

**Accessibility and non-work destination choice: A microscopic
analysis of GPS travel data**

**A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY**

Yan Huang

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy**

Dr. David M. Levinson

January 2014

**© Yan Huang 2014
ALL RIGHTS RESERVED**

ABSTRACT

The advancements of GPS and GIS technologies provide new opportunities for investigating vehicle trip generation and destination choice at the microscopic level. This research models how land use and road network structure influence non-work, non-home vehicle trip generation and non-work destination choice in the context of trip chains, using the in-vehicle GPS travel data in the Minneapolis-St. Paul Metropolitan Area. This research includes three key parts: modeling non-work vehicle trip generation, modeling non-work, single-destination choice, and modeling non-work, two-destination choice. This research contributes to methodologies in modeling single-destination choice and multiple-destination choice and tests several hypotheses which were not investigated before.

In modeling non-work vehicle trip generation, this research identifies the correlation of trips made by the same individual in the trip generation models. To control for this effect, five mixed-effects models are systematically applied: mixed-effects linear model, mixed-effects log-linear model, mixed-effects negative binomial model, and mixed-effects ordered logistic model. The mixed-effects ordered logistic model produces the highest goodness of fit for our data and therefore is recommended.

In modeling non-work, single-destination choice, this research proposes a new method to build choice sets which combines survival analysis and random sampling. A systematic comparison of the goodness of fit of models with various choice set sizes is also performed to determine an appropriate choice set size. In modeling non-work, multiple-destination choice, this research proposes and compare three new approaches to build choice sets for two-destination choice in the context of trip chains. The outcomes of these approaches are empirically compared and we recommend the major/minor-destination approach for modeling two-destination choice. The modeling procedure can be expanded to trip chains with more than two destinations.

Our empirical findings reveal that:

1. Although accessibility around home is not found to have statistically significant effects on non-work vehicle trips, the diversity of services within 10 to 15 minutes and

15 and 20 minutes from home can help reduce the number of non-work vehicle trips.

2. Accessibility and diversity of services at destinations influence destination choice but they do not exert the same level of impact. The major destination in a trip chain tends to influence the decision more than the minor destination.
3. The more dissimilar the two destinations in a trip chain are, the more attractive the trip chain is.
4. Route-specific network measures such as turn index, speed discontinuity, axis of travel, and trip chains' travel time saving ratio display statistically significant effects on destination choice.

Our findings have implications on transportation planning for creating flourishing retail clusters and reducing the amount of vehicle travel.

Acknowledgement

Then Samuel took a stone and set it up between Mizpah and Shen and called its name Ebenezer; for he said: "Till now the Lord has helped us."

—1 Samuel 7:13 (ESV Bible)

It has been a challenging period of my life to go through this Ph.D. journey. I am thankful for starting this journey and being able to finish it. Over the past years, I have experienced joy, satisfaction, disappointments, sorrows, doubts, weaknesses, and most of all, love and hope which sustained me through this trial.

First of all, I want to thank Professor David Levinson who has discipled me on how to succeed as an academic scholar in almost all aspects. His strict training and high expectation of me pushed me farther than I would have pushed myself on this route. The past days of working with him have been really rewarding and fruitful. And the greatest fruit is that he helped shape my character in facing difficult times. My former colleagues and friends Carlos, Pavithra, and Shanjiang all encouraged me to persevere and showed care when I needed it. It is wonderful to have them as friends. I want to thank Jenny, my wife and my dearest friend who encouraged, challenged, comforted, and prayed for me earnestly. I would not have been able to finish this race without her. This dissertation is as much hers as it is mine.

Above all, this dissertation is a milestone in my life which always reminds me of the love and grace of the Lord Jesus Christ. He has been teaching me to live above my circumstances and to trust Him more and more. This journey attests that He has faithfully walked me through fire and waters and will continue to do so in the days to come.

Contents

List of Tables	viii
List of Figures	ix
List of Variables	xii
1 Introduction	1
1.1 Conceptual framework	3
2 Literature review	7
2.1 Land use and vehicle trip generation	8
2.1.1 Summary of empirical studies	8
2.2 Transportation network and travel behavior	9
2.3 Modeling approaches on trip generation	13
2.4 Destination choice models	14
2.4.1 Choice set formation	14
2.4.2 Modeling shopping destination choice	17
2.5 Using GPS data to study travel behavior	18
2.6 Perception of risks and time	21
3 Data	24
3.1 In-vehicle GPS data in the Twin Cities	24
3.2 Other data sets	28
3.2.1 Block-level business data in the Twin Cities (2010)	28
3.2.2 Metropolitan planning network data	28
3.2.3 TIGER/Line network data	28

3.3	Define non-work trips	29
3.4	Algorithm for identifying home-based trip chains	31
4	Modeling non-work, non-home vehicle trip generation	33
4.1	Descriptive statistics of non-work, non-home trips	35
4.2	Independent variables	35
4.3	Model structures	37
4.4	Results and analysis	41
4.4.1	Identifying random effects	41
4.4.2	Testing the number of categories for mixed-effects ordered logit model	41
4.4.3	Comparing Poisson and negative binomial models	45
4.4.4	Modeling results	45
4.4.5	Model fit	49
4.5	Predictive results	49
4.6	Conclusions	50
5	Modeling single-destination choice	53
5.1	The modeling procedure	54
5.2	Definition of destinations	54
5.3	Independent variables	55
5.3.1	Land use measures	55
5.3.2	Transportation network measures	57
5.3.3	Axis of travel	61
5.4	Choice set formation	62
5.5	Choice set size	68
5.6	Checking repeated destinations	69
5.7	Model formulation	71
5.8	Results and analysis	71
5.8.1	Choice set size	71
5.8.2	Modeling results	74
5.9	Discussion	77

6	Modeling two-destination choice	80
6.1	Modeling procedure	80
6.2	Definition of a choice	81
6.3	Independent variables	82
6.3.1	Land use measures	82
6.3.2	Transportation network measures	83
6.3.3	Axis of travel	83
6.3.4	Travel time savings	85
6.4	Choice set formation	85
6.4.1	Approach I	87
6.4.2	Approach II	87
6.4.3	Approach III	90
6.5	Choice set size	90
6.6	Model formulation	92
6.7	Results and analysis	92
6.7.1	Deciding choice set size	92
6.7.2	Modeling results	94
6.8	Discussion	100
7	Conclusions	103
	Bibliography	107

List of Tables

2.1	Summary of selected studies on trip generation	10
2.2	Summary of selected studies on trip generation (continued)	11
2.3	Summary of selected studies on destination choice from literature	19
2.4	Summary of selected studies on destination choice from literature (continued)	20
3.1	Two types of GPS devices used in this study	26
3.2	Comparison of socio-demographics in the GPS data and the Twin Cities Metropolitan Area (Source: Minnesota DEED)	27
4.1	Hypotheses on the relationships between key independent variables and vehicle trip generation.	38
4.2	Regression results for different mixed-effects model structures (# of obs: 4988)	46
4.3	The coefficients for predicting the probability of zero trips for the mixed- effects ZINB model (using Monday as the base term)	48
4.4	Correlation of accessibility and diversity of services in [5, 10) min, [10, 15) min and [15, 20) min driving zones around home.	48
4.5	The models' mean average error (MAE) and mean average percentage error (MAPE)	51
5.1	A comparison of log likelihoods of different distributions of $\ln(T_k)$ for a single subject with GPSID 1019.	66
5.2	A comparison of different distributions of $\ln(T_k)$ for all subjects	67
5.3	Modeling single-destination choice for non-work vehicle trips	75
5.4	Correlations of key independent variables	77

5.5	Elasticity of the odds of single-destination choice for key independent variables (1% change)	77
6.1	Hypotheses of key independent variables	86
6.2	Modeling home-based, two-destination choice in the context of trip chains (using basic independent variables)	95
6.3	Correlations of land use variables in Model 3	99
6.4	Correlations of route-specific network measures in Model 3	99
6.5	Modeling home-based, two-destination choice in the context of trip chains (excluding the entropy measure)	100
6.6	Modeling home-based, two-destination choice in the context of trip chains (excluding the entropy measure)	101
6.7	Elasticity of the likelihood of destination choice for key independent variables (with 1% change)	101

List of Figures

1.1	A conceptual framework of the relationships among transportation networks, retail distribution patterns, consumers' non-work destination choice, and retail location choice.	6
2.1	The land use-transport feedback cycle.	8
2.2	A diamond of action for travel decision-making.	9
2.3	Hierarchical structure in choice sets.	15
2.4	A hypothetical market with two alternatives for a consumer to make home-based shopping trips (source: Brooks et al. (2004)).	22
3.1	The timeline of the GPS Travel data collection process.	25
3.2	An individual's shopping trip trajectory captured by an in-vehicle GPS device on October 5th, 2008 in Minneapolis, MN.	27
3.3	The percentiles of parking distances (in meters) from home and work in the GPS data.	30
3.4	The flow chart for extracting home-based non-work trip chains from the in-vehicle GPS data.	32
4.1	Probability distribution function (pdf) and cumulative distribution function (cdf) of daily non-work, non-home vehicle trips from the GPS data.	34
4.2	The [0, 5) min, [5, 10) min, [10, 15) min and [15, 20) min driving zones around one individual (GPSID 1036)'s home address. The driving zones are created based on the road network with an estimated travel speed on each road link.	37
4.3	Residual plots of the fixed-effects linear model	42

4.4	95 percent confidence intervals for the intercepts and slopes of the individual-based regressions of the number of daily trips on the day-of-week variable.	43
4.5	Comparison of the goodness of fit for selecting the number of categories of trips for the mixed-effects ordered logit model	44
4.6	Predicted and observed non-work, non-home vehicle trip patterns.	51
5.1	Examples of entropy measures with each color representing one type of service.	58
5.2	Two examples of calculating speed discontinuity.	59
5.3	Two examples of calculating turn index.	60
5.4	The analogy for using survival analysis to predict the probability of making a trip.	63
5.5	The procedure to form choice sets which combines survival analysis and random sampling.	64
5.6	The cumulative probability distribution of distances between destinations visited by the same individual in the GPS data.	70
5.7	The root mean squared error (RMSE) value and running time for models of different choice set sizes.	72
5.8	RMSE values for single-destination choice models of different choice sizes.	73
6.1	Multiple choices of home-based non-work, two-destination trip chains.	81
6.2	Examples of calculating the similarity index.	83
6.3	Examples of calculating travel time saving ratio.	86
6.4	Approach I to select two destinations to constitute a trip chain.	88
6.5	Approach II to select two destinations to constitute a trip chain.	89
6.6	Approach III to select two destinations to constitute a trip chain.	91
6.7	The RMSE values and running times for the models of different sizes developed by Approach I.	93
6.8	The RMSE values and running times for the models of different sizes developed by Approach II.	94
6.9	The RMSE values and running times for the models of different sizes developed by Approach III.	96

6.10 Comparison of root mean squared errors of different choice set sizes for models developed by the three approaches.	97
---	----

List of Variables

Chapter 4

$A_{[0,5)}$	Accessibility within the [0, 5) min driving zone from home
$A_{[5,10)}$	Accessibility within the [5, 10) min driving zone from home
$A_{[10,15)}$	Accessibility within the [10, 15) min driving zone from home
$A_{[15,20)}$	Accessibility within the [15, 20) min driving zone from home
a	Random-effect term for a subject
b	Another random-effect term for a subject
$H_{([0,5)}$	Diversity of services within the [0, 5) min driving zone from home
$H_{[5,10)}$	Diversity of services within the [5, 10) min driving zone from home
$H_{[10,15)}$	Diversity of services within the [10, 15) min driving zone from home
$H_{[15,20)}$	Diversity of services within the [15, 20) min driving zone from home
S	Vector of a subject's sociodemographic variables
U_t	Utility of making auto trips on day t
W_t	Vector of day-of-week and monthly dummy variables
Y_t	Number of trips made by a subject on day t
Λ	Vector of land use measurements around home
ϵ_t	Random component in a utility function
$\Phi(\cdot)$	Standard normal cumulative distribution

Chapter 5

A_k	Accessibility at destination k
H_k	Diversity of services at destination k
K	Total number of destinations in the study area
M	Choice set size
N	Total number of observations in a data set
N_1	Total number of chosen destinations in a data set
N_2	Total number of unchosen destinations in a data set
T_k	Travel time of the fastest route from home to k
$T_{w,k}$	Travel time between workplace and destination k
$T_{d,k}$	Travel time between destination k and the nearest downtown
U_k	Utility of choosing destination k
v_q	Travel speed on road link q
V	Total number of service types at a destination
Υ_k	Interaction terms between socio-demographics and land use/network
Θ_k	Vector of transportation network measures
κ_n	Binary dependent variable for checking whether a destination is visited
$\hat{\kappa}_n$	Predicted probability in the n th observation
ϑ_k	Turn index of the fastest route from home to destination k
$\lambda(\cdot)$	Hazard function
Λ_k	Vector of land use measurements at destination k
ρ_{kv}	Proportion of service type v at destination k
τ_k	Total number of turns of the fastest route from home to k
ψ_k	Speed discontinuity of the fastest route from home to k
Ω_k	Survival probability of visiting destination k

Chapter 6

A_j	Accessibility at destination j
A_k	Accessibility at destination k
H_j	Diversity of services at destination j
H_k	Diversity of services at destination k
g	Common category of services at destination j and k
G	Total number of common types of services at destination j and k
$T_{j,k}$	Home-based trip chain travel time
T_j	Home-based round trip travel time by a subject to visit j
T_k	Home-based round trip travel time by a subject to visit k
$T_{w,jk}$	Travel time between the major destination and workplace
$T_{d,jk}$	Travel time between the major destination and the nearest downtown
$U_{j,k}$	Utility of choosing destinations j and k
$\Delta_{j,k}$	Vector of axis of travel at destinations j and k
$\zeta_{j,k}$	Perceived traveling saving ratio of a trip chain
$\Theta_{j,k}$	Vector of transportation network measures of the trip chain's travel route
$\vartheta_{j,k}$	Turn index of a trip chain with two destinations j and k
$\Lambda_{j,k}$	Vector of land use measurements at destinations j and k
ρ_{kg}	Proportion of service type g at destination k
ρ_{jg}	Proportion of service type g at destination j
$\Upsilon_{j,k}$	Interaction terms between socio-demographics and land use/road network

Chapter 1

Introduction

This research focuses on non-work vehicle trips. According to the 2009 National Household Travel Survey, non-work trips make up approximately 90% of total trips. Non-work trips include a spectrum of trip purposes, such as social/recreational trips, shopping trips, family/personal errands, and school/church activities. Because of its significant share in daily travel, it is of interest to investigate non-work travel behavior and the influencing factors using empirical data.

Take shopping trip behavior for example. Recent data show that shopping travel behavior in North America has evolved over time. The average number of annual household visits to grocery stores in Canada declined from 85 in 1998 to 75 visits in 2001 while at the same time the average number of yearly visits to supercenters (agglomerations of stores) and dollar stores raced ahead from 14 visits in 1998 to 18 visits in 2001 ([ACNielsen, 2002](#)). In addition, trip-chaining behavior (visiting multiple destinations in a chain of trips) has become increasingly common. For example, according to the 1995 Nationwide Personal Transportation Survey data, about 61.2% of women and 46.4% of men made one or more stops from work to home. For home-to-home trip tours, the average number of stops was 2 for women and 1.8 for men ([McGuckin and Murakami, 1999](#)). It is therefore important to gain insights into trip chaining behavior and to understand the factors contributing to such decision-making processes.

In terms of mode choice for non-work trips, both our life experience and statistics attest the fact of automobile dominance. The 2009 National Household Travel Survey Data

reveal that private vehicles account for 87.8% of family/personal errand trips, 70.7% of church/school trips, and 76.9% of social/recreational trips.

The aforementioned facts on trip chaining behavior, automobile dominance, and the built environment may have intrinsic relationships. Based on the positive feedback loop theory (Levinson and Krizek, 2008), two basic feedback loops can be hypothesized. In the first loop, multi-purpose/multi-stop trip chaining behavior provides incentives for the creation of strip malls, atrium malls, and other forms of retail clusters. And the development of retail clusters further stimulates multi-purpose/multi-stop behavior. In the second loop, high vehicle use influences stores' location choice, the layout of retail clusters, and the evolution of road network structure, which on the other hand endorses the convenience of driving. The two feedback loops are connected through a variety of factors. The details are discussed in Section 5.1.

The objective of this dissertation is to contribute to understanding non-work vehicle trip behavior by investigating:

1. the impact of land use around home on vehicle trip generation,
2. the influence of land use and transportation networks on non-work, single-destination choice, and
3. the influence of land use and transportation networks on non-work, two-destination choice in the context of trip chains.

These questions are addressed at the microscopic level based on the GPS travel data collected by the Nexus Research Group at the University of Minnesota in 2008. The GPS travel data recorded 141 individuals' travel routes for about 3 months in the Minneapolis-St. Paul Metropolitan Area. The GPS data provide opportunities to examine hypotheses that were not tested before, such as the relationship between route-specific network structure patterns and destination choice in the context of trip chains and the impact of individuals' mental representations of destinations on destination choice in trip chaining behavior. To summarize, the main contributions of this research include:

- Propose a conceptual model to understand the connections between various built environment factors and complex non-work travel decisions.

- Empirically examine how land use around home influences non-work, non-home vehicle trip generation by systematically comparing different model structures.
- Propose a new method to form choice sets for single-destination, non-work destination choice.
- Propose three approaches to build choice sets for two-destination, non-work destination choice in the context of trip chains.

Our research findings shed light on the impacts of the built environment on vehicle trip generation and why people drive to certain places for non-work purposes in the context of trip chains. The broader impact would be on retail regulatory policy and the design and planning of retail clusters and road networks. Some exemplary questions that this research may provide implications on include:

- How should retail clusters be located to improve the profits of retailers?
- How should retail clusters and road networks be designed to influence non-work trip chains?
- What incentives and disincentives may be provided to reduce the amount of non-work vehicle travel?

1.1 Conceptual framework

Figure 1.1 shows the hypothetical relationships among retail distribution patterns, transportation network structure, and consumers' destination choice. It is assumed that the default travel mode is vehicle which fits the context of in-vehicle GPS data in this study. Retail distribution patterns influence the types of potential activities and the number of potential activities that consumers can engage in. These potential activities provide opportunities for multi-purpose shopping or multi-stop shopping. Road network structure, along with retail distribution patterns, influences consumers' travel distance to access destinations. Travel distance and network topological features further impact travel time and perceived reachability of destinations. On the one hand, households' sociodemographic

characteristics, multi-purpose potential, and perceived reachability of destinations influence non-work trip generation, non-work destination choice, and trip chaining behavior. On the other hand, consumers' shopping travel behavior influences retail location choice from the demand side. For example, if consumers want to purchase certain complementary goods in a trip chain, retailers selling these products may have the incentive to locate closer to each other. Also, existing retail distribution patterns affect retail market conditions by providing retailers with opportunities for competition and cooperation. After assessing retail market conditions, individual retailers select locations in order to maximize profits. Stores' location choice behavior shapes retail distribution patterns at the macroscopic level.

The above feedback loops portray the relationships among microscopic behavior and the macroscopic built environment factors. The complexity lies in the fact that it involves a large number of agents who interact with each other to make decisions in different spatial and temporal settings. My previous research modeled retail location choice and the evolution of transportation networks (Huang and Levinson, 2009, 2011; Levinson and Huang, 2012). Continuing the previous efforts, this research focuses on non-work vehicle trip generation and non-work destination choice in the context of trip chains.

Non-work destination choice in this dissertation is modeled with three steps. In the first step, an individual mentally defines locations. Some may see stores as destinations; others may view clusters of stores such as strip malls as destinations. The definitions of locations can be of large or small granularity depending on the topic of study. If available, surveyed information regarding perceptions of locations would serve as a useful reference. In the second step, individuals form choice sets according to their preferences, perceptions, trip purposes, and priorities from a bigger set of alternatives (universal set/awareness set/evoked set). In modeling, a variety of choice set formation strategies may be applied to replicate this process. It would be beneficial to complement quantitative modeling with qualitative information through surveys and/or interviews to understand this internal choice set formation process for a specific trip purpose. In the third step, individuals select their favorite destination(s) after collecting and processing a multitude of factors such as trip purpose, travel time, location-specific factors, and route-specific factors.

In modeling such travel decision-making processes, there are several avenues. The first one is utility-based models. The basic random utility model has been expanded in literature to consider increasingly complex trip decisions (Dellaert et al., 2008; Arentze et al., 2005). A detailed review of previous models can be found in Chapter 2. The second approach is behavior-based using controlled experiments. Researchers in psychology and behavioral science have indicated that utility models cannot fully account for “non-normative effects in consumers’ conception and evaluations of shopping trip alternatives” (Brooks et al., 2004; Dellaert et al., 2008). This research adopts utility-based models to understand the data because they provide actionable procedure to analyze GPS data for modeling non-work vehicle trip generation and destination choice. The models used in this research are programmed using the SAS statistical software.

The rest of this dissertation is organized as follows:

- Chapter 2 reviews literature on modeling trip generation and destination choice.
- Chapter 3 introduces the data sets used in this research and describes the GPS data collection process.
- Chapter 4 examines how land use around impacts non-work vehicle trip generation.
- Chapter 5 models home-based non-work, single-destination choice.
- Chapter 6 models home-based non-work, two-destination choice in trip chains.
- Chapter 7 summarizes key findings and discusses future work.

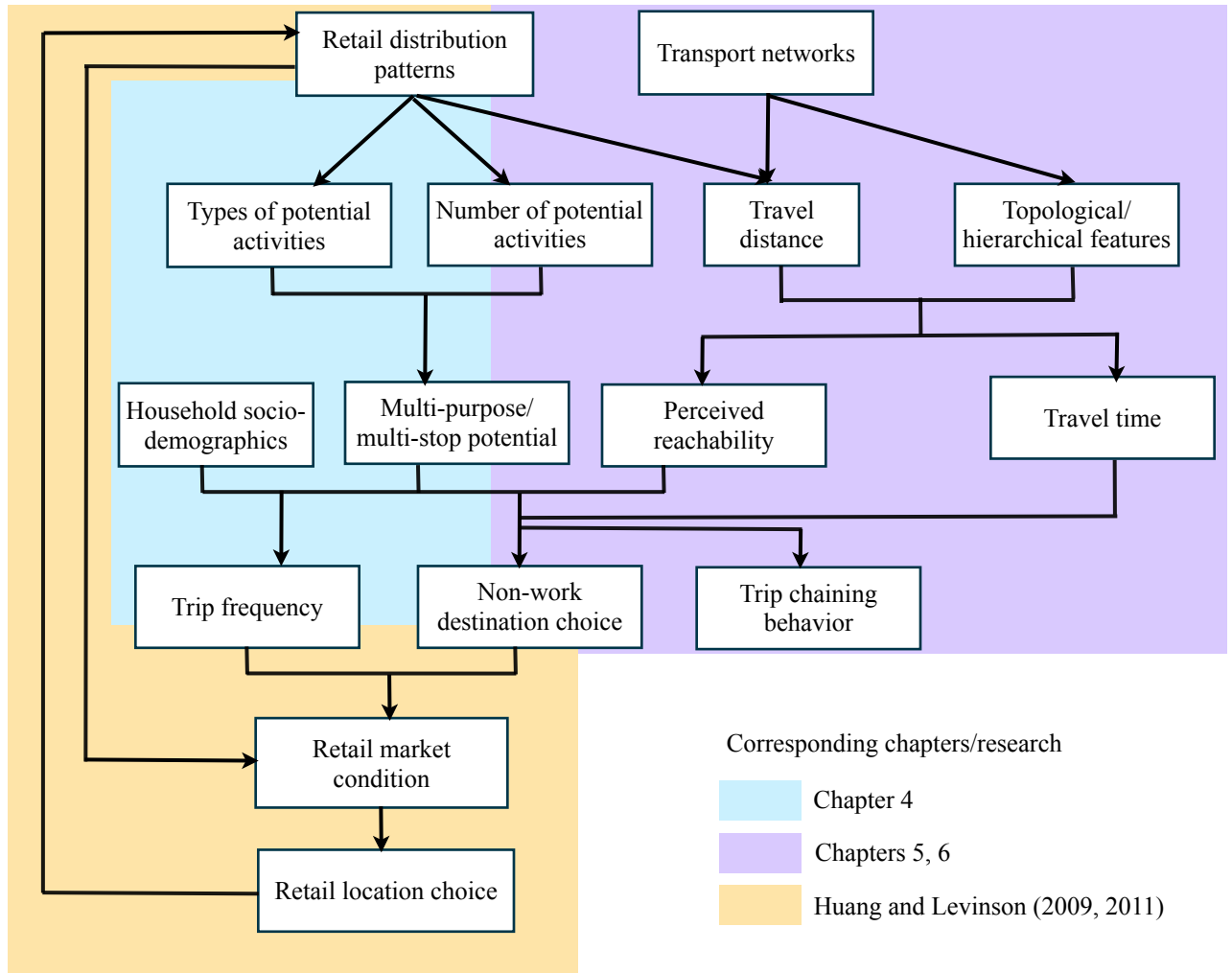


Figure 1.1: A conceptual framework of the relationships among transportation networks, retail distribution patterns, consumers' non-work destination choice, and retail location choice.

Chapter 2

Literature review

Travel behavior and the built environment have been found to be closely related ([Handy, 1996](#); [Schwanen et al., 2004](#); [Scott and Horner, 2008](#)). Accessibility is one key concept linking the two together. Accessibility measures the ease of accessing activities in a place to reflect “the desire of people or firm to overcome spatial segregation” ([Hansen, 1959](#)). Accessibility considers two aspects of the transport-land use connection. First, it evaluates the spatial distribution of amenities (e.g., stadiums and retail stores) and opportunities (e.g., jobs and services). Second, it concerns the ease of access, influenced by transportation networks, mode choice, and distance between places.

There are three types of accessibility measures: cumulative opportunities, gravity-based measures, and random utility theory ([Handy and Clifton, 2001](#)). These measures serve to characterize the relationships among opportunities to access, travel cost, and transportation and land use systems. Conceptually, such relationships can be illustrated in [Figure 2.1](#), where creating congestion, creating access, increasing land prices, building infrastructure, and travel decisions constitute a closed loop ([Levinson and Krizek, 2008](#)). Creating access to services leads to higher land prices by inducing travel; higher land prices result in creating more access through land development. Creating access can bring more congestion, which further incentivizes more investment on infrastructure and creating more access. [Figure 2.2](#) shows that individuals make travel decisions based on existing opportunities and constraints (e.g., quality of service and travel cost) from competitive and complementary service providers that they desire to patronize ([Levinson and Krizek,](#)

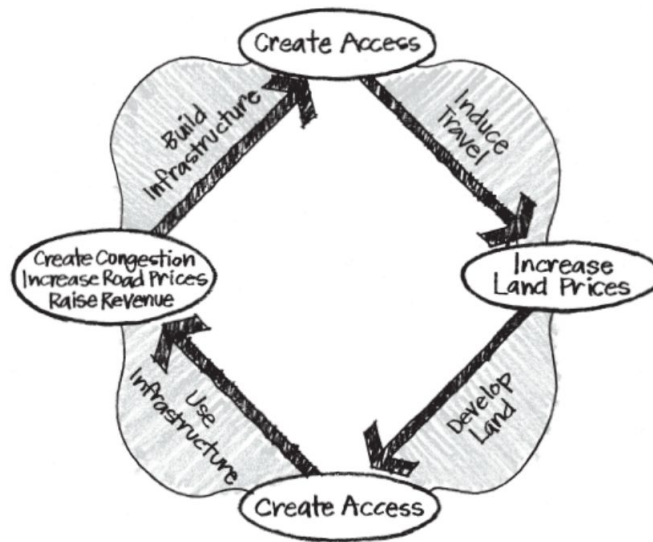


Figure 2.1: The land use-transport feedback cycle.
 (source: [Levinson and Krizek \(2008\)](#))

[2008](#)). Such relationships are assessed in modeling destination choice.

This chapter consists of five sections. The first section overviews existing empirical studies on the influence of the built environment on vehicle trip generation. The second section introduces models in modeling non-work destination choice and key research findings. The third section summarizes research on using GPS data to study travel behavior. The fourth section reviews existing transportation network measures. The reference-dependent theory in economics is introduced in the last section.

2.1 Land use and vehicle trip generation

2.1.1 Summary of empirical studies

There has been a plethora of studies which examine the influence of the built environment on vehicle trip generation. Most studies found that land use influences vehicle trip generation ([Handy, 1996](#); [Schwanen et al., 2004](#); [Scott and Horner, 2008](#); [Levinson and Krizek, 2008](#)). Independent variables that have been used include: residential/employment density, availability and quality of transit services, pedestrian accessibility, distance to desti-

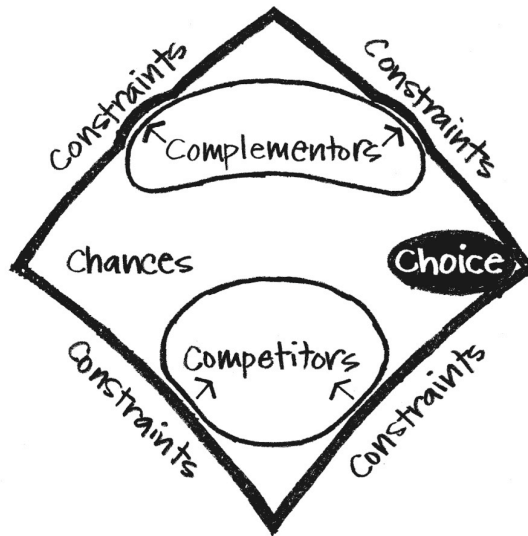


Figure 2.2: A diamond of action for travel decision-making.
 (source: [Levinson and Krizek \(2008\)](#))

nations, mixed land use, destination accessibility, parking supply and cost, vehicle ownership, socio-demographics, attitudes toward mode choice, residential location choice, and street design. Examples of dependent variables include: trip frequency, VMT, household trip generation rate, and the proportion of trips using a particular mode. Some exemplary studies on trip generation are summarized in [Table 2.1](#) and [Table 2.2](#). The key differences among these studies include data sets, dependent variables, and modeling approaches. The travel data used were mostly based on paper-and-pencil surveys where individuals were asked to reflect upon their past travel experience.

2.2 Transportation network and travel behavior

Transportation network structure also influences travel behavior. Traditional interests in transportation networks are in the fields of geography (seeing networks as an input to regional development) ([Taaffe et al., 1996](#)) and physics (focusing on the topology and spatial evolution of the networks) ([Gastner and Newman, 2006](#)). Yet the connection between transportation network structure and travel behavior has not been sufficiently in-

Table 2.1: Summary of selected studies on trip generation

Study	Data	Key findings
Cervero (1991)	83 randomly-sampled buildings in 6 US cities' suburban activity centers	(1) Lower density, single land uses, and high supplies of parking are associated with greater vehicle trip rate and that greater density is associated with more walking trips to work. (2) Clustered, mixed-used developments, high-quality transit services, and higher prices for automobile usage may reinforce each other in suburban employment settings.
Frank and Pivo (1994)	Travel survey data in Washington State	(1) Mode choice and employment density is nonlinear, and that population density has the strongest relationship with mode choice. (2) Land use mix measured at the Census tract level only revealed weak relationships with mode choice. (3) Further examining this relationship at a smaller scale is needed.
Handy (1996)	3 traditional neighborhoods and 1 modern neighborhood in California	(1) Urban form (such as distance, barriers, and pedestrian access) may influence perception of walking as a mode choice. (2) A greater range of destination choices may result in more travel.
Schimek (1996)	Florida travel survey data	After controlling for sociodemographic variables, residential density, mixed use, and accessibility do not have statistically significant effects on household trip rates.
Kitamura et al. (1997)	Five San Francisco Bay Area neighborhoods	(1) Neighborhood environment factors (such as parking space, distance to nearest bus stop, distance to nearest park) are statistically associated with the amount of vehicle travel and mode choice. (2) Individuals' attitudes toward travel (such as pro-environment and pro-transit) are even more strongly associated with their travel behavior.
Cervero and Kockelman (1997)	1990 travel diary data in the San Francisco Bay Area	(1) Proposed a 3D framework: density (population/employment density), diversity (dissimilarity index, entropy, etc) of services, design (street patterns, pedestrian and cycling provisions, etc.) to quantify the built environment. (2) The built environment factors have a statistically significant but modest effect on mode choice and VMT.

Table 2.2: Summary of selected studies on trip generation (continued)

Study	Data	Key findings
Dieleman et al. (2002)	Netherlands National Travel Survey	(1) Individuals living in the top 3 big cities, all else equal, make more fewer vehicle trips and more biking and walking trips to shop. (2) Personal/household attributes (income, car ownership, household type, and education) are statistically significant in influencing mode choice and trip distance. (3) A bigger household size is associated with longer vehicle trip distance and longer walking and biking distance.
Pinjari et al. (2007)	2000 San Francisco Bay Area household travel survey	“Residential sorting” (individuals’ ranking of different neighborhoods) should be controlled to understand the impacts of the built environment on travel choices based on their attitudes, values, life style, and travel preferences.
Cao et al. (2009)	8 neighborhoods in Northern California	(1) Household structure influences trip generation. (2) Neighborhood characteristics affect trip frequency. (3) Travel attitudes and residential preferences (self-selection effects) affect travel decisions.
Cao (2009)	8 neighborhoods in Northern California	(1) Individuals’s self-selection of residential locations affect vehicle miles driven. (2) The effect of the built environment on driving behavior outweighs that of self-selection.
Cao et al. (2010)	Regional travel data in Raleigh, NC	(1) Individuals’ vehicle miles driven are influenced by their residential location choice. (2) The self-selection effect is nontrivial.
Ewing and Cervero (2010)	Meta-analysis to summarize the literature’s findings	(1) Destination accessibility and distance to downtown have the strongest association with VMT. The next strong variable is intersection and street connectivity. (2) Density was found only to have a weak relationship with VMT. (3) Density is an intermediate variable that is often expressed by other Ds.

vestigated, particularly at the microscopic level. For example, the 3D principles (density, diversity, and design) by [Cervero and Kockelman \(1997\)](#) did not include transportation networks.

With the availability of fine-grained network data, the past decade has witnessed a growing interest in studying the connection between travel behavior and transportation networks. [Kissling \(1969\)](#) analyzed the linkage importance of a regional highway network based on accessibility. [Xie and Levinson \(2007\)](#) proposed a set of new network measures such as ringness, webness, beltiness, circuitness, and treeness. ([Xie and Levinson, 2011](#)) systematically summarized various network measures and investigated the economic models of transportation network growth. [Hess \(1997\)](#) used block size, length and completeness of sidewalk networks to indicate street network connectivity to explain the pedestrian volumes between two neighborhoods. [Jiang et al. \(2009\)](#) studied human mobility patterns in the context of a large street network. [Levinson and El-Geneidy \(2009\)](#) created a network circuitry measure (the ratio of road network distance to Euclidean distance) to understand the choice of home-work pairs. Based on the data in the Twin Cities and Portland, they found that road network pattern transitions from grid-like to tree-like as moving away from the city center. All else equal, workers are more likely choose roads with lower circuitry ([Levinson and El-Geneidy, 2009](#)). [Derrible and Kennedy \(2009, 2010\)](#) quantified the network structure of 33 metro systems, and found that the metro network structure and transit riderships are closely associated. Hierarchy, topology, morphology, and scale of road networks were found to be associated with household spatial activities, road congestion levels, trip distance, and daily vehicle kilometers traveled per capita ([Parthasarathi, 2011; Parthasarathi et al., 2012](#)). Based on a 2009 survey of students at Davis High School in Davis, CA, [Emond and Handy \(2012\)](#) concluded that students' perception of travel distance from home to school (an effect of the biking network) strongly affects their choice of biking even after controlling for actual travel distance. Individuals' attitudes toward biking and perceptions on the environment for biking (e.g., quality of bicycle routes and adjacency to work) were found to impact the amount of bicycle travel ([Schoner, 2013](#)).

2.3 Modeling approaches on trip generation

In the literature, models for trip generation include linear, log linear, Poisson/negative binomial, Tobit/logit, and factor analysis. [Barmby and Doornik \(1989\)](#) compared Poisson model and the negative binomial (NB) model in modeling shopping trip frequency. The predicted results at the aggregate level suggested that the NB model is a better fit. [Cotrus et al. \(2005\)](#) applied linear regression and Tobit model to investigate the transferability of person-level disaggregate trip generation models based on the 1984 Travel survey data in Tel Aviv and Haifa and the 1996/97 Israeli National Travel Habits Survey data. They found that Tobit models perform better than the linear models but neither model transfers well in time. [Ma and Goulias \(1999\)](#) employed the Poisson, geometric, and negative binomial models to estimate daily activity frequencies by activity type using the first 4 years of the Puget Sound Transportation Panel data. The authors found that NB fits the data better than the Poisson distribution for all activity types. [Paez et al. \(2007\)](#) adopted mixed ordered probit model to examine elderly trip generation in Hamilton, Canada. [Agyemang-Duah and Hall \(1997\)](#) analyzed spatial transferability of an ordered response model, and the results showed that a directly transferred ordered response model performs well in predicting shopping trip generation. Yet this research did not compare the model with alternative models. [Jang \(2005\)](#) compared Poisson, NB, and zero-inflated Poisson models in modeling non-home trips at the household level, and claimed that negative binomial model and the modified Poisson models can improve reliability when there are overdispersion and heterogeneity in the count data. Based on the 2009 National Household Travel data, [Lim and Srinivasan \(2011\)](#) further adopted linear, log linear, Poisson, NB, and ordered probit models to model trip generation for trips with various purposes. Their results suggested that the ordered probit model performs the best. To control for many zeros in biking trip data, [Schoner \(2013\)](#) adopted a zero-inflated negative binomial (ZINB) model to estimate the frequency of commuting trips by bicycle, and found that the ZINB model fits the data better than NB and logit models.

In summary, the aforementioned findings suggest that count-data models work better than linear models in modeling trip generation. But most of the models are fixed-effects models with only one observation for one individual. It is still unclear whether they can

be properly applied to GPS travel data with repeated observations for one individual, and the goodness of fit of these model structures is yet to be investigated for GPS travel data. The next section reviews studies on destination choice.

2.4 Destination choice models

With the invention of the discrete choice model, destination choice problems have been a topic of extensive study. This section first describes different approaches of constructing choice sets, and then reviews models used in modeling shopping destination choice.

2.4.1 Choice set formation

In a choice problem, a choice set is comprised of all alternatives considered by a traveler. Given a large number of potential destinations in an area for an individual, it is unrealistic that one individual considers all of them. Furthermore, it is computationally burdensome to incorporate all of them into a model. The key challenges lie in how to select choice alternatives and how to decide an appropriate choice set size in modeling.

In terms of selecting choice alternatives, [Ben-Akiva and Lerman \(1985\)](#) introduced a series of choice sampling approaches: simple random sampling (a sample is drawn at random from the whole population), general stratified sampling (partitioning the population into a certain number of exclusive strata, selecting sampling fractions, and randomly drawing a designated number of alternatives from each strata), exogenous sampling (defining strata by segments only on attributes and not on actual choices), choice-based samples (choices are defined and controlled for subjects), enriched sampling (the pooling of exogenously stratified samples with one more choice-based samples), and multistage sampling (multiple surveys on preferences of choices). Among these approaches, the most widely used approaches are simple random sampling and general stratified sampling. Both approaches assume that every subject has perfect knowledge of all locations. [Spiggle and Sewall \(1987\)](#) and [Shocker et al. \(1991\)](#) suggested that the decision-making process involves nested sets of alternatives from higher to lower levels: total set, awareness set, evoked set, and choice (consideration) set, where the lower level set is a subset of the higher level

set. [Spiggle and Sewall \(1987\)](#)'s research revealed that the proportions of subsets vary by individuals and by different brands/products.

[Hoogendoorn-Lanser and Van Nes \(2004\)](#) and [Zhu \(2010\)](#) argued that there exists a hierarchical structure in choice sets ([Figure 2.3](#)). All available choices comprise the universal set, of which an individual knows a proportion (subjective choice set) and only considers feasible ones (consideration set). Only some choices in the consideration set are selected into the actual choice set in a decision-making process.

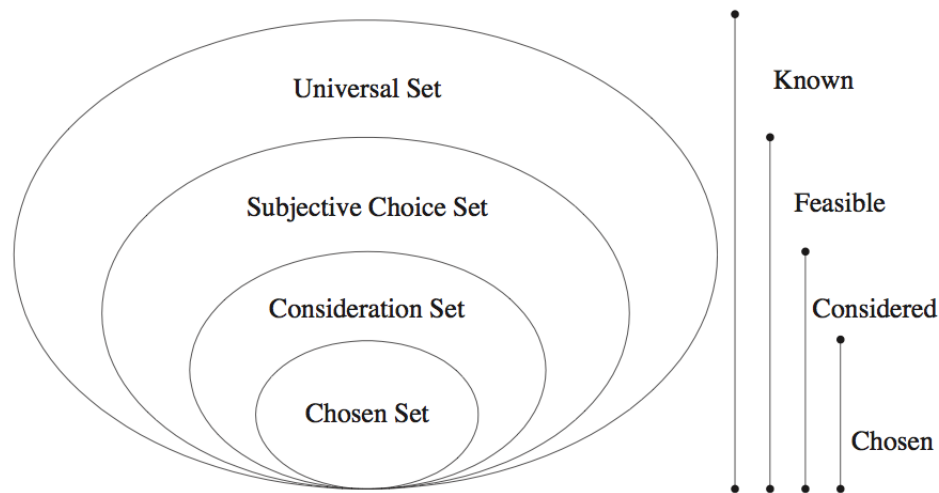


Figure 2.3: Hierarchical structure in choice sets.
(Source: [Zhu \(2010\)](#))

Empirically there are different variations of the hierarchical structure for determining a choice set. One direction is to impose time constraints on the universal choice set. Travel time budgets are assumed to be the limiting source that restricts the feasible alternatives to only a subset of the universal choice set ([Thill and Horowitz, 2002](#)). For example, [Thill and Horowitz \(1997\)](#) used travel time constraints to downsize destination choice sets. Several other models used the space-time prism to reduce choice set size given their spatial-temporal constraints in formulating choice sets for activity-based models ([Kitamura et al., 1997](#); [Kwan and Hong, 1998](#); [Auld and Mohammadian, 2011](#); [Scott and He, 2012](#)). [Horni et al. \(2009\)](#) built a simulation module to simulate activity patterns. For each destination choice, a time budget for an activity is calculated, based on which a set of destinations that can be reached within the given time is generated. However, this model is more applica-

ble to fixed activities and less applicable to flexible activities, and it does not consider the spatial relationships of destinations for destination choice. Another direction is to select alternative destinations based on certain criteria (such as distance, priorities, lifestyle, and utilities) in combination with aforementioned methods to reduce the universal choice set size ([Rashidi et al., 2012](#)). More complicated models include a model with random constraints ([Swait and Ben-Akiva, 1987](#)) and the GenL model ([Swait, 2001](#)). Nonetheless, they tend to be computationally inefficient. This area needs to be further investigated to balance the complexity of sampling techniques and computational efficiency.

[Willumsen and Ortuzar \(2001\)](#) summarized three methods of choice set formation: (1) random/stratified sampling, (2) heuristic or deterministic approach, and (3) learning about preferences about locations through surveys. Each approach has its limitations. The random sampling approach, though simple and good at incorporating all possible alternatives, treats all locations with the same weight yet does not fully consider the spatial effect. Given one trip's starting point, there are more locations that are farther away than locations nearby. Therefore, a random-selection approach tends to incorporate more farther-away destinations than nearby locations into the choice set. The deterministic choice-set generation approach aims to limit the number of destinations by defining a radius from the starting point. Nevertheless, the value of radius is hard to pinpoint and may vary by person and trip purpose thanks to inter-personal heterogeneity. The stratified importance sampling approach tries to improve simple random sampling by grouping locations with the same or close weights. However, it is challenging to justify the number of groups that should be defined, and different numbers of groups may produce different estimates of coefficients. The surveying approach, though can provide useful information regarding priorities, preferences, and tastes, suffers from that fact it is almost impossible to ask individuals to precisely document all their alternatives in their decision-making processes. Furthermore, there may exist differences between stated preference and revealed preferences.

The size of a choice set is also worth studying. [Adler and Ben-Akiva \(1979\)](#) proposed a theoretical model to combine destination choice and mode choice and the actual choice set size used for modeling can be huge. To illustrate, let's assume a household has a max-

imum of 2 destinations to visit on a given day and there are 3 travel modes available. If the household chooses not to travel, it is one alternative. If the household visits one destination and each destination is presumed to have 20 alternatives (randomly selected from all destinations), each of which can be reached by 3 possible travel modes, it becomes 60 possible destinations. If the household visits two destinations in separate trips and each trip may use a different mode, there are 570 choices. If the two destinations are visited in one tour, there are 1710 possible choices. In total, the household has 1981 ($= 1 + 270 + 1710$) possible travel patterns for one day. The advantage of this model is that it considers trip-chaining behavior as an alternative and combines destination choice with mode choice. The disadvantages of this model include: (1) It does not consider the spatial and land use connections of the destinations in a trip tour. (2) The number of alternative destinations used is a hypothetical number which needs to be justified in application. (3) If one more alternative destination is added, the choice set size will be 303 more choices than the base case. If one more travel mode is added, the choice set will have 1100 more than the base case.

It should be noted that most of destination choice studies lack a systematic investigation on the appropriate choice set size for modeling. [Nerella and Bhat \(2004\)](#) performed numerical experiments to examine the effects of choice set size on the performance of Multinomial (MNL) models and mixed multinomial logit (MMNL) models. The research recommended a choice set size of a fourth of the full choice set for MNL models and one-half of the full choice set for MMNL models. Yet these numbers may still be too large for shopping destination choice models if there are a large number of potential destinations in a metropolitan area. How to decide an appropriate choice set size for non-work trips is worth investigating.

2.4.2 Modeling shopping destination choice

Traditional utility-based models have been developed to model shopping destination choice. The basic structure is the multinomial logit model (MNL). [McFadden \(1978\)](#) showed that the MNL model can consistently estimate parameters from a sample of alternatives through maximizing the conditional likelihood function, a feature that makes MNL widely used in

modeling discrete choices. One important hypothesis about MNL is the independence of irrelevant alternatives property (IIA): the random components of the utilities are independently and identically distributed (i.i.d.) (Ben-Akiva and Lerman, 1985). Since this assumption does not hold in many cases (such as the red bus/blue bus paradox), different extensions of the model have been developed. Two kinds of models receive the most attention: generalized extreme value (GEV) class and mixed multinomial logit (MMNL) class of models (Bhat and Guo, 2004). In addition, because retail services are often clustered, there are two schools of thought: (1) A consumer considers all possible alternatives. (2) Individuals initially evaluate clusters of alternatives and then evaluate alternatives in a cluster (Fotheringham, 1988). In a choice set built using the importance sampling approach, the choice set formation process introduces sampling bias into the model, which needs to be corrected for the model specification (Ben-Akiva and Lerman, 1985). Kitamura (1984) modeled destination choice by considering trip chaining behavior. A zone's "prospective utility" function incorporates not only the expected utility of the visit to that zone but also the utilities of the adjacent zones that may be visited.

Table 2.3 and Table 2.4 list some exemplary studies on discrete choice models applied in studying shopping destination choice. Such studies tend to use traffic analysis zones or specific stores (such as big supermarkets or malls) as destinations.

2.5 Using GPS data to study travel behavior

The advancements of GPS and GIS technologies provide new opportunities and challenges for investigating trip generation. According to Draijer et al. (2000), GPS devices have the following advantages over traditional paper-and-pencil diary methods: (1) Real-time spatial and temporal information of a trip is available, such as distance, travel times, travel speed, and route information. (2) Fewer misreporting or underreporting of trips. (3) Data are stored in digital formats. (4) The subjects' burden of reporting travel information is reduced. Therefore, GPS data can better incorporate trip information for modeling trip generation than traditional data collection methods. In addition, the procedure to draw trip trajectories based on GPS points in GIS can be automated. Readers may refer to Li (2004), Quddus et al. (2005), Quddus et al. (2007), and Zhu (2010) for details on processing

Table 2.3: Summary of selected studies on destination choice from literature

Study	Data	Topic	Model	Key findings
Timmermans (1996)	Travel survey data in Eindhoven, Netherlands	Sequential mode and destination choice for shopping trips	MNL	Mode choice does not influence the choice of shopping centers. Shopping centers with greater size are more attractive than those of smaller size.
Pellegrini et al. (1997)	Phone survey data on shopping trips in Gainesville FL	Parameter sensitive to choice set specification for shopping destination choice	MNL	The stability of parameter estimates can be sensitive to choice set size and composition.
Bhat (1998)	1990 San Francisco Bay Area Household Travel Survey	Travel mode and departure time choice of shopping trips	MNL for mode choice and MNL-OGEV for departure time choice	In estimating travel time choice, nested logit model outperforms the MNL model and MNL-OGEV model outperforms the nested logit model in terms of data fit.
Leszczyc et al. (2000)	Grocery shopping data in Springfield, MO	Consumers' store choice and trip time choice	Hazard model and MNL	Store choice and shopping time choice are interdependent. Spatial competition between stores affects consumers' store choice and switching behavior.
Pozsgay and Bhat (2001)	1996 Dallas-Fort Worth household activity survey	Destination choice for home-based recreational trips	nonlinear-parameter MNL	Agglomeration effects are prominent in affecting recreational attraction-end choice.
Bernardin et al. (2009)	Household survey data in Knoxville, Tennessee	Destination choice of home based maintenance trips and home-based other trips	MNL and ACDC (agglomerating and competing destination choice models)	The ACDC model reflects the effects of trip chaining and spatial agglomeration whereas MNL cannot.
Newman and Bernardin (2010)	2000 Knoxville Urban Area Household Travel Behavior data	Mode choice and destination choice for work tours	Hierarchical ordering nested logit	Hierarchical ordering of decision nesting trees is important for modeling location and mode choice; employing a reverse ordering can be a good choice.

Table 2.4: Summary of selected studies on destination choice from literature (continued)

Study	Data	Topic	Model	Key findings
Recker and Kostyniuk (1978)	Survey data of 1500 households in Buffalo, NY	Shopping destination choice	MNL	Individuals' attitudes toward the store and its operation, perception of destinations' reachability, and the number of destinations at the destinations influence destination choice.
Fotheringham (1986)	Immigration data in 62 US Cities (1965-1970)	Hierarchical destination choice	Competing destination model	Competing destinations model perform better than gravity-based model for hierarchical destination choice. The model is limited by the lack of consideration of trip chaining behavior.
Arentze et al. (2005)	Household survey data in Northern Brabant in the Netherlands	Combined choices of trip purpose and destination	Nested logit	(1) The presence and size of different types of stores influence location choices. (2) Consumers prefer shopping centers they are more familiar with and prefer to visit a single large agglomeration of stores to conduct multi-purpose shopping.
Wang and Lo (2007)	Stated shopping preference data of Chinese immigrants in Toronto	Supermarket destination choice	MNL	Immigrant shoppers' socio-demographic attributes more affect their shop destination choice than distance and accessibility.
De Palma et al. (2010)	Numerical examples	Modeling both retail location choice and consumers' destination choice with the consideration of trip chaining behavior	Decision tree and logit model	Trip chaining option reduces the profit margins in the short run but increases welfare for firms.
Auld and Mohammadian (2011)	Activity travel survey of households in Chicago	Non-work destination choice	MNL	Choice set formation that considers planning constraints improves prediction accuracy.

GPS travel data in the GIS environment. A challenge is that one person makes multiple trips or travel decisions over a certain period of time in the GPS data, which suggests that there may exist correlations among observations. How to appropriately select a model is a key issue.

The first few studies using GPS to collect travel data date to the 1990s (Wagner, 1997; Casas and Arce, 1999; Draijer et al., 2000). Since then, the GPS technology has gained popularity in collecting travel data as its precision improves over time. Some existing research using GPS to study travel behavior include: Li et al. (2004) (inspecting travel time variability in commute trips, and its effects on departure time and route choice, including cases with trip-chaining), Li et al. (2005) (analyzing attributes determining whether to choose one or more routes in the morning commute), Wolf et al. (2003) (capturing under-reporting trips in household travel surveys), Zhang and Levinson (2008) (estimating the value of information for travelers and the impact of travel time, distance, and aesthetics), Zhu (2010) (evaluating route choice behavior after the collapse of I-35 W Bridge in Minneapolis), and Carrion (2010) (studying travel time perception, valuation of time, and route choice for work trips).

2.6 Perception of risks and time

Perception of risk may also influence destination choice. Tversky and Kahneman (1991) proposed a reference-dependent theory to explain perception of risks. There are three key features in this theory: reference dependence, loss aversion, and diminishing sensitivity. Reference dependence states that “the carriers of value are gains and losses defined relative to a reference point” (Tversky and Kahneman, 1991). Loss aversion means that the function is steeper for the loss than the gain. Diminishing sensitivity indicates that “the marginal value of both gains and losses decreases with their size” (Tversky and Kahneman, 1991). The theory argues that the subjective desirability function is concave for gains and convex for losses (S-shaped function). This theory has been applied in studying perception of waiting time (Leclerc et al., 1995) and store-choice decisions (Brooks et al., 2004).

Brooks et al. (2004) illustrated this theory using one simple example with two alternatives in Figure 2.4. A person starts a trip from home and needs to visit two stores A and

HYPOTHETICAL MARKET

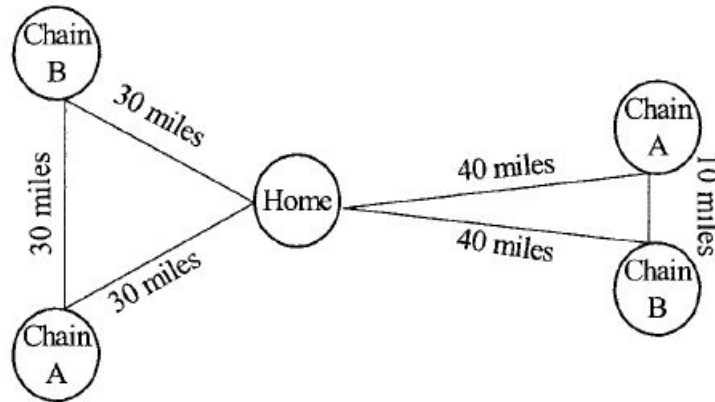


Figure 2.4: A hypothetical market with two alternatives for a consumer to make home-based shopping trips (source: Brooks et al. (2004)).

B before going back home. If the stores in both alternatives are the same and the total trip tour lengths for both routes are the same, the consumer needs to decide which alternative to choose based on the characteristics of the routes. The route to the left is comprised of 30-mile, 30-mile, and 30-mile links; the route to the right consists of 40-mile, 10-mile, and 40-mile links. Since travel is a cost, its function presumably has a convex shape. According to the feature of diminishing sensitivity for an S-shaped function, the total subjective costs of the route with three 30-mile links are more negative than the route with 40-mile, 10-mile, and 40-mile links. Therefore the route to the right is preferred.

If we consider travel time savings of chained trips, the route to the right also outweighs the route to the left. We assume that the reference point is the travel distance of making two separate trips to visit the two destinations. The reference point for the alternative to the left is 120 miles (two 60-mile round trips), and the reference point for the alternative to the right is 160 miles (two 80-mile round trips). The travel time saving percentage for the route to the left equals $(120 - 90)/90 \times 100\% = 33.33\%$. In comparison, the travel time saving percentage for the route to the right equals $(160 - 90)/90 \times 100\% = 77.78\%$. The alternative to the right is chosen because of greater perceived travel time savings/transaction utility.

In short, in chained trips consumers not only minimize the subjective costs in travel time, but also maximize the perceived savings in travel time relative to a neutral refer-

ence point (Brooks et al., 2004). Controlled experiments through surveys and simulation programs to investigate multi-destination shopping behavior confirmed such hypotheses (Brooks et al., 2004, 2008). These studies also revealed individuals' preference for high-clustered stores reached by right-hand turns.

Furthermore, research indicated that spatial characteristics of an environment influence how people perceive time. For example, Raghurir et al. (2011) performed three controlled experiments to ask individuals to estimate their travel time for home trips and non-home trips. The results showed that there exist "going-home effects": Perception of home trips' travel time is shorter than the actual travel time, but their perception of travel time for non-home trips tends to be longer than the actual time. Going home effects also exist for familiar locations (Raghurir et al., 2011). Boroditsky and Ramscar (2002) investigated the perception of time when they were waiting line for food, boarding, or picking up someone. The authors argued that people's thinking about time is related with their spatial thinking and their spatial experiences. Boroditsky (2000) suggested that abstract domains (such as time) are defined by "metaphorical mappings from more concrete and experiential domains such as space".

Chapter 3

Data

This chapter first introduces the in-vehicle GPS travel data and other data sets used in this dissertation, following which is the definition of non-work trips based on the in-vehicle GPS data. The algorithm for identifying home-based non-work trip chains is further described.

3.1 In-vehicle GPS data in the Twin Cities

The in-vehicle GPS data used in this research are the same as the data used in [Zhu \(2010\)](#) and [Carrion and Levinson \(2012\)](#). The Minneapolis-St. Paul Metropolitan Area is selected as the study area because of the availability of the unique in-vehicle GPS travel data. The original goal of collecting the GPS data was to understand commuters' travel behavior after the collapse of the old I-35 W Bridge and after the opening of the new I-35 W Bridge in Minneapolis ([Zhu, 2010](#)). In this study the same data set is used to study non-work travel behavior.

The collection process lasted from September to December of 2008, during which 141 surveyed subjects made over 20,000 trips. The in-vehicle GPS data collection process includes three stages ([Figure 3.1](#)). The first stage is to recruit the subjects. The announcements on recruiting subjects were posted on various media such as Craigslist.com and Citypages.com, and were sent out via other forms such as postcards handed out in downtown parking ramps and emails sent to about 7000 University of Minnesota staff (exclud-

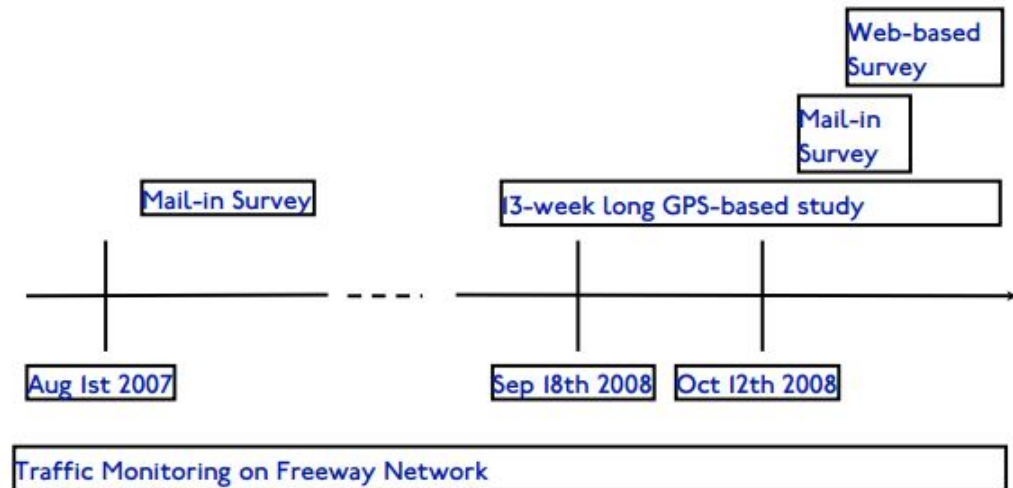


Figure 3.1: The timeline of the GPS Travel data collection process.
 (Source: [Zhu \(2010\)](#))

ing students and faculty). More than 900 people responded, from which subjects were selected based on the following criteria:

- Age between 25 and 65
- Legal driver
- Have a full-time job and follow a “common” work schedule
- Drive alone to work
- Affected by the opening of the new I-35W Mississippi River bridge

The second stage is to collect the data by installing GPS devices in participants’ vehicles. There were two types of GPS devices used in his study which are compared in Table 3.1. The first type was the real-time tracking GPS devices provided by the subcontractor Vehicle Monitoring Technologies (VMTInc). A local subcontractors was hired to install the GPS devices ([Zhu, 2010](#)). The GPS devices recorded the coordinates of the vehicle every second while the vehicle is turned on. There were 46 participants who used this type of GPS device in this study, and the data collection process lasted for 13 weeks. The participants were told to follow their regular travel routes and to complete periodic online

Table 3.1: Two types of GPS devices used in this study

Product	Feature	Frequency	No. of subjects	Diaries
Qstarz VT-1000	Logging	One point per 25 meters	97	No
VMT, Inc	Real time	One point per second	46	Yes

surveys about the trips made. The second type was the logging GPS (QSTARZ BT-Q1000p GPS Travel Recorder). Different from the previous type, the data can only be exported manually at the end of the study. The GPS frequency was one point per 25 meters. In total, 97 subjects' vehicles were equipped with this type of GPS devices. Participants were asked to periodically complete surveys about their trip purposes.

The third stage is to create GPS trip trajectories. The trip trajectories were drawn based on the GPS points in the underlying the Metropolitan Planning Network. The technical details on creating such trajectories can be found in [Zhu \(2010\)](#). Figure 3.2 shows an exemplary non-work trip trajectory on October 5th, 2008.

Table 3.2 summarizes the subjects' socio-demographic information, which is compared with the overall socio-demographic data in the Minneapolis-St. Paul Metropolitan Area (summarized in [Carrion and Levinson \(2012\)](#)). The percentage of women in our data is higher than the Twin Cities Metropolitan Area. This is probably because of a relatively high proportion of female staff from the University of Minnesota among the subjects. In addition, more people in our data hold degrees above high school than the Twin Cities Metropolitan Area, which is probably influenced by the participants from the University of Minnesota. The overall income level in the GPS data is also higher than Twin Cities Metropolitan Area. This is because one needs to own a car to be qualified for this study.

The statistics indicate that the subjects are probably not representative of the overall population in the Minneapolis-St. Paul Metropolitan Area. There are multiple reasons for that. The first one is the selection criteria. Note that the above criteria for selecting subjects are required by research in [Zhu \(2010\)](#), but are not necessarily the desired requirements for this study. The second reason is that the data collection process somewhat depends on how people responded to the recruitment ads, car ownership, and willingness to participate, etc. While we caution against generalizing the findings to the entire metropolitan area,

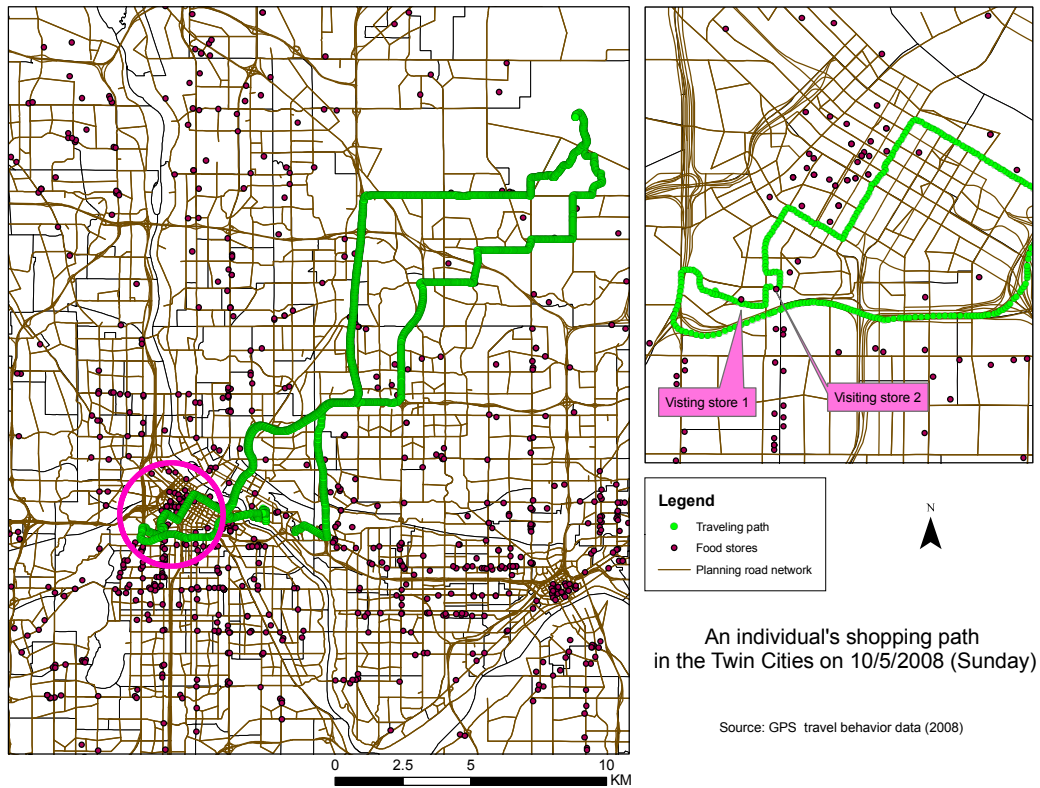


Figure 3.2: An individual's shopping trip trajectory captured by an in-vehicle GPS device on October 5th, 2008 in Minneapolis, MN.

Table 3.2: Comparison of socio-demographics in the GPS data and the Twin Cities Metropolitan Area (Source: Minnesota DEED)

Variable	Category	GPS data (%)	Twin Cities (%)
Gender	Male	41.25	49.40
	Female	58.75	50.60
Education	11th grade or less	0	9.40
	High School	13.09	49.60
	Associate	24.99	7.70
	Bachelor	45.22	23.20
	Graduate	16.69	10.10
Household Income	< \$49,999	20.20	45.20
	\$50,000 – \$74,999	30.73	23.30
	\$75,000 – \$124,999	29.44	14.60
	> \$124,999	20.16	16.90
Race	White	83.06	87.70
	Black	7.36	6.20
	Others	9.58	6.10

one main goal of this research is to propose and showcase the methodologies of using GPS travel data to investigate microscopic non-work travel behavior.

3.2 Other data sets

There are three other network data used in this study: block-level business data in the Minneapolis-St. Paul, the Metropolitan planning network, and TLG road network data.

3.2.1 Block-level business data in the Twin Cities (2010)

The business data at the block level in the Minneapolis-St. Paul Metropolitan Area documents the number of establishments ¹ categorized by 15-digit North American Industry Classification System (NAICS) codes. Out of 54378 blocks from Census 2010 in the Minneapolis-St. Paul Metropolitan Area, there are 16851 blocks with at least one establishment. This data set is used to measure land use at potential destinations ².

3.2.2 Metropolitan planning network data

The Metropolitan planning network data in the Twin Cities are used to create fastest travel routes to destinations. Zhu (2010) estimated the travel speed on the network links by three time periods (morning peak hours, afternoon peak hours, and midday) for the Metropolitan planning network based on the GPS data in the Twin Cities. This research uses the same GPS trajectories as in Zhu (2010) which created the GPS trip trajectories using the Metropolitan planning network ³.

3.2.3 TIGER/Line network data

The TIGER/Line network data from the Census Bureau's MAF/TIGER database is more comprehensive than the Metropolitan planning network because the TIGER/Line network

¹According to the US Census, an establishment is defined as: "generally a single physical location where business is conducted or where services or industrial operations are performed".

²This data set can be downloaded from <http://mn.gov/deed/data/data-tools/index.jsp>.

³This data set can be downloaded from <http://www.datafinder.org/catalog/index.asp>.

data include all local roads. The road network data are used to create walking zones around potential destinations ⁴.

3.3 Define non-work trips

The focus of this research is on non-work vehicle trips. Chapter 4 models the generation of non-work, non-home vehicle trips; Chapter 5 and Chapter 6 address the destination choice problem for home-based non-work trip chains. The first tasks are to identify non-work trips and to identify home-based non-work trip chains. To achieve this goal, the in-vehicle GPS trips with available travel diaries are analyzed. It is important to measure how far the subjects parked from home/work for the trips that were indicated as home trips or work trips. The procedure is as follows:

1. Match the time stamp of the in-vehicle GPS data and one individual's travel diary, and create a "trip purpose" attribute for the in-vehicle GPS data.
2. Select the trips whose purpose is work or home from the in-vehicle GPS data.
3. For a home trip, calculate the Euclidean distance between the trip destination and home. For a work trip, calculate the Euclidean distance between the trip destination and one's work address.

The percentiles of parking distances from home for home trips in the GPS data are calculated to identify an appropriate threshold (Figure 3.3). The maximum distance from home and the distances at the 90th threshold and above clearly subject misreporting of trip purposes. We focus on around 85th percentile where the distance is about 1190 meters and at 80th percentile is about 208 meters; the midpoint between the thresholds is about 700 meters. A further investigation of the 83rd percentile reveals a distance of 897.38 meters. Based on this information and our educated guess, we use 800 meters as the maximum parking distance from home for home trips.

Figure 3.3 further shows the percentiles of parking distances from work. The range of distances is much wider than that of distances from home. Besides the possibility of mis-

⁴This data set can be downloaded from <http://www.census.gov/geo/maps-data/data/tiger-line.html>.

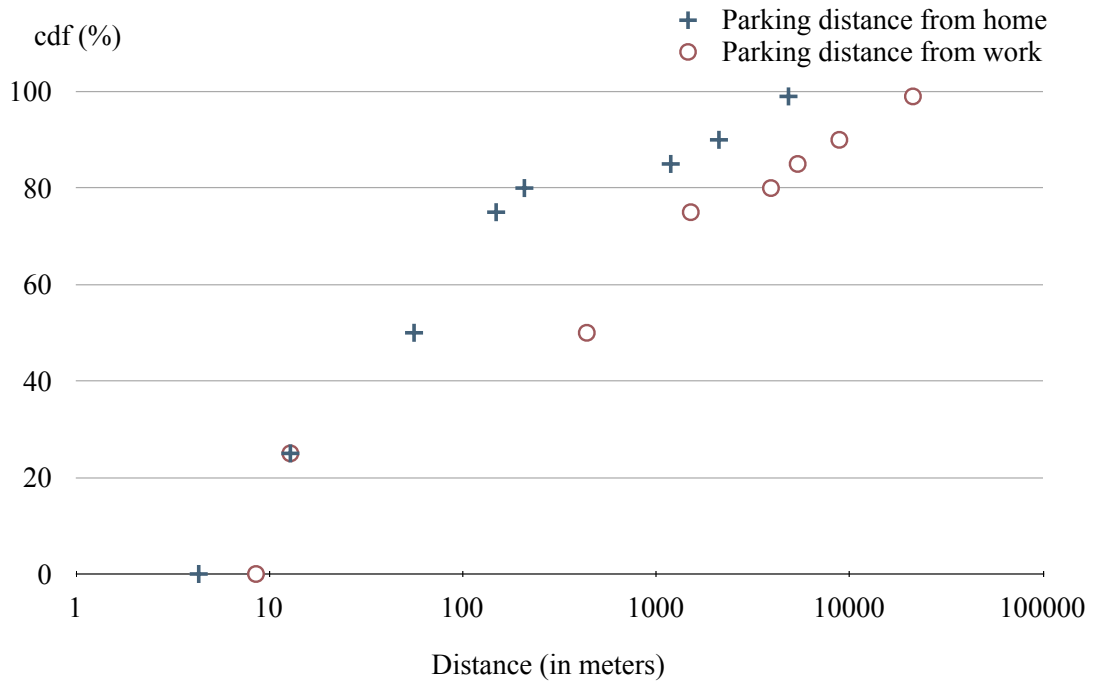


Figure 3.3: The percentiles of parking distances (in meters) from home and work in the GPS data.

reporting trip purposes, it may be due to that one did not actually go to the reported work address for work, as traveling to another site for work-related purposes is possible. Yet such information about other possible work sites is unknown to us. The distance between the 65th and 70th percentiles seems a reasonable threshold. Thus, we use 1000 meters as the maximum parking distance from work for work trips.

In summary, the threshold for parking distance from home is selected as 800 meters, and the threshold for parking distance from work is set as 1000 meters. Therefore, if a trip's parking destination is within 800 meters from home, it is defined as a home trip. If a trip's parking destination is greater than 1000 meters from work, it is defined as a non-work trip.

3.4 Algorithm for identifying home-based trip chains

An algorithm is proposed to extract home-based non-work trip chains from the in-vehicle GPS data (Figure 3.4). First, the in-vehicle GPS trips need to be sorted by the date of travel and the starting time for each individual. Then the program creates a new empty trip chain list and read the first trip's information. If the first trip starts from home, add it to the trip chain list and move to the second trip; otherwise, drop this trip and move to the second trip. The program further investigates if the second trip is a work trip. If yes, then drop it and move to the third trip; otherwise, add it to the trip chain list and move to the third trip. This loop of search continues until a trip ending at home is found. After a home trip is found, the trip chain list is closed. In the following, a new empty trip chain is created and the same process repeats. The whole program ends when all trips in the GPS data have been searched. This algorithm was implemented on the Netlogo Programming platform.

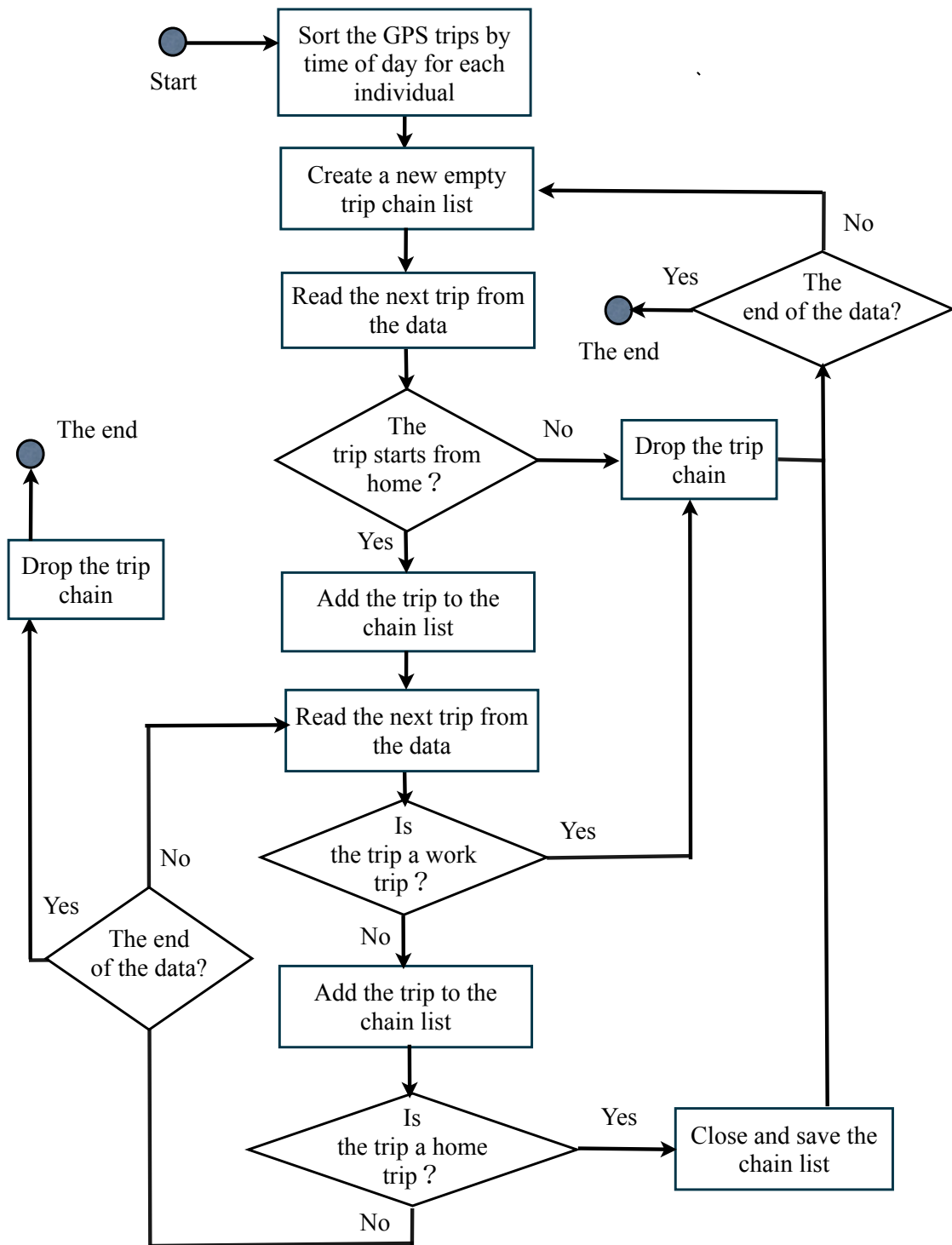


Figure 3.4: The flow chart for extracting home-based non-work trip chains from the in-vehicle GPS data.

Chapter 4

Modeling non-work, non-home vehicle trip generation

This chapter studies the impacts of land use around home on non-work, non-home vehicle trip generation using the in-vehicle GPS data. The key contributions of this research are:

- Examining the appropriateness of the mixed-effects model structures for in-vehicle GPS data with repeated observations for individuals.
- Systematically investigating five mixed-effects model structures in modeling non-work vehicle trip generation.
- Studying the impact of land use around home at the parcel level based on travel time from home on non-work vehicle trip generation.

The rest of this chapter is organized as follows. Section 4.1 describes the basic statistics about non-work, non-home vehicle trips from the GPS data, following which is the description of five model structures for modeling non-work, non-home vehicle trip generation. Section 4.4 shows the results of the five models. Section 4.5 presents the predicted trip generation data from the models at the aggregate level and compare them with the actual data. Section 4.6 discusses the results and summarizes key findings.

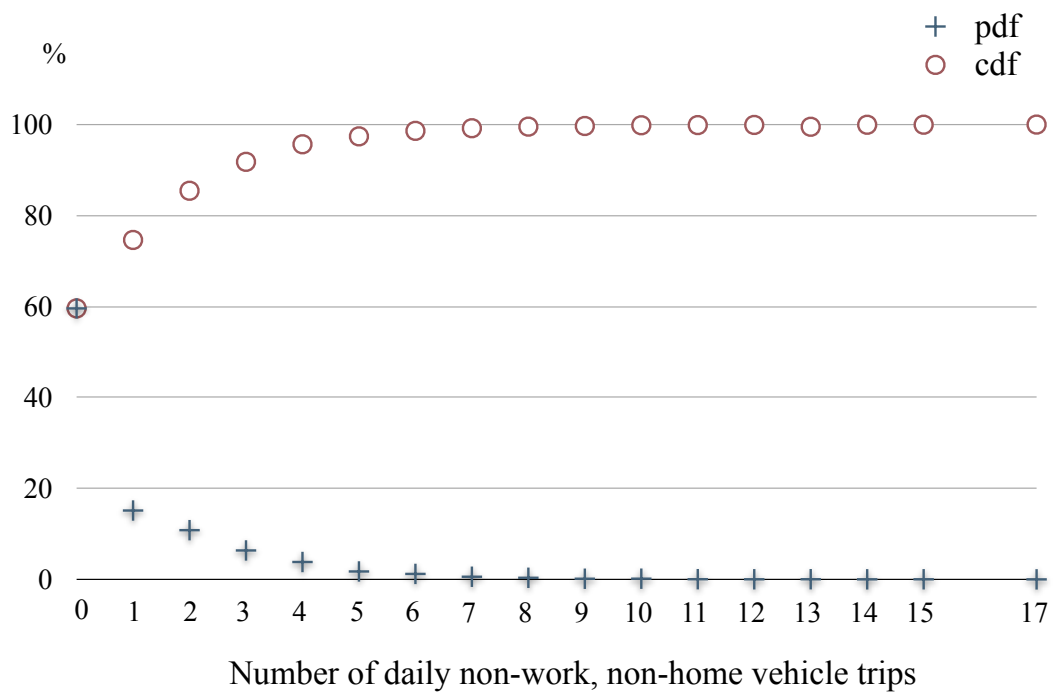


Figure 4.1: Probability distribution function (pdf) and cumulative distribution function (cdf) of daily non-work, non-home vehicle trips from the GPS data.

4.1 Descriptive statistics of non-work, non-home trips

This chapter focuses on non-work, non-home vehicle trips, suggesting that these trips' destinations should be at least 1000 meters away from work and be at least 800 meters away from home based on our findings in Chapter 3. Figure 4.1 shows the distribution of such trips from all participants. The number of daily non-work, non-home trips ranges from 0 to 17 (16 is missing). In total, there are 1832 days with zero non-work, non-home vehicle trips, which is the highest frequency. The lowest frequency is the number of days with 17 trips with only two records. The average number of daily non-work, non-home vehicle trips equals 1.57 with standard deviation 1.86. The 25th percentile is located at 0 trips and the median equals 1 trip. The 75th percentile of the data is located at 5 trips.

For the rest of this chapter, unless otherwise specified, all trips refer to non-work, non-home vehicle trips. For the simplicity of presentation, the term "trips" will be used.

4.2 Independent variables

The independent variables used in this chapter include land use measures around home, day-of-week variables, and individuals' socio-demographics such as age, gender, and income.

To measure land use around home at the microscopic level, we create [0, 5) min¹, [5,10) min, [10, 15) min, and [15, 20) min driving zones around home using the road network with an estimated travel speed on each road. One example of the created driving zones is shown in Figure 4.2. The key land use measures include accessibility and diversity of services (land use mix).

The cumulative opportunities measure which calculates the sum of the services within a zone is used to measure accessibility. An assumption of this measure is that each service in a zone has an equal opportunity (likelihood) to be visited. We argue that it is approximately true as each zone is defined within a small time interval. The empirical tests reveal that the \ln form of accessibility produces greater goodness of fit for the models. Therefore, the \ln form of the cumulative accessibility measure in the four driving zones around

¹The [0, 5) min zone is the area between 0 minutes (inclusive) and 5 minutes (exclusive) from home by vehicle.

home is employed. $A_{[0,5)}$, $A_{[5,10)}$, $A_{[10,15)}$, and $A_{[15,20)}$ respectively indicate the accessibility measures in the [0, 5) min, [5, 10) min, [10, 15) min, and [15, 20) min driving zones from a subject's home.

The diversity of services or land use mix in a zone can be measured by the entropy index (Shannon, 1948). Using the [0, 5) min driving zone from home as an example, the diversity of services in this zone ($H_{[0,5)}$) equals:

$$H_{[0,5)} = - \sum_{s=1}^S \rho_s \ln(\rho_s) \quad (4.1)$$

Where ρ_s represents the proportion of service type s within the zone and S is the total number of available service types in this zone. The greater $H_{[0,5)}$ is, the more diverse services a destination has. Similarly we can measure the diversity of services in the [5, 10) min, [10, 15) min, and [15, 20) min driving zones from home, which are indicated by $H_{[5,10)}$, $H_{[10,15)}$, and $H_{[15,20)}$.

The hypotheses about the impacts of the independent variables on the number of daily trips are summarized in Table 4.1. There are two possible arguments about the relationship between accessibility around home and the number of daily trips. The first argument is that all else equal, greater accessibility around home can help reduce the number of daily trips because some of the short-distance trips may be replaced by non-auto travel modes. The second argument is that all else equal, greater accessibility around home can increase the number of daily trips because it induces greater travel demand. For example, if there is a big shopping mall only 5 minutes' drive from home, all else equal, one may visit this mall more often (even just for window shopping or enjoying the facilities there) than the scenario where the mall is farther away.

In addition, we hypothesize that diversity of services around home can reduce the number of trips because more types of services support multi-purpose shopping and comparison shopping. By doing so, one may reduce the number of trips needed by visiting a location with multiple types of services. In addition, higher income may have a positive or negative effect on the number of trips. On the one hand, higher-income families have a greater financial capacity of making more trips. On the other hand, they may have a tighter work schedule and thus have less time for making those trips. In addition, trips are

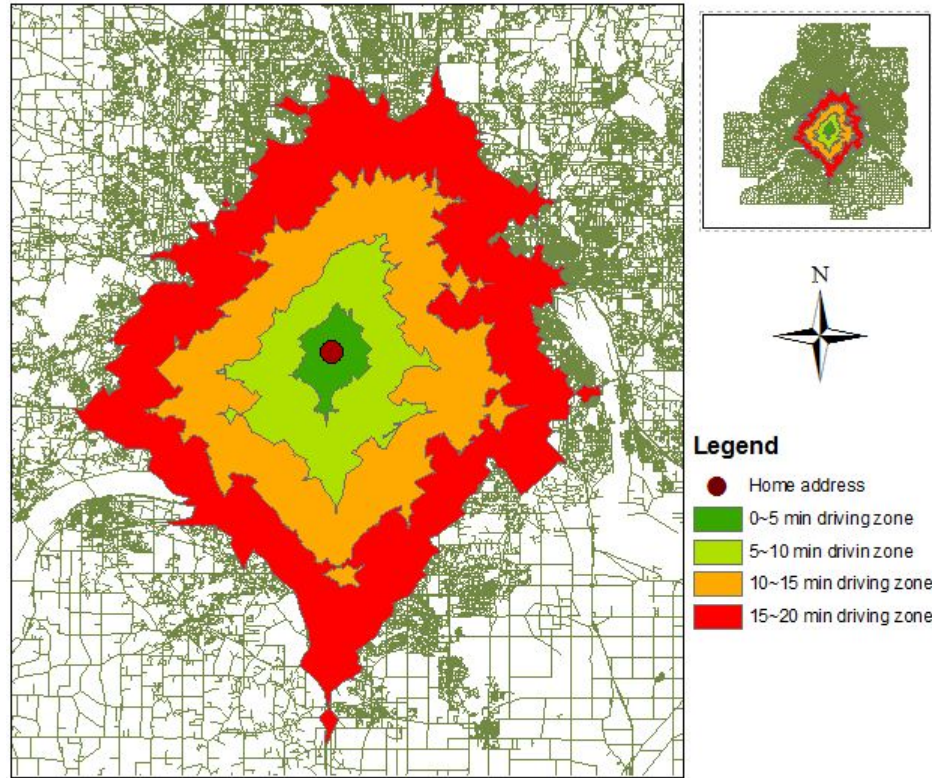


Figure 4.2: The $[0, 5)$ min, $[5, 10)$ min, $[10, 15)$ min and $[15, 20)$ min driving zones around one individual (GPSID 1036)'s home address. The driving zones are created based on the road network with an estimated travel speed on each road link.

hypothetically more likely to happen at weekends than on weekdays.

4.3 Model structures

Five model structures have been used in literature to model trip generation: linear, log linear, Poisson/Negative binomial, ordered logit, and zero-inflated Poisson/negative binomial models. Yet there is a lack of systematic comparison of the goodness of fit of various model structures in modeling vehicle trip generation with repeated observations for each subject. Traditional fixed-effects models do not consider the correlations of dependent variables for the same subject. Therefore given repeated observations for each subject, random-effects models may be a better choice as such models incorporate an extra random-effect component for each subject to control for the heterogeneity. The mixed-

Table 4.1: Hypotheses on the relationships between key independent variables and vehicle trip generation.

Variable	Impacts on vehicle trip generation
Accessibility	+/-
Diversity of services	-
Income	+/-
Weekend	+

effects linear, log linear, Poisson/negative binomial, ordered logit, and zero-inflated Poisson/negative binomial models are described as follows.

For one individual, the mixed-effects linear model can be written as:

$$Y_t = f(b, \Lambda, S, W_t) \quad (4.2)$$

Where Y_t represents the number of trips by an individual on day t . b is the added random effect term for the individual generated from a standard normal distribution with mean zero. Λ indicates a vector of land use measurements around home. W_t is a vector of day-of-week dummy variables and monthly dummy variables.

Similarly, the mixed-effects log linear model for one individual can be expressed as:

$$\ln(Y_t) = f(b, \Lambda, S, W_t) \quad (4.3)$$

In cases where Y_t equals 0, we use a small value 0.01 to replace 0 to make $\ln(Y_t)$ meaningful.

The mixed-effects Poisson model assumes that the conditional mean of the dependent variable is an exponential function of the explanatory variables (including the added random effect) and their coefficients. Consistent with the symbols in [Jang \(2005\)](#), the probability of making Y_t trips can be written as:

$$Pr(Y_t) = \frac{e^{-\lambda_t} \lambda_t^{Y_t}}{Y_t!} \quad (4.4)$$

Where Y_t is the number of trips (0, 1, 2...) and λ_t is the mean parameter of the model which is estimated as:

$$\ln(\lambda_t) = f(b, \Lambda, S, W_t) \quad (4.5)$$

If the conditional variance exceeds the conditional mean, negative binomial model serves as a better fit than the Poisson model. Compared with the Poisson model, $\ln(\lambda_t)$ in the negative binomial model has an extra unobserved heterogeneity term ϵ_t which follows Gamma distribution:

$$\ln(\lambda_t) = f(b, \Lambda, S, W_t, \epsilon_t) \quad (4.6)$$

For the mixed-effects ordered logit model, if the number of daily trips are categorized into Z groups, the utility of making Y_t trips for one individual can be written as:

$$U_t = f(b, \Lambda, S, W_t, \epsilon_t) \quad (4.7)$$

Where ϵ_t is an error term that assumes to follow the logistic distribution (See [Becker and Kennedy \(1992\)](#) for details). While we cannot observe U_t , we can observe the categories of daily trips, which can be represented as:

$$Y_t = \begin{cases} 0 & \text{if } U_t \leq \delta_0 \\ 1 & \text{if } \delta_0 \leq U_t \leq \delta_1 \\ \dots & \\ z & \text{if } \delta_{z-1} \leq U_t \leq \delta_z \\ \dots & \\ Z & \text{if } \delta_{Z-1} \leq U_t \end{cases} \quad (4.8)$$

The probability that one individual makes z trips on day t can be written as:

$$P(Y_t = z) = \Phi(\delta_z - U_t) - \Phi(\delta_{z-1} - U_t) \quad (4.9)$$

Where $\Phi(\cdot)$ represents the standard normal cumulative distribution function. The next step is to select an appropriate number of Z by systematically comparing the mixed-effects logit models with different numbers of categories of trips.

The zero-inflated Poisson/negative binomial model structure aims to model count data

with an excess number of zeros (Lambert, 1992). The model takes two steps. The first step adopts a binary logit function to predict the probability of producing 0 vehicle trips. The second step incorporates the probability of producing more than 0 vehicle trips (which is the complement of the previous result) for the non-zero data using the Poisson/negative binomial structure. In our data, the utility function for estimating 0 trips for one individual can be written as:

$$U(Y_t = 0) = f(a, W_t, \epsilon_t) \quad (4.10)$$

Where ϵ_t is a random-effect term that follows the logistic distribution and W_t is a vector of day-of-week variables. It hypothesized that a subject is more likely to make 0 trips on weekdays than on weekends. a is an extra random effect term for a subject and it follows a normal distribution with mean zero. This term is used to control for the heterogeneity of decisions made by the same subjects.

After predicting the probability of having 0 trips, the remaining trip counts occur with a probability calculated by 1 minus the probability of making 0 trips. And the remaining trip counts presumably follow the Poisson/negative binomial distribution as described before. The utility for a non-zero vehicle trip count ($y_t > 0$) equals:

$$U(Y_t = y_t) = f(b, \Lambda, S) \quad (4.11)$$

Where Λ indicates a vector of land use measurements around home. S is a vector of personal socio-demographic variables. We add another random-effect term b for each subject to control for heterogeneity among observations. Like a , b also presumably follows a normal distribution with mean zero. But the distributions of a and b are set to have different standard deviations in the estimation process.

4.4 Results and analysis

4.4.1 Identifying random effects

To examine whether there exist extra random effects, we plot the residuals versus fitted values to verify homogeneity and a histogram of the residuals for normality based on the regular fixed-effects linear model (Figure 4.3). Figure 4.3-1 shows that residuals do not symmetrically center around 0. There seems minor evidence of the histogram of the residuals being skewed to the right. The problem may be partially due to repeated observations for each individual over the sampling period. To simply observe that effect, we run separate models for each individual, where the independent variable is day of week and the dependent variable is the \ln form of the number of daily trips. The 95 percent confidence intervals for the intercepts and slopes are shown in Figure 4.4. There exist substantial variations in the intercepts among individuals, and there are also apparent individual-to-individual variations in the slopes. It signals the existence of extra random effects in the data. Therefore, mixed-effects models are considered as more appropriate than traditional fixed-effects models.

4.4.2 Testing the number of categories for mixed-effects ordered logit model

In our data, the number of trips ranges from 0 to 17 (16 is missing). Therefore for the mixed-effects ordered logit model, we can test different numbers of categories ranging from 2 to 17 with an increment of 1. The goodness of fit (measured by Nagelkerke R^2) of the mixed-effects ordered logit models with different numbers of categories of trips is shown in Figure 4.5. The results indicate that the models produce similar Nagelkerke R^2 values and similar estimates of the coefficients. Models with categories from 7 to 16 have approximately the same Nagelkerke R^2 value which is slightly higher than the models with fewer categories. In addition, the number of days with more than 5 trips account for less than 3% of all days. It seems reasonable to select 7 categories (0, 1, 2, ..., 5, > 5) by combining more than 6 daily trips as one category.

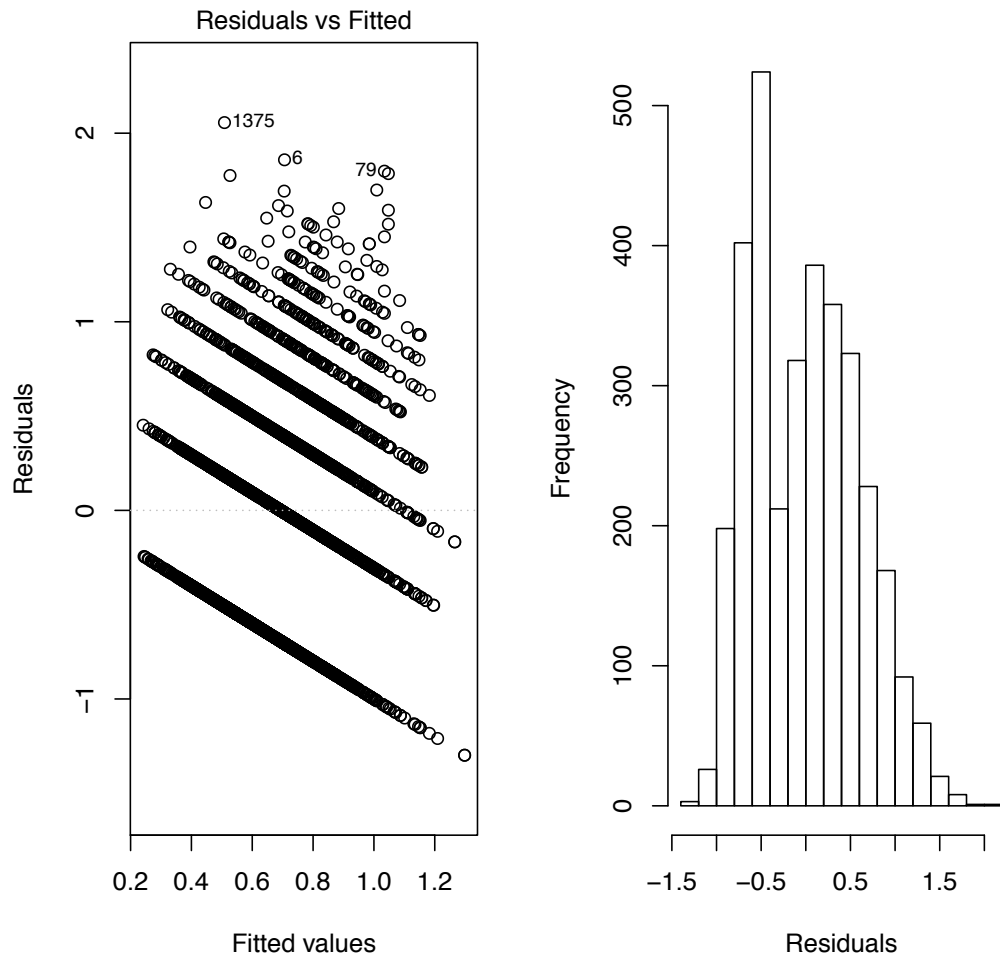


Figure 4.3: Residual plots of the fixed-effects linear model

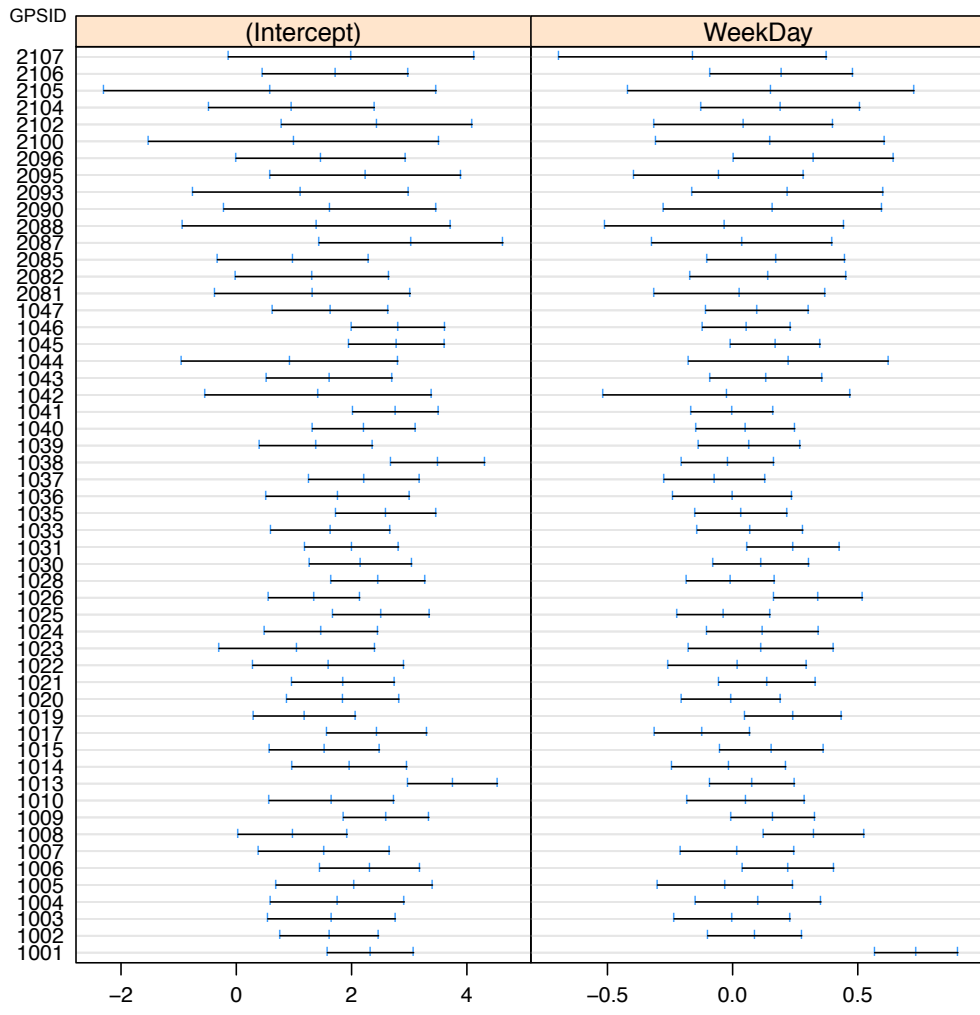


Figure 4.4: 95 percent confidence intervals for the intercepts and slopes of the individual-based regressions of the number of daily trips on the day-of-week variable.

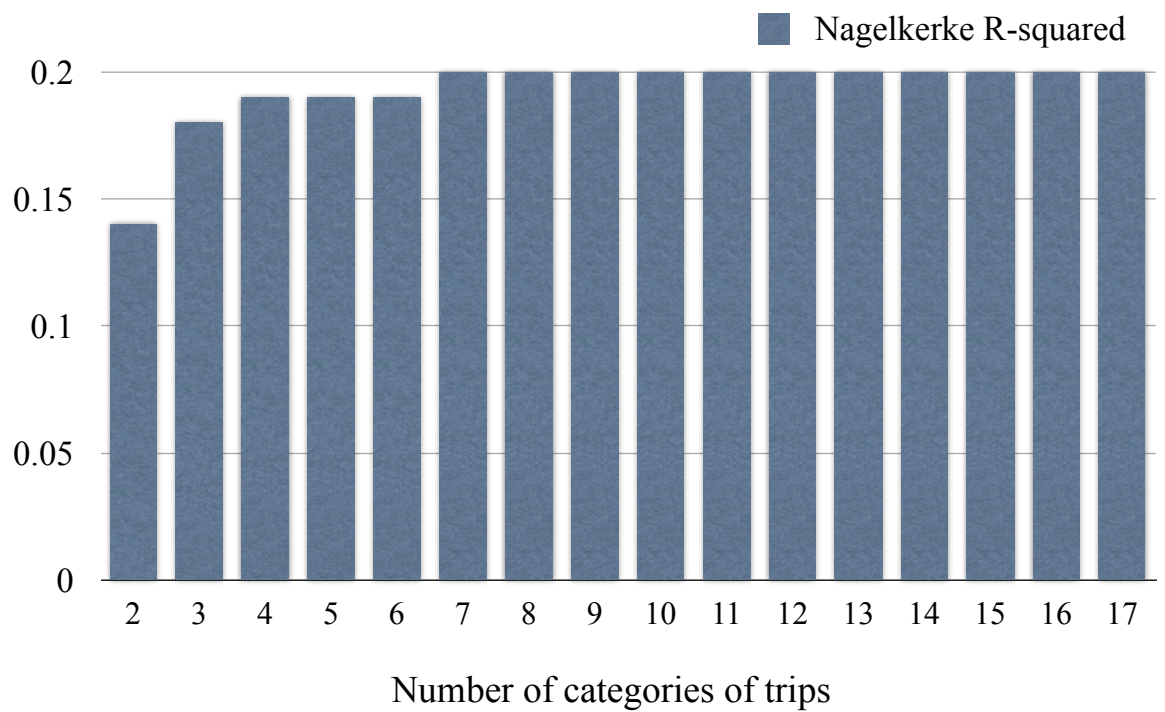


Figure 4.5: Comparison of the goodness of fit for selecting the number of categories of trips for the mixed-effects ordered logit model

4.4.3 Comparing Poisson and negative binomial models

Regarding the choice of the Poisson or negative binomial models, if the mean of the dependent variable equals the variance, then the Poisson model is preferred. But if the variance is greater than the mean, negative binomial would be a better fit. We compare the goodness of fit of the Poisson model and negative binomial model for our data. The test used is the likelihood-ratio Chi-squared test where the null hypothesis is that the dispersion parameter equals 0 (Cameron and Trivedi, 1998). We run the mixed-effects negative binomial model and obtain its log likelihood value ($LL_{nb} = -8567.505$). Then the mixed-effects Poisson model is examined and its log likelihood value is also recorded ($LL_p = -8361.32$). The next step is to calculate $\chi^2 = -2(LL_p - LL_{nb}) = 33.72$. It is statistically significant at the 1% level for one degree of freedom. The large test statistic suggests that the count data are over-dispersed and cannot be sufficiently described by the Poisson distribution. Therefore the mixed-effects negative binomial model seems a better fit than the mixed-effects Poisson model.

4.4.4 Modeling results

Table 4.2 exhibits the results from mixed-effects linear, mixed-effects log linear, mixed-effects negative binomial, and mixed-effects ordered logit, and mixed-effects zero-inflated negative binomial (ZINB) models. All models are estimated using the maximum-likelihood method. For the mixed-effects linear model, a one-unit increase in the independent variable can be interpreted as one more trip. For the mixed-effects log linear model, a one-unit increase in the independent variable can be interpreted as about 1% more trips. For mixed-effects negative binomial, ordered logit, and ZINB models, a one-unit rise in the independent variable can be interpreted as a one-unit increase in the log-odds of the number of trips.

Of the built environment variables, most entropy measures in various zones from home are negative, but only the entropy measures in the [10, 15) min zone and [15, 20) min zone are statistically significant. The results imply that greater diversity of services in these zones supports multi-purpose trip behavior and thus may help lower the number of trips. Interestingly, most accessibility measures are not statistically significant except accessibil-

Table 4.2: Regression results for different mixed-effects model structures (# of obs: 4988)

Variable		Mixed-effects models				
		Linear	Log-linear	Negative binomial	Ordered logit	ZINB
Land use	$A_{[0,5)}$	0.006	0.04	0.03	0.002	-0.02
	$A_{[5,10)}$	-0.18	-0.20	-0.06	-0.22	0.006
	$A_{[10,15)}$	0.16	0.15	0.09	0.27	0.06
	$A_{[15,20)}$	1.13*	1.07*	0.80*	1.24*	0.48*
	$H_{[0,5)}$	-0.34	-0.27	-0.29	-0.27	-0.46
	$H_{[5,10)}$	0.58	0.11	-0.03	0.23	0.62
	$H_{[10,15)}$	-7.55***	-7.53***	-5.85***	-9.52***	-3.10**
	$H_{[15,20)}$	-2.70	-1.82	-2.97**	-3.32	-1.95
Day of week	Tue	0.003	-0.01	-0.01	0.003	0.01
	Wed	0.02	0.02	0.02	0.02	0.03
	Thur	0.05	0.07	0.05	0.09	-0.002
	Fri	0.45***	0.43	0.31***	0.54***	0.24**
	Sat	0.95***	0.73	0.57***	0.98***	0.53***
	Sun	0.13	0.07	0.13***	0.14***	0.15**
Month	Sep	0.28***	0.48***	0.18***	0.47***	0.13***
	Oct	0.18***	0.29***	0.09***	0.29***	0.08***
	Nov	0.24***	0.26***	-0.13***	0.28***	0.12***
Socio-demographics	Age	-0.02*	-0.01	-0.01 ***	-0.02***	-0.01***
	Inc level 2	-0.36***	-0.24	0.20***	-0.32*	-0.18*
	Inc level 3	-0.52	-0.34	-0.27***	-0.50*	-0.27
	Male	0.10	0.19	0.12	0.20	0.04
Goodness of fit	log likelihood	-9636	-9754	-8056	-11033	-8005
	McFadden's R^2	0.05	0.04	0.06	0.05	0.06
	Cox&Snell R^2	0.20	0.15	0.19	0.22	0.20
	Nagelkerke R^2	0.20	0.15	0.19	0.22	0.20

Incllevel 1: < \$100,000 ; Incllevel 2: \$100,000 – \$149,999; Incllevel 3: > \$149,999

*** significant at 0.01; ** significant at 0.05; * significant at 0.1.

ity in the [15, 20) min zone from home. There are three possible reasons for explaining this:

1. This result may be due to its correlation with other land use variables, especially with the entropy measure (diversity of services) in the same zone.
2. The positive coefficient may suggest that more services in the [15, 20) min zone, all else equal, induce a higher travel demand.
3. The above two reasons jointly contribute to the result.

We test several alternative models by including one land use variable at a time. The results show that none of the accessibility measures in any of the zones are statistically significant. Therefore there is not enough evidence to support Reason (2) and Reason (3). Overall, the results from all the models show that land use variables in the [0, 5) and [5, 10) min zones from home do not appear to impact trip generation with sufficient statistical significance. Only the diversity of services in the [10,15) min and [15, 20) min zones can help decrease the number of daily trips. This finding may be due to that most of the subjects live in the suburbs which have relatively low accessibility and low diversity of services. Their major non-work activities are more likely to happen within the [10, 20) min zone from home.

The Pearson correlation test of key land use parameters are conducted (Table 4.4). Accessibility in the [0, 5) min zone is highly correlated with the entropy measure in the same zone (the coefficient equals 0.74). As travel time from home rises, while there still exist positive correlations between accessibility and entropy in the same zone, the level of correlation shrinks. For example, accessibility and entropy in the [15, 20) min zone only equals 0.13. In addition, there exist correlations for the same variable in different zones, and adjacent zones have higher correlation coefficients than zones that are farther away from each other. For instance, the correlation between accessibility in the [0, 5) min zone and [5, 10) min zone equals 0.69, while the correlation between accessibility in the [0, 5) min zone and [10, 15) min zone equals 0.52 and the correlation between accessibility in the [0, 5) min zone and [15, 20) min zone equals 0.46. Similar results can be found for other parameters. Geographical proximity contributes to the similarity of land use.

Regarding day-of-week variables, all else equal, Friday, Saturday, and Sundays are

Table 4.3: The coefficients for predicting the probability of zero trips for the mixed-effects ZINB model (using Monday as the base term)

Variable	Estimate
Sunday	0.19
Tuesday	0.12
Wednesday	0.15
Thursday	-0.40
Friday	-0.49
Saturday	-0.34
St. dev (random term)	0.30

Table 4.4: Correlation of accessibility and diversity of services in [5, 10) min, [10, 15) min and [15, 20) min driving zones around home.

	$A_{[0,5)}$	$A_{[5,10)}$	$A_{[10,15)}$	$A_{[15,20)}$	$H_{[0,5)}$	$H_{[5,10)}$	$H_{[10,15)}$	$H_{[15,20)}$
$A_{[0,5)}$	1.0	0.69	0.52	0.46	0.74	0.40	0.14	0.32
$A_{[5,10)}$		1.0	0.78	0.66	0.50	0.42	0.35	0.24
$A_{[10,15)}$			1.0	0.70	0.29	0.49	0.41	0.21
$A_{[15,20)}$				1.0	0.34	0.47	0.54	0.13
$H_{[0,5)}$					1.0	0.57	-0.03	0.04
$H_{[5,10)}$						1.0	0.16	-0.20
$H_{[10,15)}$							1.0	-0.23
$H_{[15,20)}$								1.0

Note: All estimated coefficients are statistically significant at the 1% level.

associated with more daily trips than other weekdays. Higher income, all else equal, seems to be associated with fewer trips. Limited time budgets for higher-income families may have played a role here. But more information about household structure would be helpful for providing further insights.

Table 4.3 further shows the estimated coefficients for predicting the probability of zero trips from the mixed-effects ZINB model. Thursday, Friday, and Saturday are less likely to have 0 trips than Monday. The individual-specific random term in this equation is assumed to follow a normal distribution with mean zero, and the standard deviation is estimated to be 0.30.

4.4.5 Model fit

As these models have different structures, there is no direct measure for comparing these models' goodness of fit to the classic R^2 in ordinary least squared regressions. It is also invalid to compare the log likelihood values across different model structures. To shed more light on this issue, we calculate several Pseudo- R^2 measures which are based on comparing the log likelihoods of the model with a null model: McFadden's Pseudo- R^2 , Cox&Snell R^2 , and Nagelkerke R^2 . As shown in Table 4.2, the five models have close Pseudo- R^2 values, though they are not ranked quite the same in terms of these Pseudo- R^2 measures. Overall, the mixed-effects ordered logistic model seems to be in the first tier because it displays higher goodness of fitness than other models in most measures. The mixed-effects ZINB, NB, and linear models are about in the second tier. The log linear model has the lowest Pseudo- R^2 estimates for all measures. Note that since all the models have close Pseudo- R^2 values, further investigation is needed to gain more insights about the models' predictability.

4.5 Predictive results

The five estimated mixed-effects models (linear, log linear, negative binomial, ordered logit, and ZINB) are employed to predict trip generation patterns at the macroscopic level. In the mixed-effects ordered logit model, for every trip, we calculate the probability of choosing each trip level (0, 1, 2, ..., 5, >5). Based on the probability of each trip level, we randomly generate a choice for every trip, based on which the predicted trip counts for all trip levels can be calculated. This process is repeated for 100 times, and the average trip frequency for each trip level is computed. For the other four models, we use the 0.5 cutoff points as the threshold. For example, if the predicted value is below 0.5, it is considered as 0 trips. If the predicted value is within [0.5, 1.0), it is considered as 1 trip. If the predicted value is within [1.5, 2.5), it is considered as 2 trips.

Figure 4.6 shows the observed and predicted trip counts. Several observations can be made:

1. The mixed-effects ordered logit model outperforms other models by matching the

actual observations the best for all numbers of daily trips.

2. The mixed-effects log linear model over-predicts the number of days with 0 or 1 trips while under-predicting the number of days with more than 1 trip.
3. The mixed-effects log linear, NB, and ZINB models over-estimate the number of days with 1 or 2 trips but considerably under-estimate the number of days with 0 trips.

The mean absolute error (MAE) measure and the mean absolute percentage error (MAPE) for each model are calculated. Mathematically, the equation for calculating MAE can be written as:

$$MAE = \frac{1}{N} \sum_{n=1}^N |y_n - f_n| \quad (4.12)$$

And the equation for computing MAPE can be expressed as:

$$MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{y_n - f_n}{y_n} \right| \times 100\% \quad (4.13)$$

Where y_n is the actual value for the n th observation in the data and f_n is the predicted value. N is the total number of observations in the data. The MAE is an average of the absolute errors and the MAPE is an average of the percentages of the absolute errors. The smaller MAE/MAPE is, the more accurate the predicted values are. The five models' MAE and MAPE values are reported in Table 4.5. The mixed-effects ordered logit model obviously produces the lowest MAE and MAPE values, which supports visual observation from Figure 4.6. Interestingly, based on the MAE measure, the mixed-effects log linear model ranks the second; nevertheless, its rank drops to the lowest according to the MAPE measure. The mixed-effects ZINB model ranks higher than the mixed-effects NB model for both measures.

4.6 Conclusions

Trip generation is typically modeled with fixed-effects models. For data sets such as the GPS travel data that feature repetitive observations among individuals, traditional fixed-effects models do not fit well. Furthermore, there is a lack of research on comparing various

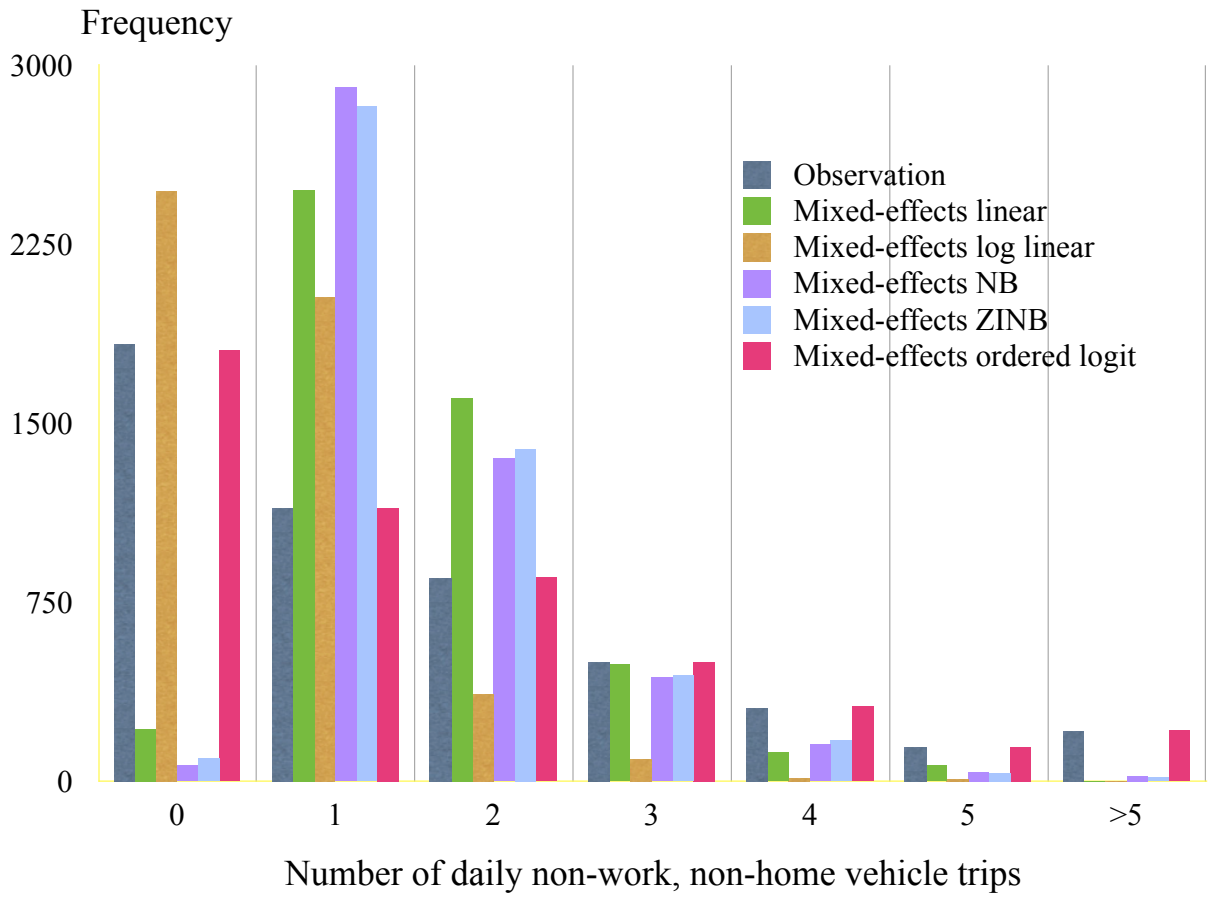


Figure 4.6: Predicted and observed non-work, non-home vehicle trip patterns.

Table 4.5: The models' mean average error (MAE) and mean average percentage error (MAPE)

Model	MAE	MAPE (%)
Mixed-effects ordered logit	6.71	1
Mixed-effects log linear	434.86	77
Mixed-effects linear	597.145	73
Mixed-effects ZINB	635.43	75
Mixed-effects negative binomial	647.71	76

alternative model structures on modeling count data with repetitive observations. To this end, this research conducts a comparative study of five mixed-effects model structures based on the in-vehicle GPS data. This research uses the parcel-level land use data around home to examine the relationship between land use and non-work vehicle trip generation.

The key findings are:

1. The mixed-effects ordered logistic model produces the highest goodness of fit of all the models tested. This finding is consistent with [Lim and Srinivasan \(2011\)](#) though the models tested in that paper are all fixed-effects models. The results indicate that traditional Poisson/NB models may not be the best choice for modeling trip generation using the GPS data.
2. The accessibility measures in the [0, 5) min, [5, 10) min, [10, 15) min, and [15, 20) min driving zones from home are not found to influence the generation of non-work vehicle trips for our data with sufficient statistical significance. Based on our hypothesis, accessibility may both induce and dampen the generation of non-work vehicle trips. It is likely that both positive and negative effects may have played a role here. The correlation among accessibility measures in adjacent zones also influence the estimates.
3. The diversity of services in the [10, 15) min, and [15, 20) min driving zones from home displays depressive effect on the number of non-work, non-home vehicle trips, a sign of trip chaining behavior. Most of the subjects in this study live in the suburbs. Our findings reflect non-work driving behavior for individuals living in the suburbs where land use in the immediate vicinity of home is frequently less diverse than zones which are more than 10 minutes away.

This research can be expanded by further examining vehicle trip generation for various non-work trip purposes, as the effects of land use on trip generation for different purposes may well be different. It is of interest to investigate such effects at the microscopic level.

Chapter 5

Modeling single-destination choice

This chapter aims to study the impacts of land use and transportation network structure on home-based non-work, single-destination choice. The contributions of this chapter include:

- Proposing a new method to form choice sets for home-based, non-work destination choice.
- Proposing a procedure to decide the choice set size for destination choice.
- Examining the impacts of land use and route-specific network structure on destination choice based on the in-vehicle GPS data.

This chapter focuses on home-based non-work, single-destination trip chains (i.e., home → destination → home) for the simplicity of illustrating our approach. The rest of this chapter is organized as follows. The modeling procedure is proposed in Section 5.1. Section 5.2 introduces the definition of destinations used in this study. The independent variables are summarized in Section 5.3. The survival analysis-based choice set formation method is illustrated in Sections 5.4 and 5.5. Section 5.7 describes the model and evaluation criteria. Section 5.8 presents and analyzes the research findings. Section 5.9 concludes this chapter.

5.1 The modeling procedure

The home-based non-work, single-destination choice process is modeled with three steps. The first step is to define destinations for individuals. The second step is to decide a feasible choice set from all destinations in the Minneapolis-St. Paul Metropolitan Area. The third step is to select a destination from the choice set. It is assumed that vehicle is the default mode choice in order to fit the context of the in-vehicle GPS travel data. An individual's trip starts from home (the starting point is within 800 meters from home) and returns home after one destination is visited. A destination needs to be at least 1000 meters away from work to be qualified as a non-work destination. There are multiple non-work destinations in the metropolitan area for selection.

In this modeling process, there are four key questions to address:

1. Defining destinations. What is a proper way to define destinations for destination choice?
2. Choice set formation. How should choice sets be properly formed based on the land use data?
3. Choice set size. How many choices should be incorporated into a choice set?
4. Model specification. How should the model be specified to fit the data so as to produce consistent and unbiased estimates?

Our solutions for these questions are described in Sections [5.4](#), [5.5](#), and [5.7](#).

5.2 Definition of destinations

In modeling home-based non-work, single-destination choice, the first step is to define destinations. In previous studies, the definitions range from counties, cities, traffic analysis zones, parcel-based locations, to store-based destinations. Since this study focuses on non-work trips (e.g., shopping, recreational, visiting friends), we prefer finer granularity of locations to larger granularity because finer granularity can provide more insights about microscopic land use.

In this research, the centroids of Census blocks data are used to define destinations for three reasons:

1. The Census block data provide better precision of locations than other larger scale definitions. The block-level data are the finest geographical definition of locations in the US Census. In the 2010 US Census, the Twin Cities have 16851 Census blocks with at least one establishment, far more than 1165 traffic analysis zones, 182 cities, and 7 counties in the metropolitan area. In addition, even though we do know which store one visited, we can measure the land use around a destination.
2. The Census block-based definition of destinations creates more precise travel paths once mapped to the road network data. The shortest travel paths are created with the ArcGIS Network Analyst tool which locates the centroid of a Census block to its nearest road. The more granular a destination is, the more accurately travel time can be calculated.
3. In addition to the 2010 business data at the US Census block level, we also have the 2005 business data at the parcel level. The 2010 business data at the US Census block level are preferred to the 2005 parcel-level data because business data at the 2010 US Census block level are more recent and closer to the year where the GPS data were collected. Given that an individual may visit multiple stores in a block or in adjacent blocks, we create a walking area around each destination to measure a destination's land use. The details are depicted in Section 5.3.

5.3 Independent variables

The independent variables used in this study include land use measures, transportation network measures, axis of travel measures, and the interactions between socio-demographics and land use/transportation network measures.

5.3.1 Land use measures

The key land use measures include accessibility and diversity of services (land use mix). In the literature, there are several accessibility measures: cumulative opportunities mea-

sure, gravity-based accessibility measure, and random utility-based measure. They offer different trade-offs between simplicity and the sophistication with which the activities and transportation system are characterized (Handy and Clifton, 2001). In this research we are interested in the accessibility of walking after parking one's vehicle at a destination. After leaving one's car, one might walk to multiple stores and might not even visit the store closest to the parking spot due to parking constraints. If that happens, it is less justified to employ the gravity model based on the distance/time impedance function. Considering the characteristics of shopping and for the simplicity of measurement, it is decided to use cumulative opportunities (A_k) to measure accessibility at destination k . The empirical tests reveal that its \ln form produces greater goodness of fit of the model. Therefore $\ln(A_k)$ is adopted as an independent variable. All else equal, a destination surrounded by more stores to visit presumably provides more opportunities for comparison/multi-purpose shopping, and thus may be more likely to be selected. It is therefore hypothesized that greater accessibility at a destination enhances the attractiveness of the destination.

The next question is to define the size of the walking area. Burke and Brown (2007) found that the 85th percentile of the walking distance to a shop is 1.24 km. If we assume the average walking speed is 5.44 km/hr (3.40 mi/hr) (Krizek et al., 2009), the walking time is around 15 minutes. The walking time considers the fact that one has to walk a little farther when there is not enough parking for the building one desires to visit. Therefore, we create a 15-min walking area around the centroid of a block where lies the parking destination. All blocks whose centroids are within this area are considered as the destination's walking area. The cumulative opportunities measurement (A_k) is calculated as the total number of establishments in the destination's 15-min walking area.

The diversity of services or land use mix at a destination (indicated by k) is typically measured by the entropy index (Shannon, 1948) which can be written as:

$$H_k = - \sum_{v=1}^V \rho_{kv} \ln(\rho_{kv}) \quad (5.1)$$

Where ρ_{kv} is the proportion of service type v in destination k 's walking area. The service type of a store is defined by the 6-digit North American Industry Classification System (NAICS) code. V is the total number of services in the destination's walking area.

The greater H_k is, the more diverse services a destination has. All else equal, a destination with greater entropy indicates greater diversity of services, which supports multi-purpose shopping and reduces the average travel time it takes to finish per task compared with making several single-destination trips. It is therefore hypothesized that greater diversity of services, all else equal, is associated with greater attractiveness of a destination.

Our further analysis shows that the entropy index and accessibility (ln form) at a destination are highly correlated, as the Pearson correlation coefficient equals 0.94. Therefore we desire to modify the traditional entropy index in order to obtain less biased estimates when incorporating the two measures in the model. We consider both dividing the traditional entropy index divided by A_k and dividing it by $\ln(A_k)$ and further examine their correlations with accessibility. The correlation between the traditional entropy index divided by A_k and accessibility ($\ln(A_k)$) equals -0.79, and the correlation between the traditional entropy measure divided by $\ln(A_k)$ and accessibility ($\ln(A_k)$) equals -0.22. It is decided that the diversity of services is defined as the traditional entropy index divided by $\ln(A_k)$ because it has a weaker correlation with accessibility and still reflects the characteristics of diversity of services. Mathematically the diversity of services measure can be written as:

$$H_k = \begin{cases} 0 & \text{if } A_k = 0 \text{ or } A_k = 1 \\ -\frac{\sum_{v=1}^V \rho_{kv} \ln(\rho_{kv})}{\ln(A_k)} & \text{if } A_k > 0 \end{cases} \quad (5.2)$$

Several simple examples of the diversity of services are shown in Figure 5.1. The natural logarithm form of H_k is used in the model for comparing its elasticity of the attractiveness of a destination with other independent variables.

5.3.2 Transportation network measures

The road network measures used in this study include speed discontinuity (Levinson and El-Geneidy, 2009) and turn index.

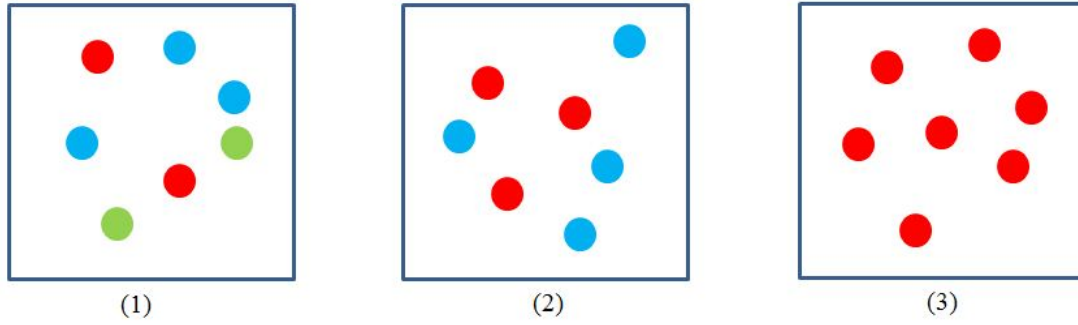


Figure 5.1: Examples of entropy measures with each color representing one type of service. In Example (1), there are three types of services and the diversity of services equals $1.08/\ln 7 = 0.56$. In Example (2) there are two types of services and the diversity of services equals $0.68/\ln 7 = 0.35$. In Example (3) there is only one type of service and the diversity of services equals 0.

Speed discontinuity

Speed discontinuity, first proposed and applied in [Xie and Levinson \(2007\)](#) and [Parthasarathi et al. \(2012\)](#), was described as the changes of speed along the fastest path between an origin and a destination divided by the length of this route. In this study travel time is used as an independent variable. In order to reduce the correlation with travel time, speed discontinuity in this study is defined as the changes of speed along the fastest path between an origin and a destination divided by trip travel time. The basic form of speed discontinuity ψ_k of the fastest route from home to destination k is written as:

$$\psi_k = \frac{\sum(|v_{q+1} - v_q|)}{T_k} \quad (5.3)$$

Where v_q is the travel speed on road link q . $|v_q - v_{q+1}|$ indicates the absolute value of the speed difference on two consecutive links q and $q + 1$. $\sum |v_q - v_{q+1}|$ measures the sum of the absolute value of the changes of speed along a route. It is further divided by the travel time of the fastest route from home to destination k to calculate ψ_k . The speed discontinuity measure has a relatively wide range and the \ln form gives a higher goodness of fit for the model than Equation (5.3). Therefore, ψ_k used as an independent variable in the model is defined as:

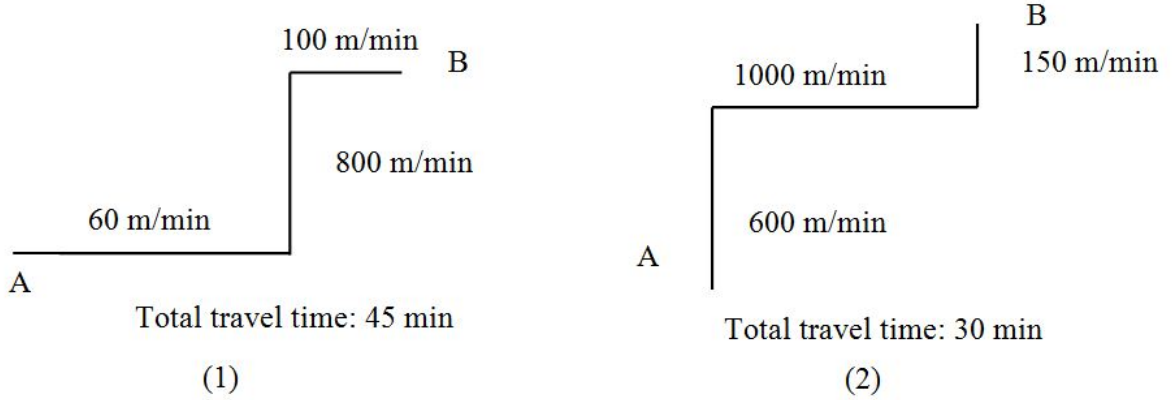


Figure 5.2: Two examples of calculating speed discontinuity.

The route in Example (1) consists of three road links. The speed discontinuity of the route can be calculated as $\ln \frac{(|800-60|+|100-800|)}{45} = 3.46$. The route in Example (2) is also comprised of three road links. The speed discontinuity of the route equals $\ln \frac{(|1000-600|+|150-1000|)}{30} = 3.73$. The greater speed discontinuity is, the more discontinuous the travel speed on the route is per unit time.

$$\psi_k = \begin{cases} \ln \frac{0.5}{T_k} & \text{if } \sum (|v_q - v_{q+1}|) = 0 \\ \ln \frac{\sum (|v_q - v_{q+1}|)}{T_k} & \text{if } \sum (|v_q - v_{q+1}|) > 0 \end{cases} \quad (5.4)$$

When $\sum (|v_q - v_{q+1}|) = 0$, we use the midpoint of 0 and 1 to replace 0 to make the definition meaningful. Two examples of measuring speed discontinuity are illustrated in Figure 5.2.

We argue that speed discontinuity is an index for measuring the perception of a destination's reachability. A trip with greater speed discontinuity is considered less comfortable and reduces the perceived reachability which reduces the desire of travel. It is therefore hypothesized that greater speed discontinuity on the route dampens the attractiveness of the trip's destination.

Turn index

We propose a new measure named turn index (ϑ_k) to quantify the perception of a destination's reachability. It measures the number of turns a drivers needs to make from home to

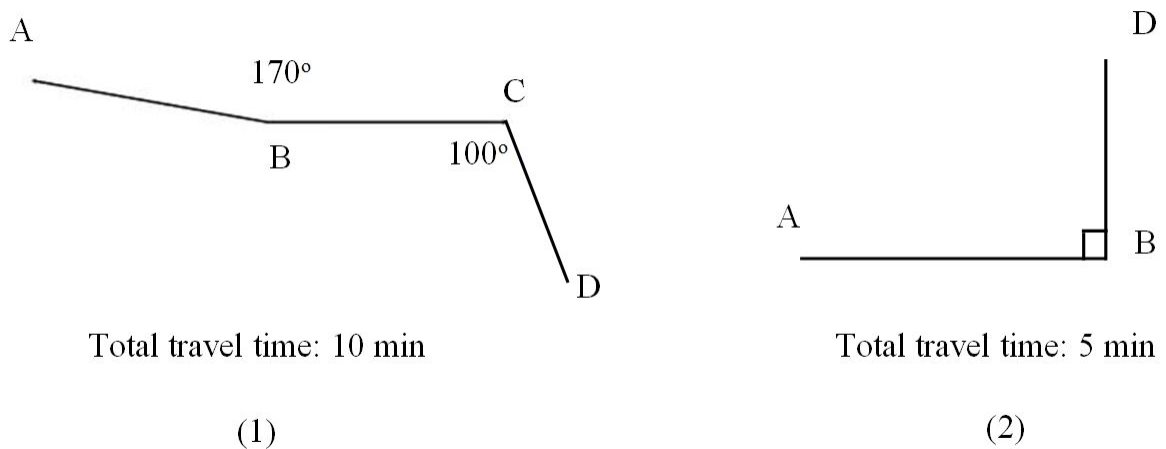


Figure 5.3: Two examples of calculating turn index.

The route in (1) consists of three links. The angle between link AB and BC equals 170 degrees; therefore there is no turning maneuver. The angle between link BC and CD equals 100 degrees; thus a driver needs to make a turn to go from BC to CD. The total number of turns equals 1. Given that the total travel time equals 10 minutes, the turn index of the route equals $\ln(1/10) = -2.3$. In Example (2), the route consists of two links. The angle between link AB and link BD equals 90 degrees; thus there is one turn between the two links. Given that the total travel time equals 5 minutes, the turn index of the route equals $\ln(1/5) = -1.6$. The greater this value is, the more turns per unit travel time a route requires.

destination. If the acute angle between every two connected road links is between 170 degrees (inclusive) and 180 degrees (inclusive), a driver is considered as not having to make any turning maneuver to transition from one link to the other; otherwise, a driver is considered as having to make a turn. Turn index (ϑ_k) is calculated as the cumulative number of turns one drivers needs to make on a route divided by the total travel time (in order to reduce the correlation with travel time). Our further test reveals that its \ln form also produces a higher log likelihood value for the model. Therefore, turn index (ϑ_k) used in this research are defined as:

$$\vartheta_k = \begin{cases} \ln \frac{0.5}{T_k} & \text{if } \Gamma_k = 0 \\ \ln \frac{\Gamma_k}{T_k} & \text{if } \Gamma_k > 0 \end{cases} \quad (5.5)$$

Where Γ_k is the total number of turns on the route to visit destination k . Two simple examples of calculating turn index are shown in Figure 5.3. A greater turn index suggests more turns one needs to make per unit time, which makes a trip less desirable. It is therefore hypothesized that a greater turn index reduces the convenience of driving on the route, and thus lowers the attractiveness of a destination.

5.3.3 Axis of travel

The measures on the axis of travel include travel time between destination and work and travel time between destination and the nearest downtown.

Travel time between destination and work

For each individual, we measure the fastest-path travel time between destination k and workplace, which is represented by $T_{w,k}$ (the symbol w represents work). This measure indicates one's familiarity with the destination. One may be more familiar with destinations adjacent to work, and therefore may be inclined to select these destinations. We hypothesize that all else equal, a non-work destination closer to work is more likely to be selected due to a person's greater familiarity with the destination.

Travel time between destination and the nearest downtown

Another axis of travel measure is the fastest-path travel time between destination k and the nearest downtown, indicated by $T_{d,k}$ (the symbol d represents downtown). We calculate the travel time between the destination and the center of downtown Minneapolis (IDS Center) and the travel time between the destination and the center of downtown St. Paul (Wells Fargo Place), and the smaller time is used for $T_{d,k}$.

This measure implies one's consideration of greater parking constraints, narrower streets, and more traffic lights which are common in the downtown area. It is hypothesized that all else equal, a destination closer to the nearest downtown is less likely to be selected because of these nuisances.

5.4 Choice set formation

The choice set formation problem concerns how to form choice sets based on all destinations in the Twin Cities Metropolitan Area.

We propose a new method of choice set formation which combines survival analysis and random sampling. Survival analysis was originally applied to studying the hazards of deaths. In recent years, several studies used the hazard-based analysis to calculate average work distance for housing location choice set (Rashidi et al., 2012), the length of stay in golf tourism (Barros et al., 2010), and the deterministic distance constraint for residential destination choice (Zolfaghari et al., 2012). To illustrate the idea, an analogy is used (Figure 5.4). For example, in a bee nest there are more than a thousand bees. Each bee was born at a random time and has a random life span. As shown in Figure 5.4-(1), a bee was born at a certain time and survived for 25 days; another bee was born at a different time and died after 38 days. Given a multitude of bees, a question rises: how likely is a bee to survive for a certain period of time? Similarly, a person makes a number of trips at different times and on different days, and each trip's travel time is different. For example, one trip on a given day began at certain time and lasted for 21 minutes; another trips started at a different time on a different day and "survived" for 59 minutes (Figure 5.4-(2)). We are interested in how likely a trip will happen given some travel time and the destination's

land use characteristics. It is hypothesized that (1) all else equal, a trip with shorter travel time is more likely to happen, and that (2) all else equal, a trip with greater accessibility at the destination is more likely to be made.

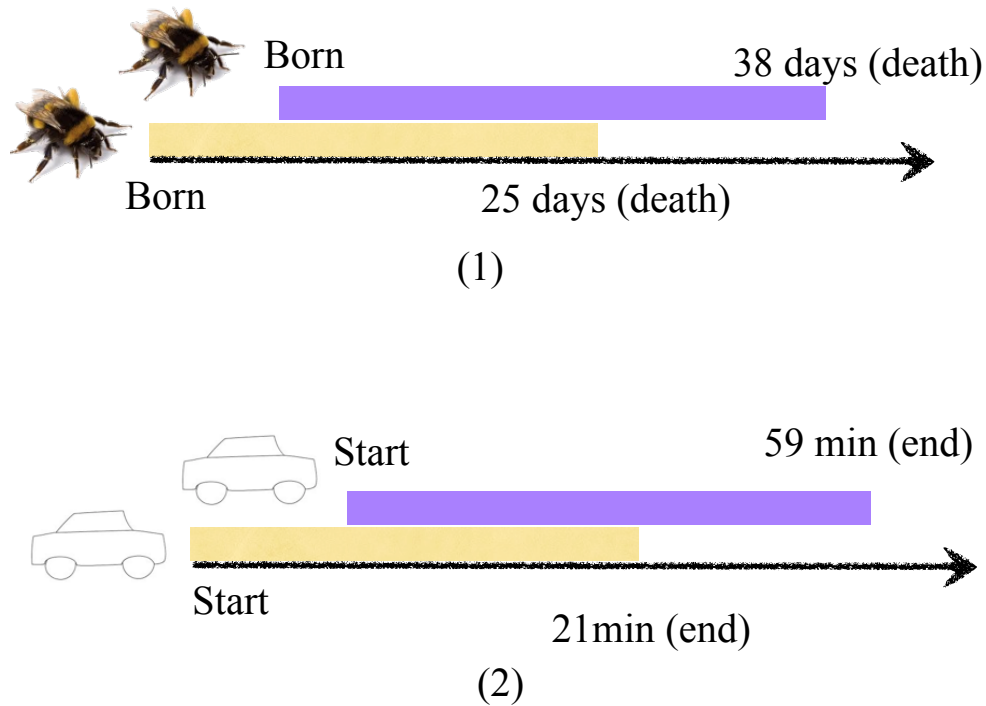


Figure 5.4: The analogy for using survival analysis to predict the probability of making a trip.

The survival model used in this research aims to produce the selection probability for each possible destination based on distance and accessibility. Although we lack information about individuals’ preferences of destinations, our intuition tells us that travel time is an important factor in destination choice. All else equal, a person is more likely to consider a closer destination. If the probability distribution function to visit various destinations for a traveler can be formulated, we can estimate the “importance” of a destination to the traveler by measuring the probability of visiting a destination. This way of understanding matches the purpose of survival analysis which is used to model the probability of the occurrence of events for longitudinal data. An overview of our approach is displayed in Figure 5.5.

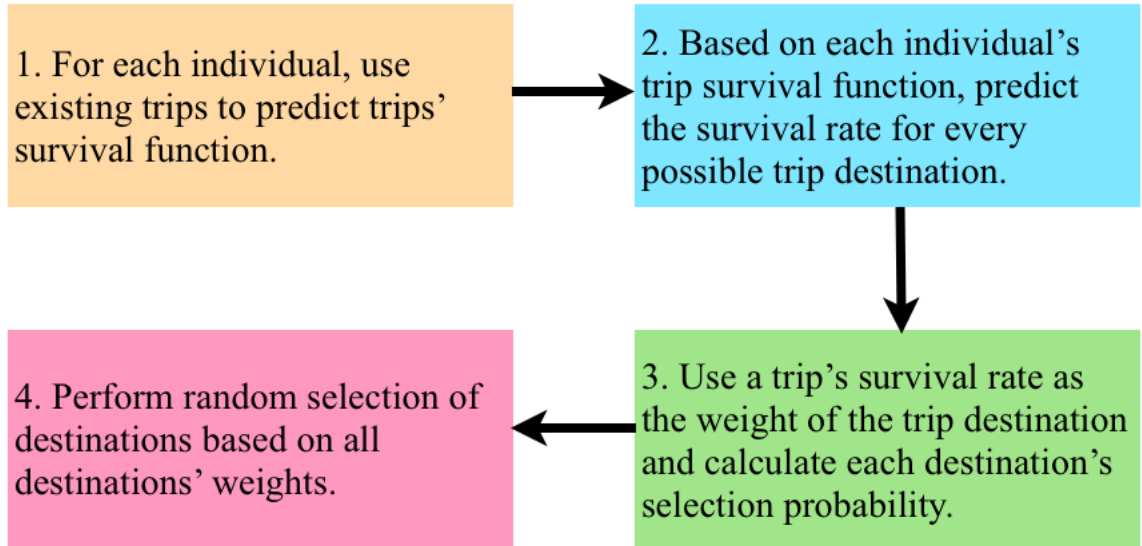


Figure 5.5: The procedure to form choice sets which combines survival analysis and random sampling.

The generic form of survival function given the duration of time (t) can be written as:

$$\Omega(t) = P(t > T) \tag{5.6}$$

Where T is a random variable denoting the time of “death” (the ending of a trip). $\Omega(t)$ indicates the probability that a trip will go beyond a certain travel time T .

Another function related with the survival function is the hazard function. Hazard function, indicated by $\lambda(t)$, represents the event rate at time t conditional on survival until ($T \leq t$).

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = -\frac{d \log \Omega(t)}{dt} \tag{5.7}$$

Integrating both sides of Equation (5.7) gives the survival function in terms of the hazard function (Allison, 2010):

$$\Omega(t) = \exp\left[-\int_0^t \lambda(t) dt\right] \tag{5.8}$$

There are two widely used models in survival analysis: the proportional hazard model

(or called the Cox model) (Cox, 1959) and the accelerated failure time (AFT) model (Cox and Oakes, 1984). The proportional hazard model assumes that the effect of a covariate is the multiplication of the hazard and a constant. The AFT model assumes that the effect of a covariate is the multiplication of the predicted event time and a constant. The greatest advantage of this AFT is that the probability distribution of the survival time (in our case travel time) can be formulated based on which we can calculate the probability of making a trip to a destination given the trip's travel time. Therefore, the AFT model is adopted to model the probability of selecting a destination.

For one individual, let T_k be the travel time of the fastest route from home to destination k . $\ln(T_k)$ is the \ln form of accessibility at destination k . It can be written as:

$$\ln(T_k) = \beta_0 + \beta_1 \ln(A_k) + \sigma \epsilon_k \quad (5.9)$$

Where ϵ_k is a random error term. β_0 and σ are parameters to be estimated. Note that if $\sigma = 1$ and $\ln(T_k)$ follows a normal distribution, the model is the same as the ordinary linear model.

The hazard function is a useful tool for describing the probability distribution for the time of event occurrence (Allison, 2010). The simplest function is that the hazard is constant over time ($h(t) = \lambda$), meaning that during any period of time with a fixed length, the expected number of event occurrences is the same. Then its survival function follows exponential distribution. If the natural logarithm of hazard presumably equals $h(t) = \mu + \alpha \log(t)$, the time of event occurrence is said to follow the Weibull distribution. Other distributions for survival analysis include log-normal distribution, log-logistic distribution, and the Gamma distribution. The exponential, Weibull, and log-normal distributions are special cases of the generalized Gamma model. In addition, the generalized Gamma model can also take a U shape or a bathtub shape. The generalized Gamma model has been found to fit the human mortality data (Allison, 2010).

The next step is to select an appropriate distribution function for $\ln(T_k)$. The tested distribution functions include: Weibull, log-normal, log-logistic, exponential, and the Gamma model. We need to test which distribution function is the best fit for each individual's trips by comparing the models' log likelihood values. We separately estimate the probability

Table 5.1: A comparison of log likelihoods of different distributions of $\ln(T_k)$ for a single subject with GPSID 1019.

Distribution of T_k	Distribution of ϵ_k	log likelihood	Nagelkerke R^2
Gamma	Log-gamma	-45.20	0.58
Log-logistic	Logistic	-48.69	0.55
Log-normal	normal	-48.13	0.54
Weibull	Extreme value (2 parameters)	-55.90	0.48
Exponential	Extreme value (1 parameter)	-74.63	0.28

density function of $\ln(T_k)$ for each individual. To illustrate the distribution function selection process, an individual with GPSID 1019 is used as an example.

Table 5.1 compares the log likelihoods of different distributions of $\ln(T_k)$ for an individual with GPSID 1019. The Gamma distribution produces the largest log likelihood, which suggests the best fit among the candidates. The next step is to test whether the differences of the log likelihoods are statistically significant by performing the log likelihood ratio test. The null hypothesis is that the log likelihood of another model equals the log likelihood of the Gamma model; the alternative hypothesis is that the log likelihood of another model is smaller than the log likelihood of the Gamma model.

To compare the goodness of fit of different distribution functions, the test statistic used is defined as twice the difference of two log likelihood values (i.e., $-2 \ln(\text{likelihood for another model}) + 2 \ln(\text{likelihood for the Gamma model})$). The value of the test statistic is later compared with Chi-squared distribution with $df = 1$ at a level of significance of 0.01. In all these tests, we reject the null hypothesis that the log likelihood of another model equals the log likelihood of the Gamma model. Therefore the Gamma distribution is chosen to fit the distribution of travel time for the individual with GPSID 1019. The generalized Gamma distribution also produces a Pseudo- R^2 of 0.58, which shows satisfactory goodness of fit compared with other distribution functions. Based on the Gamma distribution function and AFT model, the coefficient of $\ln(A_k)$ is estimated. Given the estimated probability density function, we can predict the survival probabilities for trips to all destinations (Allison, 2010).

Further, it is of interest to investigate which distribution best fits the travel time of non-work trips made by all subjects in the GPS data. Because there are repeated observations

Table 5.2: A comparison of different distributions of $\ln(T_k)$ for all subjects

Distribution of T_k	Distribution of ϵ_k	log likelihood
Gamma	Log-gamma	-123.97
Log-normal	normal	-124.19
Log-logistic	Logistic	-124.69
Weibull	Extreme value (2 parameters)	-128.77
Exponential	Extreme value (1 parameter)	-142.10

for each individual, there exist correlations among trips made by the same person. It is necessary to control for the extra random effects in the survival analysis. However, there is no existing function in the SAS Program which can test different distributions for AFT models while controlling for extra random effects. Therefore it is decided to randomly select one trip for every subject, and the survival analysis is applied to analyze the smaller data set. To ensure the consistency of the results, we perform the random selection 10 times and compare the results. All results reveal that the Gamma distribution produces the highest log likelihood value, following which are the log-normal distribution, log-logistic distribution, Weibull distribution, and exponential distribution. Table 5.2 shows the result of one smaller choice set. It is recommended to use Gamma distribution to model all individuals' trips for survival analysis.

The remaining question is how to form a choice set. We first reject several existing methods. First, the deterministic boundary setting of the selection area is not adopted because we do not have specific data to help define individuals' selection boundary. Second, the stratified sampling approach is rejected because our predicted probabilities have implied the selection weight for each destination and we do not want to add a new parameter (the number of strata) to the selection process. The random sampling of all destinations in the Twin Cities Metropolitan Area is selected because it is simple to use and it gives every destination an opportunity to be considered, as we lack more information about an individual's selection boundary. We integrate it with the survival analysis so that the random selection is based on the estimated weights of destinations. If a destination has a higher survival probability, it carries a heavier weight (i.e., a higher chance) to be selected into the choice set.

5.5 Choice set size

After the method of constructing choice sets is determined, a key question is to decide the choice set size M . A large number of destinations (Census blocks) in the Twin Cities Metropolitan Area make it computationally difficult to include all destinations in a choice set. But too small choice sets can result in inconsistent estimates (Auld and Mohammadian, 2011). It is therefore necessary to test different sizes of choice set to decide an appropriate choice set size needed for this research. In this study we propose a systematic method to test choice set size based on the weighted RMSE value of each model.

The traditional RMSE value is defined as:

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\hat{\kappa}_n - \kappa_n)^2}{N}} \quad (5.10)$$

Where $\hat{\kappa}_n$ is the predicted probability for the n th observation in a data set. κ_n is a binary dependent variable which equals 1 if the destination in observation is visited and 0 otherwise. N is the total number of observations in the data set. The smaller the RMSE is, a better fit the model is claimed to be.

The traditional RMSE has one defect. If we increase the choice set size by adding less attractive destinations (such as very far destinations), the RMSE value may be well likely to decline because such destinations has low selection probabilities anyway. Nevertheless, it does not necessarily mean that the model's actual predictability is enhanced. To control for this situation, we first separately measure the RMSE for chosen destinations and RMSE for non-chosen destinations in choice sets.

If there are N_1 actually chosen destinations in the data set, the RMSE of actual destinations can be written as:

$$RMSE_{chosen} = \sqrt{\frac{\sum_{n=1}^{N_1} (\hat{\kappa}_{chosen,n} - \kappa_{chosen,n})^2}{N_1}} \quad (5.11)$$

If there are N_2 unchosen destinations in the data set, the RMSE of all non-chosen destinations can be written as:

$$RMSE_{unchosen} = \sqrt{\frac{\sum_{n=1}^{N_2} (\hat{\kappa}_{unchosen,n} - \kappa_{unchosen,n})^2}{N_2}} \quad (5.12)$$

This function better balances the accuracy of predicting chosen destinations and the accuracy of predicting non-chosen destinations. The RMSE of the model is defined as the average of $RMSE_{chosen}$ and $RMSE_{unchosen}$. In other words, we assign 50% weight to $RMSE_{chosen}$ and the other 50% weight to $RMSE_{unchosen}$. This definition reduces the effects of having more undesirable destinations in a choice set on RMSE.

$$RMSE_{model} = p \cdot RMSE_{chosen} + (1 - p) \cdot RMSE_{unchosen} \quad (5.13)$$

Where $p = 0.5$. The choice set sizes tested range from 10 to 200, with an increment of 10. Even larger sizes such as 500, 1000, 2000, 5000, and 10000 are also investigated. The RMSE values of models with different choice set sizes are further compared to decide an appropriate choice set size.

5.6 Checking repeated destinations

It is important to check whether there are many repeated destinations visited by the same person. In the modeling destinations, if there exist repeated destinations visited by the same person, the modeling results may be biased. This is because there may be unknown reasons (such as an individual's preference for a particular store or service) that explain the choice of a repeated destination, and such information is unavailable to us. Therefore repeated destinations should be examined before applying the model.

We calculate the Euclidean distance (in meters) between destinations visited by the same persona and identify the percentiles for the whole data set (Figure 5.6). If we use 100 meters as the threshold for defining repeated destinations, repeated destinations account for only about 10% of all destinations. Thus, the effects on the modeling results can be seen as marginal. Therefore the destination choice model can be applied for our data.

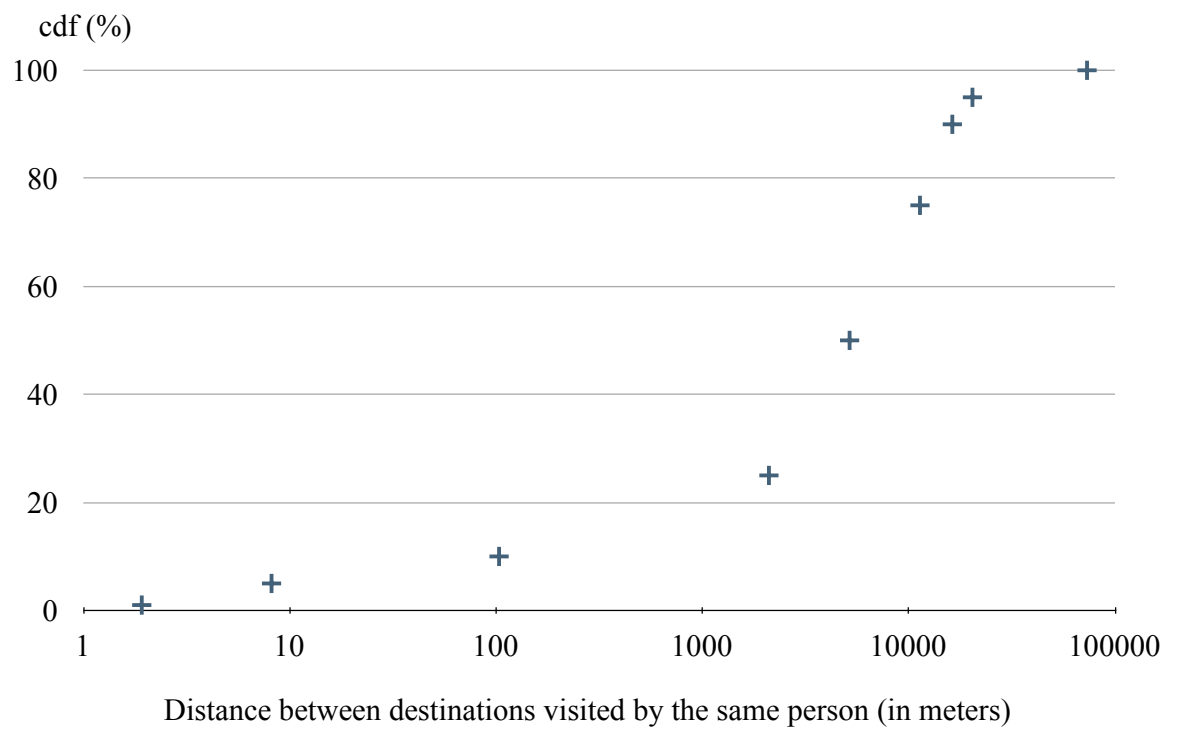


Figure 5.6: The cumulative probability distribution of distances between destinations visited by the same individual in the GPS data.

5.7 Model formulation

In modeling destination choice, the utility-based MNL model and its variations are widely used. Since the GPS data are panel data with repeated choices for individuals, there exists unobserved heterogeneity. To tackle this issue, we apply the mixed-effects logit model to investigate individuals' home-based non-work, single-destination destination choice.

The utility for one individual to visit destination k is defined as:

$$U_k = f(\ln(T_k), \Lambda_k, \Theta_k, \Upsilon_k, b) \quad (5.14)$$

Where T_k is travel time of the fastest route from home to destination k . Λ_k represents a vector of land use variables. Θ_k represents a vector of transportation network measures. Υ_k represents the interaction of the individual's socio-demographics and transportation network measures and land use at destination k . b is an extra random effect term generated from a standard normal distribution with mean 0 for this individual.

5.8 Results and analysis

5.8.1 Choice set size

Figure 5.7 shows the RMSE of the mixed-effects models given different choice set sizes. As the choice set size increases, the RMSE of the model decreases in the beginning but then floats around a certain value. The computational cost rises with the increase of choice set size.

Figure 5.8 further exhibits the RMSE values for different choice set sizes. The RMSE value floats around 0.48 as the choice set size ascends from 60 to 200. As the choice set size increases to 2000, the RMSE only lowers by 0.01 but the computational cost increases exponentially. The RMSE value for the chosen destinations shows similar values as the choice set size increases. The RMSE value for the chosen destinations also show similar values as the choice set size becomes greater than 40. After balancing the level of accuracy and computational cost, it is decided to use choice set size 60 for modeling non-work, single-destination choice, as it produces an appropriate level of accuracy with reasonable

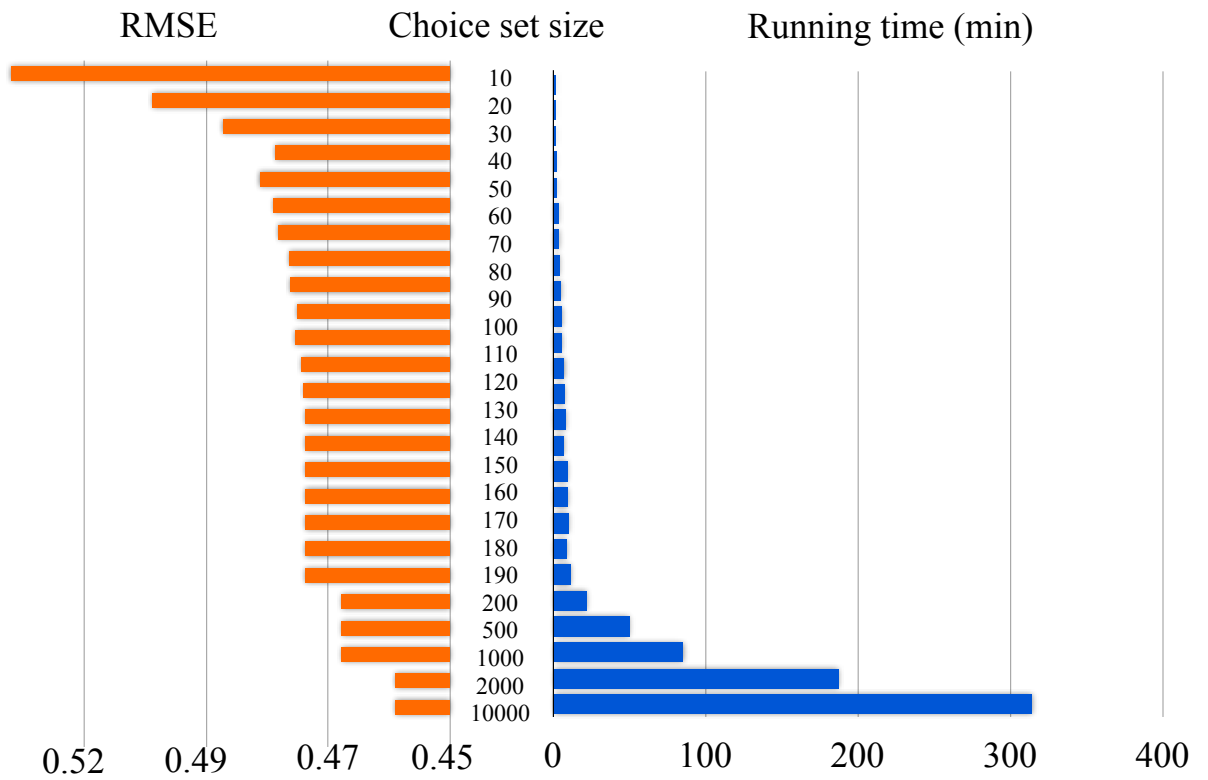


Figure 5.7: The root mean squared error (RMSE) value and running time for models of different choice set sizes.

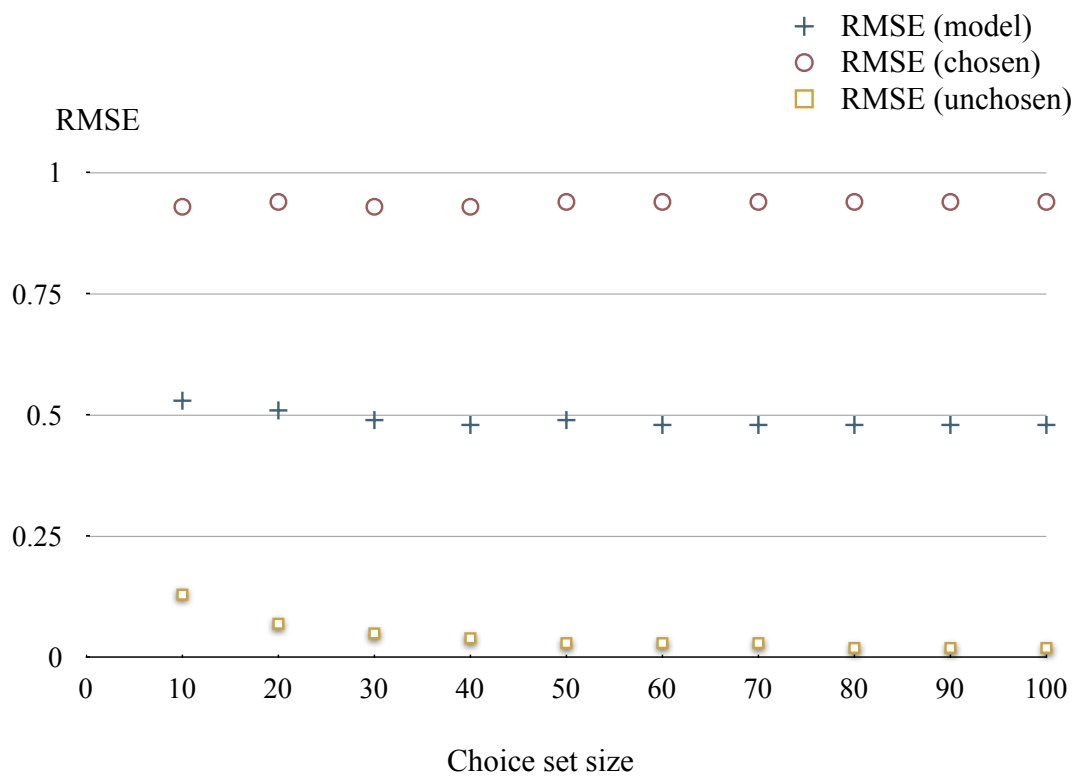


Figure 5.8: RMSE values for single-destination choice models of different choice sizes.

computational cost.

5.8.2 Modeling results

The results of the mixed-effects multinomial logit model are shown in Table 5.3. Model 1 includes all variables of interest. Model 2 excludes all interaction terms and turn index and Model 2 excludes all interaction terms and speed discontinuity, thanks to the correlation between turn index and speed discontinuity.

As shown in Model 1, travel time has a significant effect on destination choice. Longer travel time, all else equal, lessens the attractiveness of a destination. In Model 1, accessibility has a positive and statistically significant coefficient, indicating that an increase of stores at a destination makes it more attractive. The interaction term between male and accessibility has a negative coefficient, implying that a destination's increase of accessibility, all else equal, is more attractive to women than men. In addition, a destination's increase of accessibility, all else equal, is attractive to an individual with household income \$100,000 – \$149,999 than an individual whose household income is below \$100,000. We also have examined a model using continuous household income variable in the interaction terms. This model reports a smaller log likelihood value and smaller McFadden's R^2 than the using income levels as groups, and therefore is not adopted.

Though not statistically significant (but close to 0.10 level of significance), the coefficient of the diversity of services is also positive. In fact when the accessibility measure is excluded, the coefficient of the diversity of services becomes statistically significant. The interaction term between male and diversity of services has a positive sign, meaning that a destination's increase of diversity of services, all else equal, is more attractive to men than women.

In network structure measures, as hypothesized, the turn index has a negative coefficient which indicates that a destination reached via a route with more turns per unit time dampens its attractiveness. The interaction term between male and speed discontinuity has a positive coefficient, and so does the interaction term between male and turn index. The findings reveal the attractiveness of a destination drops more for a woman than for a man, as the route requires more turn per unit time. Speed discontinuity here has a nega-

Table 5.3: Modeling single-destination choice for non-work vehicle trips

Model type		Mixed-effects logit model			
Number of observations used		73381			
Independent variables		Model 1	Model 2	Model 3	Model 4
Land use	Accessibility ($\ln(A_k)$)	0.27 ***	0.17 ***	0.18 ***	0.01
	Male $\times \ln(A_k)$	-0.29 ***			
	Inclevel2 $\times \ln(A_k)$	0.10 ***			
	Inclevel3 $\times \ln(A_k)$	-0.12 ***			
	Diversity of services ($\ln(H_k)$)	0.19	0.14	0.14	0.26 **
	Male $\times \ln(H_k)$	0.11***			
	Inclevel2 $\times \ln(H_k)$	-0.36***			
	Inclevel3 $\times \ln(H_k)$	3.69 ***			
Network features	Travel time ($\ln(T_k)$)	-0.51***	-0.10 ***	-0.60 ***	
	Speed discontinuity (ψ_k)	-0.03 *	-0.67 ***		-0.03 **
	Turn index (ϑ_k)	-1.45**		-1.48 ***	-1.38 ***
	Male $\times \psi_k$	0.01			
	Inclevel2 $\times \psi_k$	0.11**			
	Inclevel3 $\times \psi_k$	-0.13***			
	Male $\times \vartheta_k$	0.19 **			
	Inclevel2 $\times \vartheta_k$	0.45 *			
Inclevel3 $\times \vartheta_k$	0.28 **				
Axis of travel	Time to work ($\ln(T_{w,k})$)	-0.80 **	-0.68 ***	-0.77 ***	
	Time to near downtown ($\ln(T_{d,k})$)	0.55 ***	0.56 ***	0.77 ***	
	Male $\times \ln(T_{w,k})$	-0.24			
	Inclevel2 $\times \ln(T_{w,k})$	-0.20			
	Inclevel3 $\times \ln(T_{w,k})$	0.25			
	Male $\times \ln(T_{d,k})$	-0.60 **			
	Inclevel2 $\times \ln(T_{d,k})$	0.93 ***			
	Inclevel3 $\times \ln(T_{d,k})$	0.24			
Time from home \times time to work (\ln form)				-0.43 ***	
Goodness of fit	AIC	11225	12134	11473	11547
	log likelihood	-5586	-6059	-5728	-5767
	McFadden's R^2	0.11	0.04	0.09	0.09
	Nagelkerke R^2	0.12	0.04	0.10	0.09

Inclevel 1: < \$100,000 ; Inclevel 2: \$100,000 – \$149,999; Inclevel 3: > \$149,999

*** significant at 0.01; ** significant at 0.05; * significant at 0.1.

tive coefficient but is not statistically significant. Further investigation reveals that it may be due to the correlation between speed discontinuity and turn index. When turn index is excluded from the model, the coefficient of speed discontinuity becomes statistically significant (see Model 2). The interaction terms also suggest that the changes of speed discontinuity have different effects on gender and income groups in single-destination choice.

Regarding the axis of travel, travel time between destination and work has a negative coefficient, meaning that a destination closer to work, all else equal, is more likely to be selected. In addition, as hypothesized, travel time between destination and the nearest downtown has a positive coefficient. It suggests that all else equal, a destination closer to the nearest downtown is less attractive which may be due to greater parking cost or limited parking space. All else equal, men are more likely to choose a destination far away from downtown than women. We further test a new variable which equals the multiplication of travel time from home and travel time to work. It aims to quantify the distance from a destination to the axis between work and home. It is hypothesized that the greater this term is, the less attractive the destination is. As this multiplicative term is correlated with trip chain's travel time and travel time to work, these two variables are excluded in the model when the multiplicative term is included. The results are shown in Model 4 in Table 5.3. The coefficient of the multiplicative term is negative which supports our hypothesis.

The Pearson coefficients of the key independent variables are shown in Table 5.4. The correlation coefficient between a destination's accessibility and travel time to downtown is about -0.67, showing that if a destination is farther away from downtown (which makes it more suburban), all else equal, its accessibility is lower. In addition, the correlation between speed discontinuity and turn index is about 0.40, suggesting a moderate positive relationship.

The elasticity of key independent variables are further calculated (Table 5.5). Considering the correlations among variables, we first run the mixed-effect model on one variable at a time, and then calculate the elasticity for each estimated coefficient. The variable that has the highest absolute value of elasticity is turn index, following which is travel time. A 1 % increase of the number of turns per unit distance for the travel route, all else equal, is associated with an about 76% decrease of the likelihood of selecting this destination. A

Table 5.4: Correlations of key independent variables

	Diversity	Travel time	Discont.	Turn index	Time to work	Time to downtown
Accessibility	-0.21	-0.12	0.08	0.11	-0.58	-0.67
Diversity	1	-0.12	0.05	-0.02	0.13	0.21
Travel time		1	0.36	-0.05	0.32	0.25
Discont.			1	0.40	-0.10	-0.10
Turn index				1	-0.15	-0.13
Time to work					1	0.83
Time to downtown						1

Note: all coefficients are statistically significant at 0.01.

Table 5.5: Elasticity of the odds of single-destination choice for key independent variables (1% change)

Variable	Elasticity of selection odds (%)	Rank by absolute value
Turn index	-76	1
Travel time	-43	2
Time to home × time to work	-32	3
Time to work	-29	4
Diversity of services	24	5
Speed discontinuity	-19	6
Accessibility	12	7
Time to downtown	-11	8

1 % increase of travel time, all else equal, is associated with an about 43% decline of the likelihood of selecting this destination. The same interpretation applies to other variables. We further test a new independent which is the multiplication of travel time to home and travel time to work. The idea is to investigate the impact of a destination's distance to the axis between home and work. Its elasticity equals -32%, suggesting that the farther away a destination is from the axis between home and work, the less attractive the destination is.

5.9 Discussion

This research proposes a new approach that combines survival analysis and random sampling to form choice set for home-based non-work, single-destination choice. A systematic

investigation of appropriate choice set sizes is also performed. Based on the in-vehicle GPS travel data in the Twin Cities, mixed-effects multinomial logistic models are used to model single-destination choice. In these models we examine the effects of land use and transportation networks on destination choice. The key findings are:

1. The two most influential factors on single-destination choice are turn index and travel time from home to destination.
2. Greater accessibility and diversity of services, all else equal, make a destination more attractive.
3. A destination reached by a route with greater changes of speed per unit time or more turns per unit time is less attractive.
4. Individuals' socio-demographics such as gender and household income, interacting with land use and route network measures, also affect destination choice.
5. The axis of travel impacts destination choice. All else equal, a non-work destination closer to work is more attractive to travelers. A destination closer to the nearest downtown is less attractive to travelers, which may be due to a greater parking cost and other nuisances near downtown.
6. The above variables have different effects on gender and income groups in destination choice.

In summary, using single-destination choice as an example, this chapter proposes a new approach to select choices and a systematic method to decide the choice set size. Further, we test some hypotheses which were not tested before. The results indicate that land use features, travel time, familiarity with destinations, consideration of parking, perception of a destination's reachability, and individuals' income and socio-demographics all together influence non-work destination choice.

For simplicity, this chapter focuses on home-based non-work, single-destination choice. In a multi-destination scenario, the spatial interactions of different destinations may influence both the choice set formation and the destination choice process in a trip chain. The

following chapter will examine this problem for non-work trip chains with two destinations.

Chapter 6

Modeling two-destination choice

This chapter aims to examine the impacts of land use and transportation network structure on home-based non-work, two-destination choice in the context of trip chains. The key contributions of this chapter include:

- Proposing a model that explicitly considers home-based two-destination choice (i.e., home \rightarrow destination 1 \rightarrow destination 2 \rightarrow home).
- Proposing and empirically testing three methods to form choice sets for modeling home-based two-destination choice.
- Empirically applying the reference-dependent theory to the two-destination choice problem.
- Investigating the impacts of land use and characteristics of travel routes on two-destination choice based on the in-vehicle GPS data.

6.1 Modeling procedure

The basic procedure for modeling home-based two-destination choice is similar to modeling home-based non-work, single-destination choice. The first step is to define destinations for individuals. The next step is to properly select two destinations to constitute a home-based trip chain. The third step is to form a choice set for each decision. The following step is to model two-destination choice. Figure 6.1 sketches the two-destination choices

of home-based non-work trip chains. A choice set consists of multiple two-destination choices. Each trip chain starts from home and ends at home after stopping at two non-work destinations.

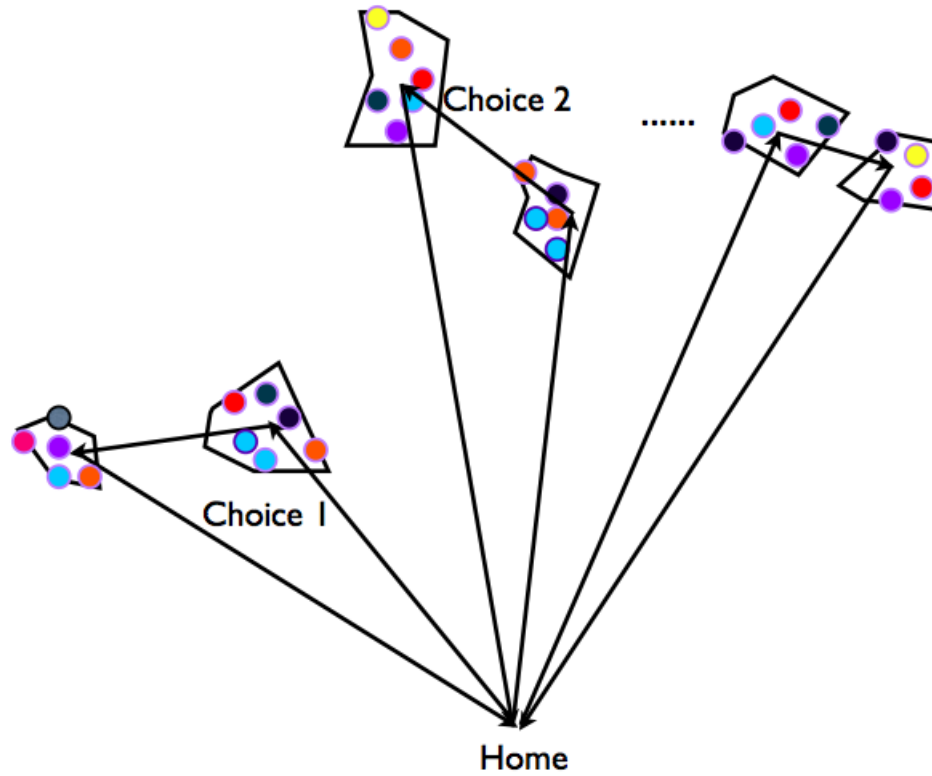


Figure 6.1: Multiple choices of home-based non-work, two-destination trip chains.

6.2 Definition of a choice

A two-destination choice problem suggests that a choice has two destinations. A destination has the same definition as in Chapter 5 which is based on the centroids of Census blocks in 2010. A 15-min walking area is created around each destination in order to characterize land use around the destination. As in Chapter 5, it is assumed that this is the area one visits after parking the car.

6.3 Independent variables

The independent variables used in this chapter include land use measures, transportation network measures, axis of travel measures, travel time saving ratio, and the interaction terms between socio-demographics and other independent variables.

6.3.1 Land use measures

The land use measures include accessibility, diversity of services, and the similarity index of two destinations in a trip chain. The definitions of accessibility and diversity of services at a destination are the same as in Chapter 5. Given two destinations j and k in a trip chain, we respectively measure the cumulative opportunities and diversity of services (entropy measure) of the two destinations. As in Chapter 5, it is hypothesized that greater accessibility and diversity of services at either one of the two destinations, all else equal, make the trip chain more attractive, because such features support multi-purpose/comparison-shopping opportunities at the destinations.

In addition, it is of interest to measure the similarity of the services at the two destinations. If two destinations have greater diversity of services, it promotes multi-purpose shopping for the whole trip chain and helps reduce the average travel cost for each task involved. Therefore, greater dissimilarity of two destinations in a trip chain makes the trip chain more favorable.

One approach to quantify the similarity of services is based on Lieberman's dissimilarity measure (Lieberman, 1969). The similarity measure ($\Xi_{j,k}$) of two destinations can be written as:

$$\Xi_{j,k} = \sum_{g=1}^G \rho_{jg} \rho_{kg} \quad (6.1)$$

Where g indicates a common category of service at destination k and destination j . G is the total number of common types of services at the two destinations. ρ_{jg} represents the proportion of service type g at destination k . ρ_{kg} refers to the proportion of service type g at destination j . The definition suggests that the more common services the two destinations have (i.e., the bigger $\sum_{g=1}^G \rho_{jg} \rho_{kg}$ is), the more similar the two destinations

are. The natural logarithm form of $\Xi_{j,k}$ is used in the model for comparing its elasticity of the odds of selection of a trip chain with other independent variables. Several examples of calculating the similarity index for two destinations are shown in Figure 6.2.

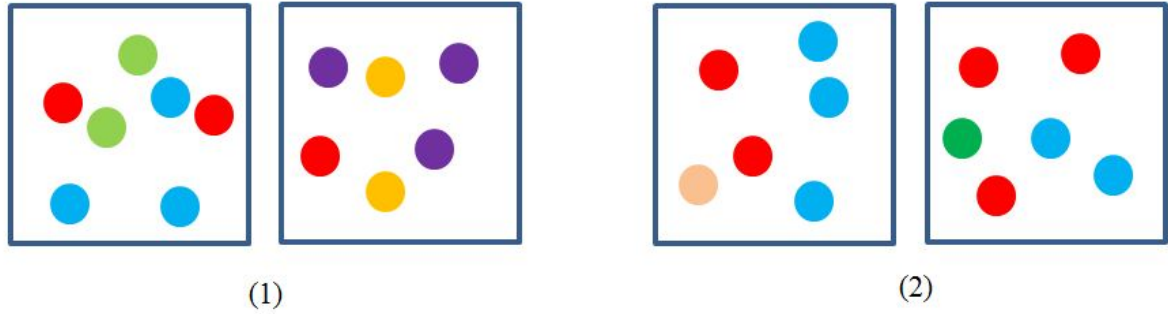


Figure 6.2: Examples of calculating the similarity index.

Circles of the same color indicate the same type of service. In Figure 6.2-(1), the two destinations share one common type of service (red). The similarity index of the two destinations can be calculated as $\frac{2}{7} \cdot \frac{1}{6} = 0.05$. In Figure 6.2-(2), the two destinations have two types of services in common (red and blue). The similarity index of the two destinations equals $\frac{2}{6} \cdot \frac{3}{6} + \frac{3}{6} \cdot \frac{2}{6} = 0.32$. The two destinations in Figure 6.2-(1) have greater similarity than the two destinations in Figure 6.2-(2).

6.3.2 Transportation network measures

For each trip chain, the whole route (i.e., home \rightarrow destination 1 \rightarrow destination 2 \rightarrow home) is analyzed. The route-specific network measures used include speed discontinuity and turn index which are defined in Chapter 5. The hypotheses are also similar to Chapter 5.

6.3.3 Axis of travel

The axis of travel measures include two variables: travel time to work ($T_{w,jk}$) and travel time to the nearest downtown ($T_{d,jk}$). The goal of measuring the axis of travel is to examine the influence of work and downtown's locations on the selection of destinations.

Travel time to work

We are interested in how travel time to work from the two destinations may influence destination choice. A variety of measures are investigated, such as the sum of travel time

from the two destinations respectively to work, the major destination's travel time to work, and the longer travel time of the two destinations from work, and have compared the models' goodness of fit given different measures. The measure that produces the highest goodness of fit is travel time from the major destination of the two destinations to a person's workplace. For actual trips, the major destination is defined as the destination that has a longer length of stay. The travel time to work measure ($T_{w,jk}$) for one individual can be written as:

$$T_{w,jk} = \begin{cases} T_{w,j} & \text{if destination } j \text{ is the major destination} \\ T_{w,k} & \text{if destination } k \text{ is the major destination} \end{cases} \quad (6.2)$$

Where $T_{w,j}$ represents the network travel time between destination j and workplace. $T_{w,k}$ indicates the network travel time between destination k and workplace. $T_{w,jk}$ is used to indicate one's familiarity with the major destination because one tends to be familiar with the area close to workplace. A smaller value indicates that one may be more familiar with the major destination. All else equal, travelers may favor such a choice with smaller travel time to work due to greater familiarity with the destination. The natural logarithm form of $T_{w,jk}$ is used in the model for comparing its elasticity of the odds of selection of a trip chain with other independent variables.

Travel time to the nearest downtown

After several different definitions are tested, travel time to the nearest downtown ($T_{d,jk}$) is defined as travel time from the major destination to the nearest downtown. Mathematically, it can be written as:

$$T_{d,jk} = \begin{cases} T_{d,j} & \text{if destination } j \text{ is the major destination} \\ T_{d,k} & \text{if destination } k \text{ is the major destination} \end{cases} \quad (6.3)$$

This measure implies one's consideration of parking constraints, narrower streets, and more traffic lights, etc. If $T_{d,jk}$ is smaller, it means the major destination is closer to the nearest downtown where parking is more limited and more costly, and thus the trip chain is less attractive to travelers for non-work trips. Therefore we hypothesize that all else

equal, smaller $T_{d,jk}$ makes a trip chain less attractive. The natural logarithm form of $T_{d,jk}$ is used in the model for comparing its elasticity of the odds of selection of a trip chain with other independent variables.

6.3.4 Travel time savings

Based on the reference-dependent theory (Tversky and Kahneman, 1991), a trip chain can be more attractive than making two separate trip chains due to travel time savings. For one individual, we use $\zeta_{j,k}$ to represent the travel time saving ratio of a trip chain with two destinations compared with making two separate round trips to visit the two destinations. The travel time saving ratio of a trip chain can be written as:

$$\zeta_{j,k} = 1 - \frac{T_{j,k}}{T_j + T_k} \quad (6.4)$$

Where $T_{j,k}$ represents the home-based trip chain's travel time. T_j refers to travel time of making a home-based round trip to visit destination j only (i.e., home \rightarrow destination $j \rightarrow$ home). T_k represents the travel time of making a home-based round trip to visit k only (i.e., home \rightarrow destination $k \rightarrow$ home). Several examples of calculating $\zeta_{j,k}$ are shown in Figure 6.3.

Travel time saving ratio reflects the value of chaining the trips. If a two-destination trip chain has a greater travel time saving ratio, all else equal, it is more worthwhile chaining the trips, and therefore the two-destination trip chain is more likely to be selected.

The hypotheses of the impacts of the independent variables on a trip chain's attractiveness are summarized in Table 6.1.

6.4 Choice set formation

By expanding the survival analysis-based random selection method in Chapter 5, we propose three approaches to build the choice sets for two-destination trip chains.

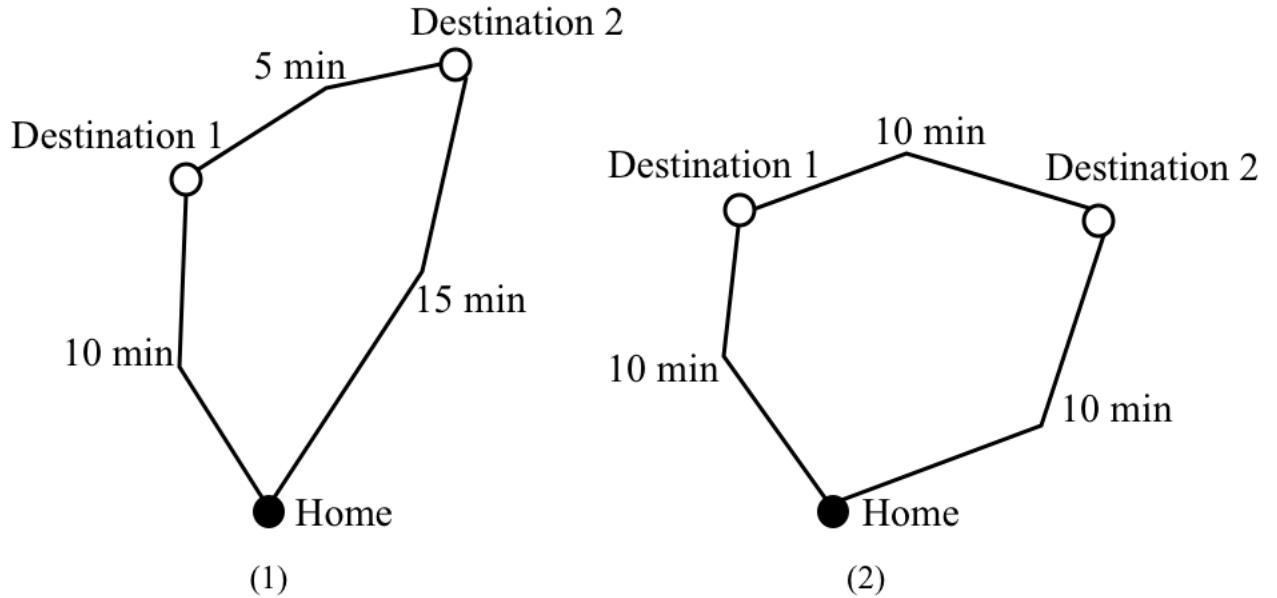


Figure 6.3: Examples of calculating travel time saving ratio.

In Scenario (1), travel time for making the home-based two-destination trip chain equals $10 + 5 + 15 = 30$ minutes. The total travel time for making two separate home-based one-destination trip chains equals $10 \times 2 + 15 \times 2 = 50$ minutes. The travel time saving ratio equals $(1 - \frac{30}{50}) \times 100\% = 40\%$. In Figure 6.3-(2), travel time for making the home-based two-destination trip chain equals 30 minutes. The total travel time for make two separate home-based one-destination trip chains equals $10 \times 2 + 10 \times 2 = 40$ minutes. The travel time saving ratio equals $(1 - \frac{30}{40}) \times 100\% = 25\%$. Therefore, Scenario (1) has a greater travel time saving ratio than Scenario (2).

Table 6.1: Hypotheses of key independent variables

Variable	A trip chain's attractiveness
Travel time	-
Travel time saving ratio	+
Accessibility	+
Diversity of services	+
Similarity of destinations	-
Speed discontinuity	-
Turn index	-
Distance to work	-
Distance to the nearest downtown	+

6.4.1 Approach I

The procedure of Approach I is shown in Figure 6.4. Given each individual's GPS trips, the first step is to calculate travel time from home to the first destination and travel time from home to the second destination (which equals travel time from home to the first destination plus travel time between the first destination and the second destination). The second step is for each individual to estimate a survival function based on travel time from home to the first destination, and to estimate another survival function based on travel time from home to the second destination. According to the estimated survival functions we can predict the weights of all destinations. Given the weights of all potential destinations from the first survival function, we can select a destination (indicated by j) using random sampling. Likewise, given the weights of all destinations from the second survival function, we can select the other destination (indicated by k) using random sampling. In several cases where a survival function does not converge, the traditional simple random selection is used.

After the two destinations are chosen, the fastest-path travel route of the home-based trip chain can be identified based on the road network by solving the Traveling Salesman Problem (Gutin and Punnen, 2002) where the sequence of visiting the two destinations is not fixed.

6.4.2 Approach II

The second approach assumes that one destination is first selected and the other destination is secondly selected based on travel time from the first destination to the second destination. The procedure of Approach II is shown in Figure 6.5. The first step is for each individual to estimate the first destination j 's survival function based on travel time from home to the first destination j . The second step is to estimate the second destination k 's survival function based on travel time between the first destination j and the second destination k . The two destinations j and k are randomly selected separately based on the weights of all possible destinations from the two survival functions. In cases where a survival function does not converge, the traditional simple random selection is used. The fastest-path travel route of the home-based trip chain to visit the two destinations are created based on the road network, given that the sequence of visiting j and k is fixed.

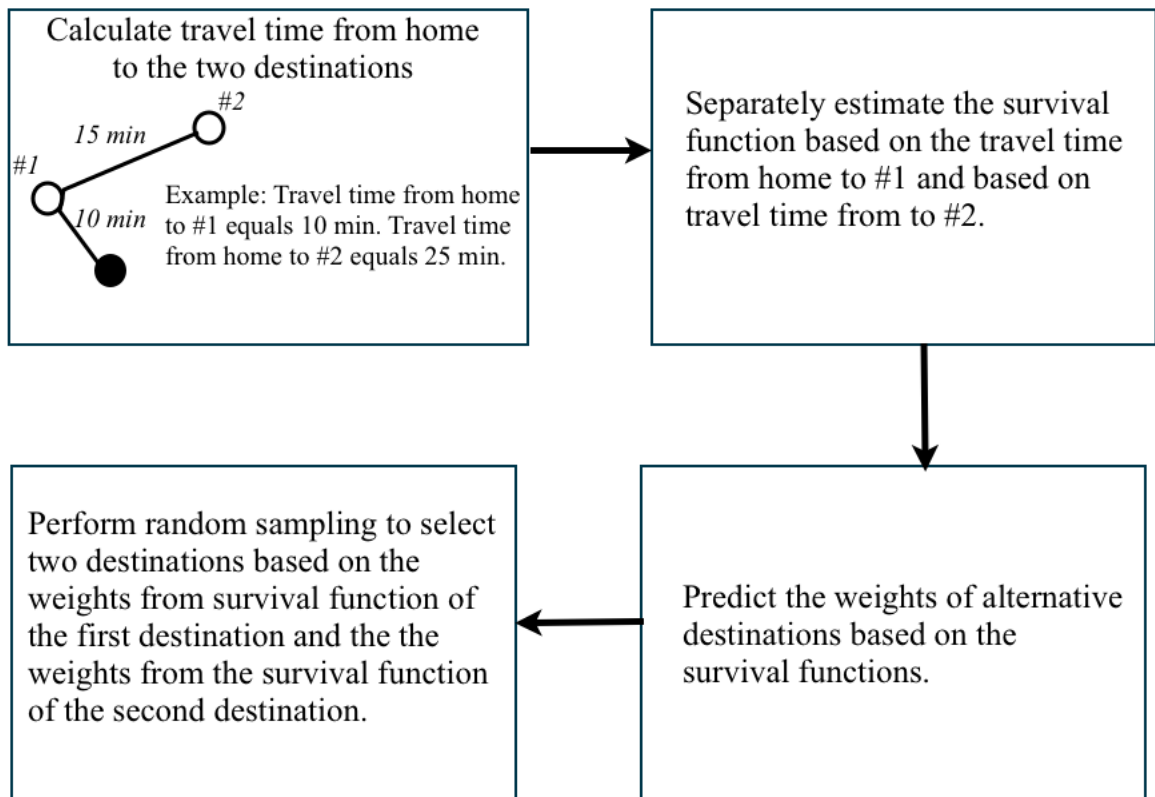


Figure 6.4: Approach I to select two destinations to constitute a trip chain.

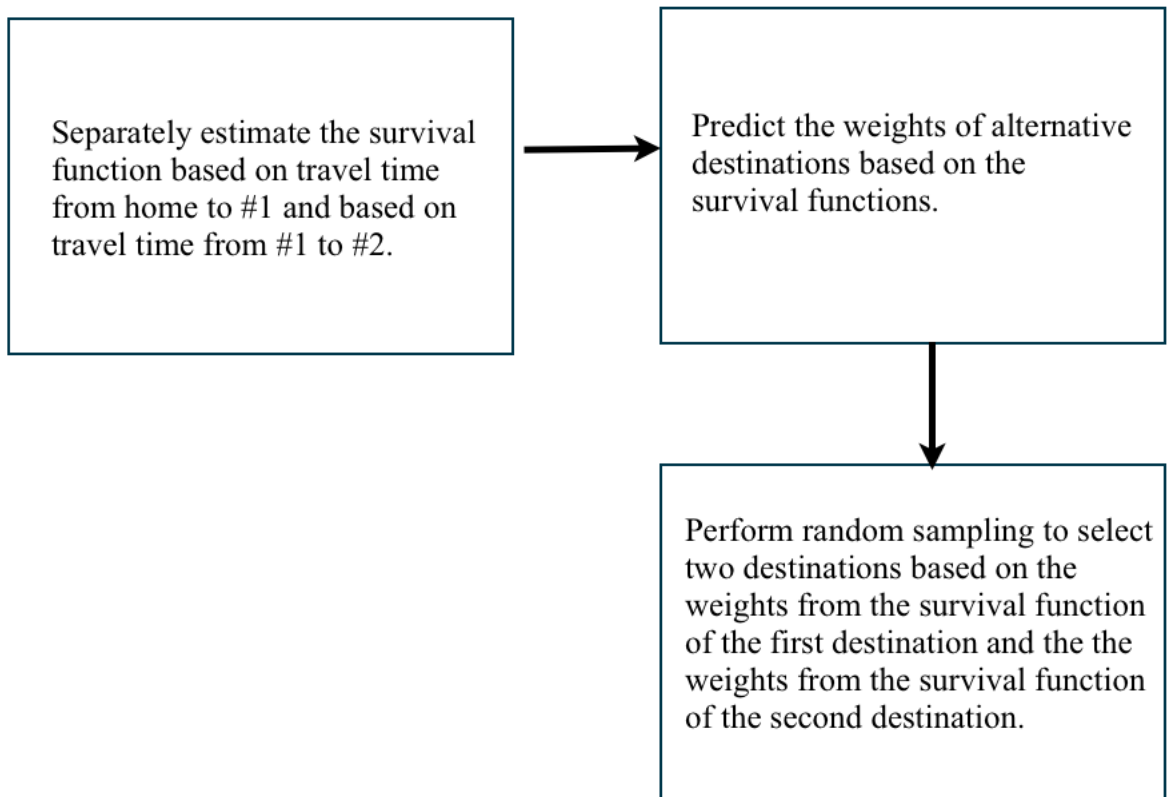


Figure 6.5: Approach II to select two destinations to constitute a trip chain.

6.4.3 Approach III

The third approach assumes that an individual selects the major destination first and then decides the minor destination. Although we do not have direct information regarding which destination is the major one, we may infer it based on the individual's length of stay at a destination. A destination associated with longer length of the stay is considered as the major of the two. While this assumption is not always true, it serves as a plausible start. The procedure of Approach III is shown in Figure 6.6. The first step includes identifying the major destination and the minor destination based the length of stay and calculating travel time from home to the major destination and travel time between the major destination and the minor destination. In the example in Figure 6.6, given that the second destination is the major and the first destination is the minor, travel time from home to the major destination equals $10 + 15 = 25$ minutes and travel time between the major destination and minor destination equals 10 minutes. The second step is for each individual to separately estimate the survival function based on travel time from home to the major destination and based on travel time between the two destinations. The third step is to separately predict the weights of all destinations based on the major destinations' survival function and the minor destinations' survival function. The last steps are to randomly select a major destination based on the weights of destinations from the major destinations' survival function, and to randomly select a minor destination based on the weights of destinations from the minor destinations' survival function. In cases where a survival function does not converge, the traditional simple random selection is used. After the two destinations are selected, the fastest-path travel route of the home-based trip chain can be created based on the road network by solving the Traveling Salesman Problem where the sequence of visiting the two destinations is not fixed (the major destinations may be visited later than the minor destination).

6.5 Choice set size

The procedure to decide the choice set size is similar to Chapter 5. For each approach we test choice set sizes from 10 to 100 with an increment of 10. We measure the root

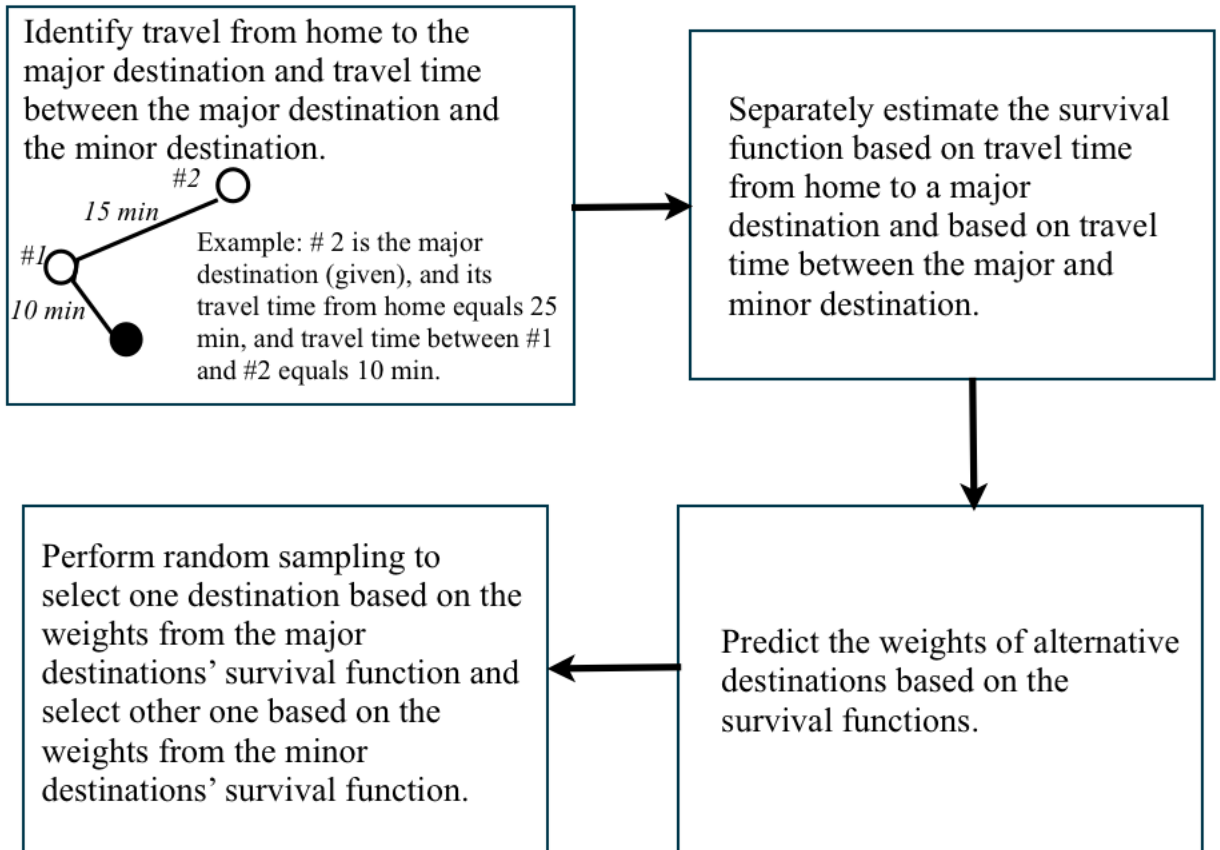


Figure 6.6: Approach III to select two destinations to constitute a trip chain.

mean square (RMSE) of the chosen trip chains and the RMSE of the unchosen trip chains. The RMSE of the model is defined as the average of $RMSE_{chosen}$ and $RMSE_{unchosen}$. In other words, we assign a 50% weight to $RMSE_{chosen}$ and the other 50% weight to $RMSE_{unchosen}$, which can be presented as:

$$RMSE_{model} = p \cdot RMSE_{chosen} + (1 - p) \cdot RMSE_{unchosen} \quad (6.5)$$

Where $p = 0.5$. The reason for doing so is to balance accuracy between predicting chosen trip chains and unchosen trip chains (see more detailed discussion in Chapter 5). The RMSEs of models with different choice set sizes developed by different approaches are further compared to decide an appropriate choice set size.

6.6 Model formulation

The mixed-effects logit model is used to model individuals' home-based two-destination choice. The utility for a subject to select a choice consisting of two destinations j and k is defined as:

$$U_{j,k} = f[\ln(T_{j,k}), \Lambda_{j,k}, \Theta_{j,k}, \Delta_{j,k}, b] \quad (6.6)$$

Where $T_{j,k}$ represents travel time of the whole trip chain of visiting the two destinations j and k . $\Lambda_{j,k}$ indicates a vector of land use variables for destination j and k including the similarity of the two destinations. $\Theta_{j,k}$ indicates a vector of transportation network measures of the trip chain's travel route. $\Delta_{j,k}$ represents a vector of axis of travel variables for trip chains. b is an extra random effect term for a subject generated from a standard normal distribution with mean zero.

6.7 Results and analysis

6.7.1 Deciding choice set size

To decide an appropriate choice set size for the models, we test various choice set sizes from 10 to 100 and measure the corresponding RMSE values, which are shown in Fig-

ure 6.10. Though there are exceptions, the generation trend is that as the choice set size increases from 10 to 60, the RMSE value drops. As the choice set size continues to climb, the RMSE value for the models developed by Approach I decreases, but it fluctuates for models developed by Approach II and Approach III. The fluctuation can be partially attributed to a certain degree of randomness in selecting trip chains. In models developed by Approach III, the model with choice set size 60 produces the lowest RMSE. In models developed by approach II, the RMSE value fluctuates around 0.16, and for models developed by Approach II, the value stays around 0.19. Figure 6.7, and Figure 6.8, and Figure 6.9 exhibit the running times versus RMSE values for various models. As the choice set size increases, the running time rises. Balancing accuracy and running time and for consistent comparison of models developed by different approaches, we decide to choose choice set size 60 for models developed by the three approaches. This is believed to have sufficient accuracy compared with other sizes and does not demand too high computational cost. A choice set size equals 60, including 1 chosen choice and 59 unchosen choices.

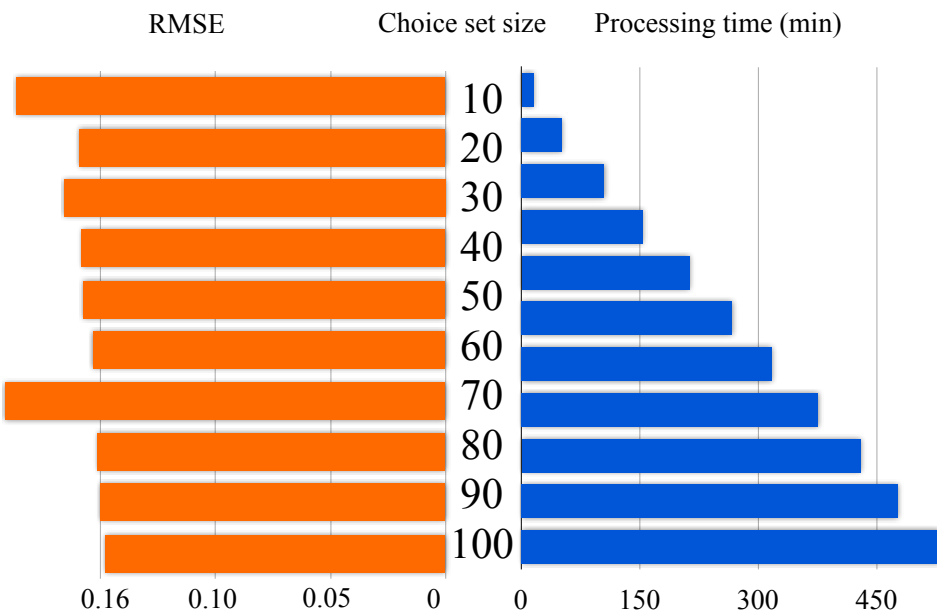


Figure 6.7: The RMSE values and running times for the models of different sizes developed by Approach I.

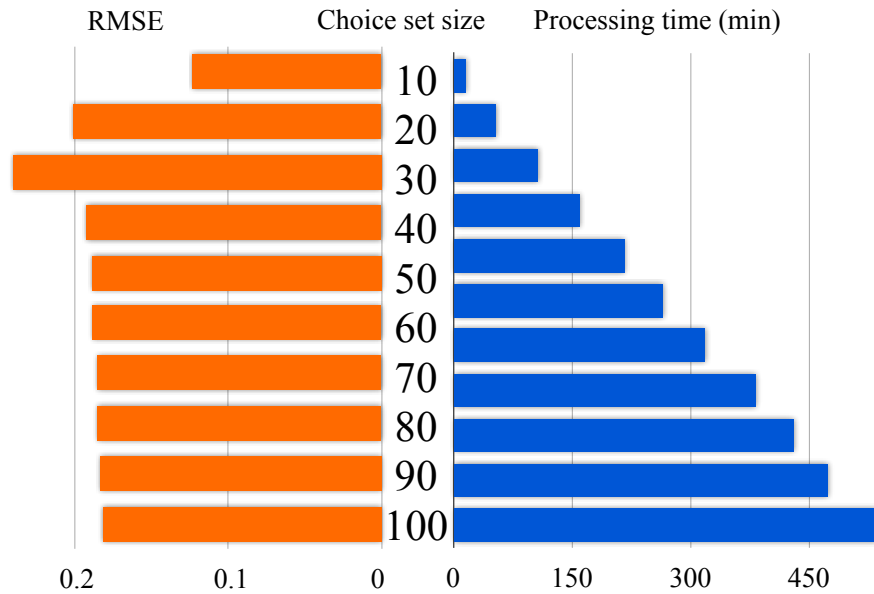


Figure 6.8: The RMSE values and running times for the models of different sizes developed by Approach II.

6.7.2 Modeling results

After each model's choice set size is decided, the modeling results of the mixed-effects logit models (without interaction terms) developed by the three approaches are shown in Table 6.2. In terms of land use, in Model 1 developed by Approach I, the coefficient of the first destination's accessibility has a positive coefficient and is statistically significant. The first destination's entropy measure has a positive coefficient which is statistically significant. Regarding the second destination, the accessibility measure is statistically significant but the entropy measure is not statistically significant. The relatively large absolute values of coefficients for the entropy measure for the first destination and the similarity measure may suggest correlation, which is confirmed by the Pearson correlation test (Table 6.3). Therefore, it is of interest to separate these measures in models to further investigate their coefficients. The coefficient of the similarity of the two destinations in a trip chain is negative and statistically significant. Greater dissimilarity of the two destinations, all else equal, makes a trip chain more attractive. It concurs with the notion that such destinations can

Table 6.2: Modeling home-based, two-destination choice in the context of trip chains (using basic independent variables)

Model type		Mixed-effects logit model		
Model's name		Model 1	Model 2	Model 3
Choice set generation method		Approach I	Approach II	Approach III
Choice set size		60	60	60
Land use at Location 1 ★	Accessibility ($\ln(A_j)$)	0.39 ***	0.30 ***	0.22 ***
	Diversity of services ($\ln(H_j)$)	9.37 ***	9.15 ***	10.69 ***
Land use at Location 2 ●	Accessibility ($\ln(A_k)$)	0.82 **	0.70 ***	1.15 ***
	Diversity of services ($\ln(H_k)$)	-0.09	-0.01	-0.69
Comparing two locations	Similarity ($\Xi_{j,k}$)	-13.00 ***	-12.70 ***	-15.27 ***
Travel/network features	Travel time ($\ln(T_{j,k})$)	-2.02 ***	-1.81 ***	-3.96 **
	Turn index ($\vartheta_{j,k}$)	-0.95 ***	-0.79 ***	-0.67***
	Speed discontinuity ($\psi_{j,k}$)	0.67 ***	0.52 **	0.57 ***
	Time saving ratio ($\zeta_{j,k}$)	7.48 ***	6.81***	9.62 ***
Axis of travel	Time to work ($\ln(T_{w,jk})$)	0.93 ***	0.71 ***	-0.40
	Time to downtown ($\ln(T_{d,jk})$)	1.22 ***	1.16 ***	1.62 ***
Goodness of fit	AIC	963	1122	223.0
	log likelihood	-468.8	-548	-364.9
	McFadden's R^2	0.88	0.86	0.90
	Nagelkerke R^2	0.89	0.87	0.91

In Model 1 and Model 2, ★ indicates the first visited location in the trip chain. In Model 3, ★ represents the major destination in the trip chain, defined as the one with longer length of stay. In Model 1 and Model 2, ● indicates the secondly visited location in the trip chain. In Model 3, ● represents the minor destination in the trip chain, defined as the one with shorter length of stay.

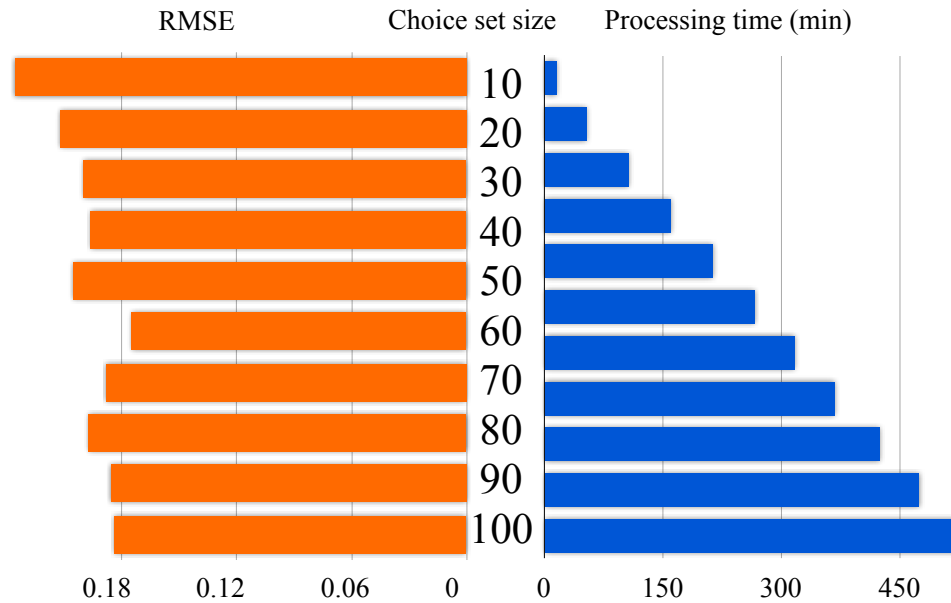


Figure 6.9: The RMSE values and running times for the models of different sizes developed by Approach III.

provide more complementary services. Total travel time of a trip chain is also an important factor in influencing a trip chain’s attractiveness. Greater travel time for the whole trip chain, all else equal, is associated with less attractiveness of a choice.

The turn index variable in all three models carries a negative sign, suggesting that more turns per unit travel time, all else equal, lower the attractiveness of the trip chain. It pinpoints the impact of routes-specific factors on two-destination choice. Contrary to our hypothesis, the coefficient of speed discontinuity has a positive sign, implying that all else equal, a route with more changes of speed per unit length along the route make a trip chain more attractive. Further investigation reveals that the Pearson correlation coefficient between turn index and speed discontinuity equals 0.80 (see Table 6.4). A relatively strong correlation may bias the estimates of the two variables’ coefficients. When the removing turn index from the model, speed discontinuity’s coefficient becomes negative but is not statistically significant.

Travel time saving ratio has a positive coefficient, indicating that all else equal, if a trip

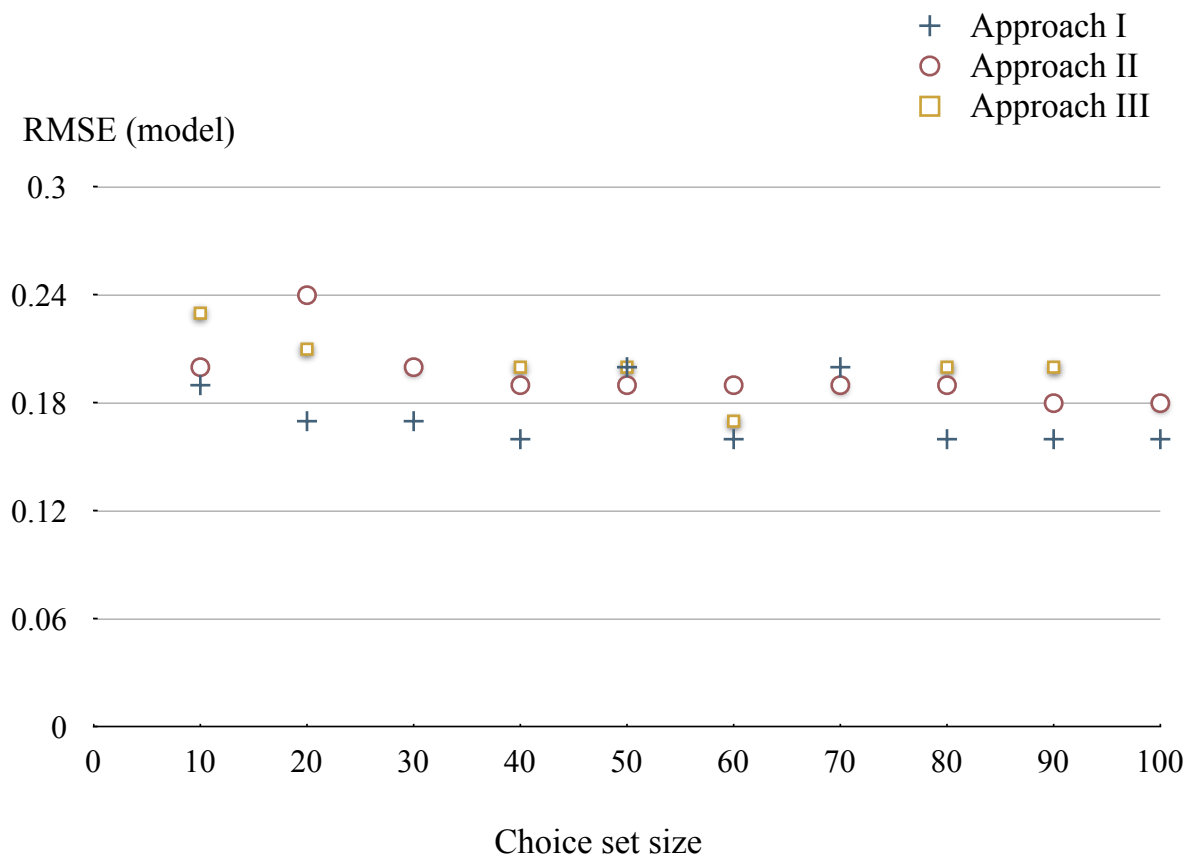


Figure 6.10: Comparison of root mean squared errors of different choice set sizes for models developed by the three approaches.

chain has a higher travel time saving ratio compared with making separate round trips, it is more attractive to travelers. This finding is an application of the reference-dependent theory in travel behavior. Individuals' perception of destinations' travel time saving influences individuals' destination choice.

In terms of the axis of travel, the distance to downtown has a positive coefficient, suggesting that favorable non-work trip chain destinations, all else equal, tend to be farther away from the closer downtown. The correlation between the two variables equals 0.75 (Table 6.4). When only one of the two variables is retained in the model, the other variable becomes statistically significant in all three models and the sign of the coefficient is consistent with our hypothesis. We further test a new variable for the axis of travel, which is the multiplication of the trip chain's travel time and the major destination's travel time to work. The idea is to investigate the impact of the destinations' distance to the axis between home and work on destination choice. It is hypothesized that the greater this term is, the less attractive a trip chain is. The natural logarithm of this multiplicative term is used in order to compare its elasticity with other independent variables. Due to this term's correlation with trip chain's travel time and major destination's travel time to work, separate regressions are run for models including this term while the trip chain's travel time and other axis of travel measures are excluded. The results are shown in Table 6.6. The results reveal that this multiplicative term has a negative coefficient in all three models, which is consistent with our hypothesis.

As shown in Table 6.2, all three models seem to have high goodness of fit, as evidenced by high McFadden's R^2 and Nagelkerke R^2 measures (note that they should not be interpreted in the same way as R^2 for OLS models). All these measures show that Model 3 seems to perform the best. Therefore we recommend the major/minor-approach for modeling two-destination choice.

Note that the Pseudo- R^2 values in the two-destination choice problem are higher than the Pseudo- R^2 values in the one-destination choice problem (Chapter 5). This is because it is easier to find the real pair of destinations from our artificially generated choice sets in two-destination choice than in one-destination choice. For a pair of destinations, if one destination is very far from home or has low accessibility/diversity of services, the prob-

Table 6.3: Correlations of land use variables in Model 3

	$\ln(A_j)$	H_j	$\ln(A_k)$	H_k	$\Xi_{j,k}$
$\ln(A_j)$	1	0.09***	0.15***	0.01***	0.37***
H_j		1	-0.007***	0.01***	-0.61***
$\ln(A_k)$			1.0	0.22***	0.08***
H_k				1.0	-0.08
$\Xi_{j,k}$					1

Table 6.4: Correlations of route-specific network measures in Model 3

	$\vartheta_{j,k}$	$\psi_{j,k}$	$\zeta_{j,k}$	$\ln(T_{w,jk})$	$\ln(T_{d,jk})$
$\vartheta_{j,k}$	1.0	0.80***	-0.11***	-0.06***	-0.06
$\psi_{j,k}$		1.0	-0.11***	-0.10***	-0.08***
$\zeta_{j,k}$			1.0	-0.21***	-0.16***
$\ln(T_{w,jk})$				1.0	0.75***
$\ln(T_{d,jk})$					1.0

ability of choosing that pair would be considerably lowered. This is especially the case when both destinations are far away or have low accessibility/diversity of services. In addition, travel time saving ratio is not considered in the choice set formation process, which might influence the easiness of finding the real pair. For example, if a pair of destinations matches the real pair in terms of travel time and land use parameters but has a low travel time saving ratio, its selection probability might still be low and thus makes the real pair more likely to be selected. Future research is needed to appropriately evaluate choice set formation approaches for the multi-destination choice problem.

We test the models without the entropy measures (Table 6.5). There is a big change in the coefficient of the similarity of the two destinations in a trip chain. Its level of magnitude concurs with our hypothesis. The elasticity of key independent variables are further calculated (Table 6.7). Considering the correlations among variables, we first run the mixed-effect model on one variable at a time, and then calculate the elasticity for each estimated coefficient. Travel time has the biggest impact on destination time. A 1% increase of a trip chain's travel time, all else equal, is associated with a 95% decrease of the likelihood of selecting the trip chain. The second influencer is travel time to work. A 1% increase of

Table 6.5: Modeling home-based, two-destination choice in the context of trip chains (excluding the entropy measure)

Model type		Mixed-effects logit model		
Model's name		Model 1	Model 2	Model 3
Choice set generation method		Approach I	Approach II	Approach III
Choice set size		60	60	60
Land use Location 1 ★	Accessibility ($\ln(A_j)$)	1.50 ***	1.56 ***	1.79 ***
	Diversity of services ($\ln(H_j)$)			
Land use Location 2 ●	Accessibility ($\ln(A_k)$)	0.61 ***	0.57 ***	0.85 ***
	Diversity of services ($\ln(H_k)$)			
Comparing two locations	Similarity ($\ln(\Xi_{j,k})$)	-2.55 ***	-2.63 ***	-3.65 ***
Travel/network features	Travel time ($\ln(T_{j,k})$)	-1.59 ***	-1.68 ***	-3.72 ***
	Turn index ($\vartheta_{j,k}$)	-1.00 **	-0.75 ***	-0.51***
	Speed discontinuity ($\psi_{j,k}$)	0.80 *	0.32 *	0.45 ***
	Time saving ratio ($\zeta_{j,k}$)	3.58 ***	2.98 ***	5.57***
Axis of travel	Time to work ($\ln(T_{w,jk})$)	-0.34 ***	-0.47 ***	-1.84 ***
	Time to downtown ($\ln(T_{d,jk})$)	1.80 ***	1.89 ***	2.81***
Goodness of fit	AIC	3719.8	3968.6	2698
	log likelihood	-1848.9	-1973.1	-1337

In Model 1 and Model 2, ★ indicates the first visited location in the trip chain. In Model 3, ★ represents the major destination in the trip chain, defined as the one with longer length of stay. In Model 1 and Model 2, ● indicates the secondly visited location in the trip chain. In Model 3, ● represents the minor destination in the trip chain, defined as the one with shorter length of stay.

the travel time between the major destination and work, all else equal, is associated with a 73% drop of the likelihood of selecting the trip chain, suggesting the influence of one's familiarity with the destinations on two-destination choice. We further test a new independent which is the multiplication of the trip chain's travel time and travel time between the major destination and work. Its elasticity carries a negative sign, suggesting that the farther away a trip chain is from the axis between home and work, the less attractive the trip chain is. All variables' elasticities are statistically significant except speed discontinuity.

6.8 Discussion

There is a lack of research that explicitly models two-destination choice and empirically addresses the problems of constructing choice sets and deciding the choice set size. This

Table 6.6: Modeling home-based, two-destination choice in the context of trip chains (excluding the entropy measure)

Model type		Mixed-effects logit model		
Model's name		Model 1	Model 2	Model 3
Choice set generation method		Approach I	Approach II	Approach III
Choice set size		60	60	60
Land use Location 1 ★	Accessibility ($\ln(A_j)$)	0.02	-0.06	-0.26 **
	Diversity of services ($\ln(H_j)$)	9.10 ***	9.07 ***	10.10 ***
Land use Location 2 ●	Accessibility ($\ln(A_k)$)	0.40 ***	0.28 ***	0.81 ***
	Diversity of services ($\ln(H_k)$)	0.18	0.27	0.33
Comparing two locations	Similarity ($\ln(\Xi_{j,k})$)	-12.36 ***	-12.36 ***	-13.92 ***
Travel/network features	Turn index ($\vartheta_{j,k}$)	-0.96 ***	-0.87 ***	-0.71***
	Speed discontinuity ($\psi_{j,k}$)	0.60 ***	0.5 ***	0.70 ***
	Time saving ratio ($\zeta_{j,k}$)	7.40 ***	6.87 ***	8.79***
Axis of travel	Time from home × time to work (\ln form)	-0.30 ***	-0.50 ***	-1.51 ***
Goodness of fit	AIC	1151	1311	1145
	log likelihood	-564.8	-644.6	-560.9

In Model 1 and Model 2, ★ indicates the first visited location in the trip chain. In Model 3, ★ represents the major destination in the trip chain, defined as the one with longer length of stay. In Model 1 and Model 2, ● indicates the secondly visited location in the trip chain. In Model 3, ● represents the minor destination in the trip chain, defined as the one with shorter length of stay.

Table 6.7: Elasticity of the likelihood of destination choice for key independent variables (with 1% change)

Variable	Elasticity of odds of selection (%)	Rank by absolute value
Travel time	-95	1
Time to work	-73	2
Time from home × time to work	-69	3
Accessibility (minor dest.)	58	4
Similarity	-48	5
accessibility (major dest.)	34	6
Entropy (minor dest.)	28	7
Turn index	-24	8
Entropy (major dest.)	14	9
Time to downtown	-11	10
Travel time saving ratio	5	11
Speed discontinuity	1.7	12

research proposes a new framework that explicitly considers non-work, two-destination choice in a home-based trip chain. Based on the new method that combines survival analysis and random sampling to form choice set proposed in Chapter 5, we further introduce and empirically test three approaches to build choice sets for two-destination trip chains. The mixed-effects logit model framework is utilized to estimate the coefficients. Our results show that all three models largely produce reasonable goodness of fit. The selection approach based on notion of major/minor destinations produces the highest goodness of fit of all three approaches. Therefore, we recommend Approach III for modeling two-destination choice in a trip chain.

This research investigates the effects of various land use and transportation network measures on destination choice in the context of trip chains. The key findings are summarized as follows:

1. Total travel time and travel time between the major destination and work influence the attractiveness of a two-destination trip chain the most.
2. Greater accessibility and diversity of services at the two destinations make them more attractive. The land uses at the two destinations do not exert the same level of influence.
3. Two destinations that are more dissimilar are more likely to be selected, which may be due to greater complementarity of services one may engage in by chaining the trips.
4. Route-specific network measures impact destination choice. The travel route with fewer turns per unit time makes a trip chain more attractive.
5. A trip chain producing a higher travel time saving ratio is more attractive to travelers.

Chapter 7

Conclusions

This dissertation contributes to current research on the connections between non-work travel behavior and the built environment. In Chapter 1, we propose a conceptual framework to illustrate the relationships among transportation networks, retail distribution patterns, and non-work destination choice. We posit that the built environment influences individuals' amount of travel and destination choice in the context of trip chains. A review of literature shows research niches in these areas. The key questions are: (1) How can we measure or find proxies for the built environment for non-work trips? (2) How can we model and verify the relationships between the built environment factors and non-work travel behavior using fine-scale spatio-temporally explicit data? Chapters 4, 5, and 6 address these questions using the in-vehicle GPS data.

The in-vehicle GPS travel data in the Twin Cities collected by the Nexus Research Group at the University of Minnesota make it possible to empirically tackle these questions. While the data were collected for another research purpose (individuals' work trip route choice after the collapse of I-35 W bridge in Minneapolis), we are able to map all the travel routes onto land use data for our research purposes. We start to analyze the data by defining home-based, non-work trips through matching some existing surveys with GIS-mapped trips.

Based on these definitions, Chapter 4 models non-work, non-home vehicle trip generation using a series of mixed-effects models. The area around home is divided into five driving zones: [0, 5) min, [5, 10) min, [10, 15), and [15, 20) min, where accessibility and

diversity of services are measured. The modeling results suggest that accessibility in these zones around home are not found to be significant in influencing non-work, non-home vehicle trip generation. The diversity of services in the [10, 15) min and [15, 20) min zones from home are negatively associated with the number of non-work, non-home vehicle trips, which implies individuals' multi-purpose shopping behavior. In terms of model structure, the mixed-effects ordered logit model generates the highest predictive accuracy and therefore is recommended for our data.

Chapter 5 focuses on modeling destination choice for home-based, non-work trips. We propose a new method that combines survival analysis and random selection to select alternative destinations and empirically examine the needed choice set size. In modeling, in addition to measuring land use around the destinations, we measure several route-specific transportation network parameters, such as turn index, speed discontinuity, and axis of travel, to quantify individuals' perception of reachability of destinations. The coefficients of land use and network measures are statistically significant and are consistent with our hypotheses.

Chapter 6 further expands Chapter 5 by modeling the home-based, two-destination choice problem. Three approaches to select alternative destinations are proposed and empirically compared. In addition to incorporating the measures used in Chapter 5, we measure the travel time saving ratio and similarity of destinations in a trip chain. The findings reveal that their effects on destination choice are statistically significant. The model structure and choice set construction process can be easily expanded to address the multi-destination choice problem.

The findings shed light on transportation and land use planning in several aspects. First, shaping an attractive retail zone needs careful planning. It not only concerns increasing the number of services and diversity of services in the zone itself, but also may be related with the types of services in other destinations to provide complementary services. Second, travel routes' network structure influences where people drive to shop. Third, a major destination's land use characteristics influence destination choice more than a minor destination in a trip chain. It may be speculated that if one's major destinations of interest are all far away from home, one has to drive a long distance for non-work trip purposes.

It provides food for thought when it comes to the policy question of how to reduce the amount of vehicle travel.

This research can be expanded in several directions:

1. It is meaningful to expand our choice set building process by understanding individuals' search radii for non-work trips. This may be done through online surveys and simulation-based surveys. By modeling randomly generated hypothetical scenarios, we may infer the search radii for various purposes by each subject.
2. When there are more data about the built environment, it would be of interest to develop other built environment measures (such as a destination's pleasantness, convenience of parking, availability and quality of service of alternative modes of transportation, stores' size, and products' quality) and to investigate whether they are related with destination choice.
3. The in-vehicle GPS data set only documents travel information from when a car's engine is on to when the engine is off. The data nonetheless cannot reveal which store one visits and through what path one reaches the store. Other GPS devices such as on-person devices or cell phone-based devices may provide insights into this realm. Such data, if available, can enable us to perform more in-depth analysis on travel behavior at the microscopic level.
4. In modeling destination choice we use 60 as the choice set size, which though seems satisfactory from the computational perspective may still be too large from a behavioral perspective. To address this issue, we may consider conduct surveys to ask individuals about the number of choices they typically consider for different trip purposes, and then test whether the model with a specific choice set size of interest can produce similar estimates and goodness of fit to the models with the choice set size we use in this research. If so, we may consider adopting a smaller choice set size.
5. This research assumes that all choices have the same choice set size in our models, which may not be the case in reality. For example, a person's choice set size for grocery shopping may well be different from the choice size for entertainment. And the choice size for the same purpose may vary by time of day, person, and location.

Further surveys and data analyses are needed to understand a spectrum of choice set formation behavior for non-work trips.

Bibliography

- ACNielsen (2002), ACNielsen study finds grocery stores continuing to lose share of customer shopping trips. In *ACNielsen News Release*. Retrieved in April 2011 from <http://www.acnielsen.com/news/american/us/2002/20020504.htm>.
- Adler, T. and Ben-Akiva, M. (1979), "A theoretical and empirical model of trip chaining behavior", *Transportation Research Part B: Methodological* , Vol. 13, pp. 243–257.
- Agyemang-Duah, K. and Hall, F. (1997), "Spatial transferability of an ordered response model of trip generation", *Transportation Research Part A: Policy and Practice* , Vol. 31, pp. 389–402.
- Allison, P. (2010), *Survival analysis using SAS: A practical guide*, SAS Institute Inc.
- Arentze, T. A., Oppewal, H. and Timmermans, H. J. (2005), "A multipurpose shopping trip model to assess retail agglomeration effects", *Journal of Marketing Research* , Vol. 42, pp. 109–115.
- Auld, J. and Mohammadian, A. (2011), "Planning-constrained destination choice in activity-based model", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 2254, pp. 170–179.
- Barmby, T. and Doornik, J. (1989), "Modelling trip frequency as a Poisson variable", *Journal of Transport Economics and Policy* , Vol. 23, pp. 309–315.
- Barros, C., Butler, R. and Correia, A. (2010), "The length of stay of golf tourism: A survival analysis", *Tourism Management* , Vol. 31, pp. 13–21.
- Becker, W. and Kennedy, P. (1992), "A graphical exposition of the ordered probit", *Econometric Theory* , Vol. 8, pp. 127–131.
- Ben-Akiva, M. and Lerman, S. R. (1985), *Discrete Choice Analysis*, Cambridge, MA: MIT press.
- Bernardin, V., Koppelman, F. and Boyce, D. (2009), "Enhanced destination choice models incorporating agglomeration related to trip chaining while controlling for spatial competition", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 2132, pp. 143–151.
- Bhat, C. (1998), "Analysis of travel mode and departure time choice for urban shopping trips", *Transportation Research Part B: Methodological* , Vol. 32, pp. 361–371.

- Bhat, C. and Guo, J. (2004), "A mixed spatially correlated logit model: Formulation and application to residential choice modeling", *Transportation Research Part B: Methodological*, Vol. 38, pp. 147–168.
- Boroditsky, L. (2000), "Metaphoric structuring: Understanding time through spatial metaphors", *Cognition*, Vol. 75, pp. 1–28.
- Boroditsky, L. and Ramscar, M. (2002), "The roles of body and mind in abstract thought", *Psychological Science*, Vol. 13, pp. 185–189.
- Brooks, C. M., Kaufmann, P. J. and Lichtenstein, D. R. (2004), "Travel configuration on consumer trip-chained store choice", *Journal of Consumer research*, Vol. 31, pp. 241–248.
- Brooks, C. M., Kaufmann, P. J. and Lichtenstein, D. R. (2008), "Trip chaining behavior in multi-destination shopping trips: A field experiment and laboratory replication", *Journal of Retailing*, Vol. 84, pp. 29–38.
- Burke, M. and Brown, A. (2007), "Distances people walk for transport", *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, Vol. 16, p. 16.
- Cameron, A. C. and Trivedi, P. K. (1998), *Regression analysis of count data*, number 30, Cambridge University Press.
- Cao, X. (2009), "Disentangling the influence of neighborhood type and self-selection on driving behavior: an application of sample selection model", *Transportation*, Vol. 36, pp. 207–222.
- Cao, X. J., Xu, Z. and Fan, Y. (2010), "Exploring the connections among residential location, self-selection, and driving: Propensity score matching with multiple treatments", *Transportation research part A: policy and practice*, Vol. 44, pp. 797–805.
- Cao, X., Mokhtarian, P. L. and Handy, S. L. (2009), "The relationship between the built environment and nonwork travel: A case study of Northern California", *Transportation Research Part A: Policy and Practice*, Vol. 43, pp. 548–559.
- Carrion, C. (2010), *Travel Time Perception Errors: Causes and Consequences*. Ph.D. Dissertation, Department of Civil Engineering, University of Minnesota.
- Carrion, C. and Levinson, D. (2012), A model of bridge choice across the Mississippi river in Minneapolis, in D. Levinson, H. Liu and M. Bell, eds, 'Network Reliability in Practice', Vol. 4, Springer, pp. 115–129.
- Casas, J. and Arce, C. (1999), Trip reporting in household travel diaries: A comparison to gps-collected data, in 'Proceedings of 78th Annual Meeting of the Transportation Research Board, Washington, DC'.
- Cervero, R. (1991), "Land uses and travel at suburban activity centers", *Transportation Quarterly*, Vol. 45, pp. 479–491.
- Cervero, R. and Kockelman, K. (1997), "Travel demand and the 3Ds: Density, diversity, and design", *Transportation Research Part D: Transport and Environment*, Vol. 2, pp. 199–219.

- Cotrus, A., Prashker, J. and Shiftan, Y. (2005), "Spatial and temporal transferability of trip generation demand models in Israel", *Journal of Transportation and Statistics* , Vol. 8, pp. 37–56.
- Cox, D. (1959), "The analysis of exponentially distributed life-times with two types of failure", *Journal of the Royal Statistical Society. Seires B (Methodological)* , Vol. 21, pp. 411–421.
- Cox, D. and Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall, London.
- De Palma, A., Dunkerley, F. and Proost, S. (2010), "Trip chaining: Who wins who loses?", *Journal of Economics & Management Strategy* , Vol. 19, pp. 223–258.
- Dellaert, B., Arentze, T. and Timmermans, H. J. (2008), "Shopping context and consumers' mental representation of complex shopping trip decision problems", *Journal of Retailing* , Vol. 84, pp. 219–232.
- Derrible, S. and Kennedy, C. (2009), "Network analysis of world subway systems using updated graph theory", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 2112, pp. 17–25.
- Derrible, S. and Kennedy, C. (2010), "Characterizing metro networks: State, form, and structure", *Transportation* , Vol. 37, pp. 275–297.
- Dieleman, F., Dijst, M. and Burghouwt, G. (2002), "Urban form and travel behaviour: Micro-level household attributes and residential context", *Urban Studies* , Vol. 39, pp. 507–527.
- Draijer, G., Kalfs, N. and Perdok, J. (2000), "Global positioning system as data collection method for travel research", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1719, pp. 147–153.
- Emond, C. R. and Handy, S. L. (2012), "Factors associated with bicycling to high school: insights from davis, ca", *Journal of Transport Geography* , Vol. 20, pp. 71–79.
- Ewing, R. and Cervero, R. (2010), "Travel and the built environment", *Journal of the American planning association* , Vol. 76, pp. 265–294.
- Fotheringham, A. (1988), "Consumer store choice and choice set definition", *Marketing Science* , Vol. 7, pp. 299–310.
- Fotheringham, A. S. (1986), "Modelling hierarchical destination choice", *Environment and Planning A* , Vol. 18, pp. 401–418.
- Frank, L. and Pivo, G. (1994), "Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking", *Transportation research record* , Vol. 1466, pp. 44–44.
- Gastner, M. and Newman, M. (2006), "The spatial structure of networks", *The European Physical Journal B* , Vol. 49, pp. 247–252.
- Gutin, G. and Punnen, A. (2002), *The Traveling Salesman Problem and Its Variations*, Springer.

- Handy, S. (1996), "Understanding the link between urban form and nonwork travel behavior", *Journal of Planning Education and Research* , Vol. 15, ACSP, p. 183.
- Handy, S. and Clifton, K. (2001), "Evaluating neighborhood accessibility: Possibilities and practicalities", *Journal of Transportation and Statistics* , Vol. 4, pp. 67–78.
- Hansen, W. (1959), "How accessibility shapes land use", *Journal of the American Planning Association* , Vol. 25, pp. 73–76.
- Hess, P. (1997), " Measures of Connectivity - Comparison of pedestrian environments in Seattle and Bellevue Suburbs, Washington", *Places* , Vol. 11, pp. 58–65.
- Hoogendoorn-Lanser, S. and Van Nes, R. (2004), "Multimodal choice set composition: Analysis of reported and generated choice sets", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1898, pp. 79–86.
- Horni, A., Scott, D., Balmer, M. and Axhausen, K. (2009), "Location choice modeling for shopping and leisure activities with matsim", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 2135, pp. 87–95.
- Huang, A. and Levinson, D. (2009), Retail Location Choice with Complementary Goods: An Agent-Based Model, in J. Zou, ed., 'Complex Sciences, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering', Vol. 4, Elsevier.
- Huang, A. and Levinson, D. (2011), "Why retailers cluster: an agent model of location choice on supply chains", *Environment and Planning B: Planning and Design* , Vol. 38, pp. 82–94.
- Jang, T. Y. (2005), "Count data models for trip generation", *ASCE Journal of Transportation Engineering* , Vol. 131, pp. 444–450.
- Jiang, B., Yin, J. and Zhao, S. (2009), "Characterizing the human mobility pattern in a large street network", *Physical Review E* , Vol. 80.
- Kissling, C. (1969), "Linkage importance in a regional highway network", *Canadian Geographer* , Vol. 13, pp. 113–127.
- Kitamura, R. (1984), "Incorporating trip chaining into analysis of destination choice", *Transportation Research Part B: Methodological* , Vol. 18, pp. 67–81.
- Kitamura, R., Chen, C. and Pendyala, R. (1997), "Generation of synthetic daily activity-travel patterns", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1607, pp. 154–162.
- Krizek, K., Iacono, M., El-Geneidy, A., Liao, C. and Johns, R. (2009), Access to destinations: Application of accessibility measures for non-auto travel modes. Access to Destinations Study Series Report: Mn/DOT 2009-24, University of Minnesota.
- Kwan, M. and Hong, X. (1998), "Network-based constraints-oriented choice set formation using GIS", *Geographical Systems* , Vol. 5, pp. 139–162.

- Lambert, D. (1992), "Zero-inflated poisson regression, with an application to defects in manufacturing", *Technometrics* , Vol. 34, pp. 1–14.
- Leclerc, F., Schmitt, B. H. and Dube, L. (1995), "Waiting time and decision making: Is time like money?", *Journal of Consumer Research* , pp. 110–119.
- Leszczyc, P., Sinha, A. and Timmermans, H. (2000), "Consumer store choice dynamics: An analysis of the competitive market structure for grocery stores", *Journal of Retailing* , Vol. 76, pp. 323–345.
- Levinson, D. and El-Geneidy, A. (2009), "The minimum circuitry frontier and the journey to work", *Regional Science and Urban Economics* , Vol. 39, pp. 732–738.
- Levinson, D. and Huang, A. (2012), "A positive theory of network connectivity", *Environment and Planning B: Planning and Design* , Vol. 39, pp. 308–325.
- Levinson, D. and Krizek, K. (2008), *Planning for Place and Plexus: Metropolitan Land use and Transport*, New York: Routledge.
- Li, H. (2004), Investigating Morning Commute Route Choice Behavior Using Global Positioning Systems and Multi-day Travel Data, PhD thesis, Georgia Institute of Technology, GA.
- Li, H., Guensler, R. and Ogle, J. (2005), "Analysis of morning commute route choice patterns using global positioning system-based vehicle activity data", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1926, pp. 162–170.
- Li, H., Guensler, R., Ogle, J. and Wang, J. (2004), "Using global positioning system data to understand day-to-day dynamics of morning commute behavior", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1895, pp. 78–84.
- Lieberman, S. (1969), "Measuring population diversity", *American Sociological Review* , Vol. 34, pp. 850–862.
- Lim, K. and Srinivasan, S. (2011), "Comparative analysis of alternate econometric structures for trip generation models", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 2254, pp. 68–78.
- Ma, J. and Goulias, K. (1999), "Application of Poisson regression models to activity frequency analysis and prediction", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1676, pp. 86–94.
- McFadden, D. (1978), Modeling the choice of residential choice, in A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, eds, 'Spatial Interaction Theory and Planning Models', Amsterdam: North Holland, pp. 75–96.
- McGuckin, N. and Murakami, E. (1999), "Examining trip-chaining behavior: Comparison of travel by men and women", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1693, pp. 79–85.

- Nerella, S. and Bhat, C. (2004), "Numerical analysis of the effect of sampling of alternatives in discrete choice models", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1894, pp. 11–19.
- Newman, J. and Bernardin, V. (2010), "Hierarchical ordering of nests in a joint mode and destination choice model", *Transportation* , Vol. 37, pp. 677–688.
- Paez, A., Scott, D., Potoglou, D., Kanaroglou, P. and Newbold, K. (2007), "Elderly mobility: Demographic and spatial analysis of trip making in the hamilton cma, canada", *Urban Studies* , Vol. 44, pp. 123–146.
- Parthasarathi, P. (2011), Network structure and travel. Ph.D. Dissertation, Department of Civil Engineering, University of Minnesota, Minneapolis.
- Parthasarathi, P., Hochmair, H. and Levinson, D. (2012), "Network structure and spatial separation", *Environment and Planning Part B* , Vol. 39, pp. 137–154.
- Pellegrini, P., Fotheringham, A. and Lin, G. (1997), "An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models", *Papers in Regional Science* , Vol. 76, pp. 257–284.
- Pinjari, A., Pendyala, R., Bhat, C. and Waddell, P. (2007), "Modeling residential sorting effects to understand the impact of the built environment on commute mode choice", *Transportation* , Vol. 34, pp. 557–573.
- Pozsgay, M. and Bhat, C. R. (2001), "Destination choice modeling for home-based recreational trips: Analysis and implications for land use, transportation, and air quality planning", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1777, pp. 47–54.
- Quddus, M., Noland, R. and Ochieng, W. (2005), "Validation of map matching algorithms using high precision positioning with gps", *Journal of Navigation* , Vol. 58, pp. 257–271.
- Quddus, M., Ochieng, W. and Noland, R. (2007), "Current map-matching algorithms for transport applications: State-of-the art and future research directions", *Transportation Research Part C: Emerging Technologies* , Vol. 15, pp. 312–328.
- Raghubir, P., Morwitz, V. G. and Chakravarti, A. (2011), "Spatial categorization and time perception: Why does it take less time to get home?", *Journal of Consumer Psychology* , Vol. 21, pp. 192–199.
- Rashidi, T., Auld, J. and Mohammadian, A. (2012), "A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection", *Transportation Research Part A: Policy and Practice* , Vol. 46, pp. 1097–1107.
- Recker, W. W. and Kostyniuk, L. P. (1978), "Factors influencing destination choice for the urban grocery shopping trip", *Transportation* , Vol. 7, pp. 19–33.

- Schimek, P. (1996), "Household motor vehicle ownership and use: How much does residential density matter?", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1552, pp. 120–125.
- Schoner, J. (2013), *Catalysts and Magnets: Built Environment Effects on Bicycle Commuting*. Master Thesis, Department of Civil Engineering, University of Minnesota, Minneapolis.
- Schwanen, T., Dijst, M. and Dieleman, F. (2004), "Policies for urban form and their impact on travel: the Netherlands experience", *Urban Studies* , Vol. 41, p. 579.
- Scott, D. and Horner, M. (2008), "Examining the role of urban form in shaping people's accessibility to opportunities: An exploratory spatial data analysis", *Journal of Transport and Land Use* , Vol. 1, pp. 89–119.
- Scott, D. M. and He, S. Y. (2012), "Modeling constrained destination choice for shopping: a gis-based, time-geographic approach", *Journal of Transport Geography* , Vol. 23, pp. 60–71.
- Shannon, C. E. (1948), "A mathematical theory of communication", *Bell System Technical Journal* , Vol. 27, pp. 379–423, 623–656.
- Shocker, A., Ben-Akiva, M., Boccara, B. and Nedungadi, P. (1991), "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions", *Marketing letters* , Vol. 2, pp. 181–197.
- Spiggle, S. and Sewall, M. (1987), "A choice sets model of retail selection", *The Journal of Marketing* , Vol. 51, pp. 97–111.
- Swait, J. (2001), "Choice set generation within the generalized extreme value family of discrete choice models", *Transportation Research Part B: Methodological* , Vol. 35, pp. 643–666.
- Swait, J. and Ben-Akiva, M. (1987), "Incorporating random constraints in discrete models of choice set generation", *Transportation Research Part B: Methodological* , Vol. 21, pp. 91–102.
- Taaffe, E., Gauthier, H. and Morton, E. (1996), *Geography of Transportation*, Upper Saddle River, NJ: Prentice-Hall.
- Thill, J. and Horowitz, J. (1997), "Travel-time constraints on destination-choice sets", *Geographical Analysis* , Vol. 29, pp. 108–123.
- Thill, J. and Horowitz, J. (2002), "Modelling non-work destination choices with choice sets defined by travel-time constraints", *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling and Computational Intelligence* , Springer, pp. 186–208.
- Timmermans, H. (1996), "A stated choice model of sequential mode and destination choice behaviour for shopping trips", *Environment and Planning A* , Vol. 28, pp. 173–184.
- Tversky, A. and Kahneman, D. (1991), "Loss aversion in riskless choice: A reference-dependent model", *The Quarterly Journal of Economics* , Vol. 106, pp. 1039–1061.

- Wagner, D. P. (1997), "Lexington area travel data collection test: GPS for personal travel surveys", *Final Report, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus* .
- Wang, L. and Lo, L. (2007), "Immigrant grocery-shopping behavior: Ethnic identity versus accessibility", *Environment and Planning A* , Vol. 39, p. 684.
- Willumsen, L. G. and Ortuzar, J. (2001), *Modeling Transport*, New York: John Wiley & Sons.
- Wolf, J., Oliveira, M. and Thompson, M. (2003), "Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 1854, pp. 189–198.
- Xie, F. and Levinson, D. (2007), "Measuring the structure of road networks", *Geographical Analysis* , Vol. 39, pp. 336–356.
- Xie, F. and Levinson, D. (2011), *Evolving Transportation Networks*, Springer.
- Zhang, L. and Levinson, D. (2008), "Determinants of route choice and the value of traveler information: A field experiment.", *Transportation Research Record: Journal of the Transportation Research Board* , Vol. 2086, pp. 81–92.
- Zhu, S. (2010), *The Roads Taken: Theory and Evidence on Route Choice in the Wake of the I-35 W Mississippi River Bridge*. Ph.D. Dissertation, Department of Civil Engineering, University of Minnesota.
- Zolfaghari, A., Sivakumar, A. and Polak, J. (2012), "Choice set pruning in residential location choice modelling: A comparison of sampling and choice set generation approaches in greater london", *Transportation Planning and Technology* , Vol. 35, pp. 87–106.