Estimating a Toronto Pedestrian Route Choice Model using Smartphone GPS Data: It's not the destination, but the journey, that matters

By

Gregory Lue

A thesis submitted in conformity with the requirements

For the degree of Master of Applied Science

Graduate Department of Civil Engineering

University of Toronto

Estimating a Toronto Pedestrian Route Choice Model using Smartphone GPS Data: It's not the destination, but the journey, that matters

Gregory Lue

Master of Applied Science

Graduate Department of Civil Engineering

University of Toronto

2017

Abstract

City planning has an emphasis on working towards creating walkable cities with boulevards, wide sidewalks, and social spaces. This study uses revealed preference GPS data collected through a smartphone-based travel survey and discrete choice modelling techniques to determine pedestrians' preferences towards street infrastructure, built environment, and land use. A path size logit model with stochastic route choice generation choice set was used for this model. The results of the model showed that distance, the number of turns, the number of signalized intersections, and distance along links with sidewalks on both sides of the street were significant variables in the route choice model. Turns are found to be equivalent to an additional 32m, signalized intersections are equivalent to a reduction of 34m, and travel along streets with sidewalks on both sides of the road is perceived as 33% shorter than streets with other sidewalk conditions.

Acknowledgments

Transportation engineering is the road I chose and I do not regret it. Even though this thesis was no walk in the park, it was still enjoyable. Just like a road during rush hour, this thesis has had its stop and go moments. I would like to take this time to thank those that supported me throughout this time. I want to especially thank my supervisor Eric Miller who gave me much guidance on this project. I would also like to thank Waterfront Toronto for the smartphone data. Thank you, Chloe, for keeping me on track and putting up with me. Thanks to my friends in the ITS lab for help and support.

Also, thanks to anyone who decides to read this. I put a lot of time and effort into this project and it makes me happy to know that someone took the time to look at it.

Table of Contents

1. Introduction ... 1
   1.1. Study Objective ... 1
   1.2. Study Motivation ... 1
   1.3. Thesis Structure ... 4
2. Literature Review ... 5
   2.1. Built Environment Measures ... 5
   2.2. Built Environment and Pedestrian Route Choice ... 6
   2.3. Choice Models ... 13
      2.3.1. Choice Set Generation ... 13
      2.3.2. Route Choice Models ... 17
   2.4. GPS Point Map-Matching ... 24
3. Network Data, Street Attributes, and Smartphone-Based Travel Survey ... 25
   3.1. Data Sources ... 25
   3.2. Smartphone Travel Survey ... 27
   3.3. Cleaning Data ... 29
      3.3.1. Network Cleaning and Merging Data ... 29
      3.3.2. Land Use data ... 31
      3.3.3. Walk Trip Data ... 35
4. Methodology ... 40
   4.1. Alternative Route Generation Algorithm ... 40
   4.2. Path Size Logit Choice Model ... 43
      4.2.1. Path Size Factor Sensitivity Analysis ... 44
   4.3. Model Testing Procedure ... 45
   4.4. Limited Observations per Individual vs Full Dataset ... 46
5. Toronto Case Study: Model Specification and Results ... 52
   5.1. Toronto Walk Trip Data ... 52
   5.2. Route Variables ... 55
   5.3. Model Results ... 58
   5.4. Path Size Factor Analysis ... 62

5.5. Route Generation Analysis 64
5.6. Discussion 65
5.6.1. Model results 65
5.6.2. GPS Data 74
5.6.3. Route Generation Analysis 76
6. Recursive Logit (RL) 78
7. Conclusions 79
8. References 81

List of Figures

Figure 2.1 - Sampling correction term sensitivity analysis 22
Figure 3.1 - Walk Score map 31
Figure 3.2 - Land use map comparison 33
Figure 3.3 - Land use segment matching 34
Figure 3.4 - Street segment composition 34
Figure 3.5 - Large gap subway trip 36
Figure 3.6 - Map-matching process 38
Figure 3.7 - Map-matching issues 39
Figure 4.1 - Phi sensitivity analysis 45
Figure 4.2 - Trip observation frequency 47
Figure 4.3 - Dataset size with limited observations 48
Figure 4.4 - Change in mean distance with observation limit 48
Figure 4.5 - Maximum percent share from an individual 48
Figure 4.6 - Change in dataset characteristics with observation limit 49

List of Tables

Table 3.1 - Walk Score Descriptions 26
Table 3.2 - Mode detection success rate (Harding et al., 2015) 29
Table 4.1 - Change in demographics with observation limit 49
Table 5.1 - Toronto network characteristics 52
Table 5.2 - Observed walk trip characteristics 53
Table 5.3 - Demographic characteristics of user sample 54
Table 5.4 - Alternative route characteristics 55
Table 5.5 - Route variables 56
Table 5.6 - Interaction variables 57
Table 5.7 - Final general model results 60
Table 5.8 - Final interaction term model 60
Table 5.9 - Distance equivalent of attributes for general model 61
Table 5.10 - Distance equivalent of attributes for interaction term model 61
Table 5.11 - Path size factor formulation summary 62
Table 5.12 - Path size factor formulation characteristics 63
Table 5.13 - Observed route draw probability 64
Table 5.14 - Observed route characteristics 65
Table 5.15 - Average difference between observed route and choice set characteristics 65

List of Appendices

Appendix A – Stochastic route generation process

Appendix B – Observed route characteristics

Appendix C – Observed route variable correlation matrix

Appendix D – Model estimation parameters

Appendix E – Distribution of path size factor

1. Introduction

1.1. Study Objective

The purpose of this study is to use revealed preference Global Positioning System (GPS) data collected through a smartphone-based travel survey and discrete choice modelling techniques to develop a pedestrian route choice model and determine pedestrians' preferences towards street infrastructure, built environment, and land use.

1.2. Study Motivation

There are four main motivations for this study:

1. Travel demand models overlook walking trip routes or have a limited representation of the routes. This research aims to develop a pedestrian route choice model suitable for use in travel demand models as well as providing a measure of accessibility for other models, such as transit access models.
2. City planning supports building walkable streets but the measures of attributes are often qualitative, not quantitative. This research provides quantitative information about walking preferences of pedestrians.
3. Smartphone-based GPS travel surveys are becoming more common for data collection. One of the motivations of this research is to test if smartphone GPS data are accurate enough to develop a meaningful route choice model
4. Explore the route choice set generation problem.

Travel demand models often overlook walking trip routes or have a limited representation of the routes (Clifton, Singleton, Muhs, & Schneider, 2016; Singleton & Clifton, 2013). A study by Singleton & Clifton (2013) found that few regional travel demand models of metropolitan planning organizations (MPOs) modeled non-motorized trips and out of the 48 MPOs included in the study, none of them assigned walking trips to the network. There is a noticeable lack of representation of pedestrian activity in travel demand models and their inclusion seems to be the logical next step for travel demand model development. Pedestrian route choice can provide travel demand models with more detailed measurements of accessibility, safety, and health.

Accessibility for pedestrians is often measured as the shortest path between origin and destination, however, studies have shown that pedestrian trips do not simply follow the shortest path and that there are other factors/considerations affecting the route choice (Borst, Miedema, de Vries, Graham, & van Dongen, 2008). This research looks to develop a pedestrian route choice model that can predict pedestrian flows along the network by considering preferences to street characteristics.

City planning has an emphasis on working towards creating walkable cities with boulevards, wide sidewalks, and social spaces (parks with benches, plazas). In general, complete sidewalks, parks, window shops are seen as attractive street attributes (Mehta, 2008). These qualities are believed to make a street more walkable, however; there is a difference between having a "walkable" street and a street where people walk. There are certain characteristics that are preferable for walking routes but the issue is that pedestrians might only be willing to deviate a small distance from the shortest route to experience them. There could be a beautiful boulevard with wide sidewalks, street furniture, large tree canopy, and window shops, but people may not go out of their way to walk along it because it requires a lengthy detour.

Some of the main factors affecting walkability are mixed use land development and high-density residential developments. These land uses create origins and destinations that are close and walkable but it is unclear whether people deviate from the shortest path to traverse streets with these land uses. Land use, street network density, and sidewalk infrastructure are seen to improve walking rates of neighbourhoods, but it is less clear whether these aspects influence what route is taken once an individual decides to walk.

Singleton and Clifton (2013) identified that the main barriers to representing non-motorized/walk travel in travel demand models were limited records of walking or cycling trips from household travel surveys and limited resources for data collection. The field of data collection is rapidly changing with the introduction of smartphones. Smartphones are ubiquitous in today's developed society and smartphone app travel diaries provide the possibility of a low-cost travel survey method. Many people always carry their smartphones, and these smartphones have GPS and accelerometer capabilities that permit tracking of their travel behaviour. However, battery life concerns prevent smartphone GPS data from being as detailed as dedicated GPS loggers. This study investigates whether the frequency and accuracy of a Smartphone GPS travel diary are sufficient for developing a pedestrian route choice model.

One of the challenges with developing a pedestrian route choice model is generating the alternative routes considered by the individual. The methods of generating alternative route choices can be seen as arbitrary and inaccurate. Often times, the observed route is not even generated through these processes. If the method cannot generate the observed route, it may be possible that the method is not even generating likely/probable alternative routes. This thesis explores a stochastic alternative route generation method and analyzes is the ability to generate the observed route.

The outcomes from this research are as follows:

- Develop a pedestrian route choice model
- Provide information about walking preferences of pedestrians that could be used for pedestrian infrastructure planning, safety analysis, or pedestrian health impact studies.
- Provide a measure of accessibility for transit or land use development by predicting walking routes.
- Determine if passive/background smartphone app GPS points provide a sufficient amount of information about walking trip route choice.
- Assess a stochastic alternative route generation method and its ability to generate the observed route.

1.3. Thesis Structure

In addition to this introduction section, this thesis follows the following structure.

Chapter 2 is the literature review

Chapter 3 discusses the network data, street attributes, and smartphone-based travel survey

Chapter 4 outlines the methodology

Chapter 5 talks about the Toronto route choice model specifications and the results

Chapter 6 explores the recursive logit model and the results

Chapter 7 discusses the limitations of the study and areas of future work

2. Literature Review

2.1. Built Environment Measures

Built environment is defined as the buildings, transportation systems, open space and land-use that support communities and impact human health (City of Toronto, 2016). The influence of built environment on active transportation has been studied from various perspectives: pedestrian health, transportation accessibility, and land use development. However, the results of these studies have been mixed. Badoe and Miller (2000) reviewed the transportation-land-use interaction in North America and found that the results were not always consistent. Some studies conclude that urban density, neighbourhood design, and land-use mix all influence variables such as auto ownership and mode choice, while other studies suggest that the effects are marginal or at times non-significant (Badoe & Miller, 2000).

A possible explanation for the contradictory results is that the measures of built environment have not been consistent. The built environment can be measured using three categories: perceived measures, observed measures, and geographic measures. Perceived measures are an individual's perception of the built environment. Data on perceived measures are collected via a survey (telephone, online, in person, etc). Perceived built environment attributes include aesthetics, sounds, and safety. Safety is largely a perception measure but can also be quantified through pedestrian-vehicle collisions, the amount of reported crime, or degree of street lighting. Perceived measures of the built environment provide valuable information about the connection between built environment and travel behaviour. Knowing the perception of built environment can help policy makers and street designers make informed decisions about where to invest money/funds. However, the issues with perceived measures of built environment are that the surveys are often lengthy (which may decrease response rates) and the results can be difficult to incorporate in transportation models.

Observed measures typically evaluate the physical features of the environment. Examples of observed measures are sidewalk width, street slope/grade, land use frontage, building architecture, or street design. Observational measure data are usually collected through trained observers traversing through neighborhoods to observe street characteristics. Observational measure data are useful for filling in information that is not present in the geographic measure data. Observers can also be trained to look for specific characteristics important for modelling. Some of the challenges with observed measures are that the observers must be properly trained to ensure consistent data and data collection can involve a large amount of manual labour.

Geographic measures usually evaluate characteristics of the area such as population density, land-use composition, and street network. This data can be collected from numerous sources on the internet or from various agencies, such as a city or municipal departments. Geographic measures are typically used in geographic information system (GIS) programs for analysis. One of the challenges with geographic measures is acquiring consistent sets of data. While databases of geographic data can be found online or from agencies, there is no single database that holds the various types of data. For example, land use, population densities, and elevations may be found from different sources. Another issue is that the data may not be updated on a regular basis so population data may be of a different year than the land use data. The varying dates of the data can cause inconsistencies within the analysis area. For example, a land use dataset may label a parcel as agricultural but more recent population data may suggest a medium density residential development occupies the parcel. However, geographic measures are a good way of categorizing areas or neighbourhoods. These measures provide an aggregated overview of neighborhood characteristics.

Built environment measures have varying scales at which they are assessed. Geographic measures are typically recorded at a zonal level. The size of these zones can include individual parcels, neighbourhoods, traffic zones, or census tracts depending on the spatial detail of the measure. Perceived measures are typically recorded on a route level since the data comes from the travellers' point of view. Observed measures can be either a zonal level or route level. Some of the observations may be applied to an area as a neighborhood attribute, while some observations may only refer to specific street characteristics. For route choice, there is some debate to how much the zonal attributes influence the choice.

2.2. Built Environment and Pedestrian Route Choice

Studies that look into built environment and pedestrian walking often combine the journey with the destination (Broach & Dill, 2015). It may be stated that pedestrians like to walk to get to a certain type of built environment (dense street networks, wide sidewalks) but pedestrians experience the built environment along the routes in a linear measure (Rodríguez et al., 2015). There are many studies on how neighborhood design, built environment, and land use affect active transportation. But these studies typically relate to the choice of walking as a mode of transportation. There are fewer studies done on the effect of built environment on the route choice of walking pedestrians.

Active Transportation and Built Environment Effects at an Area Scale

Land use mix, development density, urban green foliage, and street connectivity are often associated with higher active transportation rates and flows. For example, Agrawal and Schimet (2007) found that higher population densities are associated with higher walking rates. Sarkar et al. (2015) found that increased density of trees along the street and higher street connectivity increase the likelihood of walking. While the studies point to land use mix, development density, and street connectivity having a positive correlation with walking, the results on which has the greatest impact are not consistent.

A study by Cervero and Kockelman (1997) states that retail and commercial land use within close proximity to residential neighbourhoods has a stronger impact on mode choice for non-work trips than density.

However, when Ozbil et al. (2011) explored the link between street connectivity, land use and pedestrian flows it was found that development density is a key factor in determining pedestrian volumes while the spatial structure explains the distribution of the pedestrian flow. An interesting finding from this study is that land use mix had limited explanatory power for predicting pedestrian flow. The results suggest that the distribution of pedestrian movements often relate more to a spatial hierarchy of streets rather than land use. Ozbil et al. (2011) explain that the connection between street connectivity and land use is that street network design lays the long-term framework within which land uses change over time.

Lamiquiz and Lopez-Dominquez (2015) examined the association between built environment and walking at a neighbourhood scale. The results of their analysis showed that street network and built environment factors are associated with the amount of walking in urban spaces. Similar to the study by Ozbil et al. (2011), the street configuration was found to be more influential than development density or land use mix. However, unlike Ozbil et al. (2011), land use factors were found to be significant predictors of walking trips at the neighborhood level.

Lin and Hsia (2007) analyzed the effects of the built environment on active transportation in Taipei. The study concluded that mixed-use urban areas positively correlated with daily physical activity and negatively correlated with commuting activity. Neighbourhood environment quality and the presence of footpaths and walkways are positively correlated with active transportation. Contrary to most studies, this study found that transit access was negatively correlated with commuting.

A cross-sectional study in North Carolina investigated if an “active community environment” is promoted through land use and transportation planning. An active community environment combines high-density developments, mixed land use, and a connected system of walkways, bikeways, and transit. The study tested if active community environment scores were associated with active transportation and physical activity. Results found that more favourable community environment score was significantly associated with leisure physical activity and active transportation (Aytur, Rodriguez, Evenson, Catellier, & Rosamond, 2007).

There are some studies that suggest that travel behaviour may be less of a result of neighborhood characteristics and may be a pre-existing preference when choosing residence location. In other words, people who choose to take up residence in walkable neighbourhoods may already have a preference towards active transportation. For example, a study in Finland by Haybatollahi et al. (2015) explored the idea that people’s residential and travel choices may be indicators of their travel behaviour and that grouping individuals by neighborhood preferences may distinguish characteristics of the living environment. The results of the study supported the notion that neighbourhood preferences can be associated with active travel behaviours. By clustering, residents based on neighbourhood preferences moderated the association between some features of density measurements and travel behaviour (Haybatollahi et al., 2015).

The terms ‘walkability’ and ‘walkable neighbourhood’ have become popular in urban development and transportation. Walkability and walkable neighbourhoods are the result of land use mix, development density, urban green foliage, street connectivity, pedestrian infrastructure, and transit accessibility coming together to promote walking. However, walkability and walkable neighbourhoods have been defined in many different ways. Bauman et al.(2012) state that walkable neighbourhoods are designed such that residents can walk from home to nearby destinations. Manaugh and El-Geneidy (2011) define walkability as the match between residents' desires/expectations for various destinations and their willingness to walk a given distance and the quality of the required path. Frank et al. (2010) describe walkability as the proximity from home to non-residential destinations.

With the varying definitions of walkability, there have also been a variety of studies trying to measure walkability. Mehta (2008) performed a microscale analysis of the environmental qualities that make streets better for walking. The study examined physical, land-use, and social characteristics of the environment and pedestrians' behaviour and perceptions towards walking. Mehta (2008) identifies a hierarchy of walking needs as follows: feasibility, accessibility, usefulness, safety, comfort, sensory pleasure, and sense of belonging. Feasibility relates to mobility, time and other responsibilities of the individual when deciding to make a walking trip. Accessibility is one's ability to access the destination considering distance, street connectivity, and barriers. Usefulness is described as the state of the environment and its ability to meet the individual's day-to-day needs such as shopping, eating, or entertainment (Mehta, 2008). Safety relates to the sense of real and perceived safety. This includes crime, street configuration, traffic, or presence of people/stores. Sensory pleasure is achieved through various environmental stimuli from buildings, streets, shop windows, signage, or people. The sense of belonging describes the community feeling brought about by a neighbourhood. This could be in the form of public spaces, buildings, or culture within a neighbourhood. The results of the study found that physical improvements such as path connectivity, wide sidewalks, and trees help to make streets more pedestrian friendly while the variety of businesses and community gathering places that act as destinations work to draw pedestrians to the area.

Walk Score is a common measure of the walkability of an area. Walk Score is calculated based on proximity to amenities, with maximum points for amenities within a 5-minute walk and no points for amenities past a 30-minute walk. Walk Score also takes into consideration pedestrian friendliness of an area through measures of population density, block length, and intersection density (Walk Score, 2016). However, the exact weightings and methods for calculating Walk Score are proprietary and are not disclosed to the public. Walkability measures are generally used to study the impacts of the built environment on pedestrian behaviour. Walk Score is an accessible and free resource to use which makes it a common measure for walkability.

Carr et al. (2010) conducted a study on Walk Score as an estimate for neighborhood walkability. Neighborhood walkability was measured based on street connectivity, residential density, access to public transit provisions, and crime. The results showed a positive correlation to on-street connectivity, residential density, and access to public transit provisions, however, Walk Score was also found to have a positive correlation with crime (Carr et al., 2010). The reason for a positive correlation between Walk Score and crime is that Walk Scores are typically higher in densely populated areas which are, in turn, more likely to have crime.

Walk Score only captures some of the elements of walkability which causes it to be inconsistent in predicting walking behaviour. In some cases there are neighbourhoods with similar Walk Scores but with very different rates of walking (Herrmann, Boisjoly, Ross, & El-Geneidy, 2017).

Herrmann et al. (2017) conducted a study to explore the discrepancy between Walk Score and observed walking behaviour by analyzing an origin-destination survey in Montreal, Canada. The results showed that open-space areas (areas devoid of buildings or natural terrains) such as parking lots, were found to be negatively associated with walking rates, even in areas with high Walk Score. Also, on-street tree coverage has a positive impact on walking rates even in neighborhoods with low Walk Scores. It was determined through a regression analysis that combining Walk Score with open-space and tree cover measures helped to explain discrepancies in walking rates between areas with similar Walk Score. The open-spaces decrease the density of amenities and absence of trees lack aesthetics and comfort which overall, impacted the walkability of the area. Herrmann et al. also determined that job density, transit density, and shop density are more explanatory of walking behaviour than Walk Score alone.

Active Transportation and Built Environment Effects at a Route Scale

Just like other modes of transportation, distance to the destination is one of the most important factors influencing walking route choice. If distance was the only factor pedestrians would be walking on the shortest path all the time, however, studies have shown that the shortest path may only be chosen around 20 percent of the time (Borst et al., 2008).

A study by Rodríguez et al. (2009) examined if micro-scale features of the built environment (width of the sidewalk, benches, trash bins, crossing aids: stoplights and crosswalks) have a relationship with street segment pedestrian activity. The study found greater pedestrian activity on segments with higher development intensity, mixed land uses, and more crossing aids. It was also found that street connectivity and pedestrian friendly aids are related to higher pedestrian counts (Rodríguez et al., 2009).

Various studies have examined which street characteristics are desirable and which are not. These studies are often qualitative and give an idea of which street attributes are attractive and which are seen as barriers.

A study by Borst (2008) on the relationship between street characteristics and perceived attractiveness for elderly residents in three Dutch urban districts found that low slopes, zebra crossings, trees, gardens, bus stops, business buildings, catering establishments, city centre, and traffic volume were all perceived as attractive attributes along a route. Conversely, litter, high-rise buildings, high neighborhood density were negatively related to perceived attractiveness (Borst et al., 2008).

Similar studies have found that wider sidewalks, higher density of intersections, higher density of pedestrian friendly parcels are all attractive attributes for a route while attributes such as large street crossings, poor lighting, litter, absence of people, or steep slopes act as deterrents (Ferreira, Johansson, Sternudd, & Fornara, 2016; Ferrer, Ruiz, & Mars, 2015; Guo, 2009).

Ferrer et al. (2015) examined built environment factors on walking trips in Valencia, Spain, with less than 45 minutes in duration. The study used three focus groups consisting of participants who undertook non-shopping trips during the week. Almost all participants mentioned sidewalk width, the presence of trees, and low traffic volumes as factors that they considered. The results from the study showed that factors relating to safety, such as poor lighting or absence of people, were strong deterrents for walking. Crossing large streets or roundabouts can be deterrents because of poor coordination between traffic signals causing long wait times. Ferrer et al. (2015) found that some variables that are attractive to some pedestrians, can be seen as obstacles to others. For example, sidewalk width can be seen as a barrier. While sidewalk cafes and bollards are often seen as aesthetic improvements, some of the participants found them to be obstacles.

Agrawal et al. (2008) collected stated route preferences and chosen routes from morning commuters at five rail stations in San Francisco and Portland, Oregon. It was found that a pedestrian's primary goal in choosing a route is to minimize distance and time, but safety, crossing delays, sidewalk conditions, a presence of other pedestrians were also highly rated factors.

Koh and Wong (2013a) also found that shops, good scenery, and crowdedness became significant between shortest and actual route suggesting that people are pulled towards these street attributes.

Another study by Koh and Wong (2013b) was conducted on how pedestrians' needs and behaviours changed based on different land use environments. The study interviewed pedestrians at transit stations in Singapore inquiring about their preferences for the first/last mile of their walking trips. Pedestrians were asked to rate the importance of 12 factors: distance, comfort, rain shelters, stairs/slopes, traffic accident risk, detour, crowded walkway, security, the number of road crossings/delay, shops along the route, good scenery, and directional signs. Rain shelters, traffic accident risk, and stairs/slopes were found to be significantly more important in residential areas than in mixed land use areas. For industrial areas, the top three factors were distance, security, and rain shelters. In general, factors relating to pedestrian infrastructure facilities were more important in the industrial land than in residential land. This suggests that pedestrian infrastructure facilities could be lacking in the industrial land (Koh & Wong, 2013b). Traffic accident risk was also very important in industrial areas which Koh and Wong suggest could be due to heavy vehicle traffic. Mixed land use responses were found to be very similar to residential area responses. For residential and mixed use areas, the top three factors where rain shelters, distance, and security. In residential areas, the top reason for deciding to walk was the distance. Availability of public transit and convenience of walking also influenced the decision to walk.

While there is a modest supply of qualitative research on pedestrian route choice and the built environment, there are fewer studies that quantitatively assess the perception of street characteristics from the pedestrian's point of view.

Guo (2009) used a path choice model to analyze the effect of pedestrian environment on the utility of walking. Subway commuters' paths from the station to their workplace in downtown Boston were modeled. The model showed that an increase of one more intersection per 100m increased utility by 0.3min, increasing sidewalks by 6ft increases utility by 0.5min, and people were willing to walk 2.9 minutes to avoid hilly topography (Guo, 2009).

Dill and Broach (2015) conducted a pedestrian route choice study using revealed preference GPS data in Portland, Oregon. The study considered distance, turns, steep upslope, substandard street, busy streets, commercial neighborhoods, unsignalized arterial crossings, and unmarked collector crossings as variables. Dill and Broach (2015) found that each additional turn was equivalent to about 50 meters, upslopes of 10 percent are seen as twice as costly as less steep ground and no real benefit was found for traversing off-street paths. In addition, results showed that pedestrians would be willing to travel 70 meters to avoid an unsignalized arterial path. Busy streets such as collector roads or larger were perceived as 14% longer while commercial neighborhoods were attractive for pedestrians, being perceived as 28% shorter.

2.3. Choice Models

2.3.1. Choice Set Generation

In path choice, alternative paths need to be generated based on decision rules, but these rules are difficult to define and validate (Hoogendoorn-Lanser, 2005). Methods for generating paths can be organized into two groups: deterministic approaches and stochastic approaches.

<u>Deterministic Approaches</u>

The *labelled paths* approach by Ben-Akiva et al. (1984) takes a large number of possible alternative routes and creates a choice set of labelled "optimal" routes. Each route in the choice set optimizes some criterion function which may include travel time, distance, scenery, congestion, etc. The functions for labelling are estimated by maximizing the proportion of observed routes included in the sets of labelled paths (Ben-Akiva et al., 1984). This method is meant to consider that different travellers may have varying objectives in seeking routes. A portion of travellers may prefer to minimize travel time while others may want to minimize the number of turns. Some travellers may prefer a scenic route while other travellers may choose to avoid congestion. These preferences translate into objective functions which are used to solve for the optimal route for each objective; although, it is possible for two objective functions to produce the same route.

The *link elimination* method by Azevedo, Costa, Madeira and Martins (1993) is a procedure used to determine a specified number of shortest paths. Once the shortest path is found, the shortest path is deleted so that it is not generated again. The modified network is solved again for the shortest path. This process is repeated until the desired number of paths is generated. This process may generate paths that are overly circuitous or may generate paths that are very similar to each other.

A *link penalty* method is an approach created by de la Bara, Perez, and Anez (1993) that solves the shortest path between the origin and destination and labels the links in that path. A penalty is added to the links already traversed so that when the shortest path is solved again, the shortest path algorithm will try to avoid links that have are already in a generated route. The variation between the generated routes varies with the penalty imposed. Instead of eliminating links from the network, the link penalty method increases the impedance of links that have already been used in the shortest path. The advantage of this approach over link elimination is that it allows for previously traversed links to be used in other routes. This helps to avoid the issue of overly circuitous routes.

Constrained k-shortest paths is a procedure used by Van der Zijpp and Catalano (2005) to generate a specified number of shortest paths with constraints such as a number of transfers in a multimodal trip or the sequence of modes taken. The algorithm developed by Van der Zijpp and Catalano greatly improved the computation time for the procedure. This approach can also minimize general cost instead of the shortest path.

Stochastic Approaches

The issue with deterministic approaches is that they assume perfect information about travel conditions and street characteristics. If this assumption were true, the observed paths would always be the path of least cost. However, numerous studies have shown that the observed route is not the shortest. The observed path can range from 20-70% longer than the shortest path (Borst et al., 2008). It has also been observed that people may change their route in response to changing street/road conditions. Stochastic approaches allow random effects to influence route choice set generation. Compared to the deterministic approach, there are much fewer stochastic alternative route generation methods.

Ramming (2002) uses a simulation method which generates paths by drawing link costs from probability distributions. A Gaussian distribution is used with the mean and standard deviation taken from a model of travel time perception. Bovy and Fiorenzo-Catalano (2007) modified Ramming's simulation method by adding a generalized cost function with random parameters and random attributes.

Freijinger (2007) uses a biased random walk algorithm to generate the alternative routes. The probability of choosing a link is based on the sum of the shortest path from the origin to the sink node of the link, the cost of the link, and the cost of the shortest path from the sink node of that link. This "link cost" is compared to the shortest path cost from the source of the link to the destination. The costs for all connected links at a node are compared to determine the probability of choosing each link. If the link is on the shortest path to the destination, it is more likely to be chosen. The formula for the probability of choosing a link is given below:

$$P(i) = \frac{1 - \left(1 - \left(\frac{SP(v,D)}{SP(o,w) + \ cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}{\sum_{i \in M} 1 - \left(1 - \left(\frac{SP(v,D)}{SP(o,w) + cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}$$

[2.1]

Where:

P(i) is the probability of choosing link i out of possible outgoing links (M) with source node v and sink node w.
SP(v,D) is the shortest path/least cost path from source node v to destination D.
Cost(i) is the cost of link i.
α and β are parameters that make the probability more sensitive to increase in cost.

This method generates the routes link by link, assessing the probability of outgoing links connected to each node. Due to the randomness of the route generation, rules should be implemented in the route generation to avoid routes that loop or U-turn and to ensure that routes are a reasonable length.

Borst et al. (2008) used an ad hoc iterative fitting method with stochastic link friction factors to match link counts. The model regressed link-level built environment variables on the best fitting friction factors.

Broach and Dill (2015) use an algorithm based on the biased random walk algorithm. The cost for each link only considers the sum of the cost of the link and the shortest path from the sink node of the destination to the destination. The formula for the probability of choosing a link out of outgoing links is given below:

$$P(i) = \frac{1 - \left(1 - \left(\frac{SP(v,D)}{cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}{\sum_{i \in M} 1 - \left(1 - \left(\frac{SP(v,D)}{cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}$$

[2.2]

Where all terms are the same as in equation [2.1].

Through trial and error, the α and β values were determined to be 5 and 1 respectively.

Broach and Dill (2015) implemented rules for the route generation such that:

- No node is traversed twice
- No U-turns needed
- The route does not exceed three times the shortest network path
- The route does not pass by the destination link

Both deterministic and stochastic approaches use a fixed number of generated routes as the choice set and assume that the generated choice set reflects the individual’s actual choice set. If the generated choice set includes five routes, it assumes that the individual only considered five routes. However, empirical studies have shown that the observed route is often not generated, which suggests that this assumption is not true (Frejinger, 2007; Prato & Bekhor, 2006; Ramming, 2002).

Fosgerau, Frejinger, and Karlstrom (2013) developed a Recursive Logit (RL) model which is a link based network route choice model with an unrestricted choice set. The route choice problem is formulated as a dynamic discrete choice model where the utility maximization problem is a dynamic programming problem. Similar to the biased random walk algorithm, the route choice problem in the recursive logit formulates a route as a sequence of links. The recursive logit allows route probabilities to be estimated without specifying a limited/bounded choice set while accounting for correlation between routes and characteristics.

2.3.2. Route Choice Models

<u>Multinomial Logit</u>

The Multinomial Logit (MNL) model calculates the probability of choosing a route based on the utility of each alternative (Train, 2009). The model assumes that error terms are identically and independently distributed (i.i.d). However, this assumption is not always true for route choice where alternative routes can have correlation due to overlapping paths. That is, to the extent that shared portions of two routes will have common unobserved attributes (e.g., the common portion may be idiosyncratically slower or faster than expected), then their utilities will be correlated.

One approach to accounting for this correlation is to use a model that permits correlation among alternatives, such as the multinomial probit model (Bouthelier & Daganzo, 1979). Probit models, however, are computationally challenging for even moderate-sized networks. In order to retain the computational efficiency of the multinomial logit model, various "correction factors" have been suggested to approximately account for correlation between alternatives. These correction factors are added to the route utility equations, thereby permitting retention of the overall logit choice model. The various correction factors suggested in the literature are briefly discussed below.

C-Logit

The C-Logit was developed by Cascetta et al. (1996). It introduces a "Commonality Factor" to correct for the overlap. Each alternative has its own Commonality Factor which is proportional to the overlap with other routes in the choice set. Cascetta et al. (1996) provide three formulations for calculating the Commonality Factor; however, there is a lack of any guidance about which formulation should be used (Frejinger & Bierlaire, 2009).

Path Size Logit

Similar to the C-logit model, to account for the correlation from overlapping alternative routes, a Path Size (PS) correction term was introduced by Ben-Akiva and Bierlaire (1999) into the multinomial logit formulation to form the Path Size Logit (PSL). The utility equation for path i for an individual n is $U_{in} = V_{in} + \beta_{PS}\, lnPS_{in} + \epsilon_{in}$ and the probability of choosing a route is given by:

$$P(i|C_n) = \frac{e^{\mu(V_{in}+\ln(PS_{in}))}}{\sum_{j\in C_n} e^{\mu(V_{jn}+\ln(PS_{jn}))}}$$

[2.3]

Where:

Cn is the choice set for user n (includes chosen route)

μ is the logit scale term

V_{in} is systematic utility for alternative i for user n

PS_{in} is the path size factor for alternative i for user n

$$PS_{in} = \sum_{a\in\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \delta_{aj}}$$

[2.4]

Γi is the set of links in path i

L_a is the length of link a

L_i is the length of path i

δ_{aj} equals 1 if link a is on path j and 0 otherwise

$\Sigma_{j\in Cn}\delta_{aj}$ is the number of paths in choice set Cn sharing link a

The correction for partially overlapping paths is weighting the links by the percent contribution of the link to total path length. L_a/L_i represents the percent contribution of the link to the total route length. The remaining terms represent the number of times the link appears in the choice set. If the links are only used once, then the route is referred to as "unique" and has a correction term equal to zero. If a path has overlapping links with other alternatives, the PS factor reduces the utility for the alternative. An issue with this model formulation is that arbitrarily long paths, which are unlikely to be chosen, will overlap with other paths, thus reducing the utility for even likely paths.

Ben-Akiva and Bierlaire (1999) present another formulation of the PS factor which considers the ratio of the shortest path to the length of the alternative routes which contain the overlapping links:

$$PS_{in} = \sum_{a\epsilon\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \frac{L^*_{C_n}}{L_j} \delta_{aj}}$$

[2.5]

L^*_{Cn} is the length of the shortest path in the choice set

Ramming (2002) introduces a 'Generalized PS' formation which decreases the impact of unrealistically long paths in the choice set:

$$PS_{in} = \sum_{a\epsilon\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \left(\frac{L_i}{L_j}\right)^{\phi} \delta_{aj}}$$

[2.6]

φ is a parameter that controls the impact of route length to the correction factor. When φ=0 the equation is insensitive to length and is equal the original formulation presented by Ben-Akiva and Bierlaire (1999).

Hoogendoorn-Lanser et al. (2005) conducted a study on how to define overlap in multi-modal networks. Through their testing of the previous equation, it was found that φ=14 showed best results and that the best model fit was also found when the overlap was expressed in terms of number of legs compared to time or distance.

Bovy et al. (2008) proposed a Path Size Correction (PSC) factor which is applied to the utility equation as follows: $U_{in} = V_{in} + \beta_{PSC}\ PSC_{in} + \epsilon_{in}$

$$PSC_{in} = \sum_{a \epsilon \Gamma_i} \frac{L_a}{L_i} \ln \left(\frac{1}{\sum_{j \epsilon C_n} \delta_{aj}} \right)$$

[2.7]

The main difference in the formulation is the placement of the logarithm. Bovy et al. (2008) suggest that this formulation is more appropriate from a theoretical perspective than weighting the L_a/L_i term within the logarithm.

According to Hoogendoorn-Lanser et al. (2005), the reason for the estimation of the β_{PS} parameter is that the PS attribute can have a behavioral attribute. β_{PS} is a strictly positive parameter in order to be consistent with the theory $\beta_{PS} = 1/\mu$.

<u>Expanded Path Size Logit</u>

Frejinger, Bierlaire, and Ben-Akiva (2009) developed an Expanded Path Size (EPS) correction factor which is used to correct for correlation from overlap for choice sets generated from a stochastic choice set generation method. The EPS logit also includes the stochastic sampling correction factor from Frejinger (2007). The formula for EPS Logit is given by the follow equations:

$$P(i|C_n) = \frac{e^{\mu(V_{in} + \ln(EPS_{in})) + \ln\left(\frac{k_{in}}{q(i)}\right)}}{\sum_{j \in C_n} e^{\mu(V_{jn} + \ln(EPS_{jn})) + \ln\left(\frac{k_{jn}}{q(j)}\right)}}$$

[2.8]

Where:

C_n is the choice set for user n (includes chosen route)

μ is the logit scale term

V_{in} is systematic utility for alternative i for user n

EPS_{in} is the expanded path size factor given by equation

k_{in} is the number of times alternative i is randomly drawn. If the chosen route, k_{in}+1

q(i) is the probability of choosing a route containing the street segments. It is calculated as the product of each link choice probability

$$EPS_{in} = \sum_{a \in \Gamma i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj} \, \Phi_{jn}}$$

[2.9]

and

$$\Phi_{jn} \begin{cases} 1 & if \ \delta_{jc} = 1 \ or \ q(j)R_n \geq 1 \\ \dfrac{1}{q(j)R_n} & otherwise \end{cases}$$

[2.10]

Where:

δ_{aj} equals 1 if link a is on path j and 0 otherwise

δ_{jc}=1 for the chosen alternative

R_n is the number of draws

q(j) is the probability of generating alternative j

Φ_{jn} is an expansion factor that corrects for the stochastic sampling method. Φ_{jn} has a value of 1 if the alternative is the chosen route or is expected to be drawn at least once.

Random Sampling Correction Factor

The biased random walk algorithm creates a choice set that may have overlapping routes and may also have repeated routes. To correct for reoccurring routes a sample correction is introduced. Frejinger (2007) introduces a sampling correction factor which is calculated as the natural logarithm of the number of times the alternative is randomly drawn divided by the probability of that route being generated. The formula for the path size logit with stochastic choice set mode is defined in equation [2.8]. The probability of a route being generated is calculated as the product of the probabilities of all the links in the route.

The sampling correction factor corrects for two things: the number of times an alternative is drawn and the probability of the route being generated. Figure 2.1 illustrates how the value of the sampling correction factor changes with a number of draw times and the probability of generating route. If an alternative is drawn multiple times, it is only included in the choice set once. The correction factor accounts for an alternative having multiple draws by adding utility to that alternative. The more times an alternative is drawn, the more utility the correction factor adds. However, as the probability of a route being generated increases, the utility added by the correction factor decreases. With the stochastic route generator, there are some routes that are more likely to be generated. The correction factor will add more utility to routes that have a lower probability of being generated so that the choice set is not overpowered by routes with high probabilities of being generated.

Figure 2.1a shows how the correction factor increases as the number of times the alternative is drawn increases (with q(i) held constant). Figure 2.1b illustrates the relationship between the correction factor and the probability of route generation (with k(i) held constant).

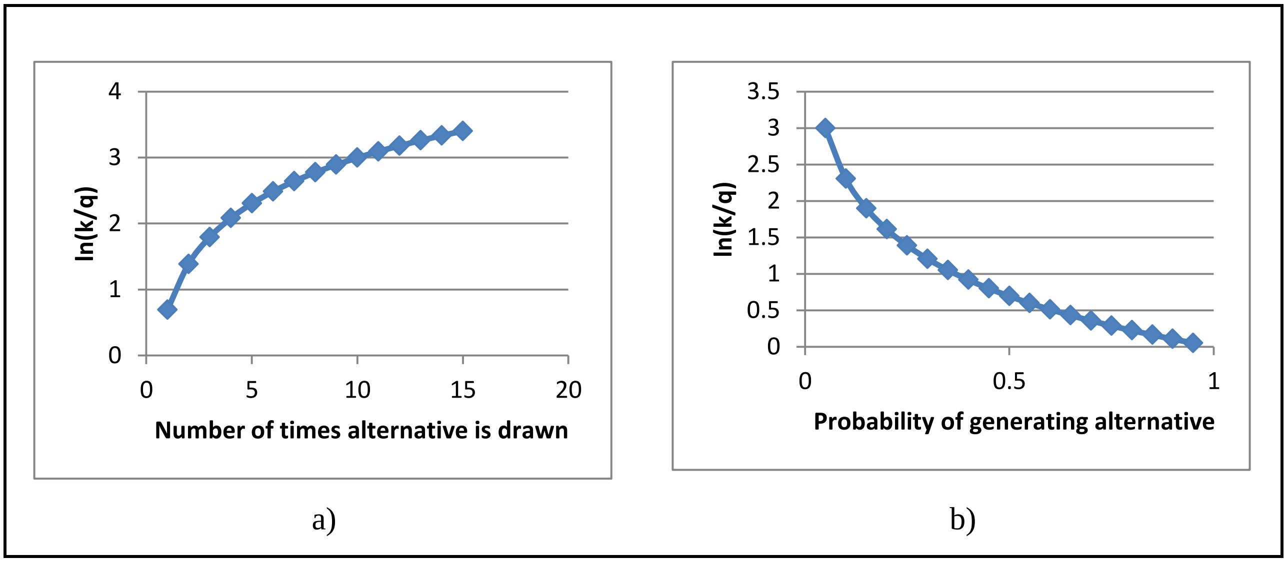


Figure 2.1 - Sampling correction term sensitivity analysis

Recursive Logit

The Recursive Logit (RL) model formulation introduced by Fosgerau et al. (2013) is a dynamic discrete choice model where the path choice is treated as a sequence of link choices. At each node in the network, link choice probabilities are determined using a multinomial logit model and expected downstream utilities (Fosgerau et al., 2013). Utilities are given by the instantaneous cost, a maximum utility to the destination, and i.i.d. extreme value type I error terms. The decision maker chooses the utility-maximizing outgoing link. Similar to the path size logit, a correction factor called the ‘link size’ is introduced to correct the utility from overlapping paths.

In the recursive logit, a path is a sequence of links ($k_0,\ldots k_1$). A deterministic utility $v_n(a|k)$ component is a function of a vector of observed characteristics v ($x_{n,a|k};\beta$) of the link pair (k,a). Path attributes are assumed to be link additive. The probability of choosing link a given state k is given by the multinomial logit model:

$$P_n^d(a|k) = \frac{e^{\frac{1}{\mu}\left(v_n(a|k)+V_n^d(a)\right)}}{\sum_{a'\in A(k)} e^{\frac{1}{\mu}\left(v_n(a\prime|k)+V_n^d(a\prime)\right)}}$$

[2.12]

The value of the logsum is given by:

$$V_n^d(k) = \begin{cases} \mu \ln\left(\sum_{a\in A} \delta(a|k)\, e^{\frac{1}{\mu}\left(v_n(a|k)+V_n^d(a)\right)}\right) & \forall k \in A \\ 0 & k = d \end{cases}$$

[2.13]

Where $\delta(a|k)$ equals one if $a \in A(k)$ and zero otherwise. (Fosgerau et al., 2013)

The RL method has the advantage of considering a potentially infinite number of paths between the origin and destination, which is similar to a multinomial logit model with an infinite number of alternatives. Being able to consider infinite paths avoids the restriction on choice sets seen in other models. The approach of using a sequential link choice model is found to be computationally efficient and permits consistent estimation of parameters.

2.4. GPS Point Map-Matching

There are very few standards in the academic field for processing smartphone GPS data. Studies that process smartphone GPS data tend to design their own procedures tailored to the characteristics of the data collected.

Dalumpines and Scott (2011) developed a set of tools in ArcGIS 10.1 for map-matching and trace processing. Their approach for map-matching consists of filling any gaps within the GPS traces, creating a buffer around the GPS points, and solving for the shortest route between the origin and destination within the buffer area using ArcGIS. If the route solving fails, the buffer area will be increased and the route solver will be attempted again.

Whereas the ArcGIS toolset developed by Dalumpines and Scott (2011) only solves the route for a specified mode, the GPS matching system developed by Sheung and Shalaby (2006) attempts to do much more. Sheung and Shalaby (2006) developed two systems for the link and mode identification using GPS data. The first system is a GPS-alone system and the second is a GPS-GIS integrated system. The GPS-alone process cleans poor quality GPS points from the set and then proceeds to break down a 1-day continuous GPS record into individual activities and trips. Trip ends are identified using a 120 second dwell time and trip mode is identified by a fuzzy logic-based algorithm looking at the change in speed (Sheung & Shalaby, 2006). The GPS-GIS system uses similar logic to the first system but adds additional link matching algorithms from GIS maps.

3. Network Data, Street Attributes, and Smartphone-Based Travel Survey

3.1. Data Sources

The City of Toronto has a wide array of data available for free in the City of Toronto Open Data catalogue. The Open data catalogue was one of the main sources of topological data in this project.

The Toronto Street GIS file (Centreline File) from the Open Data catalogue was used as the base network for this analysis. The centreline file contains polylines of the streets, walkways, rivers, railways, highways, and administrative boundaries in the City of Toronto. Each line also includes attributes such as street name, address ranges, and length of the segment. The centreline file was last updated in March 2016 and is updated on a semi-annual basis.

A separate GIS file for sidewalk inventory contains information on the location/condition of sidewalks along transportation corridors. This sidewalk shapefile aligns with the Street Network Centreline file. The inventory of sidewalk conditions was completed using aerial photography from 2011. Conditions of the sidewalk include no sidewalk, pending, the sidewalk on both sides, the sidewalk on one side only, walkway (confirmed), or walkway (unconfirmed).

The Intersection Inventory is a GIS point file with all the intersections/junctions of the polylines in the Centreline file. The intersection shapefile was last updated on September 25, 2015, and is updated semi-annually.

The Signalized Intersection Inventory is a csv file that contains the coordinates of the traffic signal as well as information on the main street, side street, and signal activation date. The signalized intersection data was last updated in 2016 is updated annually.

The Pedestrian Crossover Inventory is a csv file that contains the coordinates of the signal as well as main street, side street, and midblock distance. Signalized intersection data are updated annually.

Elevation data was retrieved from the Natural Resources Canada’s website. The elevation file was published on March 1, 1999.

Walk Score data was retrieved from the ArcGIS online database. The file was created on May 5, 2015, and contains point data located on neighborhood centroids. The Walk Score for each neighborhood was calculated based on distance/time to nearby amenities, population density, block length, and intersection density (Walk Score, 2016). Walk Scores range from 0 to 100. Table 3.1 provides descriptions of the Walk Score. Each point contains information about the area name, Walk Score, and Walk Score description.

Table 3.1 - Walk Score Descriptions

Walk Score	Description
90-100	Walker's Paradise - Daily errands do not require a car
70-89	Very Walkable - Most errands can be accomplished on foot
50-69	Somewhat Walkable - Some errands can be accomplished on foot
25-49	Car-Dependent - Most errands require a car
0-24	Car-Dependent - Almost all errands require a car

(Walk Score, 2016)

Land use information was collected through the combination of a property boundary shapefile and an address point shapefile; both from the Open Data catalogue. The property boundary information is in the form of GIS parcel shapefiles. The shapefile outlines the geographical area of all parcels in the City of Toronto. The property boundary data was last updated on March 17, 2016, and is updated semi-annually. The address point shapefile contains points representing over 500,000 addresses within the City of Toronto. Each address point includes information such as street name, street number, address type, and latitude and longitude coordinates. The address file was last updated in March 2016 and is updated on a semi-annual basis.

3.2. Smartphone Travel Survey

The smartphone travel survey was conducted as part of the Waterfront Toronto Smartphone Data Collection Project in 2014. The full detail about the development of Smartphone Travel Survey is outside the scope of this project. More information can be found in the paper by Harding, Zhang & Miller (2015).

Recruitment

Survey respondents were recruited through social media, print media, and emailing individuals working in the Waterfront area. Recruitment was also conducted in-person among customers of coffee shops in the Greater Toronto Area (GTA). For the in-person recruitment, trained interviewers were utilized to help people who eligible for recruitment install the study app. As an incentive for downloading the app and answering a demographic survey, individuals were provided a $5 gift card. In addition, all participants who installed the app and resided within the GTA were entered into weekly prize draws.

Smartphone recorded data

The survey app utilized the phone's accelerometer, gravity sensor, Bluetooth, gyroscope, and GPS. The data was collected on a regular basis through wireless networks. Participants were also prompted with an optional web validation survey to collect information on trips made, origin, destination, time, mode, and purpose. The app was designed to be a passive data collection tool which did not require the participant to start or stop the app when a trip was being carried out. Participants would simply install the app and leave it running in the background as they carried out their normal activities. The app would always be on and the GPS location was recorded only if the individual travelled more than 50m. The app that was used for this study had to balance accuracy and battery usage. Participants would be less likely to keep the app if there was significant battery usage by the app. If the GPS location was transmitted at more frequent travel intervals, there would be the better accuracy of route traces but a decrease in phone operation time.

Collection period

The full survey was launched in November 2014 with the goal of collecting data for four weeks. After the four week period was concluded, the servers remained active to allow respondents to continue to contribute data. The length of participation varies per individual. Some individuals kept the app installed for the four-week period, some uninstalled the app after a day, and some participants kept recording trips for over a month. The initial intention was to release the apps in the early fall but delays in development postponed the launch.

Post Survey Data Processing

A trip end was determined based on the dwell time at the location. After an individual stayed at the same location for three minutes, it was determined to be a trip end. Since points are only recorded if the individual travels 50m from the last point, the destination point may not be exactly where the individual ended their trip. In addition, GPS triangulation may take a few moments to communicate with satellites to acquire the location of the individual. In that time the individual may be moving and once the GPS acquires the location, it may not be situated at the exact location of the trip origin. However, for this study, it is assumed that the origin and destination reported in the dataset is sufficiently close to the real origin and destination.

Travel modes were inferred based on speed profiles and the surrounding network. For example presence of above-ground tracks or railway corridors suggest subway or trains are being used. Out of the inferred modes, walking had the most accurate results when compared with the reported mode. Comparing the inferred mode and the reported mode of travel, the overall mode detection had a success rate of 87%, while the walking mode specifically had a success rate of 100% (Harding, Zhang, & Miller, 2015). The inferred/reported mode success rate is presented in Table 3.2.

Table 3.2 - Mode detection success rate (Harding et al., 2015)

Inferred mode	Reported Mode				Overall success rate
	Walking	Bicycle	Car/passenger car	Transit	
Walking	**100.0%**	11.1%	5.4%	3.1%	100.0%
Bicycle	0.0%	**81.5%**	2.7%	2.0%	81.5%
Car/passenger car	0.0%	0.0%	**83.8%**	14.3%	83.8%
Transit	0.0%	7.4%	8.1%	**80.6%**	80.6%
					87%

There were some issues when identifying mode due to the lack of detailed accelerometer data, limited location logging, and complexity of travel in the downtown core. While the inferred/reported success rate for walking was 100%, there is a small percentage of bicycles, cars, and transit that may appear in the walking data.

The trip purpose was not collected during the survey and therefore could not be included in the analysis.

3.3. Cleaning Data

3.3.1. Network Cleaning and Merging Data

The street network shapefile included streets and walkways but also included rivers, railways, and administrative boundaries in the City of Toronto. The rivers, railways, shorelines and administrative boundaries were removed from the shapefile so that alternative walking routes would not traverse these lines. Highways and highway ramps were also removed from the network to prevent observed trips or alternative routes being incorrectly matched to them.

While the street network shapefile does include some walkways there are some paths that were not present. The missing walkways included footpaths through parks or open spaces. These pedestrian paths were manually added by comparing the network shapefile to Google Maps.

An overall dataset needed to be built which consolidated the information available from the various data files described above into the street network file. First, the sidewalk description was matched to the corresponding street link. Next, signalized intersection and pedestrian crossover information was transferred into the street network junction shapefile. New junction points were added for locations where signalized intersections or pedestrian crossovers could not be matched with a junction point. Elevation data was also transferred to the street junction file. Street junctions were then matched with the 'To' and 'From' end points of each street segment. Elevation information was transferred from the junctions to the Street network for 'To'/'From' node elevation attributes. 'To'/'From' ends of the links also had an attribute labelling if the junction was a signalized intersection or pedestrian crossover.

Walk Score points contain Walk Scores for various neighborhoods throughout Toronto. The street segments within the neighborhood were assigned the corresponding Walk Scores for that neighborhood. The link Walk Score is stable for segments near the centre of the neighborhood but can lead to large shifts in Walk Score at the boundaries of the neighborhoods as seen in Figure 3.1.

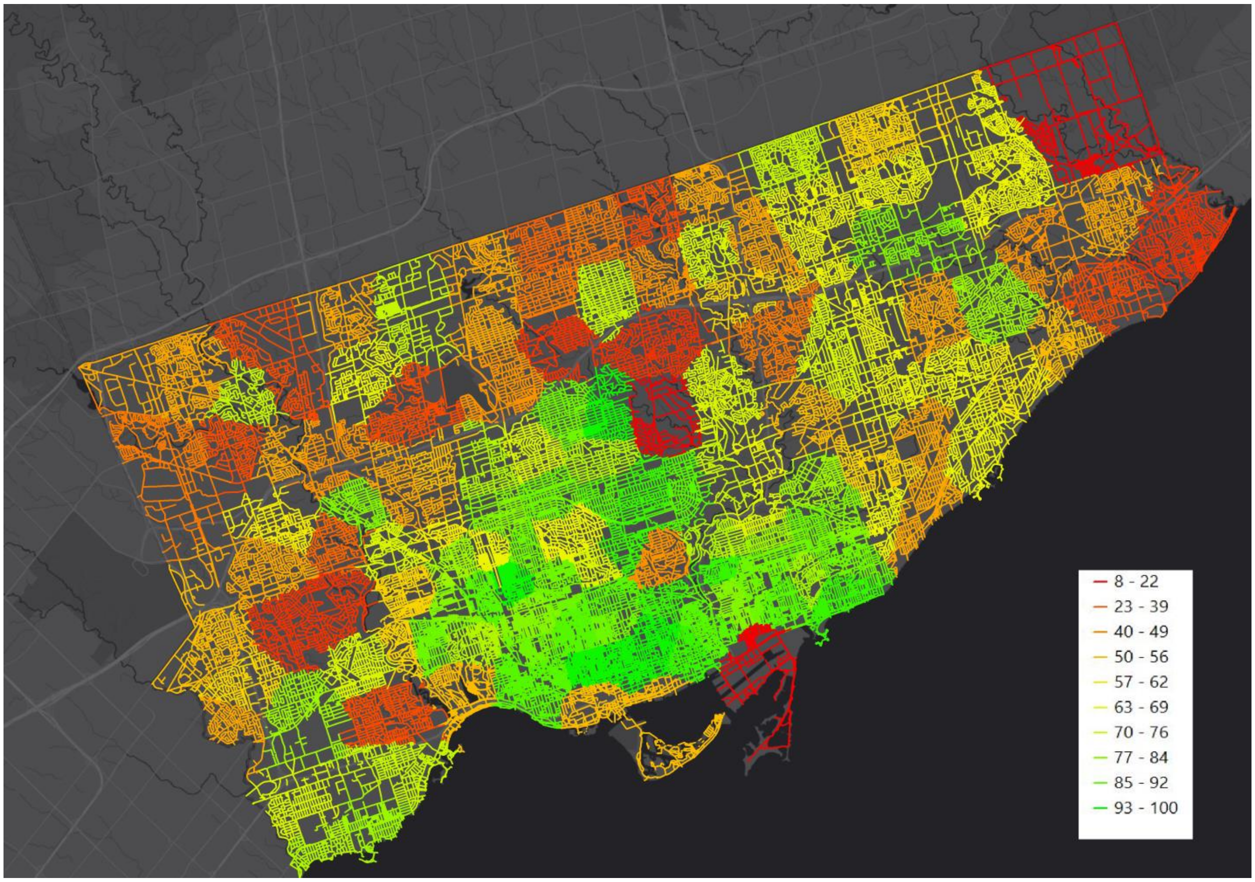


Figure 3.1 - Walk Score map

Land use data had to be converted into a frontage measurement before it could be added to the street segments.

3.3.2. Land Use data

Land use information comes in the form of address points with land use descriptions and property boundary shapefiles. The format for land use in this analysis requires the "land use frontage" along the streets to be calculated.

First, the land use information from the address points needed to be attached to the land parcels. Land parcels that were primarily occupied by roads or highways were lablled as 'road' land use and filtered out. For the remaining land use, ArcGIS was used to transfer the address land use to the parcel. If there were multiple land uses located in the same parcel, the parcel was labelled as mixed use.

Some cases of mixed-use had to be corrected such as buildings/shops within parks caused parks to be labelled as mixed use. In some cases, schools would be labelled as mixed use because there would be multiple buildings on the property.

The address data set did not include all addresses, thus, some land parcels were left unlabelled. Addresses were manually coded through inspection of the land use from Google maps and the remaining unknown land use was inferred based on the land use surrounding it. Figure 3.2 compares the City of Toronto land use map to the land use map generated through the open source data.

The generated land use map fits the overall land use location and proportions found in the City of Toronto land use map. The generated land use map has a "peppered" look because the map shows land uses of individual land parcels rather than the aggregated land use by-law zones found in the City of Toronto land use map. The generated land use map contains more mixed-use land than the zoning by-law map. This result may also be due to the scale of the generated map. A parcel of land may contain multiple addresses, such as a small shopping plaza, which would be labelled as mixed use by the map generation code. Also, the land use matching algorithm may be sensitive to land uses within a close proximity.

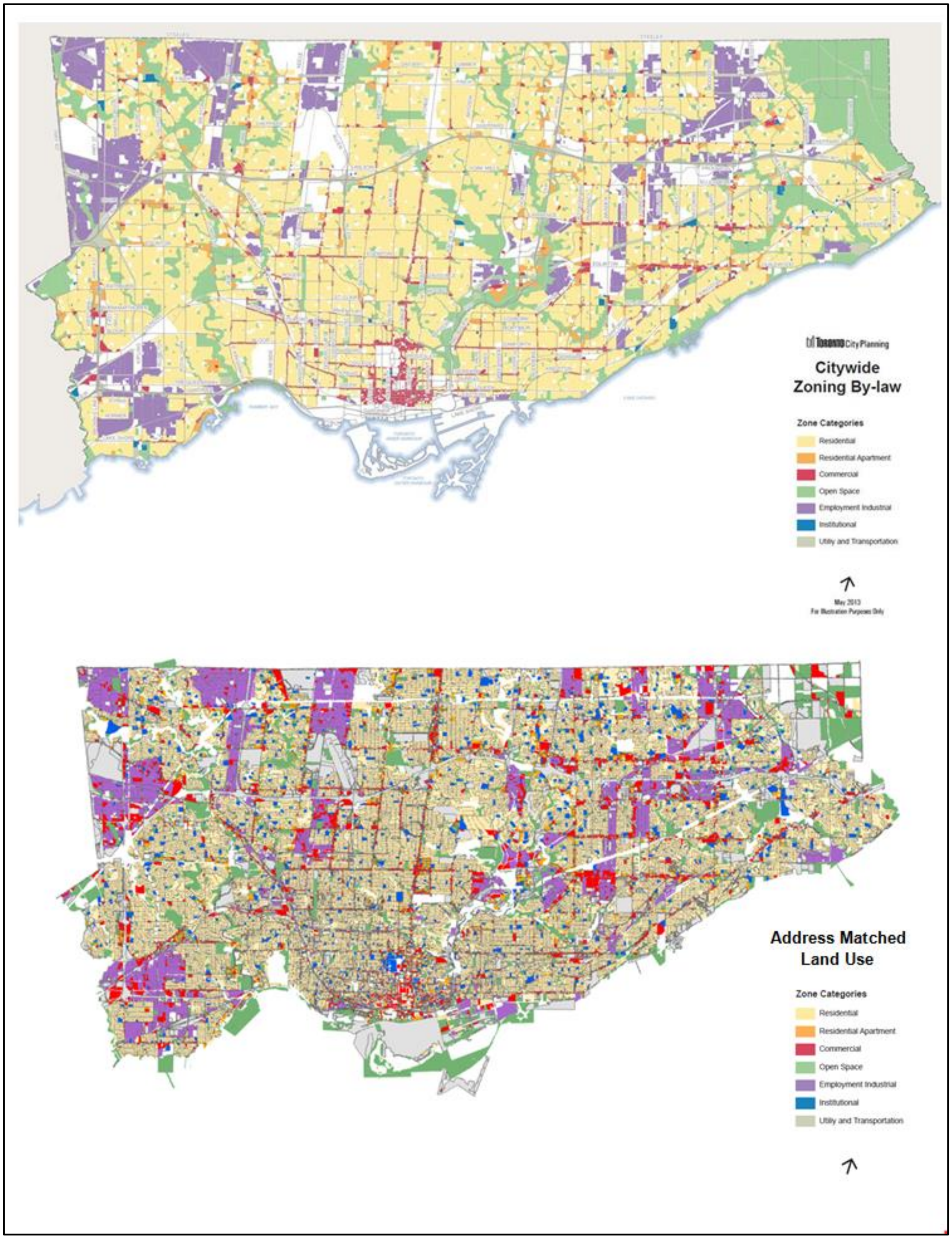


Figure 3.2 - Land use map comparison

Once the land use parcels were labelled, the composition of land use frontage needed to be calculated. A Python script was developed that utilizes ArcGIS to determine the land use frontage. Land parcels were deconstructed into only the boundaries of the parcel then the boundaries were matched to the closest street link seen in Figure 3.3a. In some cases, the closest street segment to the land use boundary is the cross street as seen in Figure 3.3b. For this reason, angles of the street and land use boundary were compared and land use boundaries were matched to the nearest parallel street segment. For each street segment, the length of each land use was summed. The land uses were then ranked based on the amount of frontage on the street segment. The top three land use codes and lengths are added as a street segment attribute as illustrated in Figure 3.4.

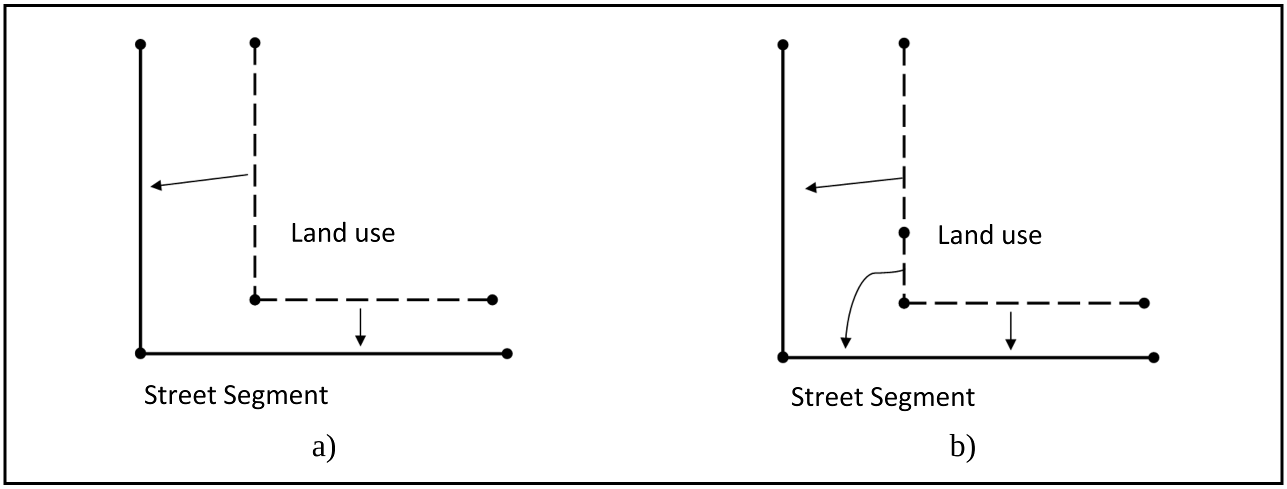


Figure 3.3 - Land use segment matching

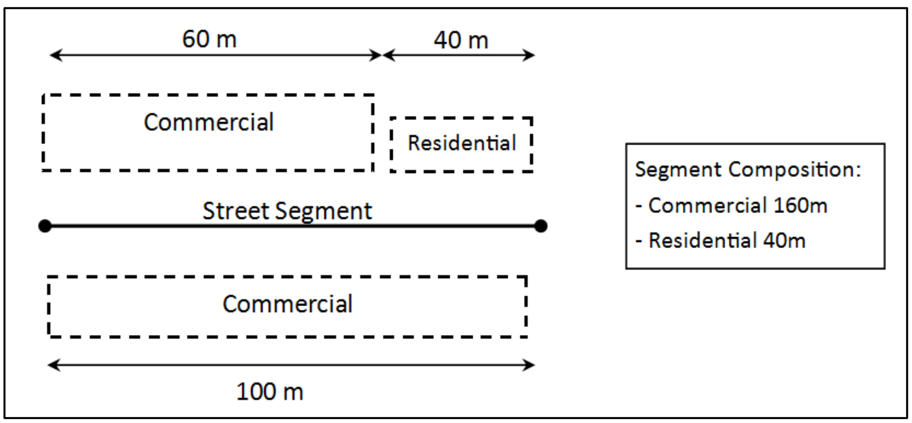


Figure 3.4 - Street segment composition

3.3.3. Walk Trip Data

The Smartphone Travel Survey collected over 5800 trips across 156 users in Toronto. Trips with travel mode other than walking were filtered out of the data set to create walk trip data set used in this study. The number of trips remaining was 3193 walking trips across 103 users.

Further cleaning of the data set was required. GPS connection combined with the minimum 50m travel distance and GPS noise resulted in some trips containing large gaps between points in the trip. Trips with gaps over 200 meters were excluded. Gaps could be filled by finding the shortest path between the points on either side of the gap but this would be inferring the observed path which may have influenced the end model. There were 195 trips removed for large gaps.

A percentage of the trips with large gaps had gaps due to taking the subway or using Toronto's underground walkway called the "PATH". These trips showed individuals walking to a subway station or PATH entrance and emerge from another subway station or PATH entrance a distance away.

Subway stations were mapped in ArcGIS with a 100m buffer area generated around the stations. Trips with large gaps were superimposed onto the map. If gaps started within the buffer area of a subway station and ended at another station, the trip would be labelled as a walk trip to transit. Figure 3.5 shows a walk trip using the subway. The walk trips to and from subway stations were applicable to be included in the data set. The trip to the subway and the trip from the subway were treated as two walking trips. However, these trips were often very short and typically would later be excluded due to lack of variation of alternative routes.

Similarly, trips were matched with PATH entrances; however, these trips sometimes did not reappear. It is assumed that from the PATH, the individual stayed within a building or the trip was determined to have ended. Even for the trips that were labelled as PATH trips, the route attributes within the PATH network differ greatly from the street characteristics of interest. In addition, the portion of the trip that was recorded was often too short for a variety of alternative routes to be generated. Thus, walk trips to/through the PATH network were excluded from the dataset.

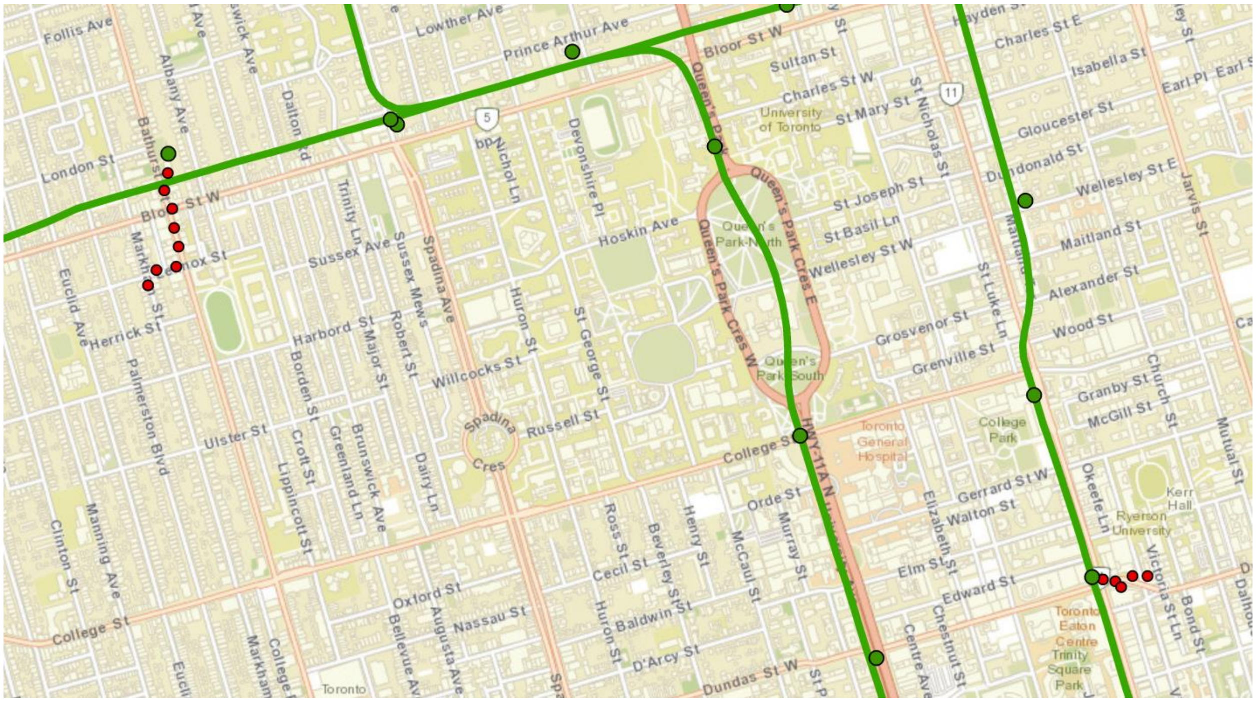


Figure 3.5 - Large gap subway trip

Since the mode of travel was determined by speed, there are some trips that were mislabelled. For instance, there were a number of ferry trips between Downtown Toronto and Toronto Island that were mislabelled as walking trips and were removed from the dataset.

There were some trips that only consisted of one or two GPS points. This may be due to the GPS connectivity, the minimum 50m travel distance, or the three-minute trip end criteria. These trips do not provide sufficient information or route variability for the model and therefore, were removed.

<u>Map-Matching GPS points</u>

A Python 2.7 script using ArcGIS 10.2 was developed to match the GPS points to the street network. While map matching algorithms already exist, such as the one developed by Dalumpines and Scott (2011) or the GPS-GIS matching system by Sheung and Shalaby (2006), it was decided to develop a map-matching process tailored to the nature of the GPS data in this study. The map matching approach used in this study is based on the tool developed by Dalumpines and Scott (2011). The map-matching approach used in this study followed the following procedure:

1. Import the trip GPS points into ArcGIS
2. If any gaps, fill in with points (points added in straight line between gap ends)
3. Create a 40 m radius buffer area around the GPS points
4. Import buffer area into ArcGIS route solver as a barrier. This restriction means that ArcGIS will solve for a route within the buffer area.
5. Import the origin and destination points and solve for the shortest path within the buffer area

The map matching process is illustrated in Figure 3.6. First, the distance between the GPS points in the trip is checked. There may be gaps of less than 200m between GPS points. This approach requires the distance between points to be 50 meters or less. If there are gaps greater than 50m the script would fill in the gap between the two points. The gap would be filled with GPS points laid in a straight line between the two points. The filler points were evenly placed to fill the gap.

Once the gaps are filled, a 40m buffer is created around the GPS points. The origin and destination GPS points are imported into the ArcGIS 10.2 route solver tool and the perimeter/outline of the buffer area is used as a restrictive barrier for route solving. The shortest route between the origin and destination points is then solved with the condition that the route cannot cross the restrictive barrier.

Figure 3.6 - Map-matching process

There are a few cases where this method fails:

- Pedestrian trips do not always occur on streets but sometimes go through buildings or across open spaces which there are no links to assign the GPS points. Some paths available to pedestrians are not necessarily included in the City of Toronto Centreline file. A map which is able to catch these trips may need to allow for travel off of the normal paths.
- With dense street networks, there is a chance that alternative routes may exist within the buffer area.
- If there are large gaps between the GPS points it is possible that the buffer area is not continuous which would cause the route solver to fail.

- Filling GPS points in a straight line may result in routes that cut through buildings or do not align properly to the street network. When the buffer is generated around the points, there may not be a solvable route between the origin and destination

a) b)

Figure 3.7 - Map-matching issues

Figure 3.7a illustrates an observed route that cuts through a building or travels along an alleyway not included in the street network. Figure 3.7b shows a case where there is a discontinuity in the barrier restriction. Attempting to solve the route within the boundary will fail. When the gap is filled using the straight line method the route solver still fails because the barrier will not include a connecting intersection that is necessary for solving the route.

Initially, instead of creating a route solving buffer area, the GPS points were treated as ‘stops’ within ArcGIS route solver. However, because of the GPS noise, the points would occasionally appear on cross streets which caused the route solver to detour or make U-turns to reach the points. The buffer approach found in Dalumpines and Scott (2011) was found to have a much more accurate route identification ability.

4. Methodology

4.1. Alternative Route Generation Algorithm

One of the most difficult parts of modelling path choice is the generation of alternative routes. Rule-based deterministic route generation approaches are difficult to validate but stochastic route generation methods may have too much of a random component within the generation (Hoogendoorn-Lanser, 2005). Another issue with alternative route choice sets is the possibility of overlap across alternatives which violates the i.i.d assumption, although there have been path size correction factors developed to correct for correlation.

A stochastic route generation approach was used for the generation of alternative routes for this model. It is hypothesized that pedestrians choose their route link by link, assessing the current situation and expected conditions ahead such as lighting, signalized crossing, or future turns. While the individual may have a general idea of the route when they start, at each junction point there is assessment and possibility for deviation. This type of route building behaviour is best represented by the biased random walk algorithm developed by Frejinger (2007) and modified by Broach and Dill (2015). Alternative routes are generated link by link from the origin to the destination. The probability of choosing a link is based on the sum of the cost of the link and the cost of the shortest path from the sink node of that link to the destination. This “link cost” is compared to the shortest path cost from the source of the link to the destination. The costs for all connected links at a node are compared to determine the probability of choosing each link. If the link is on the path of least cost to the destination, it is more likely to be chosen. As discussed in Chapter 2, the formula for the probability of choosing a link is given below:

$$P(i) = \frac{1 - \left(1 - \left(\frac{SP(v,D)}{cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}{\sum_{i \in M} 1 - \left(1 - \left(\frac{SP(v,D)}{cost(i) + SP(w,D)}\right)^{\alpha}\right)^{\beta}}$$

[4.1]

Where:

P(i) is the probability of choosing link i out of possible outgoing links (M), with source node v and sink node w.

SP(v,D) is the shortest path/least cost path from source node v to destination D.

Cost(i) is the cost of link i.

α and β are parameters that make the probability more sensitive to increase in cost.

Values of α and β parameters were tested on a sample of walking trips. As the α and β values increased, the number of unique alternatives decreased as the route generation process converged on routes that minimized cost. Conversely, as α and β were decreased, the generated routes became increasingly random, making longer and more unlikely routes. Values of 5 and 1 for α and β respectively were determined to produce acceptable route generation results. These values are also supported by other pedestrian route choice research (Broach & Dill, 2015; Frejinger, 2007).

The biased walk algorithm was applied as follows:

- Origin (O) and destination (D) were defined as the nearest points on the network links
- Paths were generated such that
    - No node is traversed twice. If a loop is detected, the route generation attempt fails.
    - U-turns are not allowed.
    - The generated path does not exceed two times the shortest path between O and D. If the generated route exceeds this length, the attempt fails.
    - The path does not pass the destination link
    - Travel on street segments that go in a direction away from the destination is heavily penalized (cost=9999m) unless they are on the shortest path from the source to the destination.
    - If a dead end is reached, the route generation attempt fails and the dead end segment is recorded so it is not considered again. After 10 attempts, the iteration is abandoned.

Since the street network in the study area is a dense grid, when the biased walk around the shortest path method was used, the generated routes would often contain numerous turns and unlikely routes. By analyzing the chosen routes, it was observed that most routes continue in straight along roads instead of weaving through the network making many turns.

For this reason, instead of biasing the generation towards the shortest path, the algorithm was biased towards "least cost". To generate routes that are plausible alternatives, a cost of 50m was added for every turn in the route. In addition, travel along streets with sidewalks on both sides was set to be 10% shorter than the actual distance. These costs biased the generated routes towards more straight routes and routes that travel along more pedestrian friendly streets. The values for turns was based on the pedestrian route choice research done by Broach and Dill (2015). The value for complete sidewalks was used to reflect the perception of attractiveness for sidewalk conditions (Ferrer et al., 2015; Rodríguez et al., 2015).

Street segments in the wrong direction are heavily penalized but are still considered because there are some cases where it is necessary to take a street segment in the wrong direction to get to the destination. An example of the route generation process can be found in Appendix A.

The alternative route generation process was implemented in a Python code which utilized route solving functions in ArcGIS 10.2. Origin and destination points were imported into ArcGIS and matched to the closest links. The origin link is used as the starting point for the algorithm while the destination link is used as the stopping criteria. The straight-line distance between the origin and destination is calculated and used as the radius of a buffer area created around both the origin and destination. Within the buffer area, the "cost" of travelling from each street junction to the destination is calculated.

When generating a route, the cost for the next possible link needs to be calculated to determine the probability of choosing each link. However, it was determined to be more computationally efficient to run the ArcGIS Origin-Destination (OD) cost matrix for the street junctions within an area compared to the route solver from the end of the each of the considered.

At each junction, the connected links are considered to be added to the route. The path of least cost from the junction is compared to the cost of the considered link plus the cost to the destination from the end of the considered link. This ratio is then used in a logit model to determine the probability of choosing each of the considered links. If a link is on the path of least cost, then it is more likely to be chosen. A Monte Carlo simulation is used to determine which street segment will be added to the route. Once a segment is chosen, the links connected to the sink node are then considered and the process is repeated. Once the destination link is reached the route generation is complete.

ArcGIS is then used to generate the route shapefile. The centroid for each link in the generated route is added to ArcGIS route solver as a 'stop' to ensure that ArcGIS creates the generated route. Previously, the junctions within the generated route were used as the stops in ArcGIS route solver but errors would occur in the shapefile. The error occurred because ArcGIS has the ability to approach stops from a specified direction. However, if the stop is in the middle of an intersection, additional links may be added to properly approach the stop. This type of problem does not occur when the stop is along a link, which is why the centroids of the links were used.

4.2. Path Size Logit Choice Model

The choice model specification used for this study was based on the PSL with a stochastic choice set generated through the biased random walk algorithm (Ben-Akiva & Bierlaire, 1999; Frejinger, 2007). As first discussed in Chapter 2, the formula for the PSL with stochastic choice set mode is given as:

$$P(i|C_n) = \frac{e^{\mu(V_{in}+\ln(PS_{in}))+\ln\left(\frac{k_{in}}{q(i)}\right)}}{\sum_{j\in C_n} e^{\mu(V_{jn}+\ln(PS_{jn}))+\ln\left(\frac{k_{jn}}{q(j)}\right)}}$$

[4.2]

Where:

C_n is the choice set for user n (includes chosen route)
μ is the logit scale term
V_{in} is systematic utility for alternative i for user n
PS_{in} is the expanded path size factor for alternative i for user n
k_{in} is the number of times alternative i is randomly drawn. If chosen route, $k_{in}+1$
q(i) is the probability of choosing a route containing the street segments. It is calculated as the product of each link choice probability

The PS factor used in this formulation was the 'Generalized PS' factor introduced by Ramming (2001). Through preliminary testing with the study data set, it was determined that other formulations of PS factors and the EPS factor were not significant in the model. The Generalized PS is given by the formula:

$$PS_{in} = \sum_{a\epsilon\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \left(\frac{L_i}{L_j}\right)^{\Phi} \delta_{aj}}$$

[4.3]

Where:

Γ_i is the set of links in path i
L_a is the length of link a
L_i is the length of path i
L_j is the length of path j
δ_{aj} equals 1 if link a is on path j and 0 otherwise
φ is a parameter that controls the impact of route length in the correction factor. A value of 14 was used based on a study by Hoogendorn-Lanser et al.(2005).

4.2.1. Path Size Factor Sensitivity Analysis

A sensitivity analysis was conducted on the value of Φ in the Generalized PS factor to see how values of Φ compare to Φ=14 as presented in Hoogendorn-Lanser et al. (2005). Φ was tested with the Toronto GPS walk data and a simple model of length and turns. Φ values ranging from 0 to 30 were tested with the rho squared of the model recorded. Figure 4.1 presents the change in rho across values of Φ.

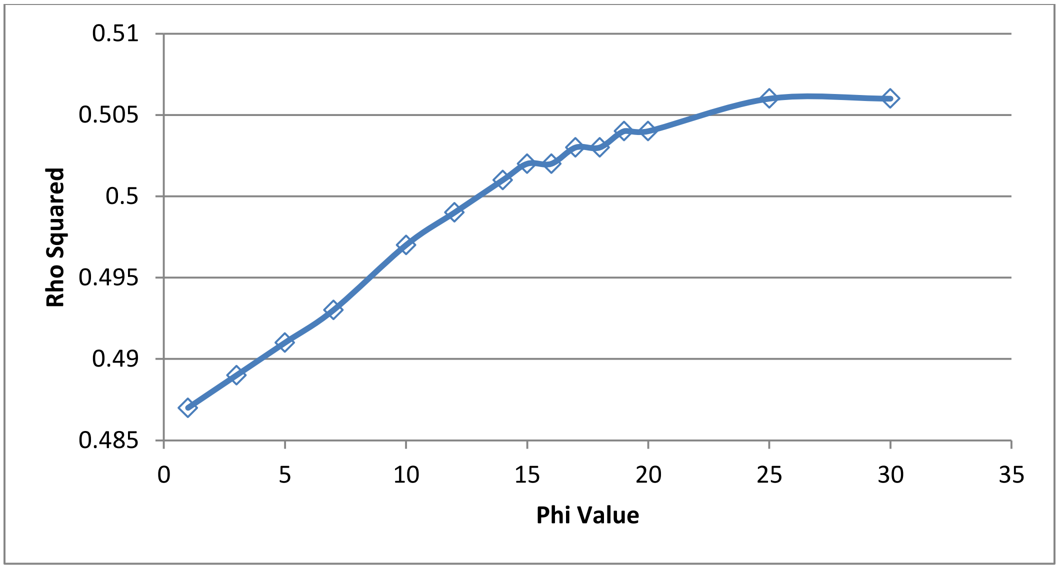


Figure 4.1 - Phi sensitivity analysis

It was found that as Φ increased the overall fit of the model increased. With the Smartphone walk trip data set, the PS correction factor was not significant for values of $\Phi \leq 1$. As phi increases, the impact of longer routes in the choice set is decreased. Decreasing the correction from overlap with longer routes increases the significance of the correction factor and improves the overall model. This suggests that out of the generated alternatives, the observed route is more likely to be one of the shorter routes. This could suggest that either the observed routes are more likely to be short routes with minor deviations or the route choice generator had a tendency to generate routes longer than the observed path. For phi values above 14 the increase to rho-squared decreases suggesting an asymptotic relationship. While the model rho-squared values do not peak at a value of Φ=14, the results of this analysis are in line with the findings of Hoogendorn-Lanser et al.(2005): using a value of Φ=14 in the Generalized PS formulation produced the best fit model. Thus, for the route choice analysis in this study, a value of Φ=14 was used.

4.3. Model Testing Procedure

PSL models were estimated using Biogeme 2.5. Biogeme is an open source freeware designed by Michael Bierlaire for the maximum likelihood estimation of parametric models (Bierlaire, 2016).

The starting base model included two route variables: length of route (meters) and the number of turns within the route. Each model also included a "ln(PS)" variable to correct for correlation from overlapping paths in the choice set and a "Sample_Correction" variable to adjust for the probability of route being generated in the stochastic route generation process.

A new variable was added to the base model and tested for significance (Significance set at 95%). If the variable was not significant, the model test number and result would be noted. Variables with significance above 80% would have the significance and sign of coefficient noted.

Tested models with variables at 95% significance or above would have the rho-squared, adjusted-rho-squared, final log likelihood, and signs of variable beta parameters recorded in a table. After all the variables in the route variable list were tested individually, the variable to be added to the base was chosen based on the rho-squared value, sign of the parameter, and a chi-squared test on the log likelihoods. The variable with a high rho-squared value, intuitive sign, and which is significantly different from the base model is added to the base and the process is repeated.

Once no more route variables could be added to the model, socioeconomic and time variables were interacted with the route variable list and further tested.

4.4. Limited Observations per Individual vs Full Dataset

The walk trip dataset includes multiple observations per user. During the survey period, individuals would have the app installed and recorded full-day tours. These tours were broken down into various trips. The number of trips recorded for each individual generally is not equal. Additionally, some users uninstalled the app before the survey period was complete, thus having much fewer recorded trips. Due to users uninstalling the app, recorded trips having large gaps, or trips not having enough variation, the number of observations per individual is very inconsistent. The large variation in number of observations causes concerns of over-representation of some individuals within the full dataset. Certain individuals have over 100 observations while other users have only one observation. Estimating a model assuming all trips to be independent would result in a user with more observations having a greater influence on the estimation of parameters than a user with a low number of observations. For example, within the 71 individuals sampled from the dataset, 46% of the users are female. However, when trips are treated as independent, the percentage of female trips decreases to 33%. The change in socioeconomic ratios has a noticeable impact on the estimation of the model when socioeconomic variables are used.

Various data set characteristics were plotted as the observation per individual limit was increased from 1 to 167; Figure 4.2 to Figure 4.6 illustrate the change. Table 4.1 illustrates how the proportion of certain socioeconomic variables change as the user observation limit increases.

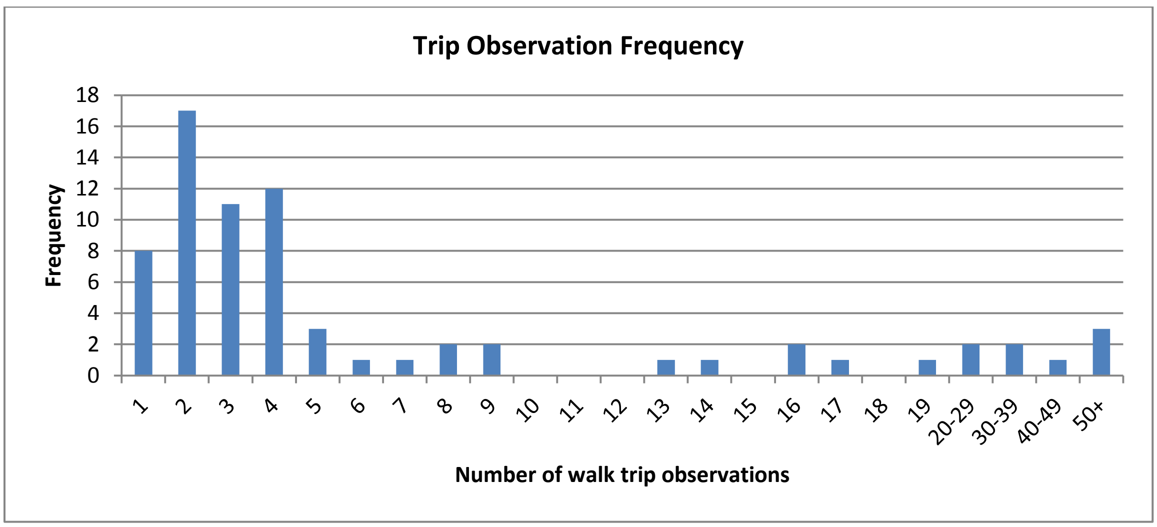


Figure 4.2 - Trip observation frequency

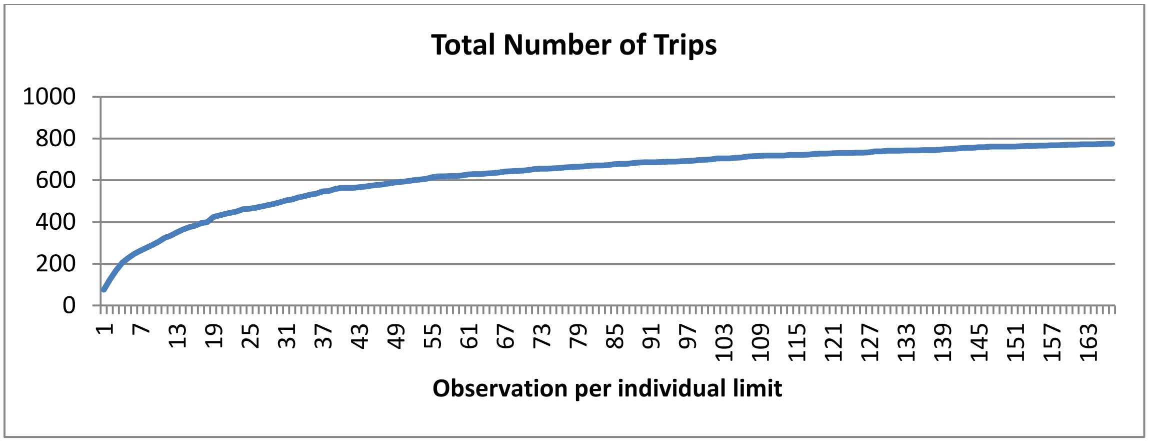


Figure 4.3 - Dataset size with limited observations

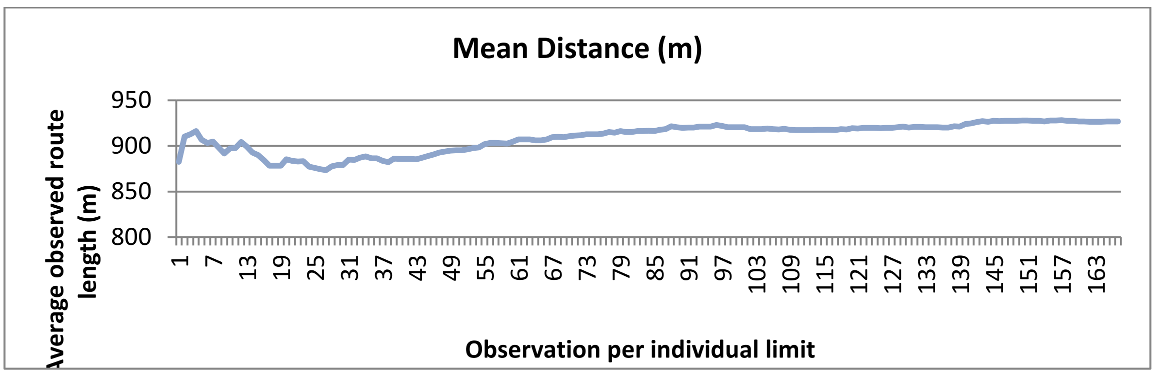


Figure 4.4 - Change in mean distance with observation limit

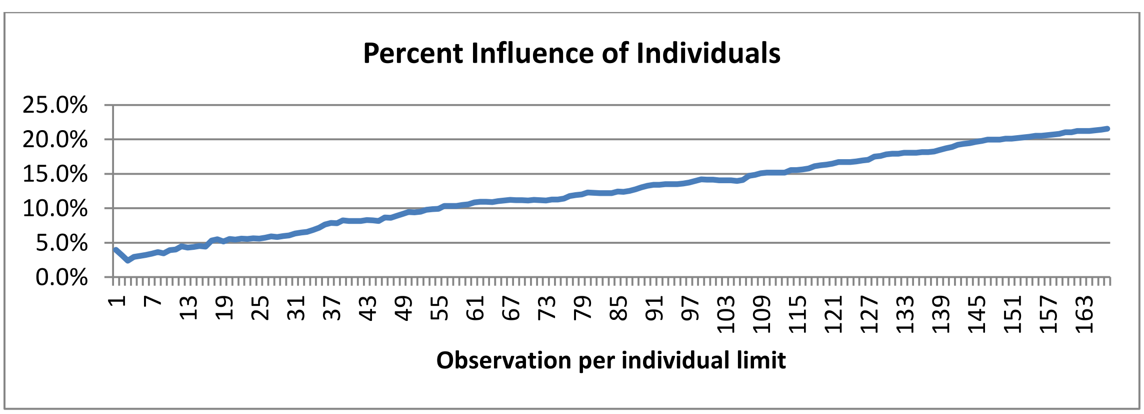


Figure 4.5 - Maximum percent share from an individual

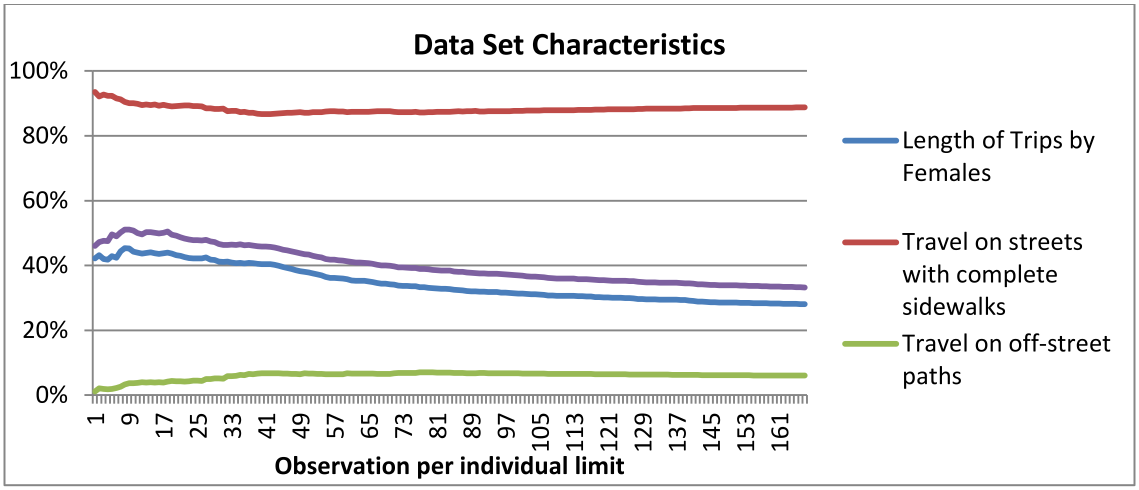


Figure 4.6 - Change in dataset characteristics with observation limit

Table 4.1 - Change in demographics with observation limit

Panel Limit	1	30	167
Total Trips	71	511	776
Total Users	71	71	71
Male	57.6%	53.4%	66.8%
Female	42.4%	46.6%	33.2%
Income0 (Less than or equal to $19,999)	7.6%	12.1%	13.5%
Income1 ($20,000-$49,999)	10.6%	18.4%	30.7%
Income2 ($50,000-$74,999)	19.7%	16.8%	12.5%
Income3 ($75,000-$124,999)	30.3%	24.1%	24.5%
Income4 (More than or equal to $125,000)	24.2%	22.1%	14.6%
Employed – Full time	51.5%	38.9%	25.6%
Employed – Part time	13.6%	11.2%	7.3%
Student – Full time	24.2%	23.7%	38.5%
Student – Part time	3.0%	2.9%	1.9%
Student (Full time or part time)	27.3%	26.6%	40.5%
Age1 (User age between 0-18)	3.0%	11.0%	8.6%

Age2 (User age between 19-25)	18.2%	20.9%	20.5%
Age3 (User age between 26-35)	45.5%	33.7%	39.6%
Age4 (User age between 36-45)	15.2%	15.1%	9.9%
Age5 (User age between 46-55)	7.6%	5.3%	3.5%
Age6 (User age over 55)	10.6%	14.1%	17.9%

Figure 4.2 graphs the frequency of observations per individual. It should be noted that even though the average number of trips per individual is 9.6, most users have less than five observed trips. There are three users with more than 50 walk trips recorded which are skewing the average.

Figure 4.3 illustrates how the size of the dataset changes with the observation limit. Above 60 observations per individual, the dataset increase is more gradual showing that there are many fewer individuals with more than 60 observations. If each of the individuals had the same amount of observations, with a limit 30 trips per individual, the size of the dataset would be 2130 trips. However, introducing a limit of 30 trips leaves only 511 trips in the dataset.

Figure 4.4 shows that the average distance travelled increases as the limit of observations per individual increases. This suggests that users with more recorded observations generally took longer walking trips.

As seen in Figure 4.5, without accounting for panel effects or limiting the number of observations per user, one user accounts for about 20% of the trips in the data set. This means that this user's route choice preferences account for 20% of the model results.

It is shown in Figure 4.6 that total distance traveled on streets with complete sidewalks and on off-street paths stays relatively the same when imposing a limit on the number of observations per individual however, the percentage of females and, by-extension, the total distance travelled by females decreases by 13% when any observation limit is removed.

As seen in Table 4.1, as the observation limit increases, the percentage of males increases, the percentage of full-time students increases, as does the share of people income level below $50,000/year. There are small increases in the number of individuals under 25 and individuals over 55. People who are employed, either full time or part time, have a large decrease as the panel limit increases.

There's a trade-off between having a limited sample which preserves the demographic characteristics of the users and having more walk trip data which would help to improve the explanatory power of the model. However, the users in this dataset represent a convenience sample collected for the purpose of exploring and analyzing the use of smartphone GPS data for pedestrian route choice. It was decided that instead of limiting the dataset to preserve the demographic characteristics of the users, the full dataset would be used for a better powered model.

Since the data includes a varying number of observations per individual, controlling the panel data effect was considered through the use of a mixed effect logit model. However, through preliminary model testing, it was found that panel effect variables were not significant in the models. A similar result was found by Broach et al. (2012), where mixed logit specifications which included individual-specific effects resulted in mean parameter values that were not sensitive to panel effects. Other research has shown that when discrete choice models with unlabelled alternatives are estimated, the formulation for controlling potential panel effects becomes extremely complex (Hsiao, 1986; Train, 2009). In addition, as the choice set becomes large, the estimation of parameters has been seen to become unstable which raises concerns about distributional assumptions and estimator consistency (Broach et al., 2012; Hensher & Greene, 2002; Nerella & Bhat, 2004). For these reasons, panel data effects were not considered and trips were assumed to be independent for this analysis.

5. Toronto Case Study: Model Specification and Results

This section describes the pedestrian route choice model estimated from smartphone GPS data in the City of Toronto. The boundaries for Toronto are south of Steeles Ave, east of Etiboke creek/Highway 427, and west of the Scarborough-Pickering Townline.

After data cleaning, the Toronto street network used in this study contains 59,951 links. Table 5.1 outlines various link characteristics of the network.

Table 5.1 - Toronto network characteristics

Total Number of Links	59951
Percent of links with…	
Sidewalk on both sides	49%
Sidewalk on one side	12%
No sidewalk	15%
Road classification…	
Major Arterial	11%
Minor Arterial	6%
Collector	11%
Local	41%
Walkway/Trail	12%
Other	19%
Number of…	
Signalized Intersections	2416
Pedestrian Crossovers	482

5.1. Toronto Walk Trip Data

The data collected from users was a continuous day tour of GPS points. The users installed an app that would remain running throughout the day. The app was always recording the change in location and was not dependent on the individual manually designating the start and stop of a trip which means that the trip data are not influenced by reporting bias; this, however, means that trip origins and destinations must be inferred. For this dataset, a trip was only considered to have ended after the user had stayed in relatively the same location for a period of three minutes.

A large portion of the walking trips recorded in the smartphone travel survey were ‘short’ walking trips. To ensure that trips used in the model showed variability in street characteristics, walk trips were only considered if three or more unique alternative routes could be generated between the trip origin and destination.

Recreational trips were also removed from the dataset dataset since these routes were often excessively long or returned to the origin. Recreational trips were identified as trips that traveled more than twice the shortest route length between the origin and destination.

After removing the short and recreational trips, 776 trips remained across 71 individuals. Trip characteristics for the observed walk trips are shown in Table 5.2 and user demographics are shown in Table 5.3.

Table 5.2 - Observed walk trip characteristics

Total Number of Trips	776
Number of Users	71
Average Number of Trips	9.6
Most Frequent Number of Trips	2
Max Number of Trips per User	167
Trips by Females	28.0%
Mean Distance (m)	926.8
Travel on streets with complete sidewalks	88.8%
Travel on off-street paths	6.0%

Table 5.3 - Demographic characteristics of user sample

User Sample Size	71
Gender	
Male	54%
Female	46%
Income	
Less than or equal to $19,999	8%
$20,000-$49,999	15%
$50,000-$74,999	20%
$75,000-$124,999	30%
More than or equal to $125,000	23%
Employment Status	
Full Time	51%
Part Time	18%
Student Status	
Full Time	23%
Part Time	4%
Age	
0-18	6%
19-25	18%
26-35	49%
36-45	13%
46-55	7%
56+	7%

This dataset was obtained through filtering walking trips from a larger dataset previously collected. This is dataset represents a convenience sample that is used to undertake an exploratory, prototype analysis of walking trips based on GPS smartphone data. Thus, the results cannot be extrapolated to the Toronto population as a whole.

<u>Alternative route characteristics</u>

A stochastic route generation approach known as the biased random walk algorithm was used to generate the choice set for this model. The biased random walk algorithm was developed by Frejinger (2007) and modified by Broach and Dill (2015). A detailed outline of the route generation formulation can be found in Section 4.1. Trip characteristics for the generated alternatives are provided in Table 5.4.

Table 5.4 - Alternative route characteristics

Mean Distance (m)	1000.6
Travel on streets with complete sidewalks	80.2%
Travel on off-street paths	4.2%
Average Number of Unique Alternatives	7.4

The generated alternatives have a slightly higher average length. Both the percent travelled on streets with complete sidewalks and percent travelled on off-street paths were lower than the observed trips.

5.2. Route Variables

The mean, standard deviation, minimum values, and maximum values of the observed route variables can be found in Appendix B. Table 5.5 describes the various route variables used in the model.

Observed trips had an average length of 927m (Standard Deviation (SD)=526m). The average length of road walked with sidewalks on both sides is 823m (SD=534m). Collector road classification had the highest average length travelled among individual road types at 338m (SD=458m).

Out of the land uses, low density residential had the highest average length of travel at 264m (SD=273m) which on average accounted for 15% of the route. On average, other land uses were traversed much less than low density residential.

Roads in areas with a Walk Score greater than or equal to 70 points were traversed on an average of 845m (SD=564m).

Table 5.5 - Route variables

Name	***Description***
Length	Total route length
Turns	Total number of turns in route
Sidewalk both sides	Length of road (m) with sidewalk on both sides
Signalized Intersection	Number of signalized intersections in route
Minor arterial road	Length of route (m) on minor arterial road
Arterial Road	Length of route (m) on major or minor arterial road
Collector road	Length of route (m) on collector road
Land commercial	Length of route (m) with commercial land use frontage
Land office	Length of route (m) with office land use frontage
Land park	Length of route (m) with park land use frontage
Percent land park	Percent of route with park land use
PS	Path size correction factor
Sample correction	Probabilistic sampling correction factor
Additional variables tested	Pedestrian crossovers, steep slopes, major arterial road, local road, incomplete sidewalk, Walk Score, low residential land, high residential land, industrial land, institutional land

Roads with slopes over 5% were not travelled often. However, this could be due to the overall topography of downtown Toronto, which does not have many steep slopes.

Pedestrian crossovers had a low number of observations within the chosen routes but this may be due to the fact that there are only 480 in the city of Toronto.

The length of the route was seen to have a strong positive correlation (0.74) to the number of signalized intersections in the route. Length also had a strong correlation to Sidewalk-Both-Sides (0.90), as expected. An interesting correlation was seen between Walk Score and the following variables: Length (0.91), Sidewalk-Both-Sides (0.91), and Signalized Intersections (0.72). A full correlation matrix for the observed route variables can be found in Appendix C.

Interaction Terms

Interaction terms were used to capture taste variation of the individuals (Train, 2009). Dummy variables are a useful tool for determining how coefficients vary across socioeconomic attributes or different times of the day. A large range of interaction terms were tested with the route variables. A full list of the socio-economic and time variables used in the analysis is outlined in Table 5.6. During the modelling procedure, some of the groups may be merged into larger groups depending on the effect.

Table 5.6 - Interaction variables

Name	***Description****
Gender_Female	User gender female
Age1	User age between 0-18
Age2	User age between 19-25
Age3	User age between 26-35
Age4	User age between 36-45
Age5	User age between 46-55
Age6	User age $\geq$ 56
Emp_Full	Full-time employment status
Emp_Part	Part-time employment status
Student_Full	Full-time student status
Student_Part	Part-time employment status
Student	Student status (full-time or part-time)
Income0	Less than or equal to $19,999
Income1	$20,000-$49,999
Income2	$50,000-$74,999
Income3	$75,000-$124,999
Income4	More than or equal to $125,000
Weekend	Trip took place on the weekend
Time_Morning1	Time of trip between 7AM-8AM
Time_Morning2	Time of trip between 8AM-9AM

Time_Morning3	Time of trip between 9AM-10AM
Time_Evening1	Time of trip between 4PM-5PM
Time_Evening2	Time of trip between 5PM-6PM
Time_Evening3	Time of trip between 6PM-7PM
Time_Night	Time of trip after 9PM

**Variables are dummy variables equal to 1 if true, 0 otherwise*

5.3. Model Results

Two models are presented in this section. The first is a general model which contains variables without interaction terms while the second model includes additional variables that were only significant after interacting with socio-economic or time of day variables.

The reasons two models are presented are due to the size of the dataset as well as the potential over-representation of certain individuals' preferences through multiple observations in the dataset. The final models' parameters are presented in Table 5.7 and Table 5.8. A list of parameters for the model iterations can be found in Appendix D.

In the final general model, the overall fit was strong and the variables were all significant. The signs for the variables were all as expected. The utility equation for route i is given by the following equation:

$$U_i = \beta_{Length} * Length_i + \beta_{Turns} * Number\ of\ Turns_i + \beta_{Sidewalk\ Both\ Sides} * Length\ Sidewalk\ Both\ Sides_i + \beta_{Siginalized\ Intersection} * Number\ of\ Signalized\ Intersections_i + \beta_{PS} * \ln(PS_i) + \ln\left(\frac{k(i)}{q(i)}\right) \quad [5.1]$$

The final model with interaction terms also had a strong fit. The variables were all significant at 5% significance. The path size factor was significant and positive in both models which suggests that its inclusion improved estimation of the models. The interaction term model's utility equation for route i is given by the following equation:

$$
\begin{aligned}
U_i = {} & \beta_{Length} * Length_i + \beta_{Length\ Female} * Length_i * Gender_{Female} + \beta_{Turns} \\
& * Number\ of\ Turns_i + \beta_{Sidewalk\ Both\ Sides} * Length\ Sidewalk\ Both\ Sides_i \\
& + \beta_{Siginalized\ Intersection} * Number\ of\ Signalized\ Intersections_i \\
& + \beta_{Minor\ Arterial\ Student} * Length\ on\ Minor\ Arterial_i * Student \\
& + \beta_{Arterial\ Age\ 25} * Length\ on\ Arterial_i * Age\ under\ 25 \\
& + \beta_{Minor\ Arterial\ Income} * Length\ on\ Minor\ Arterial_i * Income\ over\ \$75{,}000 \\
& + \beta_{Parks\ Evening} * Length\ in\ Park_i * Evening + \beta_{Commercial\ Employed} \\
& * Length\ along\ Commercial\ Land_i * Employed + \beta_{Collecter\ Age\ 45} \\
& * Length\ on\ Collector_i * Age\ over\ 45 + \beta_{Office\ Age\ 25} \\
& * Length\ along\ Office\ Land_i * Age\ under\ 25 + \beta_{Walkway\ Age\ 25} \\
& * Length\ along\ Walkways_i * Age\ under\ 25 + \beta_{PS} * \ln(PS_i) + \ln\left(\frac{k(i)}{q(i)}\right)
\end{aligned}
\quad [5.2]
$$

There was some estimated correlation between variables within the models which was expected considering the route variables were generally measured in terms of length. The correlation was within reason and did not have a significant impact on the model. The estimated variable correlations can be found in Appendix D.

In the final general model, Sidewalk-Both-Sides of the street was correlated to length of route by a value of -0.518 and length was also negatively correlated with turns by a value of -0.252. The path size factor had a significant correlation to each of the variables in the model but the correlation never exceeded a value of 0.3.

In the interaction term model, length and turns had a negative correlation of -0.238, arterial roads interacted with age less than 25 was correlated with signalized intersections with a value of 0.377, Sidewalk-Both-Sides was negatively correlated with length by a value of -0.498, and the path size factor correlated to length by 0.246.

Table 5.7 - Final general model results

	Coefficient*	t-stat
Length (m)	-0.02	-12.93
Turns	-0.645	-12.52
Length with sidewalk on both sides of the road	0.00665	7.99
Number of signalized intersections	0.669	6.99
ln(EPS)	1.53	7.06
Log-likelihood (Null)	-1488.946	
Log-likelihood (Model)	-785.99	
Rho squared	0.472	
N	776	

* all coefficients significant at 5%

Table 5.8 - Final interaction term model

	Coefficient*	t-stat
Length (m)	-0.0198	-11.81
Length (m) x Female	-0.00788	-2.55
Turns	-0.724	-13.13
Length with sidewalk on both sides of the road	0.0073	8.56
Number of signalized intersections	0.729	6.98
Length on minor arterial roads as a student	-0.00333	-4.79
Length on major or minor arterial roads when under age 25	0.00337	4.47
Length on minor arterial roads when income >$75,000/yr	0.0029	2.81
Length along parks after 4PM	-0.0214	-2.72
Length along commercial land use when employed	-0.00911	-2.61
Length on collector roads when over age 45	-0.00319	-2.87
Length along office land use when under age 25	-0.0143	-2.50
Length along walkways when under age 25	0.0145	2.88
ln(PS)	1.39	6.26
Log-likelihood (Null)	-1488.946	
Log-likelihood (Model)	-742.723	
Rho squared	0.501	
N	776	

* all coefficients significant at $p<0.05$

Distance trade-offs for attributes

By calculating the ratio of coefficients of the model parameters, the trade-offs between variables can be determined. Comparing attribute parameters to the length coefficient gives the distance equivalent of said attribute. Table 5.9 and Table 5.10 present the distance equivalent of street characteristics.

Table 5.9 - Distance equivalent of attributes for general model

Attribute	**Distance Equivalent (m)**
Per additional:	
Turn	+32.41
Signalized Intersection	-33.62
Change in perceived distance along:	
Sidewalk both sides	-33%

Table 5.10 - Distance equivalent of attributes for interaction term model

Attribute	**Distance Equivalent (m)**	
	Male	**Female**
Per additional:		
Turn	+36.57	+26.16
Signalized Intersection	-36.63	-26.34
Change in perceived distance along:		
Sidewalk both sides	-37%	-26%
Minor arterial as a student	17%	12%
Arterial road as a person under 25	-17%	-12%
Minor arterial road as a person with income over $75k/yr	-15%	-10%
Park land use after 4 PM	+108%	+77%
Commercial land use as a employed person (full or part-time)	+46%	+33%
Collector road as a person over 45	+16%	+12%
Office land use as a person under 25	+72%	+52%
Walkway land use as a person under 25	-73%	-52%

5.4. Path Size Factor Analysis

The other formulations of PS were tested with the model and only formulation 3, Generalized PS factor, was found to be significant. An analysis of the mean and standard deviation of the various PS formulations was conducted.

Table 5.11 presents a summary of the PS formulations and Table 5.12 shows the results of the PS analysis.

Table 5.11 - Path size factor formulation summary

Study	Path Size Equation	Variable Name
Ben-Akiva and Bierlaire (1999)	$PS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon C_n} \delta_{aj}}$	ln(PS1)
Ben-Akiva and Bierlaire (1999)	$PS_{in} = \sum_{a \epsilon \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon C_n} \frac{L^*_{C_n}}{L_j} \delta_{aj}}$	ln(PS2)
Ramming (2002)	$PS_{in} = \sum_{a \epsilon \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \epsilon C_n} \left(\frac{L_i}{L_j}\right)^{\phi} \delta_{aj}}$	ln(PS3)
Bovy et al. (2008)	$PSC_{in} = \sum_{a \epsilon \Gamma_i} \frac{L_a}{L_i} \ln\left(\frac{1}{\sum_{j \epsilon C_n} \delta_{aj}}\right)$	PS4
Frejinger, Bierlaire, and Ben-Akiva (2009)	$EPS_{in} = \sum_{a \in \Gamma i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj} \Phi_{jn}}$	EPS5

(A full description of the PS equations can be found in Section 2.3.2)

Table 5.12 - Path size factor formulation characteristics

PS Formulation	Mean	Standard Deviation
ln(PS1)	-0.986	0.328
ln(PS2)	-0.858	0.326
ln(PS3)	-1.006	0.354
PS4	-1.155	0.357
ln(EPS5)	-3.406	0.613

The formulations for PS factors have a maximum value of zero. When ln(PS) has a value of zero, it means that there is no overlap within the choice set.

PS1, PS2, PS3, and PS4 have similar average values around -1.0 but EPS5 is more than three times larger at a value of -3.4.

The major difference in the formulation of EPS5 is that it takes into account the probability of generating the route. If the probability of the route is very small the ln(EPS5) value becomes a large negative value. An average value of -3.4 suggests that the routes in the choice set have a small probability of being generated. The low probability of generation may be due to the rule governing the alternative route generator that penalizes links for going away from the destination.

Appendix E presents the distributions of average path size factor per individual.

PS1, PS2, PS3, and PS4 have much wider distributions. PS1, PS2, and PS4 have distributions that are slightly right-skewed while PS3 has a normal distribution. The distribution for ln(EPS5) is very narrow with the majority of values between -3.85 and -2.65.

PS3 operates similar to the EPS5 formulation but instead of decreasing the impact from unlikely routes because of generation probability, it decreases the impact because of excessive length. Possibly, PS3 corrects the overlap similar to EPS5 but is not affected by the generation probability.

5.5. Route Generation Analysis

The stochastic route generation method was used to generate 10 alternatives with various levels of calibration. The choice set size for this analysis was chosen based on the characteristics of the walk trip dataset. The dataset includes a number of short walking trips which do not allow for a high degree of variability. Generating more than 10 alternatives for these trips would not increase the number of unique alternatives for these short trips, thus it was decided to generate only 10 alternatives per observed trip.

The first set of alternatives is biased around the shortest path. The second set biased around the route of least cost where each turn is +50 meters and roads with sidewalks are perceived as 10% shorter distance. The third set of alternatives is biased around the route of least cost using the calibrated values from the route choice analysis. While the model results give an equivalent distance for signalized intersections, ArcGIS route solver is not able to handle the “negative” distance value. In order for ArcGIS route solver to solve the path of least cost, non-signalized intersections were set as extra distance while signalized intersections added zero distance. The results of the generation scenarios are outlined in Table 5.13 and Table 5.14.

Table 5.13 - Observed route draw probability

Generation scenario	Average probability of drawing observed route	Probability of drawing observed route at least once
Biased around shortest path	21.3%	53.7%
Biased around least cost	20.8%	52.1%
Biased around calibrated least cost	21.2%	51.9%

Table 5.14 - Observed route characteristics

	Value	
Number of trips where…		
Least cost probability >= shortest path probability	584	75%
Calibrated least cost probability >= shortest path probability	572	74%
Calibrated least cost probability >= least cost probability	578	74%
Average route length where…		
Least cost probability >= shortest path probability	960.2	
Least cost probability < shortest path probability	810.1	
Calibrated least cost probability >= shortest path probability	991.9	
Calibrated least cost probability < shortest path probability	730.0	
Calibrated least cost probability >= least cost probability	985.7	
Calibrated least cost probability < least cost probability	740.2	

Table 5.15 summarizes the average difference between the observed route characteristics and the average choice set characteristics.

Table 5.15 - Average difference between observed route and choice set characteristics

	Length	**Turn**	**Signalized Intersection**	**Sidewalk Both**
Shortest	6.6%	-1.37	-0.29	13.3%
General Cost	6.9%	-1.33	-0.26	14.5%
Calibrated Cost	7.6%	1.27	-0.01	29.9%

5.6. Discussion

5.6.1. Model results

Length

The length had a negative coefficient as expected. It was found that females were more sensitive to changes in route length than males; Females experienced a higher disutility from distance than males did. While length does not have the largest coefficient in the model, when considering the average length of the observed route is 926m, length is the dominant variable in the route choice.

Turns

Turns had a negative coefficient which is similar to the results of other route choice studies. An excessive number of turns can be perceived as a wayfinding issue. Straight and simple paths are more attractive than routes with many turns. The results of the general model showed that turns were perceived as the equivalent of walking an additional 32 meters. The results of the interaction term model were that turns were equivalent to an additional 37 meters for males and 26 meters for females. The quantitative values for turns were found to be similar to the research by Broach & Dill (2015) which determined that turns were perceived as an additional 54 meters to the route.

Length of Sidewalk on Both Sides

The length of sidewalk on both sides had a positive coefficient in both models showing that sidewalk completeness (i.e. sidewalk on both sides) is an attractive attribute. This result supports other research on the influence of built environment on route choice which has shown that complete sidewalk infrastructure, sidewalk quality, and sidewalk width are desirable for pedestrians (Rodríguez et al., 2015).

The results from the model put a quantitative value to the perceived benefit of sidewalks. The distance trade-off for the general model shows that links with sidewalks on both sides are perceived as 33% shorter in length. The results of the interaction term model had length of sidewalks on both sides of the street as 37% shorter for males and 26% shorter for females.

As seen previously in Table 5.1, half of the links in the Toronto network were classified as having sidewalk on both sides of the street. Areas with sidewalks on only one side of the road or no sidewalks are/were typically low-density residential areas.

Streets with sidewalks on only one side of the street may be perceived as an inconvenience or safety concern. If a pedestrian enters a street with only one sidewalk, they may have to make an additional crossing or risk walking along the road. Similarly, streets with no sidewalk can be seen as a safety concern. Pedestrians would either have to walk along the road, walk along the various properties on the street, or choose another street.

Signalized Intersections

Signalized intersections have a positive coefficient, which suggests that people prefer more signalized intersections within their route. This could be a direct result of signalized intersections providing safe crossing at busy intersections. The study by Broach & Dill (2014) found that crossing at an unsignalized arterial was equivalent to adding 73 meters to the route and crossing a collector road without a marked crosswalk was perceived as an additional 28 meters to the route.

The results from the general model show that pedestrians would travel 32 meters extra to get to a signalized intersection. The results of interaction term model found that each additional signalized intersection in the route was equivalent to a reduction of 37 meters for males and 26 meters for females. While these results are not directly comparable they do suggest that signalized intersections are desirable in a route for safe crossing.

However, there are some other possible reasons for this result that should be considered. The positive coefficient of signalized intersections could also be a by-product of network density if more walking trips are observed in an area with a dense grid network with traffic signals than in residential neighbourhoods without signalized intersections. The literature has shown that there is a strong relationship between street network density and the amount of walking that occurs (Herrmann et al., 2017; Lamíquiz & López-Domínguez, 2015; Ozbil et al., 2011; Sarkar et al., 2015). In addition, a study by Guo (2009) found that an increase of one more intersection per 100m increased route utility by 0.3min.

The positive coefficient could also be capturing a preference towards busier roads. Collector and arterials roads are more likely to have signalized intersections than residential streets. Similarly, busier roads tend to have street lighting which provides a sense of safety, higher volume of cars which can give a sense of populated streets, and in the winter the sidewalk on busy streets may be more likely to be shoveled/plowed allowing for a more comfortable walk.

Future research should use a higher GPS accuracy to determine the amount of street crossing at signalized intersections rather than the number of signalized intersections in the route.

Future work could also examine how street classification and intersection size impact crossing at signalized intersections. It would be expected that when crossing streets with high volumes of vehicles, signalized intersections would be highly desired. Similarly, signalized crossings would be preferable for wide intersections. When crossing a small street at an intersection, the benefit from the signalization may be lower since the low volume of vehicles would allow for ample gaps for crossing.

Length on Minor Arterial Roads as a Student

In the interaction term model, length on minor arterial roads as a student had a negative coefficient and was determined to be perceived as 17% longer for males and 12% longer for females. However, length of major or minor arterial roads when under age 25 had a positive coefficient and was perceived as 17% shorter for males and 12% shorter for females.

Out of the 71 individuals, there are 19 students, 17 individuals under 25, and 10 students under the age of 25. However since users have multiple trips recorded, in the dataset there were 630 student trips, 516 trips by individuals under 25, and 210 trips by students under 25.

Length on minor arterial roads as a student had a negative coefficient but length on major and minor arterial roads when under the age of 25 had a positive coefficient of similar value. If an individual is a student under the age of 25, walking along a minor arterial will have negligible influence from these two variables. Students under age 25 will experience positive utility from walking along major arterials.

This result is difficult to interpret due to the overlap between students and individuals under age 25. Students and individuals under 25 are two of the socioeconomic variables which experienced large change due to the multiple observations. This result could be due to over-representation of the two socioeconomic variables. The sample of users is comprised of 27% students whereas the walk trip sample contains 41% students. Similarly, 24% of individuals were under the age of 25 while trips made by individuals under the age of 25 represented 29% of the trips.

Length on Minor Arterial Roads when Income >$75,000/yr

The interaction term model showed that length on minor arterial roads for individuals with household income over $75,000 per year had a positive coefficient and was perceived as 15% shorter in distance for males and 10% shorter in distance for females.

These results are contrary to the findings in a study by Broach and Dill (2015) found that busy roads, collector or larger were perceived as 14% longer. Possible reasons for a difference in the relationship could stem from the interaction with income. Individuals with household income over $75,000 could be more likely to have home locations or destinations along minor arterial roads which would make minor arterial roads more preferable.

Length through Parks after 4PM

Walking through the park after 4PM was found to have a negative coefficient. Walking through the park in the evening may not be desirable because of the feeling of safety. The results of the model show that the distance of walking through a park in the evening is perceived as 108% longer in length for males and 77% longer in length for females. Whether it's the amount of lighting or an absence of people, walking through parks in the evening is not desirable. A possible reason for this result could also be that after work, people may walk along commercial/busy streets for dinner or for errands.

The results showed that length through parks in the evening has less of an impact for females than it did for males which is contrary to what may be expected. The reason for this result is because the distance trade-off was calculated by interacting gender and disutility from length. Females are more likely to minimize distance and this affects the perception of other route variables. Since females associate a higher cost to distance than males, when comparing the cost of other variables to distance, the other route variables have less of an equivalent distance for females compared to males.

This result should not be taken as a direct interpretation of gender effects with respect to perception of length through parks in the evening. This is a metric of length through parks in the evening compared with each gender's general perception of distance.

When length through parks after 4PM interacted with gender the result was not significant. It is expected that females more than males would have a higher aversion from length along parks in the evening due to safety concerns, however, this is not supported in these results. Further study with a larger sample and more detailed GPS data and park attributes would be needed for a complete analysis.

The negative coefficient from parking in the evening supports other pedestrian research where it was found that lack of lighting and absence of people were found to be deterrents for walking (Schlossberg, Agrawal, Irvin, & Bekkouche, 2007; Ferrer, Ruiz, & Mars, 2015). It would be interesting to explore the effect of lighting within parks or the size of the park on this outcome.

Some additional information that would be useful for this analysis would be the quality of the park paths. Parks can have various qualities of path ranging from a dirt path to a paved walkway. Lower quality walkways would be expected to be less desirable than fully paved walkways.

Length along Commercial Land Use when Employed

It was found that pedestrians who listed themselves as employed experienced disutility from walking along commercial land use. From the interaction term model, it was determined that length along commercial land use for employed individuals was perceived as 46% longer for males and 33% longer for females.

This result is contradictory to what is expected, which is that employed pedestrians would have more disposable income and would be more willing to walk along a commercial land use since there are more opportunities/destinations. This result could be reflecting that employed pedestrians have a high value of time and would rather not go out of their way to walk along a commercial land use.

The lack differentiation between trip purposes in the model could be affecting the results. If commercial or recreational walking trip purposes were investigated then its possible that commercial land use would have a positive coefficient because of its aesthetics and destinations. However, since recreational trips are excluded and since the trip purpose was not given, the desire to avoid increasing walking distance unnecessarily may be captured in this variable.

Length on Collector Roads when over Age 45

It was hypothesized that walking along busy streets would be perceived as a longer distance. Research by Broach and Dill (2015) found that busy streets (collector or larger) were not as attractive as other types of road classification; busy streets were perceived as 14% longer.

The findings from the interaction term model support the previous research; length on collector roads when over age 45 were perceived as 16% longer for males and 12% longer for females. Collector roads can have a large volume of traffic. Depending on the width of the sidewalk, walking along a road with a heavy volume of fast moving vehicles may feel unsafe or may expose pedestrians to a noticeable concentration of vehicle emissions.

Length along Office Land Use when under Age 25

Office land use was had a negative perception among individuals under age 25. Length along office land use when under age 25 is perceived as 72% longer for males and 52% longer for females. It was not expected that office land use would be significant with individuals under 25.

This result could be reflecting that the destinations individuals under age 25 are walking to are not the office areas in Toronto. A large percentage of the individuals under 25 are students and not employed. These individuals may be heading to school/campus rather than an office area. Walking along office land use may be out of the way from where individuals under 25 are going which may explain this result.

Length along Walkways when under Age 25

Walkways or pedestrian paths are built for pedestrian use and are typically more aesthetically pleasing. These paths have lighting, green space, and are devoid of automobile traffic. The results of the interaction term model showed that length along walkways when under age 25 is perceived as 73% shorter in distance for males and 52% shorter for females. This supports similar findings in the study by Lin and Hsia (2007) which found paths and walkways promoted active transportation.

Variables Not Included in the Final Model/Interesting Relationships

‘No sidewalk’ had a negative coefficient and was significant in early model versions but it was left out of the final models in favour for ‘Sidewalk_Both’ which produced a better fitting model. In Toronto, the number of streets with a sidewalk on both sides greatly outnumbers the number of streets with no sidewalk. Even though the model has a higher rho-squared with ‘Sidewalks_Both’, it may actually provide more useful information using ‘No sidewalk’.

Walk Score was not significant in the model. A possible reason for this may be due to the process of assigning the neighbourhood Walk Score to the network links. When transitioning from one neighbourhood to another, the Walk Score could drastically change. This would mean that two neighboring/intersecting streets could have very different Walk Scores. Also, a Walk Score is a measurement of pedestrian-friendly attributes within an area. This includes sidewalk widths and destinations. These variables are also accounted for in the route choice model so the lack of significance may be due to a relationship between the variables. Walk Score may be more influential in choosing whether to walk to a destination rather than which route to walk to get to the destination. Walk Score provides a measure of how walkable an area is, but a high Walk Score does not necessarily mean there will be a high percentage of walking (Carr et al., 2010; Herrmann et al., 2017).

Another possible reason Walk Score was not significant is that the downtown Toronto area generally has a Walk Score of 70 or above. This Walk Score is classified as ‘very walkable’. Walk Score might not be significant because the Walk Score only varies between 70 and 100 in most of the areas with observed trips. The effect on route choice from a difference in Walk Score between 70 and 100 may not be as noticeable as the effect from lower Walk Scores. While Walk Score may be useful for determining walking rates for neighborhoods, the results of this study suggest that Walk Score does not have a significant effect on people choosing routes.

The study by Broach and Dill (2015) found that upslopes of 10 percent are seen as twice as costly as less steep ground and a study by Guo (2009) found that people were willing to walk 2.9 minutes to avoid hilly topography. However, various levels of slopes were tested as variables in the models but none were significant. This could possibly be a result of the topography of the

Toronto versus the study area of Portland. Toronto is relatively flat with changes in slope being very gradual.

High-density land use and mixed use land use were expected to have a positive effect/influence on pedestrian route choice. However, neither one of these variables was significant in the final models.

Land Use

Land use variables were not significant in the model on their own but some were significant when interacted with socioeconomic variables. However, even when the land use variables were significant the coefficient was not always an expected sign. Research by Rodríguez et al. (2015) found that while trade-offs between route distance and route quality did occur, the effect of amenities was less consistent.

Similar issues with consistency occurred during the modelling process. When testing models, with a model consisting of length, turns, and sidewalk both sides, land classified with high-density residential, commercial, office, and institutional land uses had a positive coefficient (but not significant at 95%). These relationships are to be expected; however, once signalized intersections were introduced into the model, the sign for the coefficients for high density residential and commercial land became negative. Another unexpected relationship was that industrial land was appearing to have a positive coefficient in the model while the expectation is that industrial land use is unattractive to pedestrians.

While these results were not significant, they do pose questions on the effect of land use and measurement of land use for route choice. Studies on walking rates typically use land use or amenity density around the destination (Herrmann et al., 2017). Studies on route choice have used land use frontage along the route (Broach & Dill, 2015). While this study used land use frontage, the frontage was calculated from topographical land use parcels. Other route choice studies used observed data collected through trained observers traversing through neighborhoods for the analysis (Borst et al., 2009). Observational data on land use may prove to be more useful than GIS parcel information. Land use parcels may be labelled as commercial or office but from a street perspective, these land uses may not be apparent. Observation data measurements would be able to collect data on the aesthetics of the street characteristics which cannot be found in land use parcel data, such as building type or window shops/ boulevards.

It is difficult to determine whether the results of these models are representative of general pedestrian behaviour in Toronto or if the results are mainly idiosyncratic of the current small dataset. With only 71 different users, it is possible that this outcome is influenced by certain individual's habitual routes. An individual's single route may appear many times within the data set. The characteristics of habitual routes that are repeatedly recorded may be influencing the results. Students and individuals under the age of 25 were two of the socioeconomic groups that saw an increase in representation due to multiple trip observations.

5.6.2. GPS Data

Sample Size

The 71 users within the dataset represent a convenience sample for the purpose of exploratory, prototype analysis of pedestrian route choice using smartphone GPS data. While the sample does represent a subset of the population of Toronto, the results can't be fully extrapolated to the Toronto population as a whole. Further analysis should collect a much larger sample that has demographic characteristics similar to that of the Toronto population as a whole.

Trip Purpose

GPS data did not have the trip purpose which could have affected the results. Recreational trips may have a different preference for route choice compared to a work trip or shopping trip. When going to work, the priority may be minimizing travel time. However, coming home late from work or going on a shopping trip may have a route that travels on commercial streets.

The way trips were segmented by dwelling time, it is possible that walking trips to access transit or other modes of transportation were recorded. These trips may prioritize travel time rather than street infrastructure due to the timing of bus/train arrivals.

GPS Accuracy

GPS points every 50m is suitable for tracing the overall route but is not high enough accuracy to determine finer travel behaviour such as which side of the street the individual is on, or where the individual decided to cross the street. In addition, there is some error in the GPS accuracy which also affects the ability to measure detailed route behaviour. The combination of minimum 50m travel distance and GPS noise results in GPS points scattered on either side of the street of travel seeming to appear as if the individual is crossing the same street multiple times.

When solving for the observed route, a 40m radius buffer area around the GPS points is used to solve for the route. 40m was chosen because it encompasses a sufficient distance to account for the minimum 50m travel distance as well as GPS noise. However, such a large buffer area means that it is possible for alternative segments to be present within the buffer area. With alternative route segments, the route solver may generate a path that is not actually the observed route. A higher accuracy and more frequent GPS point logging would allow for the route solving buffer to be reduced and decrease the likelihood of non-observed paths being used.

For more detailed analysis, more frequent GPS traces should be collected. With more GPS points it may be possible to determine which movement is from GPS noise and which movement is from the individual crossing the street.

Mode Identification

Using speed data to infer mode may work for a percentage of the time when traffic is flowing freely, however, during peak hours/traffic congestion, it becomes increasingly difficult to determine whether a person is walking or slowly moving through stop and go traffic. In the walk trip data, there are some trips that travel over two kilometers and also happen to follow along a streetcar line. Also, given the data collection period of November/December, the weather conditions make reduce the plausibility that some of the longer walking trips recorded were indeed walking trips. Considering the mode identification accuracy had a 3.1% error for identifying transit trips as walking trips, it is possible that these trips were not walking trips.

Map-Matching

Matching the GPS points to the network and determining the chosen route was an initial challenge for this project. With the minimum 50m travel distance between points in addition to GPS noise, the points did not always fall on the correct street links. Using a buffer approach similar to Dalumpines and Scott (2011) proved to be an easy to implement and computationally efficient route solving process. The buffer size could be increased or decreased depending on the accuracy of the GPS points. While this approach did have issues where multiple routes could exist within a large buffer area, discontinuous buffer areas, or off network pedestrian travel, this process was successful at matching the majority of the observed routes to the network.

5.6.3. Route Generation Analysis

Overall, the average probability of drawing the observed route did not increase after calibration. Contrary to what was expected, the probability of drawing the observed route at least once decreased. There is still randomness within the route generation process that allows for branching sub-optimal routes to be generated which may be a reason why the average draw probability did not increase. The probability of drawing the observed route at least once likely decreased because the values of route characteristics changes with the trip length. Shorter walk trips are likely governed by distance more than other route characteristics. The influence of trip length is suggested by the results in Table 5.14. Even though the probability of drawing the observed route decreases with calibration, it is seen that the general least cost generator either performed the same or better than the biased shortest path method 75% of the time. In addition, the calibrated route generator either performed the same or better than the biased shortest path and biased least cost methods for 74% of the trip observations.

The general least cost method and the calibrated least cost method had a higher probability of drawing the observed route for longer trips. This suggests that for shorter trips, the impacts of route characteristics besides length have less of an impact. The shorter the route, the more impact deviations from the shortest path will have. For example in a 600m trip, detouring 30m to avoid a turn may be less desirable because it makes the route 5% longer. This ratio of detour length to shortest distance may be an important factor to consider when examining pedestrians' preferences to route characteristics. For low distances the ratio would be high suggesting that distance would be the only important factor but as the length of trip gets longer, the influence other route characteristics may increase.

Looking at the street characteristics from the alternative route choice sets, the results are contrary to what was expected. The shortest and general cost alternative choice sets out-performed the calibrated cost choice set in terms of length of route and distance travelled on links with sidewalks on both sides.

On average, all choice sets had longer routes and more travel on links with sidewalk on both sides than the observed route. The shortest and general cost alternative choice sets had, on average, less turns and signalized intersections than the observed route while the calibrated cost alternatives had more turns.

The calibrated cost alternative choice set had a much larger proportion of travel on sidewalks with both sides than the observed route as well as the other choice sets. Considering that sidewalks on both sides were perceived as 33% shorter after calibration, this reduction in distance could be the reason for increased number of turns and length. The route generator may be more likely to search out links with sidewalk on both sides for the distance reduction even though the route would be longer and have more turns.

While the calibrated choice set generates routes with more turns on average, the difference in turns between observed and generated routes was slightly less than the other choice sets. The area where the calibrated cost choice set noticeably performed better than the other choice sets was in number of signalized intersections. A possible way of improving the likelihood of the calibrated cost generator to generating plausible routes would be to increase the α and β parameters to make the route generator more sensitive to cost.

The biased-around-the-shortest-path algorithm has the benefit that it can generate a large number of alternative routes with a wide spread of variation. It was also found to have the highest probability of generating the observed route at least once. In terms of generating routes with variation in the street characteristics, this procedure appears to be the best among those tested.

6. Recursive Logit (RL)

The RL developed by Fosgerau et al. (2013) has the advantage of not being bound by a limited choice set but instead can consider all links in a network. The Matlab code for the RL model was generously provided by Fosgerau et al. for use in this project.

The idea of using the RL formulation was explored during this study; although, there were some issues with developing the RL for this data set. The Toronto network needed to be converted from the ArcGIS shapefile format to a matrix format. A link incidence matrix needed to be created from the GIS network for expressing which links were connected to each other. Links in the matrix format are uni-directional and since pedestrians can move in either direction on any link, the entire network had to be duplicated to represent both directions. The Toronto network, originally about 60,000 links, then became a network of about 120,000 links.

The Toronto network is very large and enumeration of possible paths was computationally intense. When attempting to run the RL model, an error would occur saying that the "requested 121950x122188(111.0 gb) array exceeds maximum array size preference". A possible solution to this error is to reduce the size of the network by removing areas where no walk trips were observed but it is unclear whether the size of that reduced network would be acceptable computationally.

Preliminary testing of the RL model occurred with a section of the network and a small sample of 20 walk trips. The model estimation requires starting beta values for the parameters; however, the final estimated beta values were influenced based on the starting values. Without a clear idea of the expected valued of beta, determining which starting values should be used would be difficult. If the beta values were not in the right range, the model estimation would fail. Using coefficients from the estimated PSL model, the RL model failed to converge. Possible reasons for the non-convergence of the model may be the result of too small of a sample being used or from specification sensitivity. For the reasons previously mentioned, the recursive logit was not used for this dataset and was left for future work.

7. Conclusions

Smartphone GPS data proved to be a viable source for pedestrian route choice. The route choice model developed from smartphone GPS data produced results consistent with other pedestrian route choice research. The results showed that the primary consideration for pedestrians is distance. Turns were found to be equivalent to an additional 32m while signalized intersections were determined to be equivalent to a reduction of 34m. Travel along streets with sidewalks on both sides of the road was perceived as 33% shorter than streets with other sidewalk conditions. Contrary to what was expected, Walk Score, high-density residential land use, and mixed use land use variables were not significant in the model.

An alternative model was estimated to test the effect of various socioeconomic variables. However, due to the small sample size, the significance of the interaction terms with route choice variables may be the result of curve fitting a small sample rather than showing behavioural tendencies of pedestrians in Toronto. It is also possible that due to the low number of individuals and trip observations some of the variables could be showing more idiosyncratic results rather than general population behaviour. Further research should be conducted using large samples to determine whether the results from this interaction model are a reflection of general pedestrian route choice behaviour or are a result of the small sample.

Through testing the ‘biased around the shortest path’ stochastic alternative route generation method, it was found that biasing the generation around shortest path was more likely to generate the alternative route than biasing the generation around a calibrated least cost route. However, the calibrated generation algorithm was more likely to generate the observed route for longer trips.

Recursive logit provides a possible way to avoid the route generation issue but preliminary testing showed issues analyzing the large Toronto network.

Areas of improvement for pedestrian route choice models include using observational data of street level attributes in addition to land use parcel data. Land use frontage calculated from parcel data gives a general idea of the street level attributes but it also overlooks detailed characteristics of the land use. Also, the trip purpose should be included in the data so that the effect on route behaviour can be explored.

Future research on pedestrian route choice should utilize a tool/app with a more frequent and accurate GPS setting to explore street crossing behaviour. The effect of distance on the perception of route variables should also be explored to determine at what distance thresholds route variables start to have an effect on route choice. Data on sidewalk width was not available within this study, but it would be interesting to see the effect of sidewalk width and sidewalk completeness on the route choice.

8. References

Agrawal, A. W., & Schimek, P. (2007). Extent and correlates of walking in the USA. *Transportation Research Part D: Transport and Environment*, *12*(8), 548–563.

Agrawal, A. W., Schlossberg, M., & Irvin, K. (2008). How far, by which route and why? A spatial analysis of pedestrian preference. *Journal of Urban Design*, *13*(1), 81–98.

Aytur, S. A., Rodriguez, D. A., Evenson, K. R., Catellier, D. J., & Rosamond, W. D. (2007). Promoting active community environments through land use and transportation planning. *American Journal of Health Promotion*, *21*(4_suppl), 397–407.

Azevedo, J., Costa, M. E. O. S., Madeira, J. J. E. S., & Martins, E. Q. V. (1993). An algorithm for the ranking of shortest paths. *European Journal of Operational Research*, *69*(1), 97–106.

Badoe, D. A., & Miller, E. J. (2000). Transportation–land-use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research Part D: Transport and Environment*, *5*(4), 235–263.

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., & Martin, B. W. (2012). Correlates of physical activity: why are some people physically active and others not? *The Lancet*, *380*(9838), 258–271. https://doi.org/10.1016/S0140-6736(12)60735-1

Ben-Akiva, M., Bergman, M. J., Daly, A. J., & Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour. In *Proceedings of the 9th International Symposium on Transportation and Traffic Theory* (pp. 299–330). VNU Science Press Utrecht, The Netherlands. Retrieved from https://books.google.ca/books?hl=en&lr=&id=nc1GCPRsH-4C&oi=fnd&pg=PA299&dq=Modelling+inter+urban+route+choice+behaviour&ots=pUdh2b4VnW&sig=J3U_3xeOZYT2H7wSqh6usLT404k

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions. In *Handbook of transportation science* (pp. 5–33). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-5203-1_2

Bierlaire, M. (2016). Biogeme. Retrieved January 29, 2017, from http://biogeme.epfl.ch/home.html

Borst, H. C., de Vries, S. I., Graham, J. M., van Dongen, J. E., Bakker, I., & Miedema, H. M. (2009). Influence of environmental street characteristics on walking route choice of elderly people. *Journal of Environmental Psychology*, *29*(4), 477–484.

Borst, H. C., Miedema, H. M., de Vries, S. I., Graham, J. M., & van Dongen, J. E. (2008). Relationships between street characteristics and perceived attractiveness for walking reported by elderly people. *Journal of Environmental Psychology*, *28*(4), 353–361.

Bouthelier, F., & Daganzo, C. F. (1979). Aggregation with multinomial probit and estimation of disaggregate models with aggregate data: A new methodological approach. *Transportation Research Part B: Methodological*, *13*(2), 133–146.

Bovy, P., Bekhor, S., & Prato, C. (2008). The Factor of Revisited Path Size: Alternative Derivation. *Transportation Research Record: Journal of the Transportation Research Board*, *2076*, 132–140. https://doi.org/10.3141/2076-15

Bovy, P. H., & Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica*, *3*(3), 173–189.

Broach, J., & Dill, J. (2015). Pedestrian Route Choice Model Estimated from Revealed Preference GPS Data. In *Transportation Research Board 94th Annual Meeting*. Retrieved from https://trid.trb.org/view.aspx?id=1338221

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740.

Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010). Walk score™ as a global estimate of neighborhood walkability. *American Journal of Preventive Medicine*, *39*(5), 460–463.

Cascetta, E., A, N., F, R., & A, V. (1996). A modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results. In *ResearchGate*. Retrieved from https://www.researchgate.net/publication/258507831_A_modified_Logit_Route_Choice_Model_Overcoming_Path_Overlapping_Problems_Specification_and_Some_Calibration_Results

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.

City of Toronto. (2016). Built Environment - Environmental Health - Toronto Public Health | City of Toronto. Retrieved January 28, 2017, from http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=d06e23bf6d481410VgnVCM10000071d60f89RCRD

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016). Representing pedestrian activity in travel demand models: Framework and application. *Journal of Transport Geography*, *52*, 111–122.

Dalumpines, R., & Scott, D. M. (2011). GIS-based Map-matching: Development and Demonstration of a Postprocessing Map-matching Algorithm for Transportation Research. In *ResearchGate* (pp. 101–120). https://doi.org/10.1007/978-3-642-19789-5_6

de la Barra, T., Perez, B., & Anez, J. (1993). Multidimensional path search and assignment (Vol. P 366). Presented at the PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom. Retrieved from https://trid.trb.org/view.aspx?id=1169872

Ferreira, I. A., Johansson, M., Sternudd, C., & Fornara, F. (2016). Transport walking in urban neighbourhoods—Impact of perceived neighbourhood qualities and emotional relationship. *Landscape and Urban Planning*, *150*, 60–69.

Ferrer, S., Ruiz, T., & Mars, L. (2015). A qualitative study on the role of the built environment for short walking trips. *Transportation Research Part F: Traffic Psychology and Behaviour*, *33*, 141–160.

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, *56*, 70–80. https://doi.org/10.1016/j.trb.2013.07.012

Frank, L. D., Sallis, J. F., Saelens, B. E., Leary, L., Cain, K., Conway, T. L., & Hess, P. M. (2010). The development of a walkability index: application to the Neighborhood Quality of Life Study. *British Journal of Sports Medicine*, *44*(13), 924–933.

Frejinger, E. (2007). Random Sampling of Alternatives in a Route Choice Context. *ResearchGate*. Retrieved from https://www.researchgate.net/publication/37456566_Random_Sampling_of_Alternatives_in_a_Route_Choice_Context

Frejinger, E., & Bierlaire, M. (2009). Route Choice Analysis: Data, Models, Algorithms and Applications. In *The 12th International Conference on Travel Behaviour Research*. Retrieved from https://infoscience.epfl.ch/record/152428/files/IATBR09_EMMA.pdf

Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Expanded Path Size attribute for route choice models including sampling correction. In *International Choice Modelling Conference 2009*. Retrieved from http://www.icmconference.org.uk/index.php/icmc/icmc2009/paper/view/52

Guo, Z. (2009). Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston. *Transportation Research Part D: Transport and Environment*, *14*(5), 343–352.

Harding, C., Zhang, Y., & Miller, E. J. (2015). Multiple purpose tours and efficient trip chaining in Toronto: an analysis of the effects of land use and transit provision on mode choice and trip chaining using smartphone data. In *ResearchGate*. Retrieved from https://www.researchgate.net/publication/281455393_Multiple_purpose_tours_and_efficient_trip_chaining_in_Toronto_an_analysis_of_the_effects_of_land_use_and_transit_provision_on_mode_choice_and_trip_chaining_using_smartphone_data

Haybatollahi, M., Czepkiewicz, M., Laatikainen, T., & Kyttä, M. (2015). Neighbourhood preferences, active travel behaviour, and built environment: an exploratory study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *29*, 57–69.

Hensher, D. A., & Greene, W. H. (2002). The mixed logit model: the state of practice and warnings for the unwary. Retrieved from https://trid.trb.org/view.aspx?id=585797

Herrmann, T., Boisjoly, G., Ross, N. A., & El-Geneidy, A. M. (2017). *The Missing Middle: Filling the Gap Between Walkability and Observed Walking Behavior*. Retrieved from https://trid.trb.org/view.aspx?id=1438284

Hoogendoorn-Lanser, S. (2005). Modelling Travel Behaviour in Multi-modal Networks. *TRAIL Thesis Series*. Retrieved from https://trid.trb.org/view.aspx?id=760541

Hsiao, C. (1986). *Analysis of Panel Data, Econometric Society Monograph No. 11*. Cambridge: Cambridge University Press.

Koh, P. P., & Wong, Y. D. (2013a). Comparing pedestrians' needs and behaviours in different land use environments. *Journal of Transport Geography*, *26*, 43–50.

Koh, P. P., & Wong, Y. D. (2013b). Influence of infrastructural compatibility factors on walking and cycling route choices. *Journal of Environmental Psychology*, *36*, 202–213.

Lamíquiz, P. J., & López-Domínguez, J. (2015). Effects of built environment on walking at the neighbourhood scale. A new role for street networks by modelling their configurational accessibility? *Transportation Research Part A: Policy and Practice*, *74*, 148–163.

Lin, J.-J., & Hsia, T.-T. (2007). Effect of built environment on active transportation: a Taipei study. In *Proceedings of the Eastern Asia Society for Transportation Studies The 7th International Conference of Eastern Asia Society for Transportation Studies, 2007* (pp. 211–211). Eastern Asia Society for Transportation Studies. Retrieved from https://www.jstage.jst.go.jp/article/eastpro/2007/0/2007_0_211/_article/-char/ja/

Manaugh, K., & El-Geneidy, A. (2011). Validating walkability indices: How do different households respond to the walkability of their neighborhood? *Transportation Research Part D: Transport and Environment*, *16*(4), 309–315.

Mehta, V. (2008). Walkable streets: pedestrian behavior, perceptions and attitudes. *Journal of Urbanism*, *1*(3), 217–245.

Nerella, S., & Bhat, C. (2004). Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record: Journal of the Transportation Research Board*, (1894), 11–19.

Ozbil, A., Peponis, J., & Stone, B. (2011). Understanding the link between street connectivity, land use and pedestrian flows. *Urban Design International*, *16*(2), 125–141.

Prato, C. G., & Bekhor, S. (2006). Applying Branch-and-Bound Technique to Route Choice Set Generation. *ResearchGate*, *1985*(1985), 19–28. https://doi.org/10.3141/1985-03

Ramming, M. S. (2002). *Network knowledge and route choice*. Massachusetts Institute of Technology. Retrieved from https://dspace.mit.edu/handle/1721.1/49797

Rodríguez, D. A., Brisson, E. M., & Estupiñán, N. (2009). The relationship between segment-level built environment attributes and pedestrian activity around Bogota's BRT stations. *Transportation Research Part D: Transport and Environment*, *14*(7), 470–478.

Rodríguez, D. A., Merlin, L., Prato, C. G., Conway, T. L., Cohen, D., Elder, J. P., … Veblen-Mortenson, S. (2015). Influence of the built environment on pedestrian route choices of adolescent girls. *Environment and Behavior*, *47*(4), 359–394.

Sarkar, C., Webster, C., Pryor, M., Tang, D., Melbourne, S., Zhang, X., & Jianzheng, L. (2015). Exploring associations between urban green, street design and walking: Results from the Greater London boroughs. *Landscape and Urban Planning*, *143*, 112–125.

Sheung, Y., & Shalaby, A. S. (2006). An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys. In *85th Annual Meeting of the Transportation Research Board*.

Singleton, P. A., & Clifton, K. J. (2013). Pedestrians in Regional Travel Demand Forecasting Models: State of the Practice. In *Transportation Research Board 92nd Annual Meeting*. Retrieved from https://trid.trb.org/view.aspx?id=1242847

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

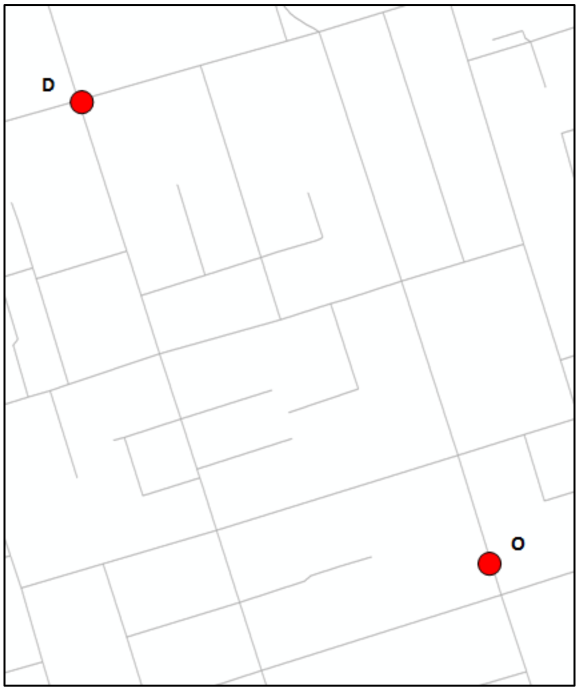
Van der Zijpp, N. J., & Catalano, S. F. (2005). Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B: Methodological*, *39*(6), 545–563.

Walk Score. (2016). How Walk Score Works. Retrieved January 28, 2017, from https://www.redfin.com/how-walk-score-works

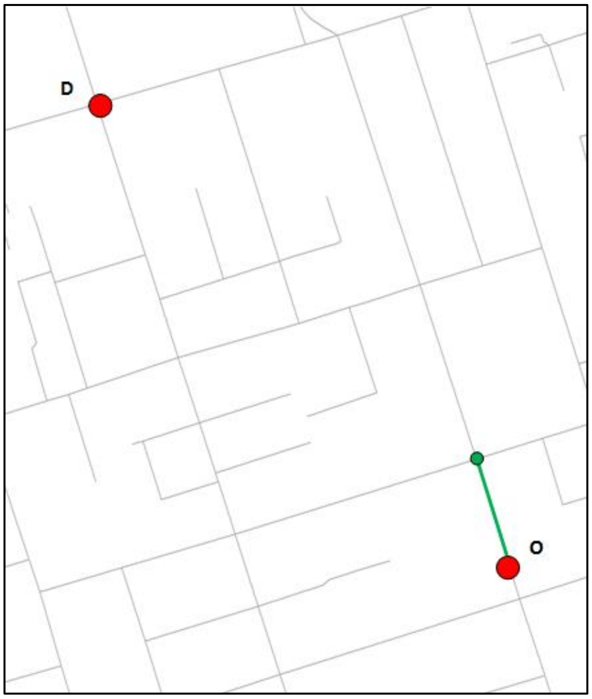
Appendix A

Stochastic Route Generation Process

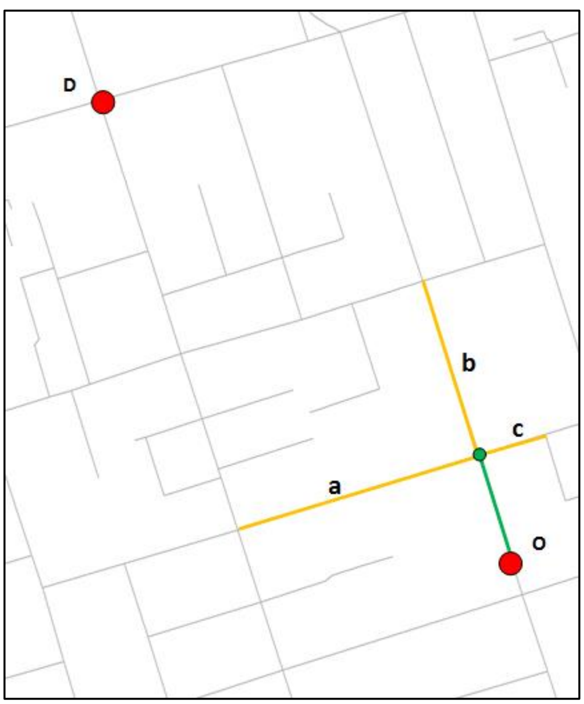
1. Import Origin and Destination



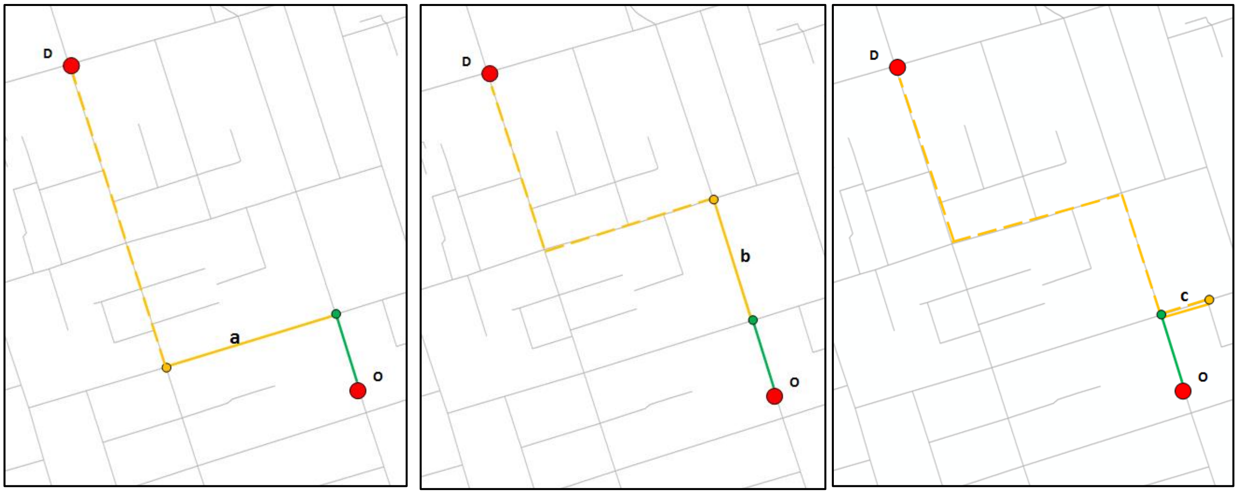
2. Determine the origin street segment



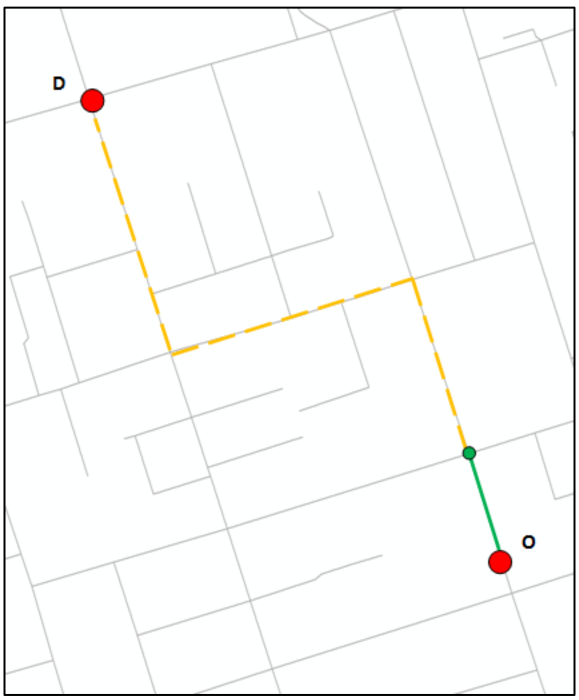
3. Find the street segments connected to the source node



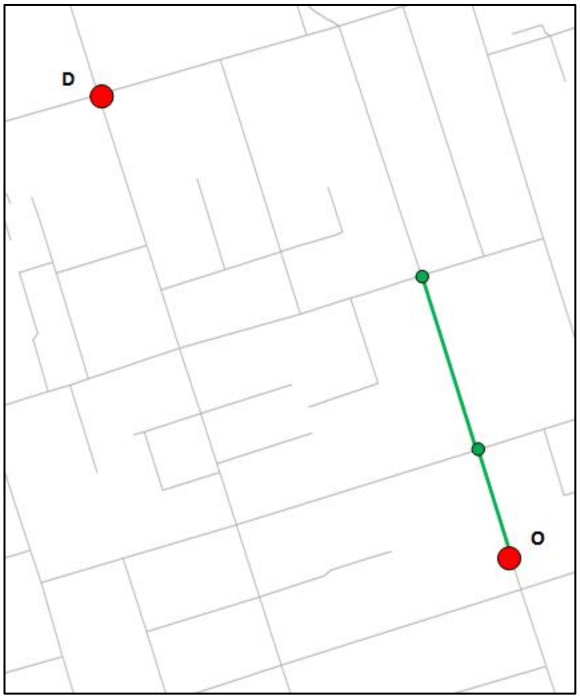
4. Determine the cost for each street segment



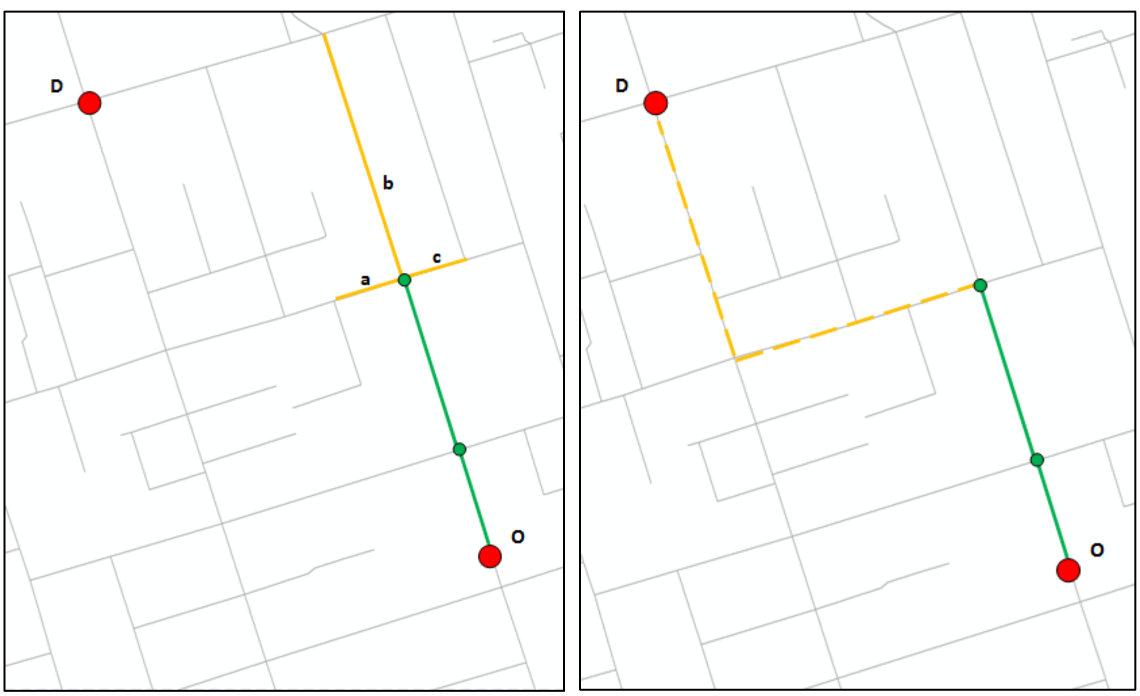
5. Determine the cost from the source node to the destination



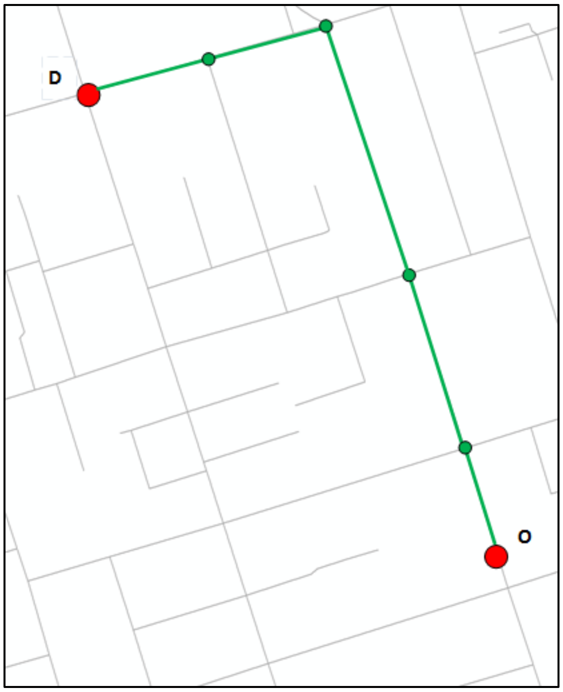
6. Calculate probabilities and use Monte Carlo simulation to select the next segment



7. Repeat Process for new selected segment and source node



8. Once the destination segment is reached, stop process and generate route



Appendix B

Observed Route Characteristics

Variable	Mean	Standard Deviation	Min	Max
Total_Length_1	926.84	525.73	151.43	4333.48
Total_Turn_1	2.56	1.77	0.00	12.00
Total_General_Cost_1	971.94	521.40	172.08	4249.37
Total_Time_1	617.89	350.49	100.95	2888.98
Total_Sig_Int_1	2.75	2.50	0.00	18.00
Ped_Crossover_1	0.36	0.73	0.00	6.00
Max_Slope_1	0.03	0.06	0.00	0.49
Steep_Slope1_1	8.01	32.96	0.00	499.50
Steep_Slope2_1	3.07	12.41	0.00	150.37
Cumul_Prob_1	0.22	0.26	0.00	1.00
Road_Collector_1	338.17	457.92	0.00	2888.71
Road_Local_1	207.40	363.55	0.00	2414.99
Road_Minor_Art_1	107.11	177.18	0.00	992.93
Road_Major_Art_1	189.08	259.30	0.00	1846.63
Road_Walkway_1	67.26	195.31	0.00	1608.30
Road_Other_1	17.81	63.71	0.00	688.42
SWLK_None_1	18.94	54.81	0.00	591.25
SWLK_BothSides_1	822.68	553.92	0.00	4333.48
SWLK_OneSide_1	6.22	30.19	0.00	281.24
SWLK_Walkway_1	55.93	197.96	0.00	1590.63
SWLK_Other_1	23.07	78.23	0.00	843.96
Land_LowRes_1	264.02	273.67	0.00	1620.41
Land_MedRes_1	0.01	0.38	0.00	10.59
Land_HighRes_1	22.88	55.01	0.00	402.94
Land_Commercial_1	36.70	73.16	0.00	549.04
Land_Industrial_1	9.29	30.22	0.00	202.95
Land_Office_1	30.47	78.62	0.00	697.03
Land_Mixed_1	15.84	31.96	0.00	225.02
Land_Park_1	10.12	35.19	0.00	321.80
Land_Institutional_1	33.74	67.36	0.00	420.83
Land_Walkway_1	43.33	76.97	0.00	459.93
Land_LowRes_Avg_1	132.01	136.83	0.00	810.20
Land_MedRes_Avg_1	0.01	0.19	0.00	5.30
Land_HighRes_Avg_1	11.44	27.51	0.00	201.47
Land_Commercial_Avg_1	18.35	36.58	0.00	274.52
Land_Industrial_Avg_1	4.64	15.11	0.00	101.47
Land_Office_Avg_1	15.24	39.31	0.00	348.51
Land_Mixed_Avg_1	7.92	15.98	0.00	112.51

Variable	Mean	Standard Deviation	Min	Max
Land_Park_Avg_1	5.06	17.60	0.00	160.90
Land_Institutional_Avg_1	16.87	33.68	0.00	210.42
Land_Walkway_Avg_1	21.66	38.49	0.00	229.97
Land_LowRes_Percent_1	0.15	0.14	0.00	0.88
Land_MedRes_Percent_1	0.00	0.00	0.00	0.01
Land_HighRes_Percent_1	0.01	0.03	0.00	0.30
Land_Commercial_Percent_1	0.02	0.04	0.00	0.32
Land_Industrial_Percent_1	0.01	0.02	0.00	0.18
Land_Office_Percent_1	0.02	0.04	0.00	0.42
Land_Mixed_Percent_1	0.01	0.02	0.00	0.20
Land_Park_Percent_1	0.01	0.02	0.00	0.24
Land_Institutional_Percent_1	0.02	0.03	0.00	0.27
Land_Walkway_Percent_1	0.02	0.04	0.00	0.27
Land_Through_Park_1	0.65	0.48	0.00	1.00
Land_Through_Mixed_1	0.14	0.35	0.00	1.00
Avg_Walkscore_1	88.82	11.33	41.00	100.00
Walkscore_Length1_1	845.49	564.05	0.00	4271.06
Walkscore_Length2_1	545.10	533.02	0.00	3186.14
Walkscore_Length3_1	81.34	240.40	0.00	1635.88
Walkscore_Length4_1	6.56	59.26	0.00	744.14
DrawnTimes_1	2.90	2.19	1.00	10.00

Appendix C

Observed Route Variable Correlation Matrix

	Total_Length_1	Total_Turn_1	Total_General_Cost_1	Total_Time_1	Total_Sig_Int_1	Ped_Crossover_1	Max_Slope_1	Steep_Slope1_1	Steep_Slope2_1	Cumul_Prob_1	Road_Collector_1
Total_Length_1	1.00										
Total_Turn_1	0.44	1.00									
Total_General_Cost_1	0.99	0.57	1.00								
Total_Time_1	1.00	0.44	0.99	1.00							
Total_Sig_Int_1	0.74	0.26	0.70	0.74	1.00						
Ped_Crossover_1	0.41	0.06	0.37	0.41	0.24	1.00					
Max_Slope_1	0.30	0.15	0.30	0.30	0.12	0.10	1.00				
Steep_Slope1_1	0.21	0.10	0.21	0.21	-0.01	0.05	0.49	1.00			
Steep_Slope2_1	0.27	0.10	0.26	0.27	0.07	0.09	0.76	0.63	1.00		
Cumul_Prob_1	-0.51	-0.41	-0.54	-0.51	-0.35	-0.23	-0.17	-0.12	-0.13	1.00	
Road_Collector_1	0.59	0.14	0.55	0.59	0.76	0.11	0.24	0.05	0.17	-0.25	1.00
Road_Local_1	0.44	0.14	0.42	0.44	0.37	0.55	-0.01	0.03	0.05	-0.21	-0.15
Road_Minor_Art_1	0.14	0.19	0.15	0.14	0.08	0.12	0.09	0.00	0.02	-0.11	-0.03
Road_Major_Art_1	0.19	0.37	0.24	0.19	-0.24	-0.10	0.06	0.23	0.10	-0.11	-0.19
Road_Walkway_1	0.10	-0.08	0.12	0.10	-0.23	-0.14	0.08	0.08	0.06	-0.14	-0.17
Road_Other_1	0.02	0.13	0.05	0.02	-0.09	-0.04	0.06	0.03	0.08	-0.03	-0.07
SWLK_None_1	0.08	0.14	0.12	0.08	-0.12	-0.08	0.03	0.00	0.05	-0.06	-0.08
SWLK_BothSides_1	0.90	0.42	0.87	0.90	0.81	0.45	0.25	0.17	0.22	-0.43	0.64
SWLK_OneSide_1	0.01	0.15	0.04	0.01	-0.09	0.00	0.09	0.06	0.08	0.01	-0.07
SWLK_Walkway_1	0.08	-0.12	0.10	0.08	-0.22	-0.13	0.01	-0.01	-0.02	-0.07	-0.15
SWLK_Other_1	0.07	0.17	0.11	0.07	-0.12	-0.06	0.14	0.20	0.18	-0.12	-0.12
Land_LowRes_1	0.41	0.25	0.40	0.41	0.12	0.24	0.06	0.11	0.05	-0.17	0.08
Land_MedRes_1	0.00	-0.01	0.00	0.00	-0.04	0.03	-0.01	-0.01	-0.01	-0.02	-0.03
Land_HighRes_1	0.29	0.18	0.29	0.29	0.32	0.16	0.11	0.03	0.03	-0.10	0.18
Land_Commercial_1	0.33	0.14	0.32	0.33	0.43	0.06	0.00	-0.01	0.01	-0.15	0.38
Land_Industrial_1	0.09	0.03	0.08	0.09	0.11	0.04	-0.03	-0.05	-0.06	-0.03	0.15
Land_Office_1	0.31	0.25	0.32	0.31	0.52	0.00	0.13	-0.04	0.00	-0.17	0.44
Land_Mixed_1	0.27	0.09	0.25	0.27	0.27	0.12	0.08	0.05	0.08	-0.18	0.22
Land_Park_1	0.14	0.01	0.14	0.14	-0.01	0.04	0.11	0.12	0.14	-0.11	0.04
Land_Institutional_1	0.31	0.18	0.31	0.31	0.28	0.21	-0.01	-0.03	-0.02	-0.17	0.18
Land_Walkway_1	0.38	0.07	0.36	0.38	0.33	0.19	0.25	0.11	0.18	-0.20	0.27
Land_LowRes_Percent_1	-0.10	0.02	-0.10	-0.10	-0.22	-0.02	-0.08	-0.01	-0.07	0.13	-0.17
Land_MedRes_Percent_1	0.00	-0.01	0.00	0.00	-0.04	0.03	-0.01	-0.01	-0.01	-0.02	-0.03
Land_HighRes_Percent_1	0.02	0.04	0.02	0.02	0.12	0.05	0.05	0.00	0.01	0.07	0.02
Land_Commercial_Percent_1	-0.01	0.01	-0.02	-0.01	0.18	-0.03	-0.08	-0.06	-0.07	0.01	0.16
Land_Industrial_Percent_1	-0.07	-0.02	-0.07	-0.07	-0.03	-0.01	-0.04	-0.05	-0.07	0.06	0.01
Land_Office_Percent_1	0.05	0.14	0.06	0.05	0.29	-0.07	0.06	-0.06	-0.04	-0.05	0.22
Land_Mixed_Percent_1	0.01	0.00	0.00	0.01	0.08	0.04	0.03	0.01	0.01	-0.05	0.08
Land_Park_Percent_1	-0.01	-0.08	-0.02	-0.01	-0.11	-0.02	0.09	0.08	0.10	-0.08	-0.04
Land_Institutional_Percent_1	0.08	0.09	0.08	0.08	0.10	0.13	-0.05	-0.05	-0.05	-0.03	0.03
Land_Walkway_Percent_1	0.06	-0.07	0.05	0.06	0.10	0.03	0.08	0.01	0.04	-0.07	0.06
Land_Through_Park_1	0.28	0.09	0.28	0.28	0.22	0.18	0.13	0.08	0.13	-0.24	0.11
Land_Through_Mixed_1	0.06	0.09	0.09	0.06	-0.09	-0.14	0.05	0.13	0.08	-0.14	-0.06
Avg_Walkscore_1	0.07	0.18	0.07	0.07	0.25	0.12	0.04	0.00	0.01	-0.08	0.27
Walkscore_Length1_1	0.91	0.46	0.89	0.91	0.72	0.43	0.29	0.21	0.26	-0.48	0.61
Walkscore_Length2_1	0.63	0.36	0.63	0.63	0.67	0.19	0.19	0.14	0.16	-0.32	0.63
Walkscore_Length3_1	0.06	-0.12	0.06	0.06	-0.07	-0.11	-0.04	-0.03	-0.03	0.01	-0.14
Walkscore_Length4_1	0.08	0.00	0.08	0.08	-0.05	0.04	0.16	0.04	0.06	0.04	-0.03

	Road_Local_1	Road_Minor_Art_1	Road_Major_Art_1	Road_Walkway_1	Road_Other_1	SWLK_None_1	SWLK_BothSides_1	SWLK_OneSide_1	SWLK_Walkway_1	SWLK_Oth
Total_Length_1										
Total_Turn_1										
Total_General_Cost_1										
Total_Time_1										
Total_Sig_Int_1										
Ped_Crossover_1										
Max_Slope_1										
Steep_Slope1_1										
Steep_Slope2_1										
Cumul_Prob_1										
Road_Collector_1										
Road_Local_1	1.00									
Road_Minor_Art_1	-0.07	1.00								
Road_Major_Art_1	-0.10	-0.13	1.00							
Road_Walkway_1	-0.11	-0.14	-0.02	1.00						
Road_Other_1	-0.02	-0.07	-0.07	0.09	1.00					
SWLK_None_1	-0.06	-0.01	0.07	0.27	0.45	1.00				
SWLK_BothSides_1	0.47	0.19	0.16	-0.31	-0.10	-0.16	1.00			
SWLK_OneSide_1	-0.02	-0.03	0.11	-0.02	0.36	0.12	-0.05	1.00		
SWLK_Walkway_1	-0.12	-0.15	0.00	0.90	0.06	0.29	-0.31	-0.01	1.00	
SWLK_Other_1	-0.02	-0.06	0.07	0.39	0.20	0.13	-0.10	0.00	0.05	1.00
Land_LowRes_1	0.25	-0.04	0.54	-0.19	-0.13	-0.11	0.47	-0.02	-0.15	-0.12
Land_MedRes_1	-0.01	-0.02	0.04	0.05	-0.01	-0.01	-0.02	-0.01	-0.01	0.14
Land_HighRes_1	0.19	0.18	-0.04	-0.09	0.02	0.05	0.31	0.01	-0.08	-0.06
Land_Commercial_1	0.12	0.00	-0.04	-0.15	-0.03	-0.03	0.38	-0.06	-0.12	-0.09
Land_Industrial_1	-0.04	0.05	0.00	-0.09	0.04	-0.03	0.12	-0.02	-0.07	-0.06
Land_Office_1	0.06	0.00	-0.14	-0.10	-0.05	-0.01	0.34	-0.04	-0.09	-0.07
Land_Mixed_1	0.15	0.01	0.02	-0.11	0.00	-0.02	0.30	0.01	-0.11	-0.03
Land_Park_1	0.05	-0.03	0.11	0.09	0.02	0.11	0.09	-0.02	0.10	0.02
Land_Institutional_1	0.26	-0.02	0.09	-0.15	-0.09	-0.11	0.37	-0.02	-0.13	-0.09
Land_Walkway_1	0.13	0.21	-0.13	0.12	0.00	0.14	0.30	-0.05	0.12	-0.01
Land_LowRes_Percent_1	-0.02	-0.13	0.44	-0.24	-0.16	-0.13	0.01	-0.05	-0.19	-0.14
Land_MedRes_Percent_1	-0.01	-0.02	0.04	0.05	-0.01	-0.01	-0.02	-0.01	-0.01	0.14
Land_HighRes_Percent_1	0.05	0.16	-0.09	-0.10	0.01	0.02	0.06	-0.02	-0.09	-0.08
Land_Commercial_Percent_1	-0.03	-0.05	-0.11	-0.16	-0.03	-0.07	0.06	-0.07	-0.14	-0.10
Land_Industrial_Percent_1	-0.07	0.04	-0.03	-0.10	0.06	-0.04	-0.03	-0.02	-0.08	-0.06
Land_Office_Percent_1	0.01	-0.05	-0.15	-0.11	-0.06	-0.04	0.10	-0.05	-0.10	-0.08
Land_Mixed_Percent_1	0.02	0.00	-0.06	-0.11	0.00	-0.01	0.05	0.01	-0.10	-0.06
Land_Park_Percent_1	-0.03	-0.08	0.05	0.10	0.04	0.12	-0.06	-0.03	0.10	0.03
Land_Institutional_Percent_1	0.16	-0.05	0.05	-0.15	-0.07	-0.10	0.15	-0.03	-0.13	-0.09
Land_Walkway_Percent_1	0.01	0.16	-0.18	0.11	0.02	0.17	0.01	-0.06	0.12	-0.02
Land_Through_Park_1	0.19	0.10	-0.18	0.25	0.10	0.16	0.15	0.03	0.20	0.12
Land_Through_Mixed_1	-0.09	-0.06	-0.05	0.50	0.29	0.21	-0.15	0.17	0.32	0.45
Avg_Walkscore_1	-0.03	0.10	-0.09	-0.38	-0.02	-0.24	0.25	-0.05	-0.43	-0.01
Walkscore_Length1_1	0.39	0.14	0.18	-0.09	0.01	-0.08	0.91	0.01	-0.14	0.07
Walkscore_Length2_1	0.12	0.17	0.00	-0.13	-0.09	-0.08	0.67	-0.06	-0.16	0.02
Walkscore_Length3_1	0.05	-0.03	0.01	0.42	0.03	0.35	-0.16	0.00	0.52	0.00
Walkscore_Length4_1	0.02	0.06	-0.04	0.21	0.12	0.13	-0.02	-0.01	0.18	0.13

	Land_LowRes_1	Land_MedRes_1	Land_HighRes_1	Land_Commercial_1	Land_Industrial_1	Land_Office_1	Land_Mixed_1	Land_Park_1	Land_Institutional_1	Land_Walkway_1	Land_LowRes_Percent_1	Land_MedRes_Percent_1	Land_HighRes_Percent_1	Land_Commercial_Percent_1	Land_Industrial_Percent_1	Land_Office_Percent_1	Land_Mixed_Percent_1	Land_Park_Percent_1	Land_Institutional_Percent_1	Land_Walkway_Percent_1	Land_Through_Park_1	Land_Through_Mixed_1	Avg_Walkscore_1	Walkscore_Length1_1	Walkscore_Length2_1	Walkscore_Length3_1	Walk
Total_Length_1																											
Total_Turn_1																											
Total_General_Cost_1																											
Total_Time_1																											
Total_Sig_Int_1																											
Ped_Crossover_1																											
Max_Slope_1																											
Steep_Slope1_1																											
Steep_Slope2_1																											
Cumul_Prob_1																											
Road_Collector_1																											
Road_Local_1																											
Road_Minor_Art_1																											
Road_Major_Art_1																											
Road_Walkway_1																											
Road_Other_1																											
SWLK_None_1																											
SWLK_BothSides_1																											
SWLK_OneSide_1																											
SWLK_Walkway_1																											
SWLK_Other_1																											
Land_LowRes_1	1.00																										
Land_MedRes_1	-0.02	1.00																									
Land_HighRes_1	0.11	0.03	1.00																								
Land_Commercial_1	0.10	-0.02	0.51	1.00																							
Land_Industrial_1	0.02	-0.01	0.23	0.28	1.00																						
Land_Office_1	-0.04	-0.01	0.27	0.29	0.05	1.00																					
Land_Mixed_1	0.07	-0.02	0.17	0.25	0.21	0.03	1.00																				
Land_Park_1	0.08	0.08	0.00	0.08	-0.03	-0.05	0.15	1.00																			
Land_Institutional_1	0.23	-0.02	0.02	0.09	0.02	0.00	0.13	-0.04	1.00																		
Land_Walkway_1	-0.14	0.01	-0.03	-0.09	-0.12	0.20	-0.05	0.09	-0.02	1.00																	
Land_LowRes_Percent_1	0.78	-0.02	-0.06	-0.03	-0.03	-0.15	-0.09	0.00	0.06	-0.31	1.00																
Land_MedRes_Percent_1	-0.02	1.00	0.03	-0.02	-0.01	-0.01	-0.02	0.08	-0.02	0.01	-0.02	1.00															
Land_HighRes_Percent_1	-0.03	0.03	0.84	0.34	0.17	0.15	0.11	-0.05	-0.02	-0.09	-0.06	0.03	1.00														
Land_Commercial_Percent_1	-0.01	-0.02	0.34	0.80	0.23	0.15	0.12	-0.02	-0.02	-0.19	-0.01	-0.02	0.35	1.00													
Land_Industrial_Percent_1	-0.05	-0.01	0.13	0.16	0.87	-0.03	0.14	-0.05	-0.02	-0.14	-0.02	-0.01	0.18	0.22	1.00												
Land_Office_Percent_1	-0.11	-0.01	0.17	0.17	-0.01	0.84	0.00	-0.08	-0.06	0.09	-0.10	-0.01	0.15	0.13	-0.05	1.00											
Land_Mixed_Percent_1	-0.05	-0.02	0.09	0.12	0.18	-0.01	0.84	0.08	0.05	-0.10	-0.09	-0.02	0.10	0.13	0.19	0.01	1.00										
Land_Park_Percent_1	0.01	0.07	-0.05	-0.02	-0.04	-0.08	0.07	0.84	-0.08	0.04	-0.02	0.07	-0.07	-0.06	-0.05	-0.09	0.08	1.00									
Land_Institutional_Percent_1	0.13	-0.02	-0.01	0.00	0.01	-0.05	0.08	-0.07	0.88	-0.09	0.08	-0.02	0.02	-0.03	0.02	-0.05	0.06	-0.09	1.00								
Land_Walkway_Percent_1	-0.26	0.01	-0.09	-0.17	-0.14	0.09	-0.11	0.06	-0.10	0.83	-0.31	0.01	-0.09	-0.22	-0.14	0.09	-0.11	0.08	-0.11	1.00							
Land_Through_Park_1	-0.20	0.03	-0.03	-0.16	-0.18	0.12	-0.02	0.04	0.06	0.40	-0.41	0.03	-0.12	-0.28	-0.25	0.07	-0.09	0.06	-0.01	0.41	1.00						
Land_Through_Mixed_1	-0.21	0.09	-0.09	-0.12	-0.06	-0.01	-0.04	0.03	-0.08	0.06	-0.26	0.09	-0.12	-0.13	-0.09	-0.07	-0.07	0.04	-0.08	0.07	0.26	1.00					
Avg_Walkscore_1	0.00	0.01	0.04	0.09	0.05	0.21	0.10	-0.18	0.16	-0.02	0.01	0.01	0.03	0.09	0.04	0.19	0.11	-0.20	0.17	-0.09	-0.02	-0.13	1.00				
Walkscore_Length1_1	0.41	0.00	0.27	0.33	0.11	0.31	0.28	0.05	0.36	0.27	-0.05	0.00	0.03	0.03	-0.03	0.08	0.04	-0.08	0.14	-0.03	0.19	0.00	0.41	1.00			
Walkscore_Length2_1	0.19	0.03	0.30	0.30	0.10	0.47	0.18	-0.04	0.23	0.28	-0.07	0.03	0.11	0.07	-0.03	0.24	0.03	-0.13	0.09	0.06	0.22	0.00	0.50	0.70	1.00		
Walkscore_Length3_1	-0.06	-0.01	0.02	-0.04	-0.06	-0.06	-0.06	0.19	-0.16	0.19	-0.10	-0.01	-0.02	-0.09	-0.09	-0.08	-0.08	0.16	-0.16	0.20	0.16	0.13	-0.81	-0.37	-0.25	1.00	
Walkscore_Length4_1	-0.03	0.00	0.03	-0.04	-0.03	-0.04	-0.03	0.01	-0.04	0.02	-0.04	0.00	0.07	-0.04	-0.03	-0.04	-0.03	-0.01	-0.04	0.01	0.01	0.06	-0.27	-0.02	-0.07	0.21	1.00

Appendix D

Model Estimation Parameters

Model	Rho	Adj Rho
001-Oct20	0.398	0.396

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0114	0.00113	-10.09	0	0.00129	-8.82	0
B_TURN	-0.9	0.0465	-19.37	0	0.0587	-15.34	0
B_EPS3	1.6	0.207	7.74	0	0.261	6.15	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho
002-Oct20	0.454	0.452

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0197	0.0015	-13.1	0	0.00211	-9.33	0
B_TURN	-0.727	0.0496	-14.66	0	0.0698	-10.42	0
B_SIDEWALK_BOTHSIDES	0.00841	0.000799	10.53	0	0.00144	5.84	0
B_EPS3	1.54	0.214	7.2	0	0.276	5.6	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
003-Oct20 (General Model)	0.472	0.469					
Name	**Value**	**Std err**	**t-test**	**p-val**	**Rob. std err**	**Rob. t-test**	**Rob. p-val**
B_LENGTH	-0.02	0.00155	-12.93	0	0.00223	-8.99	0
B_TURN	-0.645	0.0515	-12.52	0	0.0707	-9.11	0
B_SIDEWALK_BOTHSIDES	0.00665	0.000832	7.99	0	0.00146	4.57	0
B_SIG_INT	0.669	0.0956	6.99	0	0.115	5.8	0
B_EPS3	1.53	0.217	7.06	0	0.278	5.52	0
B_SAMPLE_CORRECTION	1	--fixed--					

Coeff1	Coeff2	Covariance	Correlation	t-test
B_EPS3	B_SIG_INT	0.000173	0.00834	3.66
B_EPS3	B_SIDEWALK_BOTHSIDES	-7.78E-06	-0.043	7.03
B_EPS3	B_LENGTH	0.000101	0.3	7.17
B_SIDEWALK_BOTHSIDES	B_SIG_INT	-1.93E-05	-0.243	-6.91
B_LENGTH	B_SIG_INT	-1.47E-05	-0.0996	-7.19
B_EPS3	B_TURN	0.00163	0.146	10.09
B_LENGTH	B_SIDEWALK_BOTHSIDES	-6.67E-07	-0.518	-12.68
B_LENGTH	B_TURN	-2.01E-05	-0.252	12.03
B_SIDEWALK_BOTHSIDES	B_TURN	7.87E-06	0.184	12.68
B_SIG_INT	B_TURN	0.000898	0.182	13.13

Model	Rho	Adj Rho					
004-Oct20	0.474	0.47					
Name	**Value**	**Std err**	**t-test**	**p-val**	**Rob. std err**	**Rob. t-test**	**Rob. p-val**
B_LENGTH	-0.0185	0.0016	-11.58	0	0.00221	-8.37	0
B_LENGTH_FEMALE	-0.00688	0.00307	-2.24	0.03	0.0042	-1.64	0.1
B_TURN	-0.657	0.0518	-12.69	0	0.0706	-9.31	0
B_SIDEWALK_BOTHSIDES	0.00672	0.00083	8.1	0	0.00142	4.72	0
B_SIG_INT	0.653	0.0957	6.83	0	0.115	5.7	0
B_EPS3	1.49	0.218	6.85	0	0.279	5.35	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
005-Oct20	0.481	0.476					
Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0188	0.00163	-11.52	0	0.00232	-8.12	0
B_LENGTH_FEMALE	-0.00666	0.0031	-2.15	0.03	0.00429	-1.55	0.12
B_TURN	-0.674	0.053	-12.73	0	0.0723	-9.32	0
B_SIDEWALK_BOTHSIDES	0.00694	0.000837	8.29	0	0.00146	4.75	0
B_SIG_INT	0.577	0.0987	5.85	0	0.119	4.87	0
B_MINOR_ART_STUDENT	-0.00276	0.000659	-4.19	0	0.000811	-3.4	0
B_EPS3	1.41	0.221	6.39	0	0.287	4.91	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
006-Oct20	0.486	0.481					
Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0194	0.00166	-11.7	0	0.00243	-7.97	0
B_LENGTH_FEMALE	-0.00679	0.0031	-2.19	0.03	0.00427	-1.59	0.11
B_TURN	-0.685	0.0532	-12.87	0	0.073	-9.38	0
B_SIDEWALK_BOTHSIDES	0.00712	0.000839	8.48	0	0.00147	4.83	0
B_SIG_INT	0.696	0.105	6.62	0	0.128	5.42	0
B_MINOR_ART_STUDENT	-0.00307	0.000666	-4.61	0	0.000829	-3.7	0
B_ROAD_ART_AGE12	0.00347	0.00079	4.39	0	0.00118	2.94	0
B_EPS3	1.45	0.221	6.55	0	0.284	5.12	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
007-Oct20	0.489	0.483					
Name	**Value**	**Std err**	**t-test**	**p-val**	**Rob. std err**	**Rob. t-test**	**Rob. p-val**
B_LENGTH	-0.0195	0.00165	-11.87	0	0.0024	-8.13	0
B_LENGTH_FEMALE	-0.00677	0.00309	-2.19	0.03	0.00426	-1.59	0.11
B_TURN	-0.697	0.0536	-13.01	0	0.0734	-9.49	0
B_SIDEWALK_BOTHSIDES	0.00703	0.000829	8.48	0	0.00143	4.92	0
B_SIG_INT	0.703	0.104	6.78	0	0.124	5.67	0
B_MINOR_ART_STUDENT	-0.0032	0.000672	-4.76	0	0.000844	-3.79	0
B_ROAD_ART_AGE12	0.00354	0.000792	4.48	0	0.00118	3	0
B_MINOR_ART_INCOME	0.00334	0.00102	3.26	0	0.00114	2.94	0
B_EPS3	1.48	0.221	6.7	0	0.281	5.28	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
008-Oct20	0.492	0.485					
Name	**Value**	**Std err**	**t-test**	**p-val**	**Rob. std err**	**Rob. t-test**	**Rob. p-val**
B_LENGTH	-0.0198	0.00167	-11.84	0	0.00247	-8.01	0
B_LENGTH_FEMALE	-0.00693	0.00304	-2.28	0.02	0.00406	-1.71	0.09
B_TURN	-0.706	0.0542	-13.02	0	0.0746	-9.46	0
B_SIDEWALK_BOTHSIDES	0.00712	0.000838	8.49	0	0.00146	4.88	0
B_SIG_INT	0.705	0.104	6.77	0	0.125	5.63	0
B_MINOR_ART_STUDENT	-0.0032	0.000679	-4.71	0	0.000854	-3.75	0
B_ROAD_ART_AGE12	0.00361	0.000792	4.56	0	0.00119	3.05	0
B_MINOR_ART_INCOME	0.00348	0.00103	3.39	0	0.00114	3.06	0
B_LAND_PARK_EVENING	-0.0219	0.00756	-2.89	0	0.00574	-3.81	0
B_EPS3	1.44	0.221	6.53	0	0.278	5.2	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho
009-Oct20	0.495	0.488

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0199	0.00166	-11.97	0	0.00247	-8.06	0
B_LENGTH_FEMALE	-0.00683	0.003	-2.27	0.02	0.004	-1.71	0.09
B_TURN	-0.7	0.0544	-12.86	0	0.0758	-9.23	0
B_SIDEWALK_BOTHSIDES	0.00699	0.000835	8.36	0	0.00146	4.8	0
B_SIG_INT	0.737	0.106	6.98	0	0.126	5.85	0
B_MINOR_ART_STUDENT	-0.00316	0.000679	-4.66	0	0.000855	-3.7	0
B_ROAD_ART_AGE12	0.00367	0.000792	4.63	0	0.00118	3.11	0
B_MINOR_ART_INCOME	0.00336	0.00101	3.32	0	0.00112	3.01	0
B_LAND_PARK_EVENING	-0.0238	0.00717	-3.32	0	0.00607	-3.91	0
B_LAND_PARK_PERCENT_INCOME	24.7	9.32	2.65	0.01	13	1.9	0.06
B_EPS3	1.45	0.223	6.5	0	0.283	5.12	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho
010-Oct20	0.498	0.49

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0198	0.00164	-12.07	0	0.00242	-8.19	0
B_LENGTH_FEMALE	-0.00699	0.00298	-2.35	0.02	0.00392	-1.79	0.07
B_TURN	-0.706	0.0546	-12.93	0	0.0761	-9.28	0
B_SIDEWALK_BOTHSIDES	0.00709	0.000848	8.36	0	0.00149	4.77	0
B_SIG_INT	0.759	0.105	7.25	0	0.124	6.12	0
B_MINOR_ART_STUDENT	-0.00317	0.000682	-4.64	0	0.000861	-3.68	0
B_ROAD_ART_AGE12	0.0038	0.000794	4.78	0	0.00118	3.21	0
B_MINOR_ART_INCOME	0.00315	0.00102	3.08	0	0.00115	2.75	0.01
B_LAND_PARK_EVENING	-0.0235	0.00731	-3.21	0	0.00638	-3.68	0
B_LAND_PARK_PERCENT_INCOME	26.3	10	2.63	0.01	16.4	1.6	0.11
B_LAND_COMMERCIAL_EMP	-0.00905	0.00347	-2.61	0.01	0.00404	-2.24	0.03
B_EPS3	1.45	0.222	6.56	0	0.278	5.24	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho
011-Oct20	0.5	0.591

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0198	0.00164	-12.05	0	0.00242	-8.19	0
B_LENGTH_FEMALE	-0.0069	0.00297	-2.32	0.02	0.00389	-1.77	0.08
B_TURN	-0.716	0.0549	-13.03	0	0.077	-9.3	0
B_SIDEWALK_BOTHSIDES	0.00719	0.000837	8.58	0	0.00144	4.99	0
B_SIG_INT	0.772	0.105	7.34	0	0.125	6.19	0
B_MINOR_ART_STUDENT	-0.00317	0.000684	-4.63	0	0.000865	-3.66	0
B_ROAD_ART_AGE12	0.00387	0.000796	4.86	0	0.00118	3.27	0
B_MINOR_ART_INCOME	0.0028	0.00101	2.76	0.01	0.00107	2.62	0.01
B_LAND_PARK_EVENING	-0.0235	0.00729	-3.23	0	0.00635	-3.71	0
B_LAND_PARK_PERCENT_INCOME	27.1	10	2.7	0.01	16.6	1.64	0.1
B_LAND_COMMERCIAL_EMP	-0.00934	0.00348	-2.69	0.01	0.00407	-2.3	0.02
B_COLLECTOR_AGE56	-0.0033	0.0011	-3	0	0.000869	-3.79	0
B_EPS3	1.43	0.223	6.41	0	0.281	5.09	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho
012-Oct20	0.502	0.492

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
B_LENGTH	-0.0199	0.00165	-12.06	0	0.00244	-8.16	0
B_LENGTH_FEMALE	-0.00709	0.00301	-2.36	0.02	0.004	-1.77	0.08
B_TURN	-0.718	0.0554	-12.97	0	0.0781	-9.2	0
B_SIDEWALK_BOTHSIDES	0.00732	0.000845	8.66	0	0.00147	4.98	0
B_SIG_INT	0.776	0.106	7.33	0	0.126	6.16	0
B_MINOR_ART_STUDENT	-0.00322	0.000696	-4.62	0	0.0009	-3.57	0
B_ROAD_ART_AGE12	0.00354	0.000762	4.65	0	0.00108	3.27	0
B_MINOR_ART_INCOME	0.00278	0.00101	2.74	0.01	0.00107	2.6	0.01
B_LAND_PARK_EVENING	-0.0237	0.00734	-3.23	0	0.00647	-3.66	0
B_LAND_PARK_PERCENT_INCOME	27.2	10.1	2.7	0.01	16.7	1.63	0.1
B_LAND_COMMERCIAL_EMP	-0.00946	0.00349	-2.71	0.01	0.00411	-2.3	0.02
B_COLLECTOR_AGE56	-0.00334	0.0011	-3.03	0	0.000871	-3.83	0
B_LAND_OFFICE_AGE12	-0.0125	0.00578	-2.16	0.03	0.00563	-2.21	0.03
B_EPS3	1.4	0.224	6.24	0	0.285	4.91	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
013-Oct20	0.504	0.494					
Name	**Value**	**Std err**	**t-test**	**p-val**	**Rob. std err**	**Rob. t-test**	**Rob. p-val**
B_LENGTH	-0.0199	0.00166	-11.99	0	0.00245	-8.11	0
B_LENGTH_FEMALE	-0.00775	0.00306	-2.53	0.01	0.00411	-1.89	0.06
B_TURN	-0.718	0.0553	-13	0	0.0775	-9.27	0
B_SIDEWALK_BOTHSIDES	0.00719	0.00085	8.46	0	0.00148	4.85	0
B_SIG_INT	0.762	0.106	7.2	0	0.126	6.03	0
B_MINOR_ART_STUDENT	-0.00329	0.000695	-4.74	0	0.00091	-3.61	0
B_ROAD_ART_AGE12	0.00343	0.000754	4.55	0	0.00105	3.26	0
B_MINOR_ART_INCOME	0.0028	0.00102	2.75	0.01	0.00108	2.61	0.01
B_LAND_PARK_EVENING	-0.0235	0.00744	-3.16	0	0.0065	-3.62	0
B_LAND_PARK_PERCENT_INCOME	27	10	2.69	0.01	16.5	1.63	0.1
B_LAND_COMMERCIAL_EMP	-0.00931	0.0035	-2.66	0.01	0.00413	-2.25	0.02
B_COLLECTOR_AGE56	-0.00328	0.0011	-2.98	0	0.00087	-3.77	0
B_LAND_OFFICE_AGE12	-0.0143	0.00572	-2.5	0.01	0.0059	-2.42	0.02
B_LAND_WALKWAY_AGE12	0.0143	0.005	2.86	0	0.00649	2.21	0.03
B_EPS3	1.4	0.224	6.25	0	0.283	4.95	0
B_SAMPLE_CORRECTION	1	--fixed--					

Model	Rho	Adj Rho					
014-Oct20 (Interaction Model)	0.501	0.491					
Name	**Value**	**Std err**	**t-test**	**p-val**	**Rob. std err**	**Rob. t-test**	**Rob. p-val**
B_LENGTH	-0.0198	0.00168	-11.81	0	0.00248	-8.01	0
B_LENGTH_FEMALE	-0.00788	0.00309	-2.55	0.01	0.00417	-1.89	0.06
B_TURN	-0.724	0.0551	-13.13	0	0.0767	-9.44	0
B_SIDEWALK_BOTHSIDES	0.0073	0.000853	8.56	0	0.00149	4.91	0
B_SIG_INT	0.729	0.104	6.98	0	0.127	5.75	0
B_MINOR_ART_STUDENT	-0.00333	0.000695	-4.79	0	0.000912	-3.65	0
B_ROAD_ART_AGE12	0.00337	0.000754	4.47	0	0.00106	3.18	0
B_MINOR_ART_INCOME	0.0029	0.00103	2.81	0.01	0.00111	2.62	0.01
B_LAND_PARK_EVENING	-0.0214	0.00789	-2.72	0.01	0.00619	-3.46	0
B_LAND_COMMERCIAL_EMP	-0.00911	0.00349	-2.61	0.01	0.00408	-2.23	0.03
B_COLLECTOR_AGE56	-0.00319	0.00111	-2.87	0	0.000894	-3.57	0
B_LAND_OFFICE_AGE12	-0.0143	0.00574	-2.5	0.01	0.00596	-2.41	0.02
B_LAND_WALKWAY_AGE12	0.0145	0.00502	2.88	0	0.00649	2.23	0.03
B_EPS3	1.39	0.222	6.26	0	0.278	5.01	0
B_SAMPLE_CORRECTION	1	--fixed--					

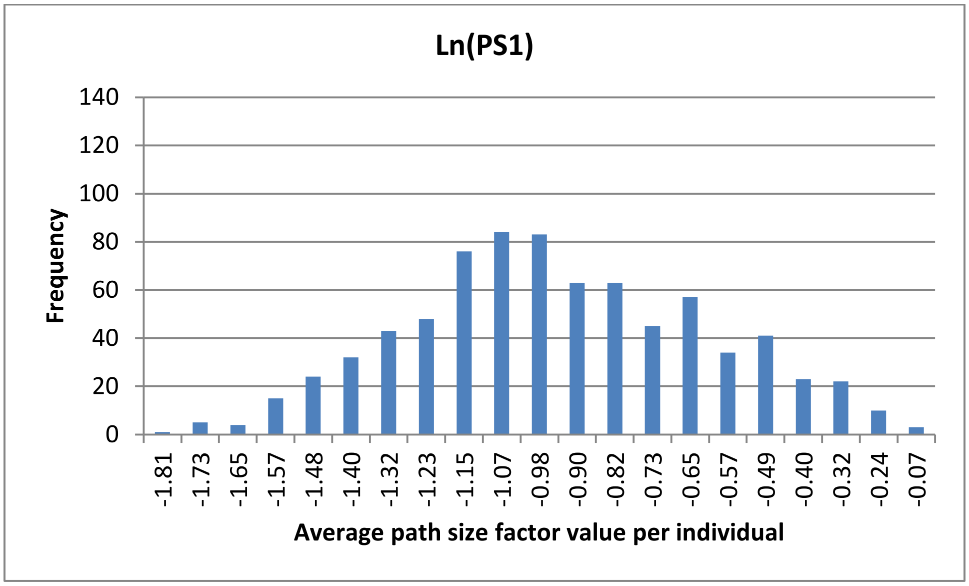
Coeff1	Coeff2	Covariance	Correlation	t-test
B_COLLECTOR_AGE56	B_MINOR_ART_STUDENT	3.28E-09	0.00424	0.11
B_LAND_COMMERCIAL_EMP	B_LENGTH_FEMALE	4.21E-07	0.0391	-0.27
B_LAND_PARK_EVENING	B_LENGTH	4.35E-07	0.0328	-0.20
B_MINOR_ART_INCOME	B_ROAD_ART_AGE12	3.37E-08	0.0433	-0.37
B_LAND_COMMERCIAL_EMP	B_LAND_OFFICE_AGE12	3.98E-07	0.0199	0.79
B_LAND_OFFICE_AGE12	B_LAND_PARK_EVENING	5.79E-07	0.0128	0.73
B_LAND_OFFICE_AGE12	B_LENGTH	6.27E-07	0.065	0.94
B_LAND_OFFICE_AGE12	B_LENGTH_FEMALE	1.37E-06	0.0773	-1.02
B_LENGTH_FEMALE	B_MINOR_ART_STUDENT	-6.10E-08	-0.0284	-1.43
B_COLLECTOR_AGE56	B_LENGTH_FEMALE	2.18E-08	0.00633	1.43
B_LAND_WALKWAY_AGE12	B_SIDEWALK_BOTHSIDES	-6.36E-08	-0.0149	1.40
B_LAND_COMMERCIAL_EMP	B_MINOR_ART_STUDENT	8.28E-09	0.00342	-1.62
B_COLLECTOR_AGE56	B_LAND_COMMERCIAL_EMP	1.62E-07	0.0417	1.63
B_LAND_WALKWAY_AGE12	B_ROAD_ART_AGE12	-4.77E-07	-0.126	2.15
B_LAND_COMMERCIAL_EMP	B_LAND_PARK_EVENING	-7.22E-07	-0.0262	1.41
B_LAND_WALKWAY_AGE12	B_MINOR_ART_INCOME	5.56E-08	0.0107	2.26
B_COLLECTOR_AGE56	B_LAND_OFFICE_AGE12	1.37E-07	0.0215	1.91
B_LAND_OFFICE_AGE12	B_MINOR_ART_STUDENT	9.88E-08	0.0248	-1.91
B_LAND_PARK_EVENING	B_LENGTH_FEMALE	2.06E-06	0.0846	-1.65
B_EPS3	B_SIG_INT	0.000444	0.0191	2.72
B_MINOR_ART_INCOME	B_SIDEWALK_BOTHSIDES	-5.49E-08	-0.0622	-3.18
B_LENGTH	B_LENGTH_FEMALE	-1.54E-06	-0.296	-3.04
B_LAND_COMMERCIAL_EMP	B_LENGTH	1.07E-07	0.0183	2.79
B_ROAD_ART_AGE12	B_SIDEWALK_BOTHSIDES	7.00E-08	0.109	-3.66
B_LENGTH_FEMALE	B_MINOR_ART_INCOME	1.08E-07	0.0339	-3.34
B_LENGTH_FEMALE	B_ROAD_ART_AGE12	1.82E-08	0.00782	-3.54
B_LAND_WALKWAY_AGE12	B_MINOR_ART_STUDENT	-3.30E-07	-0.0947	3.47
B_COLLECTOR_AGE56	B_LAND_WALKWAY_AGE12	9.74E-08	0.0175	-3.45
B_LAND_WALKWAY_AGE12	B_LENGTH_FEMALE	-1.68E-06	-0.108	3.62
B_LAND_OFFICE_AGE12	B_LAND_WALKWAY_AGE12	-5.10E-06	-0.177	-3.48
B_LAND_OFFICE_AGE12	B_MINOR_ART_INCOME	4.26E-08	0.00717	-2.96
B_LAND_COMMERCIAL_EMP	B_ROAD_ART_AGE12	-1.92E-07	-0.0732	-3.44
B_LAND_PARK_EVENING	B_MINOR_ART_STUDENT	1.07E-07	0.0195	-2.29
B_COLLECTOR_AGE56	B_LAND_PARK_EVENING	8.00E-08	0.00911	2.29
B_LAND_COMMERCIAL_EMP	B_MINOR_ART_INCOME	3.25E-07	0.0902	-3.39
B_LAND_OFFICE_AGE12	B_ROAD_ART_AGE12	1.18E-06	0.272	-3.17
B_LAND_COMMERCIAL_EMP	B_LAND_WALKWAY_AGE12	1.04E-07	0.00592	-3.87
B_LENGTH_FEMALE	B_SIDEWALK_BOTHSIDES	-2.24E-07	-0.0848	-4.64
B_LAND_OFFICE_AGE12	B_SIDEWALK_BOTHSIDES	-5.70E-07	-0.116	-3.67
B_LAND_COMMERCIAL_EMP	B_SIDEWALK_BOTHSIDES	-1.12E-07	-0.0378	-4.53
B_LAND_PARK_EVENING	B_MINOR_ART_INCOME	-2.55E-07	-0.0313	-3.04

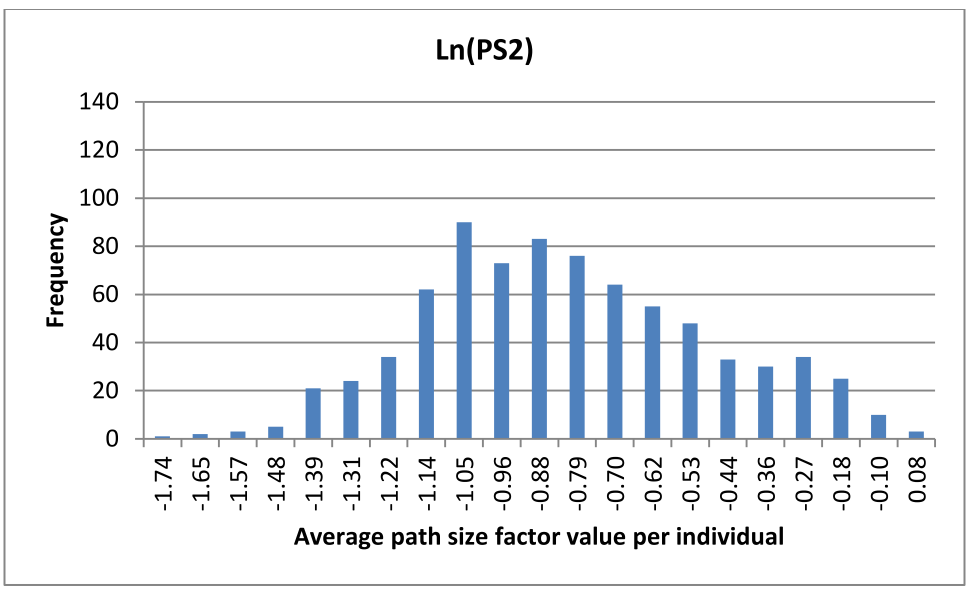
B_LAND_PARK_EVENING	B_ROAD_ART_AGE12	-1.81E-07	-0.0304	-3.12
B_LAND_PARK_EVENING	B_LAND_WALKWAY_AGE12	1.77E-07	0.00447	-3.85
B_MINOR_ART_INCOME	B_MINOR_ART_STUDENT	-4.53E-08	-0.063	4.86
B_MINOR_ART_STUDENT	B_ROAD_ART_AGE12	-1.04E-07	-0.198	-5.97
B_LAND_PARK_EVENING	B_SIDEWALK_BOTHSIDES	-1.49E-07	-0.0221	-3.61
B_COLLECTOR_AGE56	B_ROAD_ART_AGE12	-5.84E-08	-0.0696	-4.73
B_COLLECTOR_AGE56	B_MINOR_ART_INCOME	1.11E-07	0.0967	-4.22
B_LAND_WALKWAY_AGE12	B_LENGTH	-4.56E-08	-0.00541	6.47
B_EPS3	B_LAND_WALKWAY_AGE12	1.47E-05	0.0132	6.20
B_EPS3	B_SIDEWALK_BOTHSIDES	-1.24E-05	-0.0653	6.23
B_EPS3	B_ROAD_ART_AGE12	8.25E-06	0.0492	6.25
B_EPS3	B_MINOR_ART_INCOME	1.41E-05	0.0611	6.25
B_COLLECTOR_AGE56	B_EPS3	1.02E-05	0.0412	-6.28
B_EPS3	B_MINOR_ART_STUDENT	7.56E-06	0.0489	6.28
B_EPS3	B_LENGTH_FEMALE	5.98E-05	0.087	6.30
B_EPS3	B_LAND_COMMERCIAL_EMP	-2.90E-06	-0.00373	6.30
B_EPS3	B_LAND_OFFICE_AGE12	0.000105	0.0826	6.34
B_EPS3	B_LENGTH	9.19E-05	0.246	6.36
B_EPS3	B_LAND_PARK_EVENING	9.79E-05	0.0558	6.37
B_LAND_WALKWAY_AGE12	B_SIG_INT	-3.81E-05	-0.0727	-6.81
B_SIDEWALK_BOTHSIDES	B_SIG_INT	-1.71E-05	-0.192	-6.89
B_MINOR_ART_STUDENT	B_SIDEWALK_BOTHSIDES	-1.95E-08	-0.033	-9.51
B_MINOR_ART_INCOME	B_SIG_INT	7.78E-06	0.072	-6.95
B_ROAD_ART_AGE12	B_SIG_INT	2.97E-05	0.377	-6.96
B_COLLECTOR_AGE56	B_SIG_INT	-9.89E-06	-0.0851	-7.00
B_MINOR_ART_STUDENT	B_SIG_INT	2.09E-06	0.0287	-7.01
B_LAND_COMMERCIAL_EMP	B_SIG_INT	-4.27E-05	-0.117	-7.03
B_COLLECTOR_AGE56	B_SIDEWALK_BOTHSIDES	-8.31E-08	-0.0876	-7.19
B_LENGTH_FEMALE	B_SIG_INT	3.23E-05	0.0999	-7.07
B_LAND_OFFICE_AGE12	B_SIG_INT	-1.11E-05	-0.0186	-7.10
B_LENGTH	B_SIG_INT	-3.25E-05	-0.185	-7.14
B_LAND_PARK_EVENING	B_SIG_INT	-1.88E-05	-0.0228	-7.15
B_COLLECTOR_AGE56	B_LENGTH	2.76E-08	0.0148	8.32
B_LENGTH	B_MINOR_ART_STUDENT	8.81E-08	0.0755	-9.33
B_LENGTH	B_SIDEWALK_BOTHSIDES	-7.13E-07	-0.498	-12.17
B_EPS3	B_TURN	0.00201	0.164	9.61
B_LENGTH	B_ROAD_ART_AGE12	-1.83E-07	-0.144	-11.98
B_LENGTH	B_MINOR_ART_INCOME	-1.36E-07	-0.078	-11.15
B_LENGTH	B_TURN	-2.21E-05	-0.238	12.67
B_LAND_PARK_EVENING	B_TURN	3.18E-05	0.073	12.75
B_LAND_OFFICE_AGE12	B_TURN	9.74E-06	0.0308	12.84
B_LENGTH_FEMALE	B_TURN	1.87E-05	0.11	13.05

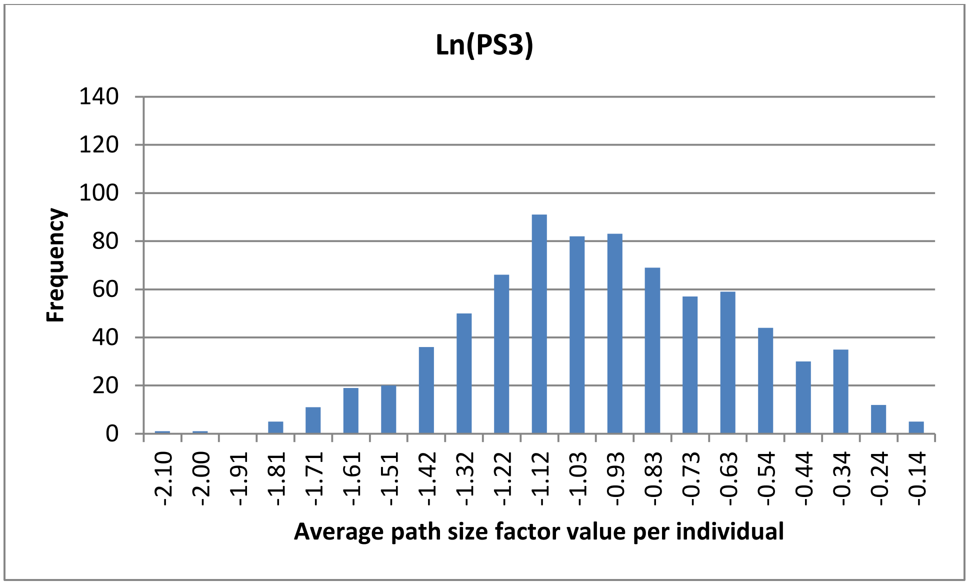
B_LAND_COMMERCIAL_EMP	B_TURN	1.08E-05	0.0561	12.99
B_MINOR_ART_STUDENT	B_TURN	3.96E-06	0.103	13.09
B_COLLECTOR_AGE56	B_TURN	5.73E-06	0.0933	13.10
B_MINOR_ART_INCOME	B_TURN	-2.40E-06	-0.042	13.17
B_ROAD_ART_AGE12	B_TURN	-3.46E-06	-0.0833	13.18
B_LAND_WALKWAY_AGE12	B_TURN	-3.95E-06	-0.0143	13.32
B_SIDEWALK_BOTHSIDES	B_TURN	6.18E-06	0.131	13.29
B_SIG_INT	B_TURN	0.000834	0.145	13.11

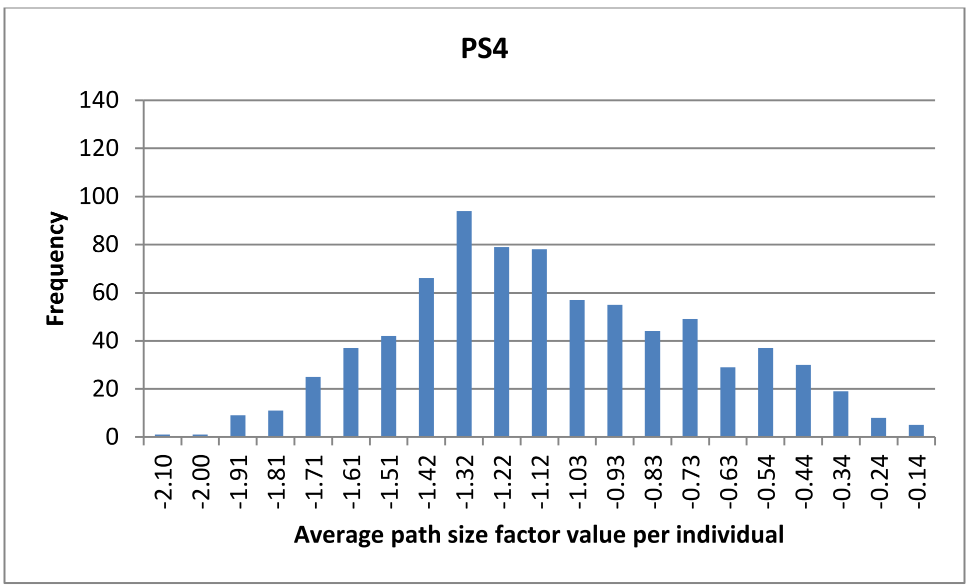
Appendix E

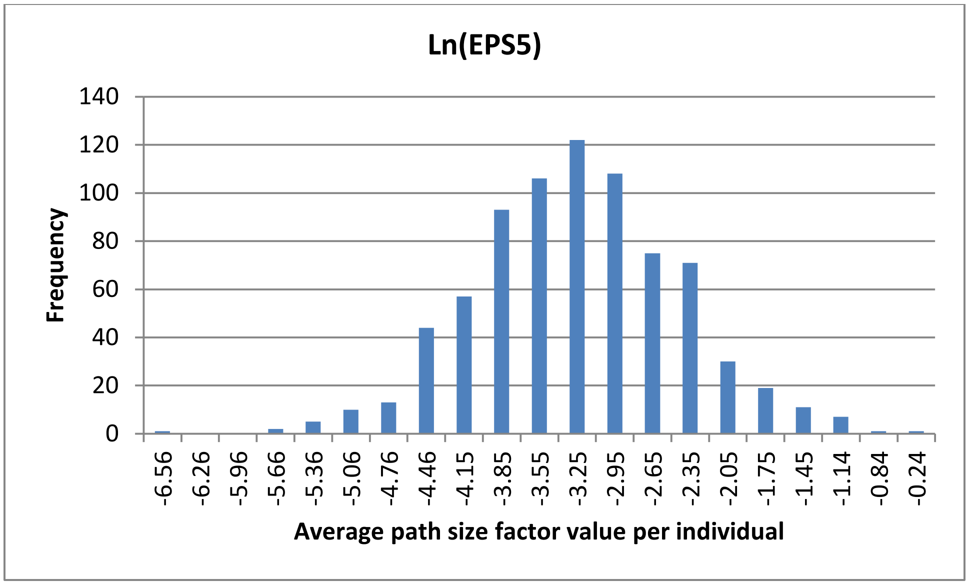
Distribution of Path Size Factor






Ln(PS3)
140
120
100
80
60
40
20
0
Frequency
-2.10
-2.00
-1.91
-1.81
-1.71
-1.61
-1.51
-1.42
-1.32
-1.22
-1.12
-1.03
-0.93
-0.83
-0.73
-0.63
-0.54
-0.44
-0.34
-0.24
-0.14
Average path size factor value per individual


PS4
140
120
100
80
60
40
20
0
Frequency
-2.10
-2.00
-1.91
-1.81
-1.71
-1.61
-1.51
-1.42
-1.32
-1.22
-1.12
-1.03
-0.93
-0.83
-0.73
-0.63
-0.54
-0.44
-0.34
-0.24
-0.14
Average path size factor value per individual



Ln(EPS5)
Frequency
140
120
100
80
60
40
20
0
-6.56
-6.26
-5.96
-5.66
-5.36
-5.06
-4.76
-4.46
-4.15
-3.85
-3.55
-3.25
-2.95
-2.65
-2.35
-2.05
-1.75
-1.45
-1.14
-0.84
-0.24
Average path size factor value per individual