporates various estimated models for different trips of a tour and tours of day in addition to the home-based model developed in this paper. This overall framework is planned to be implemented in the ADAPTS (31) dynamic activity-based travel demand framework.

## Acknowledgments

This research was completed through funding provided by Urban Transportation Center (UTC) at University of Illinois at Chicago (UIC). Funding was secured through National Center for Transit Research (NCTR), a US DOT-OST University Transportation Center and by funding from Illinois Department of Transportation (IDOT). The research team would like to thank Dr. P.S. Sriraj for providing guidance and participation in discussions. The views within this document are those of the author(s) alone.

## References

1. Bhat C, Koppelman F., (2003) Activity-Based Modeling of Travel Demand. In: Hall R, editor. Handbook of Transportation Science. International Series in Operations Research & Management Science. (56):39-65.
2. Bradley M, Bowman JL, Griesenbeck B., (2010) SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. Journal of Choice Modelling. 3(1):5-31.
3. Lemp J, McWethy L, Kockelman K., (2007) From Aggregate Methods to Microsimulation: Assessing Benefits of Microscopic Activity-Based Models of Travel Demand. Transportation Research Record: Journal of the Transportation Research Board. (1994):80-88.
4. Raney B, Cetin N, Völlmy A, Vrtic M, Axhausen K, Nagel K., (2003) An Agent-Based Microsimulation Model of Swiss Travel: First Results. Networks and Spatial Economics. 3(1):23-41.
5. Recker W, Parimi A., (1998) Development of a Microscopic Activity-Based Framework for Analyzing the Potential Impacts of Transportation Control Measures on Vehicle Emissions.
6. Walker J., (2005) Making Household Microsimulation of Travel and Activities Accessible to Planners. Transportation Research Record: Journal of the Transportation Research Board. 1931:38-48.
7. Malayath M, Verma A., (2013) Activity based travel demand models as a tool for evaluating sustainable transportation policies. Research in Transportation Economics. 38(1):45-66.
8. Yagi S, Mohammadian A., (2010) An Activity-Based Microsimulation Model of Travel Demand in the Jakarta Metropolitan Area. Journal of Choice Modelling. 3(1):32-57.
9. Yang, M., W. Wang, X. Chen, T. Wan, and R. Xu., (2007) Empirical analysis of commute trip chaining: Case study of Shangyu, China. In Transportation Research Record: Journal of the Transportation Research Board, (2038):139-147.
10. Zhou, X., X. Yang, & Y. Lao., (2007) Public transport service level influence on travel mode choice. Paper presented at the International Conference on Transportation Engineering, Chengdu, China.
11. Xiong, C., X. Chen, X. He, W. Guo, L. Zhang, (2015). The analysis of dynamic travel mode choice: a heterogeneous hidden Markov approach. In Transportation Research Board 94rd Annual Meeting, Number 15-5211.
12. Koppelman, F. S. (1983). Predicting transit ridership in response to transit service changes. Journal of Transportation Engineering 109(4), 548–564.
13. Bhat, C.R., N. Eluru, and R.B. Copperman, (2008), Flexible Model Structures for Discrete Choice Analysis, Handbook of Transport Modelling, 2nd edition, Chapter 5, pp. 75-104, edited by D.A. Hensher and K. J. Button, Elsevier Science.
14. McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. Journal of applied Econometrics 15(5), 447–470.

15. Vovsha, P. (1997). Application of cross-nested logit model to mode choice in Tel-aviv, Israel, metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board* 1607(1), 6–15.
16. Wen, C.-H. and F. S. Koppelman (2001). The generalized nested logit model. *Transportation Research Part B: Methodological* 35(7), 627–641.
17. Bhat, C.R. (1997), Work Travel Mode Choice and Number of Non-work Commute Stops, *Transportation Research Part B*, Vol. 31, No. 1, pp. 41-54.
18. Bhat, C.R., and R. Sardesai (2006), The Impact of Stop-Making and Travel Time Reliability on Commute Mode Choice, *Transportation Research Part B*, Vol. 40, No. 9, pp. 709-730.
19. Van Wee, B., H. Holwerda, et al. (2003), Preferences for modes, residential location and travel behavior: the relevance of land-use impacts on mobility." *European Journal of Transport and Infrastructure Research* 22 (3-4): 305-315.
20. Pinjari, A. R., R. M. Pendyala, C.R. Bhat and P.A. Waddell (2007). Modeling residential sorting effects to understand the impact of the built environment on commute mode choice. *Transportation* 34(5): 557-583
21. Frank, L., M. Bradley, et al. (2008), Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation* 35(1): 37-54.
22. Small, K. A., C. Winston, et al. (2005). Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica* 73(4): 1367-1382.
23. Li, Z., D. A. Hensher, and J.M. Rose (2010). Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation Review* 46(3): 384-403.
24. Manski, C. (1977), The structure of random utility models. *Theory and Decision*, Vol.8, pp. 229-254.
25. Swait, J., and M. Ben Akiva. (1987) Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil. *Transportation Research Part B*, Vol. 21, pp. 103-115.
26. Ben Akiva, M., and B. Boccara, (1995) Discrete choice models with latent choice sets. *International Journal of Research and Marketing*, Vol. 12, pp. 9-24.
27. Kunhikrishnan, K.P., K. Srinivasan (2015). Choice Set Variability and Contextual Heterogeneity in Work Trip Mode Choice in Chennai City. In *Transportation Research Board 94rd Annual Meeting*, Number 14-5203.
28. Parsons Brinckerhoff and Northwestern University (2013), CMAP Microsimulation Extension of the Activity-Based Model (ABM), Retrieved from [http://www.cmap.illinois.gov/documents/10180/98867/201309\\_NetworkMicrosimulation.pdf/cb044f3f-dfd3-4e1b-b81b-27a5c4a76e3c](http://www.cmap.illinois.gov/documents/10180/98867/201309_NetworkMicrosimulation.pdf/cb044f3f-dfd3-4e1b-b81b-27a5c4a76e3c)
29. Ben-Akiva, M., Lerman, S. (1985), *Discrete Choice Analysis*. MIT Press, London.

30. Coldren, G.M., Koppelman, F.S., (2005), Modeling the competition among air travel itinerary shares: GEV model development. *Transp. Res. A: Policy Pract.* 39 (4), 345–365.
31. Auld J, Mohammadian A.,(2012), Activity planning processes in the Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) model. *Transportation Research Part A: Policy and Practice*: 46(8):1386-403.